Eindhoven University of Technology

Eindhoven University of Technology

MASTER

Contextual Anomaly Detection through Multivariate Monitoring of Regression Profiles

van Dalen, Okki

Award date:
2020

Link to publication

Technische Universiteit
Eindhoven
University of Technology

Department of Mathematics and Computer Science

Master's Thesis

Contextual Anomaly Detection through Multivariate Monitoring of Regression Profiles

Author:
Okki van Dalen
0960299

Supervisor:
Dr. Alessandro Di Bucchianico

Eindhoven, September 2020

Abstract

Statistical monitoring methods are currently widely used to control the quality of a process. Most real-life processes, e.g. wind turbines, require a multivariate approach so that multiple crucial quality characteristics can be monitored simultaneously. Many commonly used multivariate methods, like the Hotelling T^2 chart, rely on the multivariate normality assumption which is hardly ever satisfied for process data. In this thesis we therefore propose a monitoring approach based on copula theory which allows for efficient monitoring of non-normal multivariate data. Copulas allow to separate the multivariate dependence structure from the marginals. The dependence structure described by copulas captures dependence details that go beyond the covariance matrix. We perform a simulation study to compare the copula-based approach with conventional monitoring methods. The results show that the copula-based approach outperforms conventional methods in several out-of-control cases.

In light of the wind turbines, we additionally present a case study where we apply the copula-based approach and the Hotelling T^2 chart to a real-life data set of a wind turbine. In order to correctly interpret the results, we have to introduce the notion of contextual anomalies, which arise when changes in or extreme values from observed processes can only be interpreted correctly if other information is taken into account. These contextual anomalies can be identified by first applying the multivariate regression model and subsequently monitor the deviation of the model. The case study shows the promising potential of the copula-based approach for effective monitoring of the condition of the wind turbine, which contributes to reducing repair costs and downtime.

Keywords: condition based monitoring, statistical process control, linear regression, multivariate monitoring, copula modeling.

Preface

The following Master's thesis is part of my Master's Final Project at the department of Mathematics and Computer Science. It has been written to fulfill my Master of Science study at the Eindhoven University of Technology. I have been conducting research and analyzing data for this thesis from November 2019 to September 2020, including a three month break for an internship at prognostica GmbH in Würzburg, Germany.

I would like to thank everybody at prognostica for providing me the opportunity for the internship, especially Dr. Kristina Krebs for the excellent guiding and supervision. It has been a great experience and the internship contributed massively to the content of this thesis. Also, I would like to thank Prof. Dr. Rainer Göb and Andrew Easton for the extensive "Arbeitsgemeinschaft" sessions during and after the internship. Without them I would not have been able to obtain the results of this thesis.

Furthermore, I would particularly like to thank my supervisor Dr. Alessandro Di Bucchianico for his guidance and extensive support during this project and for providing me this project and the opportunity for the internship. Without his help and useful feedback I would not have been able to finish this project. The research was challenging at some times, but the regular meetings clarified uncertainties and gave me a step in the right direction.

Last but not least, I would like to thank my mother, my father, my sister and my lovely girlfriend for their unconditional love, support and understanding throughout this project. You kept me motivated if I ever lost interest. Finally, I would like to thank all my friends for keeping me sane and happy.

I hope you enjoy reading this thesis.

Okki van Dalen

September 2020

Contents

List of Figures

6

8

List of Tables

1 | Introduction

Data science is nowadays becoming increasingly popular, since it offers great career opportunities and it contributes to improving corporate and industrial processes. It is even declared as the "sexiest" job of the 21st century. Data science is enforced by large data sets containing extensive information which are easily and quickly accessible these days. The incoming flow of data is enormous and new data is generated every second. Most of these data sets are naturally multivariate, that is, variables of interest should be considered simultaneously. For instance, in the light of sensor data it is necessary to monitor multiple temperatures simultaneously to account for dependencies among the individual temperatures. Such dependencies could arise when two sensors are closely located to each other. Univariate analyses turn out to be invalid or inefficient as they neglect dependencies between the temperatures. Therefore, we need multivariate analyses to account for dependencies among the variables under investigation whenever independence cannot be assumed.

In this thesis we mainly focus on multivariate monitoring methods. Specifically, we investigate multivariate monitoring of regression profiles, which is inspired by the univariate analyses presented in Kenbeek (2016). These univariate analyses are about contextual anomaly detection through monitoring of regression profiles, including a specific industrial example by applying the theory to wind turbine data. Contextual anomalies arise when changes in or extreme values from observed processes can only be interpreted correctly if other information is taken into account. In other words, observed values are sometimes not abnormal by themselves, but this condition is dependent on the context. A concrete example are wind turbines where the interpretation of sensor values for temperatures and vibrations have to take into account environmental factors like ambient temperature and wind speed (see, e.g., Kenbeek et al. (2016)). To this purpose, we use a monitoring approach that is based on regression analysis. By means of appropriate regression models we can adjust observed values based on external factors, which enables us to detect contextual anomalies. In the univariate case, we can monitor the deviation of the regression model in the spirit of Brown et al. (1975), Chu et al. (1996) or Zeileis et al. (2001). First steps into this monitoring approach were taken in Kenbeek (2016), and follow-up studies were presented in Meeuwis (2017) and van Dalen (2018).

The monitoring techniques are based on several popular control charts that follow the Statistical Process Control (SPC) methodology, which originates from Shewhart (1925). SPC is being used to improve the quality of industrial and business processes by effectively detecting deviations from normal behavior, based on which optimal decisions can be made. The basic concepts of SPC and several commonly used (univariate) control charts, including derivation, properties and performance, are thoroughly documented in Qiu (2013, Chapter 1-5).

The main goal of this thesis is to extend the univariate monitoring approach to the multivariate case, where we need to define the multivariate regression model (see, e.g., Monahan (2008, Chapter 9)) and develop an effective multivariate monitoring method. Most conventional multivariate monitoring methods, such as the Hotelling T^2 control chart, are based on the assumptions that observed values are uncorrelated and multivariate normally distributed. Under these assumptions, several monitoring techniques as in Hawkins (1993) have become standard and been successful in a variety of applications. In fact, these techniques have become so popular that they are often applied without carefully checking the assumptions. However, the assumptions are hardly ever satisfied which can question the efficiency of these methods. Moreover, most conventional multivariate methods are able to timely detect changes in mean or variance of a multivariate distribution that is supposed to represent a particular process. Apart from changes in mean or variance, the distribution of the process can also change in dependence structure. Unfortunately,

most conventional multivariate methods fail to detect such changes in dependence structure in a timely manner. This is where the modern theory of copulas comes into play. The main appeal of copulas is that they allow to separate a multivariate distribution into an appropriate copula, which describes the dependence structure, and all the marginals (see, e.g., Rüschendorf (2009)). This provides a flexible framework in multivariate modeling.

In order to accomplish the objectives, this thesis is organized as follows. In Chapter 2 we explain the monitoring foundation and subsequently we properly define the multivariate multiple linear regression model including some interesting properties based on Monahan (2008, Chapter 9). In Chapter 3 we introduce the basic copula model and explain vine copula structures based on core literature Nelsen (2007) and Joe (2014). Furthermore, in Chapter 4 we briefly explain some conventional multivariate monitoring methods and introduce a modern copula-based monitoring approach that was developed by Mühlig (2017). We continue to analyze the performance of this copula-based method compared to some conventional methods in Chapter 5, where we make use of performance measures presented in Frisén (2007). In Chapter 6 we present the results of applying the copula-based method to wind turbine data in a case study. Finally, in Chapter 7 we discuss our conclusions and make some interesting recommendations for further research. In the end we should be able to answer the following research questions:

- How does the multivariate regression model relate to the univariate regression model and can we derive similar properties of estimators?

- How can we design a multivariate control chart based on a given vine copula and what are its main benefits?

- How can we fairly compare the performance of control charts to various changes in a process?

- How does the vine copula-based control chart perform compared to conventional multivariate monitoring methods?

2 | Regression model

In this chapter we explain how to combine the usual monitoring approach with regression profiles. Furthermore, we introduce the multivariate multiple linear regression model based on Monahan (2008, Chapter 9), following the notation style of van Dalen (2018). Many results follow the univariate case, however, some cases require additional remarks. We will cover the most important properties of the regression model, which includes parameter estimation and corresponding distribution, and the predictive residual vector.

2.1 Monitoring with regression

The usual monitoring approach follows the SPC methodology, which describes a set of statistical methods that is used to improve the quality of industrial and business processes by accurate and timely detection of deviations from normal behavior. The deviations should then subsequently be traced back to the initial root cause, based on which optimal decisions can be made in order to restore the quality of the process. Note that the term "control" in SPC may be misleading, since SPC is really about monitoring and not referring to feedback control. SPC originates from the time of upcoming industry, when Shewhart introduced the concept (see Shewhart (1925)). Originally, SPC was used to monitor a production line in a factory. Typically, small subgroups of observations at fixed time points are collected. Then, statistical methods are applied to the observations to monitor the process and to take action when the process deviates from its normal behavior. Such behavior is called normal if it is stable, predictable and only has natural random variations. When a process behaves normally, it is called in-control. If a process does not behave normally, the process is not functioning well, and it is called out-of-control.

The most widely known tools to detect deviations from normal behavior are Shewhart control charts. Such control charts are a way to visualize the deviation of a process. The deviation can then be compared to control limits by repeated hypothesis testing. The control limits represent boundaries for in-control behavior and can be determined from historical data. This historical data is assumed to represent an in-control process. The training process of calculating control limits based on historical data is referred to as phase I. Once phase I is finished, we can move on to the online monitoring phase, the so-called phase II. During this phase, real-time observations are coming in and we have to decide if these can be classified as in-control or out-of-control. We compare these incoming observations with the control limits determined in phase I in order to classify the observations. A control charts gives a warning signal, which indicates an out-of-control situation, when the test statistic exceeds or falls below one of the control limits.

To detect changes in the mean of single observations we usually apply the Shewhart chart for individual observations, which is exemplified in Figure 2.1. We denote the individual observations by X_i, $i = 1, 2, \ldots$. It is assumed that the observations are independent identically distributed (i.i.d.) from a normal distribution. The Shewhart chart for individual observations is basically a plot of the observations X_i versus time, including 3 horizontal lines that indicate the mean, upper control limit (UCL) and lower control limit (LCL). Assume that we have a quality characteristic with mean $\mathbb{E}[X] = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$. Then, as X_i are samples of the quality characteristic, we also have $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ for all i. The control limits are now usually set at $\mu \pm 3\sigma$, and the control chart signals an alarm if an observation falls outside the control limits. In case μ and σ are unknown, we can estimate these quantities from historical data (phase I). Usually, we take $\widehat{\mu} = \overline{X}$ and $\widehat{\sigma} = \overline{\mathrm{MR}}/1.128$, where $\overline{X}$ is the average of the observations and $\overline{\mathrm{MR}}$ is the average of the moving ranges $\mathrm{MR}_i = |X_i - X_{i-1}|$. For more information on the constant 1.128, see

Qiu (2013, Chapter 3). Note that we make use of the moving range, since the sample variance for subgroups of size 1 is undefined.
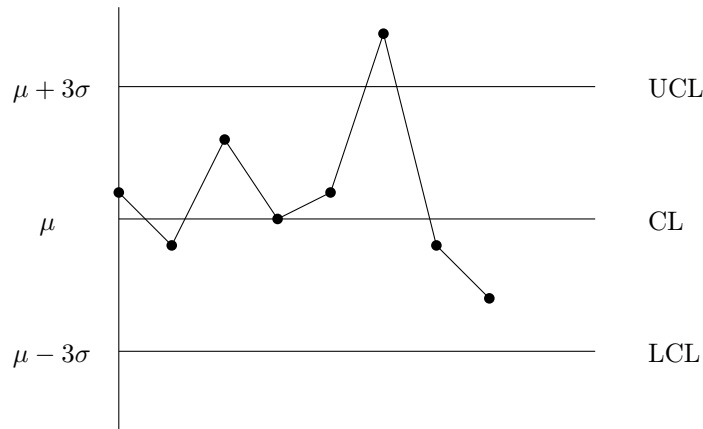
Figure 2.1: Shewhart chart for individual observations with control limits.

As described in Chapter 1, we need to base our monitoring approach on regression analysis in order to detect contextual anomalies. More specifically, by using regression models we are able to describe observed values by external factors. Subsequently, we can monitor the deviation of the regression model by monitoring the residuals, which are basically estimated error terms, with suitable control charts. An abnormal residual value should be equivalent to an abnormal response value, which means that there could be a problematic situation. In our context we, therefore, take the residuals from the regression model as "observations" for phase I and phase II analysis.

2.2 Multivariate multiple linear regression model

As previously explained, it is often crucial to use regression profiles to correct for external factors. The (univariate) multiple linear regression model for one response variable is thoroughly explained in van Dalen (2018). In this section we generalize the linear regression model for one response variable to the case of multiple response variables based on Monahan (2008, Chapter 9), such that multiple quality characteristics can be monitored simultaneously. From now on, we will call the model for multiple responses the multivariate multiple linear regression (MMLR) model. As mentioned in the previous section, we need n fixed observations in order fit the in-control model, which should be representative for the actual in-control process. Note that these observations are now actually observation vectors, since we have multiple responses for each observation in time. Based on the in-control model, we are able to accurately predict new observations. Subsequently, we can monitor the residual vectors of these new observations in order to detect deviations from expected behavior. Before we can derive statistical properties, we need to properly define the MMLR model.

Consistent with the notation defined in van Dalen (2018, Section 2.2), we define the real-valued observations $Y_{i\bullet}$ as observations from the following MMLR model consisting of p regressors and q responses:

$$Y_{i\bullet} = X_{\infty}^{(i)} B + E_{i\bullet}, \quad i = 1, 2, \ldots, \tag{2.1}$$

where $Y_{i\bullet} = \begin{pmatrix} y_{i1} & y_{i2} & \ldots & y_{iq} \end{pmatrix}$, $X_{\infty}^{(i)} = \begin{pmatrix} 1 & x_{i1} & x_{i2} & \ldots & x_{ip} \end{pmatrix}$, $B = \begin{pmatrix} \beta_{01} & \beta_{02} & \ldots & \beta_{0q} \\ \beta_{11} & \beta_{12} & \ldots & \beta_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \beta_{p2} & \ldots & \beta_{pq} \end{pmatrix}$

and $E_{i\bullet}^{T} = \begin{pmatrix} \varepsilon_{i1} & \varepsilon_{i2} & \ldots & \varepsilon_{iq} \end{pmatrix}^{T} \overset{\text{iid}}{\sim} \mathcal{N}_q(\mathbf{0}, \Sigma)$. The unknown $q \times q$ covariance matrix Σ corresponds to the unknown variance σ^2 in the univariate case and is assumed to be positive definite. Here, each $Y_{i\bullet}$-th observation vector in time is explained by the regressors defined by the row vector $X_{\infty}^{(i)}$,

and the corresponding regression coefficients for each response defined by the matrix of unknown coefficients B. We assume that the coefficient matrix B is constant for all observations coming from an in-control period. Furthermore, the vector $X_\infty^{(i)}$ represents the i-th row of the infinite matrix X_∞, which contains all observations of the regressors in time and is defined by

$$
X_\infty = \begin{pmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1p} \\
1 & x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \cdots & x_{np} \\
1 & x_{(n+1)1} & x_{(n+1)2} & \cdots & x_{(n+1)p} \\
1 & x_{(n+2)1} & x_{(n+2)2} & \cdots & x_{(n+2)p} \\
\vdots & \vdots & \vdots & \ddots & \vdots
\end{pmatrix}
\tag{2.2}
$$

In order to fit the in-control model we consider the first n observations, which are assumed to be from an in-control process. By means of the following matrix notation, also known as the multivariate Gauss-Markov model, we are able to express the in-control model altogether:

$$
Y_{[1:n,1:q]} = X_n B + E_{[1:n,1:q]},
\tag{2.3}
$$

where

$$
Y_{[1:n,1:q]} = \begin{pmatrix}
y_{11} & y_{12} & \cdots & y_{1q} \\
y_{21} & y_{22} & \cdots & y_{2q} \\
\vdots & \vdots & \ddots & \vdots \\
y_{n1} & y_{n2} & \cdots & y_{nq}
\end{pmatrix}, \qquad
X_n = \begin{pmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1p} \\
1 & x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \cdots & x_{np}
\end{pmatrix},
$$

$$
E_{[1:n,1:q]} = \begin{pmatrix}
\varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1q} \\
\varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2q} \\
\vdots & \vdots & \ddots & \vdots \\
\varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nq}
\end{pmatrix},
$$

and B is as defined above.

Here, $Y_{[1:n,1:q]}$ is the $n \times q$ *response matrix*, X_n is the $n \times (p+1)$ *design matrix*, B is the $(p+1) \times q$ *coefficient matrix* and $E_{[1:n,1:q]}$ is the $n \times q$ *error matrix*. Note that we assume that the design matrix X_n is the same for all q responses. We can break the in-control model of (2.3) apart by either row (observation) or column (response). The error vectors here are usually assumed to be uncorrelated across observations, but the responses are not uncorrelated within an observation. Dissecting by rows, we simply obtain the model defined in (2.1) for the i-th observation, $i = 1, \ldots, n$. To discuss the Gauss-Markov form, we transpose these row vectors, which results into:

$$
Y_{i\bullet}^T = B^T \left(X_\infty^{(i)} \right)^T + E_{i\bullet}^T, \quad i = 1, \ldots, n.
\tag{2.4}
$$

For estimation purposes, we work more generally and assume for now that the errors do not follow a particular distribution. Hence, we simply obtain that

$$
\mathbb{E}\left[E_{i\bullet}^T \right] = \mathbf{0}, \text{ and } \mathrm{Cov}\left(E_{i\bullet}^T \right) = \Sigma.
$$

The mean and covariance within an observation can thus be written directly in terms of the response as

$$
\mathbb{E}\left[Y_{i\bullet}^T \right] = B^T \left(X_\infty^{(i)} \right)^T, \text{ and } \mathrm{Cov}\left(Y_{i\bullet}^T \right) = \Sigma.
\tag{2.5}
$$

Breaking the model of (2.3) apart by columns gives a different view of this model, as we have for the j-th response variable

$$
Y_{\bullet j} = X_n B_{\bullet j} + E_{\bullet j}, \quad j = 1, \ldots, q,
\tag{2.6}
$$

where

$$
\mathbb{E}[E_{\bullet j}] = \mathbf{0}, \text{ and } \mathrm{Cov}\left(E_{\bullet j} \right) = \Sigma_{jj} I_n.
$$

Note that this basically equals the univariate case when considering the first n observations altogether, except that the errors within response j now have variance $\sigma_j^2 := \Sigma_{jj}$. Overall, we are able to capture the covariance relationships across both rows and columns by

$$\text{Cov}(E_{ij}, E_{st}) = \begin{cases} 0 & \text{if } i \neq s \\ \Sigma_{jt} & \text{if } i = s \end{cases} \quad \text{or equivalently} \quad \text{Cov}(Y_{ij}, Y_{st}) = \begin{cases} 0 & \text{if } i \neq s \\ \Sigma_{jt} & \text{if } i = s \end{cases}, \quad (2.7)$$

where E_{ij} and Y_{ij} denote the (i, j)-th entry of the matrices $E_{[1:n,1:q]}$ and $Y_{[1:n,1:q]}$, respectively.

2.2.1   Least squares estimation

We have seen that the responses within an observation are correlated. If we want to estimate the coefficient matrix B, we cannot simply use ordinary least squares, since we have to take into account the introduced covariance across responses. Therefore, we use the multivariate version of generalized least squares which estimates B by minimizing the following function:

$$Q(B) = \sum_{i=1}^{n} \left(Y_{i\bullet} - X_\infty^{(i)} B \right) \Sigma^{-1} \left(Y_{i\bullet} - X_\infty^{(i)} B \right)^T \quad (2.8)$$

Now, note that (2.8) is simply the sum of the diagonal elements of the $n \times n$ square matrix

$$\left(Y_{[1:n,1:q]} - X_n B \right) \Sigma^{-1} \left(Y_{[1:n,1:q]} - X_n B \right)^T,$$

from which we deduce that

$$Q(B) = \text{Tr} \left[\left(Y_{[1:n,1:q]} - X_n B \right) \Sigma^{-1} \left(Y_{[1:n,1:q]} - X_n B \right)^T \right]. \quad (2.9)$$

Two directions are suggested in minimizing (2.9). The first route, following column dissection, suggests a series of q univariate least squares problems, solving normal equations of the form

$$X_n^T X_n B_{\bullet j} = X_n^T Y_{\bullet j}, \quad j = 1, \ldots, q.$$

Following the univariate case, it is obvious that the solution for the j-th response yields

$$\widehat{B}_{\bullet j} = \left(X_n^T X_n \right)^{-1} X_n^T Y_{\bullet j}, \quad j = 1, \ldots, q.$$

Alternatively, a second route which includes the computation of derivatives appears to be rather complicated. However, solving the multivariate normal equations is supported by Lemma A.5, using a matrix $\widehat{B}$ that solves to the multivariate normal equations:

$$X_n^T X_n B = X_n^T Y_{[1:n,1:q]}. \quad (2.10)$$

The solution to the multivariate normal equations clearly yields

$$\widehat{B} = \left(X_n^T X_n \right)^{-1} X_n^T Y_{[1:n,1:q]}, \quad (2.11)$$

and by Lemma A.5 we conclude that (2.11) is the least squares estimator of B.

Subsequently, estimation of the covariance matrix uses a generalization of SSE with the error sum of squares and cross-products matrix F, defined by

$$F = \left(Y_{[1:n,1:q]} - X_n \widehat{B} \right)^T \left(Y_{[1:n,1:q]} - X_n \widehat{B} \right). \quad (2.12)$$

By Lemma A.6 we obtain that the commonly used covariance matrix estimator

$$\widehat{\Sigma} = \frac{1}{n - p - 1} F \quad (2.13)$$

is an unbiased estimator of Σ, as $\text{rank}(X_n)$ is assumed to be equal to $p + 1$.

Estimability in the multivariate case follows the univariate case, but we have to be cautious since we are dealing with matrices. For this purpose, let $m \in \mathbb{R}^q$ be a $q \times 1$ vector representing the

coefficients of the linear combination across responses. With these coefficients we can select which responses to take into account and to which extent by means of a weight.

Estimating the scalar quantity $\lambda^T B m$ suggests taking the linear combination across responses expressed by the vector m:

$$Y_{[1:n,1:q]}m = X_n B m + E_{[1:n,1:q]}m. \tag{2.14}$$

For instance, if we want to only take into account the first response with weight 1, we simply choose $m = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}^T$ such that (2.14) simplifies to

$$Y_{\bullet 1} = X_n B_{\bullet 1} + E_{\bullet 1}.$$

We now have a response vector $Y_{[1:n,1:q]}m$, with the same design matrix X_n, but with a coefficient vector Bm and error $E_{[1:n,1:q]}m$, whose covariance matrix is

$$\text{Cov}\left(E_{[1:n,1:q]}m\right) = \text{Cov}\begin{pmatrix} m^T E_{1\bullet}^T \\ \vdots \\ m^T E_{n\bullet}^T \end{pmatrix} = \left(m^T \Sigma m\right) I_n = \text{Cov}\left(Y_{[1:n,1:q]}m\right). \tag{2.15}$$

Note that since $m^T \Sigma m \in \mathbb{R}$, the model in (2.14) now appears as a standard univariate linear model, and all the usual results can be applied.

Definition 2.1. *A linear estimator of a scalar quantity $\lambda^T B m$ in the multivariate case is an estimator of the form $t^T Y_{[1:n,1:q]}m$, where $t \in \mathbb{R}^n$ is an $n \times 1$ vector which can be used to select the observations to take into account.*

Definition 2.2. *The scalar quantity $\lambda^T B m$ in the multivariate case is estimable if and only if a linear unbiased estimator of it exists.*

By Lemma A.7, we know that the scalar quantity is estimable if and only if $\lambda \in \mathcal{C}\left(X_n^T\right)$. Furthermore, Lemma A.8 ensures that if $\lambda^T B m$ is estimable, then $\lambda^T \widehat{B} m$ is constant for all solutions $\widehat{B}$ of the multivariate normal equations (2.10).

As usual, we are able to compute the mean and variance of this generalization of the least squares estimator.

$$\mathbb{E}\left[\lambda^T \widehat{B} m\right] = \lambda^T \mathbb{E}\left[\widehat{B}\right]m = \lambda^T \left(X_n^T X_n\right)^{-1} X_n^T \mathbb{E}\left[Y_{[1:n,1:q]}\right]m \tag{2.16}$$

$$= \lambda^T \left(X_n^T X_n\right)^{-1} X_n^T X_n B m = \lambda^T B m$$

$$\text{Var}\left(\lambda^T \widehat{B} m\right) = \lambda^T \left(X_n^T X_n\right)^{-1} X_n^T \text{Cov}\left(Y_{[1:n,1:q]}m\right) X_n \left(\left(X_n^T X_n\right)^{-1}\right)^T \lambda \tag{2.17}$$

$$= \lambda^T \left(X_n^T X_n\right)^{-1} X_n^T \left(m^T \Sigma m\right) I_n X_n \left(\left(X_n^T X_n\right)^{-1}\right)^T \lambda$$

$$= \left(m^T \Sigma m\right) \lambda^T \left(X_n^T X_n\right)^{-1} \lambda$$

By Lemma A.9, we conclude that $\lambda^T \widehat{B} m$ is the best linear unbiased estimator (BLUE) of $\lambda^T B m$.

2.2.2 Maximum likelihood estimation

We can now extend the Gauss-Markov model to include the assumption that the error distribution follows the multivariate normal distribution, leading to $Y_{i\bullet}$ independent of each other and

$$Y_{i\bullet}^T \sim \mathcal{N}_q\left(B^T \left(X_\infty^{(i)}\right)^T, \Sigma\right), \quad i = 1, \ldots, n. \tag{2.18}$$

This easily leads to the joint density and a likelihood function of the form

$$\log L(B, \Sigma) = \ell(B, \Sigma)$$

$$= -\frac{nq}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}\left(Y_{i\bullet} - X_\infty^{(i)}B\right)\Sigma^{-1}\left(Y_{i\bullet} - X_\infty^{(i)}B\right)^T$$

$$= -\frac{nq}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|) - \frac{1}{2} \operatorname{Tr} \left[\left(Y_{[1:n,1:q]} - X_n B \right) \Sigma^{-1} \left(Y_{[1:n,1:q]} - X_n B \right)^T \right]$$

$$= -\frac{nq}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|) - \frac{1}{2} \operatorname{Tr} \left[\Sigma^{-1} \left(Y_{[1:n,1:q]} - X_n B \right)^T \left(Y_{[1:n,1:q]} - X_n B \right) \right]$$

$$= -\frac{nq}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|)$$

$$- \frac{1}{2} \operatorname{Tr} \left[\Sigma^{-1} \left(Y_{[1:n,1:q]} - X_n \widehat{B} \right)^T \left(Y_{[1:n,1:q]} - X_n \widehat{B} \right) \right]$$

$$- \frac{1}{2} \operatorname{Tr} \left[\Sigma^{-1} \left(\widehat{B} - B \right)^T X_n^T X_n \left(\widehat{B} - B \right) \right],$$

where we make use of the fact that two matrices in a trace of a product can be switched without changing the result, and that the following relation holds: $X_n^T X_n \widehat{B} = X_n^T Y_{[1:n,1:q]}$.

Now, it should be obvious that the log-likelihood $\ell(B, \Sigma)$ can be maximized as a function of B by taking $B = \widehat{B}$. In this way the negative term in the end, which is the only term depending on B, vanishes. This leads to the function

$$\ell_*(\Sigma) = \ell\left(\widehat{B}, \Sigma \right)$$

$$= -\frac{nq}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|) - \frac{1}{2} \operatorname{Tr} \left[\Sigma^{-1} F \right],$$

where F is the error sum of squares matrix as defined in (2.12). Maximizing ℓ_* as function of Σ is rather difficult, but we use a clever route that avoids derivatives. Namely, when we apply Lemma A.11 to $\ell_*(\Sigma)$, we conclude that the maximum likelihood estimators of B and Σ are the solution to the multivariate normal equations $\widehat{B}$ (2.11) and $\widehat{\Sigma}_{\text{ML}} = \frac{1}{n} F$. Note that therefore the least squares and maximum likelihood estimator of B are the same, whereas $\widehat{\Sigma}$ and $\widehat{\Sigma}_{\text{ML}}$ are slightly different.

It now simply follows from (2.16) and (2.17) that

$$\lambda^T \widehat{B} m \sim \mathcal{N} \left(\lambda^T B m, \left(m^T \Sigma m \right) \lambda^T \left(X_n^T X_n \right)^{-1} \lambda \right)$$

Also, by Lemma A.12 we have that $\lambda^T \widehat{B} m$ is the minimum-variance unbiased estimator (MVUE) of $\lambda^T B m$.

The distribution of the matrix F, however, requires some new definitions, including the *Wishart distribution*, which is a generalization of the chi-square distribution.

Definition 2.3. *Let $Z^{(i)}, i = 1, \ldots, m$, be iid $\mathcal{N}_q(\mathbf{0}, \Sigma)$, then the $q \times q$ non-negative definite matrix $W = \sum_{i=1}^m Z^{(i)} \left(Z^{(i)} \right)^T$ has the Wishart distribution with m degrees of freedom and scale matrix Σ, or equivalently $W \sim W_q(m, \Sigma)$.*

We will briefly state the most important results of the Wishart distribution. By Lemma A.14, we know that if $W \sim W_q(m, \Sigma)$, then for any $q \times 1$ vector a,

$$\frac{a^T W a}{a^T \Sigma a} \sim \chi_m^2.$$

Furthermore, from the characterization in Definition 2.3 it directly follows that if $W_1 \sim W_q(m_1, \Sigma)$ and, independently, $W_2 \sim W_q(m_2, \Sigma)$, then

$$W_1 + W_2 \sim W_q(m_1 + m_2, \Sigma).$$

And finally, by Lemma A.15 we know that if $W \sim W_q(m, \Sigma)$, and A is a $p \times q$ matrix, then

$$A W A^T \sim W_p \left(m, A \Sigma A^T \right).$$

We are now able to determine the distribution of the matrix F by the following theorem.

Theorem 2.1. *The error sum of squares matrix F, defined by*

$$F = \left(Y_{[1:n,1:q]} - X_n \widehat{B} \right)^T \left(Y_{[1:n,1:q]} - X_n \widehat{B} \right),$$

follows the $W_q(n - p - 1, \Sigma)$ distribution.

Proof. This proof basically generalizes the proof of Theorem B.1 in van Dalen (2018) to the multivariate model. For convenience, we denote $Y_{[1:n,1:q]}$ by Y during this proof. Then, we simply have that $X_n \widehat{B} = X_n \left(X_n^T X_n\right)^{-1} X_n^T Y = PY$, where $P = X_n \left(X_n^T X_n\right)^{-1} X_n^T$. Lemma A.17 ensures that P and $I_n - P$ are symmetric and idempotent. Subsequently, by using that $I_n - P$ is symmetric, we deduce that

$$
\begin{aligned}
F &= (Y - PY)^T (Y - PY) \\
&= \left((I_n - P) Y\right)^T \left((I_n - P) Y\right) \\
&= Y^T (I_n - P) (I_n - P) Y \\
&= Y^T (I_n - P)^{1/2} (I_n - P) (I_n - P)^{1/2} Y \\
&= \widetilde{Y}^T \widetilde{P} \widetilde{Y},
\end{aligned}
$$

where $\widetilde{Y} = (I_n - P)^{1/2} Y$ and $\widetilde{P} = (I_n - P)$.

Then, we define $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ to be the base vector with all components equal to 0, except the i-th, which is equal to 1. Subsequently, we have for the i-th row of $\widetilde{Y}$, $i = 1, \ldots, n$, that

$$
\begin{aligned}
\mathbb{E}\left[\widetilde{Y}_{i\bullet}^T\right] &= \mathbb{E}\left[\widetilde{Y}^T e_i\right] \\
&= \mathbb{E}\left[Y^T\right] (I_n - P)^{1/2} e_i \\
&= B^T X_n^T (I_n - P) (I_n - P)^{-1/2} e_i \\
&= \left(B^T X_n^T - B^T X_n^T P\right) (I_n - P)^{-1/2} e_i \\
&= \left(B^T X_n^T - B^T X_n^T X_n \left(X_n^T X_n\right)^{-1} X_n^T\right) (I_n - P)^{-1/2} e_i \\
&= \mathbf{0}.
\end{aligned}
$$

and, by using that $I_n - P$ is symmetric and idempotent and by Lemma A.18, we have that

$$
\begin{aligned}
\mathrm{Var}\left(\widetilde{Y}_{i\bullet}^T\right) &= \mathrm{Var}\left(\widetilde{Y}^T e_i\right) \\
&= \mathrm{Var}\left(Y^T (I_n - P)^{1/2} e_i\right) \\
&= \mathrm{Var}\left(Y^T (I_n - P) (I_n - P)^{-1/2} e_i\right) \\
&= \Sigma \left((I_n - P) (I_n - P)^{-1/2} e_i\right)^T (I_n - P) (I_n - P)^{-1/2} e_i \\
&= \Sigma \left(e_i^T (I_n - P)^{-1/2} (I_n - P) (I_n - P) (I_n - P)^{-1/2} e_i\right) \\
&= \Sigma \left(e_i^T (I_n - P)^{-1/2} (I_n - P) (I_n - P)^{-1/2} e_i\right) \\
&= \Sigma.
\end{aligned}
$$

Therefore, we conclude that $\widetilde{Y}_{i\bullet}^T$ are independent of each other and

$$
\widetilde{Y}_{i\bullet}^T \sim \mathcal{N}_q \left(\mathbf{0}, \Sigma\right), \quad i = 1, \ldots, n.
$$

Let us now define $A_1 := \widetilde{P}$ and $A_2 := P$, which are symmetric, idempotent matrices with ranks $n - p - 1$ and $p + 1$, respectively. Since $A_1 + A_2 = I_n$, we obtain by Lemma A.16 that

$$
F = \widetilde{Y}^T \widetilde{P} \widetilde{Y} \sim W_q(n - p - 1, \Sigma).
$$

$\square$

Theorem 2.1 confirms that the usual estimator of Σ as stated in (2.13) is indeed unbiased, since by properties of the Wishart distribution we simply have that

$$
\mathbb{E}\left[\widehat{\Sigma}\right] = \frac{1}{n - p - 1} \mathbb{E}[F] = \frac{1}{n - p - 1}(n - p - 1)\Sigma = \Sigma.
$$

Note that the maximum likelihood estimator of Σ, however, is biased. The bias of the maximum likelihood estimator is given by

$$\mathbb{E}\left[\widehat{\Sigma}_{\mathrm{ML}}\right] - \Sigma = \frac{n-p-1}{n}\Sigma - \Sigma = -\frac{(p+1)}{n}\Sigma \to 0_{q\times q} \quad \text{as } n \to \infty,$$

which means that $\widehat{\Sigma}_{\mathrm{ML}}$ is asymptotically unbiased.

2.2.3   Prediction

Now that we have completely defined the in-control model (2.3) and derived its most important properties, we can use the model to predict new observations. More specifically, we fit the MMLR model for a fixed in-control period consisting of n observations of q responses. Then, we use the estimators from the in-control model to predict new observations for the online monitoring period, which are called fitted values. These predictions enable us to calculate the predictive residuals, which are the differences between the real observations and the corresponding fitted values.

First, we introduce the MMLR model for m new observations for phase II analysis. We break the model apart by row (observation), so we basically obtain the model defined in (2.1) for the i-th observation, $i = n+1, \ldots, n+m$. That is, we define the new observation vector $Y_{(n+j)\bullet}$ as observations from the following MMLR model consisting of p regressors and q responses:

$$Y_{(n+j)\bullet} = X_\infty^{(n+j)} B + E_{(n+j)\bullet}, \quad j = 1, \ldots, m, \tag{2.19}$$

where $Y_{(n+j)\bullet}$, $X_\infty^{(n+j)}$, B and $E_{(n+j)\bullet}$ are as defined previously. Note that the coefficient matrix B is constant for all new observations. Since the error distribution is assumed to follow the multivariate normal distribution, that is $E_{(n+j)\bullet}^T \overset{\text{iid}}{\sim} \mathcal{N}_q(\mathbf{0}, \Sigma)$, it obviously follows that $Y_{(n+j)\bullet}$ are independent of each other and

$$Y_{(n+j)\bullet}^T \sim \mathcal{N}_q\left(B^T\left(X_\infty^{(n+j)}\right)^T, \Sigma\right), \quad j = 1, \ldots, m. \tag{2.20}$$

Now, we simply use the estimator $\widehat{B}$ (2.11), which is based on the n fixed observations, for B in order to predict new observations. Hence, the fitted value of a new observation $Y_{(n+j)\bullet}$ becomes

$$\widehat{Y}_{(n+j)\bullet} = X_\infty^{(n+j)}\widehat{B} = X_\infty^{(n+j)}\left(X_n^T X_n\right)^{-1} X_n^T Y_{[1:n,1:q]} = P_{(n,j)}Y_{[1:n,1:q]},$$

where $P_{(n,j)} = X_\infty^{(n+j)}\left(X_n^T X_n\right)^{-1} X_n^T$.

By simple calculations and using Lemma A.18, we obtain that

$$\widehat{Y}_{(n+j)\bullet}^T \sim \mathcal{N}_q\left(B^T\left(X_\infty^{(n+j)}\right)^T, \Sigma\left(X_\infty^{(n+j)}\left(X_n^T X_n\right)^{-1}\left(X_\infty^{(n+j)}\right)^T\right)\right). \tag{2.21}$$

Subsequently, we want to compare these fitted values with the real, actual observations. Therefore, we define the difference of these quantities as the predictive residual vector. The $(n+j)$-th predictive residual vector becomes

$$\widehat{E}_{(n+j)\bullet} = Y_{(n+j)\bullet} - \widehat{Y}_{(n+j)\bullet} = Y_{(n+j)\bullet} - P_{(n,j)}Y_{[1:n,1:q]}. \tag{2.22}$$

From (2.20) and (2.21) it easily follows that

$$\begin{aligned}
\mathbb{E}\left[\widehat{E}_{(n+j)\bullet}^T\right] &= B^T\left(X_\infty^{(n+j)}\right)^T - B^T\left(X_\infty^{(n+j)}\right)^T \\
&= \mathbf{0},
\end{aligned}$$

and by using that $Y_{(n+j)\bullet}$ and $\widehat{Y}_{(n+j)\bullet}$ are independent as $\widehat{Y}_{(n+j)\bullet}$ only depends on the first n observations, we obtain that

$$\begin{aligned}
\mathrm{Var}\left(\widehat{E}_{(n+j)\bullet}^T\right) &= \mathrm{Var}\left(Y_{(n+j)\bullet}^T\right) + \mathrm{Var}\left(\widehat{Y}_{(n+j)\bullet}^T\right) \\
&= \Sigma + \Sigma\left(X_\infty^{(n+j)}\left(X_n^T X_n\right)^{-1}\left(X_\infty^{(n+j)}\right)^T\right)
\end{aligned}$$

$$= \Sigma \left(1 + X_\infty^{(n+j)} \left(X_n^T X_n \right)^{-1} \left(X_\infty^{(n+j)} \right)^T \right).$$

Finally, the MMLR model (2.19) assumes that the error terms are independent from each other. Now, we check if the residuals satisfy this assumption. Consider two residual vectors $\widehat{E}_{(n+j)\bullet}$ and $\widehat{E}_{(n+k)\bullet}$, where $j, k \in \{1, \ldots, m\}$, $j \neq k$. Then, we have that

$$\begin{aligned}
\text{Cov}\left(\widehat{E}_{(n+j)\bullet}^T, \widehat{E}_{(n+k)\bullet}^T \right) &= \text{Cov}\left(Y_{(n+j)\bullet}^T - Y_{[1:n,1:q]}^T P_{(n,j)}^T, Y_{(n+k)\bullet}^T - Y_{[1:n,1:q]}^T P_{(n,k)}^T \right) \\
&= \text{Cov}\left(Y_{[1:n,1:q]}^T P_{(n,j)}^T, Y_{[1:n,1:q]}^T P_{(n,k)}^T \right) \\
&= \text{Cov}\left(\begin{pmatrix} P_{(n,j)} Y_{\bullet 1} \\ \vdots \\ P_{(n,j)} Y_{\bullet q} \end{pmatrix}, \begin{pmatrix} P_{(n,k)} Y_{\bullet 1} \\ \vdots \\ P_{(n,k)} Y_{\bullet q} \end{pmatrix} \right) \\
&= \begin{pmatrix} P_{(n,j)} \Sigma_{11} P_{(n,k)}^T & P_{(n,j)} \Sigma_{12} P_{(n,k)}^T & \cdots & P_{(n,j)} \Sigma_{1q} P_{(n,k)}^T \\ P_{(n,j)} \Sigma_{21} P_{(n,k)}^T & P_{(n,j)} \Sigma_{22} P_{(n,k)}^T & \cdots & P_{(n,j)} \Sigma_{2q} P_{(n,k)}^T \\ \vdots & \vdots & \ddots & \vdots \\ P_{(n,j)} \Sigma_{q1} P_{(n,k)}^T & P_{(n,j)} \Sigma_{q2} P_{(n,k)}^T & \cdots & P_{(n,j)} \Sigma_{qq} P_{(n,k)}^T \end{pmatrix} \\
&= \Sigma P_{(n,j)} P_{(n,k)}^T \\
&= \Sigma \left(X_\infty^{(n+j)} \left(X_n^T X_n \right)^{-1} \left(X_\infty^{(n+k)} \right)^T \right).
\end{aligned}$$

Therefore, we conclude that $\widehat{E}_{(n+j)\bullet}^T$ are dependent on each other and that

$$\widehat{E}_{(n+j)\bullet}^T \sim \mathcal{N}_q \left(\mathbf{0}, \Sigma \left(1 + X_\infty^{(n+j)} \left(X_n^T X_n \right)^{-1} \left(X_\infty^{(n+j)} \right)^T \right) \right), \quad j = 1, \ldots, m.$$

Note, however, that most conventional multivariate monitoring charts, like Hotelling T^2 require that the predictive residual vectors have equal variance. Therefore, we first have to standardize the predictive residuals such that they have equal variance.

2.3   Summary

- The univariate multiple linear regression model is generalized to the case of multiple responses for each observation to the multivariate multiple linear regression model.
- Standard estimation results generalize easily to the multivariate case.
- The least squares and maximum likelihood estimator of B are equal and the resulting estimator $\widehat{B}$ (2.11) is unbiased.
- The chi-square distribution is generalized to the Wishart distribution, and similar properties can be derived.
- The matrix F (2.12) follows the $W_q(n - p - 1, \Sigma)$ distribution.
- The maximum likelihood estimator $\widehat{\Sigma}_{\text{ML}}$ is biased. An unbiased estimator of Σ is given by $\widehat{\Sigma}$ (2.13).
- Unlike the error terms, the residuals as defined in (2.22) are dependent on each other.

3 | Copula model

Most conventional multivariate control charts are based on the multivariate normality assumption. In practice, this assumption is hardly ever satisfied which can question the efficiency of these methods. The modern theory of copulas offers a great solution since it allows to construct a monitoring method which is not based on multivariate normality. Copula modeling is becoming increasingly popular and it is used in a wide variety of applications. Originally, the theory of copulas was mainly applied in the financial sector. However, it quickly became clear that the theory was very useful in other fields of application as well. In this chapter we explain the foundation of the copula model, which is based on the core literature Nelsen (2007) and Joe (2014). For a more detailed view on the application of copulas, see Hofert et al. (2018). First, we introduce the main idea behind copulas and its most important properties. Then, we thoroughly describe bivariate and vine copula models.

3.1 An introduction to copulas

A copula is a multivariate cumulative distribution function with all univariate marginals being uniform on the unit interval. This can be easily achieved since any continuous random variable can be transformed to a uniform random variable over the unit interval by its probability integral transformation. The name "copula" is derived from the fact that a copula "couples" a multivariate distribution function to its marginals. More specifically, copulas allow to combine the marginals of univariate random variables to arrive at the multivariate distribution of these random variables, which makes them very useful to model non-normal multivariate data. Furthermore, copulas can be used to efficiently describe the dependence structure of a multivariate distribution, as they define non-parametric measures of dependence between the individual components. The usefulness of copulas is based on Sklar's theorem, which allows to separate a multivariate distribution into two components: an appropriate copula which describes the dependence structure and all the marginals. This provides great flexibility in multivariate modeling.

We will go into more detail and introduce the mathematical foundation of copulas, which is based on Nelsen (2007) and Hofert et al. (2018). First of all, we need to define a copula in probabilistic terms.

Definition 3.1. *(Copula).* $C : [0,1]^d \to [0,1]$ *is a d-dimensional copula if C is a joint CDF of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals.*

Additionally, we need to define some analytic conditions for a copula. The above defined function C is a d-dimensional copula if:

- $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_p) = 0$, the copula equals 0 if at least one coordinate is 0.

- $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$, the copula is equal to u_i if all coordinates are 1 except u_i.

- C is d-increasing. Before we are able to define this concept, we need additional definitions. First, for any $a, b \in [0,1]^d$, $a_i \leq b_i$, $i = 1, \ldots, d$, let $(a,b]$ denote the hyper-rectangle with lower end point a and upper end point b defined by $\{u \in [0,1]^d : a_i < u_i \leq b_i, i = 1, \ldots, d\}$. Then, for any hyper-rectangle $(a,b]$ in $[0,1]^d$, define its C-volume as

$$\Delta_{(a,b]}C = \sum_{i \in \{0,1\}^d} (-1)^{\sum_{j=1}^d i_j} C\left(a_1^{i_1} b_1^{1-i_1}, \ldots, a_d^{i_d} b_d^{1-i_d}\right),$$

where the summation is taken over all 2^d-many vectors $i = (i_1, \ldots, i_d)$ for $i_1, \ldots, i_d \in \{0,1\}$.

If it holds that

$$\Delta_{(a,b]}C \geq 0 \quad \text{for all } a, b \in [0,1]^d, a \leq b,$$

then C is called d-increasing.

Now, we consider a random vector $(X_1, \ldots, X_d)$ and suppose that the marginal cumulative distribution functions $F_i(x) = \Pr[X_i \leq x]$, $i = 1, \ldots, d$, are continuous. Then, by applying the probability integral transform (see Lemma A.1) to each component, the random vector

$$(U_1, \ldots, U_d) = (F_1(X_1), \ldots, F_d(X_d))$$

has uniformly distributed marginals on the closed interval $[0,1]$. The copula C of $(X_1, \ldots, X_d)$ is now defined as the joint cumulative distribution function of $(U_1, \ldots, U_d)$:

$$C(u_1, \ldots, u_d) = \Pr[U_1 \leq u_1, \ldots, U_d \leq u_d]. \tag{3.1}$$

The copula C now contains all information on the dependence structure between the components of $(X_1, \ldots, X_d)$, whereas the marginal cumulative distribution functions F_i contain all information on the marginal distributions. The importance of the way copulas are defined above is that the procedure can be reversed to generate pseudo-random samples from general classes of multivariate probability distributions. Specifically, given a sample $(U_1, \ldots, U_d)$ from the copula distribution generated with the above procedure, the required sample from the multivariate probability distribution can be constructed as:

$$(X_1, \ldots, X_d) = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)).$$

Note that the inverse functions F_i^{-1} do not cause problems since the marginal cumulative distribution functions F_i were assumed to be continuous. Formula (3.1) can now be rewritten as:

$$\begin{aligned} C(u_1, \ldots, u_d) &= \Pr[F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d] \\ &= \Pr[X_1 \leq F_1^{-1}(u_1), \ldots, X_d \leq F_d^{-1}(u_d)]. \end{aligned}$$

The above derivation leads to Sklar's theorem, which provides the theoretical foundation of the application of copulas.

Theorem 3.1. *(Sklar's theorem).*

1. *For any d-dimensional distribution function H with univariate marginals $F_1, \ldots, F_d$, there exists a d-dimensional copula C such that*

$$H(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right), \quad (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{3.2}$$

   If all marginals $F_1, \ldots, F_d$ are continuous, the copula C is uniquely defined and given by

$$C(u_1, \ldots, u_d) = H\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) \tag{3.3}$$

2. *Conversely, given a d-dimensional copula C and univariate distribution functions $F_1, \ldots, F_d$, H defined by (3.2) is a d-dimensional distribution function with marginals $F_1, \ldots, F_d$ and "dependence structure" C.*

Proof. See the proof of Theorem 2.2 of Rüschendorf (2009). $\qquad\square$

By Sklar's theorem it is clear that copulas are those functions which combine the univariate marginals $F_1, \ldots, F_d$ to form the d-dimensional distribution function H. For estimation of H from data, this offers great flexibility as it allows one to model the marginals separately from the dependence structure. Copulas are thus exactly the functions to investigate if one is interested in the dependence structure between the components of a random vector.

A copula is a multivariate distribution function and if it is continuous the corresponding density function is defined by

$$c(u_1, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C(u_1, \ldots, u_d) \tag{3.4}$$

When we assume that $f_1, \ldots, f_d$ are the univariate marginal densities corresponding to $F_1, \ldots, F_d$, we have by (3.2) and (3.4) that

$$h(x_1, \ldots, x_d) = c\left(F_1(x_1), \ldots, F_d(x_d)\right) \cdot f_1(x_1) \cdot \ldots \cdot f_d(x_d) \tag{3.5}$$

Furthermore, since copulas are able to describe correlation of a multivariate distribution, they are bounded by an upper and lower bound. These bounds are called the Fréchet-Hoeffding bounds and they are one of the cornerstones of copula theory. These bounds are comparable to the $+1$ and -1 bounds of the linear correlation coefficient.

Theorem 3.2. *(Fréchet-Hoeffding bounds). For any d-dimensional copula C,*

$$W(u) \leq C(u) \leq M(u), \quad u \in [0, 1]^d,$$

where

$$W(u) = \max\left\{\sum_{j=1}^{d} u_j - d + 1, 0\right\} \quad and \quad M(u) = \min_{1 \leq j \leq d}\{u_j\}, \quad u \in [0, 1]^d.$$

Proof. See the proof of Proposition 1 of Genest et al. (1999). $\qquad\square$

Note that W is a copula only if $d = 2$, whereas M is a copula for all $d \geq 2$.

3.2 Copula families

Multivariate data often exhibit several characteristic properties that should be described as precisely as possible using a multivariate distribution function. Therefore, there exist various copula families that are able to express specific distributional characteristics like heavy tails or asymmetry. It is crucial to choose the right copula family for a given data set in order to receive a well-fitting model that covers the characteristic properties of the data. In this section the most popular and commonly used copulas will be described.

Definition 3.2. *(Independence copula). One of the simplest copulas is the independence copula, which corresponds to independent standard uniform components, defined by*

$$\Pi(u) = \prod_{j=1}^{d} u_j, \quad u \in [0, 1]^d.$$

3.2.1 Elliptical copulas

Elliptical copulas are among the most widely used copulas in practice, describing the dependence of the multivariate normal distribution in terms of the Normal copula and the dependence of the multivariate Student-t distribution in terms of the Student-t copula. The main advantage of elliptical copulas is that they are able to describe different levels of correlation between the components. The disadvantages are that they do not have closed form expressions and that they are restricted to have radial symmetry. Due to their implicit construction by Sklar's theorem, properties of elliptical copulas can be derived from the properties of the corresponding elliptical distributions. For detailed properties and results about elliptical distributions, see, e.g., Fang (2018). The general density function of an elliptical distribution is given by

$$f_g(x) = |\Sigma|^{-1/2} g\left((x - \mu)^T \Sigma^{-1} (x - \mu)\right), \quad x \in \mathbb{R}^d.$$

By declaring the function $g(t)$ it is possible to define some well-known density functions. For instance, taking $g(t) = (2\pi)^{-d/2} \exp\{-t/2\}$ results in the multivariate normal distribution, and taking $g(t) = c\left(1 + t/\nu\right)^{-(d+\nu)/2}$ where c is a constant, results in the multivariate Student-t distribution with ν degrees of freedom. The contour of such elliptical density functions has the form of an ellipse. Now, by Hofert et al. (2018), we can define the corresponding elliptical copulas.

Definition 3.3. *(Normal copula).* *The d-dimensional Normal copula C_P^n is the copula obtained by Sklar's theorem from the multivariate normal distribution $\mathcal{N}_d(\mathbf{0}, P)$ with correlation matrix P. If Φ_P denotes the distribution function of the latter, we have*

$$C_P^n(u) = \Phi_P\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right), \quad u \in [0,1]^d, \tag{3.6}$$

where Φ^{-1} denotes the quantile function of $\mathcal{N}(0,1)$.

When P is positive definite, the corresponding copula density function is denoted by

$$c_P^n(u) = \frac{\varphi_P\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right)}{\prod_{j=1}^d \varphi\left(\Phi^{-1}(u_j)\right)}, \quad u \in (0,1)^d.$$

Bivariate copulas are often easy to interpret and visualize and, they are used in constructing vine copulas. Therefore, we consider the bivariate Normal copula by setting $d = 2$. Formula (3.6) now simplifies to

$$C_P^n(u_1, u_2) = \Phi_P\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right), \quad 0 \le u_1, u_2 \le 1,$$

where Φ_P denotes the distribution function of the bivariate normal distribution $\mathcal{N}_2(\mathbf{0}, P)$ with correlation matrix $P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where ρ is Pearson's correlation coefficient that measures linear correlation between two variables. For more information on Pearson's correlation coefficient, see Section 3.3. Figure 3.1 shows the bivariate Normal copula for several parameters ρ.
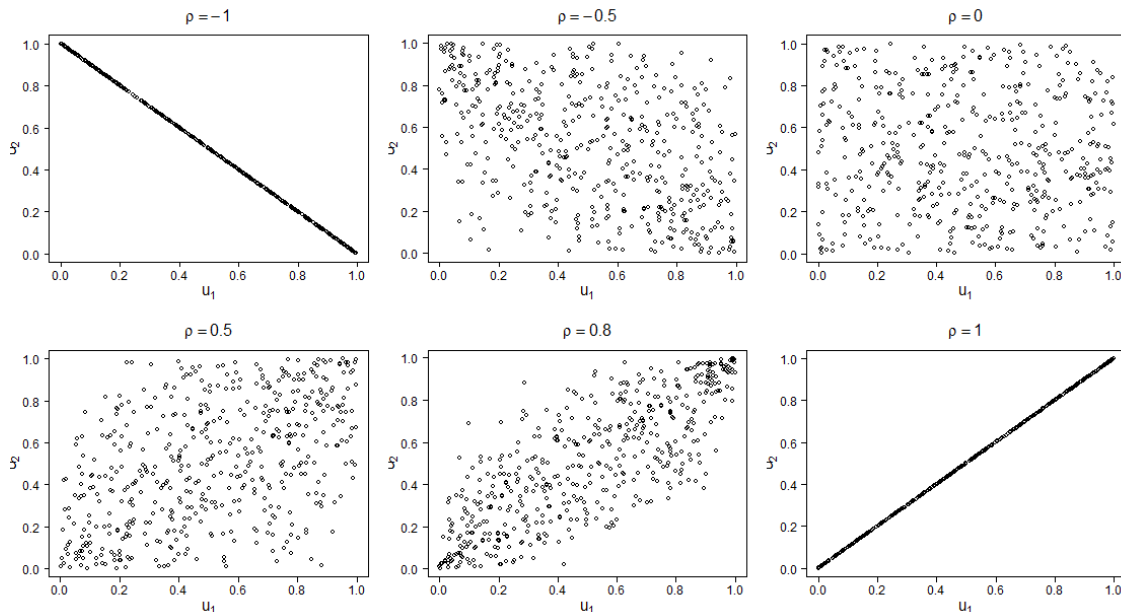
Figure 3.1: Bivariate Normal copula for different parameters ρ.

Before we are able to define the Student-t copula, we need to define the multivariate Student-t distribution.

Definition 3.4. *(Multivariate Student-t distribution).* *Let Y and u be independent from each other and distributed as $\mathcal{N}(\mathbf{0}, \Sigma)$ and χ_ν^2, respectively. Then,*

$$X = \frac{Y}{\sqrt{\frac{u}{\nu}}} + \mu$$

is said to be distributed as a multivariate Student-t distribution with location vector μ, scale matrix Σ and degrees of freedom ν.

Definition 3.5. *(Student-t copula).* *The d-dimensional Student-t copula $C_{P,\nu}^t$ is the copula obtained by Sklar's theorem from the multivariate Student-t distribution with location vector $\mathbf{0}$, scale*

matrix P and $\nu > 0$. If $T_{P,\nu}$ denotes the distribution function of the latter, we have

$$C^t_{P,\nu}(u) = T_{P,\nu}\left(T^{-1}_\nu(u_1), \ldots, T^{-1}_\nu(u_d)\right), \quad u \in [0,1]^d, \tag{3.7}$$

where T^{-1}_ν denotes the quantile function of the univariate Student-t distribution with ν degrees of freedom.

If P is positive definite, the corresponding copula density function is denoted by

$$c^t_{P,\nu}(u) = \frac{t_{P,\nu}\left(T^{-1}_\nu(u_1), \ldots, T^{-1}_\nu(u_d)\right)}{\prod^d_{j=1} t_\nu\left(T^{-1}_\nu(u_j)\right)}, \quad u \in (0,1)^d.$$

For $d = 2$ we consider the bivariate Student-t copula, which simply follows from (3.7) and becomes

$$C^t_{P,\nu}(u_1, u_2) = T_{P,\nu}\left(T^{-1}_\nu(u_1), T^{-1}_\nu(u_2)\right), \quad 0 \le u_1, u_2 \le 1,$$

where $T_{P,\nu}$ denotes the distribution function of the bivariate Student-t distribution with location vector $\mathbf{0}$, scale matrix $P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\nu > 0$. Note that the bivariate Student-t copula has two parameters instead of one, namely ρ and ν. Figures 3.2 and 3.3 show the bivariate Student-t copula for various values of ρ and ν.
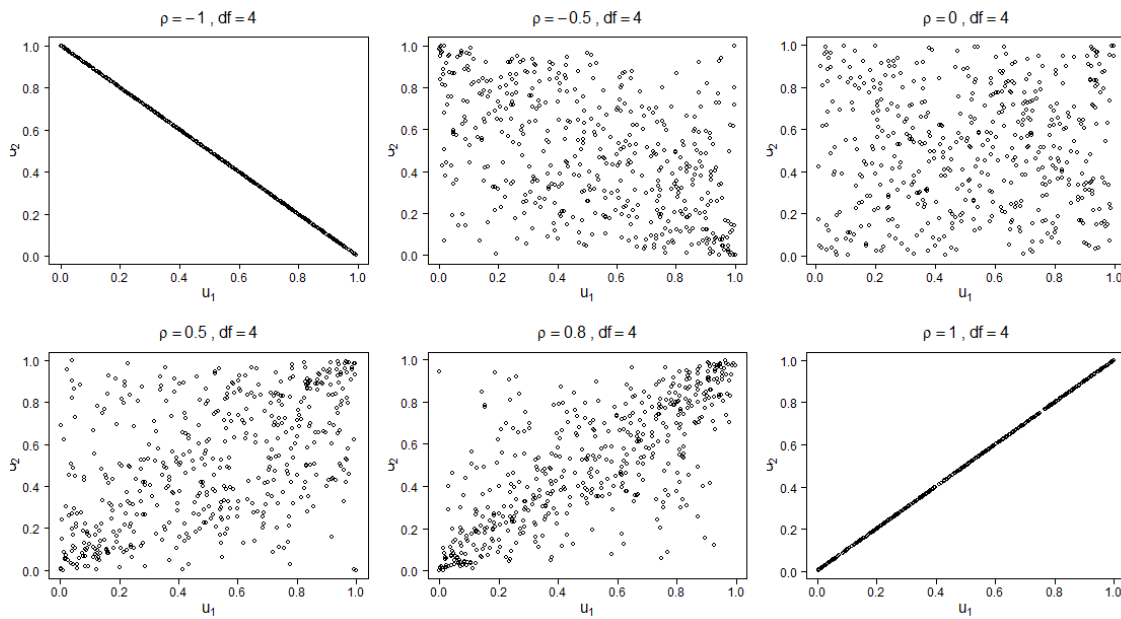
Figure 3.2: Bivariate Student-t copula for different parameters ρ.
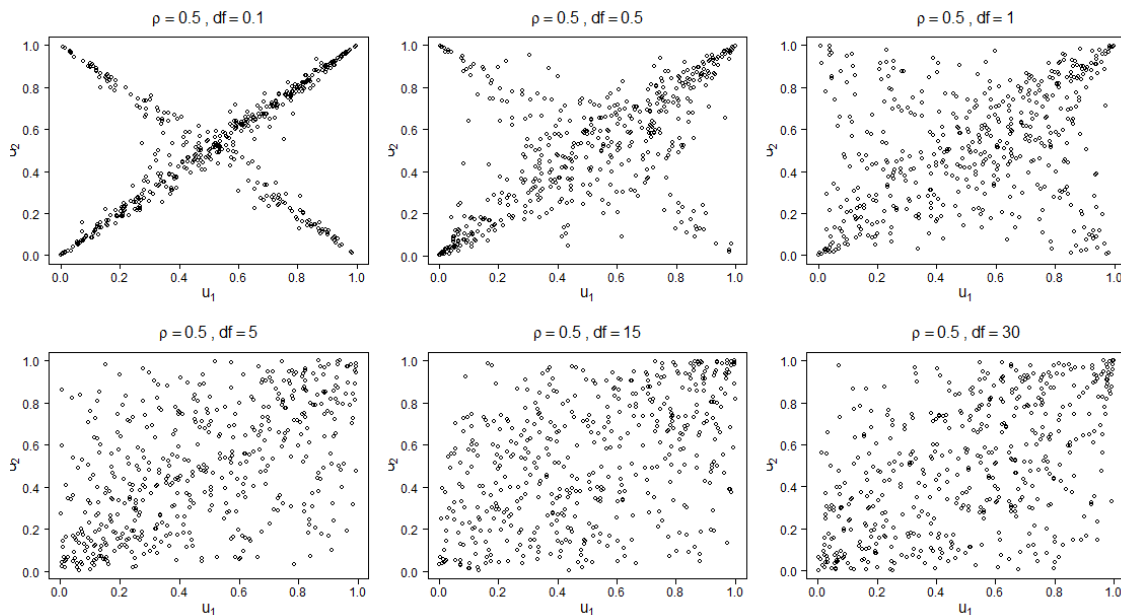
Figure 3.3: Bivariate Student-t copula for several degrees of freedom ν.

3.2.2 Archimedean copulas

Archimedean copulas are important in copula modeling since they can be constructed with ease and they offer a wide choice of families and, therefore, many distributional properties can be covered. Before we can define the general Archimedean copula we need to define the concept pseudo-inverse.

Definition 3.6. *For a continuous and strictly decreasing function $\phi : [0,1] \to [0,\infty)$ with $\phi(1) = 0$, the pseudo-inverse is defined as*

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \le t \le \phi(0), \\ 0, & \phi(0) \le t \le \infty. \end{cases}$$

The pseudo-inverse of ϕ is continuous, non-increasing on $[0,\infty)$ and even strictly decreasing on $[0, \phi(0)]$. Furthermore, it holds that

$$\phi\left(\phi^{[-1]}(t)\right) = \begin{cases} t, & 0 \le t \le \phi(0), \\ \phi(0), & \phi(0) \le t \le \infty, \end{cases}$$

and on $[0,1]$ it holds that $\phi^{[-1]}(\phi(u)) = u$. Additionally, we have that $\phi^{[-1]} = \phi^{-1}$ if $\phi(0) = \infty$.

Now, we have all the information we need to define the general Archimedean copula.

Definition 3.7. *(Archimedean copula). Let ϕ be a function as defined above and let $\phi^{[-1]}$ be its pseudo-inverse. Then, a d-dimensional Archimedean copula is a copula of the form*

$$C_\theta(u) = \phi^{[-1]}\left(\phi(u_1) + \ldots + \phi(u_d)\right), \quad u \in [0,1]^d, \tag{3.8}$$

where ϕ is the so-called generator of the copula.

Again, we set the dimension $d = 2$ and, by (3.8), a bivariate Archimedean copula is defined as

$$C_\theta(u_1, u_2) = \phi^{[-1]}\left(\phi(u_1) + \phi(u_2)\right), \quad 0 \le u_1, u_2 \le 1. \tag{3.9}$$

By choosing appropriate generators ϕ we can define various families in the extensive class of Archimedean copulas. Within these families we can distinguish by a different choice of the parameter θ. In this section we only discuss the most important Archimedean copulas. Many more families are described in Nelsen (2007) and Joe (2014).

Definition 3.8. *(Clayton copula).* *The Clayton family is defined by the generator* $\phi_\theta(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$ *where* $\theta \in [-1, \infty) \setminus \{0\}$. *The bivariate Clayton copula is*

$$C_\theta^c(u_1, u_2) = \left[\max\left\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\right\}\right]^{-\frac{1}{\theta}}, \quad 0 \le u_1, u_2 \le 1. \tag{3.10}$$

Figure 3.4 exemplifies the bivariate Clayton copula for various parameters θ. From the figure it becomes clear that the Clayton copula is asymmetric, exhibiting greater dependence in the negative tail than in the positive tail.
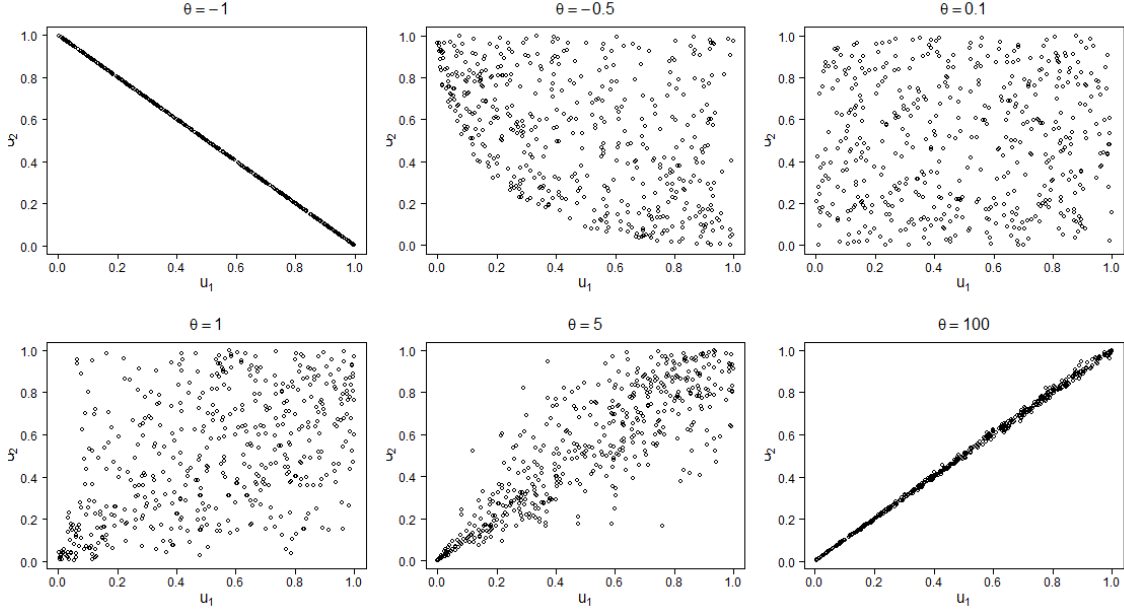
Figure 3.4: Bivariate Clayton copula for various parameters θ.

Definition 3.9. *(Gumbel copula).* *The Gumbel family is defined by the generator* $\phi_\theta(t) = (-\ln t)^\theta$ *where* $\theta \in [1, \infty)$. *The bivariate Gumbel copula is*

$$C_\theta^g(u_1, u_2) = \exp\left\{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{\frac{1}{\theta}}\right\}, \quad 0 \le u_1, u_2 \le 1. \tag{3.11}$$

Figure 3.5 illustrates the bivariate Gumbel copula for various parameters θ. From the figure it becomes clear that the Gumbel copula is asymmetric, exhibiting greater dependence in the positive tail than in the negative tail.
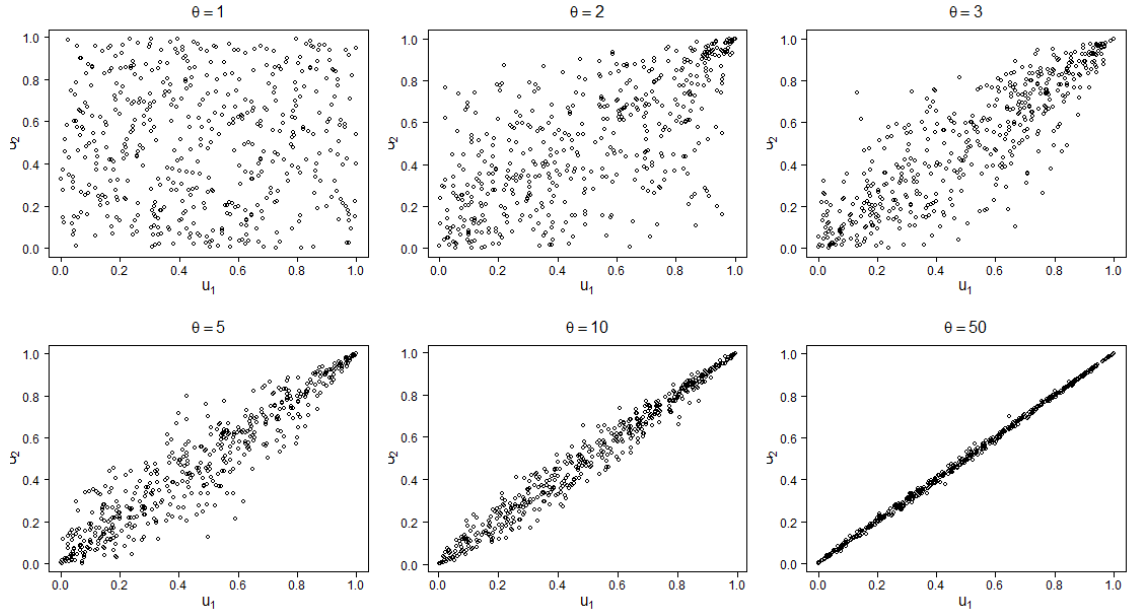
Figure 3.5: Bivariate Gumbel copula for different parameters θ.

Definition 3.10. *(Frank copula). The Frank family is defined by the generator* $\phi_\theta(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$ *where* $\theta \in (-\infty, \infty) \setminus \{0\}$. *The bivariate Frank copula is*

$$C_\theta^f(u_1, u_2) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u_1}-1\right)\left(e^{-\theta u_2}-1\right)}{e^{-\theta}-1}\right), \quad 0 \le u_1, u_2 \le 1. \tag{3.12}$$

Figure 3.6 shows the bivariate Frank copula for various parameters θ.
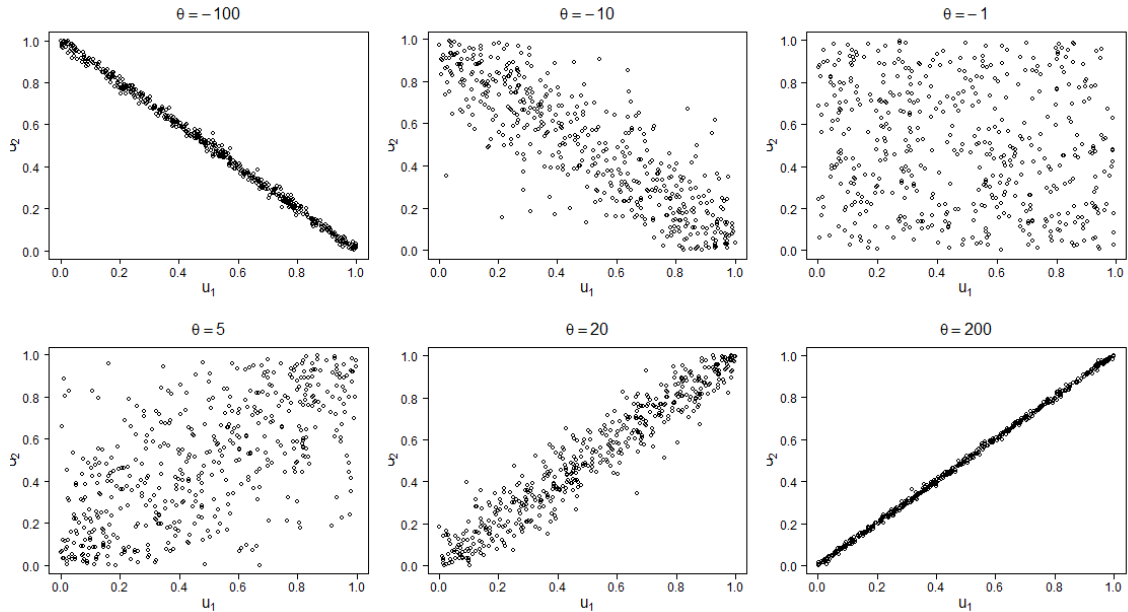
Figure 3.6: Bivariate Frank copula for several parameters θ.

3.3 Rank correlation

From the application point of view, it is often desirable to summarize the dependence between components of a random vector by a real number. Such numerical summaries of dependence are

known as measures of association and are mostly studied in the bivariate case. In this section several measures of association are discussed, based on Hofert et al. (2018) and Mühlig (2017).

Recall that one widely known measure of association is Pearson's (or the linear) correlation coefficient defined, for a random vector (X_1, X_2) whose components have finite variances, by

$$\rho(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)}\sqrt{\operatorname{Var}(X_2)}} \tag{3.13}$$

The most important properties of Pearson's correlation coefficient are summarized below.

1. $-1 \leq \rho(X_1, X_2) \leq 1$.

2. If X_1 and X_2 are independent, then $\rho(X_1, X_2) = 0$.

3. For any $a_1 > 0$, $a_2 > 0$, or any $a_1 < 0$, $a_2 < 0$, and for any $b_1, b_2 \in \mathbb{R}$,

$$\rho(a_1 X_1 + b_1, a_2 X_2 + b_2) = \rho(X_1, X_2).$$

4. In particular, Pearson's correlation coefficient is invariant under strictly increasing linear transformations.

The use of Pearson's correlation coefficient as a measure of association, however, involves some severe limitations:

1. $\rho(X_1, X_2)$ does not exist for all random vectors (X_1, X_2) (only for those with finite second moments).

2. $\rho(X_1, X_2)$ depends on the marginal distribution functions of (X_1, X_2), even when the latter are continuous. The correlation can thus not be expressed in terms of the unique underlying copula alone. We explain this by Sklar's theorem. As a consequence of Sklar's theorem, all the dependence information between two continuous random variables is captured by the underlying copula. If we keep the copula unchanged, and hence the dependence unchanged, we can change ρ according to (3.13) by simply changing one of the marginals.

3. $\rho(X_1, X_2)$ is not invariant under strictly increasing transformations in general.

By only depending on the underlying copula in the case of continuous random vectors, rank correlation coefficients overcome the aforementioned limitations of Pearson's correlation coefficient. We elaborate on the two best-known examples of rank correlation measures: Kendall's tau and Spearman's rho.

Definition 3.11. *(Kendall's tau, Spearman's rho). Let (X_1, X_2) be a bivariate random vector with continuous marginal distribution functions F_1 and F_2.*

1. *Spearman's rho is defined by*

$$\rho_s(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)),$$

where ρ is Pearson's correlation coefficient.

2. *Let (X_1', X_2') be an independent copy of (X_1, X_2). Kendall's tau is defined by*

$$\tau(X_1, X_2) = \Pr[(X_1 - X_1')(X_2 - X_2') > 0] - \Pr[(X_1 - X_1')(X_2 - X_2') < 0].$$

Lemma A.19 provides representations of Kendall's tau and Spearman's rho in terms of the underlying copula C, namely

$$\tau(C) = 1 - 4 \int_{[0,1]^2} \frac{\partial}{\partial u_1} C(u_1, u_2) \frac{\partial}{\partial u_2} C(u_1, u_2) du_1 du_2,$$

and

$$\rho_s(C) = 12 \int_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3.$$

From these expressions it becomes clear that Kendall's tau and Spearman's rho only depend on the underlying copula C, and thus not on the marginals F_1 and F_2. Therefore, we can conclude that Kendall's tau and Spearman's rho

1. always exist.

2. are invariant under strictly increasing transformations.

3. can reach any value in $[-1, 1]$.

From a copula modeling point of view, one advantage of Kendall's tau over Spearman's rho is that there are analytical formulas known for the classes of elliptical and Archimedean copulas. For a detailed comparison of these rank correlation measures, see Nelsen (2007, Section 5.1.3).

3.4 Parameter estimation

In order to fit the correct copula family to cover characteristic properties of the data, one has to estimate the unknown parameters of the copula model based on the available data. In this section we only briefly describe two methods which are often used in practice.

First, we describe the method of maximum likelihood estimation based on the exposition in Choroś et al. (2010). Recall formula (3.5), which states that the density h of the d-dimensional distribution function H with univariate marginals $F_1, \ldots, F_d$ and corresponding univariate densities $f_1, \ldots, f_d$ can be represented as

$$h(x_1, \ldots, x_d) = c\left(F_1(x_1), \ldots, F_d(x_d)\right) \prod_{i=1}^{d} f_i(x_i),$$

where $c(u_1, \ldots, u_d)$ is the density of the d-dimensional copula $C(u_1, \ldots, u_d)$ as defined in (3.4). We consider a random sample of (i.i.d.) vectors $x^{(j)} = \left(x_1^{(j)}, \ldots, x_d^{(j)}\right)$, $j = 1, \ldots, n$, with the density h. Furthermore, we assume that the copula C belongs to a family of copulas indexed by a parameter $\theta : C = C(u_1, \ldots, u_d; \theta)$, and the marginals F_i and the corresponding densities f_i are indexed by parameters $\alpha_i : F_i = F_i(x_i; \alpha_i)$, $f_i = f_i(x_i; \alpha_i)$. The parameter vector $(\alpha_1, \ldots, \alpha_d, \theta)$ can then be estimated by the maximum likelihood estimator, that is, the maximizer $\left(\widehat{\alpha}_1^{\mathrm{ML}}, \ldots, \widehat{\alpha}_d^{\mathrm{ML}}, \widehat{\theta}^{\mathrm{ML}}\right)$ of the log-likelihood function

$$\ell(\alpha_1, \ldots, \alpha_d, \theta) = \sum_{j=1}^{n} \log c\left(F_1\left(x_1^{(j)}; \alpha_1\right), \ldots, F_d\left(x_d^{(j)}; \alpha_d\right); \theta\right) + \sum_{i=1}^{d} \sum_{j=1}^{n} \log f_i\left(x_i^{(j)}; \alpha_i\right).$$

In practice, the marginal distributions are often unknown and, therefore, we have to adapt the above parametric method. One solution is to use a semiparametric approach that is based on the pseudo-maximum likelihood. In this approach the marginals F_i are replaced by non-parametric estimators $\widehat{F}_i$, which could be the empirical distribution functions, for instance. As a result, the parameter θ can be estimated by the pseudo-maximum likelihood estimator, that is, the maximizer $\widehat{\theta}^{\mathrm{PML}}$ of the log-likelihood function

$$\ell(\theta) = \sum_{j=1}^{n} \log c\left(\widehat{F}_1\left(x_1^{(j)}\right), \ldots, \widehat{F}_d\left(x_d^{(j)}\right); \theta\right).$$

An alternative method to estimate the copula parameters is the use of rank correlation coefficients which are introduced in Section 3.3. These rank correlation coefficients are particularly useful for bivariate copulas with only one parameter to estimate. As shown in Hofert et al. (2018, Section 4.1.2), it is often possible to define an explicit function between the copula parameters and rank correlation coefficients. This relation can be used to estimate the copula parameter directly by the estimated correlation. For instance, if we consider bivariate copula families and Kendall's tau, Table 3.1 shows some of the relations that hold.

Family	Normal	Student-t	Clayton	Gumbel
Parameter	$\widehat{\rho} = \sin\left(\frac{\pi}{2}\widehat{\tau}\right)$	$\widehat{\rho} = \sin\left(\frac{\pi}{2}\widehat{\tau}\right)$	$\widehat{\theta} = \frac{2\widehat{\tau}}{1-\widehat{\tau}}$	$\widehat{\theta} = \frac{1}{1-\widehat{\tau}}$

Table 3.1: Estimation of copula parameters with Kendall's tau.

3.5 Vine copulas

It should be clear that for the bivariate case, there exist many well-investigated copula families which can be used in order to find the best-fitting copula model. For higher dimensions, however, the choice of usable copula families is somewhat limited, since these copula families are restrictive in terms of flexibility and dependence modeling. To overcome this drawback, we will make use of vine copulas which provide a flexible graphical model for describing multivariate copulas composed of a cascade of bivariate copulas, so-called pair-copulas. The main idea of this construction is to decompose a multivariate probability density into blocks of pair-copulas, where each pair can be chosen independently from the others (see, e.g., Brechmann and Schepsmeier (2013)).

3.5.1 Pair-copula decomposition

Vines are a graphical representation to specify pair-copula as introduced by Aas et al. (2009). We consider a vector $X = (X_1, \ldots, X_n)$ of random variables with a joint density function $f(x_1, \ldots, x_n)$. This density can be decomposed as

$$f(x_1, \ldots, x_n) = f_n(x_n) \cdot f_{n-1}(x_{n-1}|x_n) \cdot f_{n-2}(x_{n-2}|x_{n-1}, x_n) \cdot \ldots \cdot f_1(x_1|x_2, \ldots, x_n). \tag{3.14}$$

For the bivariate case it holds by (3.5) that

$$f(x_1, x_2) = c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1) \cdot f_2(x_2),$$

where c_{12} is the appropriate pair-copula density for the pair of transformed variables $F_1(x_1)$ and $F_2(x_2)$. It now simply follows that, if $f_2(x_2) \neq 0$,

$$f_1(x_1|x_2) = c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1), \tag{3.15}$$

for the same pair-copula. The density decomposition (3.14) now becomes

$$f(x_1, x_2) = f_2(x_2) \cdot c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1),$$

which turns out to be trivial.

Similarly, for three random variables X_1, X_2 and X_3 we have that

$$f_2(x_2|x_3) = c_{23}\left(F_2(x_2), F_3(x_3)\right) \cdot f_2(x_2),$$

where c_{23} is the appropriate pair-copula density for the pair of transformed variables $F_2(x_2)$ and $F_3(x_3)$. Furthermore,

$$f_1(x_1|x_2, x_3) = c_{12|3}\left(F(x_1|x_3), F(x_2|x_3)\right) \cdot f_1(x_1|x_3), \tag{3.16}$$

where $c_{12|3}$ is the appropriate pair-copula density for the pair of transformed variables $F(x_1|x_3)$ and $F(x_2|x_3)$. Alternatively, (3.16) can be decomposed as

$$f_1(x_1|x_2, x_3) = c_{13|2}\left(F(x_1|x_2), F(x_3|x_2)\right) \cdot f_1(x_1|x_2), \tag{3.17}$$

where $c_{13|2}$ is different from the pair-copula $c_{12|3}$. We can decompose (3.17) further by using (3.15), which leads to

$$f_1(x_1|x_2, x_3) = c_{13|2}\left(F(x_1|x_2), F(x_3|x_2)\right) \cdot c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1),$$

where two pair-copulas are present. The density decomposition (3.14) now becomes

$$\begin{aligned} f(x_1, x_2, x_3) &= f_3(x_3) \cdot f_2(x_2|x_3) \cdot f_1(x_1|x_2, x_3) \\ &= f_3(x_3) \cdot c_{23}\left(F_2(x_2), F_3(x_3)\right) \cdot f_2(x_2) \\ &\quad \cdot c_{13|2}\left(F(x_1|x_2), F(x_3|x_2)\right) \cdot c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1). \end{aligned}$$

It has now become clear that each term in (3.14) can be decomposed into the appropriate pair-copula times a (conditional) marginal density. In this way we can iteratively construct a product of pair-copulas and marginal densities for the joint density function.

In general, for a d-dimensional vector $\boldsymbol{v}$, it holds that

$$f_x(x|\boldsymbol{v}) = c_{xv_j|\boldsymbol{v}_{-j}}\left(F\left(x|\boldsymbol{v}_{-j}\right), F\left(v_j|\boldsymbol{v}_{-j}\right)\right) \cdot f_x(x|\boldsymbol{v}_{-j}),$$

where v_j is one arbitrarily chosen component of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the vector $\boldsymbol{v}$, excluding the component v_j. In this decomposition we use marginal conditional distributions of the form $F(x|\boldsymbol{v})$. Joe (1996) showed that, for every j,

$$F(x|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}\left(F\left(x|\boldsymbol{v}_{-j}\right), F\left(v_j|\boldsymbol{v}_{-j}\right)\right)}{\partial F\left(v_j|\boldsymbol{v}_{-j}\right)}, \tag{3.18}$$

where $C_{ij|\boldsymbol{k}}$ is a bivariate copula distribution function. If $\boldsymbol{v}$ is univariate, (3.18) becomes

$$F(x|v) = \frac{\partial C_{x,v}\left(F(x), F(v)\right)}{\partial F\left(v\right)}.$$

If x and v are uniform, this conditional distribution function can simply be expressed as

$$h(x, v, \Theta) = F(x|v) = \frac{\partial C_{x,v}(x, v, \Theta)}{\partial v}, \tag{3.19}$$

where the second parameter of h always corresponds to the conditioning variable and Θ denotes the set of parameters for $C_{x,v}$.

3.5.2 Vines

The decomposition of a multivariate density from Section 3.5.1, however, is not unambiguous. Therefore, a graphical model which is able to distinguish between the possibilities is needed (see, e.g., Brechmann and Schepsmeier (2013)). For this purpose, Bedford and Cooke (2002) introduced the model of regular vines, so-called R-vines, which are based on a set of trees. For a random vector consisting of d components, we obtain a d-dimensional vine. This vine can then be represented as a set of $d-1$ trees with $d(d-1)/2$ edges in total, where each edge stands for the corresponding pair-copula density. In general, the nodes of a tree are equivalent to the edges of the tree above. There is, however, one exception to this rule that only applies to the first tree. Namely, the nodes of the first tree are represented by the d variables. In order to get an idea of what such R-vine could look like, a possible graphical representation of a 5-dimensional R-vine is shown in Figure 3.7.
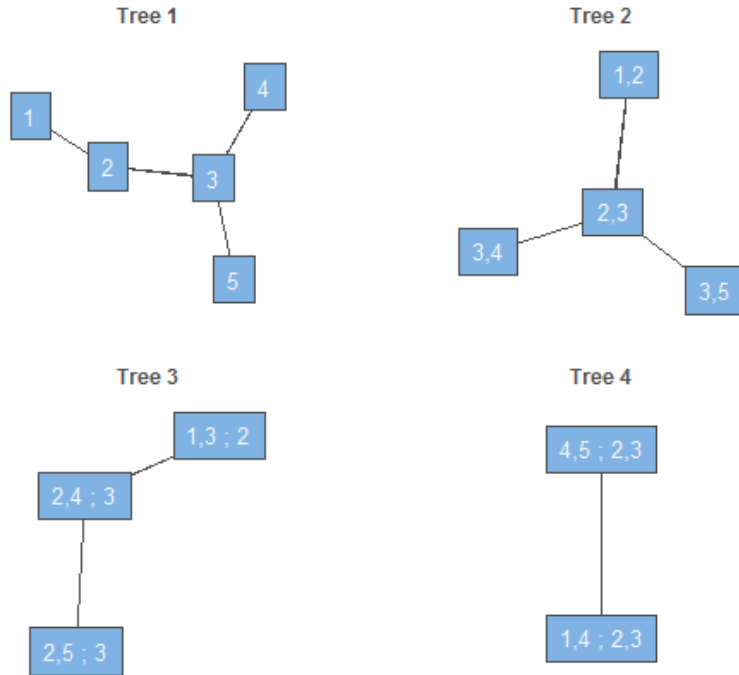
Figure 3.7: Possible 5-dimensional R-vine tree structure.

The above figure shows a possible R-vine for a 5-dimensional vector, consisting of four trees. The first tree simply consists of the nodes $1, 2, 3, 4, 5$. Subsequently, the second tree is constructed such that the nodes are the edges of the first tree. Then, the third tree is constructed in such a way that the nodes are the edges of the second tree. This procedure is repeated until the fourth and final tree, in which only one edge is left. The edge label describes the index of the corresponding pair-copula density. For instance, the node $(4, 5; 2, 3)$ of the fourth tree is equivalent to the bottom-left edge of the third tree, which represents the copula density of the fourth and fifth variable given the values of the second and third variable, denoted as $c_{4,5|2,3}$.

Following Bedford and Cooke (2002) we are able to define vines and R-vines in a mathematical way.

Definition 3.12. *(Vine.) If*

1. *$V = (T_1, \ldots, T_m)$ where T_i is a tree.*

2. *T_1 has nodes $N_1 = \{1, \ldots, n\}$ and a set of edges denoted by E_1.*

3. *For $i = 2, \ldots, m$, T_i has nodes $N_i \subset N_1 \cup E_1 \cup E_2 \cup \cdots \cup E_{i-1}$ and an edge set E_i.*

hold, then V is called a vine on n elements.

Definition 3.13. *(R-vine.) If*

1. *$m = n$.*

2. *T_i is a connected tree with edge set E_i and $n - i + 1$ nodes where $N_i = E_{i-1}$, for $i = 1, \ldots, n$.*

3. *For $i = 2, \ldots, n - 1$, if $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are two nodes in N_i linked with an edge, then the number of elements in $a \cap b = 1$.*

hold, then V is called an R-vine on n elements.

Following Czado (2010) we denote the edges in tree T_i by $jk|D$ where $j < k$ and D is the conditioning set. If the conditioning set D is empty, the edge is simply denoted by jk. The notation of an edge e in tree T_i will depend on the two adjacent nodes and, hence, on the two edges in T_{i-1} that have a common node. We now denote these edges by $a = j(a), k(a)|D(a)$ and $b = j(b), k(b)|D(b)$, where we define $V(a) := \{j(a), k(a), D(a)\}$ and $V(b) := \{j(b), k(b), D(b)\}$, respectively. The edges a and b in tree T_{i-1} are nodes in tree T_i connected by the edge $e = j(e), k(e)|D(e)$, where

$$j(e) := \min \{i : i \in (V(a) \cup V(b)) \setminus D(e)\}$$
$$k(e) := \max \{i : i \in (V(a) \cup V(b)) \setminus D(e)\}$$
$$D(e) := V(a) \cap V(b).$$

Furthermore, there exist two special cases of R-vine structures. First, an R-vine is called drawable vine, or so-called D-vine, if in each tree there exist two edges for each node at most. Second, an R-vine is called canonical vine, or so-called C-vine, if in each tree there exists one unique node with the number of edges equal to the total number of nodes in that particular tree, minus one. In order to clarify these descriptions, a possible graphical representation of a 5-dimensional D-vine and C-vine is shown in Figures 3.8 and 3.9, respectively.
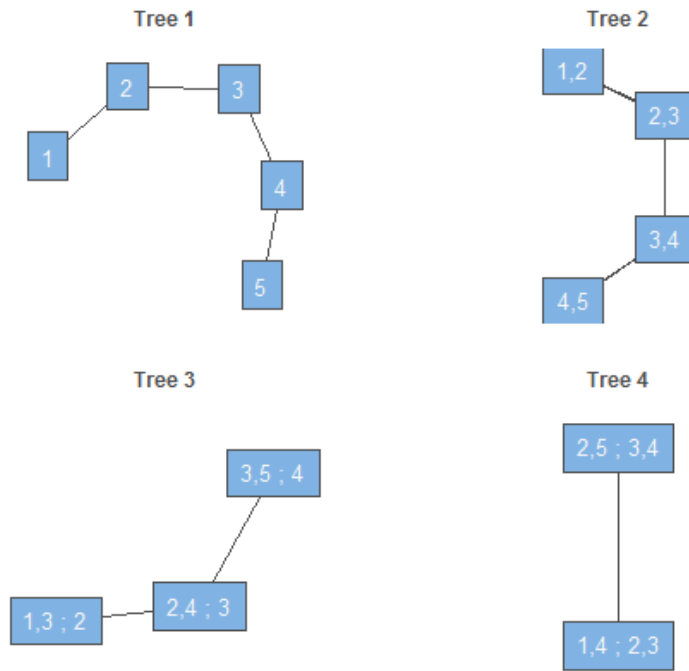
Figure 3.8: Possible 5-dimensional D-vine tree structure.
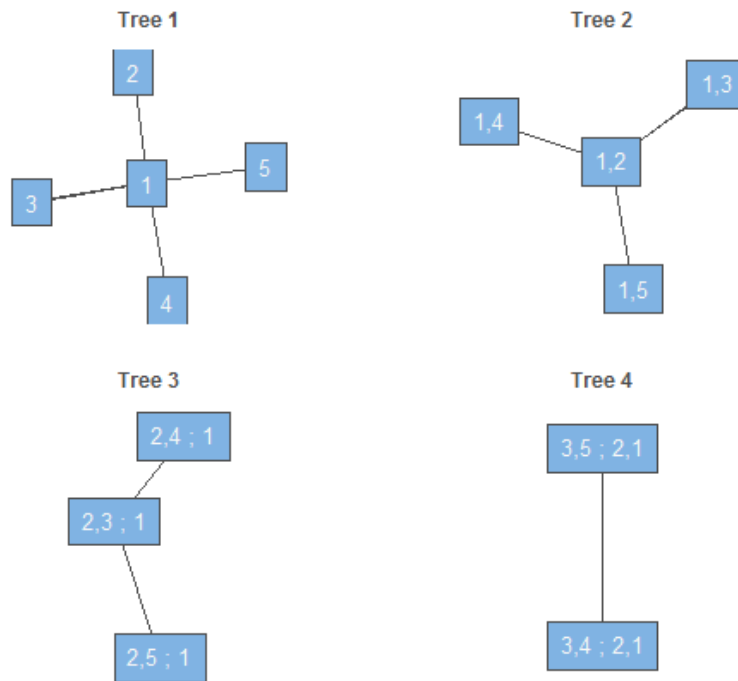
Figure 3.9: Possible 5-dimensional C-vine tree structure.

We are now able to properly construct a statistical model on an R-vine tree with nodes $\mathcal{N} := \{N_1, \ldots, N_{d-1}\}$ and edges $\mathcal{E} := \{E_1, \ldots, E_{d-1}\}$ based on Czado (2010). Each edge $e = j(e), k(e)|D(e)$ in E_i corresponds with a bivariate copula density $c_{j(e),k(e)|D(e)}$. Subsequently, let $X_{D(e)}$ be the subvector of X indicated by the indices contained in $D(e)$. Now, a vine distribution is defined as the distribution of the random vector $X = (X_1, \ldots, X_d)$ with marginal densities f_i, $i = 1, \ldots, d$, and the conditional density of $(X_{j(e)}, X_{k(e)})$ given the variables $X_{D(e)}$ specified as $c_{j(e),k(e)|D(e)}$ for

the R-vine tree with nodes $\mathcal{N}$ and edges $\mathcal{E}$. By Kurowicka and Cooke (2006) we know that the joint density of X is uniquely determined and expressed by

$$f(x_1,\ldots,x_d) = \prod_{r=1}^{d} f(x_r) \cdot \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)} \left(F\left(x_{j(e)}|x_{D(e)}\right), F\left(x_{k(e)}|x_{D(e)}\right) \right), \qquad (3.20)$$

where $x_{D(e)}$ is the subvector of x indicated by the indices contained in $D(e)$.

Furthermore, by Aas et al. (2009) we are able to express the joint density of X corresponding to a D-vine or C-vine tree. Regarding the D-vine structure, it holds that

$$f(x_1,\ldots,x_d) = \prod_{r=1}^{d} f(x_r) \cdot \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}$$
$$(F(x_i|x_{i+1},\ldots,x_{i+j-1}), F(x_{i+j}|x_{i+1},\ldots,x_{i+j-1})),$$

where index j identifies the trees, while index i runs over the edges in each tree.

And finally, for the C-vine structure, we have that

$$f(x_1,\ldots,x_d) = \prod_{r=1}^{d} f(x_r) \cdot \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\ldots,j-1} (F(x_j|x_1,\ldots,x_{j-1}), F(x_{j+i}|x_1,\ldots,x_{j-1})).$$

3.5.3 Model selection

To select the best-fitting vine copula model for a given dataset, we must process the following separate points. First, we have to select the R-vine structure, i.e., selecting which pairs of variables to use. Second, we have to choose a bivariate copula family for each selected pair. And finally, we need to estimate the parameters for each pair-copula. These steps can be done based on a sequential, heuristic method proposed by Dissmann et al. (2013). The structure of the tree is basically determined by spanning trees with an edge weight represented by, for instance, Kendall's tau or Spearman's rho. The procedure based on Kendall's tau starts at the first tree $T_1 = (N_1, E_1)$, where the term

$$\max \sum_{e=\{j,k\} \text{ in spanning tree}} |\widehat{\tau}_{j,k}|$$

must be solved and the resulting is selected. Note that the tree is selected in such a way that the chosen pairs model the strongest pairwise dependencies present. When the first tree structure is selected, we can choose an appropriate bivariate copula family for each edge of the tree by means of the AIC. Alternatively, one could use the BIC or cAIC as decision criterion or a goodness-of-fit test can be performed. Besides that, the independence copula is selected if independence is reasonable to assume by tests based on, for instance, Kendall's tau. Subsequently, we are able to estimate the parameters for each pair-copula in the tree with the methods described in Section 3.4. Next, the data is transformed using the fitted copula and (3.19). Now, we move forward to the next tree and repeat the steps. A small remark is that for the remaining trees T_i, $i = 2,\ldots,n-1$, Kendall's tau must be determined for the conditional variable pairs, so the corresponding optimization problem becomes

$$\max \sum_{e=\{j,k|D\} \text{ in spanning tree}} |\widehat{\tau}_{j,k|D}|,$$

where D is the conditioning set. In the end, we have defined all trees in order to construct a suitable R-vine model.

3.5.4 Matrix representation

The graphical model of vine copulas is not feasible for statistical algorithms, as storage of the trees is too expensive and too complicated. Therefore, a convenient way of representing an R-vine structure is to use matrix notation. In this subsection we introduce the R-vine matrix representation based on

the notation and definitions of Dissmann et al. (2013). Before we formally define an R-vine matrix we need to introduce the following two sets. For a lower-triangular matrix $M = (m_{i,j})_{i,j=1,\ldots,n}$ set for $i = 1, \ldots, n-1$,

$$B_M(i) := \{(m_{i,i}, D) | k = i+1, \ldots, n; D = \{m_{k,i}, \ldots, m_{n,i}\}\},$$
$$\widetilde{B}_M(i) := \{(m_{k,i}, D) | k = i+1, \ldots, n; D = \{m_{i,i}\} \cup \{m_{k+1,i}, \ldots, m_{n,i}\}\}.$$

Definition 3.14. *(R-vine matrix). A lower triangular matrix $M = (m_{i,j})_{i,j=1,\ldots,n}$ is an R-vine matrix if for $i = 1, \ldots, n-1$ and for all $k = i+1, \ldots, n-1$ there is an $j \in \{i+1, \ldots, n-1\}$ with*

$$(m_{k,i}, \{m_{k+1,i}, \ldots, m_{n,i}\}) \in B_M(j) \quad or \quad \in \widetilde{B}_M(j).$$

From the above definition it can be shown that every column contains all elements from the column to the right, i.e., $\{m_{i,i}, \ldots, m_{n,i}\} \subset \{m_{j,j}, \ldots, m_{n,j}\}$ for $1 \leq j < i \leq n$. Furthermore, there is a different element on the diagonal in every column, i.e., $m_{i,i} \notin \{m_{i+1,i+1}, \ldots, m_{n,i+1}\}$ for $i = 1, \ldots, n-1$.

In order to fully describe an R-vine we need two additional matrices $T = (t_{i,j})_{i,j=1,\ldots,n}$ and $P = (p_{i,j})_{i,j=1,\ldots,n}$ that contain information about the families and parameters of the pair-copulas, respectively. For all $j = 1, \ldots, n-1$, $i = j+1, \ldots, n$, the entry $m_{i,j}$ represents the pair-copula of the variables indexed by $m_{j,j}$ and $m_{i,j}$ conditional on the variables indexed by $\{m_{i+1,j}, \ldots, m_{n,j}\}$, $t_{i,j}$ represents the family corresponding to this copula and $p_{i,j}$ contains the parameters of this copula.

For instance, if we look at the R-vine from Figure 3.7, the corresponding matrices are

$$M = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 \\ 4 & 1 & 3 & 0 & 0 \\ 2 & 2 & 1 & 2 & 0 \\ 3 & 3 & 2 & 1 & 1 \end{pmatrix}, T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ t_{21} & 0 & 0 & 0 & 0 \\ t_{31} & t_{32} & 0 & 0 & 0 \\ t_{41} & t_{42} & t_{43} & 0 & 0 \\ t_{51} & t_{52} & t_{53} & t_{54} & 0 \end{pmatrix}, P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ p_{21} & 0 & 0 & 0 & 0 \\ p_{31} & p_{32} & 0 & 0 & 0 \\ p_{41} & p_{42} & p_{43} & 0 & 0 \\ p_{51} & p_{52} & p_{53} & p_{54} & 0 \end{pmatrix}$$

First, consider the red circled elements. Here, the copula with conditioned variables indexed by $\{5, 4\}$ and conditioning variables indexed by $\{2, 3\}$, i.e., $c_{5,4|23}$, has copula family t_{31} with corresponding parameter p_{31}. Then, consider the green circled elements. The copula $c_{3,2}$ belongs to the family t_{53} and has parameter p_{53}.

Furthermore, we consider the two examples of a possible D-vine and C-vine as illustrated in Figures 3.8 and 3.9, respectively. The R-vine matrix corresponding to the D-vine structure is given by

$$M = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 \\ 2 & 1 & 3 & 0 & 0 \\ 3 & 2 & 1 & 2 & 0 \\ 4 & 3 & 2 & 1 & 1 \end{pmatrix},$$

and the R-vine matrix corresponding to the C-vine structure is denoted by

$$M = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 4 & 4 & 0 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 2 & 2 & 2 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

By means of the three matrices M, T and P we are able to store all necessary information of an R-vine distribution in a convenient way. Subsequently, we can determine the joint density of the R-vine specification by an algorithm, which can be found in Dissmann et al. (2013).

3.6   Summary

- Copula modeling is robust against non-normal multivariate data.
- Elliptical copulas are widely used and they are able to specify different levels of correlation between the components.
- Archimedean copulas are very popular as they can be constructed with ease and they can express many distributional properties.
- For bivariate copulas there often exists an explicit relation between the parameters and rank correlation coefficients.
- Vine copula structures provide a flexible model for copulas in dimension ≥ 3.

4 | Monitoring approach

In Chapter 2 we briefly described the Shewhart chart for individual observations, which is a univariate monitoring method. As stated in Chapter 1, most processes require multivariate monitoring methods to take into account possible dependencies between the components under investigation. To this purpose, we first discuss two conventional multivariate monitoring approaches in this chapter, namely the Hotelling T^2 chart and the MCUSUM chart. These approaches, however, rely on assumptions that are hardly ever satisfied in practice. To overcome these limitations, we explain the modern vine copula-based monitoring method based on Mühlig (2017). The performance of this vine copula-based approach compared to the conventional methods will be thoroughly analyzed in the next chapter.

4.1 Hotelling T^2 control chart

The Hotelling T^2 chart is one of the most, if not the most, popular control chart for multivariate monitoring. In this section we briefly describe the Hotelling T^2 chart following the exposition in Qiu (2013, Chapter 7). We assume that we have a random sample $(X_1, \ldots, X_n)$ from a p-dimensional population with the distribution $\mathcal{N}_p(\mu, \Sigma)$. Furthermore, let $\overline{X}$ and S^2 denote the sample mean and sample covariance matrix, respectively. Then, the random variable

$$T^2 = n \left(\overline{X} - \mu \right)^T \left(S^2 \right)^{-1} \left(\overline{X} - \mu \right)$$

is called the Hotelling's T^2 statistic, which is basically a generalization of the univariate t-test statistic. The T^2 statistic was first applied for multivariate SPC by Hotelling (1947). For the phase I problem, we assume that we have a phase I dataset $(X_1, \ldots, X_M)$ obtained from a p-dimensional process. We assume that the phase I data is representative for the actual in-control process and that it follows the distribution $\mathcal{N}_p(\mu_0, \Sigma_0)$. If μ_0 and Σ_0 are known, we consider the statistic

$$T_{0,i}^2 = (X_i - \mu_0)^T \Sigma_0^{-1} (X_i - \mu_0), \tag{4.1}$$

where i denotes the i-th time point.

When the process is in-control at the i-th time point, it holds that

$$T_{0,i}^2 \sim \chi_p^2,$$

and, typically, the chart signals when

$$T_{0,i}^2 > \chi_{1-\alpha,p}^2, \tag{4.2}$$

where $\alpha \in [0,1]$ is a given significance level and $\chi_{1-\alpha,p}^2$ is the $(1-\alpha)$-th quantile of the χ_p^2 distribution. In this setup, if we assume that all observations are independent and the process is in-control, the average in-control run length is equal to $1/\alpha$. The average in-control run length is simply the average number of observations which are being measured until a false alarm occurs, see Section 5.1.

In practice, however, the values of μ_0 and Σ_0 are rarely known and they need to be estimated by the sample mean $\overline{X}$ and the sample covariance matrix S^2. Subsequently, the charting statistic becomes

$$T_{1,i}^2 = \left(X_i - \overline{X} \right)' \left(S^2 \right)^{-1} \left(X_i - \overline{X} \right).$$

By Tracy et al. (1992) we know that

$$\frac{M}{(M-1)^2}T_{1,i}^2 \sim Beta(p/2, (M-p-1)/2),$$

and, therefore, the chart would give a signal when

$$T_{1,i}^2 > \frac{(M-1)^2}{M}Beta_{1-\alpha}(p/2, (M-p-1)/2),$$

where $\alpha \in [0,1]$ is a given significance level and $Beta_{1-\alpha}(p/2, (M-p-1)/2)$ is the $(1-\alpha)$-th quantile of the $Beta(p/2, (M-p-1)/2)$ distribution.

Then, for the phase II problem we assume that $X_1, X_2, \ldots$ are incoming observations obtained from a p-dimensional process with in-control distribution $\mathcal{N}_p(\mu_0, \Sigma_0)$. In case both μ_0 and Σ_0 are known, we can simply use the chart (4.1)-(4.2) for online monitoring. In most cases both μ_0 and Σ_0 are unknown and they must be estimated from an in-control dataset. We assume that we have an in-control sample of M observations, and μ_0 and Σ_0 are estimated by the sample mean and sample covariance matrix denoted as $\widehat{\mu}_0$ and $\widehat{\Sigma}_0$, respectively. Then, we can use the following statistic

$$T_{2,i}^2 = (X_i - \widehat{\mu}_0)^T \widehat{\Sigma}_0^{-1} (X_i - \widehat{\mu}_0),$$

where i denotes the i-th point in time.

When the process is in-control at the i-th time point, Tracy et al. (1992) have shown that

$$\frac{(M-p)M}{p(M-1)(M+1)}T_{2,i}^2 \sim F_{p,M-p},$$

where $F_{p,M-p}$ denotes the F distribution with the numerator degrees of freedom p and the denominator degrees of freedom $M-p$. Typically, the resulting chart gives a signal when

$$T_{2,i}^2 > \frac{p(M-1)(M+1)}{(M-p)M}F_{1-\alpha,p,M-p},$$

where $F_{1-\alpha,p,M-p}$ is the $(1-\alpha)$-th quantile of the $F_{p,M-p}$ distribution.

4.2 MCUSUM control chart

The Hotelling T^2 chart uses test statistics for a given time point which only depend on the observed data at that particular time point. Just as in the univariate case, such control charts are effective for detecting relatively large and transient shifts. In this section we introduce the multivariate CUSUM (MCUSUM) control chart, which seems to be more effective for detecting relatively small and persistent shifts (see, e.g., Qiu (2013)). There exist many different versions of the MCUSUM control chart. In this section we only focus on one particular version that receives much attention in literature, namely the chart proposed by Crosier (1988) since this chart is demonstrably effective for detecting shifts in the process mean. Alternative versions of the MCUSUM chart can be found in Qiu (2013, Chapter 7). We will only discuss the phase II problem as these charts are typically used in phase II analysis, although they can also be used in phase I analysis.

We assume that $X_1, X_2, \ldots$ are phase II observations obtained from a p-dimensional process with in-control distribution $\mathcal{N}_p(\mu_0, \Sigma_0)$, where both μ_0 and Σ_0 are known. The charting statistic C_n, for $n \geq 1$, depends on the n-th observation as well as earlier observations and is defined as follows. Let

$$U_n = \begin{cases} 0, & \text{if } Y_n \leq k \\ (U_{n-1} + X_n - \mu_0)(1 - k/Y_n), & \text{otherwise,} \end{cases}$$

where $U_0 = 0$, $k > 0$ is an allowance constant, and

$$Y_n = \left[(U_{n-1} + X_n - \mu_0)^T \Sigma_0^{-1} (U_{n-1} + X_n - \mu_0) \right]^{1/2}.$$

Furthermore, the chart signals at the n-th time point if

$$C_n = \left(U_n^T \Sigma_0^{-1} U_n \right)^{1/2} > h,$$

where $h > 0$ is the control limit chosen in order to reach a pre-specified in-control average run length.

In cases where both μ_0 and Σ_0 are unknown, they can simply be estimated by the sample mean $\widehat{\mu}_0$ and sample covariance matrix $\widehat{\Sigma}_0$ of an in-control dataset. Furthermore, when we consider a mean shift, k is usually set at $\frac{1}{2}\delta$ where δ is the size of the shift in standard deviation units. Since we obviously do not know the shift size in advance, the allowance parameter k should be chosen such that we detect a desirable shift as soon as possible. If we want to quickly detect relatively small shifts, we could for instance set $k = 0.25$. On the other hand, if we want to detect relatively large shifts, an optimal choice could be to set $k = 0.75$. Once k is selected we should choose h to reach the desired in-control average run length. For more information on the allowance parameter k, see, e.g., Qiu (2013, Section 4.2.2).

4.3   Vine copula-based control chart

The conventional multivariate monitoring methods, like the ones described in the previous sections, are based on the assumptions that observed values follow the multivariate normal distribution and do not exhibit serial correlation. Real-life data sets almost never satisfy these assumptions and the efficiency of these conventional methods will be harmed. Therefore, we introduce the modern vine copula-based approach developed by Mühlig (2017) which is suitable for non-normal multivariate data. The aim of the approach is to timely detect changes in the dependence structure of a multivariate distribution, as well as changes in its mean or covariance matrix. This refreshing approach focuses on combining vine copula structures with the SPC methodology and basically extends the ideas of Baíllo and Cuevas (2006) and Verdier (2013).

4.3.1   Tolerance region approach

The basic idea of the vine copula-based monitoring approach follows Baíllo and Cuevas (2006). We assume that the quality characteristics vector $X = \left(X^{(1)}, \ldots, X^{(d)} \right)$ is a d-dimensional random vector with continuous marginal distribution functions $F_1, \ldots, F_d$, which describes a d-dimensional process independent over time. When the process is in-control, we assume that H and h denote the in-control distribution function and density of X, respectively. When a change in the process occurs, the process does no longer follow the in-control distribution and can be declared as out-of-control. For each new observation, we have to decide whether it was generated from H (the process is still in-control) or from $G \neq H$ (the process is out-of-control). We base our decision on a set of i.i.d. observations $X_1, \ldots, X_n$ drawn from H. More specifically, we construct a tolerance region by means of density level set estimation. If a new observation falls outside the tolerance region, the process is statistically out-of-control. This tolerance region is constructed such that the false alarm probability is equal to a pre-specified level α. Following Verdier (2013), the tolerance region is defined as

$$D(c_\alpha) = \left\{ x \in \mathbb{R}^d | h(x) \geq c_\alpha \right\},$$

where c_α satisfies

$$I = P\left[X \in D(c_\alpha) \right] = \int_{D(c_\alpha)} h(x) dx = 1 - \alpha.$$

In practice, the density function h is often unknown and has to be estimated. We assume for now that the unknown function h can be estimated by h_n. Then, the value c_α will be estimated by c_n, which satisfies

$$I_n = \int_{\{h_n \geq c_n\}} h_n(x) dx = 1 - \alpha \quad \text{a.s.}$$

Consequently, the estimated tolerance region is

$$D_n(c_n) = \left\{ x \in \mathbb{R}^d | h_n(x) \geq c_n \right\}.$$

By this construction it follows that

$$P\left(h_n(X) \geq c_n\right) = 1 - \alpha,$$

and therefore c_n is the α-th quantile of $h_n(X)$. The main objective is to determine the value of c_n. Cadre et al. (2013) propose to use the α-th quantile of the empirical distribution of $h_n(X_1), \ldots, h_n(X_n)$ and the resulting estimator is denoted as $\widehat{c}_n$. Furthermore, Verdier (2013) improved the estimation and proposes to use an empirical quantile computed on N random variates obtained from the estimated density h_n, where N is a large number. For this purpose, let $\widehat{F}_{h_n(X)}$ be the distribution function of a random sample of $h_n(X)$, denoted by $h_n(X_{1,n}), \ldots, h_n(X_{N,n})$. Then, it holds that

$$\widehat{F}_{h_n(X)}(y) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left(h_n\left(X_{i,n}\right) \leq y\right), \tag{4.3}$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Subsequently, the estimator of c_α is obtained by

$$\widehat{c}_{N,n} = \inf\left\{c \in \mathbb{R} : \widehat{F}_{h_n(X)}(c) \geq \alpha\right\}.$$

Following Cadre et al. (2013), we define the following set and assumption:

$$\mathscr{H}_0 = \left\{c \in (0; \sup h) : \inf_{\{h=c\}} ||\nabla h|| = 0\right\}$$

Assumption 4.1.

1. *The density h is of class C^2 with a bounded Hessian matrix, and $h(x) \to 0$ as $||x|| \to \infty$.*

2. *$\mathscr{H}_0$ has Lebesgue measure 0.*

3. *$\lambda(\{h = c\}) = 0$ for all $c > 0$.*

See Cadre et al. (2013) for a more detailed explanation concerning Assumption 4.1.

The estimation approach by Verdier (2013) relies on the following theorem.

Theorem 4.1. *Suppose h satisfies Assumption 4.1 and that*

$$\sup_{x \in \mathbb{R}^d} |h_n(x) - h(x)| \to 0 \quad p.s.$$

Then, for almost all $c \in (0, 1)$ (see Verdier (2013) for details),

$$\widehat{c}_{N,n} \to c_\alpha \quad in\ probability\ for\ N \to \infty.$$

Proof. See the proof of Theorem 2 of Verdier (2013). $\square$

4.3.2 Monitoring with vine copulas

We now have enough knowledge to build a statistical framework around the copula methodology. In this section we construct a monitoring approach based on vine copulas, which provide more flexibility than regular copula models, see also Section 3.5. The monitoring framework will follow the tolerance region approach by Verdier (2013) as described above, combined with a Shewhart-like control chart approach (see Section 2.1).

We assume that a d-dimensional process is described by the quality characteristics vector $X = \left(X^{(1)}, \ldots, X^{(d)}\right)$ with continuous marginals $F_1, \ldots, F_d$ and corresponding densities $f_1, \ldots, f_d$. Furthermore, let H and h denote the in-control distribution function and density of X, respectively. First, we need a suitable data set for phase I, in which we assume that the process is in-control. We assume that we have a phase I sample $X_1, \ldots, X_n$ consisting of n observation vectors from X, which follow the in-control distribution. For convenience of notation, we assume that the sample is stored in a data matrix $R \in \mathbb{R}^{n \times d}$, where row $R_{i\bullet} \in \mathbb{R}^{1 \times d}$ corresponds to the i-th observation vector, $i = 1, \ldots, n$, and column $R_{\bullet j} \in \mathbb{R}^{n \times 1}$ corresponds to the j-th quality characteristic,

$j = 1, \ldots, d$. Note that we often have to transform the sample by the probability integral transform to satisfy the requirement of standard uniform marginals. That is, we transform the sample of the j-th quality characteristic $R_{\bullet j}$ by

$$U_{\bullet j} := F_j(R_{\bullet j}), \quad j = 1, \ldots, d. \tag{4.4}$$

When the marginals $F_1, \ldots, F_d$ are unknown, we can use so-called pseudo-observations which basically apply the empirical distribution functions to the data and scale the result such that the transformed sample falls inside the unit interval. For this approach we assume that $d \geq 2$, since most processes require to monitor multiple quality characteristics simultaneously. Note that for $d = 2$ we can simply use a bivariate copula model, since the vine copula model reduces to only one regular bivariate pair-copula. Now, we are able to determine the best-fitting vine copula model by the procedure described in Section 3.5.3. Based on the selected vine copula, we can estimate the joint density function h by (3.20) and the resulting estimated density function is denoted by h_n. Subsequently, we calculate a random sample of $h_n(X)$, denoted by $h_n(X_{1,n}), \ldots, h_n(X_{N,n})$, where N is really large. This density sample serves as the base of the quantile estimation by Verdier (2013) as described in the section above. The density sample follows the distribution function $\widehat{F}_{h_n(X)}$ as defined by (4.3) and we can calculate the corresponding control limits dependent on whether we want to use the one-sided or two-sided version of the control chart. For the one-sided version, the (lower) control limit $\widehat{c}_{N,n}$ is defined as the α-th quantile of the density sample $h_n(X_{1,n}), \ldots, h_n(X_{N,n})$, where α is the pre-specified false alarm probability. For the two-sided version, the upper control limit $\widehat{u}_{N,n}$ and lower control limit $\widehat{l}_{N,n}$ are defined as the $(1 - \frac{\alpha}{2})$-th quantile and $(\frac{\alpha}{2})$-th quantile of the density sample $h_n(X_{1,n}), \ldots, h_n(X_{N,n})$, respectively. For phase I analysis, the test statistics corresponding to the sample $X_1, \ldots, X_n$ are given by $h_n(U_{1\bullet}), \ldots, h_n(U_{n\bullet})$ and an alarm for the one-sided version is raised when

$$h_n(U_{i\bullet}) < \widehat{c}_{N,n}, \quad i = 1, \ldots, n,$$

and an alarm for the two-sided version is raised when

$$h_n(U_{i\bullet}) < \widehat{l}_{N,n} \quad \text{or} \quad h_n(U_{i\bullet}) > \widehat{u}_{N,n}, \quad i = 1, \ldots, n.$$

For phase II analysis, we have incoming observation vectors $X_1, X_2, \ldots$ from a d-dimensional process and we need to decide if they are generated from the in-control distribution. Again, we first need to transform the incoming observations by either the probability integral transform or pseudo observations. Similar to (4.4), let $U_{k\bullet}$, $k = 1, 2, \ldots$, denote the transformation of the k-th observation vector. Then, the test statistics corresponding to the incoming observations $X_1, X_2, \ldots$ are given by $h_n(U_{1\bullet}), h_n(U_{2\bullet}), \ldots$, where h_n is the estimated joint density function obtained from phase I. Subsequently, we compare these density evaluations with the control limits that we determined in phase I. For the one-sided version, we raise an alarm when

$$h_n(U_{k\bullet}) < \widehat{c}_{N,n}, \quad k = 1, 2, \ldots,$$

and for the two-sided version we raise an alarm when

$$h_n(U_{k\bullet}) < \widehat{l}_{N,n} \quad \text{or} \quad h_n(U_{k\bullet}) > \widehat{u}_{N,n}, \quad k = 1, 2, \ldots$$

The one-sided and two-sided version of the vine copula-based control chart seem to focus on different aspects of the multivariate distribution regarding the detection of shifts, which we clarify with an example. In this example we consider $d = 2$ for visualization purposes, so the vine copula model described above simplifies to the bivariate copula model. We assume that an in-control 2-dimensional process can be described by the Gumbel copula with parameter $\theta = 3$ and that we have perfect knowledge of the in-control distribution. We look at the phase I problem, where we have an in-control sample $X_1, \ldots, X_{1000}$ generated from the in-control distribution. For the calculation of the control limits, we take $N = 10^6$ and $\alpha = 0.05$. We obtain the test statistics corresponding to the in-control sample and appropriate control limits by applying the copula-based monitoring method that is described above. Subsequently, we identify the out-of-control points of the in-control sample and mark these points in red. The results for the one-sided and two-sided version are illustrated in Figures 4.1 and 4.2, respectively. From these figures we observe that the one-sided version signals observations which are too far away from the mean line and the two-sided

version signals observations that are even further away from the mean line plus observations that are suspicious in the upper tail and lower tail of the distribution. Therefore, it seems that the lower control limit aims at detecting deviations from the mean line and the upper control limit aims at detecting deviations in tail dependence of the distribution. Mühlig (2017) studied the working of the one-sided and two-sided version in more detail, including some illustrative examples. The conclusion of the above example is in line with the findings of Mühlig (2017). In the following chapter we study the advantages and disadvantages of the one-sided and two-sided version by means of an extensive performance study.
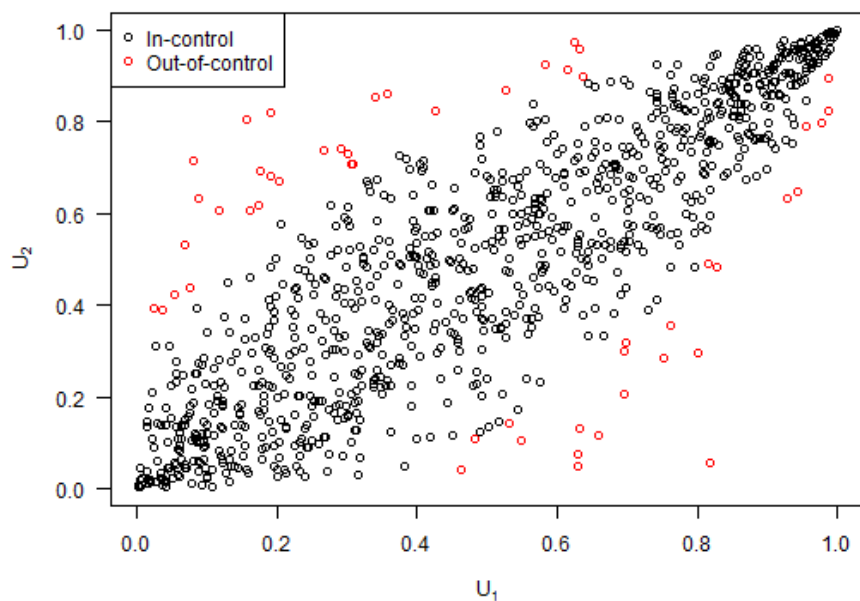

Figure 4.1: Phase I analysis with the one-sided version of the copula-based control chart.
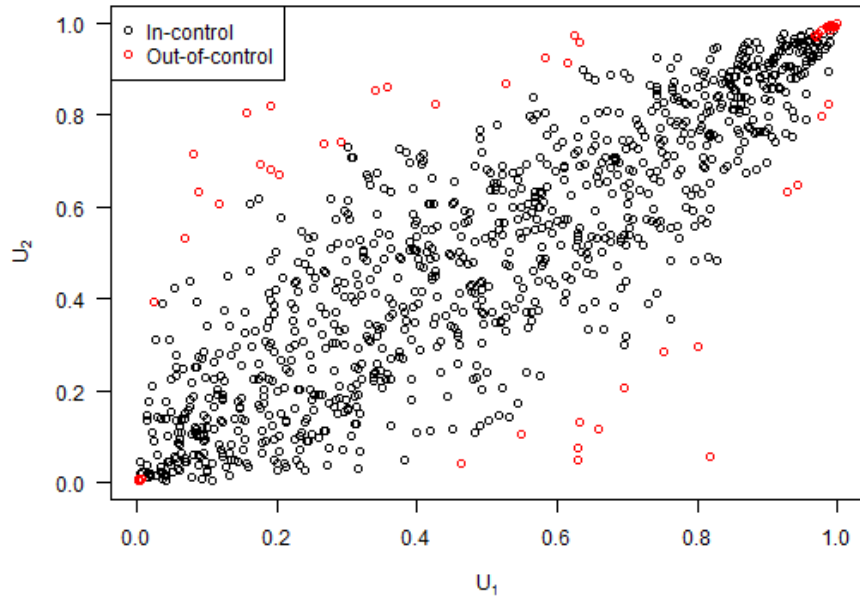
Figure 4.2: Phase I analysis with the two-sided version of the copula-based control chart.

4.4 Summary

- It is possible to construct a Shewhart-like control chart based on vine copulas.
- The vine copula-based control chart is a distribution free method and therefore suitable for non-normal multivariate data.
- The two-sided version of the vine copula-based control chart aims at detecting deviations from the mean line or deviations in tail dependence of a multivariate distribution.
- The one-sided version of the vine copula-based control chart only aims at detecting deviations from the mean line of a multivariate distribution.

5 | Simulation study

In this chapter we conduct a simulation study to assess the performance of the recently introduced copula-based control chart, where we make use of the average run length measure. First, we compare the performance of the one-sided version of the copula-based control chart with the two-sided version. In this way we are able to recommend which version one should use in order to detect a particular type of shift in a process. Subsequently, we analyze the performance of the copula-based control chart compared to the conventional Hotelling T^2 and MCUSUM control charts. These analyses follow the ideas of Mühlig (2017), but are more comprehensive in the sense that more possibilities are included. A comparable simulation study is performed in Sukparungsee et al. (2017), although the context is somewhat different. The similarities include a performance analysis based on the average run length, where several copulas are fitted and various controlled out-of-control cases are analyzed. There are also some significant differences. First of all, Sukparungsee et al. (2017) simulated a three-dimensional in-control process from the exponential distribution. Furthermore, they only investigated the possibility of mean shifts in the marginal distributions. Finally, they only analyzed the performance of the MCUSUM control chart and hence were not able to compare control charts.

5.1 Performance measures

When we apply control charts in phase I we conduct retrospective analysis. During this phase we mainly focus on calculating appropriate control limits for phase II, such that the false alarm rate will be low. Additionally, control charts in phase II should be able to quickly detect an out-of-control situation. In order to measure these performance requirements, we first need to introduce the concept of average run length (ARL). For an in-control process, the ARL equals the average number of observations that are measured until a false alarm occurs and is denoted by ARL_0. For an out-of-control process, the ARL equals the average number of observations that are measured until an out-of-control situation is detected and is denoted by ARL_1. Ideally, we want the ARL_0 to be large and the ARL_1 to be small to ensure a low false alarm rate and fast detection of out-of-control situations. However, there seems to be a trade-off between the ARL_0 and the ARL_1. Clearly, increasing the ARL_0 by aiming for wider control limits results in an undesired increase of the ARL_1. On the other hand, decreasing the ARL_1 by aiming for tighter control limits results in an unwanted decrease of the ARL_0. Therefore, we need to find an optimal balance between the ARL_0 and ARL_1. In SPC optimality is often stated as minimal ARL_1 for a pre-specified ARL_0, see, e.g., Frisén (2007). Advantages of the ARL criterion are that it is very well known, widely used and simple, and that no assumption of a distribution for the time of change is used. However, these measures do not tell the whole story as run length distributions are often very skewed. Additionally, estimation of parameters might influence the run length calculations. Alternatively, one could use the conditional expected delay (CED) of detection, which represents the delay from the first opportunity to detect a change. For a more extensive discussion on this topic, see Frisén (2003) and Kenett and Pollak (2012).

Next, we use the basic univariate Shewhart chart for individual observations defined in Section 2.1 as an example to show how its ARL_0 and ARL_1 values can be computed. It is not hard to conclude that the distribution of the ARL_0 is the geometric distribution $Geom(\alpha)$, where α denotes the false alarm probability. We assume that both the in-control mean μ_0 and the in-control standard deviation σ are known. Under the normality assumptions, the false alarm probability can be easily

computed by

$$\alpha = P(X_i < \mu_0 - 3\sigma) + P(X_i > \mu_0 + 3\sigma)$$
$$= P\left(\frac{X_i - \mu_0}{\sigma} < -3\right) + 1 - P\left(\frac{X_i - \mu_0}{\sigma} < 3\right)$$
$$= \Phi(-3) + 1 - \Phi(3)$$
$$\approx 0.0027.$$

By the formula of the mean of a geometrically distributed random variable, we simply obtain that

$$ARL_0 = \frac{1}{\alpha} \approx 370.$$

For computing the ARL_1, we now assume that the mean of the quality characteristic shifts from μ_0 to $\mu_1 = \mu_0 + k\sigma$. It is obvious that the distribution of the ARL_1 is the geometric distribution $Geom(1 - \beta)$, where β is the probability that an observation collected at a later time point is within the upper and lower control limits. This probability can be expressed in terms of k as

$$\beta = P\left(\mu_0 - 3\sigma \le X_i \le \mu_0 + 3\sigma\right)$$
$$= P\left(-k - 3 \le \frac{X_i - \mu_1}{\sigma} \le -k + 3\right)$$
$$= \Phi\left(-k + 3\right) - \Phi\left(-k - 3\right).$$

By the formula of the mean of a geometrically distributed random variable, we know that

$$ARL_1 = \frac{1}{1 - \beta}.$$

The values of ARL_1 for $k \in [0, 3]$ are shown in Figure 5.1, from which we can conclude that the Shewhart chart for individual observations is only effective in detecting large shifts in the mean of a process. The calculations in this example are quite straightforward. In the multivariate case, however, these performance measures can very often not be expressed in closed form. To overcome this issue we can make use of Monte Carlo simulation, which is explained in the following section.
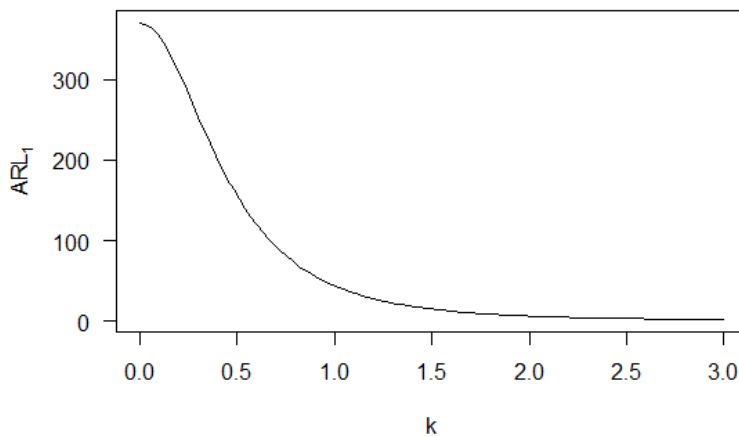
Figure 5.1: The values of ARL_1 for the Shewhart chart for individual observations when $k \in [0, 3]$.

5.2 Bivariate performance analysis

In this section we create empirical results to investigate the performance of the copula-based control chart for non-normal bivariate data. Firstly, we explain the simulation setup in detail. Secondly, we

decide on which version of the copula-based control chart to use in order to detect a shift of desired type. Finally, we compare the performance of the copula-based control chart with conventional methods such as Hotelling T^2 and MCUSUM control charts.

5.2.1 Simulation approach

We assume that some in-control bivariate process can best be described by a bivariate copula family $\in$ {Gumbel, Clayton, Frank, Normal, Student-t}. From now on, we will call this bivariate copula family the in-control copula distribution. We assume that we have perfect knowledge of the in-control copula distribution. The simulation procedure for a selected monitoring method $\in$ {Hotelling T^2, MCUSUM, vine copula} is as follows:

Phase I

For $t = 0.1, 0.2, \ldots, 0.9$:
1. Draw 2,000 random samples from the in-control copula distribution with $\tau = t$ to build the training data.
2. Based on the training data and selected monitoring method, estimate required quantities and calculate the corresponding test statistics.
3. Calculate the control limit(s) based on the training data and selected monitoring method, where we set $\alpha = 0.0027$. For the MCUSUM method the control limit is simply equal to the chosen value of the parameter h.
4. Determine the run length value, which is equal to the index of the first occurrence of an out-of-control signal.
5. Repeat steps 1-4 1,000 times to obtain an accurate estimate of the ARL_0.

Here, the value of 0.0027 is the false alarm probability for the basic univariate Shewhart chart for individual observations in order to reach $ARL_0 \approx 370$, see Section 5.1. Note that for the MCUSUM method we do not use α, but a pre-specified control limit h instead. In this case we should choose h such that we reach the desired ARL_0, see also Section 4.2. We can often determine a suitable value for h by trial and error. Furthermore, we vary the input parameter of the in-control copula distribution over a sequence of possible values. More specifically, we repeat the phase I procedure for Kendall's tau $\tau \in \{0.1, 0.2, \ldots, 0.9\}$ such that we are able to indicate the strength of dependence. The corresponding input parameter values can be easily calculated by an explicit function, see Section 3.4. Also, in this way we are able to determine the τ-value for which the ARL_0 of the methods that we want to compare are approximately equal. Subsequently, we use this particular τ-value to reach our pre-specified ARL_0 and move on to phase II, in which we manually implement three possible shifts in the process.

Phase II
1. Use the 2,000 random samples from the in-control copula distribution with τ such that the pre-specified ARL_0 is reached, and manually implement a desired shift in the process to build the test data.
2. Based on the selected monitoring method, use the estimated quantities from phase I and calculate the corresponding test statistics for the test data.
3. Use the control limit(s) that are calculated in phase I. Note that for the Hotelling T^2 chart, however, the control limit is slightly different and needs to be recalculated.
4. Determine the run length value, which is equal to the index of the first occurrence of an out-of-control signal.
5. Repeat steps 1-4 1,000 times to obtain an accurate estimate of the ARL_1.

The three possible shifts include a shift in the mean, variance and dependence structure of the in-control copula distribution. After such distributional shift the process is assumed to be out-of-control. Clearly, since the ARL_0 of the control charts that we want to compare are approximately equal, the chart with minimum ARL_1 value performs best. We will now elaborate on the three possible type of shifts and how they can be implemented.

Shift in mean

The first potential shift that we are going to investigate is a shift in the mean of the in-control copula distribution. We can implement such a shift by simply changing the level of one or both variables. More precisely, we use the 2,000 random samples $\mathbf{U} = (U_1, U_2)$ from the in-control copula distribution and subsequently implement a level shift, which is constructed as follows:

$$\mathbf{U}_{\text{shift}} = (U_{\text{shift},1}, U_{\text{shift},2}) = ((1 + s_1) \cdot U_1, (1 + s_2) \cdot U_2),$$

where s_j $(j = 1, 2)$ denotes the percentage by which the level of the j-th variable is shifted. The level increases when s_j is positive and, conversely, the level decreases when s_j is negative. Note that we have to ensure that the shifted samples remain from a copula distribution. Therefore, points that fall outside the unit square $[0, 1]^2$ are relocated to the boundary of the unit square.

In order to visualize what such a level shift looks like in terms of copula and its components, we look at an example where we assume that the in-control distribution follows a Gumbel copula with parameter $\theta = 5$. Various shifts in the level of the first variable U_1, obtained according to the procedure described above, are shown in Figure 5.2. From a numerical point of view, the mean and Kendall's τ correlation coefficient of the two variables are shown in Table 5.1. From this table we conclude that, in general, the mean indeed shifts according to the intended percentage while the correlation remains the same. In case of an increase in the level, the mean and correlation slightly differ from the expected values. This can be explained by the points that fall outside the unit square, as relocating them to the boundary affects the original mean and correlation.

For an extensive simulation study we repeat the phase II procedure for $s_j \in \{10, 20, \ldots, 90\}$ and $s_j \in \{-90, -80, \ldots, -10\}$ to cover both level increases and decreases. In this way we can distinguish between small and large shifts in the mean of the in-control copula distribution.

(a) Level increased by 20%.

(b) Level increased by 80%.

(c) Level decreased by 20%.

(d) Level decreased by 80%.
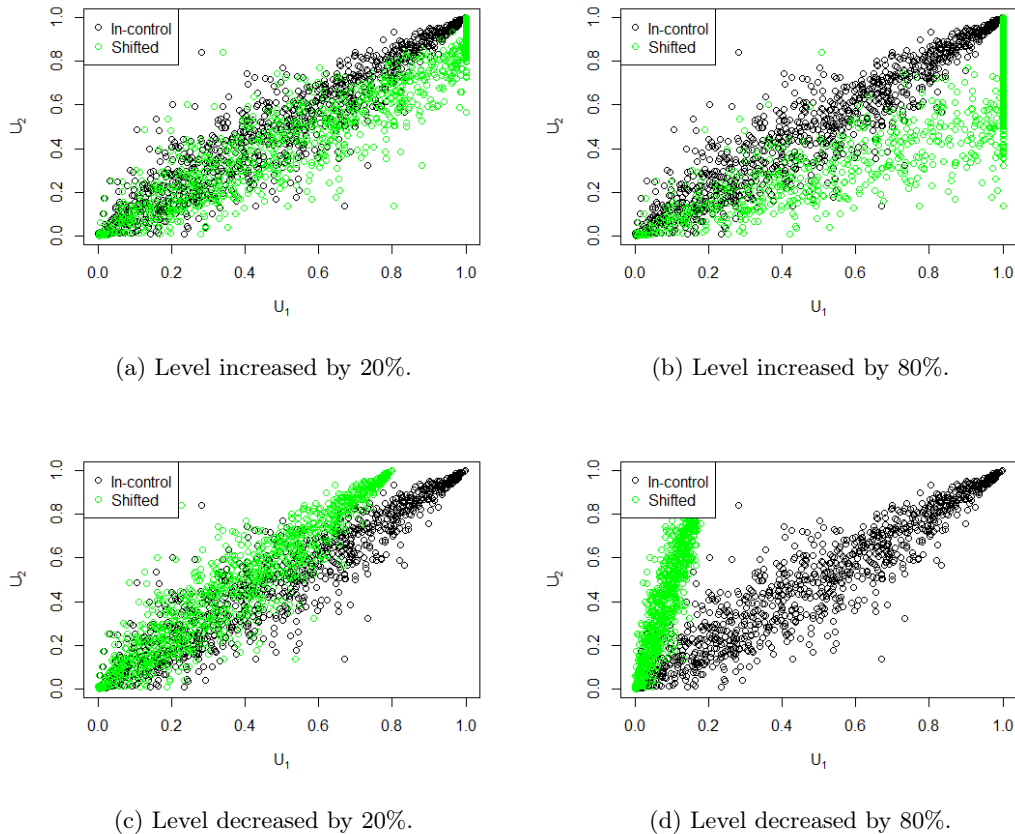
Figure 5.2: Illustration of a shift in the mean by changing the level of the first variable.

	Mean	Correlation (τ)
In-control	$(0.5009, 0.4982)$	0.7901
20% increase	$(0.5851, 0.4982)$	0.7821
80% increase	$(0.7267, 0.4982)$	0.7242
20% decrease	$(0.4007, 0.4982)$	0.7901
80% decrease	$(0.1002, 0.4982)$	0.7901

Table 5.1: Mean and correlation coefficient after various shifts in the level of the first variable.

Shift in variance

The second potential shift is a shift in the variance of the in-control copula distribution. Before we go into more details, we elaborate on how we describe the variance of the distribution. The variance can either be described by covariance or correlation. Both terms are able to measure the relationship and dependency between two variables. Covariance measures the direction of the linear relationship between two variables. Correlation, however, measures both the direction and strength of the linear relationship between two variables. Basically, correlation is the standardized version of covariance and is therefore dimensionless. Another difference is that correlation coefficients lie between -1 and $+1$, while covariance varies from $-\infty$ to $+\infty$. The sign of a correlation coefficient represents the direction of the correlation and the closer it is to $+1$ or -1, the stronger the dependence between the two variables. Additionally, we have seen that correlation coefficients can be easily related to the input parameter of a copula distribution (see Section 3.4 for more details). For the above reasons, we express the variance of a copula distribution by correlation coefficients from now on.

If we are dealing with n random variables $X_1, \ldots, X_n$, the correlation coefficients ρ_{ij} between different pairs (i, j) of random variables are expressed in a correlation matrix. Correlation matrices must satisfy the following properties $(i, j = 1, \ldots, n)$:

- All entries must lie in the interval $[-1, 1]$: $-1 \leq \rho_{ij} \leq 1$
- The diagonal terms are equal to one: $\rho_{ii} = 1$
- The matrix is symmetric: $\rho_{ij} = \rho_{ji}$
- The correlation matrix Z is positive *semidefinite*: $x^T Z x \geq 0$ for all $x \in \mathbb{R}^n$

It is not straightforward to implement a shift in the correlation matrix, since the requirements can be easily unsatisfied after a shift. Also, a 10% increase in a strong correlation does have much more impact than a 10% increase in a weak correlation. We will illustrate this with a simple example.

Consider 1,000 random samples from a bivariate Normal copula. First, we assume that Kendall's τ equals 0.1. When τ increases with 10%, this coefficient becomes 0.11. Second, we assume that Kendall's tau is equal to 0.9. In case τ increases with 10%, this coefficient becomes 0.99. Both shifts are also visualized in Figure 5.3. From this figure, we confirm that the same percentage increase does have a significantly higher impact on strong correlation rather than weak correlation.
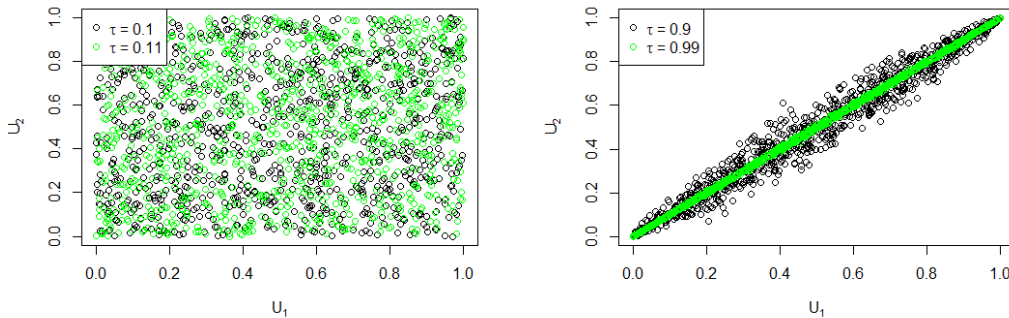
(a) $\tau = 0.1$ increased by 10%.

(b) $\tau = 0.9$ increased by 10%.

Figure 5.3: Illustration of an increase in Kendall's τ correlation coefficient.

To overcome these difficulties, we propose a method to shift the correlation matrix that is based on Galeeva et al. (2012). Suppose that we have a correlation matrix $M = \begin{pmatrix} 1 & \tau_{12} \\ \tau_{12} & 1 \end{pmatrix}$ based on Kendall's τ correlation coefficient. We consider the following steps to shift the correlation matrix M:

1. Decompose M into terms of its eigensystem (D, U) via

$$M = UDU^{-1} = UDU^T, \tag{5.1}$$

where U is the square 2×2 matrix whose i-th column is the eigenvector u_i of M, and D is the diagonal matrix whose diagonal elements D_{ii} are the corresponding eigenvalues λ_i. Furthermore, we assume that $\lambda_1 \geq \lambda_2 \geq 0$. Note that since M is real and symmetric, the matrix U is orthogonal, i.e., $U^{-1} = U^T$.

2. Shift the diagonal matrix D such that

$$D^* = \begin{pmatrix} \lambda_1^* & 0 \\ 0 & \lambda_2 \end{pmatrix}, \tag{5.2}$$

where $\lambda_1^* = (1 + s_1) \cdot \lambda_1$, and s_1 denotes the percentage by which the largest eigenvalue λ_1 is shifted.

3. Construct the shifted matrix $M^* = \begin{pmatrix} \tau_{11}^* & \tau_{12}^* \\ \tau_{21}^* & \tau_{22}^* \end{pmatrix}$ by

$$M^* = UD^*U^T. \tag{5.3}$$

4. To ensure that the shifted matrix is a correlation matrix, define

$$V = \begin{pmatrix} \sqrt{\tau_{11}^*} & 0 \\ 0 & \sqrt{\tau_{22}^*} \end{pmatrix}, \text{ and construct } M_{\text{shift}} = V^{-1}M^* \left(V^{-1} \right)^T. \tag{5.4}$$

Note that since V is symmetric, V^{-1} is symmetric as well and hence $\left(V^{-1} \right)^T = V^{-1}$.

It can be easily verified that the shifted matrix M_{shift} indeed satisfies the requirements of a correlation matrix, see Lemma A.2.

The variance shift procedure is as follows. We use the 2,000 random samples $\mathbf{U} = (U_1, U_2)$ from the in-control copula distribution and estimate the correlation matrix M of this sample. Subsequently, we shift the correlation matrix M according to the procedure described above to obtain M_{shift}. We extract the shifted correlation coefficient from the matrix M_{shift}, which can then be easily transformed to the corresponding shifted input parameter. Then, we simply draw 2,000 random samples from the shifted copula distribution with shifted parameter to obtain $\mathbf{U}_{\text{shift}} = (U_{\text{shift},1}, U_{\text{shift},2})$.

To illustrate what such a shift looks like in terms of copula and its components, we consider the following example. Assume that the in-control distribution follows a Gumbel copula with parameter $\theta = 5$. Figure 5.4 exemplifies several shifts in the correlation matrix by changing the largest eigenvalue according to the procedure described above. We see that, as one would expect, the observations become more correlated in case of an increase and less correlated in case of a decrease. The numerical values of the mean and Kendall's τ correlation coefficient of the two variables are shown in Table 5.2. From this table we conclude that the correlation indeed shifts according to the change of the largest eigenvalue, while the mean stays approximately the same.

To present an extensive simulation study we repeat the phase II procedure for $s_1 \in \{10, 20, \ldots, 90\}$ and $s_1 \in \{-90, -80, \ldots, -10\}$ to cover both correlation increases and decreases. In this way we are able to distinguish between small and large shifts in the correlation coefficient.

(a) Largest eigenvalue increased by 20%.

(b) Largest eigenvalue increased by 80%.

(c) Largest eigenvalue decreased by 20%.

(d) Largest eigenvalue decreased by 80%.
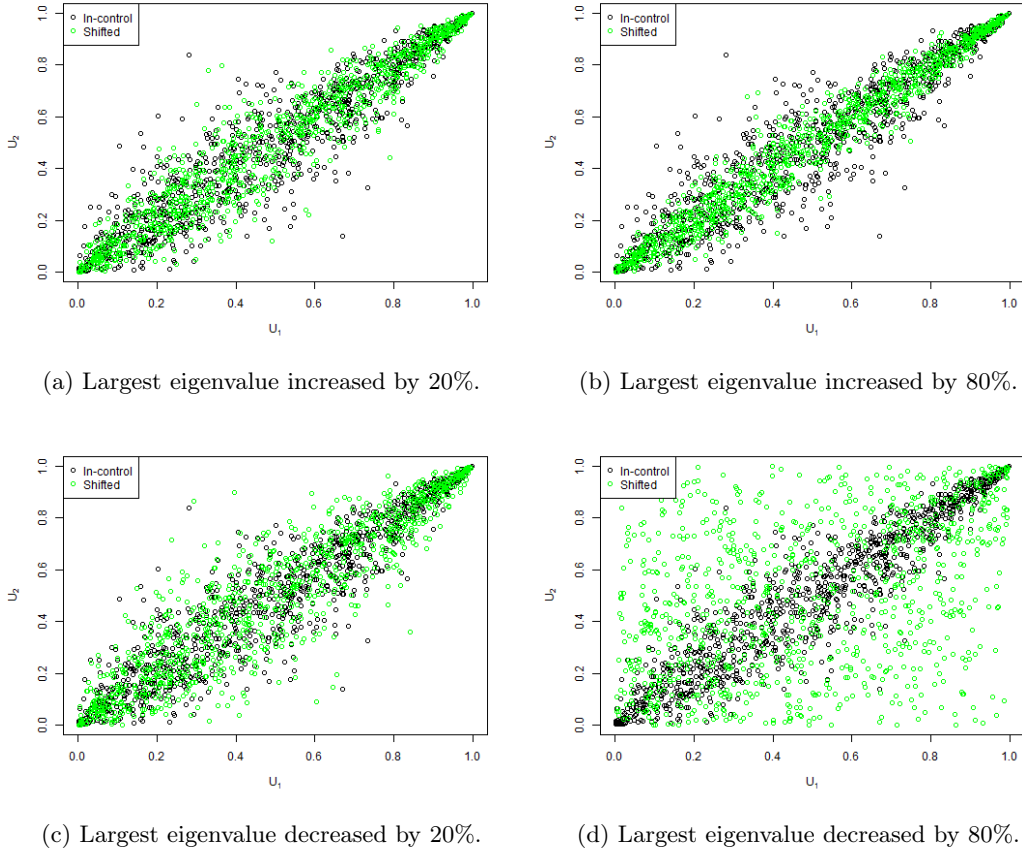
Figure 5.4: Illustration of a shift in the correlation matrix by changing the largest eigenvalue.

	Mean	**Correlation** (τ)
In-control	$(0.5009, 0.4982)$	0.7901
20% increase	$(0.5015, 0.5025)$	0.8291
80% increase	$(0.5029, 0.5035)$	0.8830
20% decrease	$(0.4999, 0.5012)$	0.7560
80% decrease	$(0.4933, 0.4964)$	0.2728

Table 5.2: Mean and correlation coefficient after various shifts in the largest eigenvalue of the correlation matrix.

Shift in dependence structure

The third and final possible shift that we will analyze is a shift in the dependence structure of the in-control copula distribution. This shift can be realized by simply changing the underlying copula structure from the in-control copula to a different, out-of-control, copula. More specifically, we use the 2,000 random samples $\mathbf{U} = (U_1, U_2)$ from the in-control copula distribution. Then, we determine the correlation coefficient τ based on which we calculate the input parameter corresponding to a particular shifted copula distribution. Subsequently, we draw 2,000 random samples from this shifted copula distribution to obtain $\mathbf{U}_{\text{shift}} = (U_{\text{shift},1}, U_{\text{shift},2})$.

We look at a concrete example to visualize what such a shift looks like in terms of copula and its components. We assume that the in-control distribution follows a Gumbel copula with parameter $\theta = 5$. Figure 5.5 illustrates what several shifts in the underlying copula structure look like. Also, the mean and Kendall's τ correlation coefficient of the two variables are shown in Table 5.3. From this table we conclude that the mean and correlation remain approximately the same after a shift in copula structure has occurred.

(a) Shift to Clayton copula.

(b) Shift to Frank copula.

(c) Shift to Normal copula.

(d) Shift to Student-t copula with 3 degrees of freedom.
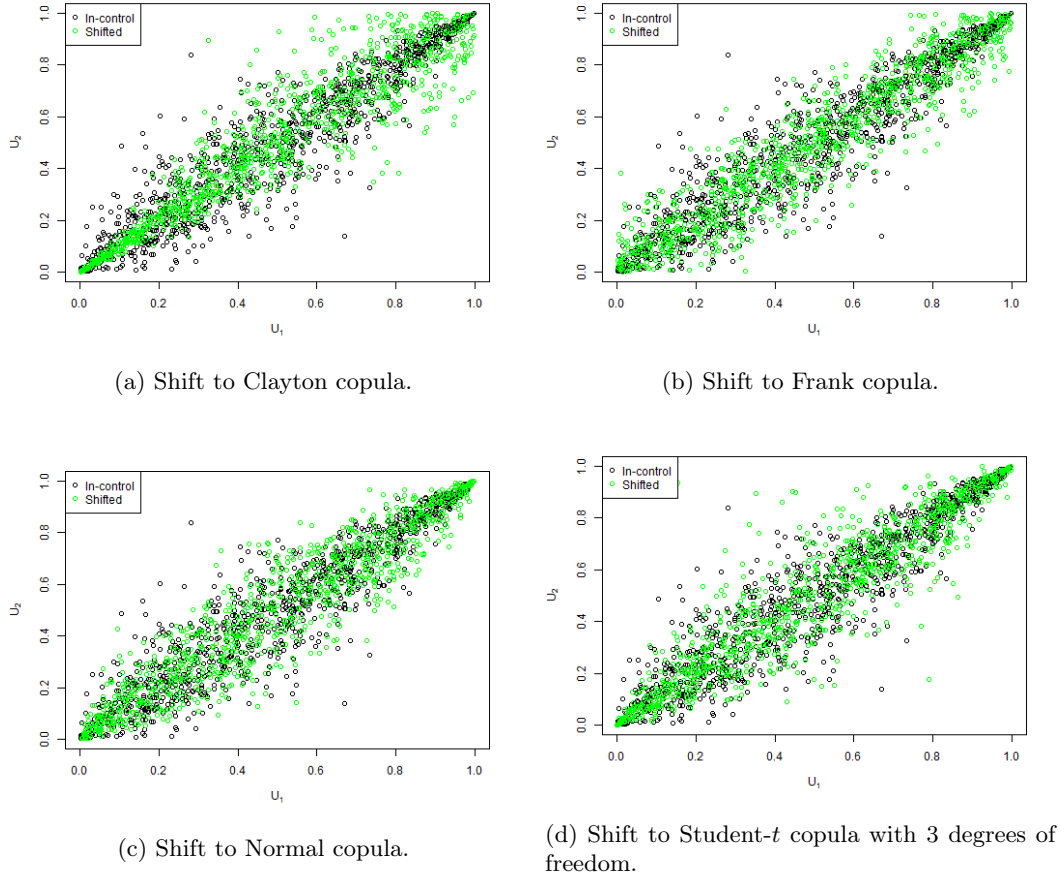
Figure 5.5: Illustration of a shift in the underlying copula structure.

	Mean	**Correlation (τ)**
In-control	$(0.5009, 0.4982)$	0.7901
Clayton	$(0.4935, 0.4986)$	0.7877
Frank	$(0.4935, 0.4902)$	0.7862
Normal	$(0.4974, 0.4962)$	0.7877
Student-t (df $= 3$)	$(0.4981, 0.4995)$	0.7856

Table 5.3: Mean and correlation coefficient after various shifts in the underlying copula structure.

Design matrix

We are now able to give an overview of the simulations in the design matrix, in which the following parameters are used:

- Copula: the in-control copula distribution. Note that we use 3 degrees of freedom when the Student-t copula is specified.

- τ: the Kendall's τ correlation coefficient of the two variables.

- Phase: the phase that we execute.

- Shift: the type of shift that we implement. The possibilities are shift in mean (M), shift in variance (V) and shift in dependence structure (C). Furthermore, 1LVL, 2LVLs, $+$ and $-$, respectively, stand for "1 level", "2 levels", "increase" and "decrease". In case of an increase, the shift size varies from 10 to 90 percent, by steps of 10 percent. Conversely, in case of a decrease, the shift size varies from -90 to -10 percent, by steps of 10 percent.

- M1: the first method of the methods that we want to compare. The options are two-sided copula-based chart (Cop. 2S), one-sided copula-based control chart (Cop. 1S), Hotelling T^2

chart (Hot. T^2) and MCUSUM chart.

- M2: the second method of the methods that we want to compare. The options are two-sided copula-based chart (Cop. 2S), one-sided copula-based control chart (Cop. 1S), Hotelling T^2 chart (Hot. T^2) and MCUSUM chart.
- α: the false alarm probability.
- h: the value of the control limit for the MCUSUM chart.
- k: the allowance parameter of the MCUSUM chart.

The design matrix of the simulations is shown in Table 5.4. Note that for this design matrix the potential shift always occurs at the first point of phase II, hence we investigate persistent shifts in a bivariate distribution. The performance comparison procedure of the one-sided and two-sided copula-based control chart for several in-control copulas is covered in cases 1 through 55. Subsequently, the comparison procedure of the Hotelling T^2 chart and two-sided copula-based chart for several in-control copulas is included in cases 56 through 110. Finally, we compare the performance of the two-sided copula-based chart and the MCUSUM chart with allowance parameter 0.25, 0.50 and 0.75 for various in-control copulas with cases 111 through 165, 166 through 220 and 221 through 275, respectively.

Case	Copula	τ	Phase	Shift	M1	M2	α	h	k
1	Gumbel	$\{0.1, \ldots, 0.9\}$	I	x	Cop. 2S	Cop. 1S	0.0027	x	x
2	Clayton	$\{0.1, \ldots, 0.9\}$	I	x	Cop. 2S	Cop. 1S	0.0027	x	x
3	Frank	$\{0.1, \ldots, 0.9\}$	I	x	Cop. 2S	Cop. 1S	0.0027	x	x
4	Normal	$\{0.1, \ldots, 0.9\}$	I	x	Cop. 2S	Cop. 1S	0.0027	x	x
5	Student-t	$\{0.1, \ldots, 0.9\}$	I	x	Cop. 2S	Cop. 1S	0.0027	x	x
6	Gumbel	0.80	II	M: 1LVL+	Cop. 2S	Cop. 1S	0.0027	x	x
7	Gumbel	0.80	II	M: 2LVLs+	Cop. 2S	Cop. 1S	0.0027	x	x
8	Gumbel	0.80	II	M: 1LVL−	Cop. 2S	Cop. 1S	0.0027	x	x
9	Gumbel	0.80	II	M: 2LVLs−	Cop. 2S	Cop. 1S	0.0027	x	x
10	Gumbel	0.80	II	V: +	Cop. 2S	Cop. 1S	0.0027	x	x
11	Gumbel	0.80	II	V: −	Cop. 2S	Cop. 1S	0.0027	x	x
12	Gumbel	0.80	II	C: Clayton	Cop. 2S	Cop. 1S	0.0027	x	x
13	Gumbel	0.80	II	C: Frank	Cop. 2S	Cop. 1S	0.0027	x	x
14	Gumbel	0.80	II	C: Normal	Cop. 2S	Cop. 1S	0.0027	x	x
15	Gumbel	0.80	II	C: Student-t	Cop. 2S	Cop. 1S	0.0027	x	x
16	Clayton	0.80	II	M: 1LVL+	Cop. 2S	Cop. 1S	0.0027	x	x
17	Clayton	0.80	II	M: 2LVLs+	Cop. 2S	Cop. 1S	0.0027	x	x
18	Clayton	0.80	II	M: 1LVL−	Cop. 2S	Cop. 1S	0.0027	x	x
19	Clayton	0.80	II	M: 2LVLs−	Cop. 2S	Cop. 1S	0.0027	x	x
20	Clayton	0.80	II	V: +	Cop. 2S	Cop. 1S	0.0027	x	x
21	Clayton	0.80	II	V: −	Cop. 2S	Cop. 1S	0.0027	x	x
22	Clayton	0.80	II	C: Gumbel	Cop. 2S	Cop. 1S	0.0027	x	x
23	Clayton	0.80	II	C: Frank	Cop. 2S	Cop. 1S	0.0027	x	x
24	Clayton	0.80	II	C: Normal	Cop. 2S	Cop. 1S	0.0027	x	x
25	Clayton	0.80	II	C: Student-t	Cop. 2S	Cop. 1S	0.0027	x	x
26	Frank	0.60	II	M: 1LVL+	Cop. 2S	Cop. 1S	0.0027	x	x
27	Frank	0.60	II	M: 2LVLs+	Cop. 2S	Cop. 1S	0.0027	x	x
28	Frank	0.60	II	M: 1LVL−	Cop. 2S	Cop. 1S	0.0027	x	x
29	Frank	0.60	II	M: 2LVLs−	Cop. 2S	Cop. 1S	0.0027	x	x
30	Frank	0.60	II	V: +	Cop. 2S	Cop. 1S	0.0027	x	x
31	Frank	0.60	II	V: −	Cop. 2S	Cop. 1S	0.0027	x	x
32	Frank	0.60	II	C: Gumbel	Cop. 2S	Cop. 1S	0.0027	x	x
33	Frank	0.60	II	C: Clayton	Cop. 2S	Cop. 1S	0.0027	x	x
34	Frank	0.60	II	C: Normal	Cop. 2S	Cop. 1S	0.0027	x	x
35	Frank	0.60	II	C: Student-t	Cop. 2S	Cop. 1S	0.0027	x	x
36	Normal	0.40	II	M: 1LVL+	Cop. 2S	Cop. 1S	0.0027	x	x
37	Normal	0.40	II	M: 2LVLs+	Cop. 2S	Cop. 1S	0.0027	x	x
38	Normal	0.40	II	M: 1LVL−	Cop. 2S	Cop. 1S	0.0027	x	x
39	Normal	0.40	II	M: 2LVLs−	Cop. 2S	Cop. 1S	0.0027	x	x
40	Normal	0.40	II	V: +	Cop. 2S	Cop. 1S	0.0027	x	x
41	Normal	0.40	II	V: −	Cop. 2S	Cop. 1S	0.0027	x	x
42	Normal	0.40	II	C: Gumbel	Cop. 2S	Cop. 1S	0.0027	x	x
43	Normal	0.40	II	C: Clayton	Cop. 2S	Cop. 1S	0.0027	x	x
44	Normal	0.40	II	C: Frank	Cop. 2S	Cop. 1S	0.0027	x	x
45	Normal	0.40	II	C: Student-t	Cop. 2S	Cop. 1S	0.0027	x	x
46	Student-t	0.40	II	M: 1LVL+	Cop. 2S	Cop. 1S	0.0027	x	x
47	Student-t	0.40	II	M: 2LVLs+	Cop. 2S	Cop. 1S	0.0027	x	x

Case	Copula	τ	Phase	Shift	M1	M2	α	h	k
48	Student-t	0.40	II	M: 1LVL−	Cop. 2S	Cop. 1S	0.0027	x	x
49	Student-t	0.40	II	M: 2LVLs−	Cop. 2S	Cop. 1S	0.0027	x	x
50	Student-t	0.40	II	V: +	Cop. 2S	Cop. 1S	0.0027	x	x
51	Student-t	0.40	II	V: −	Cop. 2S	Cop. 1S	0.0027	x	x
52	Student-t	0.40	II	C: Gumbel	Cop. 2S	Cop. 1S	0.0027	x	x
53	Student-t	0.40	II	C: Clayton	Cop. 2S	Cop. 1S	0.0027	x	x
54	Student-t	0.40	II	C: Frank	Cop. 2S	Cop. 1S	0.0027	x	x
55	Student-t	0.40	II	C: Normal	Cop. 2S	Cop. 1S	0.0027	x	x
56	Gumbel	$\{0.1,\ldots,0.9\}$	I	x	Hot. T^2	Cop. 2S	0.0027	x	x
57	Clayton	$\{0.1,\ldots,0.9\}$	I	x	Hot. T^2	Cop. 2S	0.0027	x	x
58	Frank	$\{0.1,\ldots,0.9\}$	I	x	Hot. T^2	Cop. 2S	0.0027	x	x
59	Normal	$\{0.1,\ldots,0.9\}$	I	x	Hot. T^2	Cop. 2S	0.0027	x	x
60	Student-t	$\{0.1,\ldots,0.9\}$	I	x	Hot. T^2	Cop. 2S	0.0027	x	x
61	Gumbel	0.56	II	M: 1LVL+	Hot. T^2	Cop. 2S	0.0027	x	x
62	Gumbel	0.56	II	M: 2LVLs+	Hot. T^2	Cop. 2S	0.0027	x	x
63	Gumbel	0.56	II	M: 1LVL−	Hot. T^2	Cop. 2S	0.0027	x	x
64	Gumbel	0.56	II	M: 2LVLs−	Hot. T^2	Cop. 2S	0.0027	x	x
65	Gumbel	0.56	II	V: +	Hot. T^2	Cop. 2S	0.0027	x	x
66	Gumbel	0.56	II	V: −	Hot. T^2	Cop. 2S	0.0027	x	x
67	Gumbel	0.56	II	C: Clayton	Hot. T^2	Cop. 2S	0.0027	x	x
68	Gumbel	0.56	II	C: Frank	Hot. T^2	Cop. 2S	0.0027	x	x
69	Gumbel	0.56	II	C: Normal	Hot. T^2	Cop. 2S	0.0027	x	x
70	Gumbel	0.56	II	C: Student-t	Hot. T^2	Cop. 2S	0.0027	x	x
71	Clayton	0.59	II	M: 1LVL+	Hot. T^2	Cop. 2S	0.0027	x	x
72	Clayton	0.59	II	M: 2LVLs+	Hot. T^2	Cop. 2S	0.0027	x	x
73	Clayton	0.59	II	M: 1LVL−	Hot. T^2	Cop. 2S	0.0027	x	x
74	Clayton	0.59	II	M: 2LVLs−	Hot. T^2	Cop. 2S	0.0027	x	x
75	Clayton	0.59	II	V: +	Hot. T^2	Cop. 2S	0.0027	x	x
76	Clayton	0.59	II	V: −	Hot. T^2	Cop. 2S	0.0027	x	x
77	Clayton	0.59	II	C: Gumbel	Hot. T^2	Cop. 2S	0.0027	x	x
78	Clayton	0.59	II	C: Frank	Hot. T^2	Cop. 2S	0.0027	x	x
79	Clayton	0.59	II	C: Normal	Hot. T^2	Cop. 2S	0.0027	x	x
80	Clayton	0.59	II	C: Student-t	Hot. T^2	Cop. 2S	0.0027	x	x
81	Frank	0.53	II	M: 1LVL+	Hot. T^2	Cop. 2S	0.0027	x	x
82	Frank	0.53	II	M: 2LVLs+	Hot. T^2	Cop. 2S	0.0027	x	x
83	Frank	0.53	II	M: 1LVL−	Hot. T^2	Cop. 2S	0.0027	x	x
84	Frank	0.53	II	M: 2LVLs−	Hot. T^2	Cop. 2S	0.0027	x	x
85	Frank	0.53	II	V: +	Hot. T^2	Cop. 2S	0.0027	x	x
86	Frank	0.53	II	V: −	Hot. T^2	Cop. 2S	0.0027	x	x
87	Frank	0.53	II	C: Gumbel	Hot. T^2	Cop. 2S	0.0027	x	x
88	Frank	0.53	II	C: Clayton	Hot. T^2	Cop. 2S	0.0027	x	x
89	Frank	0.53	II	C: Normal	Hot. T^2	Cop. 2S	0.0027	x	x
90	Frank	0.53	II	C: Student-t	Hot. T^2	Cop. 2S	0.0027	x	x
91	Normal	0.87	II	M: 1LVL+	Hot. T^2	Cop. 2S	0.0027	x	x
92	Normal	0.87	II	M: 2LVLs+	Hot. T^2	Cop. 2S	0.0027	x	x
93	Normal	0.87	II	M: 1LVL−	Hot. T^2	Cop. 2S	0.0027	x	x
94	Normal	0.87	II	M: 2LVLs−	Hot. T^2	Cop. 2S	0.0027	x	x
95	Normal	0.87	II	V: +	Hot. T^2	Cop. 2S	0.0027	x	x
96	Normal	0.87	II	V: −	Hot. T^2	Cop. 2S	0.0027	x	x
97	Normal	0.87	II	C: Gumbel	Hot. T^2	Cop. 2S	0.0027	x	x
98	Normal	0.87	II	C: Clayton	Hot. T^2	Cop. 2S	0.0027	x	x
99	Normal	0.87	II	C: Frank	Hot. T^2	Cop. 2S	0.0027	x	x
100	Normal	0.87	II	C: Student-t	Hot. T^2	Cop. 2S	0.0027	x	x
101	Student-t	0.44	II	M: 1LVL+	Hot. T^2	Cop. 2S	0.0027	x	x
102	Student-t	0.44	II	M: 2LVLs+	Hot. T^2	Cop. 2S	0.0027	x	x
103	Student-t	0.44	II	M: 1LVL−	Hot. T^2	Cop. 2S	0.0027	x	x
104	Student-t	0.44	II	M: 2LVLs−	Hot. T^2	Cop. 2S	0.0027	x	x
105	Student-t	0.44	II	V: +	Hot. T^2	Cop. 2S	0.0027	x	x
106	Student-t	0.44	II	V: −	Hot. T^2	Cop. 2S	0.0027	x	x
107	Student-t	0.44	II	C: Gumbel	Hot. T^2	Cop. 2S	0.0027	x	x
108	Student-t	0.44	II	C: Clayton	Hot. T^2	Cop. 2S	0.0027	x	x
109	Student-t	0.44	II	C: Frank	Hot. T^2	Cop. 2S	0.0027	x	x
110	Student-t	0.44	II	C: Normal	Hot. T^2	Cop. 2S	0.0027	x	x
111	Gumbel	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	10.1	0.25
112	Clayton	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	10.1	0.25
113	Frank	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	10.0	0.25
114	Normal	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	10.0	0.25
115	Student-t	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	10.1	0.25
116	Gumbel	0.80	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	10.1	0.25

Case	Copula	τ	Phase	Shift	M1	M2	α	h	k
117	Gumbel	0.80	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	10.1	0.25
118	Gumbel	0.80	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	10.1	0.25
119	Gumbel	0.80	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	10.1	0.25
120	Gumbel	0.80	II	V: +	MCUSUM	Cop. 2S	0.0027	10.1	0.25
121	Gumbel	0.80	II	V: −	MCUSUM	Cop. 2S	0.0027	10.1	0.25
122	Gumbel	0.80	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	10.1	0.25
123	Gumbel	0.80	II	C: Frank	MCUSUM	Cop. 2S	0.0027	10.1	0.25
124	Gumbel	0.80	II	C: Normal	MCUSUM	Cop. 2S	0.0027	10.1	0.25
125	Gumbel	0.80	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	10.1	0.25
126	Clayton	0.60	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	10.1	0.25
127	Clayton	0.60	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	10.1	0.25
128	Clayton	0.60	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	10.1	0.25
129	Clayton	0.60	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	10.1	0.25
130	Clayton	0.60	II	V: +	MCUSUM	Cop. 2S	0.0027	10.1	0.25
131	Clayton	0.60	II	V: −	MCUSUM	Cop. 2S	0.0027	10.1	0.25
132	Clayton	0.60	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	10.1	0.25
133	Clayton	0.60	II	C: Frank	MCUSUM	Cop. 2S	0.0027	10.1	0.25
134	Clayton	0.60	II	C: Normal	MCUSUM	Cop. 2S	0.0027	10.1	0.25
135	Clayton	0.60	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	10.1	0.25
136	Frank	0.80	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	10.0	0.25
137	Frank	0.80	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	10.0	0.25
138	Frank	0.80	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	10.0	0.25
139	Frank	0.80	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	10.0	0.25
140	Frank	0.80	II	V: +	MCUSUM	Cop. 2S	0.0027	10.0	0.25
141	Frank	0.80	II	V: −	MCUSUM	Cop. 2S	0.0027	10.0	0.25
142	Frank	0.80	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	10.0	0.25
143	Frank	0.80	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	10.0	0.25
144	Frank	0.80	II	C: Normal	MCUSUM	Cop. 2S	0.0027	10.0	0.25
145	Frank	0.80	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	10.0	0.25
146	Normal	0.80	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	10.0	0.25
147	Normal	0.80	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	10.0	0.25
148	Normal	0.80	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	10.0	0.25
149	Normal	0.80	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	10.0	0.25
150	Normal	0.80	II	V: +	MCUSUM	Cop. 2S	0.0027	10.0	0.25
151	Normal	0.80	II	V: −	MCUSUM	Cop. 2S	0.0027	10.0	0.25
152	Normal	0.80	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	10.0	0.25
153	Normal	0.80	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	10.0	0.25
154	Normal	0.80	II	C: Frank	MCUSUM	Cop. 2S	0.0027	10.0	0.25
155	Normal	0.80	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	10.0	0.25
156	Student-t	0.60	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	10.1	0.25
157	Student-t	0.60	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	10.1	0.25
158	Student-t	0.60	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	10.1	0.25
159	Student-t	0.60	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	10.1	0.25
160	Student-t	0.60	II	V: +	MCUSUM	Cop. 2S	0.0027	10.1	0.25
161	Student-t	0.60	II	V: −	MCUSUM	Cop. 2S	0.0027	10.1	0.25
162	Student-t	0.60	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	10.1	0.25
163	Student-t	0.60	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	10.1	0.25
164	Student-t	0.60	II	C: Frank	MCUSUM	Cop. 2S	0.0027	10.1	0.25
165	Student-t	0.60	II	C: Normal	MCUSUM	Cop. 2S	0.0027	10.1	0.25
166	Gumbel	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	6.1	0.50
167	Clayton	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	6.1	0.50
168	Frank	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	6.1	0.50
169	Normal	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	6.0	0.50
170	Student-t	$\{0.1,\ldots,0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	6.1	0.50
171	Gumbel	0.44	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
172	Gumbel	0.44	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
173	Gumbel	0.44	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
174	Gumbel	0.44	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
175	Gumbel	0.44	II	V: +	MCUSUM	Cop. 2S	0.0027	6.1	0.50
176	Gumbel	0.44	II	V: −	MCUSUM	Cop. 2S	0.0027	6.1	0.50
177	Gumbel	0.44	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	6.1	0.50
178	Gumbel	0.44	II	C: Frank	MCUSUM	Cop. 2S	0.0027	6.1	0.50
179	Gumbel	0.44	II	C: Normal	MCUSUM	Cop. 2S	0.0027	6.1	0.50
180	Gumbel	0.44	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	6.1	0.50
181	Clayton	0.40	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
182	Clayton	0.40	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
183	Clayton	0.40	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
184	Clayton	0.40	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
185	Clayton	0.40	II	V: +	MCUSUM	Cop. 2S	0.0027	6.1	0.50

Case	Copula	τ	Phase	Shift	M1	M2	α	h	k
186	Clayton	0.40	II	V: −	MCUSUM	Cop. 2S	0.0027	6.1	0.50
187	Clayton	0.40	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	6.1	0.50
188	Clayton	0.40	II	C: Frank	MCUSUM	Cop. 2S	0.0027	6.1	0.50
189	Clayton	0.40	II	C: Normal	MCUSUM	Cop. 2S	0.0027	6.1	0.50
190	Clayton	0.40	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	6.1	0.50
191	Frank	0.47	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
192	Frank	0.47	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
193	Frank	0.47	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
194	Frank	0.47	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
195	Frank	0.47	II	V: +	MCUSUM	Cop. 2S	0.0027	6.1	0.50
196	Frank	0.47	II	V: −	MCUSUM	Cop. 2S	0.0027	6.1	0.50
197	Frank	0.47	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	6.1	0.50
198	Frank	0.47	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	6.1	0.50
199	Frank	0.47	II	C: Normal	MCUSUM	Cop. 2S	0.0027	6.1	0.50
200	Frank	0.47	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	6.1	0.50
201	Normal	0.40	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	6.0	0.50
202	Normal	0.40	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	6.0	0.50
203	Normal	0.40	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	6.0	0.50
204	Normal	0.40	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	6.0	0.50
205	Normal	0.40	II	V: +	MCUSUM	Cop. 2S	0.0027	6.0	0.50
206	Normal	0.40	II	V: −	MCUSUM	Cop. 2S	0.0027	6.0	0.50
207	Normal	0.40	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	6.0	0.50
208	Normal	0.40	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	6.0	0.50
209	Normal	0.40	II	C: Frank	MCUSUM	Cop. 2S	0.0027	6.0	0.50
210	Normal	0.40	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	6.0	0.50
211	Student-t	0.53	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
212	Student-t	0.53	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	6.1	0.50
213	Student-t	0.53	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
214	Student-t	0.53	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	6.1	0.50
215	Student-t	0.53	II	V: +	MCUSUM	Cop. 2S	0.0027	6.1	0.50
216	Student-t	0.53	II	V: −	MCUSUM	Cop. 2S	0.0027	6.1	0.50
217	Student-t	0.53	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	6.1	0.50
218	Student-t	0.53	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	6.1	0.50
219	Student-t	0.53	II	C: Frank	MCUSUM	Cop. 2S	0.0027	6.1	0.50
220	Student-t	0.53	II	C: Normal	MCUSUM	Cop. 2S	0.0027	6.1	0.50
221	Gumbel	$\{0.1, \ldots, 0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	4.1	0.75
222	Clayton	$\{0.1, \ldots, 0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	4.1	0.75
223	Frank	$\{0.1, \ldots, 0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	4.1	0.75
224	Normal	$\{0.1, \ldots, 0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	4.1	0.75
225	Student-t	$\{0.1, \ldots, 0.9\}$	I	x	MCUSUM	Cop. 2S	0.0027	4.2	0.75
226	Gumbel	0.41	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
227	Gumbel	0.41	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
228	Gumbel	0.41	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
229	Gumbel	0.41	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
230	Gumbel	0.41	II	V: +	MCUSUM	Cop. 2S	0.0027	4.1	0.75
231	Gumbel	0.41	II	V: −	MCUSUM	Cop. 2S	0.0027	4.1	0.75
232	Gumbel	0.41	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	4.1	0.75
233	Gumbel	0.41	II	C: Frank	MCUSUM	Cop. 2S	0.0027	4.1	0.75
234	Gumbel	0.41	II	C: Normal	MCUSUM	Cop. 2S	0.0027	4.1	0.75
235	Gumbel	0.41	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	4.1	0.75
236	Clayton	0.40	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
237	Clayton	0.40	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
238	Clayton	0.40	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
239	Clayton	0.40	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
240	Clayton	0.40	II	V: +	MCUSUM	Cop. 2S	0.0027	4.1	0.75
241	Clayton	0.40	II	V: −	MCUSUM	Cop. 2S	0.0027	4.1	0.75
242	Clayton	0.40	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	4.1	0.75
243	Clayton	0.40	II	C: Frank	MCUSUM	Cop. 2S	0.0027	4.1	0.75
244	Clayton	0.40	II	C: Normal	MCUSUM	Cop. 2S	0.0027	4.1	0.75
245	Clayton	0.40	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	4.1	0.75
246	Frank	0.40	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
247	Frank	0.40	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
248	Frank	0.40	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
249	Frank	0.40	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
250	Frank	0.40	II	V: +	MCUSUM	Cop. 2S	0.0027	4.1	0.75
251	Frank	0.40	II	V: −	MCUSUM	Cop. 2S	0.0027	4.1	0.75
252	Frank	0.40	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	4.1	0.75
253	Frank	0.40	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	4.1	0.75
254	Frank	0.40	II	C: Normal	MCUSUM	Cop. 2S	0.0027	4.1	0.75

Case	Copula	τ	Phase	Shift	M1	M2	α	h	k
255	Frank	0.40	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	4.1	0.75
256	Normal	0.52	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
257	Normal	0.52	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	4.1	0.75
258	Normal	0.52	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
259	Normal	0.52	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	4.1	0.75
260	Normal	0.52	II	V: +	MCUSUM	Cop. 2S	0.0027	4.1	0.75
261	Normal	0.52	II	V: −	MCUSUM	Cop. 2S	0.0027	4.1	0.75
262	Normal	0.52	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	4.1	0.75
263	Normal	0.52	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	4.1	0.75
264	Normal	0.52	II	C: Frank	MCUSUM	Cop. 2S	0.0027	4.1	0.75
265	Normal	0.52	II	C: Student-t	MCUSUM	Cop. 2S	0.0027	4.1	0.75
266	Student-t	0.40	II	M: 1LVL+	MCUSUM	Cop. 2S	0.0027	4.2	0.75
267	Student-t	0.40	II	M: 2LVLs+	MCUSUM	Cop. 2S	0.0027	4.2	0.75
268	Student-t	0.40	II	M: 1LVL−	MCUSUM	Cop. 2S	0.0027	4.2	0.75
269	Student-t	0.40	II	M: 2LVLs−	MCUSUM	Cop. 2S	0.0027	4.2	0.75
270	Student-t	0.40	II	V: +	MCUSUM	Cop. 2S	0.0027	4.2	0.75
271	Student-t	0.40	II	V: −	MCUSUM	Cop. 2S	0.0027	4.2	0.75
272	Student-t	0.40	II	C: Gumbel	MCUSUM	Cop. 2S	0.0027	4.2	0.75
273	Student-t	0.40	II	C: Clayton	MCUSUM	Cop. 2S	0.0027	4.2	0.75
274	Student-t	0.40	II	C: Frank	MCUSUM	Cop. 2S	0.0027	4.2	0.75
275	Student-t	0.40	II	C: Normal	MCUSUM	Cop. 2S	0.0027	4.2	0.75

Table 5.4: Design matrix.

5.2.2 R software

R is a free software environment for statistical computing and graphics which is supported by the **R** Foundation for Statistical Computing (R Core Team (2020)). The programming language **R** is very popular among statisticians for developing statistical software and data analysis. Both the simulation study of this chapter and the case study of Chapter 6 will be analyzed using **R**. The **R** software includes a wide variety of libraries which implement statistical and graphical techniques. Furthermore, it contains several packages which are extremely useful for our research objectives. In particular, we make use of the packages `copula` (see Hofert et al. (2020)), `VineCopula` (see Schepsmeier et al. (2019)) and `rvinecopulib` (see Nagler and Vatter (2020)) for copula modeling and monitoring. Besides, we make use of the package `MSQC` (see Santos-Fernández (2013)) for conventional multivariate monitoring methods.

5.2.3 One-sided vs. two-sided version

Gumbel (case 1 & cases $6 - 15$)

As stated in Section 5.1, our conclusions cannot be based solely on the ARL as run length distributions are often very skewed. Therefore, we provide additional kernel density plots to substantiate our statements.

According to Figures 5.6 and 5.7, we can set $\tau = 0.80$ to achieve that both methods reach approximately the same ARL_0. In this way we have a fair comparison when analyzing shifts in phase II. Note that a τ value of around 0.77 would have been a better choice. We observe that the ARL_0 of the two-sided version is around 370 and the ARL_0 of the one-sided version varies around 360, which could indicate that the ARL_0 is equal to $1/\alpha$. Note that the ARL_0 values of both methods do not variate that much for any input parameter τ, that is, for any level of dependence between the two variables.
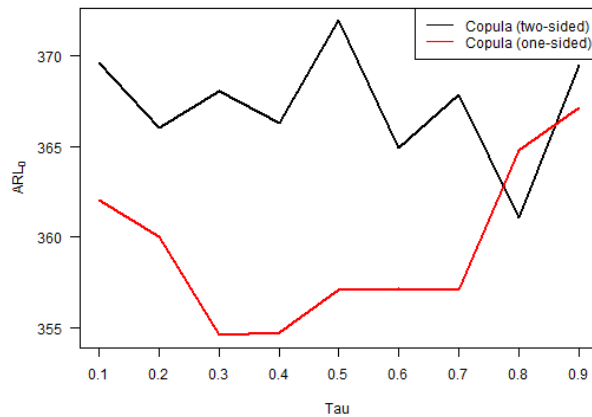
Figure 5.6: ARL_0 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Gumbel copula.
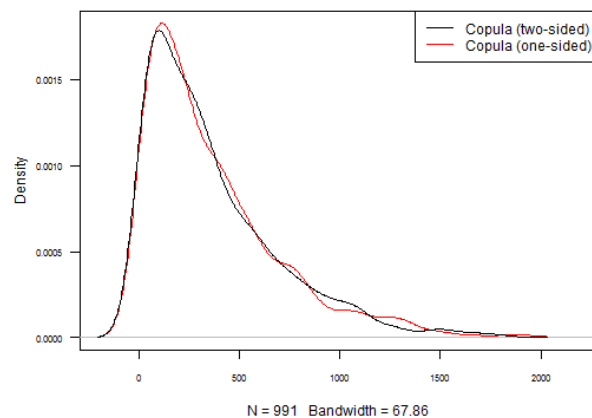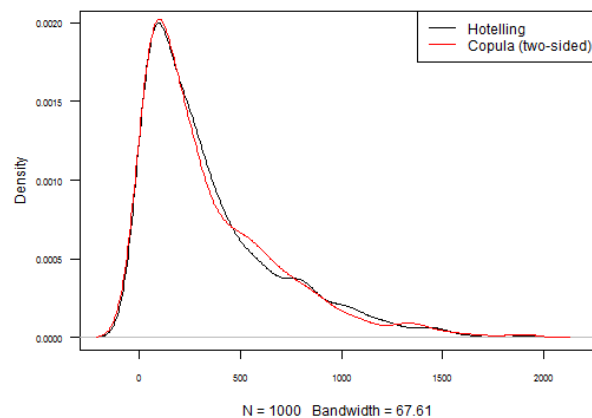

Figure 5.7: Kernel density plot of the in-control run length for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Gumbel copula, and $\tau = 0.80$.

From Figure 5.8 we conclude that both methods perform equally well when the level of one or two variables is increased, although the one-sided version seems to perform a bit better for small shifts. When the level of the first variable is decreased, we see that the one-sided version performs slightly better for small shifts. The main difference is in the case when the level of both variables is decreased. Note that when no value of the ARL is plotted, the estimated value is unknown but for sure higher than 2,000. Both methods perform poorly in this case. However, the one-sided version seems to be unable to detect large negative shifts in the level of both variables.
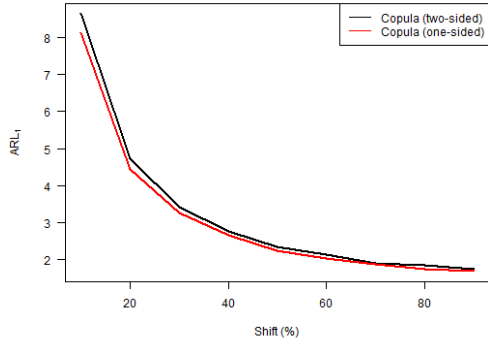
(a) Level of the first variable increased.
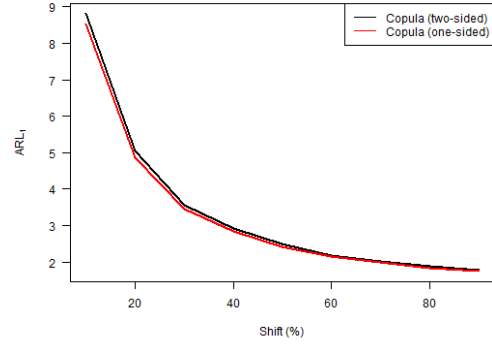
(b) Level of both variables increased.
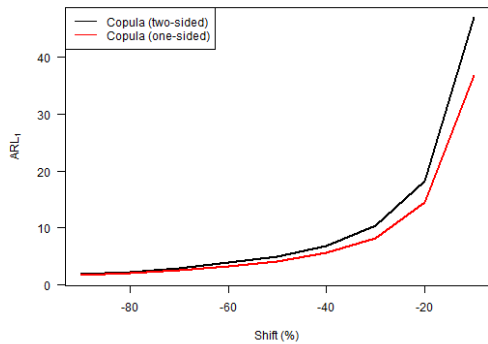
(c) Level of the first variable decreased.
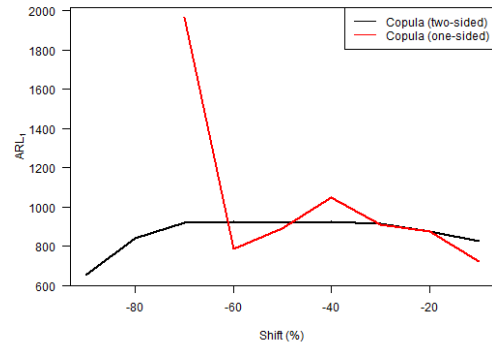
(d) Level of both variables decreased.

Figure 5.8: ARL_1 vs. level shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Gumbel copula.

When looking at Figure 5.9, we notice that the two-sided version performs better if the largest eigenvalue of the correlation matrix is increased, which means that the two-sided version is better in detecting positive shifts in the correlation matrix in general. Note that the ARL_1 values, however, are unacceptably high. On the other hand, when there is a negative shift in the correlation matrix, the one-sided version seems to perform a bit better when the shift size is small or medium.
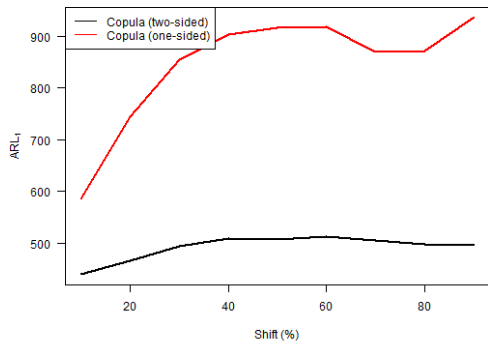

(a) Largest eigenvalue increased.
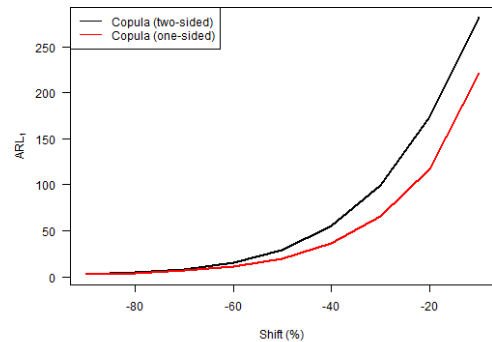
(b) Largest eigenvalue decreased.

Figure 5.9: ARL_1 vs. largest eigenvalue shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Gumbel copula.

From Figure 5.10 and the fact that we set $\tau = 0.80$, we conclude that both methods perform

equally well when the underlying copula structure shifts to the Clayton or Frank copula, although the one-sided version might perform slightly better in the latter case. In case the underlying copula shifts to the Normal or Student-t copula, the one-sided version performs significantly better. This could indicate that both methods perform equally well when the underlying copula remains an Archimedean copula, and that the one-sided version performs better when the underlying copula shifts to an elliptical copula. Note that the ARL_1 values for a shift to the Normal copula are undesirable.
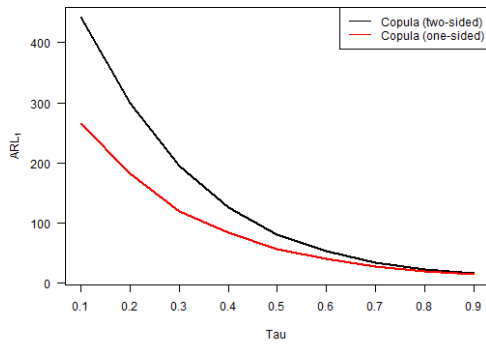

(a) Shift to Clayton copula.
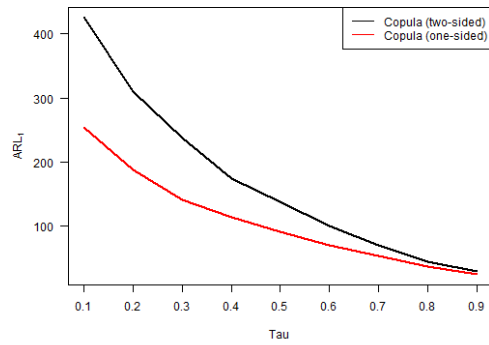
(b) Shift to Frank copula.
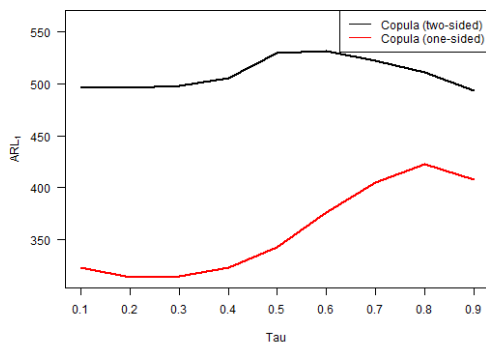
(c) Shift to Normal copula.
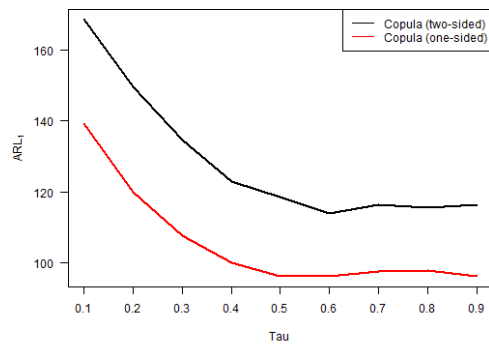
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.10: ARL_1 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Gumbel copula and shifts to a different copula.

Normal (case 4 & cases $36 - 45$)

According to Figure 5.11, there is no value of τ for which the ARL_0 values are exactly equal. However, the ARL_0 values of both methods do not differ that much. Note that the ARL_0 values of both methods are constant for any input parameter τ and the values are around 350. We simply choose $\tau = 0.40$ in order to reach approximately the same ARL_0 for both methods, which is verified by Figure 5.12, and proceed to phase II analysis.
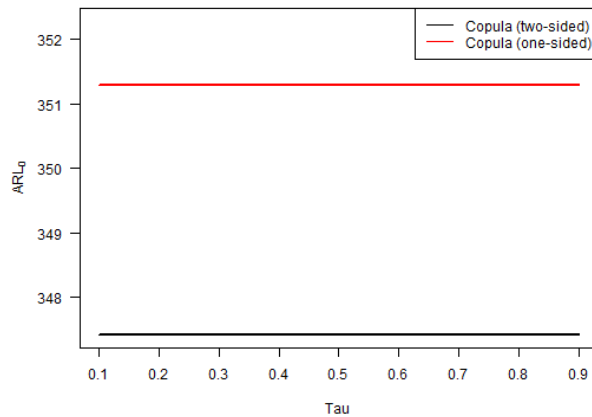
Figure 5.11: ARL_0 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Normal copula.
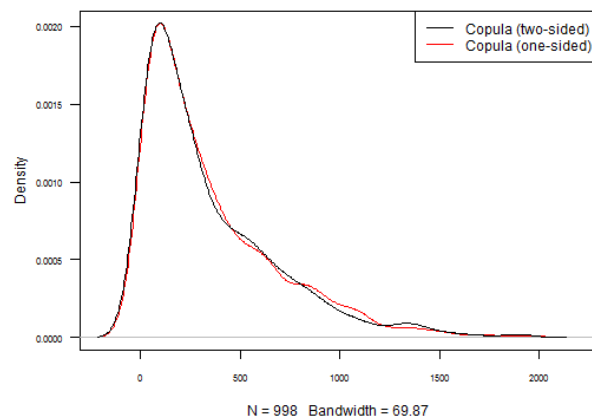
Figure 5.12: Kernel density plot of the in-control run length for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Normal copula, and $\tau = 0.40$.

From Figure 5.13 we conclude that both methods perform equally well when the level of one or two variables is increased. When the level of the first variable is decreased, we observe that the one-sided version performs better in general. The main difference is in the case when the level of both variables is decreased. Both methods perform quite poorly in this case. However, the one-sided version is unable to detect negative shifts in the level of both variables when the shift size is medium or large.
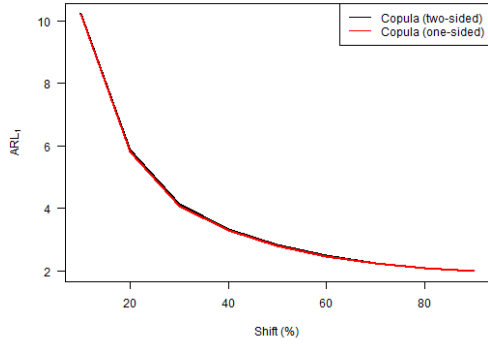
(a) Level of the first variable increased.
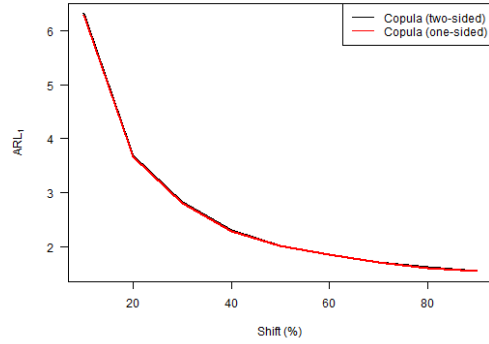
(b) Level of both variables increased.
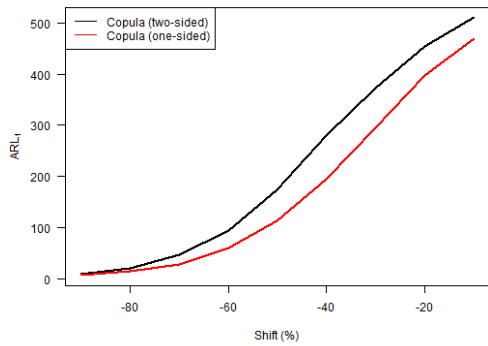
(c) Level of the first variable decreased.
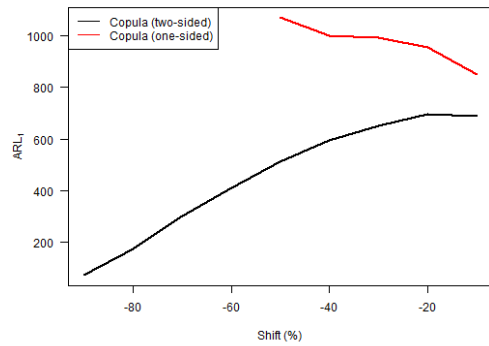
(d) Level of both variables decreased.

Figure 5.13: ARL_1 vs. level shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Normal copula.

When looking at Figure 5.14, we notice that the two-sided version is way better in detecting positive shifts in the correlation matrix for any shift size. Note that the ARL_1 values, however, are very high. Furthermore, when there is a negative shift in the correlation matrix, the one-sided version seems to perform better when the shift size is small or medium.
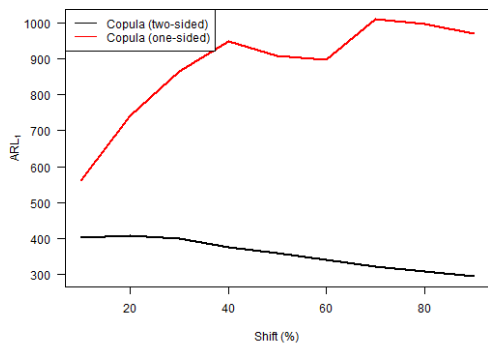
(a) Largest eigenvalue increased.
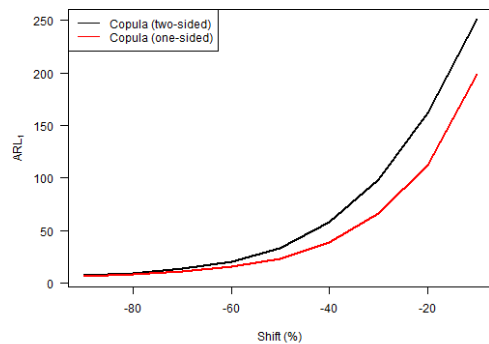
(b) Largest eigenvalue decreased.

Figure 5.14: ARL_1 vs. largest eigenvalue shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Normal copula.

From Figure 5.15 and the fact that we set $\tau = 0.40$, we conclude that the one-sided version performs slightly better when the underlying copula shifts to the Gumbel or Clayton copula. Moreover, if

65

the underlying copula shifts to the Frank or Student-t copula, the one-sided version performs better as well. Note, however, that we could have chosen a different value of τ, see Figure 5.11. If we, for instance, would have chosen a smaller τ, the two-sided version would perform better when the copula shifts to the Gumbel or Clayton copula. For really small τ, the two-sided version performs also better when the copula shifts to the Student-t copula. And if we would have chosen a larger τ, the one-sided version seems to outperform the two-sided version when the copula shifts to a different copula in general.

(a) Shift to Gumbel copula.

(b) Shift to Clayton copula.

(c) Shift to Frank copula.

(d) Shift to Student-t copula with 3 degrees of freedom.
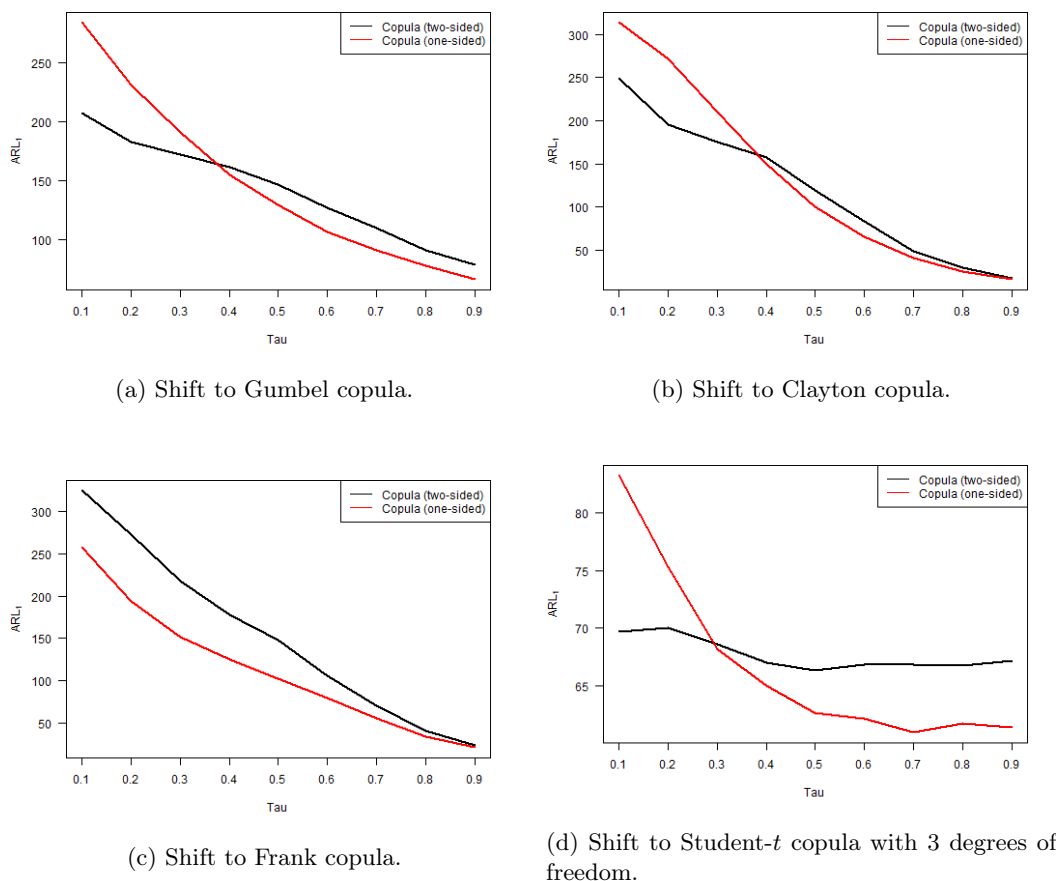
Figure 5.15: ARL_1 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Normal copula and shifts to a different copula.

Overall, we see that our main conclusions are similar in case the in-control distribution follows a Gumbel or Normal copula. Basically, we conclude that both methods perform equally well when the level is increased. If the level of only one variable is decreased, it seems that the one-sided version performs a bit better. However, when the level of both variables is decreased, the one-sided version is unable to detect large shifts and performs equal to or worse than the two-sided version for small shifts. Furthermore, the two-sided version performs way better for positive shifts in the correlation matrix, whereas the one-sided version performs slightly better for small or medium negative shifts in the correlation matrix. For large negative shifts in the correlation matrix, we observe that the two methods perform equally well. In case the underlying copula shifts to a different copula, the one-sided version seems to perform better in general. Most of these findings are in line with the results in case the in-control distribution follows a Clayton, Frank or Student-t copula, which can be found in Appendix B.1. There are some minor differences, but these can be regarded as insignificant. For instance, consider the case that the in-control distribution is a Student-t copula. Then, the one-sided version seems to perform slightly better for small shifts when the level of both variables is decreased. Besides that, when there is a negative shift in the correlation matrix, the one-sided version performs better for any shift size. Furthermore, only

when the in-control distribution follows a Frank copula, the two-sided version performs better if the underlying copula shifts to a different copula.

We advise to use the two-sided version instead of the one-sided version for detecting mean shifts in general, as the two-sided version is able to detect shifts of any size. Also, we propose to use the two-sided version in order to detect arbitrary shifts in the correlation matrix, as the large difference in performance for positive shifts outweigh the smaller difference for negative shifts. Finally, if one desires to detect a shift in the dependence structure, we recommend to use the one-sided version since it performs better than the two-sided version. All in all, we advise to use the two-sided version in general to detect shifts of arbitrary type. There are, however, particular cases in which the one-sided version performs significantly better.

5.2.4 Copula-based control chart vs. Hotelling T^2

Gumbel (case 56 & cases $61 - 70$)

According to Figures 5.16 and 5.17, we can choose $\tau = 0.56$ to achieve that both methods reach approximately the same ARL_0. Again, we observe that the ARL_0 of both methods is close to 370, which suggests that the ARL_0 could be equal to $1/\alpha$. Note that the ARL_0 values of the copula-based chart are roughly constant for any input parameter τ, whereas the ARL_0 values of the T^2 chart are highly influenced by the input parameter τ.
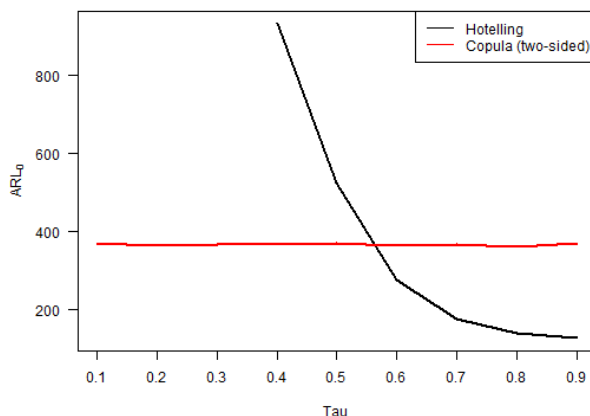
Figure 5.16: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Gumbel copula.
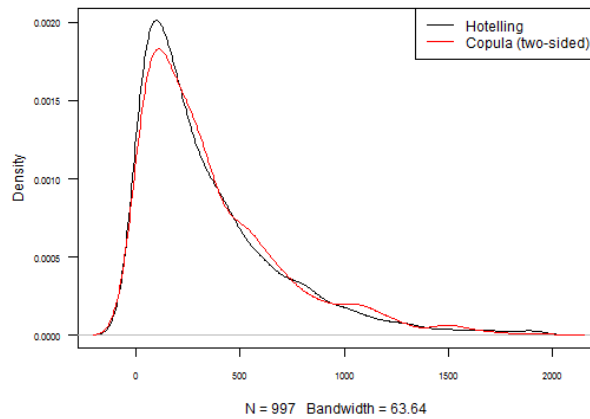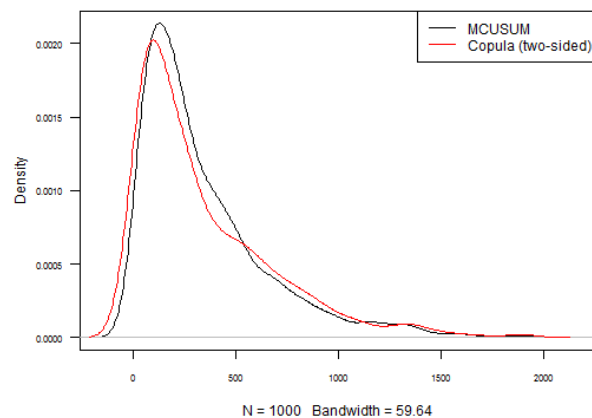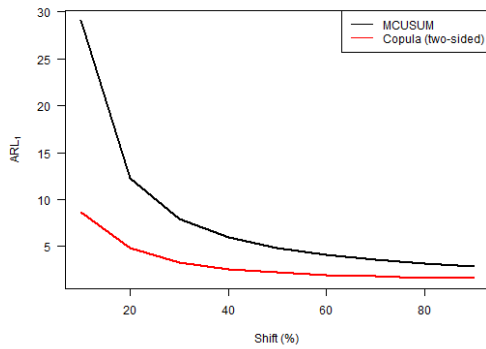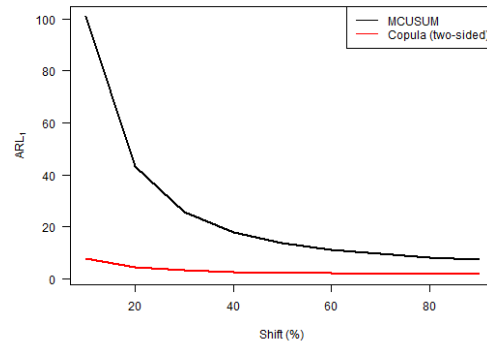
Figure 5.17: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Gumbel copula, and $\tau = 0.56$.

From Figure 5.18 we conclude that the copula-based chart performs better when the level of one or two variables is increased, especially for small shift sizes. When the level of the first variable is decreased, we conclude that the copula-based chart performs better in general. Furthermore, when the level of both variables is decreased, the T^2 chart is unable to detect negative shifts in the level of both variables when the shift size is medium or large. Only really small shift sizes can be detected in this case. The copula-based chart is able to detect these kinds of shifts, however, the resulting ARL_1 values are very high.
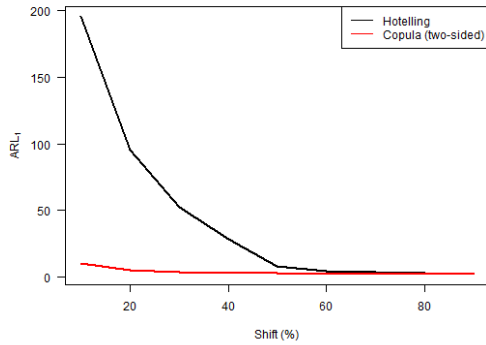
(a) Level of the first variable increased.
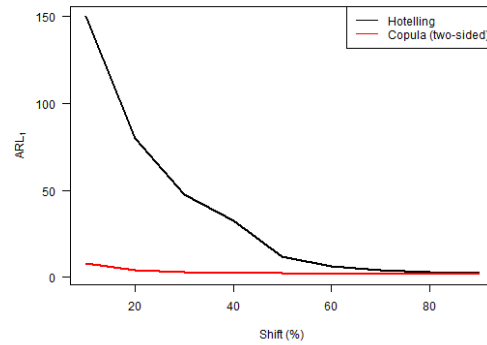
(b) Level of both variables increased.
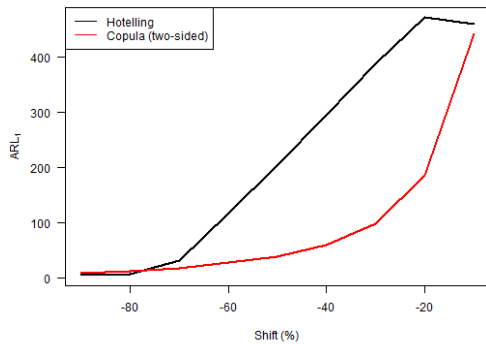
(c) Level of the first variable decreased.
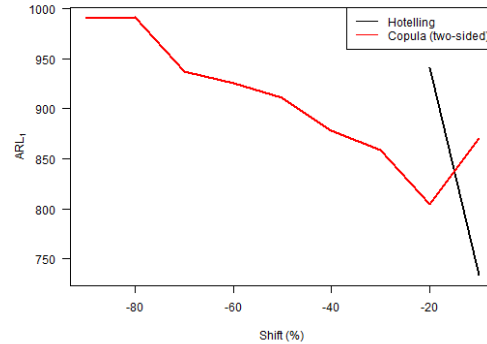
(d) Level of both variables decreased.

Figure 5.18: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Gumbel copula.

When observing Figure 5.19, we see that the copula-based chart is way better in detecting positive shifts in the correlation matrix for any shift size. Note that the ARL_1 values, however, are quite high. Furthermore, when there is a negative shift in the correlation matrix, the T^2 chart seems to perform better, especially for small shift sizes.
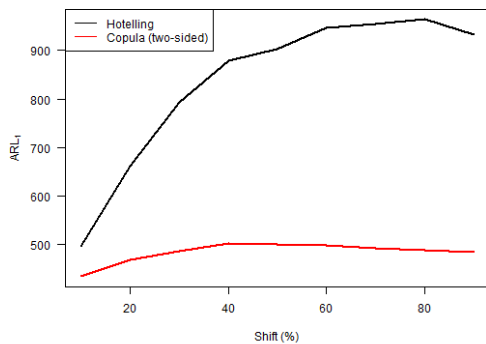
(a) Largest eigenvalue increased.
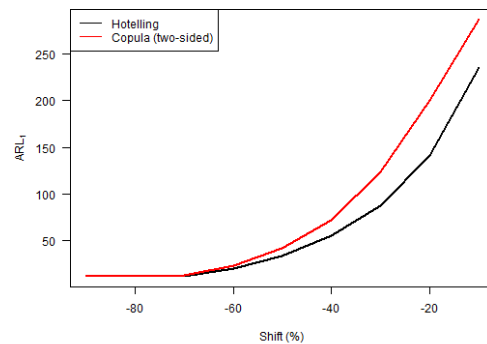
(b) Largest eigenvalue decreased.
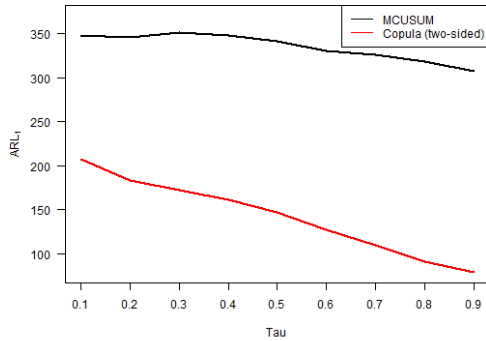
Figure 5.19: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Gumbel copula.

From Figure 5.20 and the fact that we set $\tau = 0.56$, we conclude that the copula-based chart massively outperforms the T^2 chart when the underlying copula structure shifts to a Clayton,

Frank or Normal copula. In case the copula shifts to a Normal copula, the resulting ARL_1 values are unacceptably high. If the underlying copula shifts to the Student-t copula, the copula-based chart seems to perform slightly better. Hence, it does not make any difference whether the copula family remains Archimedean or switches to an elliptical type.
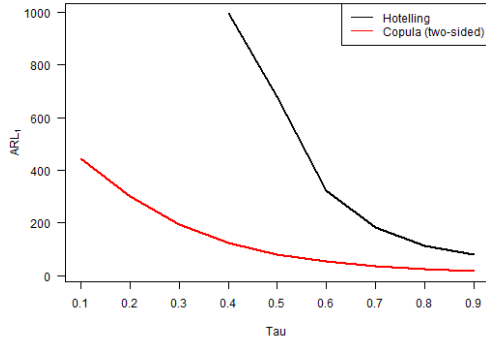

(a) Shift to Clayton copula.
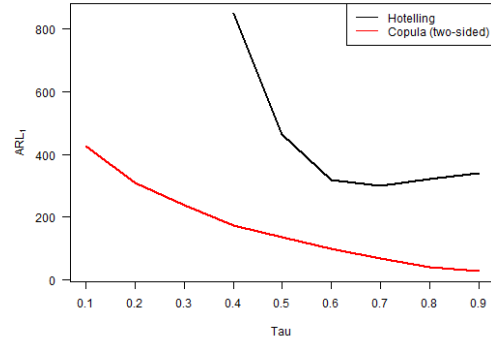
(b) Shift to Frank copula.
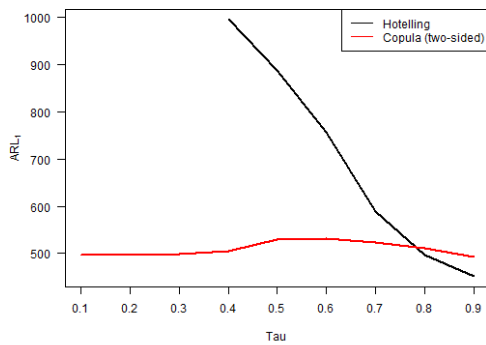
(c) Shift to Normal copula.
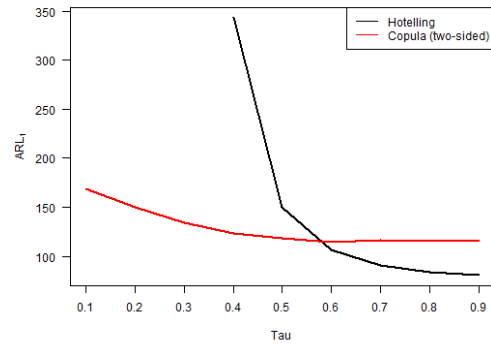
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.20: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Gumbel copula and shifts to a different copula.

Normal (case 59 & cases 91 − 100)

According to Figures 5.21 and 5.22, we can choose $\tau = 0.87$ to achieve that both methods reach approximately the same ARL_0. This corresponds to a high level of dependence between the variables, and the resulting ARL_0 is close to 370. Again, we note that the ARL_0 values of the copula-based chart are constant for any input parameter τ, whereas the ARL_0 values of the T^2 chart are highly influenced by the input parameter τ.
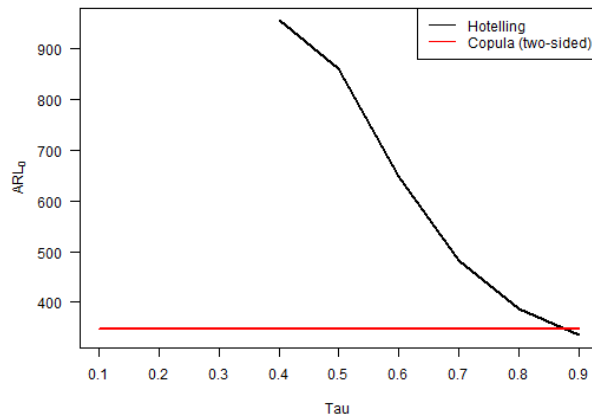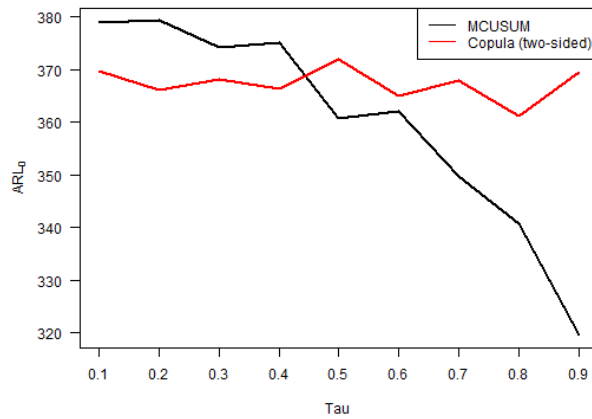
Figure 5.21: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Normal copula.
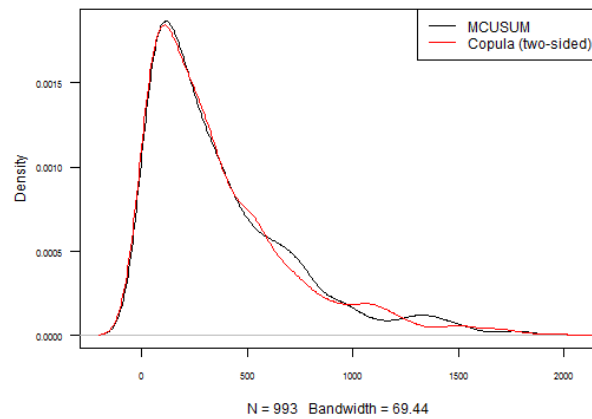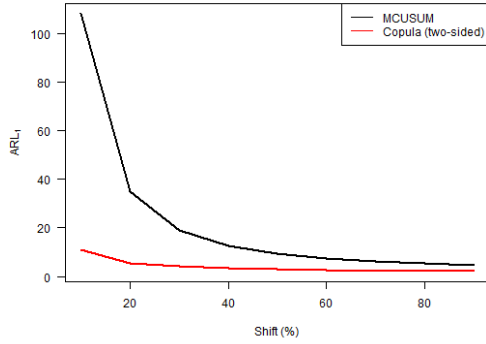

Figure 5.22: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Normal copula, and $\tau = 0.87$.

From Figure 5.23 we conclude that the copula-based chart performs better when the level of one or two variables is increased, especially for small shifts. When the level of the first variable is decreased, both methods seem to perform equally well except for small shift sizes, in which case the copula-based chart performs better. Furthermore, when the level of both variables is decreased, the T^2 chart is unable to detect negative shifts in the level of both variables when the shift size is medium or large. The copula-based chart, however, is able to detect these kinds of shifts.
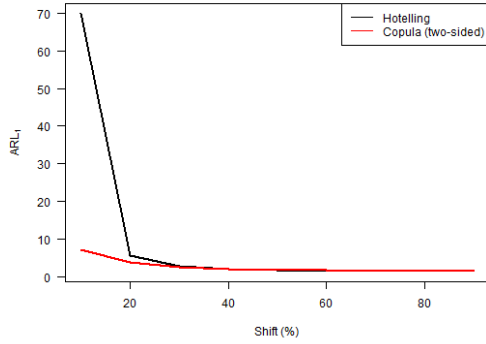
(a) Level of the first variable increased.
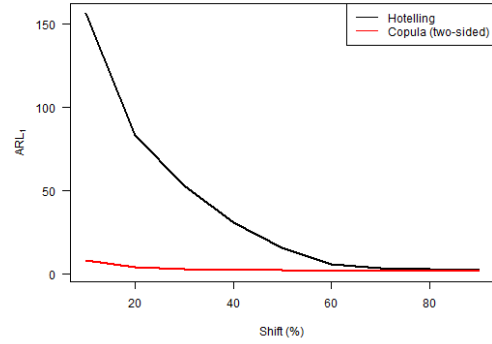
(b) Level of both variables increased.
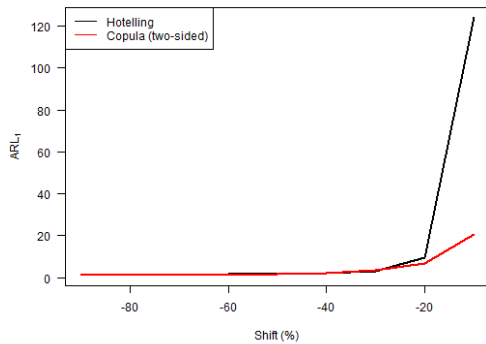
(c) Level of the first variable decreased.
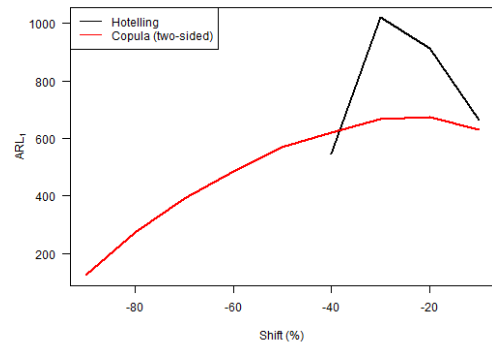
(d) Level of both variables decreased.

Figure 5.23: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Normal copula.

When looking at Figure 5.24, we see that the copula-based chart is significantly better in detecting positive shifts in the correlation matrix for any shift size. Note that the ARL_1 values, however, are quite high. Furthermore, when there is a negative shift in the correlation matrix, the T^2 chart seems to perform slightly better for small shift sizes.
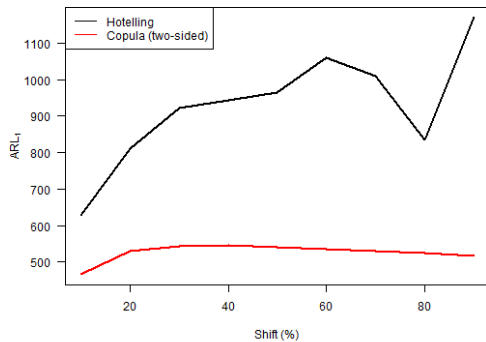
(a) Largest eigenvalue increased.
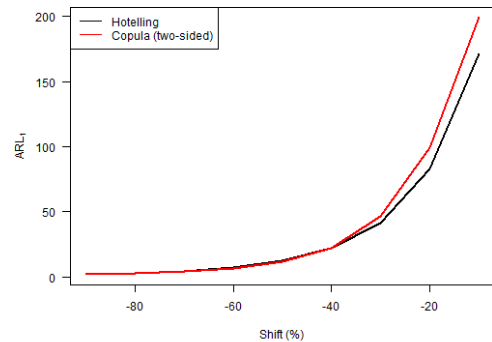
(b) Largest eigenvalue decreased.

Figure 5.24: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Normal copula.

All results in Figure 5.25 more or less exhibit the same behavior. From the fact that we set $\tau = 0.87$, we conclude that the copula-based chart performs significantly better when the underlying copula

shifts to a Gumbel, Clayton or Frank copula. If the underlying copula shifts to the Student-t copula, both methods seem to perform equally well. This could indicate that both methods perform equally well when the underlying copula remains an elliptical copula, and that the copula-based chart performs better when the underlying copula shifts to an Archimedean copula.
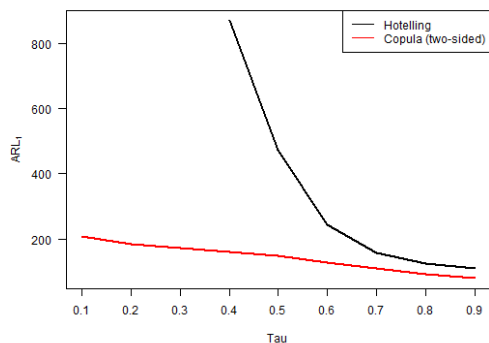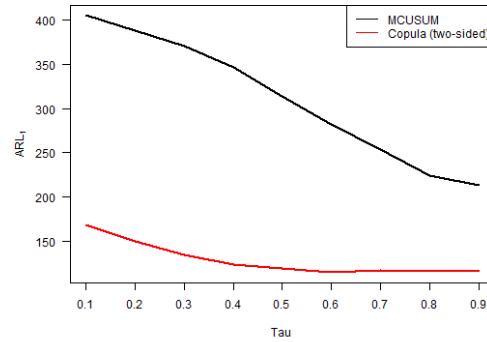

(a) Shift to Gumbel copula.
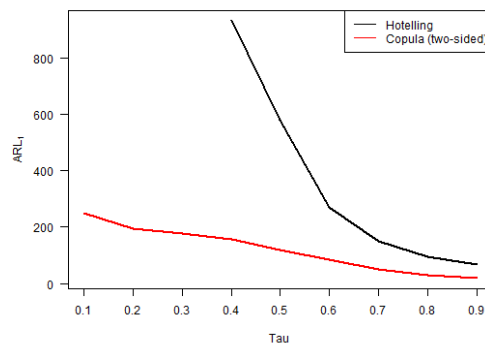
(b) Shift to Clayton copula.
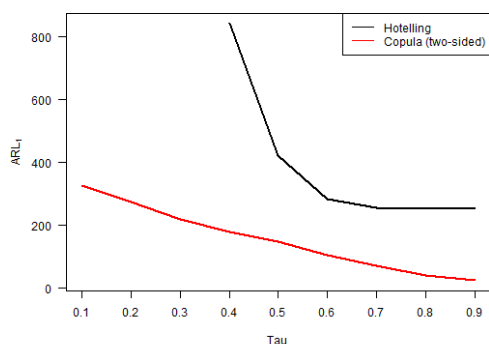
(c) Shift to Frank copula.
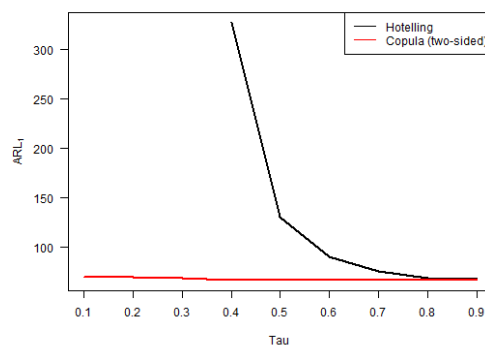
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.25: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Normal copula and shifts to a different copula.

Altogether, we conclude that our main findings are similar in case the in-control distribution follows a Gumbel or Normal copula. In general, we see that the copula-based chart performs better when the level is increased, particularly for small shift sizes. If the level of only one variable is decreased, the copula-based chart seems to perform equal to or better than the T^2 chart in all cases. When the level of both variables is decreased, the T^2 chart is unable to detect medium or large shift sizes. The copula-based chart seems to be able to detect these types of shifts, although for small shifts the resulting ARL_1 is occasionally higher than the ARL_1 of the T^2 chart. Furthermore, the copula-based chart performs significantly better for positive shifts in the correlation matrix, whereas the T^2 chart performs slightly better for small or medium negative shifts in the correlation matrix. Additionally, we notice that both methods perform equally well for large negative shifts in the correlation matrix. In case the underlying copula shifts to a different copula, no matter which copula family it belongs to, the copula-based chart seems to perform equal to or better than the T^2 chart. Most of these findings are in line with the results in case the in-control distribution follows a Clayton, Frank or Student-t copula, which can be found in Appendix B.2. There are some differences, but these do not influence our conclusions. For instance, when the in-control distribution follows a Clayton or Frank copula, the T^2 chart seems to perform better in general if the level of one variable is decreased. Also, when the in-control distribution follows a Frank or Student-t copula, the T^2 chart performs better in general for negative shifts in the correlation matrix.

Consequently, one should use the copula-based chart instead of the T^2 chart in general to detect mean shifts of arbitrary type. Also, we advise to use the copula-based chart in order to detect arbitrary shifts in the correlation matrix, as the significant difference in performance for positive shifts outweigh the smaller difference for negative shifts. Finally, we recommend to use the copula-based chart in order to detect a shift in the dependence structure, since it performs equal to or better than the T^2 chart. Overall, we propose to use the copula-based chart in general to detect shifts of arbitrary type. Note, however, that there are some cases in which the T^2 chart is more effective.

5.2.5 Copula-based control chart vs. MCUSUM

$k = 0.25$

Gumbel (case 111 & cases $116 - 125$)

According to Figures 5.26 and 5.27, we can set $\tau = 0.80$ to achieve that both methods reach approximately the same ARL_0. Note that we could have chosen different values for τ. Again, we observe that the ARL_0 of both methods is approximately equal to 370, which suggests that the ARL_0 could be equal to $1/\alpha$. Furthermore, the ARL_0 values of both methods do not differ that much for any input parameter τ.
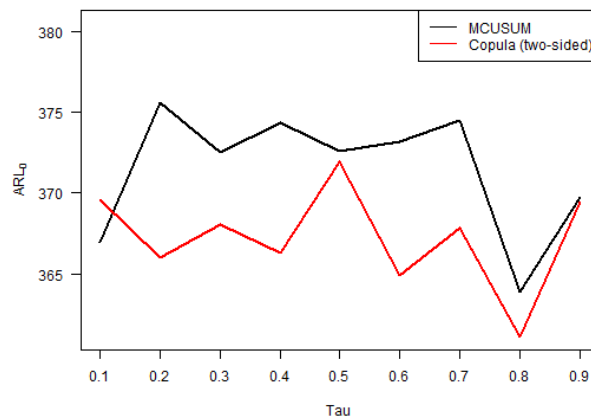
Figure 5.26: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Gumbel copula.
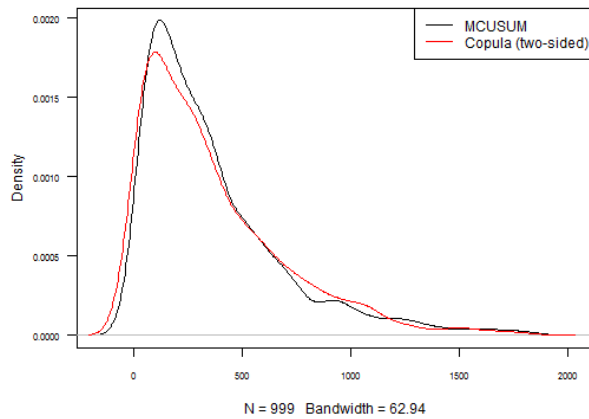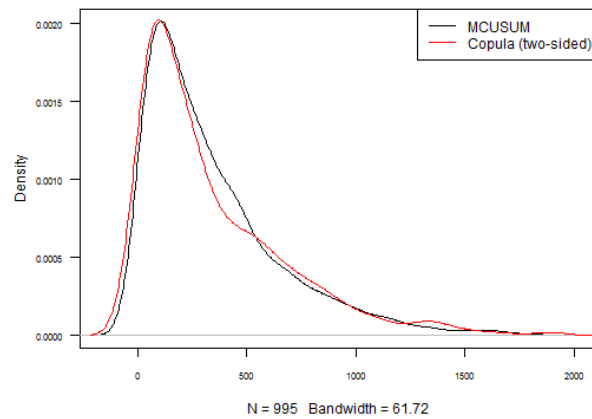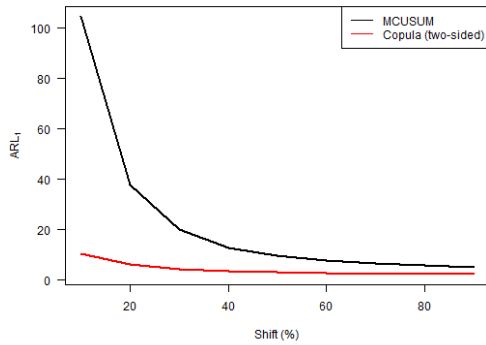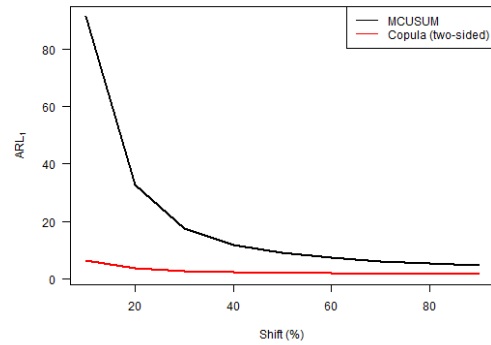
Figure 5.27: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Gumbel copula, and $\tau = 0.80$.

From Figure 5.28 we conclude that the copula-based chart performs better when the level of one or two variables is increased, especially for small shifts. When the level of the first variable is decreased, both methods seem to perform equally well except for small shift sizes, in which case the MCUSUM chart performs slightly better. Furthermore, when the level of both variables is decreased, the MCUSUM chart massively outperforms the copula-based chart for any shift size.
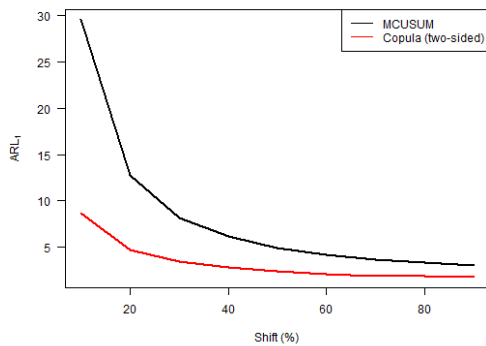
(a) Level of the first variable increased.
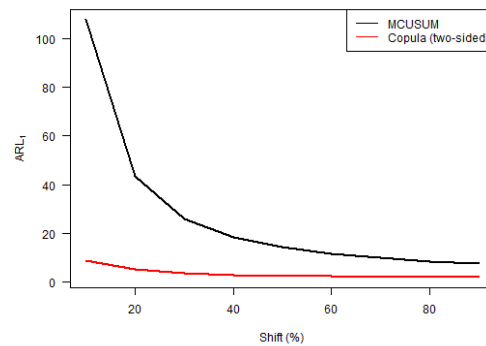
(b) Level of both variables increased.
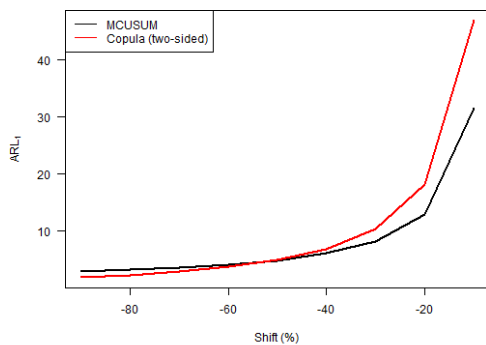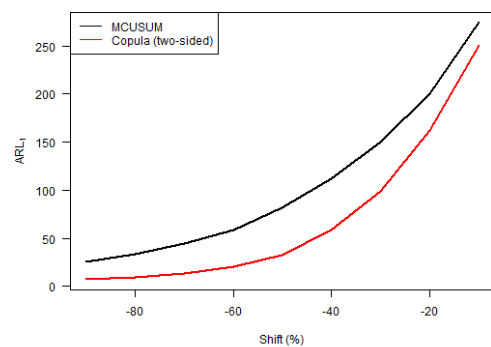

(c) Level of the first variable decreased.
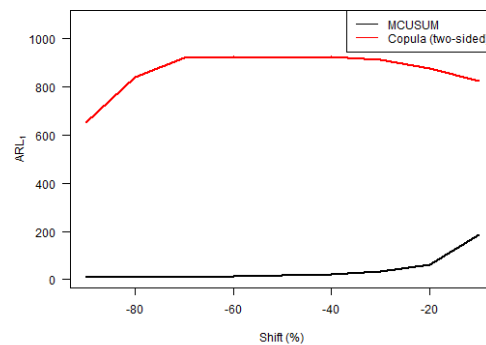
(d) Level of both variables decreased.
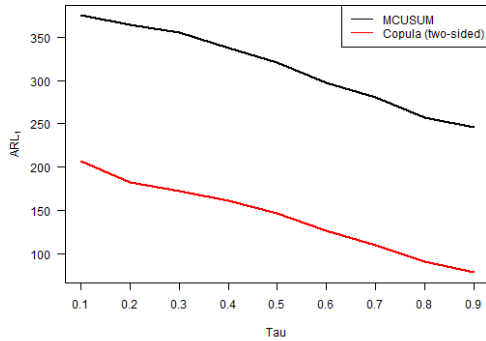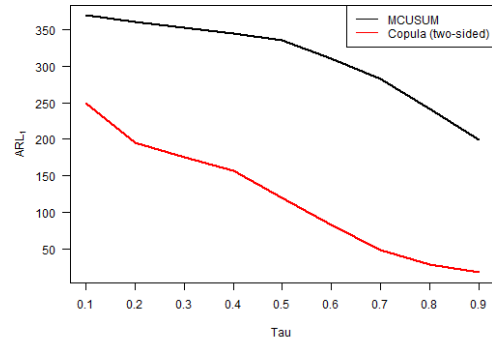
Figure 5.28: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Gumbel copula.

75

When observing Figure 5.29, we conclude that the copula-based chart is better in detecting positive shifts in the correlation matrix for any shift size. Note that the ARL_1 values, however, are higher than the ARL_0. Furthermore, the copula-based chart seems to perform better when there is a negative shift in the correlation matrix except for small shifts, in which case the MCUSUM chart performs slightly better.

(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.
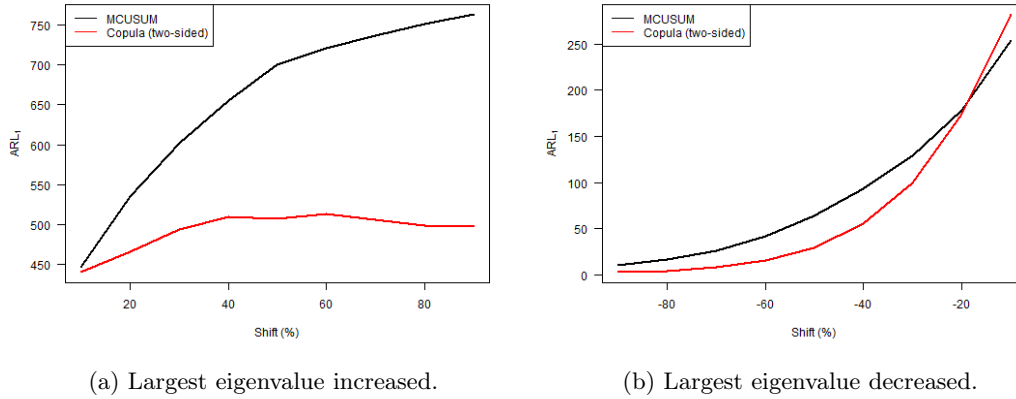
Figure 5.29: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Gumbel copula.

From Figure 5.30 and the fact that we set $\tau = 0.80$, we conclude that the copula-based chart massively outperforms the MCUSUM chart when the underlying copula shifts to a Clayton, Frank or Student-t copula. If the underlying copula shifts to the Normal copula, the MCUSUM chart seems to perform better. The resulting ARL_1 values, however, are higher than the ARL_0. We could have chosen a different value for τ, however, this will most likely not influence our findings. Only if we would have chosen a very small τ, the MCUSUM chart performs slightly better in case the copula shifts to a Clayton or Frank copula.
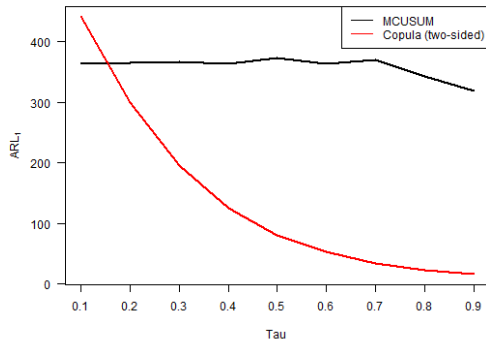
(a) Shift to Clayton copula.
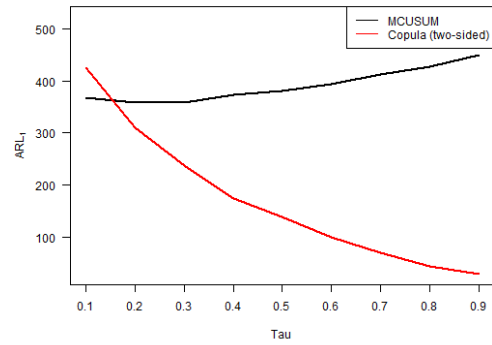
(b) Shift to Frank copula.
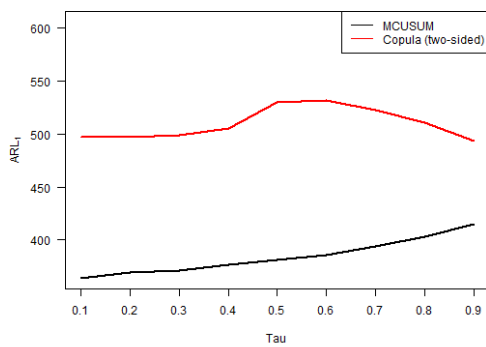
(c) Shift to Normal copula.
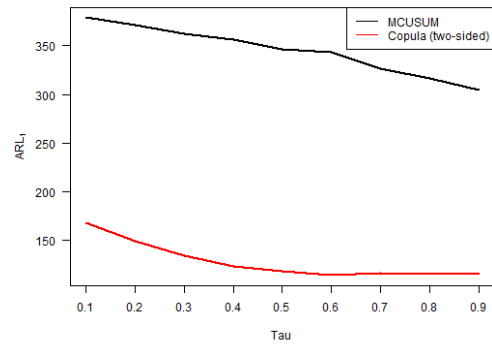
(d) Shift to Student-t copula with 3 degrees of freedom.
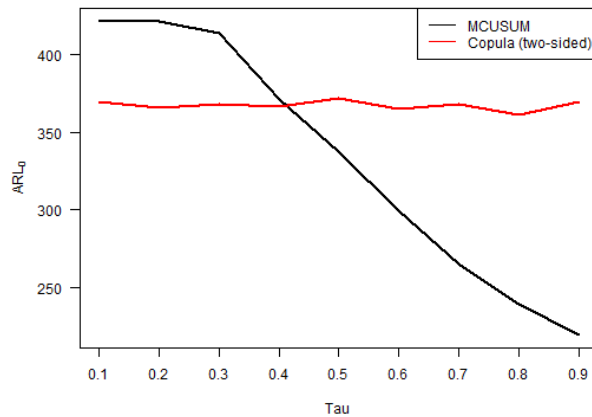
Figure 5.30: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Gumbel copula and shifts to a different copula.

Normal (case 114 & cases $146 - 155$)

According to Figures 5.31 and 5.32, we can set $\tau = 0.80$ in order to achieve that both methods reach approximately the same ARL_0. Note that we could have chosen more appropriate values for τ. We see that the ARL_0 of both methods is approximately equal to 347. Furthermore, the ARL_0 values of the copula-based chart are constant, whereas the ARL_0 values of the MCUSUM chart vary for different τ.
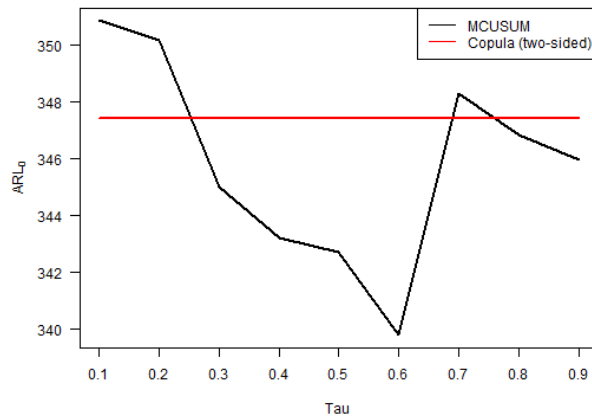
Figure 5.31: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Normal copula.

Figure 5.32: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Normal copula, and $\tau = 0.80$.

From Figure 5.33 we conclude that the copula-based chart performs better when the level of one or two variables is increased, particularly for small shifts. When the level of the first variable is decreased, the copula-based chart performs slightly better except for small shifts, in which case the MCUSUM chart seems to perform better. The main difference occurs when the level of both variables is decreased. In this case the MCUSUM chart massively outperforms the copula-based chart for any shift size.

(a) Level of the first variable increased.

(b) Level of both variables increased.
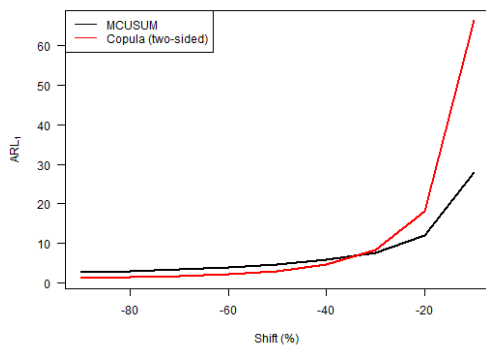
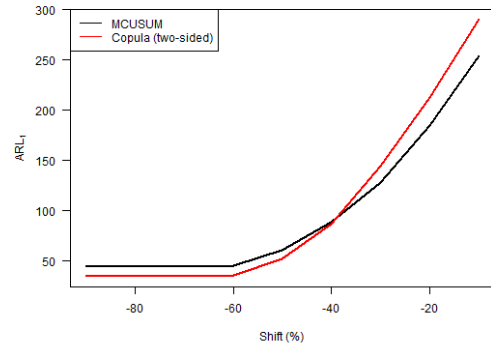(c) Level of the first variable decreased.
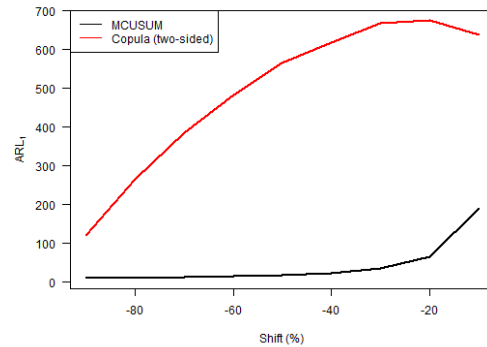
(d) Level of both variables decreased.

Figure 5.33: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Normal copula.

From Figure 5.34, we conclude that the copula-based chart is better in detecting positive shifts in the correlation matrix, except for small shift sizes. Again, note that for this type of shift the ARL_1 values are quite high. Moreover, the copula-based chart seems to perform better in general when there is a negative shift in the correlation matrix.
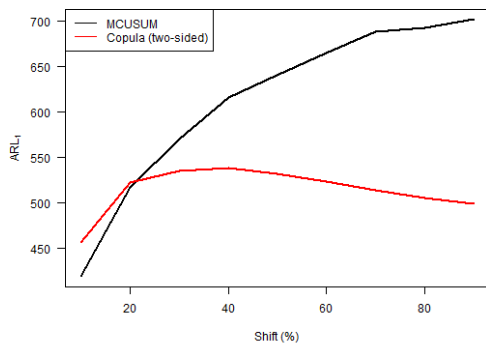

(a) Largest eigenvalue increased.
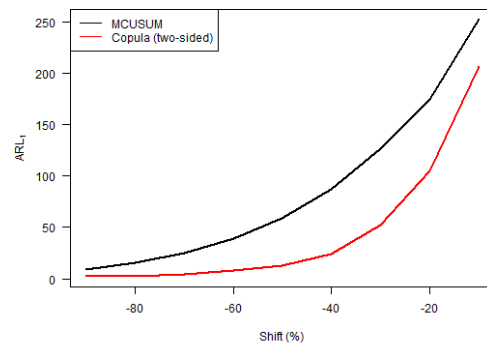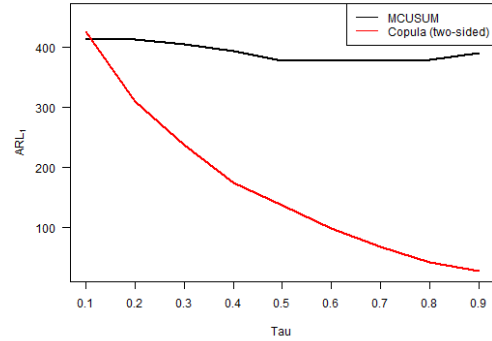
(b) Largest eigenvalue decreased.

Figure 5.34: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Normal copula.

From Figure 5.35 and the fact that we set $\tau = 0.80$, we deduce that the copula-based chart massively

outperforms the MCUSUM chart when the underlying copula shifts to a different copula. We could have chosen a different value for τ, however, this will not influence our conclusions at all.
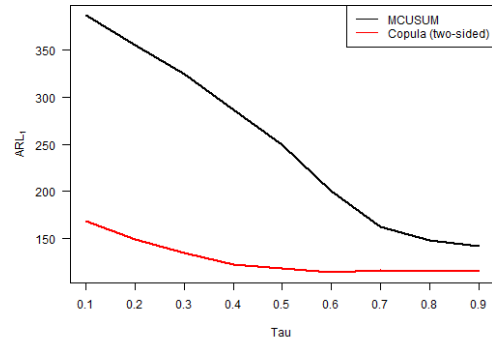

(a) Shift to Gumbel copula.
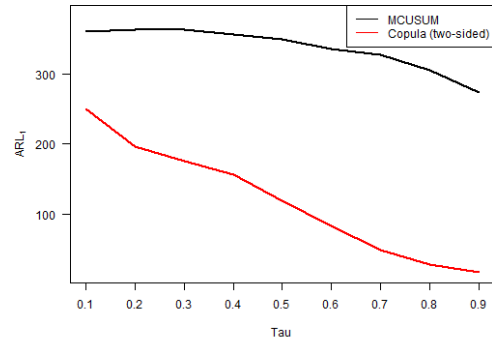
(b) Shift to Clayton copula.
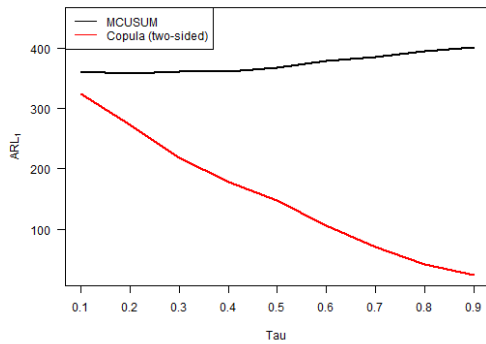
(c) Shift to Frank copula.
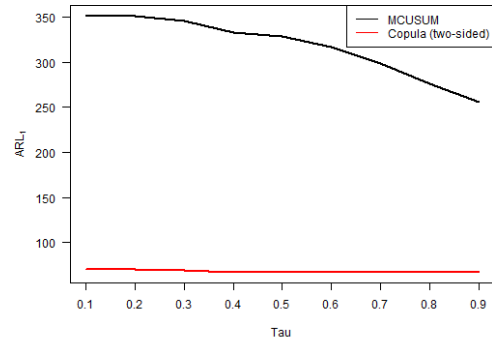
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.35: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Normal copula and shifts to a different copula.

$k = 0.50$

Gumbel (case 166 & cases 171 − 180)

According to Figures 5.36 and 5.37, we can choose $\tau = 0.44$ to achieve that both methods reach approximately the same ARL_0. The resulting ARL_0 is close to 370. Furthermore, we note that the ARL_0 values of the copula-based chart are roughly constant for any input parameter τ, whereas the ARL_0 values of the MCUSUM chart are influenced by the input parameter τ.
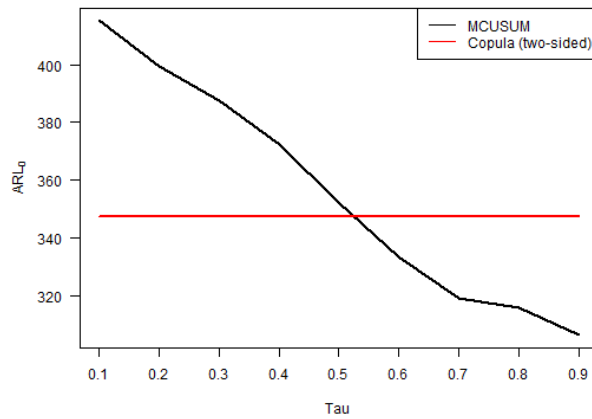
Figure 5.36: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Gumbel copula.
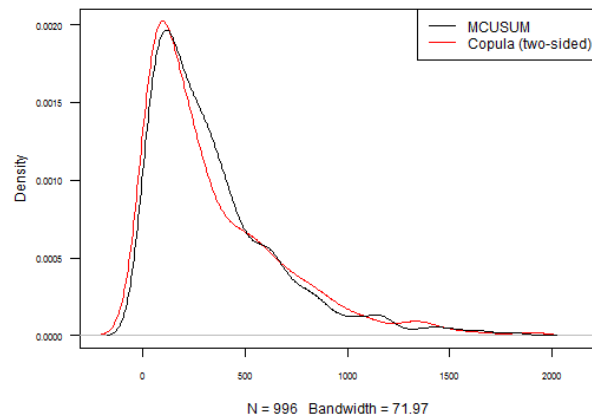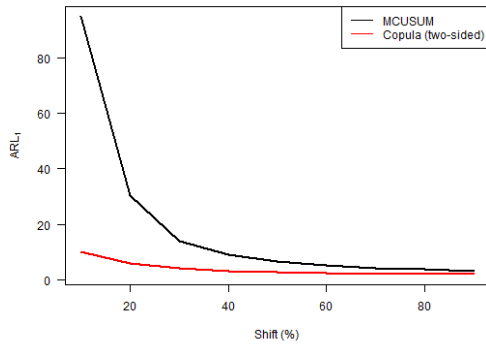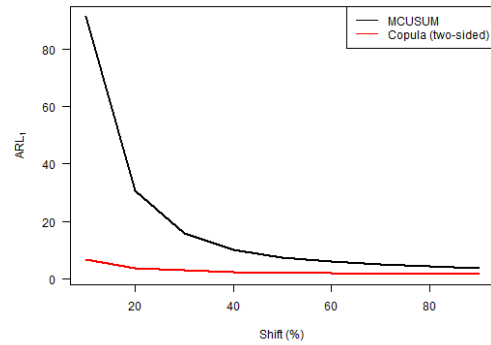

Figure 5.37: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Gumbel copula, and $\tau = 0.44$.

From Figure 5.38 we deduce that the copula-based chart performs better when the level of one or two variables is increased, especially for small shift sizes. On the other hand, the MCUSUM chart performs significantly better in general if the level of one or two variables is decreased.
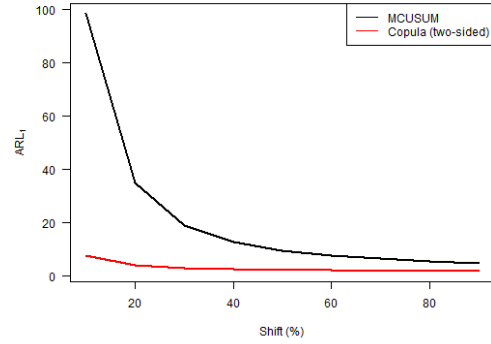
(a) Level of the first variable increased.

(b) Level of both variables increased.
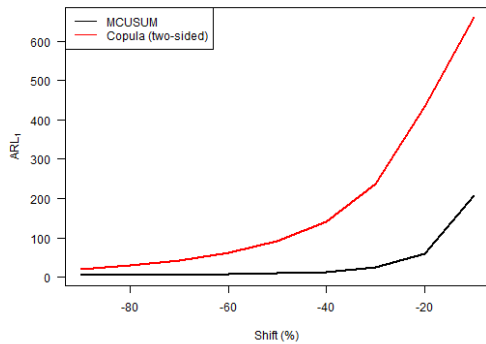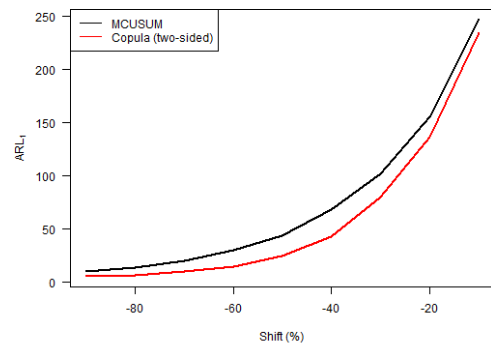

(c) Level of the first variable decreased.
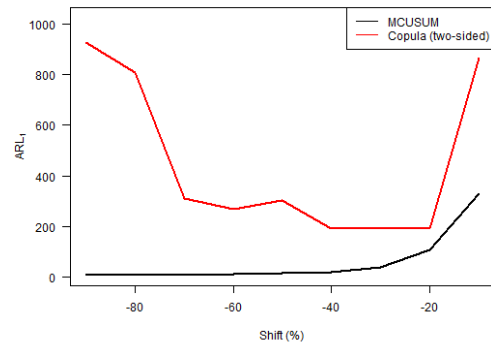
(d) Level of both variables decreased.

Figure 5.38: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Gumbel copula.

From Figure 5.39, we conclude that the copula-based chart is better in detecting positive shifts in the correlation matrix, especially when the shift size is medium or large. Note that once again the ARL_1 values are higher than the ARL_0. Furthermore, the copula-based chart seems to perform better in general when there is a negative shift in the correlation matrix.
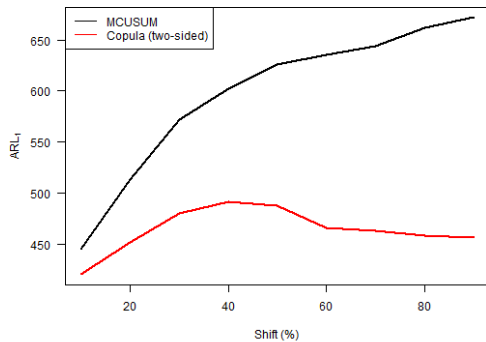

(a) Largest eigenvalue increased.
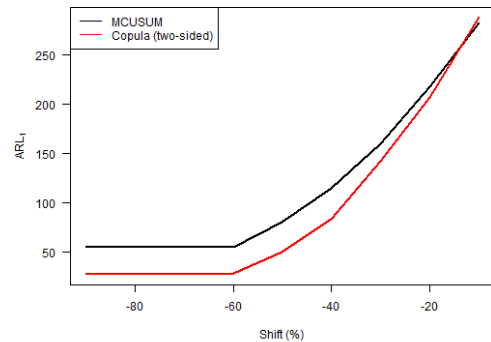
(b) Largest eigenvalue decreased.

Figure 5.39: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Gumbel copula.

From Figure 5.40 and the fact that we set $\tau = 0.44$, we see that the copula-based chart performs

better than the MCUSUM chart when the underlying copula shifts to a Clayton, Frank or Student-t copula. Furthermore, if the underlying copula shifts to the Normal copula, the MCUSUM chart seems to perform better although the resulting ARL_1 is quite poor.
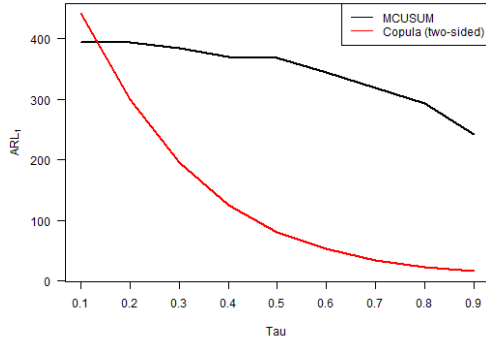
(a) Shift to Clayton copula.
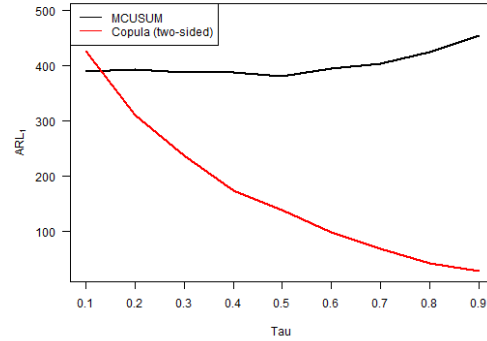
(b) Shift to Frank copula.
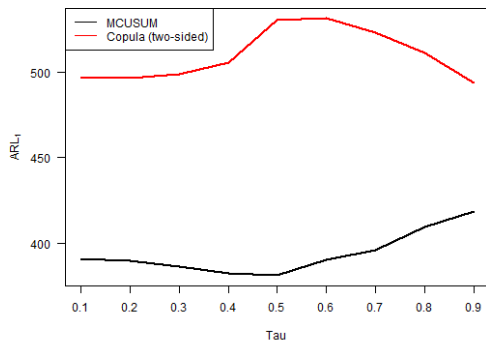
(c) Shift to Normal copula.

(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.40: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Gumbel copula and shifts to a different copula.

Normal (case 169 & cases 201 − 210)

According to Figures 5.41 and 5.42, we can set $\tau = 0.40$ to accomplish that both methods reach approximately the same ARL_0. The resulting ARL_0 is a little less than 350. Furthermore, we note that the ARL_0 values of the copula-based chart are constant for any input parameter τ, whereas the ARL_0 values of the MCUSUM chart are significantly influenced by the input parameter τ.
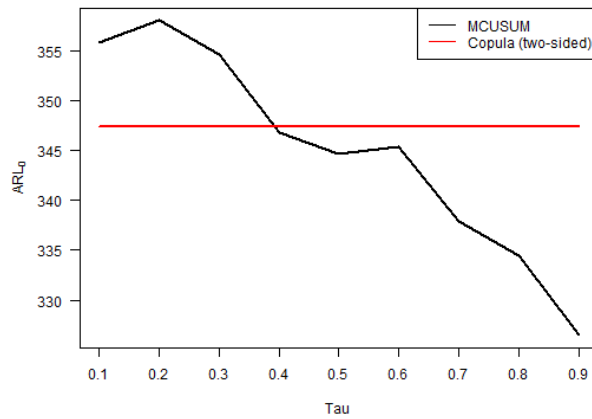
Figure 5.41: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Normal copula.

Figure 5.42: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Normal copula, and $\tau = 0.40$.

From Figure 5.43 we conclude that the copula-based chart performs better when the level of one or two variables is increased, especially for small shift sizes. On the other hand, the MCUSUM chart massively outperforms the copula-based chart when the level of one or two variables is decreased, particularly if the shift size is small or medium.

(a) Level of the first variable increased.

(b) Level of both variables increased.
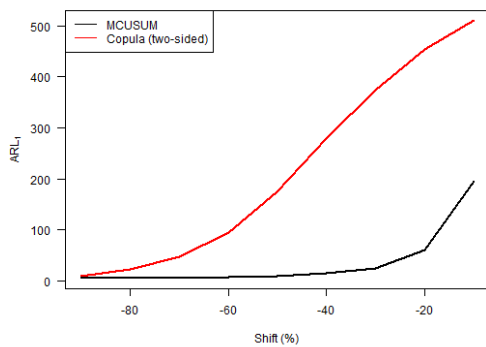
(c) Level of the first variable decreased.
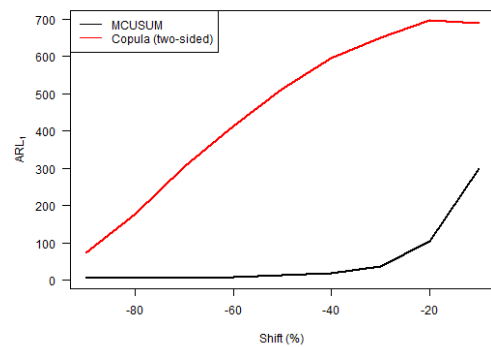
(d) Level of both variables decreased.

Figure 5.43: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Normal copula.

When observing Figure 5.44, we conclude that the copula-based chart is better in detecting positive shifts in the correlation matrix, especially when the shift size is medium or large. Note that once again the ARL_1 values are on the high side. Besides that, the copula-based chart performs better in general when there is a negative shift in the correlation matrix.
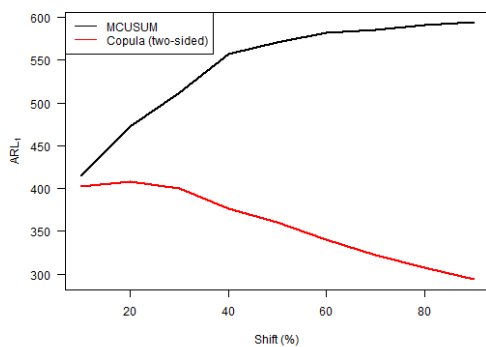
(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure 5.44: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Normal copula.

The results in Figure 5.45 are quite obvious. We simply conclude that the copula-based chart

performs significantly better when the underlying copula shifts to a different copula. This holds for any input parameter τ.

(a) Shift to Gumbel copula.

(b) Shift to Clayton copula.
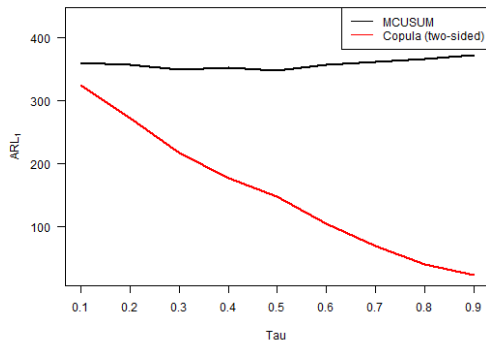
(c) Shift to Frank copula.
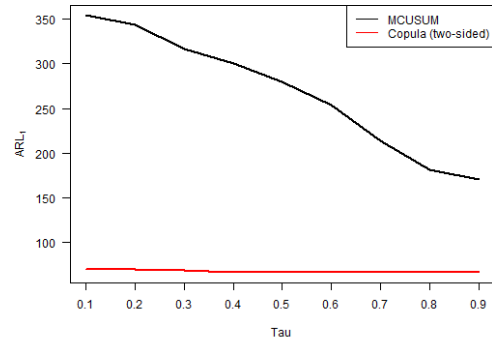
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.45: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Normal copula and shifts to a different copula.

$k = 0.75$

Gumbel (case 221 & cases $226 - 235$)

When looking at Figures 5.46 and 5.47, we can set $\tau = 0.41$ to accomplish that both methods reach approximately the same ARL_0. The resulting ARL_0 is close to 370 in this case. Furthermore, we observe that the ARL_0 values of the copula-based chart are roughly constant for any input parameter τ, whereas the ARL_0 values of the MCUSUM chart are highly influenced by the input parameter τ.

Figure 5.46: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Gumbel copula.
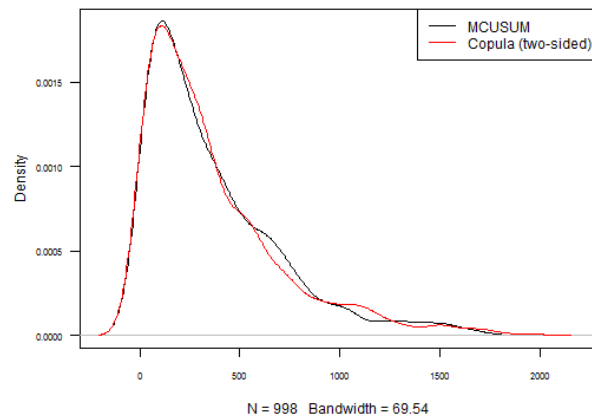
Figure 5.47: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Gumbel copula, and $\tau = 0.41$.

From Figure 5.48 we deduce that the copula-based chart performs better when the level is increased, especially for small shift sizes. On the other hand, the MCUSUM chart performs significantly better in general if the level is decreased. The copula-based chart is even unable to detect small shifts in case the level of both variables is decreased.
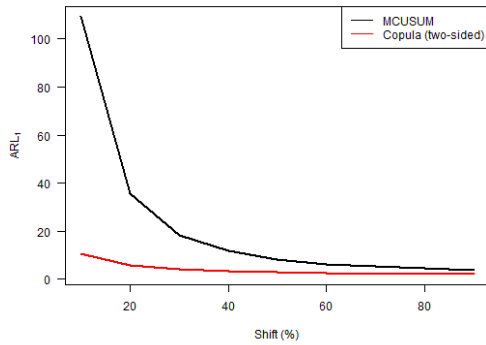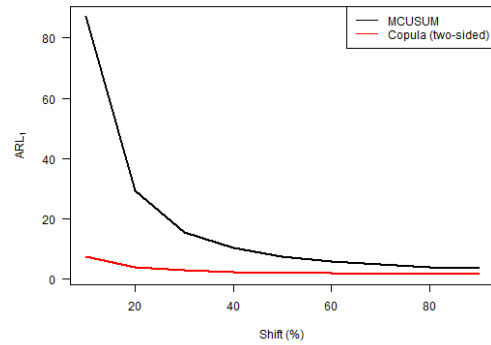
(a) Level of the first variable increased.

(b) Level of both variables increased.
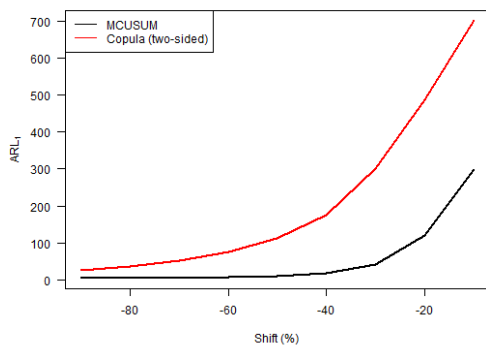
(c) Level of the first variable decreased.
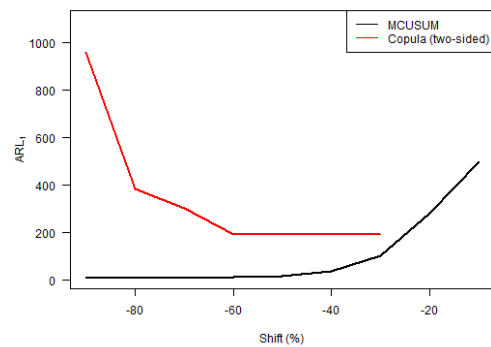
(d) Level of both variables decreased.

Figure 5.48: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Gumbel copula.

When observing Figure 5.49, we conclude that the copula-based chart is better in detecting positive shifts in the correlation matrix for any shift size. Again, the ARL_1 values are higher than the ARL_0. Furthermore, the copula-based chart seems to perform slightly better when there is a negative shift in the correlation matrix except for small shifts, in which case the MCUSUM chart performs slightly better.
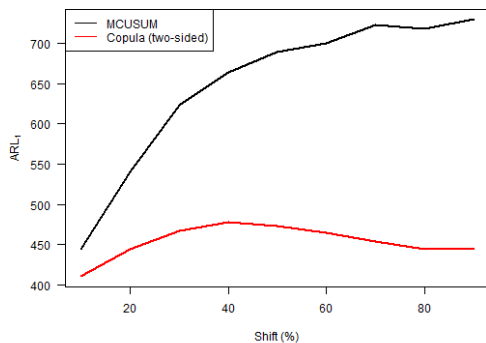
(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure 5.49: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Gumbel copula.

From Figure 5.50 and the fact that we set $\tau = 0.41$, we see that the copula-based chart performs

significantly better than the MCUSUM chart when the underlying copula shifts to a Clayton, Frank or Student-t copula. Furthermore, if the underlying copula shifts to the Normal copula, the MCUSUM chart seems to perform better although the resulting ARL_1 is undesirable.
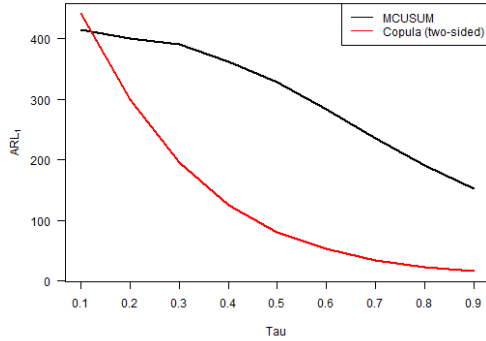
(a) Shift to Clayton copula.

(b) Shift to Frank copula.
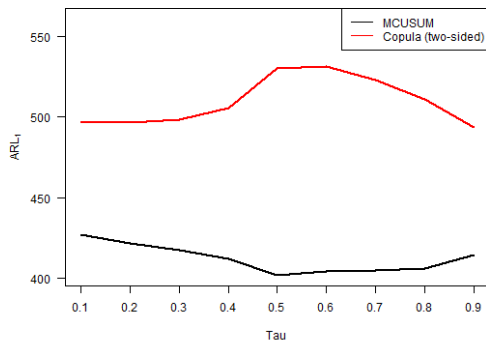
(c) Shift to Normal copula.

(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.50: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Gumbel copula and shifts to a different copula.

Normal (case 224 & cases 256 − 265)

According to Figures 5.51 and 5.52, we can set $\tau = 0.52$ to accomplish that both methods reach approximately the same ARL_0. Again, when the in-control distribution is a Normal copula, the resulting ARL_0 is close to 350. Furthermore, we note that the ARL_0 values of the copula-based chart are constant for any input parameter τ, whereas the ARL_0 values of the MCUSUM chart are influenced by the input parameter τ.

Figure 5.51: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Normal copula.

Figure 5.52: Kernel density plot of the in-control run length for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Normal copula, and $\tau = 0.52$.

From Figure 5.53 we conclude that the copula-based chart performs better when the level of one or two variables is increased, especially for small shift sizes. On the other hand, the MCUSUM chart significantly outperforms the copula-based chart when the level of one or two variables is decreased.

(a) Level of the first variable increased.

(b) Level of both variables increased.
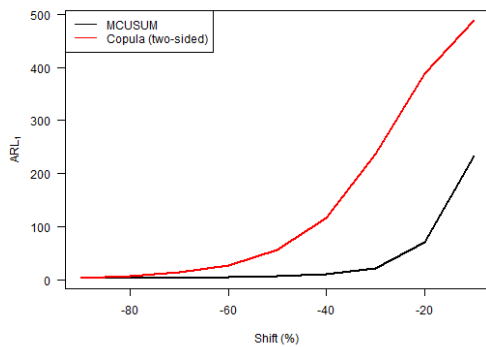
(c) Level of the first variable decreased.
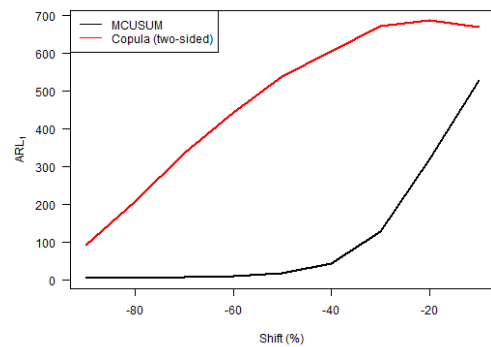
(d) Level of both variables decreased.

Figure 5.53: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Normal copula.

When observing Figure 5.54, we conclude that the copula-based chart is significantly better in detecting positive shifts in the correlation matrix. Note that once again the ARL_1 values are unacceptable. Furthermore, the copula-based chart seems to perform better in general when there is a negative shift in the correlation matrix.
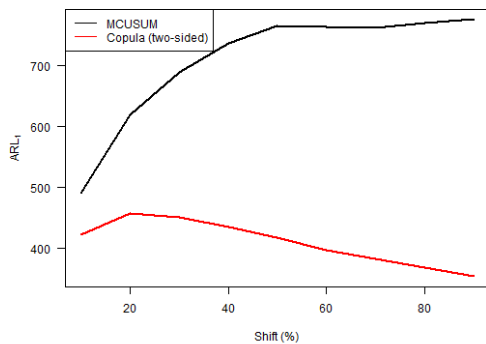
(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure 5.54: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Normal copula.

The results in Figure 5.55 are very straightforward. We simply conclude that the copula-based

chart performs significantly better when the underlying copula shifts to a different copula. This holds for any input parameter τ.
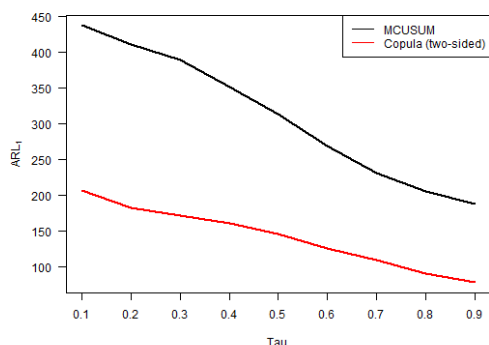
(a) Shift to Gumbel copula.
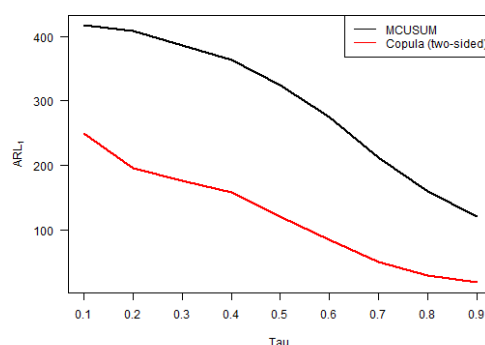
(b) Shift to Clayton copula.
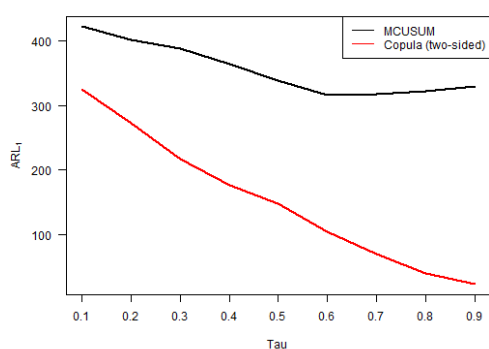
(c) Shift to Frank copula.
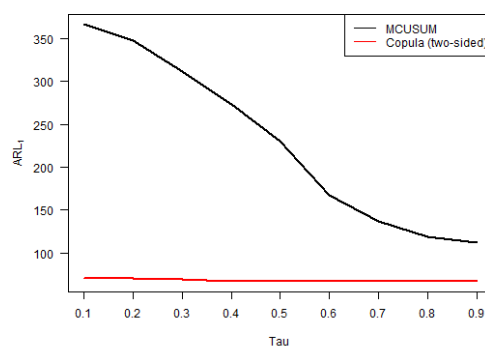
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure 5.55: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Normal copula and shifts to a different copula.

Overall, we conclude that our main findings are similar for all investigated values of k in case the in-control distribution follows a Gumbel or Normal copula. In general, we see that the copula-based chart performs better when the level of one or two variables is increased, particularly for small shift sizes. When the level of just one variable is decreased, the MCUSUM chart seems to perform better than or equal to the copula-based chart in general. Also, when the level of both variables is decreased, the MCUSUM chart massively outperforms the copula-based chart. Furthermore, the copula-based chart performs better for positive shifts in the correlation matrix in general. For negative shifts in the correlation matrix, the copula-based chart performs better except for small shifts, in which case the MCUSUM chart occasionally performs slightly better. Finally, when the underlying copula shifts to a different copula, the copula-based chart performs significantly better than the MCUSUM chart, except when the copula shifts to the Normal copula. In this case both methods perform poorly, but the MCUSUM chart seems to be the most effective chart. Most of these findings are in line with the results in case the in-control distribution follows a Clayton, Frank or Student-t copula, which can be found in Appendix B.3. There are some differences which are worth mentioning. For instance, when the in-control distribution is a Clayton or Frank copula, the copula-based chart performs better if the underlying copula shifts to a Normal copula. Besides that, the resulting ARL_1 values are reasonable. Also, when the in-control distribution follows a Frank copula and $k = 0.75$, the MCUSUM chart seems to perform better for negative shifts in the correlation matrix. Furthermore, if the in-control distribution is a Student-t copula, the MCUSUM chart performs significantly better for negative shifts in the correlation matrix. Finally, when the in-control distribution is a Student-t copula and $k = 0.25$, the MCUSUM chart performs slightly

better if the underlying copula shifts to a Gumbel copula. The resulting ARL_1 values, however, are undesirable.

All things considered, we recommend to choose the MCUSUM chart with $k = 0.25$ over the copula-based chart to detect mean shifts of arbitrary type, since this MCUSUM chart performs significantly better when the level is decreased and is able to detect level increases within a reasonable time as well. Furthermore, we advise to use the copula-based chart in order to detect arbitrary shifts in the correlation matrix, as it performs better for positive and negative shifts in general except on a few occasions. Finally, we propose to use the copula-based chart in order to detect a shift in the dependence structure, since it significantly outperforms the MCUSUM chart. The only exception is when the copula shifts to the Normal copula, in which case the MCUSUM chart performs better on some occasions. Overall, we recommend to use the copula-based chart in general to detect shifts of arbitrary type. Note, however, that there seem to be quite some cases in which the MCUSUM chart is much more effective, particularly in detecting mean shifts.

5.3   Summary

- The MCUSUM chart with $k = 0.25$ seems to be a suitable choice for detecting arbitrary mean shifts in an efficient way.
- The two-sided version of the copula-based control chart is the obvious choice in order to detect arbitrary shifts in the correlation matrix.
- The one-sided version of the copula-based control chart is the optimal choice, performance-wise, if one desires to detect a shift in the dependence structure.
- The two-sided version of the copula-based control chart seems to be more efficient than the one-sided version in detecting shifts of arbitrary type.
- The two-sided copula-based control chart outperforms the conventional control charts in detecting shifts of almost any type.
- All of the investigated methods seem to have great difficulty with timely detection of positive shifts in the correlation matrix.
- In case the in-control distribution follows a Gumbel or Student-t copula, all of the investigated methods have a lot of trouble with timely detection when the dependence structure shifts to a Normal copula.

6 | Wind turbine case study

In this chapter we apply the two-sided version of the vine copula-based monitoring approach, which is explained in Section 4.3, to a real-life data set of a wind turbine. More specifically, we combine the univariate multiple linear regression models found by Kenbeek (2016) to obtain suitable multivariate multiple linear regression models, which are discussed in Chapter 2. Subsequently, we apply the two-sided version of the vine-copula based chart and the Hotelling T^2 chart to the residuals and investigate the differences in performance. Finally, we compare the results with the findings in van Dalen (2018) to analyze the difference between univariate and multivariate monitoring of the wind turbine data.

6.1 Data description

The wind turbine data is provided by a Dutch energy supplier and consists of 229,680 observations on 110 variables. This data was collected in the period from 19/06/2013 18:32:00 to 18/03/2015 23:56:00 and the observations were measured every four minutes. The most important variables can be subdivided into two categories:

1. **Environmental variables:** variables that describe environmental factors like wind speed and environmental temperature.

2. **Conditional variables:** variables that describe the state of several components of the wind turbine. These variables can be further subdivided into:

   (a) **Speed:** rotor speed and generator speed.

   (b) **Temperatures:** measurements of bearing, gearbox, primary generator, secondary generator, nacelle and oil temperatures.

   (c) **Vibrations:** different vibration measurements on various components.

   (d) **Operational:** power output, operating state, pitch angle and yaw measurements.

In order to get a better understanding of the inside of a wind turbine, most of the related components are visualized in Figure 6.1. Additionally, the most important variable names are listed in Table 6.1. We follow the choice of Kenbeek (2016) to only use the overall vibration readings and do not consider subdivision of the vibration readings into bands of frequency levels. Subsequently, we process the data on the condition that the wind turbine is running and that the primary generator is in use. In this way we should be able to predict failures more precisely, as the wind turbine behaves very differently when the starting generator is in use. After data pre-processing we are left with 30,775 observations. We will analyze the data with the regular four-minute intervals, but also a subsample with intervals of four hours to reduce possible autocorrelation. In Figure 6.2 we present plots of several individual variables versus time based on the subsample with four-hour intervals. Note that there are some large gaps in the data, which correspond to maintenance periods or malfunctioning of sensors. We know that the wind turbine suffered a mechanical failure of the primary generator on December 2014. The identification of the root cause and the magnitude of the problem became apparent only during inspection after the failure. Therefore, it is crucial to adopt an effective monitoring strategy which is able to timely detect indications of imminent failure. For an in-depth analysis of the provided wind turbine data, see Kenbeek (2016).
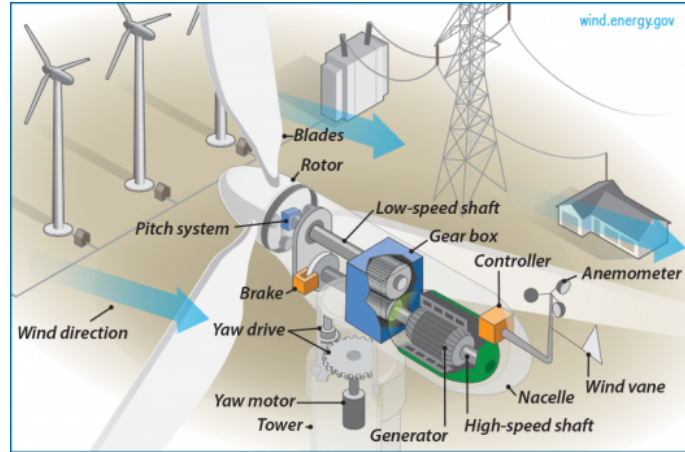
Figure 6.1: Basic overview of the inside of a wind turbine. Photo courtesy of the U.S. Department of Energy.

	Name	Name	
1	Taken	NarcellTemp	23
2	BrgTemp	OilTemp	24
3	EnvTemp	OpState	25
4	GbxISA10KHzOvr	PitchAngle	26
5	GbxISA1KHzOvr	PlanGbxA10KHzOvr	27
6	GbxMSA10KHzOvr	PlanGbxA200HzOvr	28
7	GbxMSA1KHzOvr	PlanGbxD500HzOvr	29
8	GbxMSD500HzOvr	Power	30
9	GbxMSV1KHzOvr	RotorSpeed	31
10	GbxOSA10KHzOvr	ServiceState	32
11	GbxOSA2KHzOvr	ShftGbxA10KHzOvr	33
12	GbxOSD1KHzOvr	ShftGbxA50HzOvr	34
13	GbxTemp	ShftGbxD50HzOvr	35
14	Gen1Temp	ShftGbxV1KHzOvr	36
15	Gen2Temp	ShftPropA10KHzOvr	37
16	GenDEA10KHzOvr	ShftPropA50HzOvr	38
17	GenDED1KHzOvr	ShftPropD50HzOvr	39
18	GenDEV1KHzOvr	ShftPropV1KHzOvr2	40
19	GenNDEA10KHzOvr	WindSpeed	41
20	GenNDED1KHzOvr	YawCCW	42
21	GenNDEV1KHzOvr	YawCW	43
22	GenSpeed	-	44

Table 6.1: The available variables in the data set excluding subdivision of the vibration readings into bands.

95
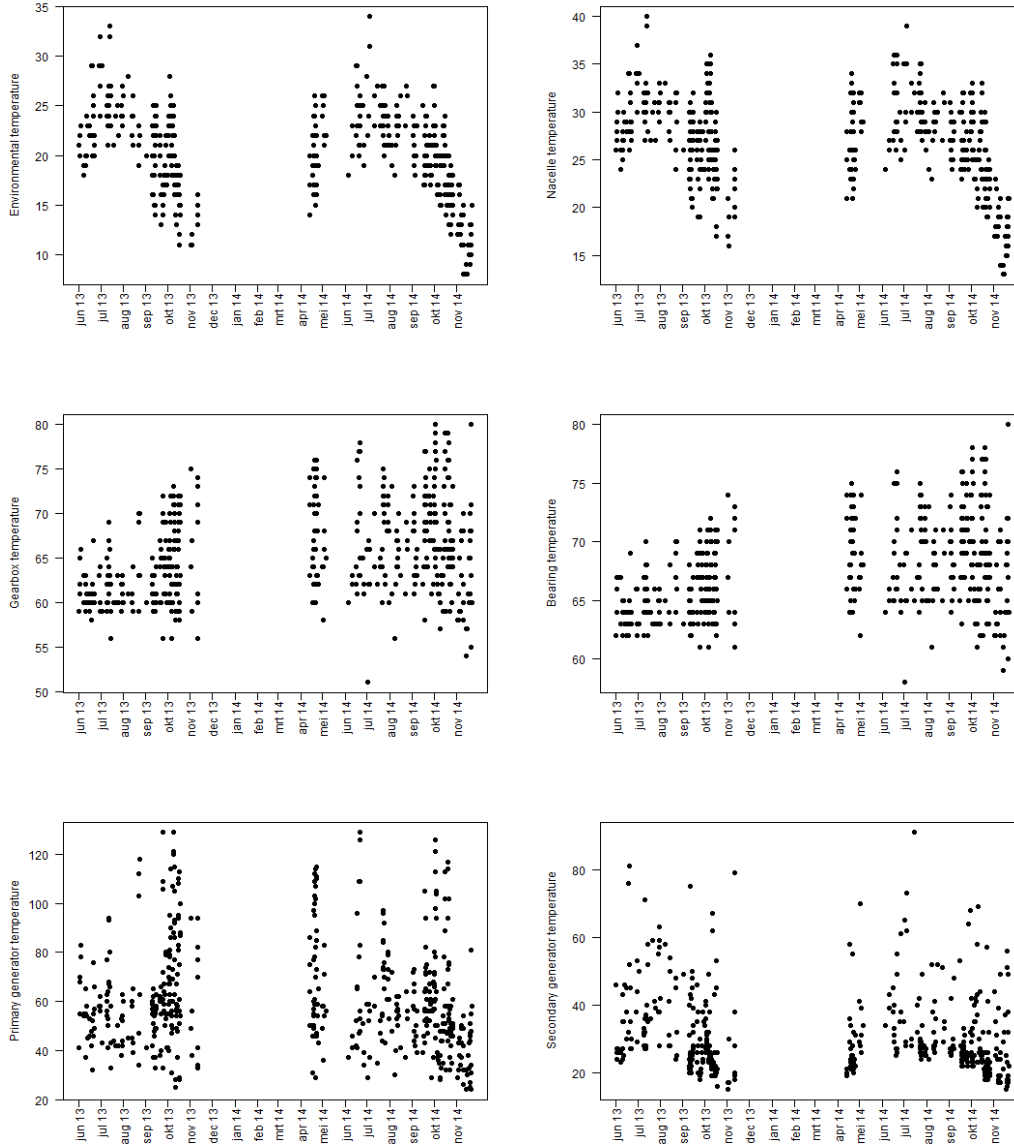
Figure 6.2: The environmental temperature, nacelle temperature, gearbox temperature, bearing temperature, primary generator temperature and secondary generator temperature versus time.

6.2 Model

As explained in Chapter 1, the observations are often not abnormal by themselves, but this state depends on external factors like environmental temperature and wind speed. By first applying a suitable regression model we can describe the observations by these external factors, which enables us to identify contextual anomalies by monitoring the residuals of the model. Since we want to monitor multiple quality characteristics simultaneously, we will fit multivariate multiple linear regression models which we are able to fully understand due to the theoretical work done in Chapter 2. Kenbeek (2016) already performed extremely valuable research and found six suitable univariate multiple linear regression models which are given below. Here, the variables before the "∼" sign denote the response variables and the variables after the "∼" sign denote the regressors of the linear regression models.

1. Nacelle Temperature ∼ Environmental Temperature.
2. Oil Temperature ∼ Environmental Temperature.

3. Bearing Temperature $\sim$ Environmental Temperature + Wind Speed + (Wind Speed)^2 + (Wind Speed)^3 + (Generator Speed)^2 + (Generator Speed)^3.

4. Gearbox Temperature $\sim$ (Environmental Temperature)^3 + (Wind Speed)^2 + Generator Speed + (Generator Speed)^2 + GbxMSD500HzOvr + GbxOSA10KHzOvr + GbxOSA2KHzOvr + GbxOSD1KHzOvr.

5. Primary Generator Temperature $\sim$ Environmental Temperature + Wind Speed + Generator Speed.

6. Secondary Generator Temperature $\sim$ Environmental Temperature + Wind Speed + Generator Speed + Power.

In addition, we assume normally distributed error terms for all of the above mentioned models.

Now, we want to monitor quality characteristics of the wind turbine at the same time, which leads to two obvious multivariate multiple linear regression models.

Model 1 consists of 2 response variables and 1 regressor:

$$\begin{pmatrix} \text{Nacelle Temperature} \\ \text{Oil Temperature} \end{pmatrix} \sim \text{Environmental Temperature}$$

Model 2 consists of 2 response variables and 4 regressors:

$$\begin{pmatrix} \text{Primary Generator Temperature} \\ \text{Secondary Generator Temperature} \end{pmatrix} \sim \begin{matrix} \text{Environmental Temperature} + \text{Wind Speed} + \\ \text{Generator Speed} + \text{Power} \end{matrix}$$

In order to fit the models, we need a large sample of phase I data in which we assume the process to be in-control. Therefore, we will continue this case study with the phase I period determined in Kenbeek (2016), which corresponds to the period from 19/06/2013 up to 12/10/2013. After we fitted the multivariate multiple linear regression models, we apply the two-sided version of the vine-copula based chart and the Hotelling T^2 chart, which are discussed in Chapter 4, to the residuals of the models for both phase I and phase II analysis.

6.3   Results

In this section we create phase I and phase II results for the two models when applying the two-sided version of the vine-copula based chart and the conventional Hotelling T^2 chart to the residuals, using both the regular 4-minute interval and 4-hour interval subsample of the wind turbine data. For both monitoring charts we set $\alpha = 0.0027$, which is the false alarm probability for the univariate Shewhart chart for individual observations in order to reach $\text{ARL}_0 \approx 370$, see Section 5.1. We investigate the signals, which could be indicators of the upcoming failure, and analyze the differences between both methods. Finally, we compare the results with the findings of the case study in van Dalen (2018) in order to investigate the difference between the univariate and multivariate approach of the wind turbine data. In the following figures, the vertical green line indicates the phase I limit determined by Kenbeek (2016).

6.3.1   Model 1

4-minute intervals

When looking at Figure 6.3, we observe that some test statistics for the vine copula-based chart are very high and that most test statistics are grouped in the lower region. In order to get a more detailed view of the test statistics we modify the range in the y-axis, which is visualized in Figures 6.3b and 6.4a. We combine the results of phase I and phase II after applying the vine copula-based chart and the Hotelling T^2 chart, see Figure 6.4. From this figure we conclude that the vine copula-based chart signals way less than the Hotelling T^2 chart. In fact, for the vine copula-based chart, 4 out of 6617 observations are outside the control limits in phase I, and 18 out of 24,158 observations are outside the control limits in phase II. This corresponds to approximately 0.06% and 0.07%, respectively. In total, this leads to 22 out of 30,775 observations outside the control limits, or approximately 0.07%. For the Hotelling T^2 chart, 110 out of 6617 observations, or approximately 1.66%, and 2343 out of 24,158 observations, or approximately 9.70%, are outside

the limit in phase I and phase II, respectively. This corresponds to a total of 2453 out of 30,775 observations, or approximately 7.97%, marked as out-of-control. Hence, based on the vine copula-based chart there are no significant indications that the process is out-of-control. If we assume that, indeed, the process is in-control during phase I, the vine copula-based chart has a lower false alarm rate than desired in phase I. Besides that, given that the wind turbine suffered a failure on December 2014, we would expect more signals in phase II, especially in the period prior to the failure. This could, however, also mean that we have to take these signals very seriously. The first signal during phase II is at 13/10/2013 12:56:00. On the other hand, according to the Hotelling T^2 chart there are clear indications that the process is not functioning as expected. As we would expect, we see many signals in phase II and peaks of signals in the period prior to the failure. However, the Hotelling T^2 chart gives too many false alarms than desired in phase I. In phase II, the Hotelling T^2 chart signals for the first time at 13/10/2013 03:20:00 which is a bit earlier than the vine copula-based chart.

(a) Full range in the y-axis.

(b) Modified range in the y-axis.
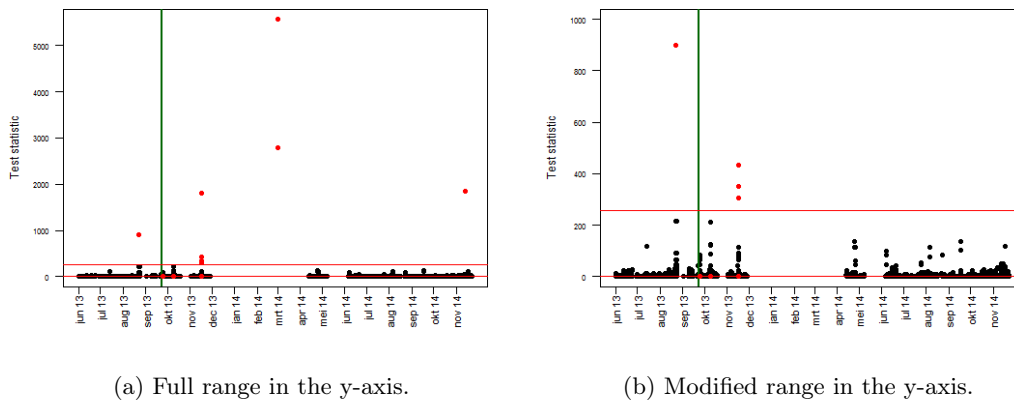
Figure 6.3: Vine copula-based control chart for model 1 when using a phase I period to calculate the regression model, and using 4-minute intervals.

(a) Vine copula-based control chart.

(b) Hotelling T^2 control chart.
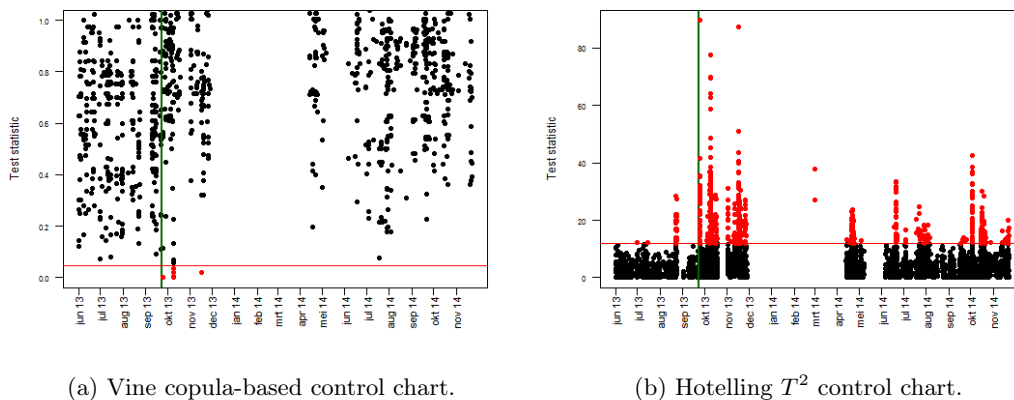
Figure 6.4: Multivariate control charts for model 1 when using a phase I period to calculate the regression model, and using 4-minute intervals.

4-hour intervals

We performed a similar analysis for the subsample based on 4-hour intervals and the corresponding results are visualized in Figure 6.5. From this figure we conclude that the results are comparable. This time the vine copula-based chart does not even signal once which is remarkable. Based on 4-hour intervals the vine copula-based chart is therefore not able to detect the upcoming failure. For the Hotelling T^2 chart, 2 out of 112 observations, or approximately 1.79%, and 28 out of 388 observations, or approximately 7.22%, are outside the limit in phase I and phase II, respectively.

This corresponds to a total of 30 out of 500 observations, or 6.00%, marked as out-of-control. The first signal during phase II is at 13/10/2013 08:00:00. As we would expect, we see clear indications of upcoming failure during phase II. However, the false alarm rate is still higher than desired in phase I.

(a) Vine copula-based control chart.

(b) Hotelling T^2 control chart.
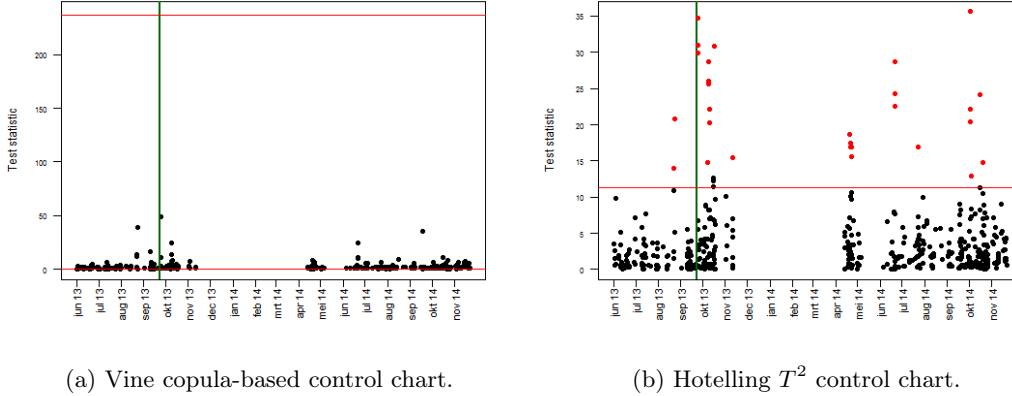
Figure 6.5: Multivariate control charts for model 1 when using a phase I period to calculate the regression model, and using 4-hour intervals.

6.3.2 Model 2

4-minute intervals

The results for model 2 using the regular 4-minute intervals can be seen in Figure 6.6. From this figure we conclude that again the vine copula-based chart signals way less than the Hotelling T^2 chart. However, this time the vine copula-based chart produces more signals than for model 1. In fact, 18 out of 6617 observations are outside the control limits in phase I, and 506 out of 24,158 observations are outside the control limits in phase II. This corresponds to approximately 0.27% and 2.09%, respectively. In total, this leads to 524 out of 30,775 observations outside the control limits, or approximately 1.70%. For the Hotelling T^2 chart, 180 out of 6617 observations, or approximately 2.72%, and 2625 out of 24,158 observations, or approximately 10.87%, are outside the limit in phase I and phase II, respectively. This corresponds to a total of 2805 out of 30,775 observations, or approximately 9.11%, marked as out-of-control. Hence, based on the vine copula-based chart there are some significant indications that the process is out-of-control. Additionally, note that the false alarm rate is approximately 0.27%, as one would expect by setting $\alpha = 0.0027$. Besides, it seems that the amount of signals accumulates in the period prior to the failure which is in line with our expectation. The first signal during phase II is at 12/10/2013 23:16:00 which is also the time of the first observation in phase II. On the other hand, according to the Hotelling T^2 chart there are very clear indications that the process is not functioning as expected. Once again, we see many signals in phase II and peaks of signals in the period prior to the failure. However, the Hotelling T^2 chart for model 2 still produces too many false alarms than desired in phase I. In phase II, the Hotelling T^2 chart signals for the first time at 13/10/2013 05:52:00 which is a bit later than the vine copula-based chart.
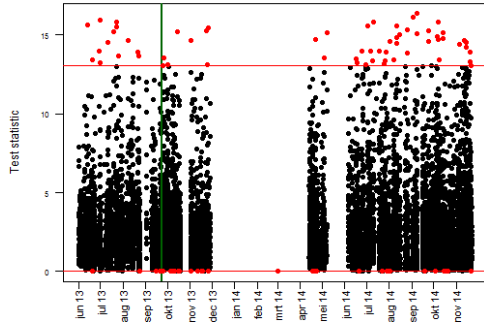
(a) Vine copula-based control chart.
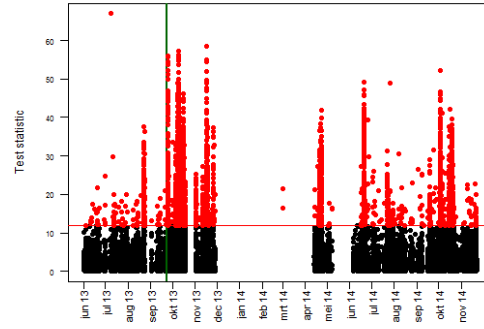
(b) Hotelling T^2 control chart.

Figure 6.6: Multivariate control charts for model 2 when using a phase I period to calculate the regression model, and using 4-minute intervals.

4-hour intervals

The results for the subsample based on 4-hour intervals are visualized in Figure 6.7. From this figure we conclude that the results are similar to the above mentioned findings. For the vine copula-based chart, all of the 112 observations in phase I are within the control limits. Furthermore, 5 out of 388 observations are outside the control limits in phase II, which corresponds to approximately 1.29%. In total, this leads to 5 out of 500 observations outside the control limits, or 1.00%. This is slightly more than expected from a well-behaved process, and this could be an indication that things are not working properly. The first signal during phase II is at 13/10/2013 12:00:00. For the Hotelling T^2 chart, 3 out of 112 observations, or approximately 2.68%, and 34 out of 388 observations, or approximately 8.76%, are outside the limit in phase I and phase II, respectively. This corresponds to a total of 37 out of 500 observations, or 7.40%, marked as out-of-control. This is much more than expected from a well-behaved process, and this is definitely an indication that things are not working properly. The first signal during phase II is at 13/10/2013 08:00:00 which is one observation earlier than the vine copula-based chart. The false alarm rate, however, is still higher than desired in phase I.
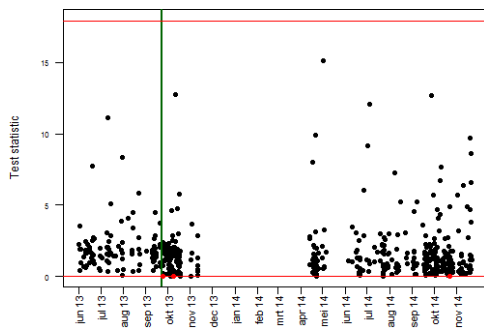
(a) Vine copula-based control chart.
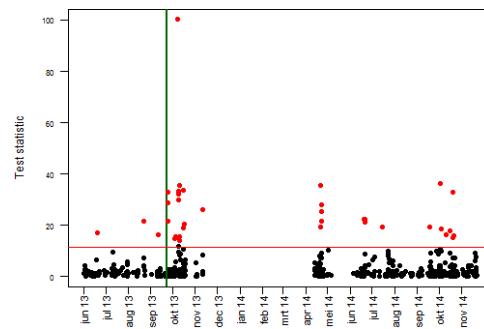
(b) Hotelling T^2 control chart.

Figure 6.7: Multivariate control charts for model 2 when using a phase I period to calculate the regression model, and using 4-hour intervals.

Overall, we conclude that the vine copula-based chart performs very good in phase I given the low false alarm rates. For the first model, it is questionable whether this chart is powerful enough to detect the upcoming failure as the amount of signals is not significant. For the 4-minute intervals, the time of the first signal in phase II is acceptable and all signals must be seriously taken into consideration. The Hotelling T^2 chart is powerful enough to detect the upcoming failure since the

amount of signals is much more than expected from a well-behaved process. The time of the first signal in phase II is about 9 hours earlier than the vine copula-based chart. However, the false alarm rate for the Hotelling T^2 chart is unacceptable, which questions the true meaning behind the signals. It is known that autocorrelation might cause many false alarms for the Hotelling T^2 chart. By using the 4-hour interval subsample, we reduce the autocorrelation and consequently the relative amount of signals, however, the false alarm rate remains significant. For the second model, the vine copula-based chart is powerful enough to detect the upcoming failure as the amount of signals is significant. For the 4-minute intervals, the chart signals instantly and for the 4-hour intervals, the chart signals almost instantly. The Hotelling T^2 chart is also powerful enough to detect the upcoming failure since the amount of signals is much more than expected from a well-behaved process. For the 4-minute intervals, the time of the first signal in phase II is about 7 hours later than the vine copula-based chart. For the 4-hour intervals, the time of the first signal in phase II is 4 hours earlier than the vine copula-based chart. However, the false alarm rate for the Hotelling T^2 chart is, once again, much higher than desired for both the 4-minute interval data and 4-hour interval subsample. When comparing these results with the findings in van Dalen (2018), we observe that the Hotelling T^2 chart performs quite similar as the univariate chart. Both charts are able to quickly detect indications that the process is out-of-control. The Hotelling T^2 chart often signals a bit earlier than the univariate chart. However, the false alarm rates of both charts are unacceptable. The vine copula-based chart offers a potential solution as it achieves an acceptable false alarm rate and is able to quickly detect the upcoming failure based on model 2.

6.4 Summary

- The vine copula-based chart seems unable to detect the upcoming failure based on the model that monitors the nacelle temperature and oil temperature simultaneously.
- The Hotelling T^2 chart is able to detect the upcoming failure based on the model that monitors the nacelle temperature and oil temperature simultaneously, however, the false alarm rate is undesirable.
- The vine copula-based chart is powerful enough to detect the upcoming failure based on the model that monitors the primary generator temperature and secondary generator temperature simultaneously, while still achieving an acceptable false alarm rate.
- The Hotelling T^2 chart is able to detect the upcoming failure based on the model that monitors the primary generator temperature and secondary generator temperature simultaneously, however, the false alarm rate is much higher than desired.
- The vine copula-based chart seems less prone to highly autocorrelated observations than the Hotelling T^2 chart.

7 | Conclusions

7.1 Summary

The main goal of this thesis was to develop an effective multivariate monitoring method, as conventional multivariate monitoring methods rely on assumptions which are rarely satisfied for process data. In order to achieve this goal, we introduced a monitoring approach based on the modern theory of copulas and investigated the performance of this method compared to conventional multivariate monitoring methods in an extensive simulation study. The results showed that the copula-based approach has great potential in becoming a well-established monitoring method, as the copula-based approach significantly outperforms conventional methods in a variety of controlled out-of-control cases. Additionally, we deviated from the controlled environment and analyzed the performance of the copula-based approach and the Hotelling T^2 chart in a case study. In this case study, several quality characteristics of a wind turbine were modeled by multivariate multiple linear regression models after which both monitoring methods were applied to monitor the deviation of the models. It immediately became clear that the copula-based approach provided a low false alarm rate, while it was also able to quickly detect indications of the upcoming failure based on one of the regression models.

7.2 Conclusions

In this work we were interested in answering the research questions which are given below. For each question the main findings are stated.

How does the multivariate regression model relate to the univariate regression model and can we derive similar properties of estimators?

The univariate regression model generalizes to the case of multiple responses for each observation which leads to the multivariate regression model. The estimation results for the univariate model generalize easily to the multivariate case, although we have to be careful with notation style and deriving results as we are dealing with matrices instead of vectors. Furthermore, in order to estimate the covariance matrix we need knowledge of the Wishart distribution, which is basically a generalization of the chi-square distribution.

How can we design a multivariate control chart based on a given vine copula and what are its main benefits?

The design of the vine copula-based control chart follows the ideas of Verdier (2013) by constructing a tolerance region based on density level set estimation. The control limits are determined by a Shewhart-like approach. The need for vine copula structures is crucial to obtain a flexible model, especially for high dimensions. Furthermore, this vine copula-based control chart is distribution free and therefore extremely useful for non-normal multivariate data. Additionally, this approach seems to be a potential solution for autocorrelated data as well.

How can we fairly compare the performance of control charts to various changes in a process?

The average run length is a commonly used measure to assess the performance of control charts. However, we have to be cautious since the comparison cannot be based solely on the average run length value, as run length distributions are often very skewed. To gain additional insights in the run length distribution one should additionally compare the kernel density plots of the run length. In order to have a fair comparison we should first make sure that the average in-control run

lengths are approximately equal by calibrating the parameters of the charts. Only when the kernel density plots do not show significant differences in distribution, we may conclude that the average in-control run lengths are equal. Subsequently, we manually implement an intended change in the process after which we are able to compare the average out-of-control run lengths. Now, we are only able to conclude a difference in performance when the kernel density plots do show significant differences in distribution.

How does the vine copula-based control chart perform compared to conventional multivariate monitoring methods?

First of all, the vine copula-based chart outperforms the conventional methods in the majority of the implemented out-of-control cases. There are, however, specific out-of-control cases in which the conventional methods perform significantly better. In case one is only interested in detecting arbitrary shifts in the mean of a multivariate process, we could say that the MCUSUM chart with $k = 0.25$ seems to perform better than the vine copula-based chart in general. Furthermore, in case one is interested in detecting arbitrary shifts in the variance or dependence structure of a multivariate process, the vine copula-based chart seems to be the obvious choice. Overall, we therefore conclude that the vine copula-based chart performs better than the conventional methods in general for detecting arbitrary shifts in a multivariate process, assuming that the shift could be of any type.

7.3   Future research

The application of copula theory within statistical process control is still in a very early stage and therefore there is considerable room for improvement. In order to further improve this statistical framework, we state the most important recommendations for future research.

In the simulation study we investigated the performance of multivariate control charts by means of the average run length. Additionally, we provided some kernel density plots to give substance to the statements about the in-control run length. It is also necessary to include kernel density plots of the out-of-control run length to make substantial statements about performance in out-of-control cases. On the other hand, we could have looked at confidence intervals for the average run length in order to make solid statistical-based conclusions about difference in performance. To investigate confidence intervals for a geometric distribution, one could look into Byrne (2005).

Furthermore, in the simulation study we assessed the performance of control charts subject to a pre-defined type of change which could be a shift in the mean, variance or dependence structure of a multivariate process. In practice, most shifts are a combination of the three types mentioned above. Besides that, we only investigated sudden shifts which occur, for instance, when a piece of equipment breaks. However, it is also quite possible that a shift gradually develops over a period of time, for instance, when a piece of equipment wears over time. Therefore, it would be of great interest to also investigate more complex types of shifts.

Then, we analyzed the performance of the vine copula-based chart compared to conventional charts by simulating an in-control process directly from a copula distribution and subsequently implement controlled shifts in the process. However, the simulated in-control process is very often not multivariate normally distributed. This could question the base of the comparison as conventional charts heavily rely on the multivariate normality assumption. Therefore, we recommend to continue the simulation study by simulating an in-control process from a multivariate normal distribution, the home of conventional charts, and implement intended shifts in the process. Then, one should apply a conventional chart, like Hotelling T^2, directly to the observed process. On the other hand, one should translate the observed process to the unit hypercube by means of an appropriate copula model and subsequently apply the vine copula-based chart to the translated process. When the vine copula-based chart outperforms the conventional chart in some specific out-of-control cases, this will even further improve the position of the vine copula-based chart with respect to conventional methods.

Next, multivariate monitoring methods seem to have several advantages over univariate methods as they take into account dependencies between quality characteristics. However, when a multivariate chart signals, it is unclear where the signal originated from. In order to find the quality characteristic(s) that causes the signal, one could apply a univariate chart or the leave-one-out

method. We suggest to consult Mason et al. (2011) to obtain more information about this signal background detection.

Finally, from the simulation study it becomes clear that the in-control run length of the Hotelling T^2 chart is massively affected by the input parameter value. Also, the in-control run length of the MCUSUM chart seems to be slightly affected by the input parameter. This could indicate that the conventional methods are significantly affected by estimation error of, for instance, the mean vector and variance-covariance matrix. The in-control run length of the vine copula-based chart, on the other hand, seems to be roughly constant for any input parameter. This could indicate that the vine-copula based chart is less affected by estimation error which would be of great practical significance. Therefore, we highly recommend to further investigate the effects of estimation error on the run length for both in-control and out-of-control cases.

Bibliography

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

A. Baíllo and A. Cuevas. Parametric versus nonparametric tolerance regions in detection problems. *Computational Statistics*, 21(3-4):523–536, 2006.

T. Bedford and R.M. Cooke. Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068, 2002.

E. Brechmann and U. Schepsmeier. Cdvine: Modeling dependence with C-and D-vine copulas in R. *Journal of Statistical Software*, 52(3):1–27, 2013.

R.L. Brown, J. Durbin, and J.M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192, 1975.

J. Byrne. A new short exact geometric confidence interval. *Australian & New Zealand Journal of Statistics*, 47(4):563–569, 2005.

B. Cadre, B. Pelletier, and P. Pudlo. Estimation of density level sets with a given probability content. *Journal of Nonparametric Statistics*, 25(1):261–272, 2013.

B. Choroś, R. Ibragimov, and E. Permiakova. Copula estimation. In P. Jaworski, F. Durante, W.K. Härdle, and T. Rychlik, editors, *Copula Theory and Its Applications*, pages 77–91. Springer, 2010.

C.S.J. Chu, M. Stinchcombe, and H. White. Monitoring structural change. *Econometrica*, 64(5):1045–1065, 1996.

R.B. Crosier. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3):291–303, 1988.

C. Czado. Pair-copula constructions of multivariate copulas. In P. Jaworski, F. Durante, W.K. Härdle, and T. Rychlik, editors, *Copula Theory and Its Applications*, pages 93–109. Springer, 2010.

J. Dissmann, E.C. Brechmann, C. Czado, and D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69, 2013.

K. W. Fang. *Symmetric Multivariate and Related Distributions*. CRC Press, 2018.

M. Frisén. Statistical surveillance. Optimality and methods. *International Statistical Review*, 71(2):403–434, 2003.

M. Frisén. Properties and use of the Shewhart method and its followers. *Sequential Analysis*, 26(2):171–193, 2007.

R. Galeeva, J. Hoogland, A. Eydeland, and C.M. Stanley. Measuring correlation risk for energy derivatives. *Risk Management in Commodity Markets: From Shipping to Agriculturals and Energy*, pages 81–89, 05 2012. doi: 10.1002/9781118467381.ch6.

C. Genest, J.Q. Molina, J.R. Lallena, and C. Sempi. A characterization of quasi-copulas. *Journal of Multivariate Analysis*, 69(2):193–205, 1999.

D.M. Hawkins. Regression adjustment for variables in multivariate quality control. *J. Qual. Technology*, 25:170–182, 1993.

M. Hofert, I. Kojadinovic, M. Mächler, and J. Yan. *Elements of Copula Modeling with R*. Springer Use R! Series, 2018. ISBN 978-3-319-89635-9. URL `http://www.springer.com/de/book/9783319896342`.

M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula: Multivariate Dependence with Copulas*, 2020. URL `https://CRAN.R-project.org/package=copula`. R package version 1.0-0.

H. Hotelling. Multivariate quality control. *Techniques of Statistical Analysis*, 1947.

H. Joe. Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141, 1996. ISSN 07492170. URL `http://www.jstor.org/stable/4355888`.

H. Joe. *Dependence Modeling with Copulas*. CRC press, 2014.

T. Kenbeek. Condition based prognostics and diagnostics for wind turbines. Bachelor's thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, 2016.

T. Kenbeek, S. Kapodistria, and A. Di Bucchianico. Data-driven online monitoring of wind turbines. *arXiv preprint arXiv:1702.05047*, 2016.

R. Kenett and M. Pollak. On assessing the performance of sequential procedures for detecting a change. *Qual. Reliab. Eng. Int.*, 28(5):500–507, 2012.

D. Kurowicka and R.M. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons, 2006.

R.L. Mason, Y.M. Chou, and J.C. Young. Detection and interpretation of a multivariate signal using combined charts. *Communications in Statistics—Theory and Methods*, 40(5):942–957, 2011.

L. Meeuwis. Monitoring wind turbines with regression analysis. Bachelor's thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, 2017.

J.F. Monahan. *A Primer on Linear Models*. CRC Press, 2008.

B. Mühlig. Multivariate process monitoring based on copula structures. Master's thesis, Julius-Maximilians-Universität Würzburg, Fakultät für Mathematik und Informatik, Lehrstuhl für Statistik, 2017.

T. Nagler and T. Vatter. *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*, 2020. URL `https://cran.r-project.org/package=rvinecopulib`. R package version 0.5.3.1.0.

R.B. Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2007.

P. Qiu. *Introduction to Statistical Process Control*. CRC press, 2013.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org`.

L. Rüschendorf. On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.

E. Santos-Fernández. *Multivariate Statistical Quality Control Using R*, volume 14. Springer, 2013. ISBN 9781461454533. URL `http://www.springer.com/statistics/computational+statistics/book/978-1-4614-5452-6`.

U. Schepsmeier, J. Stoeber, E.C. Brechmann, B. Graeler, T. Nagler, T. Erhardt, C. Almeida, A. Min, C. Czado, M. Hofmann, M. Killiches, H. Joe, and T. Vatter. *VineCopula: Statistical Inference of Vine Copulas*, 2019. URL `https://cran.r-project.org/package=VineCopula`. R package version 2.3.0.

J. Shao. *Mathematical Statistics*. Springer Science & Business Media, 2003.

W.A. Shewhart. The application of statistics as an aid in maintaining quality of a manufactured product. *Journal of the American Statistical Association*, 20(152):546–548, 1925.

S. Sukparungsee, S. Kuvattana, P. Busababodhin, and Y. Areepong. Multivariate copulas on the MCUSUM control chart. *Cogent Mathematics*, 4(1):1342318, 2017.

N.D. Tracy, J.C. Young, and R.L. Mason. Multivariate control charts for individual observations. *Journal of Quality Technology*, 24(2):88–95, 1992.

O. van Dalen. Statistical monitoring of wind turbines. Bachelor's thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, 2018.

G. Verdier. Application of copulas to multivariate control charts. *Journal of Statistical Planning and Inference*, 143(12):2151–2159, 2013.

A. Zeileis, F. Leisch, K. Hornik, and C. Kleiber. strucchange. An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1–38, 2001.

A | Lemmas

Lemma A.1. *If a random variable X has a continuous distribution for which the CDF is F_X, then the random variable Y defined as $Y = F_X(X)$ has a uniform distribution on the interval $[0,1]$.*

Proof. Given any continuous random variable X, define $Y = F_X(X)$. Then, we have that

$$
\begin{aligned}
F_Y(y) &= \Pr[Y \le y] \\
&= \Pr[F_X(X) \le y] \\
&= \Pr[X \le F_X^{-1}(y)] \\
&= F_X(F_X^{-1}(y)) \\
&= y
\end{aligned}
$$

$\square$

Lemma A.2. *The matrix M_{shift} as defined in (5.4) is a correlation matrix.*

Proof. Symmetry easily follows from (5.3) and (5.4):

$$
\begin{aligned}
M_{\text{shift}}^T &= \left(V^{-1} M^* \left(V^{-1} \right)^T \right)^T \\
&= V^{-1} \left(U D^* U^T \right)^T \left(V^{-1} \right)^T \\
&= V^{-1} M^* \left(V^{-1} \right)^T \\
&= M_{\text{shift}},
\end{aligned}
$$

since D^* is a diagonal matrix and hence symmetric.

Furthermore, by construction, we have that

$$
\begin{aligned}
M_{\text{shift}} &= V^{-1} M^* V^{-1} \\
&= \begin{pmatrix} \frac{1}{\sqrt{\tau_{11}^*}} & 0 \\ 0 & \frac{1}{\sqrt{\tau_{22}^*}} \end{pmatrix} \begin{pmatrix} \tau_{11}^* & \tau_{12}^* \\ \tau_{21}^* & \tau_{22}^* \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\tau_{11}^*}} & 0 \\ 0 & \frac{1}{\sqrt{\tau_{22}^*}} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{\tau_{12}^*}{\sqrt{\tau_{11}^* \tau_{22}^*}} \\ \frac{\tau_{21}^*}{\sqrt{\tau_{11}^* \tau_{22}^*}} & 1 \end{pmatrix},
\end{aligned}
$$

which ensures that the diagonal elements are equal to one. Also, if we define $U = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$, we obtain by (5.2) and (5.3) that

$$
\begin{aligned}
M^* &= U D^* U^T \\
&= \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \begin{pmatrix} \lambda_1^* & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1^* u_{11}^2 + \lambda_2 u_{12}^2 & \lambda_1^* u_{11} u_{21} + \lambda_2 u_{12} u_{22} \\ \lambda_1^* u_{11} u_{21} + \lambda_2 u_{12} u_{22} & \lambda_1^* u_{21}^2 + \lambda_2 u_{22}^2 \end{pmatrix},
\end{aligned}
$$

so that

$$\tau_{12}^* = \tau_{21}^* = \lambda_1^* u_{11} u_{21} + \lambda_2 u_{12} u_{22}$$

$$\begin{aligned}
\tau_{11}^* \tau_{22}^* &= \left(\lambda_1^* u_{11}^2 + \lambda_2 u_{12}^2\right)\left(\lambda_1^* u_{21}^2 + \lambda_2 u_{22}^2\right) \\
&= (\lambda_1^*)^2 u_{11}^2 u_{21}^2 + \lambda_2^2 u_{12}^2 u_{22}^2 + \lambda_1^* \lambda_2 \left(u_{11}^2 u_{22}^2 + u_{12}^2 u_{21}^2\right).
\end{aligned}$$

To show that all elements of M_{shift} lie in the interval $[-1, 1]$, we only have to show that

$$\left| \frac{\tau_{12}^*}{\sqrt{\tau_{11}^* \tau_{22}^*}} \right| \leq 1, \text{ or equivalently, } |\tau_{12}^*| \leq \left| \sqrt{\tau_{11}^* \tau_{22}^*} \right|, \tag{A.1}$$

as we have already shown that M_{shift} is symmetric. We are allowed to square both sides of (A.1) since both sides are positive. By Lemma A.3, we can deduce that

$$\begin{aligned}
|\tau_{12}^*|^2 &= (\lambda_1^* u_{11} u_{21} + \lambda_2 u_{12} u_{22})^2 \\
&= (\lambda_1^*)^2 u_{11}^2 u_{21}^2 + \lambda_2^2 u_{12}^2 u_{22}^2 + 2\lambda_1^* \lambda_2 u_{11} u_{22} u_{12} u_{21} \\
&\leq (\lambda_1^*)^2 u_{11}^2 u_{21}^2 + \lambda_2^2 u_{12}^2 u_{22}^2 + \lambda_1^* \lambda_2 \left(u_{11}^2 u_{22}^2 + u_{12}^2 u_{21}^2\right) \\
&= \left| \sqrt{\tau_{11}^* \tau_{22}^*} \right|^2,
\end{aligned}$$

which implies that all entries lie in the interval $[-1, 1]$.

Finally, we have to show that the matrix M_{shift} is positive semidefinite. We calculate its eigenvalues and show that they are non-negative:

$$\begin{aligned}
\left| M_{\text{shift}} - \lambda I \right| &= \begin{vmatrix} 1 - \lambda & \frac{\tau_{12}^*}{\sqrt{\tau_{11}^* \tau_{22}^*}} \\ \frac{\tau_{21}^*}{\sqrt{\tau_{11}^* \tau_{22}^*}} & 1 - \lambda \end{vmatrix} \\
&= (1 - \lambda)^2 - \frac{(\tau_{12}^*)^2}{\tau_{11}^* \tau_{22}^*} \qquad \Rightarrow \lambda = 1 \pm \frac{\tau_{12}^*}{\sqrt{\tau_{11}^* \tau_{22}^*}}
\end{aligned}$$

Since we have shown that (A.1) holds, we know that $\lambda \geq 0$ and by Lemma A.4 we conclude that M_{shift} is positive semidefinite. □

Lemma A.3. *For all $a, b \in \mathbb{R}$ it holds that $a^2 + b^2 \geq 2ab$.*

Proof. Let $a, b \in \mathbb{R}$. Then, we simply have that

$$0 \leq (a - b)^2 = a^2 + b^2 - 2ab$$

□

Lemma A.4. *A symmetric matrix is positive semidefinite if and only if all eigenvalues are non-negative.*

Proof. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Furthermore, let $x \in \mathbb{R}^n$ be an arbitrary vector. Then, by the eigendecomposition of A, we have

$$x^T A x = x^T \left(U D U^T\right) x = \left(x^T U\right) D \left(x^T U\right)^T = \sum_{i=1}^n \lambda_i \left([x^T U]_i\right)^2, \tag{A.2}$$

where U is the orthogonal matrix whose i-th column is the eigenvector u_i of A, and D is the diagonal matrix whose diagonal elements D_{ii} are the corresponding eigenvalues λ_i. The expression (A.2) is non-negative for all $x \in \mathbb{R}^n$ if and only if $\lambda_i \geq 0$ for all $i = 1, \ldots, n$. □

Lemma A.5. *If the matrix $\widehat{B}$ solves the multivariate normal equations (2.10), then $\widehat{B}$ minimizes $Q(B)$ as defined in (2.8).*

Proof. See the proof of Result 9.1 of Monahan (2008, Chapter 9), where $Y := Y_{[1:n,1:q]}$ and $X := X_n$. $\qquad\qquad\square$

Lemma A.6. *The covariance matrix estimator*

$$\widehat{\Sigma} = \frac{1}{n - rank(X_n)} F = \frac{1}{n - rank(X_n)} \left(Y_{[1:n,1:q]} - X_n\widehat{B}\right)^T \left(Y_{[1:n,1:q]} - X_n\widehat{B}\right)$$

is an unbiased estimator of Σ.

Proof. See the proof of Result 9.2 of Monahan (2008, Chapter 9), where $Y := Y_{[1:n,1:q]}$ and $X := X_n$. $\qquad\qquad\square$

Lemma A.7. *The scalar quantity $\lambda^T Bm$ is estimable if and only if $\lambda \in \mathcal{C}\left(X_n^T\right)$.*

Proof. First, assume that $\lambda^T Bm$ is estimable. Then, by definitions 2.1 and 2.2, there exists an $a \in \mathbb{R}^n$ such that

$$\mathbb{E}\left[a^T Y_{[1:n,1:q]} m\right] = \lambda^T Bm \quad \text{for all } B, m.$$

It also holds that

$$\mathbb{E}\left[a^T Y_{[1:n,1:q]} m\right] = a^T X_n Bm \quad \text{for all } B, m.$$

Hence, $\left(\lambda^T - a^T X_n\right) Bm = 0$ for all B, m and therefore $\lambda = X_n^T a$. So, $\lambda \in \mathcal{C}\left(X_n^T\right)$.

Second, assume that $\lambda \in \mathcal{C}\left(X_n^T\right)$. Then for some a,

$$\lambda^T Bm = a^T X_n Bm = a^T \mathbb{E}\left[Y_{[1:n,1:q]} m\right] = \mathbb{E}\left[a^T Y_{[1:n,1:q]} m\right].$$

So $a^T Y_{[1:n,1:q]} m$ is an unbiased estimator of $\lambda^T Bm$, and therefore $\lambda^T Bm$ is estimable. $\qquad\square$

Lemma A.8. *If $\lambda^T Bm$ is estimable, then $\lambda^T \widehat{B}m$ is constant for all solutions $\widehat{B}$ of the multivariate normal equations (2.10).*

Proof. See the proof of Result 9.4 of Monahan (2008, Chapter 9), where $X := X_n$. $\qquad\square$

Lemma A.9. *If $\widehat{B}$ solves the multivariate normal equations (2.10), then the best linear unbiased estimator (BLUE) of estimable $\lambda^T Bm$ is $\lambda^T \widehat{B}m$.*

Proof. Assume that $\widehat{B}$ is a solution of the multivariate normal equations (2.10). To prove that $\lambda^T \widehat{B}m$ is the BLUE of $\lambda^T Bm$, we have to show that

$$\mathrm{Var}\left(\lambda^T \widehat{B}m\right) \leq \mathrm{Var}\left(\lambda^T \widetilde{B}m\right), \tag{A.3}$$

for any constant vectors λ, m, and any other linear unbiased estimator $\widetilde{B}$.

We can write condition (A.3) as

$$\lambda^T \left(\mathrm{Var}\left(\widetilde{B}m\right) - \mathrm{Var}\left(\widehat{B}m\right)\right) \lambda \geq 0,$$

and therefore, condition (A.3) is satisfied if and only if

$$\mathrm{Var}\left(\widetilde{B}m\right) - \mathrm{Var}\left(\widehat{B}m\right) \tag{A.4}$$

is a positive semidefinite matrix. Before we prove that (A.4) is positive semidefinite, we need to derive some properties. Similar to the derivations in (2.16) and (2.17), we can show that

$$\mathbb{E}\left[\widehat{B}m\right] = Bm$$

$$\mathrm{Var}\left(\widehat{B}m\right) = \left(m^T \Sigma m\right) \left(X_n^T X_n\right)^{-1}$$

Now, let $\widetilde{B}m = CY_{[1:n,1:q]}m$ be another linear estimator of Bm with $C = D + \left(X_n^T X_n\right)^{-1} X_n^T$, where D is a $(p+1) \times n$ non-zero matrix. Then, we can write

$$
\begin{aligned}
\widetilde{B}m &= CY_{[1:n,1:q]}m \\
&= DY_{[1:n,1:q]}m + \left(X_n^T X_n\right)^{-1} X_n^T Y_{[1:n,1:q]}m \\
&= DY_{[1:n,1:q]}m + \widehat{B}m
\end{aligned}
$$

Subsequently, we calculate

$$
\begin{aligned}
\mathbb{E}\left[\widetilde{B}m\right] &= D\mathbb{E}\left[Y_{[1:n,1:q]}m\right] + \mathbb{E}\left[\widehat{B}m\right] \\
&= DX_n Bm + Bm \\
&= \left(DX_n + I_{p+1}\right) Bm
\end{aligned}
$$

Hence, $\widetilde{B}m$ is unbiased if and only if $DX_n = 0$. Then, by using (2.15) and the fact that $DX_n = 0$, we have

$$
\begin{aligned}
\mathrm{Var}\left(\widetilde{B}m\right) &= \mathrm{Var}\left(CY_{[1:n,1:q]}m\right) \\
&= C\,\mathrm{Var}\left(Y_{[1:n,1:q]}m\right) C^T \\
&= \left(m^T \Sigma m\right) CC^T \\
&= \left(m^T \Sigma m\right) DD^T + \left(m^T \Sigma m\right) DX_n \left(X_n^T X_n\right)^{-1} + \left(m^T \Sigma m\right) \left(X_n^T X_n\right)^{-1} X_n^T D^T \\
&\quad + \left(m^T \Sigma m\right) \left(X_n^T X_n\right)^{-1} X_n^T X_n \left(X_n^T X_n\right)^{-1} \\
&= \left(m^T \Sigma m\right) DD^T + \left(m^T \Sigma m\right) \left(X_n^T X_n\right)^{-1} \\
&= \left(m^T \Sigma m\right) DD^T + \mathrm{Var}\left(\widehat{B}m\right)
\end{aligned}
$$

As a consequence,

$$
\mathrm{Var}\left(\widetilde{B}m\right) - \mathrm{Var}\left(\widehat{B}m\right) = \left(m^T \Sigma m\right) DD^T \tag{A.5}
$$

is positive semidefinite by the property that the covariance matrix Σ is positive semidefinite and by Lemma A.10. This is true for any linear unbiased estimator $\widetilde{B}m$ and, therefore, $\lambda^T \widehat{B}m$ is the BLUE of $\lambda^T Bm$. $\qquad\square$

Lemma A.10. *For any real matrix $Z \in \mathbb{R}^{n \times k}$, the product ZZ^T is positive semidefinite.*

Proof. Let $Z \in \mathbb{R}^{n \times k}$ be a real matrix and let $x \in \mathbb{R}^n$ be an arbitrary vector. Then,

$$
x^T ZZ^T x = \left(Z^T x\right)^T \left(Z^T x\right) = ||Z^T x||_2^2 \geq 0
$$

by definition of the norm. $\qquad\square$

Lemma A.11. *Let $f(A) = |A|^{-k/2}\exp\left\{-\frac{1}{2}\mathrm{Tr}\left[A^{-1}F\right]\right\}$, where F is a symmetric, positive definite matrix of order p. Then $f(A)$ is maximized over positive definite matrices at $A = \frac{1}{k}F$ where $f\left(\frac{1}{k}F\right) = |\frac{1}{k}F|^{-k/2}e^{-kp/2}$.*

Proof. See the proof of Lemma 9.1 of Monahan (2008, Chapter 9). $\qquad\square$

Lemma A.12. *If $\lambda^T Bm$ is estimable, then $\lambda^T \widehat{B}m$ is the minimum-variance unbiased estimator (MVUE) of $\lambda^T Bm$.*

Proof. Assume that $Y_{i\bullet}$ independent and

$$
Y_{i\bullet}^T \sim \mathcal{N}_q\left(B^T \left(X_\infty^{(i)}\right)^T, \Sigma\right), \quad i = 1, \ldots, n.
$$

We are able to factorize the joint density as

$$f\left(y_{1\bullet}^T,\ldots,y_{n\bullet}^T|B,\Sigma\right) = \prod_{i=1}^{n}\left((2\pi)^q|\Sigma|\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(y_{i\bullet}-X_nB\right)\Sigma^{-1}\left(y_{i\bullet}-X_nB\right)^T\right\}$$

$$= \left((2\pi)^q|\Sigma|\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(y_{i\bullet}-X_nB\right)\Sigma^{-1}\left(y_{i\bullet}-X_nB\right)^T\right\}$$

$$= \left((2\pi)^q|\Sigma|\right)^{-\frac{n}{2}}$$
$$\cdot\exp\left\{-\frac{1}{2}\operatorname{Tr}\left[\Sigma^{-1}\left(y_{[1:n,1:q]}-X_n\widehat{B}\right)^T\left(y_{[1:n,1:q]}-X_n\widehat{B}\right)\right]\right.$$
$$\left.-\frac{1}{2}\operatorname{Tr}\left[\Sigma^{-1}\left(\widehat{B}-B\right)^T X_n^T X_n\left(\widehat{B}-B\right)\right]\right\}$$

$$= \left((2\pi)^q|\Sigma|\right)^{-\frac{n}{2}}$$
$$\cdot\exp\left\{-\frac{1}{2}\operatorname{Tr}\left[\Sigma^{-1}\left(y_{[1:n,1:q]}^T y_{[1:n,1:q]} - y_{[1:n,1:q]}^T X_n\widehat{B} - \widehat{B}^T X_n^T y_{[1:n,1:q]}\right.\right.\right.$$
$$\left.\left.\left.+\widehat{B}^T X_n^T X_n\widehat{B}\right)\right]-\frac{1}{2}\operatorname{Tr}\left[\Sigma^{-1}\left(\widehat{B}-B\right)^T X_n^T X_n\left(\widehat{B}-B\right)\right]\right\}$$

$$=: g\left(y_{[1:n,1:q]}^T y_{[1:n,1:q]}, X_n^T y_{[1:n,1:q]}|B,\Sigma\right)h\left(y_{1\bullet}^T,\ldots,y_{n\bullet}^T\right),$$

where $h(\cdot)=1$.

By Fisher-Neyman's factorization theorem and the fact that $X_n\widehat{B} = X_n\left(X_n^T X_n\right)^{-1}X_n^T Y_{[1:n,1:q]}$ is a function of $X_n^T Y_{[1:n,1:q]}$, we know that $\left(Y_{[1:n,1:q]}^T Y_{[1:n,1:q]}, X_n^T Y_{[1:n,1:q]}\right)$ is sufficient for (B,Σ). Furthermore, by Lemma A.13, the statistic $\left(Y_{[1:n,1:q]}^T Y_{[1:n,1:q]}, X_n^T Y_{[1:n,1:q]}\right)$ is also complete for (B,Σ). Note that $\lambda^T\widehat{B}m$ is a function of $X_n^T Y_{[1:n,1:q]}$ and, hence, a function of the complete sufficient statistic. If $\lambda^T Bm$ is estimable, then $\lambda^T\widehat{B}m$ is an unbiased estimator for $\lambda^T Bm$ (see (2.16)). By the Lehmann-Scheffé theorem, we now conclude that $\lambda^T\widehat{B}m$ is the MVUE of $\lambda^T Bm$. $\square$

Lemma A.13. *If P is in an exponential family of full rank with p.d.f.'s given by*

$$f_\eta(x) = \exp\left\{\eta^T T(x) - \zeta(\eta)\right\}h(x),$$

then $T(X)$ is complete for $\eta \in \Xi$.

Proof. See the proof of Proposition 2.1 of Shao (2003). $\square$

Lemma A.14. *If $W \sim W_q(m,\Sigma)$, then for any $q\times 1$ vector a,*

$$\frac{a^T W a}{a^T \Sigma a} \sim \chi_m^2.$$

Proof. See the proof of Result 9.8 of Monahan (2008, Chapter 9). $\square$

Lemma A.15. *If $W \sim W_q(m,\Sigma)$, and A is a $p\times q$ matrix, then*

$$AWA^T \sim W_p\left(m, A\Sigma A^T\right).$$

Proof. See the proof of Result 9.10 of Monahan (2008, Chapter 9). $\square$

Lemma A.16. *(Generalization of Cochran's Theorem) Let E follow the error distribution in the MMLR model (2.18), that is, $E_{i\bullet}^T \overset{iid}{\sim} \mathcal{N}_q(\mathbf{0},\Sigma)$, $i=1,\ldots,n$, where Σ is non-singular. Also let A_i, $i=1,\ldots,k$, be symmetric, idempotent matrices with ranks s_i, respectively. If $\sum_{i=1}^{k}A_i = I_n$, then $E^T A_i E$ are independently distributed as $W_q(s_i,\Sigma)$, and $\sum_{i=1}^{k}s_i = n$.*

Proof. See the proof of Theorem 9.1 of Monahan (2008, Chapter 9). $\square$

Lemma A.17. *If P is the matrix defined as $P = X_n \left(X_n^T X_n \right)^{-1} X_n^T$, then P and $I_n - P$ are symmetric and idempotent.*

Proof. See the proof of Lemma A.6 of van Dalen (2018). $\qquad\square$

Lemma A.18. *Let $Y_{[1:n,1:q]}$ be as defined in (2.3). If $m = \begin{pmatrix} m_1 & \dots & m_n \end{pmatrix}^T$ is an $n \times 1$ constant vector, then*

$$\mathrm{Var}\left(Y_{[1:n,1:q]}^T m \right) = \Sigma m^T m.$$

Proof. Consider the model as defined in (2.3) and let $m = \begin{pmatrix} m_1 & \dots & m_n \end{pmatrix}^T$ be an $n \times 1$ constant vector. Then, we deduce by using (2.7) that

$$
\begin{aligned}
\mathrm{Var}\left(Y_{[1:n,1:q]}^T m \right) &= \mathrm{Var} \begin{pmatrix} m^T Y_{\bullet 1} \\ \vdots \\ m^T Y_{\bullet q} \end{pmatrix} \\
&= \begin{pmatrix} m^T \Sigma_{11} m & m^T \Sigma_{12} m & \dots & m^T \Sigma_{1q} m \\ m^T \Sigma_{21} m & m^T \Sigma_{22} m & \dots & m^T \Sigma_{2q} m \\ \vdots & \vdots & \ddots & \vdots \\ m^T \Sigma_{q1} m & m^T \Sigma_{q2} m & \dots & m^T \Sigma_{qq} m \end{pmatrix} \\
&= \Sigma m^T m,
\end{aligned}
$$

since $m^T m \in \mathbb{R}$. $\qquad\square$

Lemma A.19. *Let (X_1, X_2) be a bivariate random vector with continuous marginals and copula C. Then, it holds that*

$$\tau(C) = 1 - 4 \int_{[0,1]^2} \frac{\partial}{\partial u_1} C(u_1, u_2) \frac{\partial}{\partial u_2} C(u_1, u_2) du_1 du_2,$$

and

$$\rho_s(C) = 12 \int_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3.$$

Proof. See the proof of Theorem 2.3.1 and Theorem 2.3.2 of Mühlig (2017). $\qquad\square$

B | Additional results

B.1 One-sided vs. two-sided version

Clayton (case 2 & cases $16 - 25$)
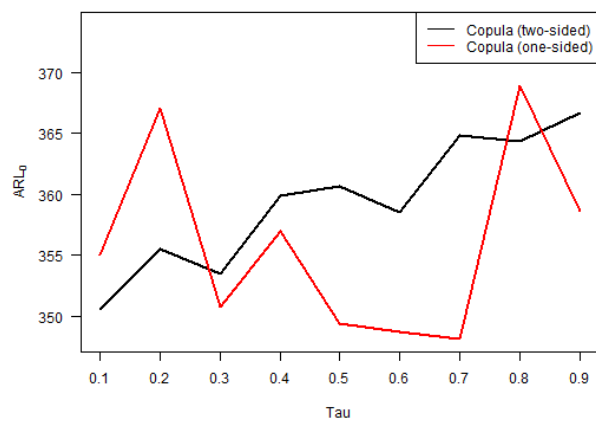
Figure B.1: ARL_0 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Clayton copula.
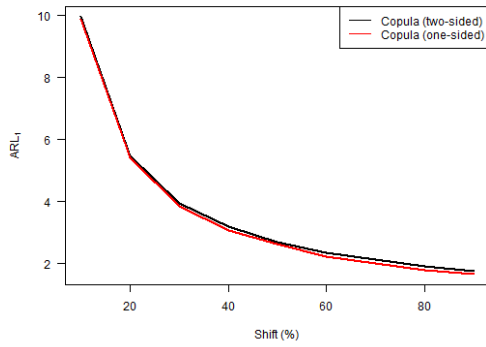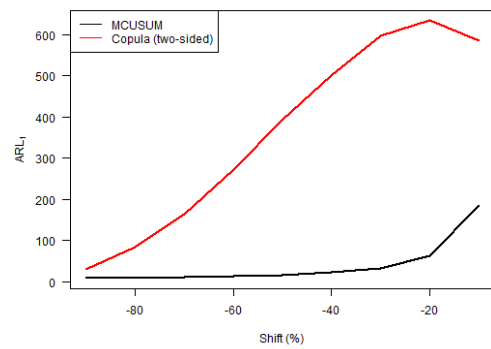
(a) Level of the first variable increased.
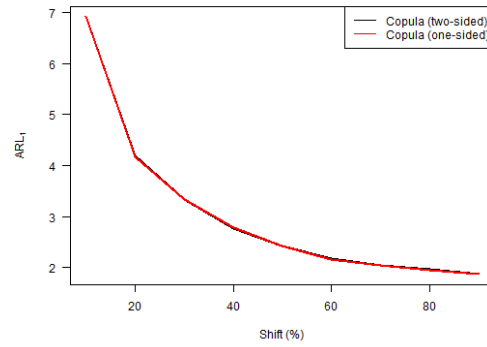
(b) Level of both variables increased.
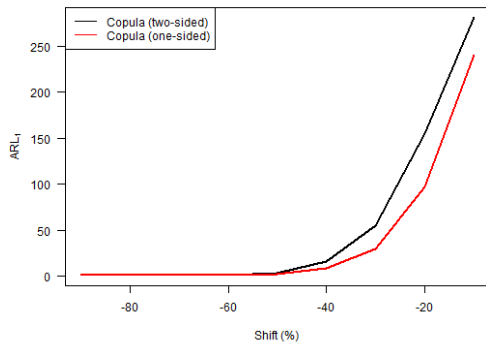
(c) Level of the first variable decreased.
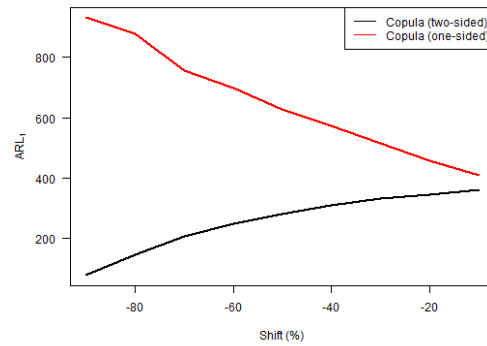
(d) Level of both variables decreased.

Figure B.2: ARL_1 vs. level shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Clayton copula.
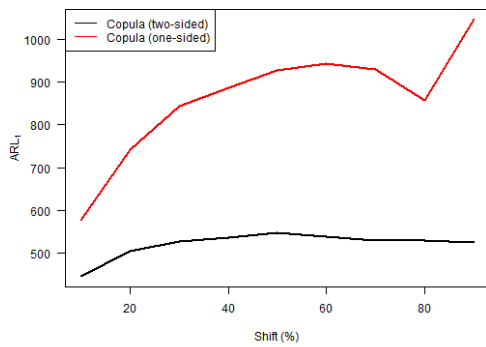
(a) Largest eigenvalue increased.
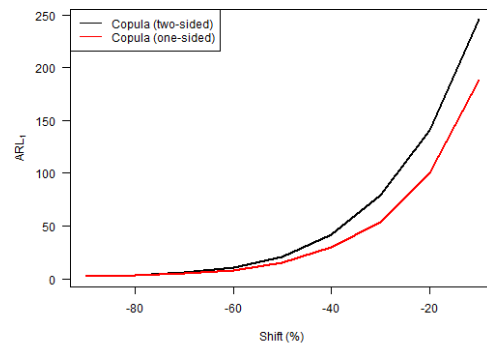
(b) Largest eigenvalue decreased.

Figure B.3: ARL_1 vs. largest eigenvalue shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Clayton copula.
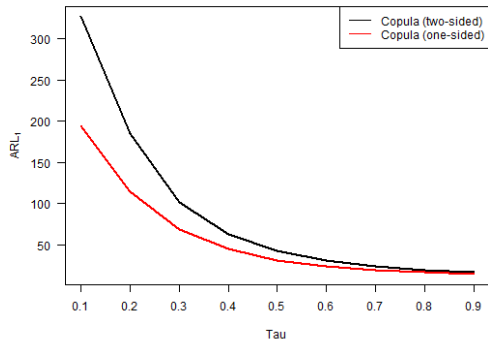
(a) Shift to Gumbel copula.
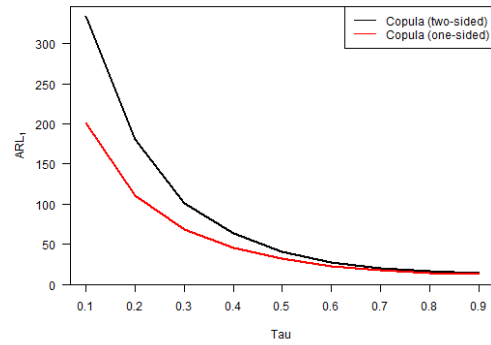
(b) Shift to Frank copula.
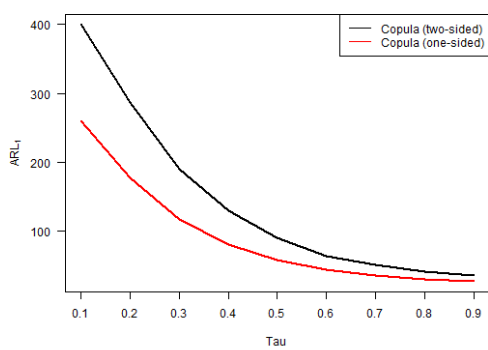
(c) Shift to Normal copula.
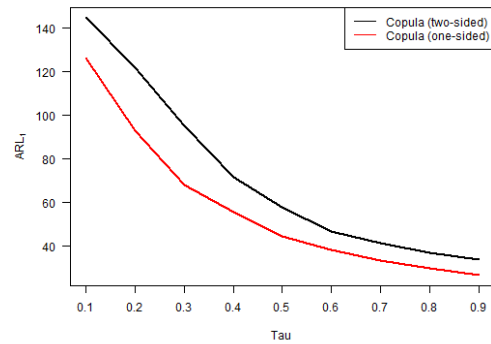
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.4: ARL_1 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Clayton copula and shifts to a different copula.

Frank (case 3 & cases $26 - 35$)
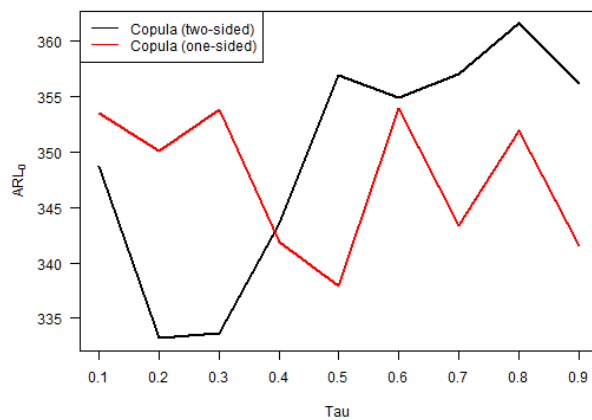
Figure B.5: ARL_0 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Frank copula.
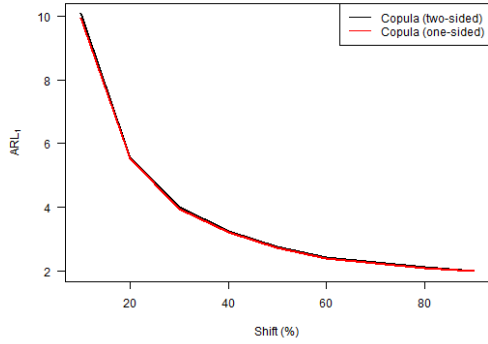
(a) Level of the first variable increased.
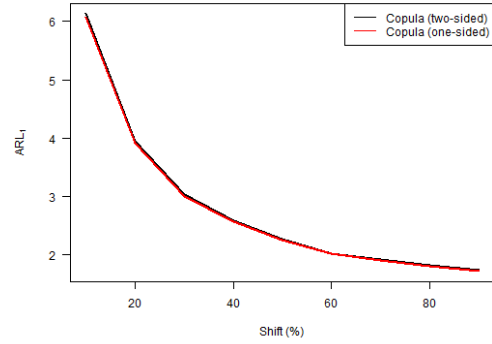
(b) Level of both variables increased.
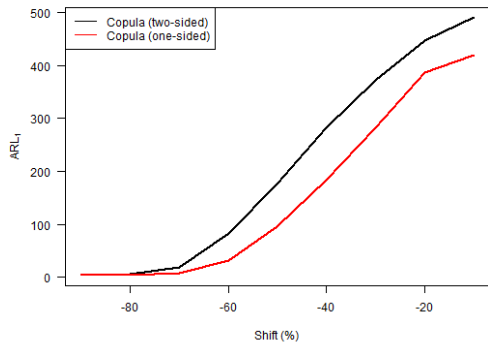
(c) Level of the first variable decreased.
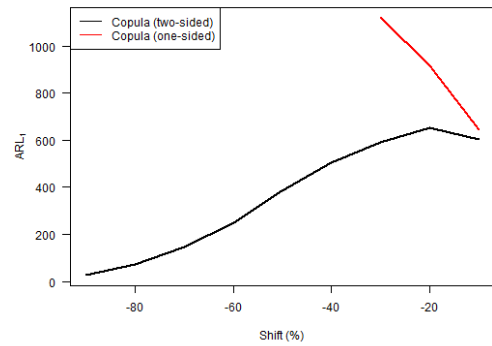
(d) Level of both variables decreased.

Figure B.6: ARL_1 vs. level shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Frank copula.
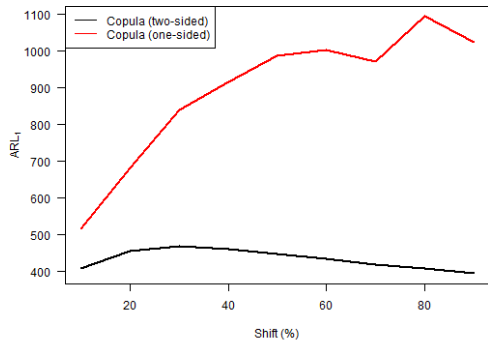
(a) Largest eigenvalue increased.
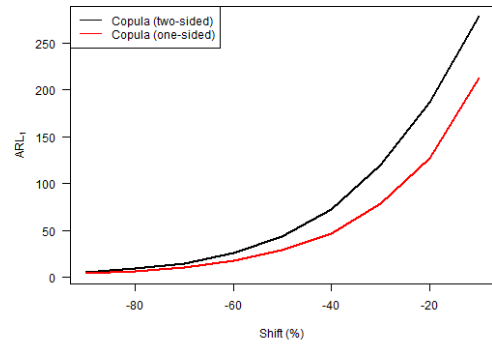
(b) Largest eigenvalue decreased.

Figure B.7: ARL_1 vs. largest eigenvalue shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Frank copula.
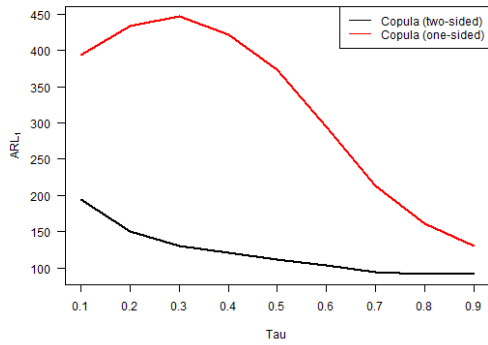
(a) Shift to Gumbel copula.
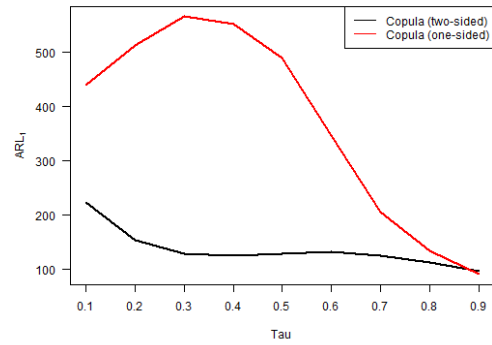
(b) Shift to Clayton copula.
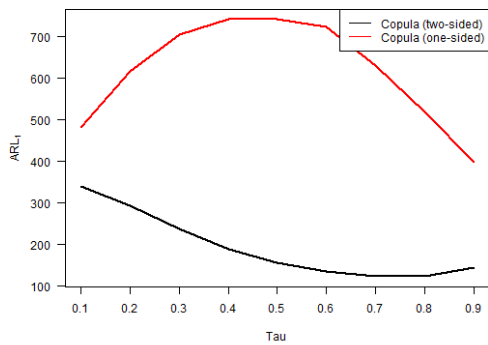
(c) Shift to Normal copula.
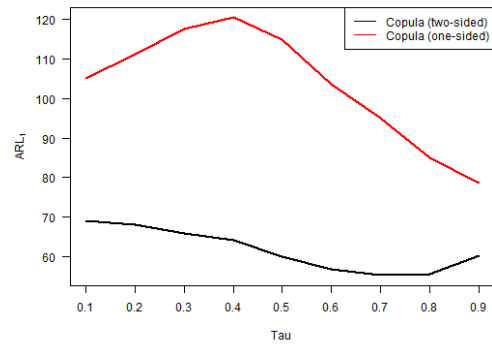
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.8: ARL_1 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Frank copula and shifts to a different copula.
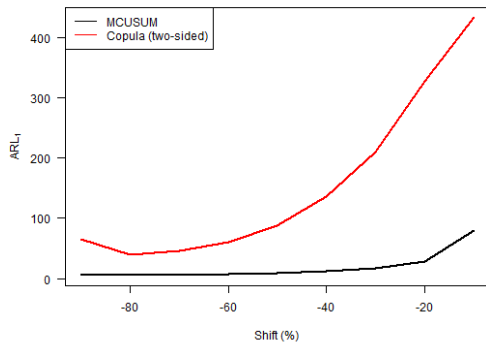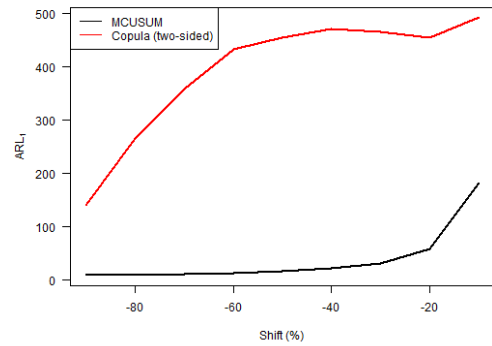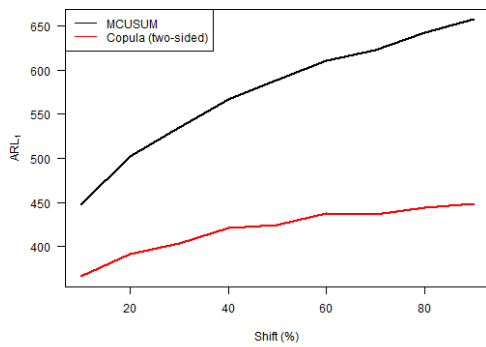
Student-t (case 5 & cases $46 - 55$)
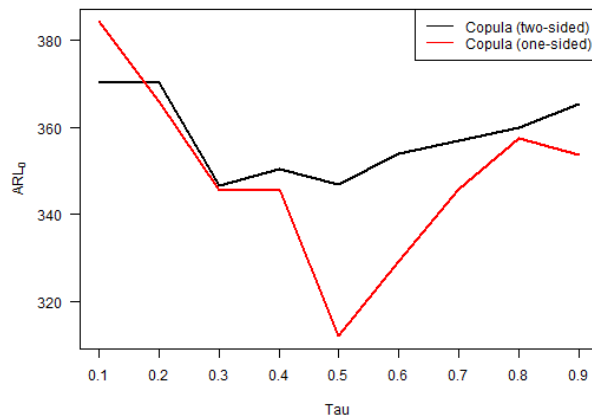
Figure B.9: ARL_0 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Student-t copula with 3 degrees of freedom.
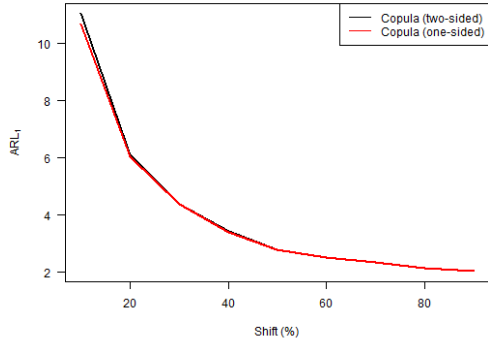
(a) Level of the first variable increased.
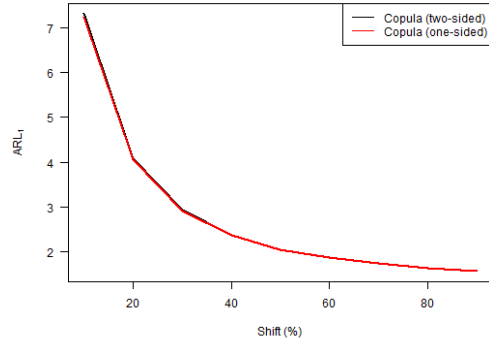
(b) Level of both variables increased.
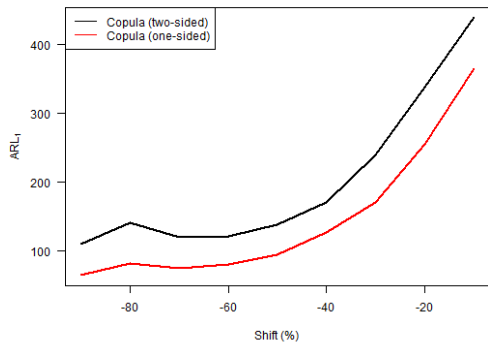
(c) Level of the first variable decreased.
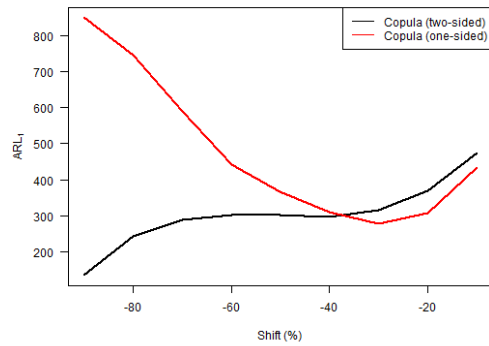
(d) Level of both variables decreased.

Figure B.10: ARL_1 vs. level shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Student-t copula with 3 degrees of freedom.
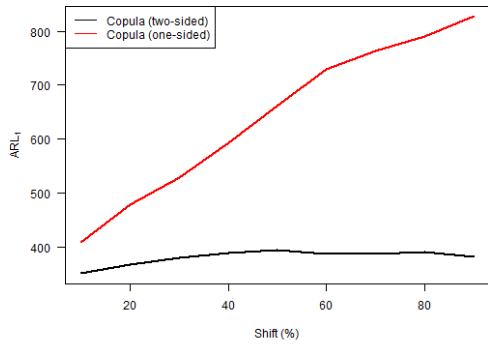

(a) Largest eigenvalue increased.
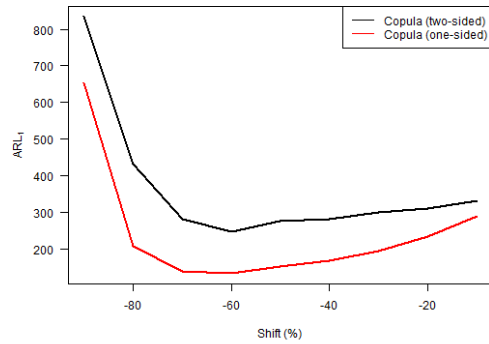
(b) Largest eigenvalue decreased.

Figure B.11: ARL_1 vs. largest eigenvalue shift for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Student-t copula with 3 degrees of freedom.
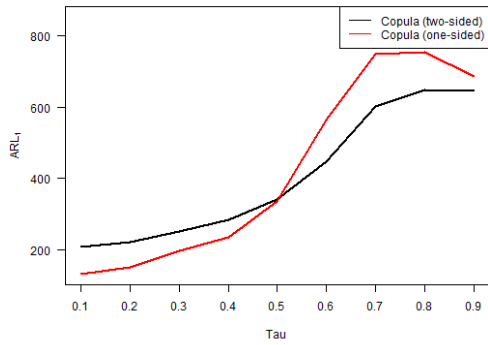
(a) Shift to Gumbel copula.
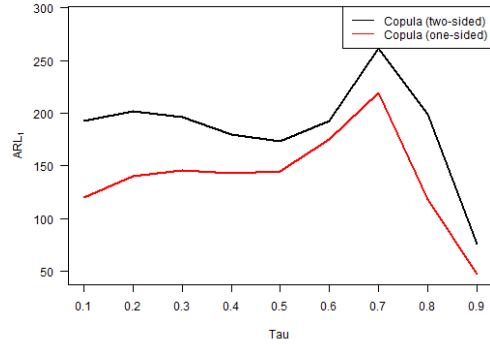
(b) Shift to Clayton copula.
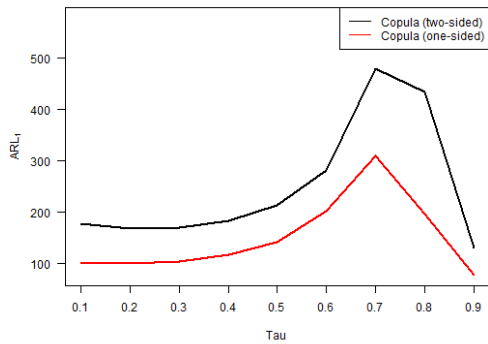
(c) Shift to Frank copula.
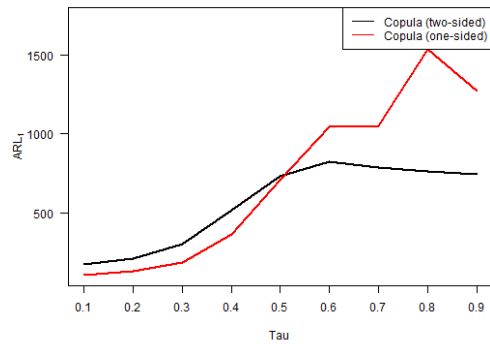
(d) Shift to Normal copula.

Figure B.12: ARL_1 vs. Kendall's τ for the one-sided and two-sided version of the copula-based control chart, where the in-control distribution follows a Student-t copula with 3 degrees of freedom and shifts to a different copula.

B.2 Copula-based vs. Hotelling T^2
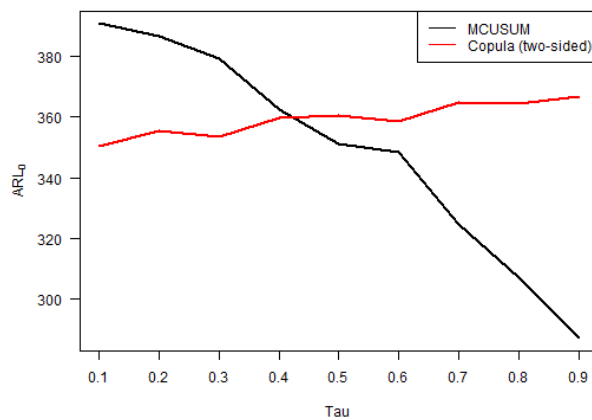
Clayton (case 57 & cases $71-80$)
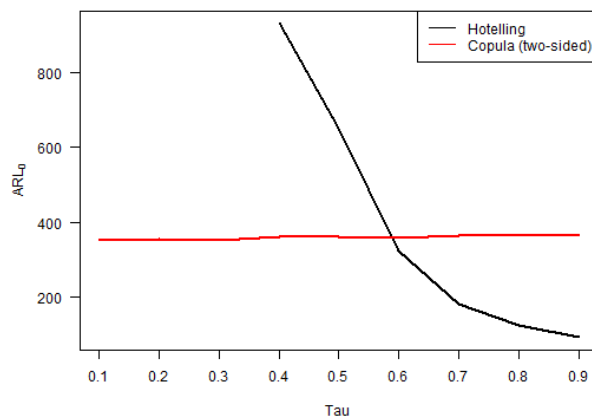
Figure B.13: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Clayton copula.
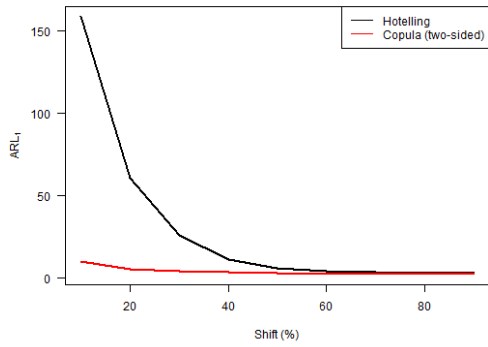
(a) Level of the first variable increased. (b) Level of both variables increased.
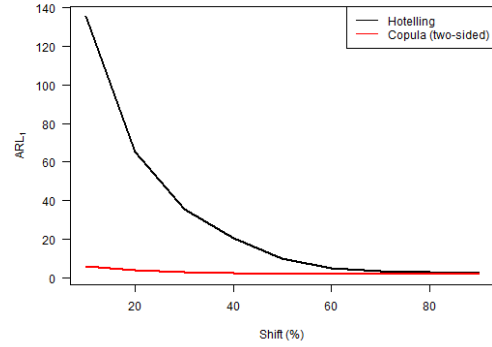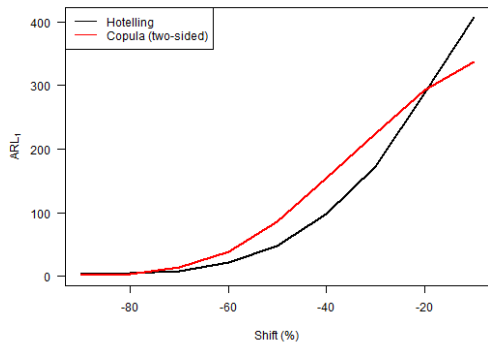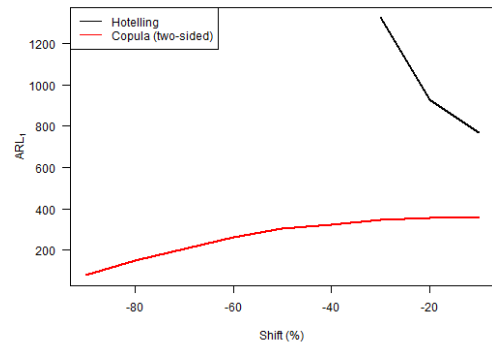
(c) Level of the first variable decreased. (d) Level of both variables decreased.

Figure B.14: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Clayton copula.
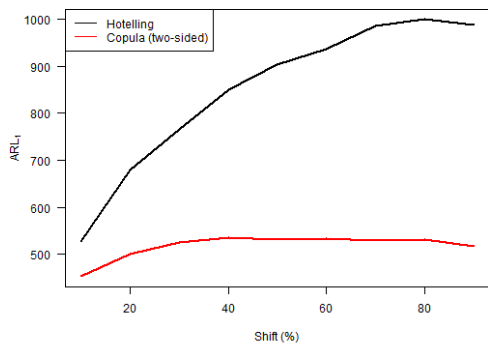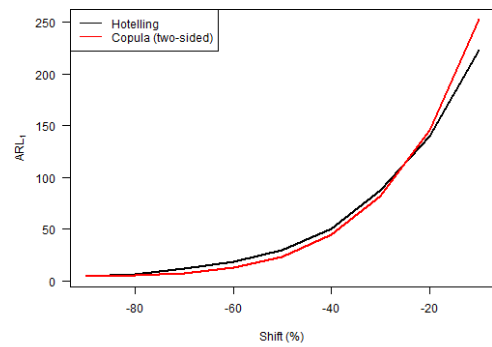
(a) Largest eigenvalue increased. (b) Largest eigenvalue decreased.

Figure B.15: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Clayton copula.
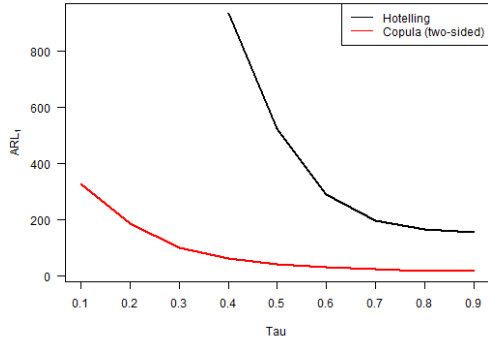
(a) Shift to Gumbel copula.
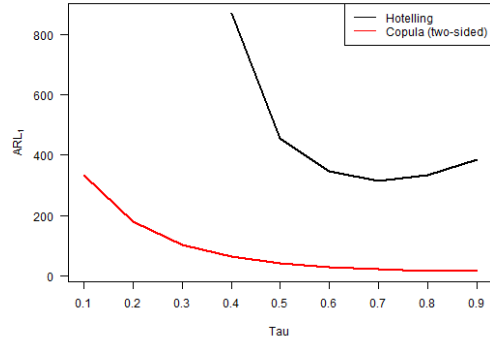
(b) Shift to Frank copula.
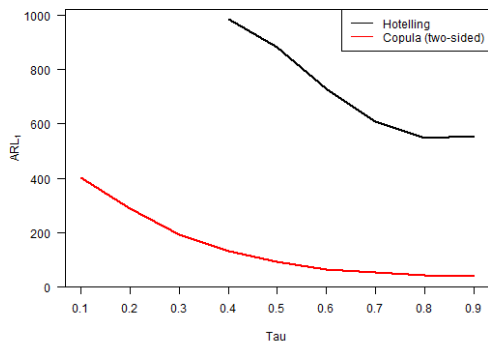
(c) Shift to Normal copula.
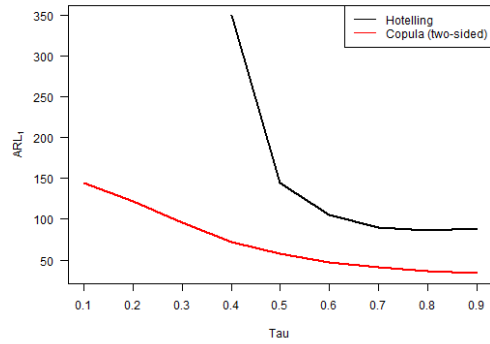
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.16: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Clayton copula and shifts to a different copula.

Frank (case 58 & cases 81 − 90)
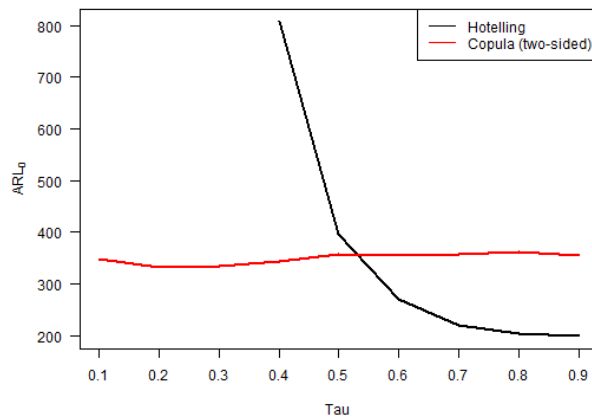
Figure B.17: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Frank copula.
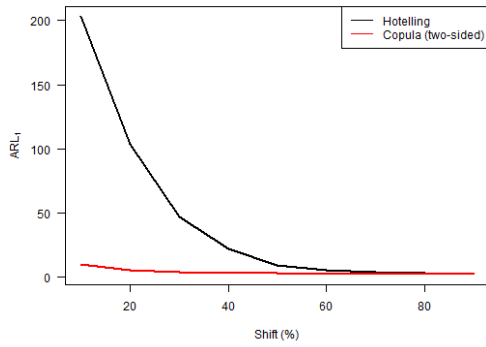
(a) Level of the first variable increased.
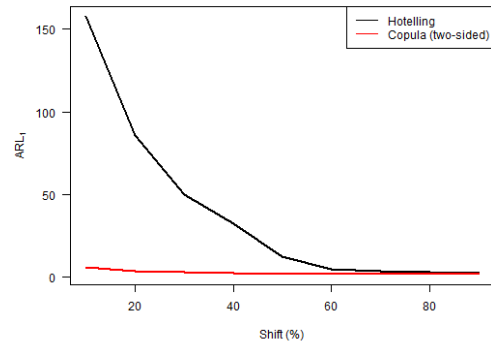
(b) Level of both variables increased.
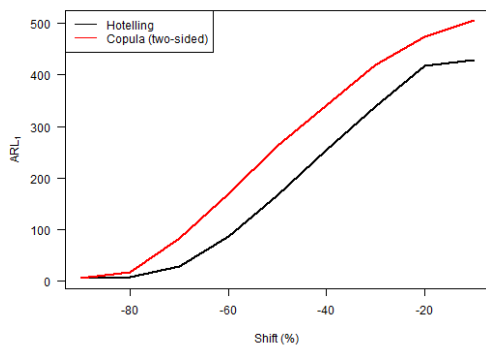

(c) Level of the first variable decreased.
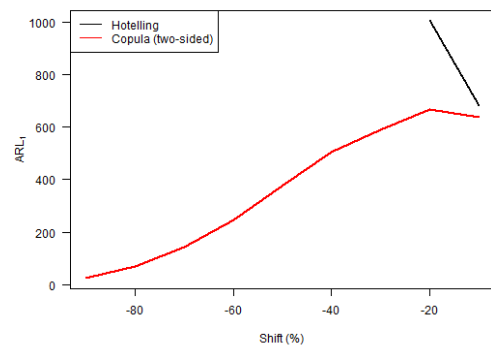
(d) Level of both variables decreased.

Figure B.18: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Frank copula.
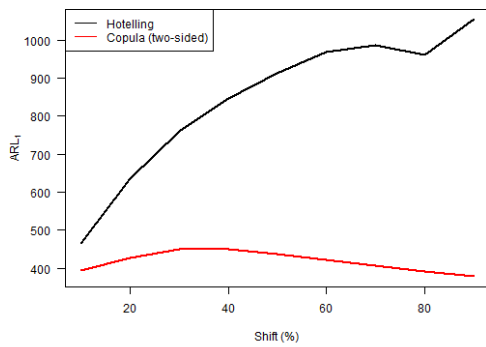

(a) Largest eigenvalue increased.
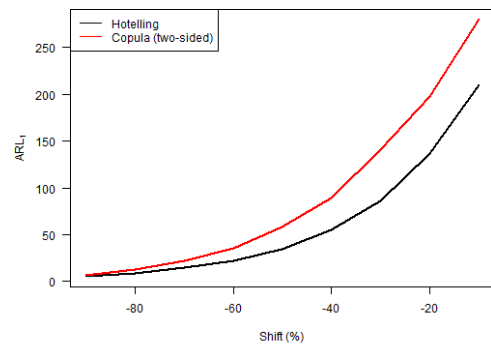
(b) Largest eigenvalue decreased.

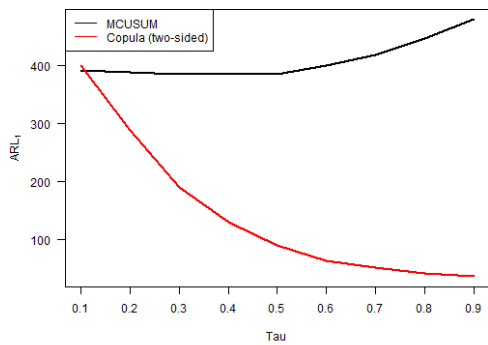Figure B.19: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Frank copula.
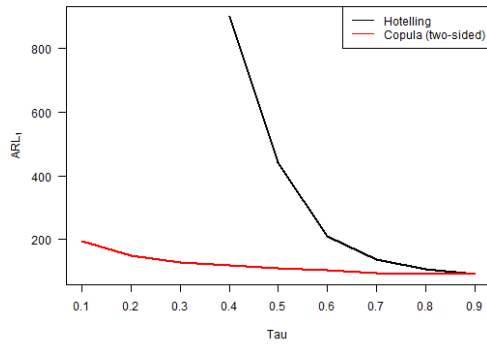
(a) Shift to Gumbel copula.
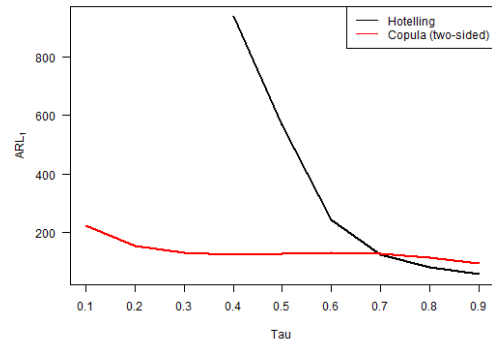
(b) Shift to Clayton copula.
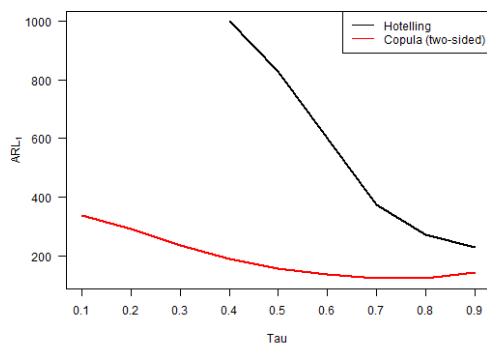
(c) Shift to Normal copula.
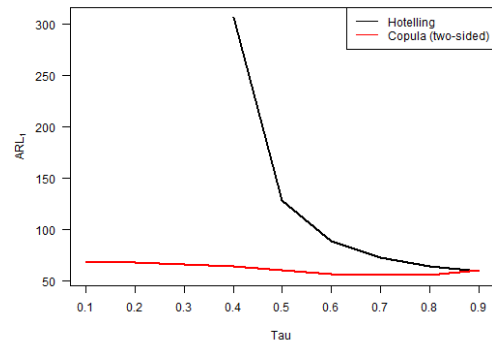
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.20: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Frank copula and shifts to a different copula.

Student-t (case 60 & cases $101 - 110$)
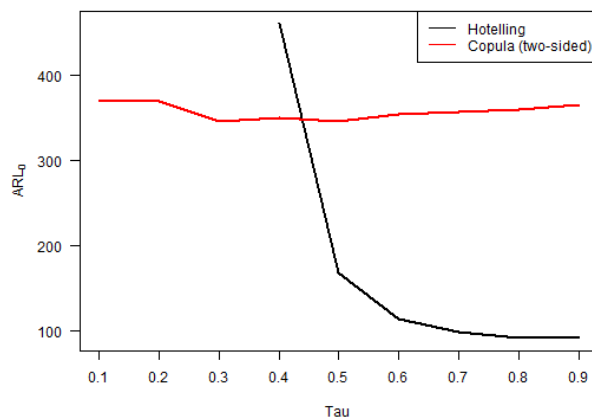
Figure B.21: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Student-t copula.
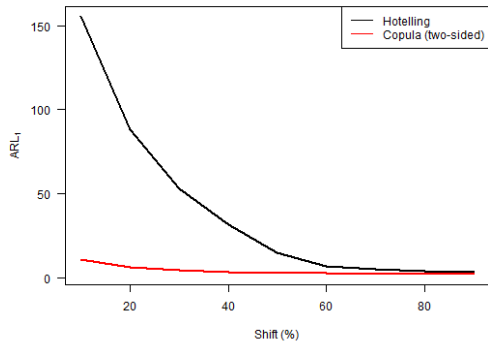
(a) Level of the first variable increased.
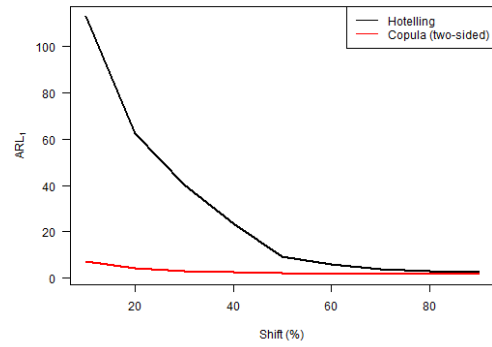
(b) Level of both variables increased.
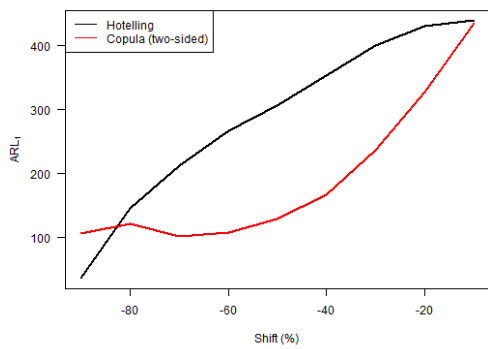
(c) Level of the first variable decreased.
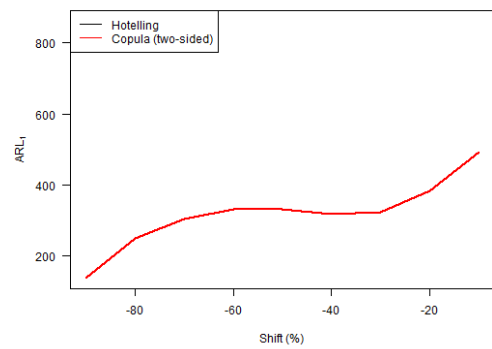
(d) Level of both variables decreased.
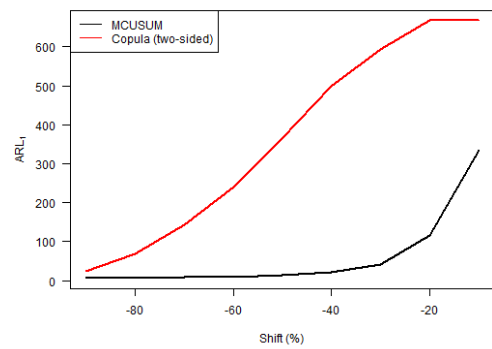
Figure B.22: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Student-t copula.
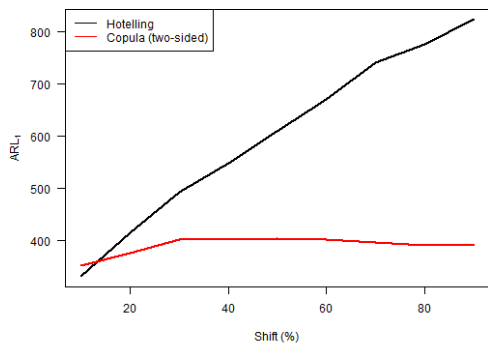
(a) Largest eigenvalue increased.
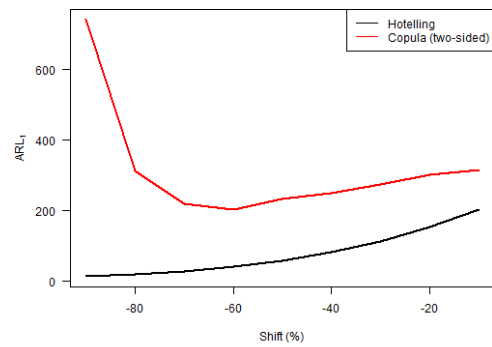
(b) Largest eigenvalue decreased.

Figure B.23: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Student-t copula.
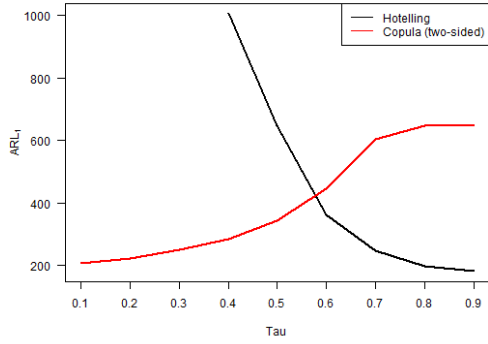
(a) Shift to Gumbel copula.
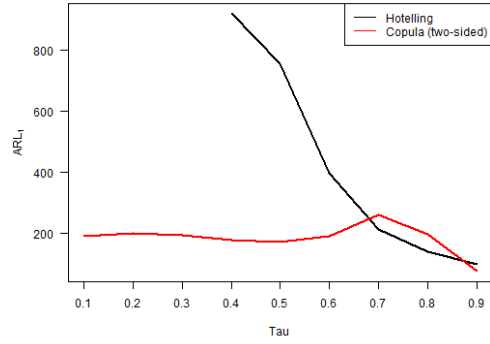
(b) Shift to Clayton copula.
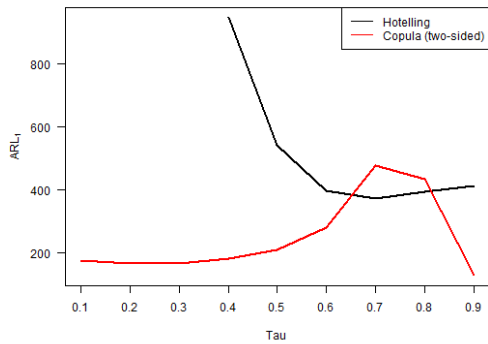
(c) Shift to Frank copula.
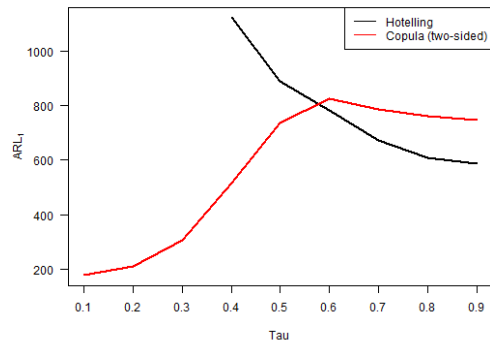
(d) Shift to Normal copula.

Figure B.24: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and Hotelling T^2 chart, where the in-control distribution follows a Student-t copula and shifts to a different copula.

B.3 Copula-based vs. MCUSUM

$k = 0.25$

Clayton (case 112 & cases 126 − 135)
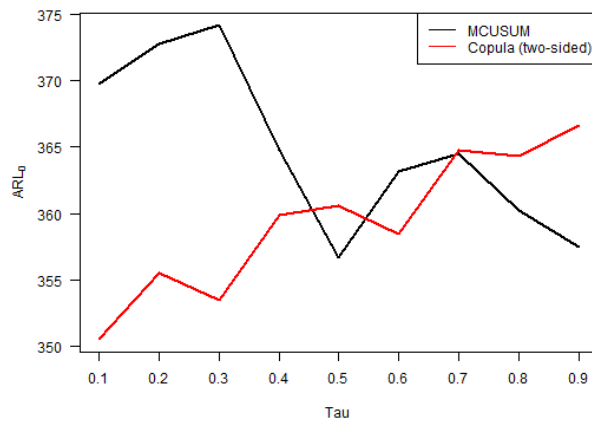
Figure B.25: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Clayton copula.
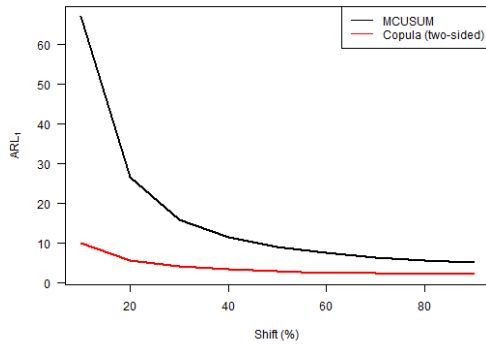
(a) Level of the first variable increased.
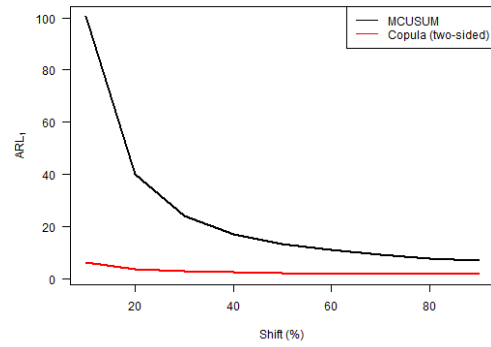
(b) Level of both variables increased.
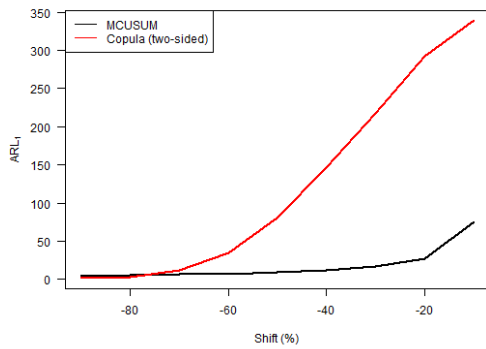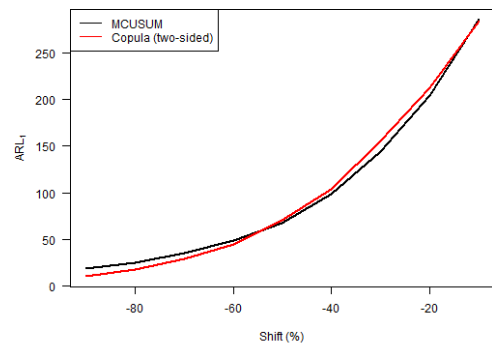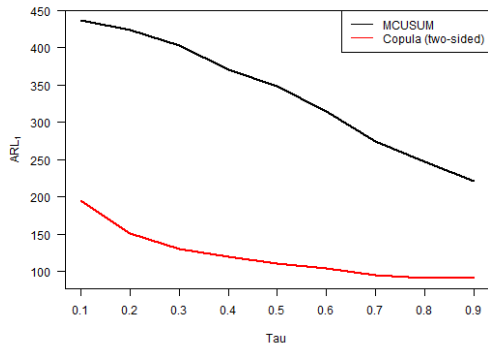

(c) Level of the first variable decreased.
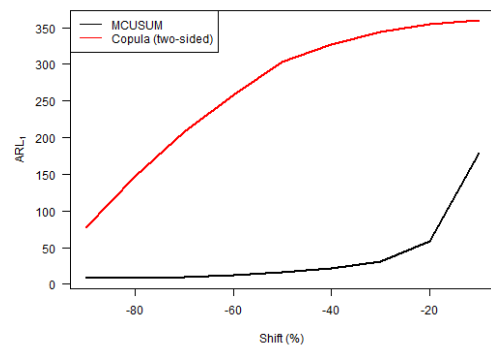
(d) Level of both variables decreased.

Figure B.26: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Clayton copula.
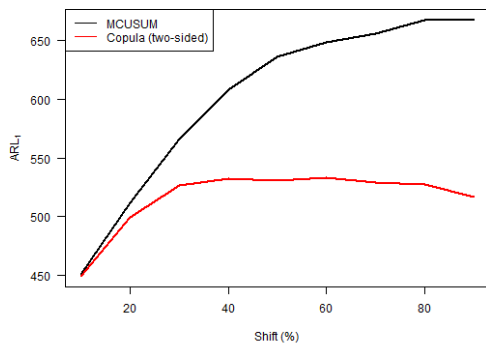

(a) Largest eigenvalue increased.
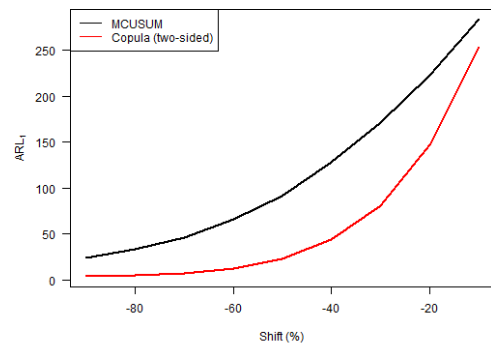
(b) Largest eigenvalue decreased.

Figure B.27: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Clayton copula.
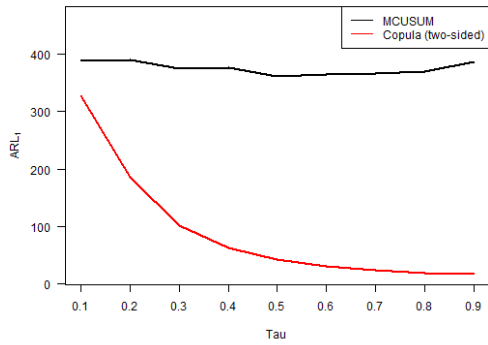
(a) Shift to Gumbel copula.
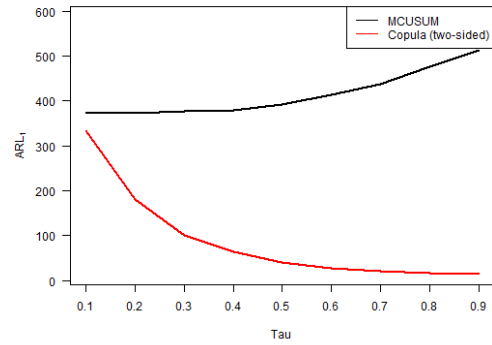
(b) Shift to Frank copula.
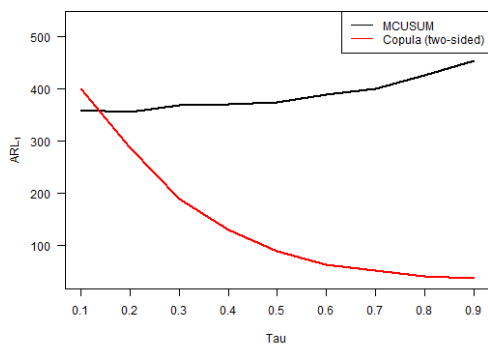
(c) Shift to Normal copula.
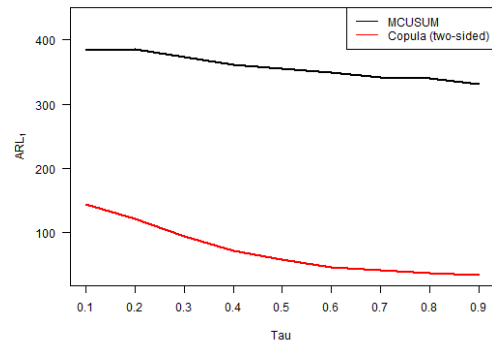
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.28: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Clayton copula and shifts to a different copula.
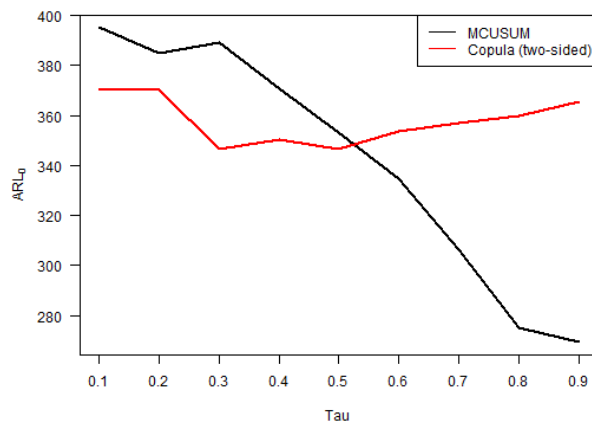
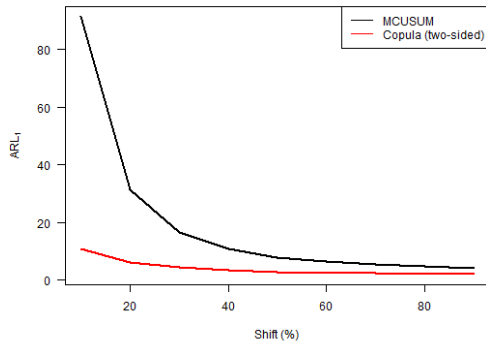**Frank (case 113 & cases 136 − 145)**
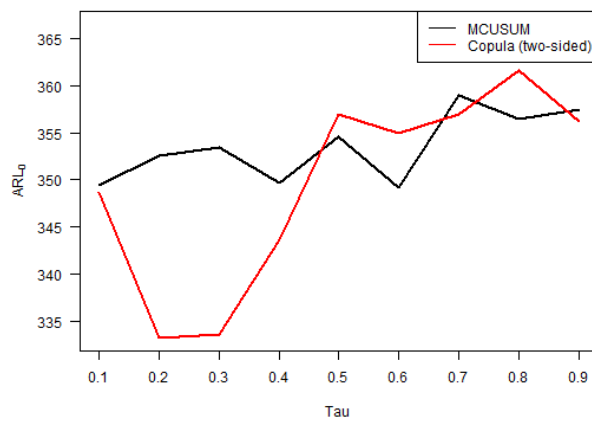
Figure B.29: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Frank copula.
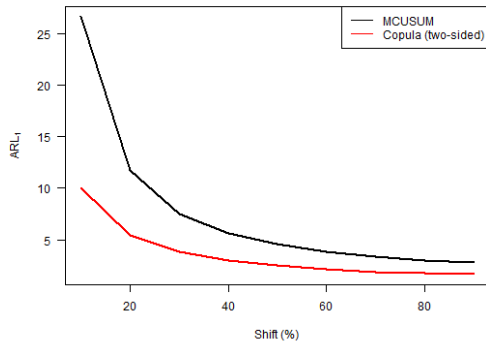
(a) Level of the first variable increased.
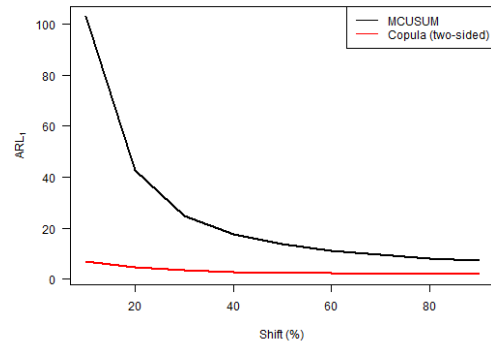
(b) Level of both variables increased.
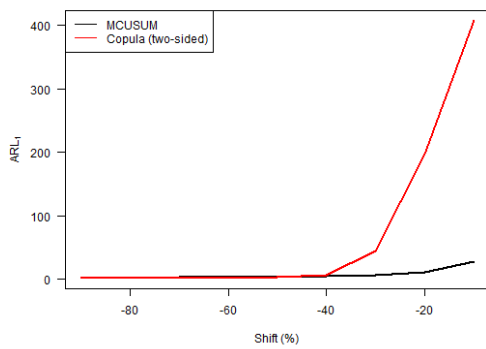
(c) Level of the first variable decreased.

(d) Level of both variables decreased.

Figure B.30: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Frank copula.
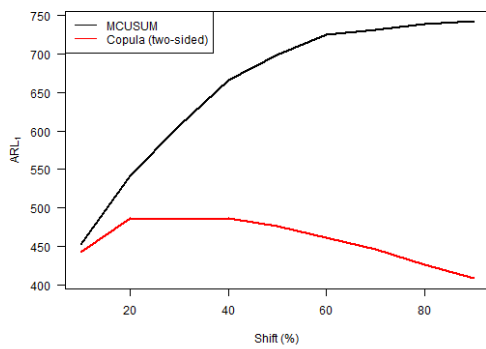
(a) Largest eigenvalue increased.
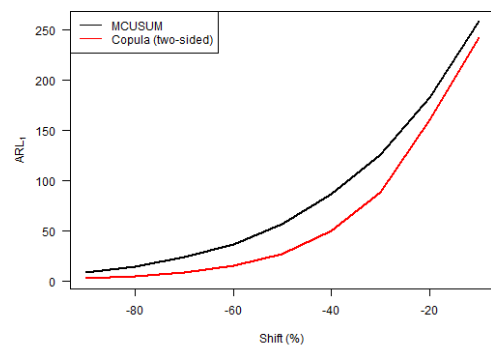
(b) Largest eigenvalue decreased.

Figure B.31: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Frank copula.
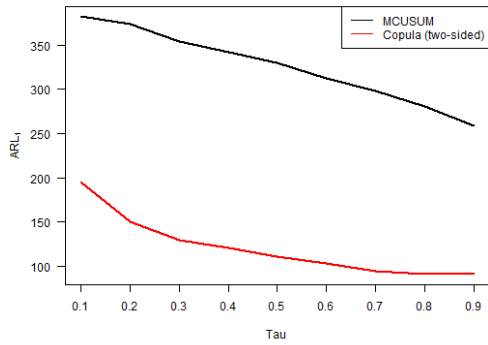
(a) Shift to Gumbel copula.
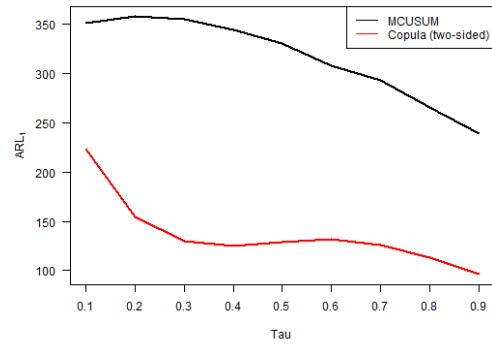
(b) Shift to Clayton copula.
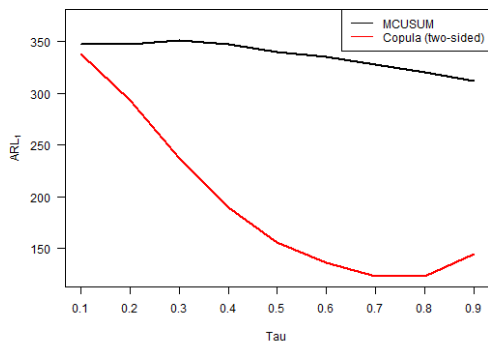
(c) Shift to Normal copula.
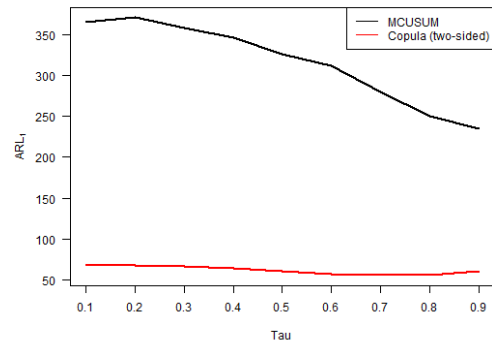
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.32: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Frank copula and shifts to a different copula.
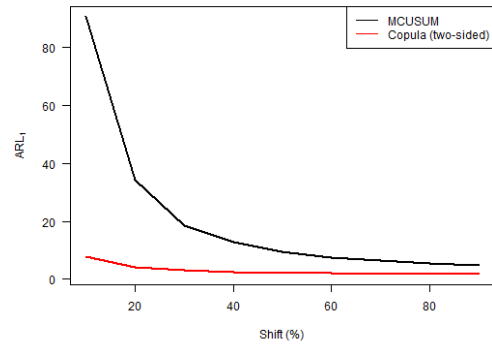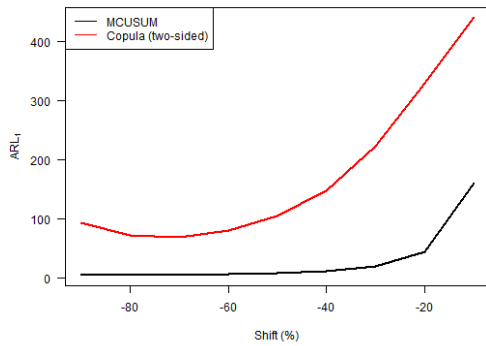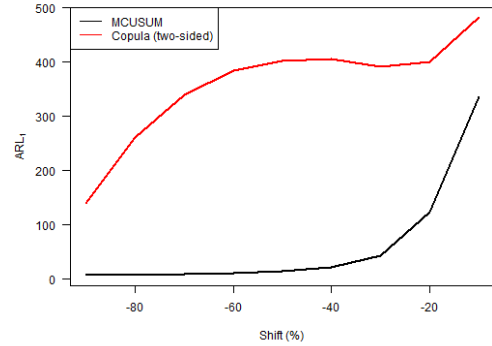
Student-t (case 115 & cases 156 − 165)
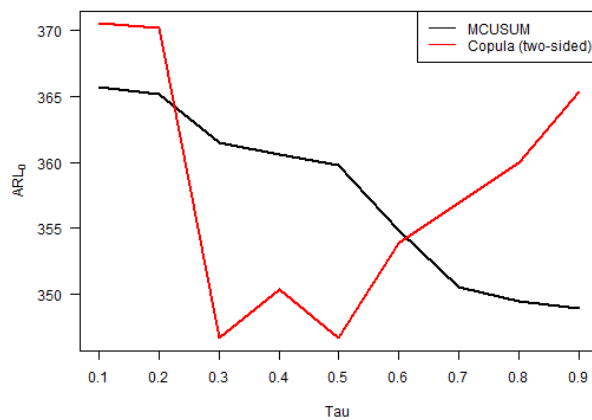
Figure B.33: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Student-t copula.
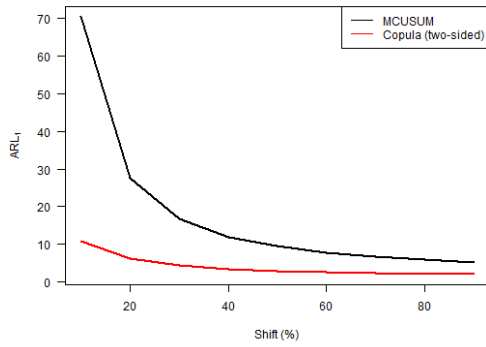
(a) Level of the first variable increased.
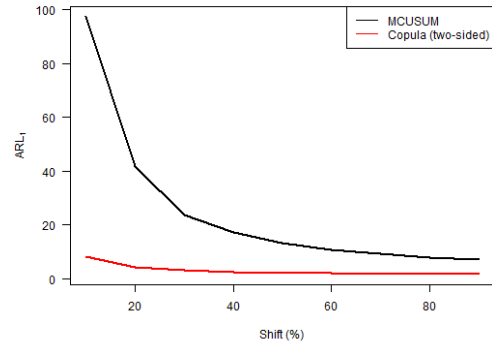

(b) Level of both variables increased.
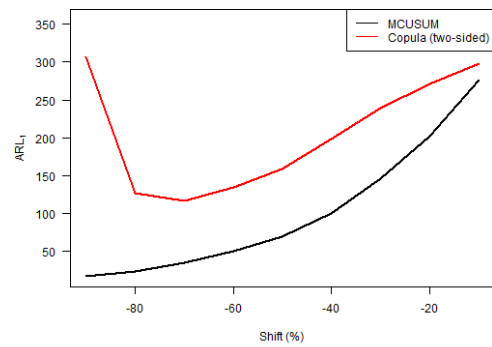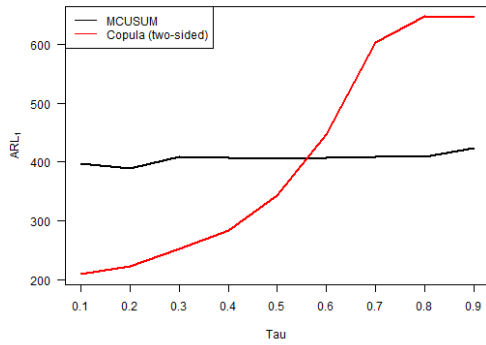

(c) Level of the first variable decreased.

(d) Level of both variables decreased.

Figure B.34: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Student-t copula.


(a) Largest eigenvalue increased.
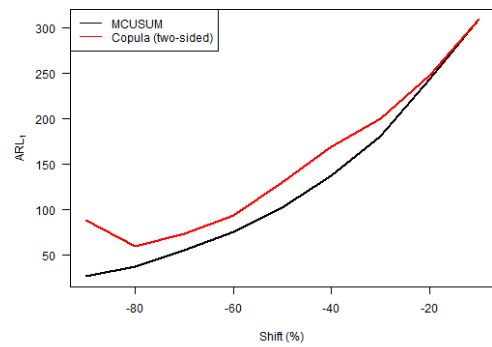
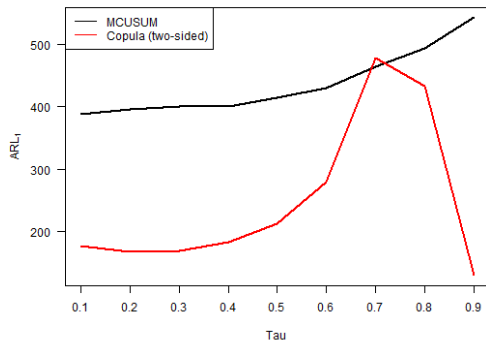

(b) Largest eigenvalue decreased.

Figure B.35: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Student-t copula.
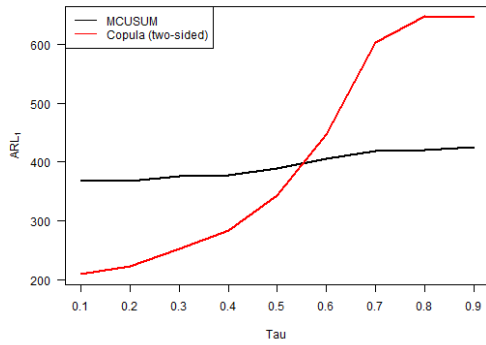
(a) Shift to Gumbel copula.
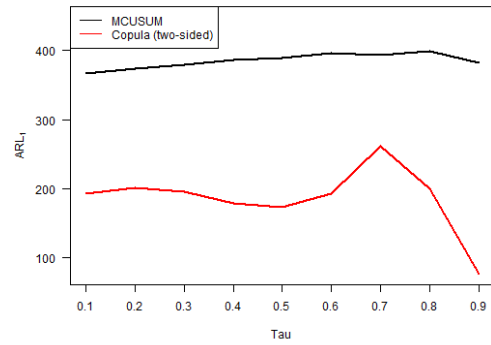
(b) Shift to Clayton copula.
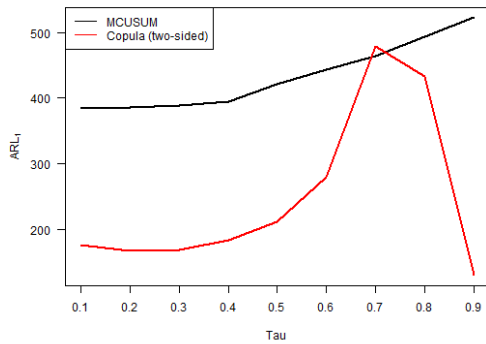
(c) Shift to Frank copula.
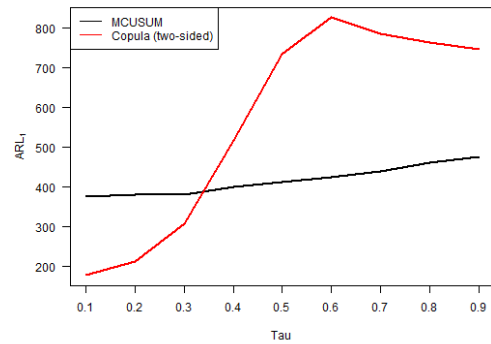
(d) Shift to Normal copula.

Figure B.36: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.25$, where the in-control distribution follows a Student-t copula and shifts to a different copula.
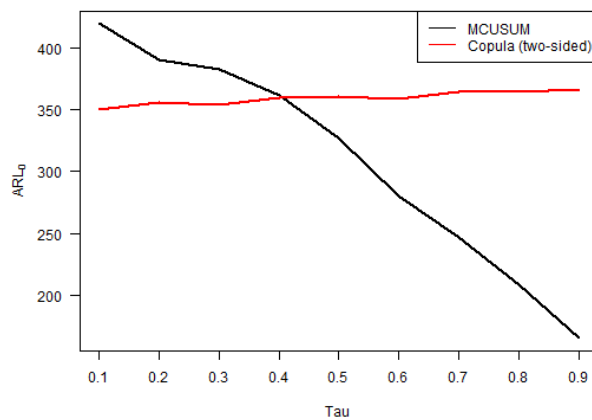
$k = 0.50$

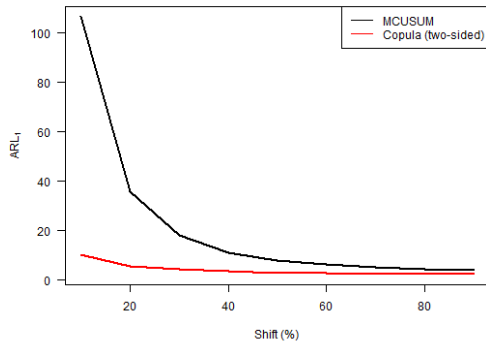### Clayton (case 167 & cases $181 - 190$)

Figure B.37: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Clayton copula.
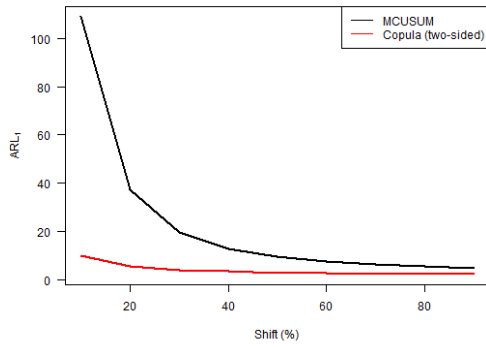
(a) Level of the first variable increased.
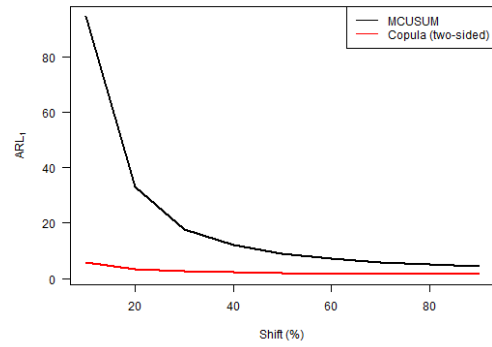
(b) Level of both variables increased.
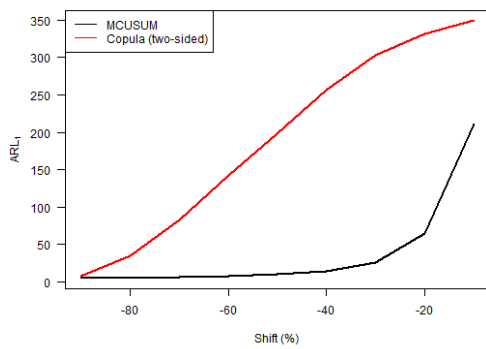
(c) Level of the first variable decreased.
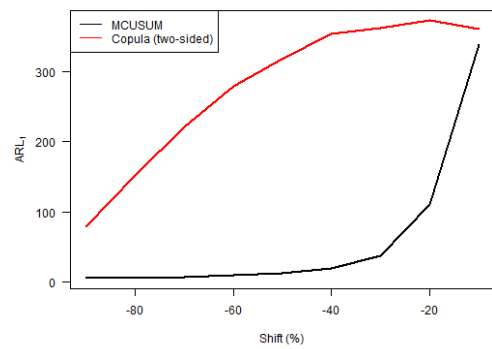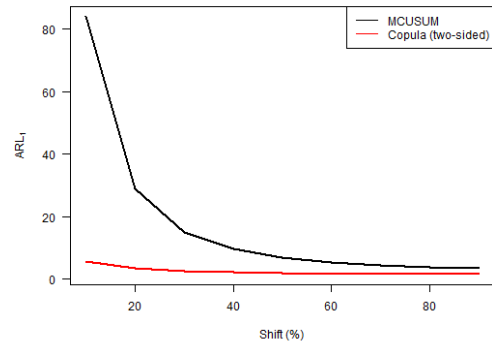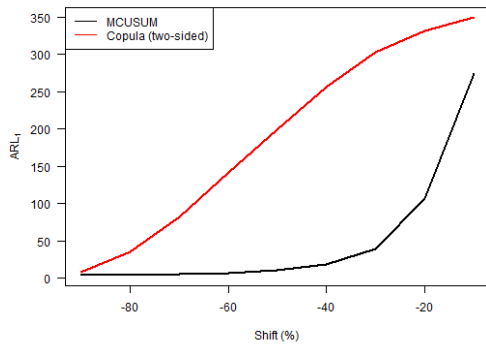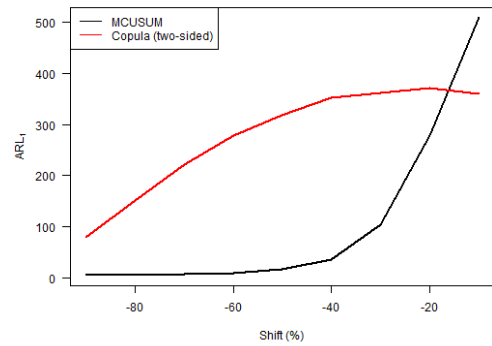
(d) Level of both variables decreased.

Figure B.38: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Clayton copula.
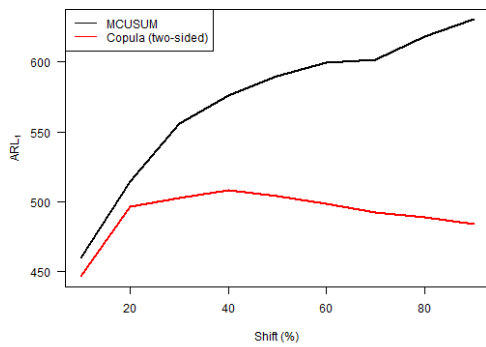
(a) Largest eigenvalue increased.
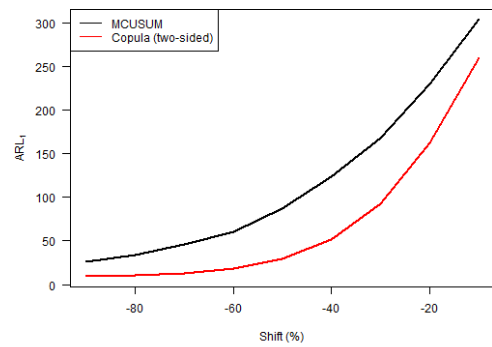
(b) Largest eigenvalue decreased.

Figure B.39: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Clayton copula.
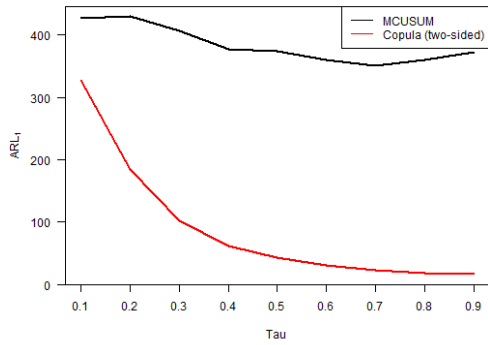
(a) Shift to Gumbel copula.
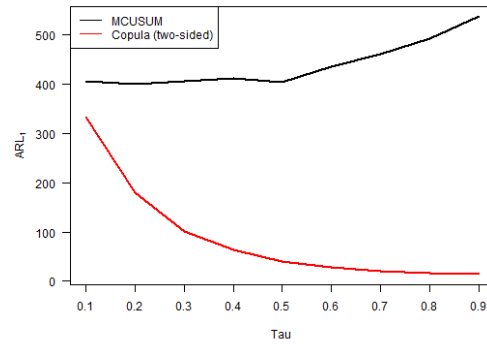
(b) Shift to Frank copula.

(c) Shift to Normal copula.
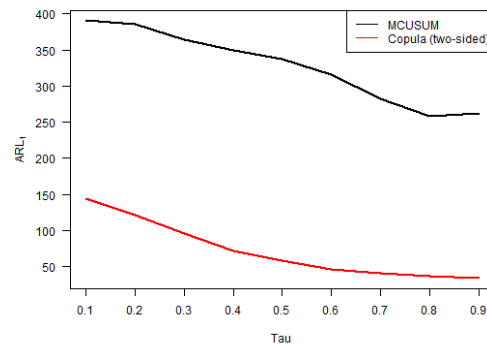
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.40: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Clayton copula and shifts to a different copula.

Frank (case 168 & cases $191 - 200$)
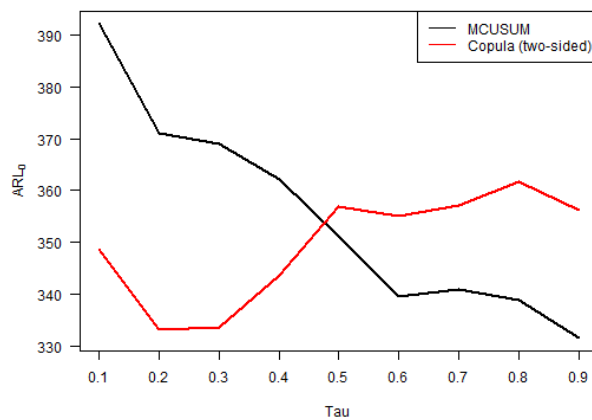
Figure B.41: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Frank copula.
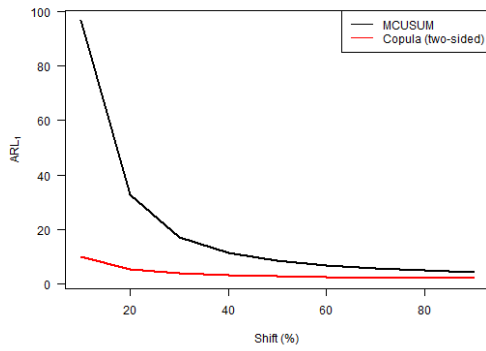
(a) Level of the first variable increased.
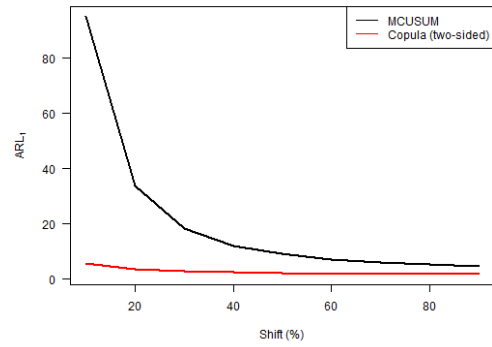
(b) Level of both variables increased.
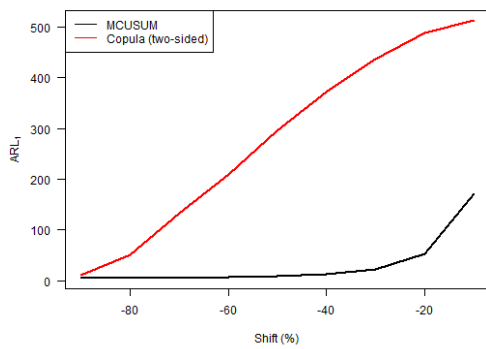
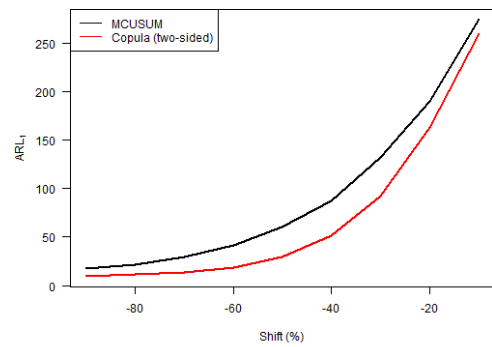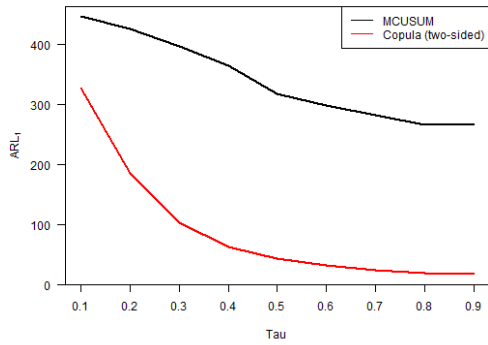(c) Level of the first variable decreased.

(d) Level of both variables decreased.

Figure B.42: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Frank copula.
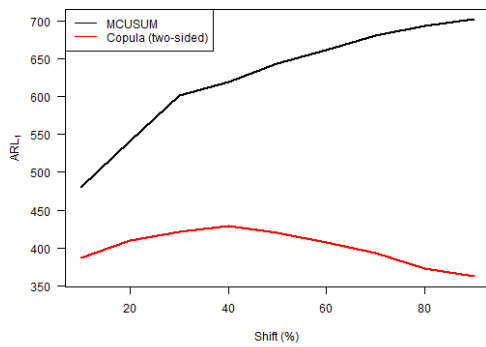

(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure B.43: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Frank copula.

(a) Shift to Gumbel copula.
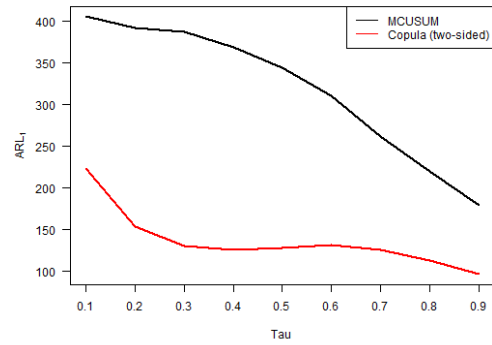
(b) Shift to Clayton copula.
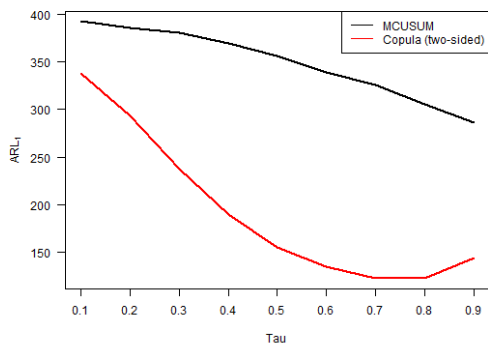
(c) Shift to Normal copula.
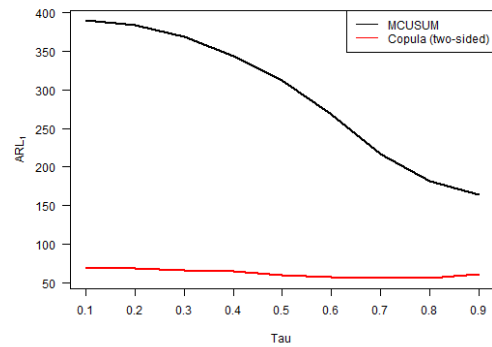
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.44: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Frank copula and shifts to a different copula.

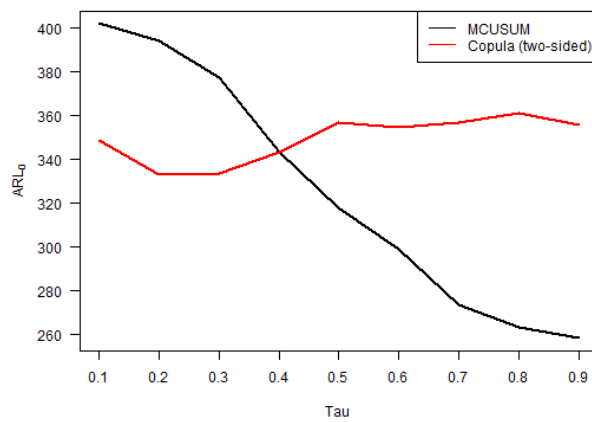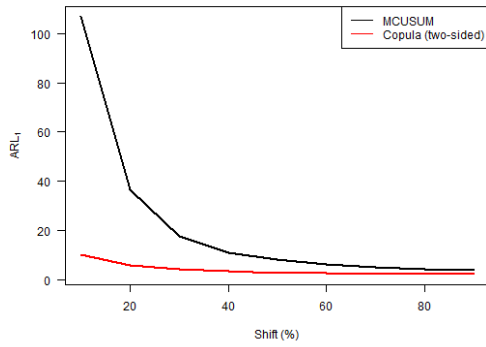**Student-t (case 170 & cases $211 - 220$)**

Figure B.45: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Student-t copula.

(a) Level of the first variable increased.

(b) Level of both variables increased.

(c) Level of the first variable decreased.
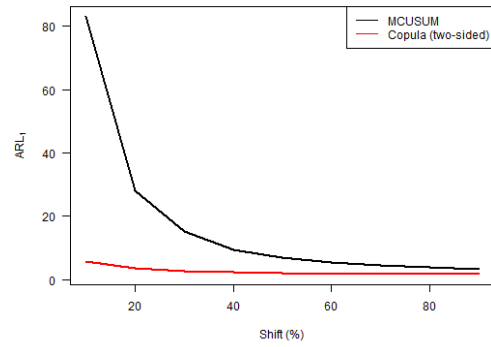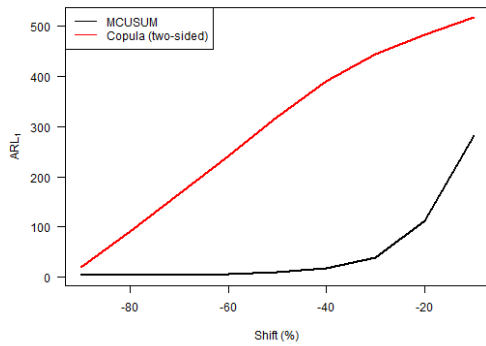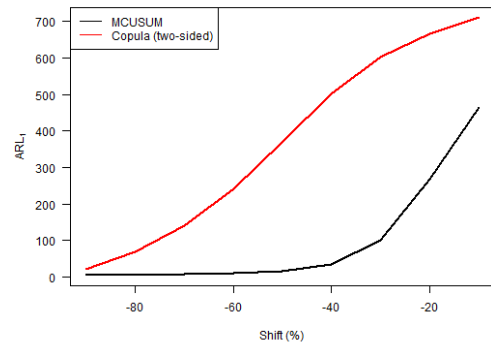
(d) Level of both variables decreased.

Figure B.46: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Student-t copula.
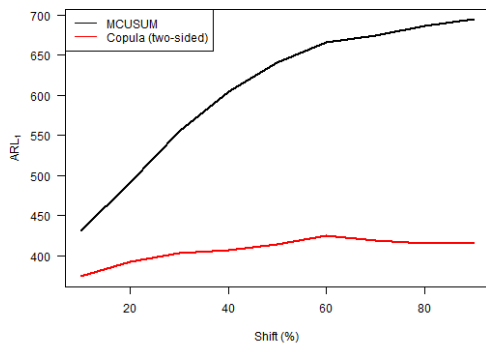
(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure B.47: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Student-t copula.

(a) Shift to Gumbel copula.
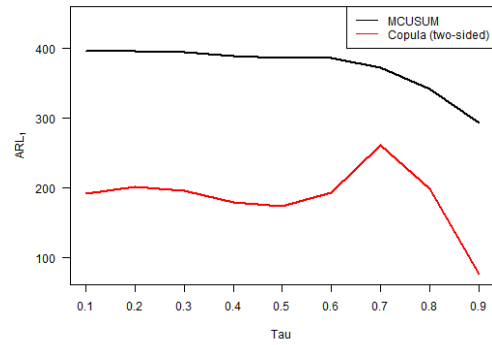
(b) Shift to Clayton copula.

(c) Shift to Frank copula.
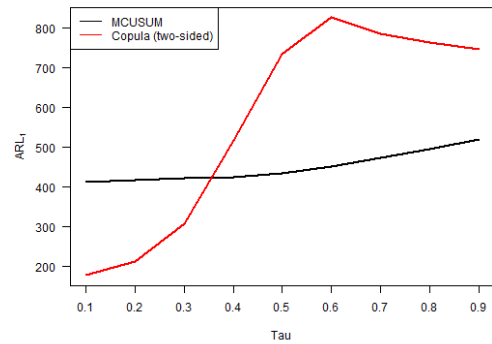
(d) Shift to Normal copula.

Figure B.48: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.50$, where the in-control distribution follows a Student-t copula and shifts to a different copula.

$k = 0.75$

Clayton (case 222 & cases $236 - 245$)

Figure B.49: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Clayton copula.

(a) Level of the first variable increased.

(b) Level of both variables increased.

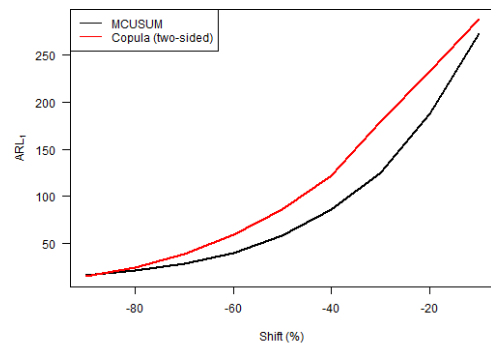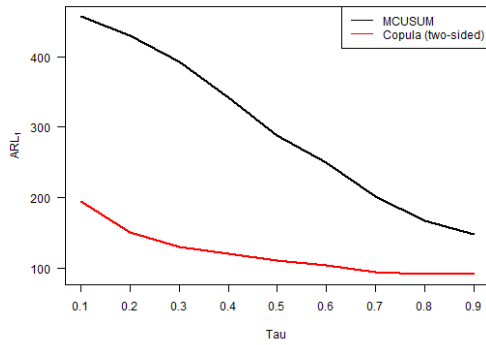(c) Level of the first variable decreased.

(d) Level of both variables decreased.

Figure B.50: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Clayton copula.
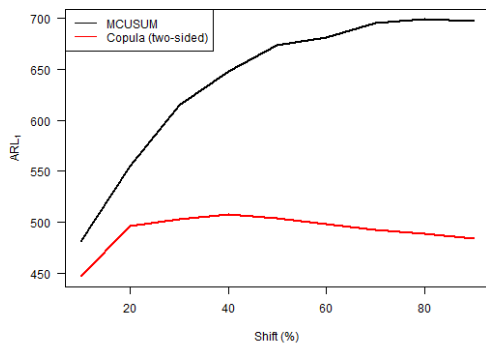

(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure B.51: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Clayton copula.

(a) Shift to Gumbel copula.
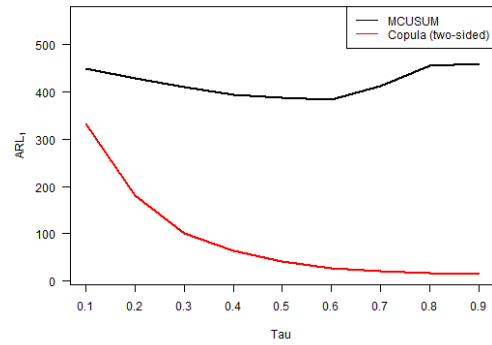
(b) Shift to Frank copula.
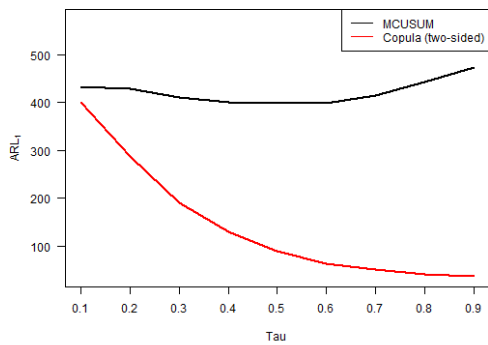
(c) Shift to Normal copula.
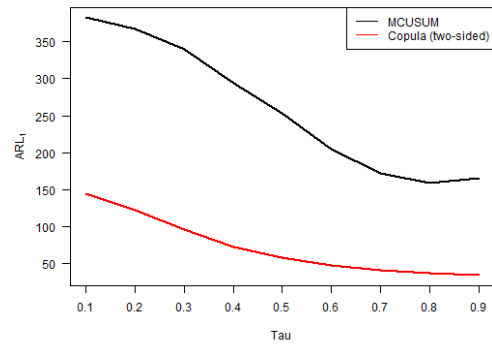
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.52: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Clayton copula and shifts to a different copula.
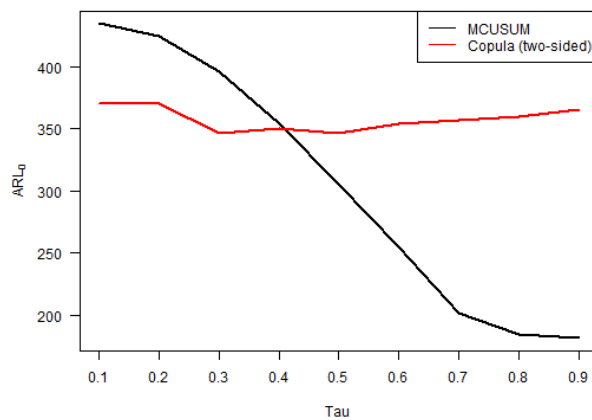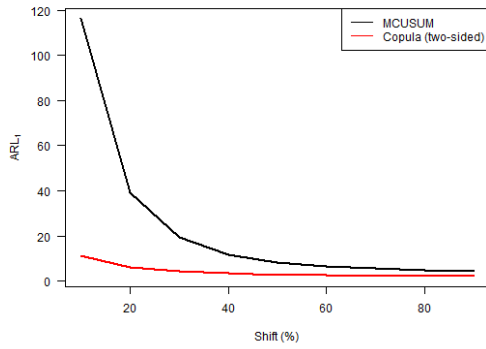
Frank (case 223 & cases 246 − 255)

Figure B.53: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Frank copula.

(a) Level of the first variable increased.

(b) Level of both variables increased.

(c) Level of the first variable decreased.

(d) Level of both variables decreased.

Figure B.54: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Frank copula.
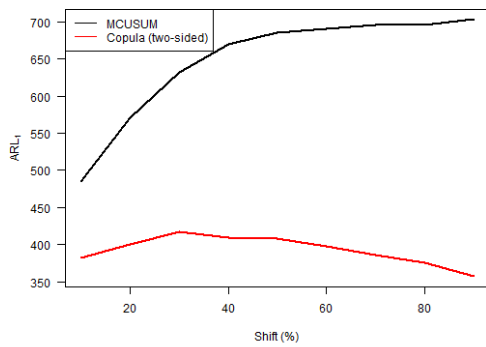
(a) Largest eigenvalue increased.

(b) Largest eigenvalue decreased.

Figure B.55: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Frank copula.

(a) Shift to Gumbel copula.
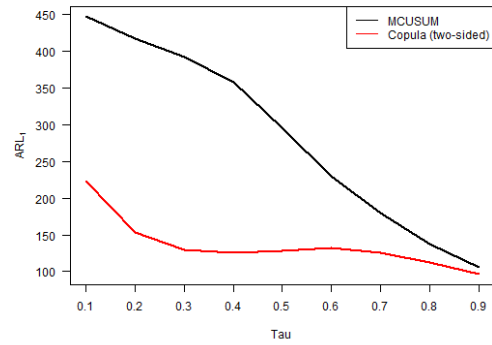
(b) Shift to Clayton copula.
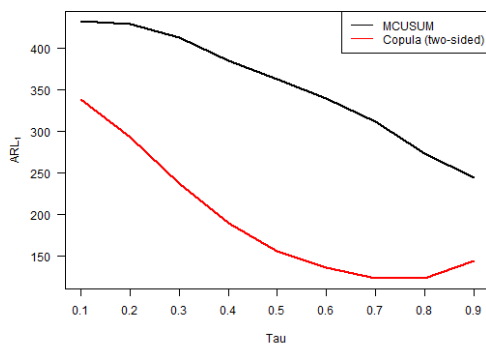
(c) Shift to Normal copula.
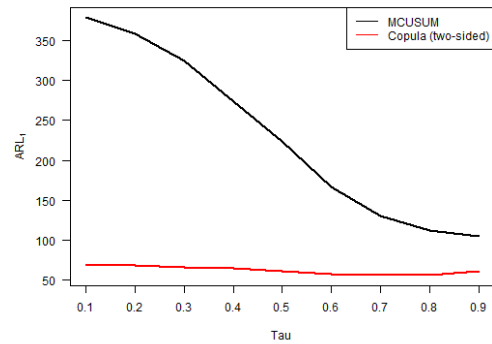
(d) Shift to Student-t copula with 3 degrees of freedom.

Figure B.56: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Frank copula and shifts to a different copula.
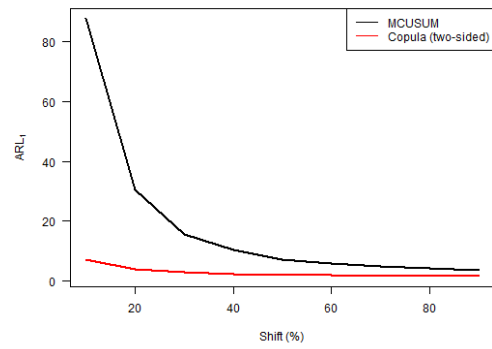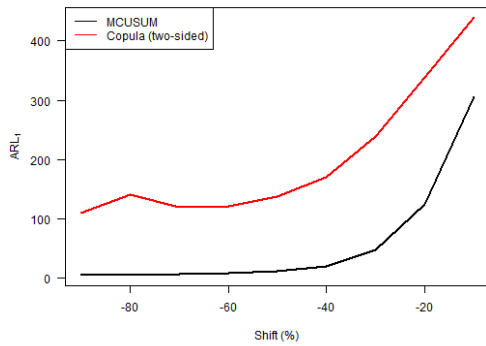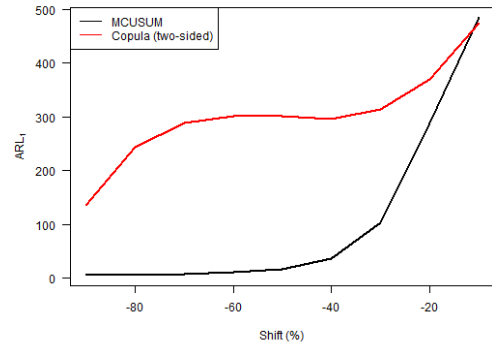
Student-t (case 225 & cases 266 − 275)

Figure B.57: ARL_0 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Student-t copula.

(a) Level of the first variable increased.

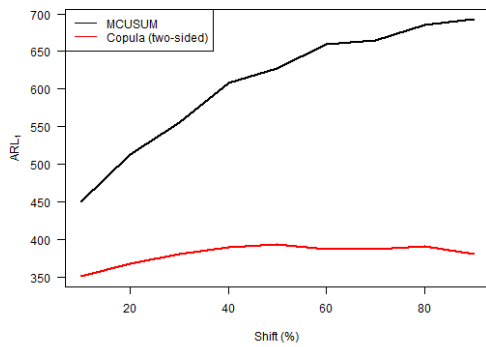(b) Level of both variables increased.
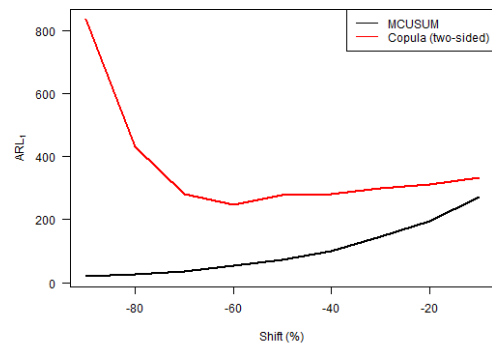

(c) Level of the first variable decreased.

(d) Level of both variables decreased.

Figure B.58: ARL_1 vs. level shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Student-t copula.
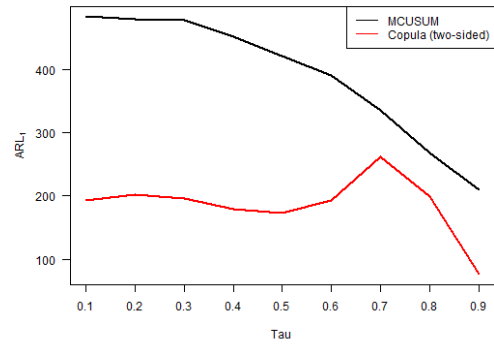

(a) Largest eigenvalue increased.

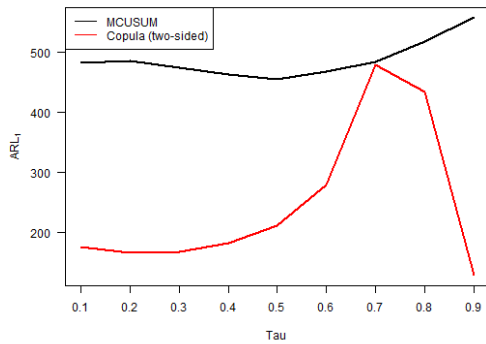(b) Largest eigenvalue decreased.

Figure B.59: ARL_1 vs. largest eigenvalue shift for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Student-t copula.
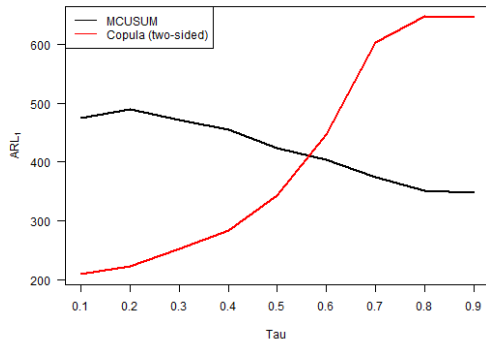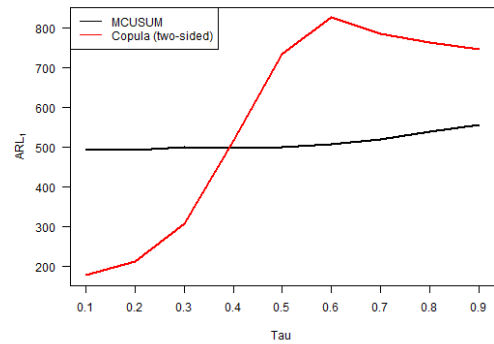
(a) Shift to Gumbel copula.

(b) Shift to Clayton copula.

(c) Shift to Frank copula.

(d) Shift to Normal copula.

Figure B.60: ARL_1 vs. Kendall's τ for the two-sided version of the copula-based control chart and the MCUSUM chart with $k = 0.75$, where the in-control distribution follows a Student-t copula and shifts to a different copula.