Fast inference for high-dimensional one-factor copula models and methods for multivariate change-point detection with applications to stock market data

Alex Verhoijsen

Fast inference for high-dimensional one-factor copula models and methods for multivariate change-point detection with applications to stock market data

Alex Verhoijsen

ORCID Identifier: 0000-0003-0649-2280

Thesis submitted in total fulfillment of the requirements of the joint degree of Doctor of Philosophy (PhD) of

The University of Melbourne
School of Mathematics and Statistics

Université de Pau et des Pays de l'Adour
Laboratoire de Mathémathiques et de Leurs Applications

June 6, 2024

Committee Members

Pavel Krupskiy	Senior Lecturer (University of Melbourne)	Supervisor
Ivan Kojadinovic	Professor (Université de Pau)	Supervisor
Elif Acar	Associate Professor (University of Manitoba)	Reviewer
Christian Paroissin	Maître de Conférences (Université de Pau)	Reviewer
Marcy Robertson	Associate Professor (University of Melbourne)	Chair
Ostap Okhrin	Professor (Technische Universität Dresden)	Examiner
Jean-François Quessy	Professor (Université du Québec à Trois-Rivières)	Examiner

Fast inference for high-dimensional one-factor copula models and methods for multivariate change-point detection with applications to stock market data

Alex Verhoijsen

Abstract

Copula models are often used to model dependence between random variables in a multivariate, low-dimensional, setting. However, practical problems arise when the dimension becomes very large. A different problem, when working with multivariate time series, is the assumption of stationarity, which is often violated, and adequate methods are required to detect departures from stationary.
We begin the thesis by introducing the technical tools related to copula models and change-point detection that are used in the remainder of this work.
The first original contribution of the thesis is a fast inference methods for high-dimensional one-factor copula models. We propose a model with Gaussian factors and residual dependence modelled using a one-factor copula model. The corresponding estimation procedure allows the statistician to identify the factor model parameters, the Gaussian factors, the factor in the one-factor copula model, and the one-factor copula parameter. Asymptotic properties for a model with up to three Gaussian factors are established, and finite sample properties are illustrated using Monte Carlo simulations. A practical application using SP500 stock market data demonstrates how the proposed model can be used to capture dependence in real-world data.
The second original contribution is a nonparametric sequential open-end change-point detection scheme using the empirical distribution function of possibly multivariate observations. We establish the asymptotic properties of the detector, and perform large-scale Monte Carlo simulations in a univariate and multivariate, low-dimensional, setting. To illustrate the procedure, we conclude with a real-world application by monitoring for changes in the log-returns of the NASDAQ composite index. The code used to implement the monitoring procedure is included in the R package `npcp`.

Inférence rapide pour des modèles de copules à un facteur en grande dimension et méthodes de détection de rupture multivariées avec applications aux données de marchés financiers

Alex Verhoijsen

Résumé

Les modèles à base de copules sont souvent utilisés pour modéliser la dépendance entre variables lorsque la dimension des données est faible. Lorsque la dimension augmente, de nombreux problèmes pratiques apparaissent malheureusement. Dans le cas où les données sont des séries temporelles, il est en plus nécessaire de tester leur stationnarité préalablement à toute modélisation, ce qui nécessite des méthodes adéquates capables de détecter des écarts à la stationnarité.
Nous commençons ce document par une introduction aux outils mathématiques nécessaires à la compréhension des modèles à base de copules et des méthodes de détection de rupture.
La première contribution de ce travail est la proposition d'une méthode d'inférence rapide pour des modèles de copules à un facteur en grande dimension. Nous proposons un modèle à un facteur dont la dépendance résiduelle est modélisée par un modèle de copules à un facteur. La méthode d'estimation associée permet d'identifier les facteurs Gaussiens sous-jacents, le facteur du modèle de copules ainsi que le paramètre de la copule paramétrique correspondante. Les propriétés asymptotiques pour un modèle avec un, deux ou trois facteurs Gaussiens sont établies et sont complétées par des simulations de Monte Carlo dans le cas d'échantillons de tailles finies. Une application à des données boursières liées aux SP500 montre comment le modèle proposé peut être utilisé pour modéliser la dépendance entre un grand nombre de variables dans des situations réalistes.
La deuxième contribution de ce travail est la proposition d'une méthode de surveillance non paramétrique n'ayant pas un horizon de surveillance fini et permettant de détecter des changements dans la distribution de données multivariées à l'aide de la fonction de répartition empirique. Nous établissons les propriétés asymptotiques du détecteur sous-jacent et nous présentons de nombreuses simulations de Monte Carlo en faible dimension. Nous illustrons l'intérêt de la procédure par une application à la surveillance de retours logarithmiques du NASDAQ composite. Le code implémentant la procédure est inclus dans le package `R` `npcp`.

Declaration

This is to certify that:

i. the thesis comprises only my original work towards the PhD except where indicated in the preface;

ii. due acknowledgement has been made in the text to all other material used; and

iii. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed,

Alex Verhoijsen

June 6, 2024

Acknowledgements

Doing a PhD is never a one-person endeavour. Doing a joint PhD at two universities in two different continents during a worldwide, once in a century pandemic makes the journey a bit more complicated. I would like to express my gratitude to a number of people who encouraged me to do, to continue doing, and to finish this PhD.

To begin with, I would like to thank my main supervisors Pavel Krupskiy and Ivan Kojadinovic for all the support they provided during my PhD journey. In Melbourne, Pavel was always available to provide advice and answer my many questions with great patience. He gave me freedom while encouraging me to broaden my knowledge in the field of multivariate statistics, and supported me with all administration-related tasks. During the third year of my PhD I visited the University of Pau, where Ivan, with great patience and availability, taught me how to do rigorous research in the field of nonparametric statistics, and kept encouraging me, even when I felt completely lost. In addition, while writing code to include in the R, Ivan taught me how to become a better statistical programmer.

I would like to thank Jean-François Quessy and Ostap Okhrin for carefully reviewing my thesis and their supportive comments and suggestions, which greatly improved the quality of the thesis. Furthermore I would also like to thank Howard Bondell, for his work as chair of my PhD committee, Mark Holmes for agreeing to co-supervising my PhD, and Elif Acar, Christian Paroissin, and Marcy Robertson for their work as committee members during the examination stage of my PhD. I am grateful to the University of Melbourne for providing financial support in the form of the Melbourne Research Scholarship and the Top-up scholarship in mathematics and statistics.

Last but not least I want to thank the people closest to me, and to whom I would like to dedicate this thesis. My parents gave me all the opportunities which allowed me to do this PhD. My parents and Adrian they provided their unlimited support during this journey, for which I will be forever grateful. I want to thank Nithila, who has seen the daily highs, lows, and everything in between, from the early COVID lockdown days to long after. You were there for me every step of the way to celebrate the wins, and help me during setbacks, for which I can never thank you enough. I hope to one day repay you all the love and patience you have shown me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

To evaluate the relationship between the components of a pair of random variables (X, Y), students in statistics are often first introduced to the Pearson correlation coefficient $\rho(X,Y) = \mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$, which is widely used in statistical applications. Nevertheless, Pearson's correlation coefficient suffers from some deficiencies (Embrechts et al., 2002). One disadvantage is that it does not always capture the entire dependence structure between random variables. An example of this is when one has at hand a bivariate random vector (X, Y), where $X \sim N(0,1)$, and $Y = X^2$. Even though Y is clearly dependent on X, one can show that $\rho(X,Y) = 0$.

A different approach to model dependence between random variables is by the use of *copulas*. Recall that the distribution function (d.f.) F of a vector of random variables $\boldsymbol{X} = (X_1, \ldots, X_d)$ is

$$F(\boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d), \tag{1.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d) \in [-\infty, \infty]^d$. Let $U_j = F_j(X_j)$, where F_j is a d.f., for $j \in \{1, \ldots, d\}$. A copula C is the d.f. of a vector of random variables $\boldsymbol{U} = (U_1, \ldots, U_d)$, with $U_j \sim U(0,1)$, for $j \in \{1, \ldots, d\}$.

The seminal work of Sklar (1959) shows that for a d-variate d.f. F and univariate margins $F_1, \ldots, F_d$, a copula C binds together the margins into their joint distribution and thus captures the entire dependence structure between random variables. If the margins are continuous, the copula will be unique, and in what follows in the rest of the thesis we assume that this is the case. Vice-versa, Sklar's theorem states that a d-variate joint d.f. F can be decomposed into a copula C and its d univariate margins $F_1, \ldots, F_d$. Copulas can be used in practical applications to obtain an estimate of the joint d.f. F by first separately estimating the univariate d.f.s $F_1, \ldots, F_d$, before estimating the copula C. This approach is useful in statistical applications such as in economics (Patton, 2012), in finance (Cherubini et al., 2004; McNeil et al., 2005), and in hydrology (Subbarao et al., 1996). For an overview of the rise in popularity of copulas in the finance literature, see Genest et al. (2009).

Even though classical copula models provide a straightforward method to model dependence between random variables, problems with estimation of the parameters arise when the dimension becomes too large. To address this, Krupskii and Joe (2013) propose a factor copula model in which dependence between random variables is caused by common, unobserved, factors. For example, in finance, one can assume that a change in the underlying economic situation affects stock prices. This factor

copula approach was extended in Krupskii and Joe (2015) to allow for a hierarchical structure, and Krupskii and Joe (2022) proposed an estimation procedure for factor copula models with non-overlapping groups.

The first contribution of the thesis in Chapter 3 extends existing inference methods for one-factor copula models. We design a new inference method that allows for fast computation of the model parameters in a very high-dimensional one-factor copula model, and use this approach to model the dependence between stock prices in different industry sectors. Chapter 3 and the corresponding Appendix A are based on Verhoijsen and Krupskiy (2022), which has been published in *Dependence Modeling*.

Assume one has at hand a sample of a stationary time series $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ of size m, called the learning sample, and a monitoring sample $\boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots$, where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$ is a random vector and $i = 1, 2, \ldots$. The idea of sequential change-point detection is to monitor a time series for departures from stationarity by comparing a detector to a threshold. In sequential change-point detection, observations $\boldsymbol{X}_k$, where $k > m$, arrive one at a time, and at each step the detector is calculated and compared to the threshold, with the goal of identifying departures from stationarity "as quickly as possible". If the null hypothesis of stationarity is not rejected, a new observation is collected and the monitoring procedure starts again from the beginning. In open-end sequential change-point detection monitoring can in theory continue indefinitely as long as the hypothesis of stationarity is not rejected, while in closed-end change-point detection, monitoring eventually stops after the arrival of observation $\boldsymbol{X}_n$, with $n > m$.

To monitor for changes in the d.f. at any point $k^\star > m$, one can verify if

$$\mathbb{P}\left(\boldsymbol{X}_1 \leq \boldsymbol{x}\right) = \cdots = \mathbb{P}\left(\boldsymbol{X}_{k^\star} \leq \boldsymbol{x}\right) \neq \mathbb{P}\left(\boldsymbol{X}_{k^\star+1} \leq \boldsymbol{x}\right) = \mathbb{P}\left(\boldsymbol{X}_{k^\star+2} \leq \boldsymbol{x}\right) = \ldots,$$

where $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The second contribution of the thesis can be found in Chapter 4 and the corresponding Appendix B, and proposes an open-end nonparametric sequential change-point detection scheme to monitor for changes in the empirical d.f. of a possibly multivariate dataset. Changes in the joint d.f. of a multivariate dataset can be the result of changes in the margins, or of changes in the dependence structure between the random variables. The monitoring scheme is illustrated with an application to stock market data. The results in this chapter have been published in the *Journal of Time Series Analysis* (Holmes et al., 2023), and the corresponding R code has been included in the package `npcp` (Kojadinovic and Verhoijsen, 2022) available on CRAN.

The rest of the thesis is organised as follows. In Chapter 2 multivariate models in the form of Gaussian factor models and (factor) copula models, along with their estimation methods are introduced. Furthermore, we explain the different approaches to change-point detection. These techniques are necessary for the rest of the thesis. The first original work of this thesis is in Chapter 3, which proposes a fast inference method for high-dimensional one-factor copula models. Chapter 4 proposes an open-end nonparametric sequential change-point detection method for changes in the empirical d.f. of possibly multivariate observations. Chapter 5 discusses ongoing investigations, summarises the work in the main body of the thesis, and concludes with directions for future research. To improve the flow of the main text, proofs and additional materials for Chapters 3 and 4 have been moved to Appendix A and B, respectively.

Chapter 2

Multivariate models and change-point detection

In this chapter, the main tools used in the thesis are presented. To begin with, techniques to model dependence in multivariate data by means of Gaussian factor models are introduced. While factor models are easy to understand from a conceptual perspective, their practical use is limited due to the assumption of Gaussianity. Copula models provide an elegant way of capturing a more general dependence structure in multivariate data, and allow the statistician to separately model the margins and the dependence structure. Basic copula concepts are introduced, as well as measures of association and one-factor copulas. One-factor copula models assume the existence of a common, unobserved, factor as the source of dependence between observable variables, allowing for more parsimonious models than standard copula models. Parametric inference methods for copula models are discussed both when using parametric and nonparametric models for the margins. Next, nonparametric inference methods by means of the empirical copula process are outlined.

The distribution function of a vector of random variables is not necessarily stable over time. Change-point detection techniques can be used to monitor for changes in the distribution function, and the goal is to identify departures from stationary in the underlying time series. We discuss the different approaches to change-point detection and its practical relevance.

In what follows, for $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{y} = (y_1, \ldots, y_d)$, two vectors in $\mathbb{R}^d$, the notation for the vector inequality $\boldsymbol{x} \leq \boldsymbol{y}$ implies that each of the componentwise inequalities $x_j \leq y_j$ holds, for $j \in \{1, \ldots, d\}$. That is,

$$\boldsymbol{x} \leq \boldsymbol{y} \iff x_j \leq y_j, \quad j \in \{1, \ldots, d\}\,.$$

Furthermore, let $\boldsymbol{x}^{(j)} = (1, \ldots, 1, x_j, 1, \ldots, 1)$ be the vector $\boldsymbol{x} = (x_1, \ldots, x_d)$ where $x_i = 1$, for $i \in \{1, \ldots, d\} \setminus j$. Lastly, let $\overline{\mathbb{R}} = [-\infty, \infty]$, and let the componentwise minima between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ be given by $\boldsymbol{x} \wedge \boldsymbol{y} = (\min(x_1, y_1), \ldots, (x_d, y_d))$.

2.1 Gaussian factor models

Assume one has at hand a vector of d observable random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Let

$$\boldsymbol{X} = \boldsymbol{V}^{-\frac{1}{2}}(\boldsymbol{Y} - \boldsymbol{\mu})$$

be the standardised version of $\boldsymbol{Y}$, where $\boldsymbol{V}^{-\frac{1}{2}} = \operatorname{diag}(1/\sigma_1, \ldots, 1/\sigma_d)$ is a diagonal matrix, with $\sigma_j^2 = \operatorname{Var}(Y_j)$, for $j \in \{1, \ldots, d\}$. The goal of factor analysis is to model the dependence between the elements of the vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ as a linear combination of an unobservable standard Gaussian vector of p factors $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top$, and an unobservable standard Gaussian vector of d *error terms* $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)$, where in practical applications, p is smaller then d, and $\operatorname{Cov}(\boldsymbol{Z}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$. For $j \in \{1, \ldots, d\}$, the Gaussian p-factor model is

$$X_j = \sum_{k=1}^{p} \lambda_{jk} Z_k + \gamma_j \varepsilon_j. \tag{2.1}$$

The observed variables X_j load heterogeneously on the unobserved common factors Z_k via the *factor loadings* λ_{jk}, where $Z_k, \varepsilon_j \overset{iid}{\sim} N(0,1)$, and $\gamma_j = \sqrt{1 - \sum_{k=1}^{p} \lambda_{jk}^2}$ such that $\operatorname{Var}(X_j) = 1$, for $j \in \{1, \ldots, d\}$ and $k \in \{1, \ldots, p\}$. Casting (2.1) into matrix notation yields

$$\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{Z} + \boldsymbol{\Gamma}^{\frac{1}{2}}\boldsymbol{\varepsilon}, \tag{2.2}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$ is the $d \times p$ matrix of factor loadings of p column vectors $\boldsymbol{\lambda}_k = (\lambda_{1k}, \ldots, \lambda_{dk})^\top$, and $\boldsymbol{\Gamma}^{\frac{1}{2}} = \operatorname{diag}(\gamma_1, \ldots, \gamma_d)$. An illustration of a one-factor model is given in Figure 2.1.

Recall that the common factors $\boldsymbol{Z}$ and the error terms $\boldsymbol{\varepsilon}$ are uncorrelated standard Gaussian vectors, that is $\operatorname{Cov}(\boldsymbol{Z}) = I_p$, $\operatorname{Cov}(\boldsymbol{\varepsilon}) = I_d$, and $\operatorname{Cov}(\boldsymbol{Z}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$. Then the variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ of $\boldsymbol{X}$ in (2.2) is

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Gamma}. \tag{2.3}$$

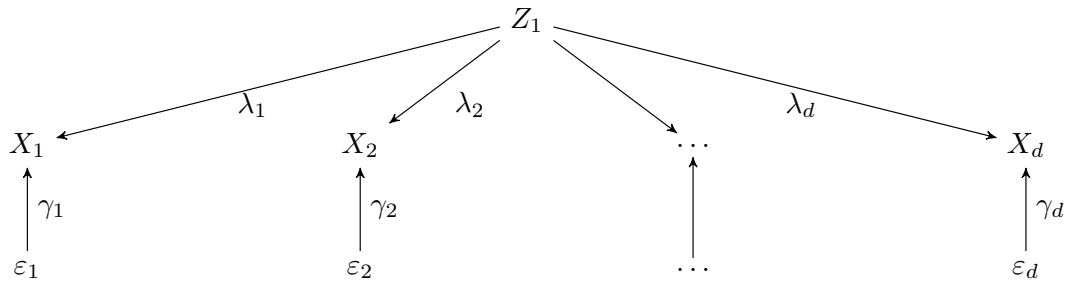


Figure 2.1: Gaussian one-factor model with observable variables $X_1, \ldots, X_d$, common factor Z_1, specific errors $\varepsilon_1, \ldots, \varepsilon_d$, factor loadings $\lambda_1, \ldots, \lambda_d$, and $\gamma_j = \sqrt{1 - \lambda_j^2}$, for $j \in \{1, \ldots, d\}$.

For each $j \in \{1, \ldots, d\}$, the proportion of variance of X_j explained by the p common factors is also known as the jth *communality* $\sum_{k=1}^{p} \lambda_{jk}^2$, while the proportion of the variance of X_j resulting from the error term is called the *uniqueness* or *specific variance* γ_j^2 (Johnson and Wichern, 2002; Tsay, 2005; Härdle et al., 2007). Here, one can clearly see that factor models are particularly useful when p is (much) smaller than d, since the $d(d+1)/2$ elements of $\mathbf{\Sigma}_{\boldsymbol{X}}$ can be described by the pd model parameters in $\mathbf{\Lambda}$, as the remaining parameters in $\mathbf{\Gamma}$ are defined by $\gamma_j = \sqrt{1 - \sum_{k=1}^{p} \lambda_{jk}^2}$, for $j \in \{1, \ldots, d\}$.

A drawback of the Gaussian factor model is that the factor loadings cannot be uniquely identified using (2.3) without imposing $p(p-1)/2$ additional restrictions on $\mathbf{\Lambda}$. To see this, let $\mathbf{A}$ be a $p \times p$ orthogonal matrix such that $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top \mathbf{A} = I_p$, and let $\boldsymbol{Z}' = \mathbf{A}^\top \boldsymbol{Z}$, and $\mathbf{\Lambda}' = \mathbf{\Lambda}\mathbf{A}$. Then (2.2) can be rewritten as

$$\boldsymbol{X} = \mathbf{\Lambda}' \boldsymbol{Z}' + \boldsymbol{\varepsilon},$$

and (2.3) becomes

$$\mathbf{\Sigma}_{\boldsymbol{X}} = \mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Gamma} = \mathbf{\Lambda}' \left(\mathbf{\Lambda}'\right)^\top + \mathbf{\Gamma}.$$

This phenomenon is called *factor rotation*, as factor loadings can only be identified up to an orthogonal matrix. Since $\mathbf{\Lambda}\mathbf{\Lambda}^\top = \mathbf{\Lambda}' \left(\mathbf{\Lambda}'\right)^\top$, the communalities are not affected by a rotation.

When the common factors $\boldsymbol{Z}$ and the idiosyncratic errors $\boldsymbol{\varepsilon}$ are multivariate Gaussian, estimates of the model parameters $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ can be obtained by maximising the log-likelihood

$$L_n\left(\mathbf{\Lambda}, \mathbf{\Gamma}\right) = -\frac{n}{2} \left\{ \ln \left| 2\pi \left(\mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Gamma} \right) \right| + \operatorname{tr}\left(\left(\mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Gamma} \right)^{-1} \mathbf{\Sigma}_{\boldsymbol{X}_n} \right) \right\}, \tag{2.4}$$

with respect to the unknown model parameters $\mathbf{\Lambda}$, and $\mathbf{\Gamma}$, where $\operatorname{tr}(\cdot)$ is the trace of a matrix, and $\mathbf{\Sigma}_{\boldsymbol{X}_n}$ is the sample correlation matrix. The log-likelihood in (2.4) requires a numerical solution, for which a practical computation scheme can be found in Supplement 9A of Johnson and Wichern (2002).

After obtaining estimates for the factor loadings and the uniqueness, it can be of interest to the statistician to obtain estimates of the common factors $\boldsymbol{Z}$, called

factor scores. In (2.2), one can treat $\boldsymbol{\varepsilon}$ as error terms with heteroskedastic variance $\text{Var}(\varepsilon_j) = \gamma_j^2$, for $j \in \{1, \ldots, d\}$. Using a weighted least squares approach, the sum of squared error terms weighted by their respective variances

$$\boldsymbol{\varepsilon}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\varepsilon},$$

can be minimised to obtain the factor scores

$$\hat{\boldsymbol{Z}} = \left(\boldsymbol{\Lambda}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}\right)^{-1} \boldsymbol{\Lambda}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{X}. \tag{2.5}$$

Gaussian factor models provide a straightforward approach to model dependence in multivariate data due to the limited set of assumptions, which is why they are often used in applied statistics, such as in econometrics, (Bai and Ng, 2002, 2008; Stock and Watson, 2002; Pesaran, 2006), in the finance literature, (McNeil et al., 2005; Tsay, 2005), and in hydrology (Subbarao et al., 1996).

Nevertheless, the assumption of multivariate Gaussianity is a major limitation of factor models. In the next section, *copula models* are introduced as a means of capturing a more general dependence structure in a multivariate dataset.

2.2 Copula models

Gaussian factor models as introduced in Section 2.1 provide a straightforward method to model dependence between random variables. Nevertheless, the assumption of Gaussianity is often violated in practice, thus limiting the usefulness of factor models in practical applications. In this section, a more flexible model of multivariate dependence is presented by means of *copulas*. Copulas and their main properties are introduced, before delving deeper into measures of association, one-factor copulas, and parametric and nonparametric inference methods for (one-factor) copula models. Modelling multivariate data with copulas allows one to separate the margins from the dependence structure, which is established by Sklar's seminal theorem (Sklar, 1959), and explains the popularity of copulas in a wide range of research fields, such as in finance, see Cherubini et al. (2004), Chapter 5 in McNeil et al. (2005), in economics (Patton, 2012), and in hydrology (Subbarao et al., 1996). Since Gaussianity usually is a strong assumption in the finance literature, copulas provide a natural extension to allow for heavy tails and skewness in the data. A more exhaustive overview of copulas and detailed proofs of theorems, lemmas, and propositions in this sections can be found in Chapter 5 in McNeil et al. (2005), Nelsen (2006), Joe (2015), Embrechts and Hofert (2013), and Hofert et al. (2018). Lastly, we refer the interested reader to additional work by Schreyer et al. (2017), and Genest (2021).

2.2.1 Distribution functions

To understand copulas as presented in forthcoming Section 2.2.2, properties of distribution functions, generalised inverse functions, quantile transforms, and probability integral transforms are introduced.

Copulas rely on the distribution function (df) of a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$.

Definition 2.2.1 (Distribution function)**.** *The* distribution function *of a d-dimensional random vector* $\boldsymbol{X} = (X_1, \ldots, X_d)$ *is the function*

$$F(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}), \quad \boldsymbol{x} = (x_1, \ldots, x_d) \in \overline{\mathbb{R}}^d.$$

The marginal dfs *or* margins *of* F *are the univariate dfs* $F_1, \ldots, F_d$ *given by*

$$F_j(x_j) = F(\infty, \ldots, x_j, \ldots, \infty) = \mathbb{P}(X_j \leq x_j), \quad x_j \in \mathbb{R},$$

for $j \in \{1, \ldots, d\}$.

The *generalised inverse function* (gif) carefully formulates how to obtain the inverse of a df.

Definition 2.2.2 (Generalised inverse function)**.** *The generalised inverse function of a right-continuous univariate df,* F*, is defined as*

$$F^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \quad u \in (0, 1),$$

with the convention that $\inf \emptyset = \infty$.

This implies that the generalised inverse function is always left-continuous. If F is continuous and strictly increasing, the gif $F^{\leftarrow}$ will coincide with F^{-1}, the *quantile function* of F. For a detailed discussion of the properties of gifs, see Embrechts and Hofert (2013). Using the *quantile transform*, a standard uniform random variable can be transformed into a random variable with arbitrary df F.

Proposition 2.2.3 (Quantile transform)**.** *Let* F *be a univariate df and let* U *be a standard uniform random variable. Then* $\mathbb{P}(F^{\leftarrow}(U) \leq x) = F(x)$.

On the other hand, starting from a random variable X with arbitrary continuous df F, the *probability integral transform* (pit) can be used to obtain a standard uniformly distributed random variable.

Proposition 2.2.4 (Probability integral transform)**.** *Let* X *be a random variable with continuous df* F*. Then* $F(F^{\leftarrow}(u)) = u$ *for all* $u \in [0, 1]$*, and as a consequence, the distribution of* $F(X)$ *is uniform on* $[0, 1]$*, that is*

$$\mathbb{P}(F(X) \leq u) = u, \quad u \in [0, 1].$$

Proofs of Proposition 2.2.3 and Proposition 2.2.4 can be found in the proof of Proposition 5.2 in McNeil et al. (2005).

2.2.2 Copula definition and properties

The following definition of copulas, which relies on Definition 2.2.1 of the distribution function, will be key in the rest of this work.

Definition 2.2.5 (Copula)**.** *A d-dimensional* copula $C : [0, 1]^d \to [0, 1]$ *is a multivariate df of* $\boldsymbol{U} = (U_1, ..., U_d)$ *whose margins are standard uniform dfs:*

$$C(\boldsymbol{u}) = \mathbb{P}(\boldsymbol{U} \leq \boldsymbol{u}), \quad \boldsymbol{u} \in [0, 1]^d, \tag{2.6}$$

and

$$\mathbb{P}(U_j \leq u_j) = u_j, \quad u_j \in [0, 1],$$

for $j \in \{1, \ldots, d\}$.

In other words, the copula C is a mapping from the unit hypercube into the unit interval. The following properties must hold for a copula C.

Proposition 2.2.6 (Copula properties)**.** *Let C be a copula, $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$ and $j \in \{1, \ldots, d\}$. Then,*

(i) if some component u_j of $\boldsymbol{u}$ is 0, then $C(\boldsymbol{u}) = 0$,

(ii) $C\left(\boldsymbol{u}^{(j)}\right) = u_j$, if $u_j \in [0,1]$,

(iii) for all $(a_1, \ldots, a_d) \in [0,1]^d$, and $(b_1, \ldots, b_d) \in [0,1]^d$, with $a_j \leq b_j$,

$$0 \leq \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C\left(u_{1i_1}, \ldots, u_{di_d}\right), \tag{2.7}$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$,

(iv) C is Lipschitz and hence continuous: for $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$,

$$|C(\boldsymbol{u}) - C(\boldsymbol{v})| \leq \sum_{j=1}^{d} |u_j - v_j|.$$

Proposition 2.2.6(i) is required for any df, while Proposition 2.2.6(ii) follows from the fact that the margins are standard uniformly distributed. A proof of Proposition 2.2.6(iv) can be found in the proof of Lemma 2.1.5 of Nelsen (2006). For $d = 2$, (2.7) becomes $C(u_{12}, u_{22}) - C(u_{12}, u_{21}) - C(u_{11}, u_{22}) + C(u_{11}, u_{21}) \geq 0$, which is akin to stating that $\mathbb{P}(a_1 \leq U_1 \leq b_1, a_2 \leq U_2 \leq b_2) \geq 0$, or in the general case that $\mathbb{P}(a_1 \leq U_1 \leq b_1, \ldots, a_d \leq U_d \leq b_d) \geq 0$.

In Definition 2.2.5, the copula is formally introduced. The seminal theorem of Sklar (1959) shows the usefulness of copulas in practical applications.

Theorem 2.2.7 (Sklar, 1959)**.** *Let C be a d-dimensional copula and let $F_1, ..., F_d$ be univariate dfs. Then the function*

$$F(\boldsymbol{x}) = F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)) \tag{2.8}$$

is a d-dimensional df with margins $F_1, \ldots, F_d$.

The previous theorem shows one can generate a joint distribution F by coupling together univariate margins $F_1, \ldots, F_d$ using a copula C.

Theorem 2.2.8 (Sklar, 1959)**.** *If F is a d-dimensional df with univariate dfs $F_1, \ldots, F_d$, then there exists a copula C such that (2.8) holds. If the margins $F_1, \ldots, F_d$ are continuous, C is unique and equal to*

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d)), \; u_j \in [0,1].$$

In what follows in this thesis, we exclusively work with continuous margins and the copulas used hereafter will be unique. When the margins are not continuous, C is uniquely defined on $\operatorname{Ran} F_1, \ldots, \operatorname{Ran} F_d$, where $\operatorname{Ran} F_j = F_j(\overline{\mathbb{R}})$ is the range of F_j, for $j \in \{1, \ldots, d\}$. A proof of Sklar's theorem can be found in the proof of Theorem

5.3 in McNeil et al. (2005), in the proof of Theorem 2.3.3 in Nelsen (2006), or in the proof of Theorem 1.1 in Joe (2015).

The following examples introduce a selection of copula families. The properties of each copula family are illustrated with a bivariate sample in Figure 2.2 for which the R package `copula` was used (R Core Team, 2022). As will be outlined in the forthcoming Section 2.2.3, all copula parameters are chosen such that their Kendall's $\tau = 0.7$.

Example 2.2.9 (Independence copula). A pair of random variables, (X_1, X_2) is independent if and only if its copula is the independence copula C_Π, where

$$C_\Pi(u_1, u_2) = u_1 u_2.$$

A sample of size $n = 1000$ from the bivariate independence copula is depicted in the upper left panel of Figure 2.2. As expected, the sampled cloud of points from the independence copula is uniformly spread out over the unit square.

Example 2.2.10 (Gaussian copula). Recall the Gaussian factor model in (2.1). Assume $p = 1$ such that

$$X_j = \lambda_j Z + \gamma_j \varepsilon_j,$$

then the d-variate copula of $\boldsymbol{X} = (X_1, \ldots, X_d)$ implicitly generated by this Gaussian one-factor model is the Gaussian copula C_{Gau} with $d \times d$ correlation matrix $\boldsymbol{\Sigma}$. In other words, it is the copula of the d-variate Gaussian distribution $N_d(\mathbf{0}, \boldsymbol{\Sigma})$, for which no closed-form expression exists:

$$C_{\text{Gau}}(\boldsymbol{u}) = \Phi_{\boldsymbol{\Sigma}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right),$$

where $\Phi_{\boldsymbol{\Sigma}}$ is the d-variate Normal df with correlation matrix $\boldsymbol{\Sigma}$, and Φ^{-1} is the quantile function of the univariate Normal df. The correlation matrix $\boldsymbol{\Sigma}$ has off-diagonal elements $\rho_{jk} = \lambda_j \lambda_k$ and diagonal elements $\rho_{jj} = 1$, for $j, k \in \{1, \ldots, d\}$, and $j \neq k$.

A sample of size $n = 1000$ from the bivariate Gaussian copula with dependence parameter $\rho = 0.89$ is depicted in the upper right panel of Figure 2.2. The cloud of points from the bivariate Gaussian copula is symmetric, and there is strong dependence between the variables.

Example 2.2.11 (Clayton copula). For n independent and identically distributed (iid) random variables $X_1, \ldots, X_n$ with continuous df F, the copula of the random pair $(\max(X_1, \ldots, X_n), -\min(X_1, \ldots, X_n))$ is the bivariate Clayton copula with parameter $\theta = -1/n$:

$$C_{\text{Cl}}(u_1, u_2) = \begin{cases} \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, & \theta > 0 \\ u_1 u_2, & \theta = 0, \\ \max\left(u_1^{-\theta} + u_2^{-\theta} - 1, 0\right)^{-1/\theta}, & -1 \leq \theta < 0. \end{cases}$$

The parameter θ captures the strength of dependence between the random variables. A sample of size $n = 1000$ from a bivariate Clayton copula with dependence parameter $\theta = 4.67$ is represented in the lower left panel of Figure 2.2. Compared to the symmetric shape of the sample from the bivariate Gaussian copula, the cloud of points sampled from the Clayton copula exhibits dependence in the lower tail and no dependence in the upper tail.

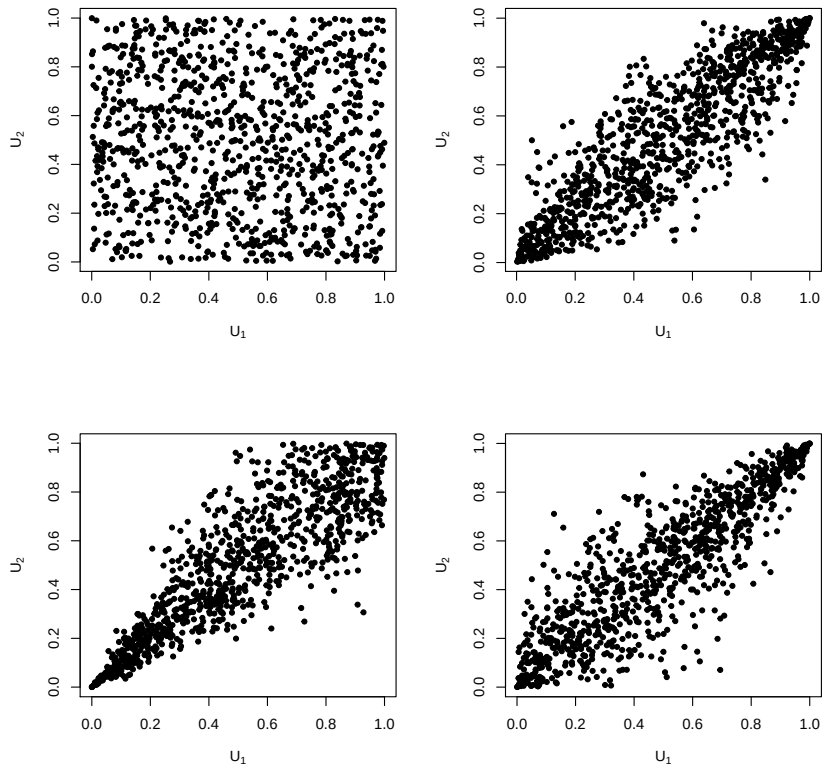


Figure 2.2: Samples of size $n = 1000$ from bivariate copula models. Top left: independence copula (Kendall's $\tau = 0$). Top right: Normal copula (Kendall's $\tau = 0.7$). Bottom left: Clayton copula (Kendall's $\tau = 0.7$). Bottom right: Gumbel copula (Kendall's $\tau = 0.7$).

Example 2.2.12 (Gumbel copula). The bivariate Gumbel copula is given by

$$C_{\mathrm{Gu}}(u_1, u_2) = \exp\left[-\left\{(-\log u_1)^\theta + (-\log u_2)^\theta\right\}^{1/\theta}\right], \quad \theta \geq 1.$$

A sample of size $n = 1000$ from a bivariate Gumbel copula with $\theta = 3.33$ is depicted in the lower right panel of Figure 2.2. In contrast to the sample from the Clayton copula, the cloud of points sampled from the bivariate Gumbel copula is characterised by dependence in the upper tail and lack of dependence in the lower tail.

The copula density is key in parametric estimation of the copula by means of the maximum likelihood, as discussed in the forthcoming Section 2.3.1.

Proposition 2.2.13 (Copula density I)**.** *If C is d times continuously differentiable on $(0,1)^d$, then C has density c given by*

$$c(\boldsymbol{u}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(\boldsymbol{u}), \quad \boldsymbol{u} \in (0,1)^d.$$

If C and $F_1, \ldots, F_d$ are absolutely continuous, the copula density can also be recovered from (2.8)*. Let F be a d-variate df with margins $F_1, \ldots, F_d$ and copula C. If*

the margins have densities $f_1, \ldots, f_d$ and if the copula C has a density c, then F admits a joint density

$$f(\boldsymbol{x}) = c\left(F_1(x_1), \ldots, F_d(x_d)\right) f_1(x_1) \cdots f_d(x_d),$$

where $\boldsymbol{x} \in \mathbb{R}$. Furthermore, the copula density can be found by

$$c\left(\boldsymbol{u}\right) = \frac{f\left(F_1^{\leftarrow}\left(u_1\right), \ldots, F_d^{\leftarrow}\left(u_d\right)\right)}{f_1(F_1^{\leftarrow}\left(u_1\right)) \cdots f_d(F_d^{\leftarrow}\left(u_d\right))},$$

where $\boldsymbol{u} \in (0,1)^d$.

The conditional copula is of paramount importance for the definition of one-factor copulas in Section 2.2.4. First, the density of a standard uniform variable is constant and equal to 1. As a result, the conditional density of $\boldsymbol{U}_{-j} = (U_i : i \in \{1, \ldots, j-1, j+1, \ldots, d\})$ given $U_j = u_j$ is c itself:

$$c_{\boldsymbol{U}_{-j}|U_j}(\boldsymbol{u}_{-j}|u_j) = c(\boldsymbol{u}), \quad \boldsymbol{u} \in (0,1)^d.$$

Then, the df of the conditional distribution of $\boldsymbol{U}_{-j}$ given $U_j = u_j$ is

$$\mathbb{P}\left(\boldsymbol{U}_{-j} \leq \boldsymbol{u}_{-j} | U_j = u_j\right) = \frac{\partial}{\partial u_j} C(\boldsymbol{u}) = \dot{C}_j(\boldsymbol{u}).$$

While copulas are a general way to describe dependence in multivariate data, it can be useful to capture association between random variables using a real number, by means of measures of association.

2.2.3 Measures of association

In this section several measures of association are presented as a means to summarise dependence between random variables, where different measures of association capture different characteristics of the data. For some copulas there is a one-on-one relation between the parameter of the copula model and measures of association. The first two measures of association are rank-based, and provide an alternative to the popular *Pearson correlation*, which has some well-known drawbacks as discussed in Embrechts et al. (2002).

Consider two points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$. The points are said to be *concordant* if $(x_2 - x_1)(y_2 - y_1) > 0$ and *discordant* if $(x_2 - x_1)(y_2 - y_1) < 0$.

Definition 2.2.14 (Kendall's tau)**.** *Let F be a continuous bivariate df. Its Kendall's tau is defined as*

$$\tau(F) = \mathbb{P}\left((X_2 - X_1)(Y_2 - Y_1) > 0\right) - \mathbb{P}\left((X_2 - X_1)(Y_2 - Y_1) < 0\right),$$

where (X_1, Y_1) and (X_2, Y_2) are iid F.

Kendall's tau is the probability of concordant random pairs minus the probability of discordant random pairs. Since the ordering of X_1 and X_2 is the same as the one of $U_1 = F(X_1)$ and $U_2 = F(X_2)$, it follows that $\tau(F) = \tau(C)$.

Definition 2.2.15 (Spearman's rho)**.** *Let (X, Y) be a bivariate random vector with continuous margins F_1 and F_2, then*

$$\rho_S(X, Y) = \mathrm{Cor}\left(F_1(X), F_2(Y)\right)$$

In other words, Spearman's rho is Pearson's linear correlation applied to the margins transformed to standard uniform variables. Since the df of $F_1(X)$ and $F_2(Y)$ only depends on the copula C, it follows that ρ_S only depends on C and not on the margins. Both Kendall's tau and Spearman's rho take values in the closed interval $[-1, 1]$.

Tail dependence coefficients summarise the degree to which variables are associated in the tails of the distribution.

Definition 2.2.16 (Tail dependence coefficients)**.** *Let (X_1, X_2) be a random vector with margins F_1 and F_2, respectively. Provided the limits exist, the* coefficients of lower *and* upper tail dependence *of X_1 and X_2 are defined by*

$$\lambda_\ell = \lim_{q \downarrow 0} \mathbb{P}\left(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q)\right),$$
$$\lambda_u = \lim_{q \uparrow 1} \mathbb{P}\left(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q)\right).$$

If the margins F_1 and F_2 are continuous, the coefficients of tail dependence are measures of association that only depend on the underlying copula.

Proposition 2.2.17 (Copula representation of tail dependence coefficients)**.** *Let (X_1, X_2) be a bivariate random vector with continuous margins and copula C. Then,*

$$\lambda_\ell = \lim_{w \downarrow 0} \frac{C(w, w)}{w},$$
$$\lambda_u = \lim_{w \downarrow 0} \frac{2w - 1 + C(1 - w, 1 - w)}{w}.$$

For many copula families, there is a one-to-one mapping between a measure of association and the copula dependence parameter, as shown in the following examples.

Example 2.2.18 (Measures of association for the bivariate Gaussian copula)*.* It can be shown that Kendall's tau and Spearman's rho of the bivariate Gaussian copula with correlation parameter ρ are given by

$$\tau(C) = \frac{\arcsin(\rho)}{\pi/2}, \qquad \rho_S(C) = \frac{\arcsin(\rho/2)}{\pi/6},$$

where $\rho \in [-1, 1]$. Furthermore, the Gaussian copula does not exhibit any dependence in the tails unless there is perfect dependence between the variables:

$$\lambda_\ell(C) = \lambda_u(C) = \begin{cases} 0, & \text{if } |\rho| < 1, \\ 1, & \text{if } \rho = 1. \end{cases}$$

Example 2.2.19 (Measures of association for the bivariate Clayton copula)*.* One can show that Kendall's tau for the Clayton copula is

$$\tau(C) = \frac{\theta}{2 + \theta}.$$

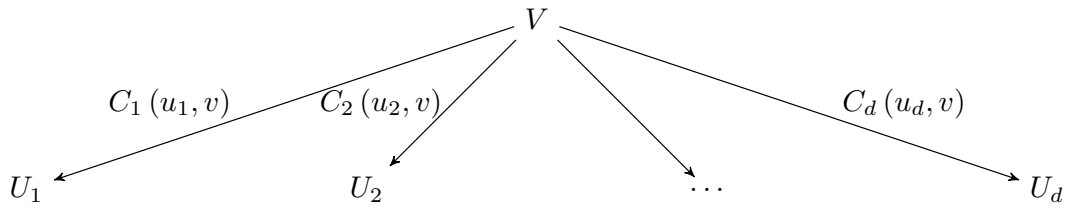


Figure 2.3: One-factor copula model of Krupskii and Joe (2013). The unobservable factor V affects the observable variables U_j via bivariate linking copulas $C_j(u_j, v) = C_{U_j,V}(u_j, v)$, for $j \in \{1, \ldots, d\}$.

where $\theta \in (0, \infty)$. In addition, the Clayton copula exhibits lower tail dependence, but no upper tail dependence:

$$\lambda_\ell(C) = 2^{-1/\theta}, \qquad \lambda_u(C) = 0.$$

Example 2.2.20 (Measures of association for the bivariate Gumbel copula). Kendall's tau for the bivariate Gumbel copula is

$$\tau(C) = \frac{\theta - 1}{\theta}.$$

where $\theta \in [1, \infty)$. In addition, the Gumbel copula exhibits upper tail dependence, but cannot capture dependence in the lower tail:

$$\lambda_\ell(C) = 0, \qquad \lambda_u(C) = 2 - 2^{1/\theta}.$$

In Figure 2.2, the samples of the bivariate Normal, Clayton and Gaussian copulas all have a Kendall's tau of $\tau = 0.7$. Furthermore, the samples from the bivariate Clayton and Gumbel copula clearly exhibit lower and upper tail dependence, respectively, while the sample from the bivariate Normal copula is characterised by a lack of tail dependence in both the lower and the upper tail.

2.2.4 One-factor copula models

As the dimension d of a dataset increases, it becomes more challenging to model the dependence between random variables using copula models that have a limited number of parameters, such as the Clayton and Gumbel copula in Example 2.2.11, and Example 2.2.12, respectively. On the other hand, if the number of copula parameters increases with the dimension, estimation becomes quickly computationally cumbersome, as for the Normal copula in Example 2.2.10 with $d(d-2)/2$ parameters.

To deal with these challenges, Krupskii and Joe (2013) propose a different class of copula models. Assume a common (unobservable) standard uniform variable $V \sim \text{Unif}(0, 1)$ is the source of dependence between d observable standard uniform variables $\boldsymbol{U} = (U_1, \ldots, U_d)$, where $U_j \sim \text{Unif}(0, 1)$ for $j \in \{1, \ldots, d\}$. This is the so-called *one-factor copula model* with $O(d)$ number of dependence parameters and is illustrated in Figure 2.3.

Dependence between the latent factor V and the observable variable U_j is captured by bivariate linking copulas $C_j(u, v)$, for $j \in \{1, \ldots, d\}$. Conditional on the latent variable V, the observable variables are independent:

$$C\left(u_1, \ldots, u_d\right) = \int_0^1 \prod_{j=1}^{d} \mathbb{P}\left(U_j \leq u_j | V = v\right) \mathrm{d}v = \int_0^1 \prod_{j=1}^{d} C_{j|V}(u_j|v)\mathrm{d}v, \tag{2.9}$$

where $C_{j|V} = \partial C_{U_j,V}(u_j, v)/\partial v$.

The one-factor copula density can be obtained by differentiating the copula with respect to the observable variables:

$$c\left(\boldsymbol{u}\right) = \frac{\partial C(\boldsymbol{u})}{\partial u_1 \cdots \partial u_d} = \int_0^1 \prod_{j=1}^{d} c_{j,V}(u_j, v)\mathrm{d}v, \tag{2.10}$$

where $c_{j,V} = c_{U_j,V}$.

Krupskii and Joe (2013) show that the one-factor copula model inherits tail dependence properties from their respective bivariate linking copulas. Thus, by appropriately choosing the bivariate linking copulas C_j, with $j \in \{1, \ldots, d\}$, tail dependence structures can be captured by the one-factor copula model.

To sample from the one-factor copula model, the conditional distribution method in the following algorithm can be used.

Procedure 2.2.21 (Sampling algorithm for one-factor copula models)**.** *Denote the inverse of the conditional copula $C_{j|V}\left(u_j|V\right)$ by $C_{j|V}^{-1}\left(u_j|V\right)$.*

1. *Generate $d+1$ random variables $V, W_1, \ldots, W_d \overset{iid}{\sim} Unif(0,1)$, where V is the unobserved common factor.*
2. *For each $j = 1, \ldots, d$, let $U_j = C_{j|V}^{-1}\left(W_j|V\right)$.*
3. *The vector $\boldsymbol{U} = \left(U_1, \ldots, U_d\right)^\top$ has the one-factor copula C as df.*

If the closed-form formula is known for $C_{j|V}^{-1}$, it can be used. If not, numerical methods are required to determine U_j for which $C_{j|V}\left(U_j|V\right) - W_j = 0$.

Example 2.2.22 (Gaussian one-factor copula model). Let the bivariate linking copulas C_j be Gaussian, with correlation coefficient λ_j, for $j \in \{1, \ldots, d\}$. Then the one-factor copula model is

$$\begin{aligned} C\left(\Phi(x_1), \ldots, \Phi(x_d)\right) &= \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) \\ &= \int_0^1 \prod_{j=1}^{d} \Phi\left(\frac{x_j - \lambda_j \Phi^{-1}(v)}{\sqrt{1-\lambda_j^2}}\right) \mathrm{d}v \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^{d} \Phi\left(\frac{x_j - \lambda_j z}{\sqrt{1-\lambda_j^2}}\right) \phi(z)\mathrm{d}z, \end{aligned}$$

where ϕ is the standard Gaussian probability density function. The one-factor copula model with Gaussian bivariate linking copulas coincides with the multivariate

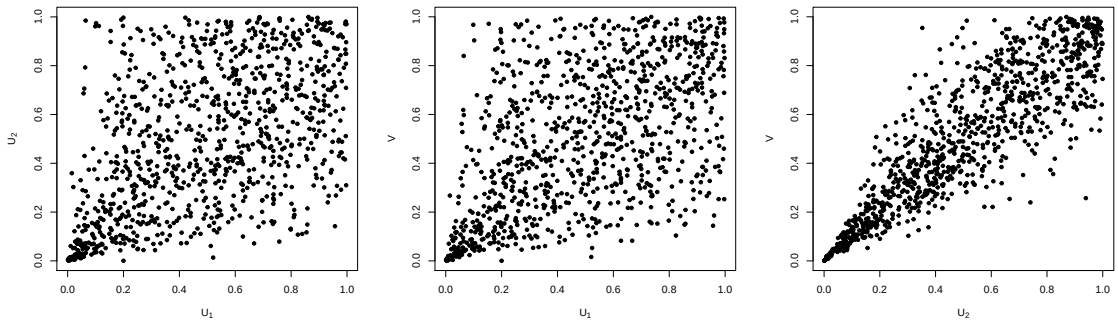


Figure 2.4: Sample of size $n = 1000$ from a one-factor copula model with Clayton bivariate linking copulas and Kendall's tau of $\tau = 0.4$ and $\tau = 0.7$, respectively, where U_1 and U_2 are observable variables, and V is a latent factor.

Gaussian model generated by the Gaussian factor model in Section 2.1 with one Gaussian common factor, as the multivariate d.f. comes from

$$X_j = \lambda_j Z + \gamma_j \varepsilon_j,$$

with $\gamma_j = \sqrt{1 - \lambda_j^2}$, $Z \sim N(0,1)$, $\varepsilon_j \overset{iid}{\sim} N(0,1)$ and $j \in \{1, \ldots, d\}$.

Example 2.2.23 (One-factor copula models). A sample of size 1000 is generated for three different factor copula models with $d = 2$. The first model is constructed using two Clayton linking copulas with dependence parameter of the first and second Clayton copula chosen such that $\tau = 0.4$ and $\tau = 0.7$, respectively. The scatterplots of (U_1, U_2), (U_1, V), and (U_2, V) are given in Figure 2.4, and illustrate how lower tail dependence is inherited from the tail dependence in the bivariate Clayton linking copulas, while dependence between the pair (U_2, V) is stronger than that between the pair (U_1, V).

Similarly, the second factor copula model uses two Gumbel linking copulas with dependence parameter of the first and second Gumbel copula chosen such that $\tau = 0.4$ and $\tau = 0.7$, respectively. The scatterplots of (U_1, U_2), (U_1, V), and (U_2, V) are given in Figure 2.5. Again, the upper tail dependence properties of the factor model are inherited from the upper tail dependence in the bivariate Gumbel copulas, which depends on their Kendall's τ.

The sample from the third factor-copula model is generated using a Gumbel linking copula and a Clayton linking copula with dependence parameters of both linking copulas chosen such that Kendall's $\tau = 0.7$. The scatterplots of (U_1, U_2), (U_1, V), and (U_2, V) are given in Figure 2.6. While the samples from the bivariate linking copulas clearly exhibit upper and lower tail dependence, respectively, there is no apparent tail dependence in the sample from the one-factor copula model.

The one-factor copula model can further be extended to allow for a more general p-factor copula model (Krupskii and Joe, 2013). Furthermore, it is also possible to impose a hierarchichal structure on the factors (Krupskii and Joe, 2015). As these factor copula models will not be used in the rest of thesis, we refer the interested reader to Krupskii and Joe (2013), and Krupskii and Joe (2015).

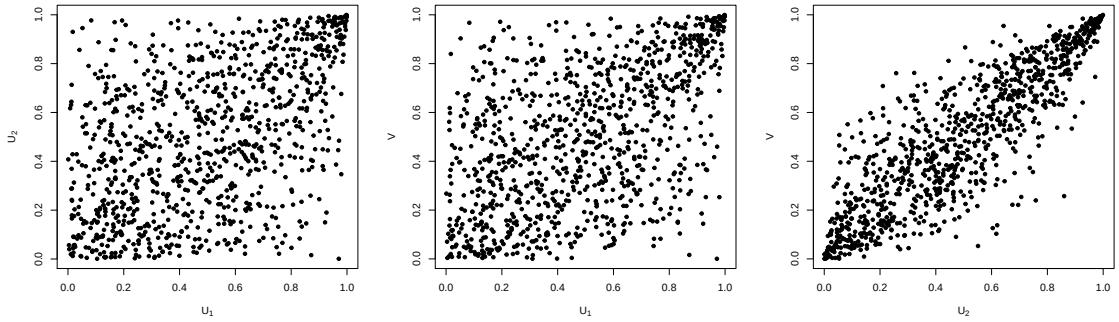


Figure 2.5: Sample of size $n = 1000$ from a one-factor copula model with Gumbel bivariate linking copulas and Kendall's tau of $\tau = 0.4$ and $\tau = 0.7$, respectively, where U_1 and U_2 are observable variables, and V is a latent factor.

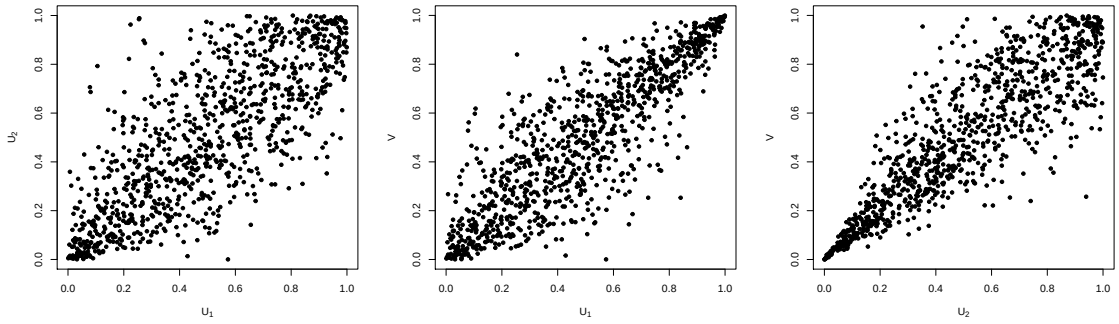


Figure 2.6: Sample of size $n = 1000$ from a one-factor copula model with Gumbel bivariate linking copula for (U_1, V) and Clayton bivariate linking copula for (U_2, V), and Kendall's tau of $\tau = 0.7$ for both bivariate linking copulas, where U_1 and U_2 are observable variables, and V is a latent factor.

A different type of factor copula model was proposed by Oh and Patton (2017). As this factor copula model will not be used in the rest of this thesis, we refer the interested reader to Oh and Patton (2013, 2017).

2.3 Parametric inference for copula models

When the parametric form of both the margins and the copula model is known, the model parameters can be estimated using a full maximum likelihood approach. A different approach is the two-step estimation procedure known as inference functions for margins (Joe and Xu, 1996; Joe, 1997, 2005), where in a first step, the margins are parametrically estimated, and the resulting sample of *pseudo-observations* is plugged into the parametric copula likelihood. When the margins are unknown, a sample of pseudo-observations can be formed using the *ranks* of the observations, which is subsequently used in the parametric copula likelihood, thus yielding a semiparametric pseudo-maximum likelihood estimator. However, when both the

margins and the copula model are unknown, a fully nonparametric approach for both the margins and the copula is warranted, which will be discussed in Section 2.4. An exhaustive overview of estimation methods for copula models can be found in Chapter 4 of Hofert et al. (2018).

2.3.1 Parametric inference for the margins

Let the copula $C_{\boldsymbol{\theta}}$ and the margins $F_j = F_j(\cdot; \boldsymbol{\xi}_j)$ be absolutely continuous, where $\boldsymbol{\theta}$ and $\boldsymbol{\xi}_j$ are unknown parameter vectors, $j \in \{1, \dots, d\}$. Using Proposition **??**, the joint density function is

$$f(\boldsymbol{x}; \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_d) = c_{\boldsymbol{\theta}}(F_1(x_1; \boldsymbol{\xi}_1), \dots, F_d(x_d; \boldsymbol{\xi}_d)) f_1(x_1; \boldsymbol{\xi}_1) \cdots f_d(x_d; \boldsymbol{\xi}_d),$$

where the parametric model can be different for each margin F_j, $j \in \{1, \dots, d\}$. The log-likelihood associated with a dataset $\boldsymbol{\mathcal{X}}_{1:n} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ is

$$\begin{aligned} L_n(\boldsymbol{\theta}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_d) &= \sum_{i=1}^{n} \ln f(\boldsymbol{X}_i; \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_d) \\ &= \sum_{i=1}^{n} \ln c_{\boldsymbol{\theta}}(F_1(X_{i1}; \boldsymbol{\xi}_1), \dots, F_d(X_{id}; \boldsymbol{\xi}_d)) + \sum_{j=1}^{d} \sum_{i=1}^{n} \ln f_j(X_{ij}; \boldsymbol{\xi}_j). \end{aligned} \tag{2.11}$$

Thus, estimates of the parameters $(\boldsymbol{\theta}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_d)$ can be obtained by maximising the log-likelihood in (2.11) with respect to the parametes $\boldsymbol{\theta}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_d$. However, there are two major drawbacks to this approach. First, if at least one of the margins is misspecified, the estimated copula parameter will no longer be unbiased. Second, if the dimension d of the dataset is very large, the log-likelihood is maximised over a large-dimensional space, which may lead to numerical problems (Hofert et al., 2018).

To solve the issue associated with maximisation over a large-dimensional space, one can use a two-step estimator also known as the *inference functions for margins* (ifm) estimator, see Joe and Xu (1996), Joe (1997), and Joe (2005) for more information. In a first step, for $j \in \{1, \dots, d\}$, the $F_j(\cdot; \boldsymbol{\xi}_j)$ are modelled separately and the corresponding log-likelihood

$$\sum_{i=1}^{n} \ln f_j(X_{ij}; \boldsymbol{\xi}_j),$$

is maximised with respect to the parameter vector $\boldsymbol{\xi}_j$ to obtain estimates $\hat{\boldsymbol{\xi}}_j$. In a second step, for each $j \in \{1, \dots, d\}$, the estimated marginal parameter vectors $\hat{\boldsymbol{\xi}}_j$ are used to form a sample of pseudo-observation $F_j\left(X_{ij}; \hat{\boldsymbol{\xi}}_j\right)$, $j \in \{1, \dots, d\}$, which is subsequently plugged into the copula density. The corresponding copula log-likelihood

$$\sum_{i=1}^{n} \ln c_{\boldsymbol{\theta}}\left(F_1\left(X_{i1}; \hat{\boldsymbol{\xi}}_1\right), \dots, F_d\left(X_{id}; \hat{\boldsymbol{\xi}}_d\right)\right)$$

can be maximised to obtain an estimate of the copula parameter vector $\boldsymbol{\theta}$. While this approach solves the problem of numerically maximising over a high-dimensional

space, the ifm estimator for $\boldsymbol{\theta}$ will still be biased if the margins are misspecified (Fermanian and Scaillet, 2005). The latter problem can be addressed by nonparametric estimation of the margins.

2.3.2 Nonparametric inference for the margins

Having at hand a sample $\boldsymbol{\mathcal{X}}_{1:n} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ and the margins $F_1, \ldots, F_d$, one can use the Probability Integral Transform in Proposition 2.2.4 to obtain a true sample $\boldsymbol{U}_i = (F_1(X_{i1}), \ldots, F_d(X_{id}))$ of the copula C, for $i \in \{1, \ldots, n\}$. However, in practice the margins $F_1, \ldots, F_d$ are often unknown, and need to be estimated from the sample $\boldsymbol{\mathcal{X}}_{1:n}$ using the empirical df.

Definition 2.3.1 (Empirical distribution function)**.** *For a d-variate random vector* $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$ *with df* F*, the* empirical distribution function *of a sample* $\boldsymbol{\mathcal{X}}_{1:n} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ *of size* n *is*

$$F_{1:n}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}),$$

where $\boldsymbol{x} \in \mathbb{R}^d$*. The* marginal empirical distribution function *is, for each* $j \in \{1, \ldots, d\}$*,*

$$F_{1:n,j}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_{ij} \leq x),$$

where $x \in \mathbb{R}$*.*

Example 2.3.2 (Empirical distribution function)*.* Let F be the univariate standard Normal distribution, from which we draw 10 iid samples, each of size $n = 50$. For each sample, we calculate the empirical df, which is illustrated in the left-hand panel of Figure 2.7.

Let $R_{ij}^{1:n} = nF_{1:n,j}(X_{ij})$ be the (maximal) rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$. Then for $i \in \{1, \ldots, n\}$,

$$\boldsymbol{R}_i^{1:n} = \left(R_{i1}^{1:n}, \ldots, R_{id}^{1:n}\right), \tag{2.12}$$

$$\hat{\boldsymbol{U}}_i^{1:n} = \frac{\boldsymbol{R}_i^{1:n}}{n+1}. \tag{2.13}$$

These are the multivariate ranks and the pseudo-observations (as multivariate rescaled ranks), respectively, obtained from $\boldsymbol{\mathcal{X}}_{1:n}$. While asymptotically negligible, division of the ranks by $n + 1$ instead of by n ensures that the pseudo-observations fall into the unit hypercube $(0, 1)^d$. This is important in pseudo-maximum likelihood estimation, as some copula densities may explode to infinity at the boundaries.

Thus, if the margins are unknown, a semiparametric pseudo-maximum likelihood approach is used, where the pseudo-observations in (2.13) are plugged into the parametric copula model, for which an estimate of the copula parameter $\boldsymbol{\theta}$ is obtained by maximising the copula likelihood

$$\sum_{i=1}^{n} \ln c_{\boldsymbol{\theta}} \left(\hat{U}_{i1}^{1:n}, \ldots, \hat{U}_{id}^{1:n}\right).$$

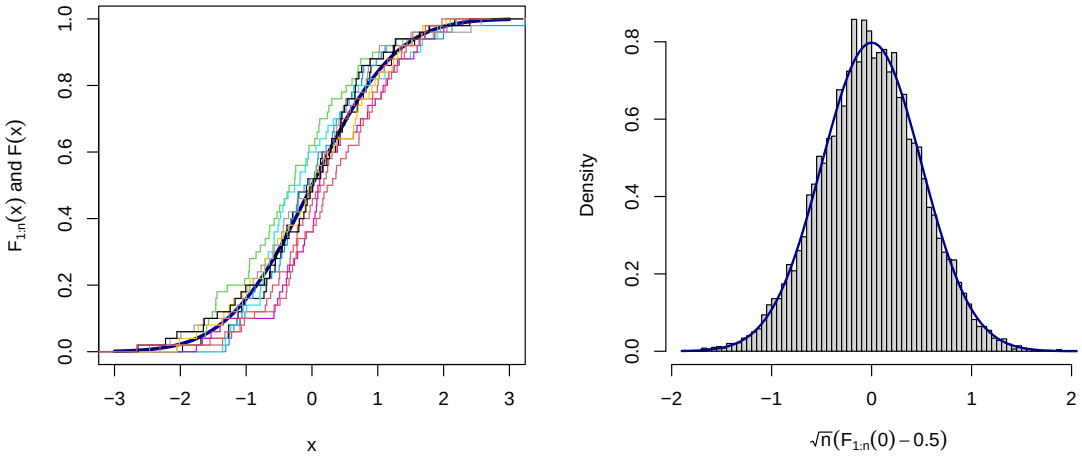


Figure 2.7: Left: 10 empirical dfs for iid samples of size $n = 50$ drawn from a standard Normal distribution depicted in blue. Right: histogram of values of $\sqrt{n}\,(F_{1:n}(0) - 0.5)$ for 10000 independent draws of sample size $n = 10000$ from a standard Normal distribution, with the density of $N\,(0, 0.25)$ depicted in blue.

A last method to parametrically estimate the copula parameter using nonparametrically estimated margins is a method-of-moments estimator based on measures of association such as Kendall's tau and Spearman's rho. Since this will not be used in the rest of the thesis, we refer the interested reader to Chapter 4 in Hofert et al. (2018).

2.3.3 Inference for one-factor copula models

Let $\boldsymbol{U}_i, \ldots, \boldsymbol{U}_n$ be a sample of a one-factor copula. Then the one-factor copula likelihood is

$$L_n\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{n} \ln\left(\int_0^1 \prod_{j=1}^{d} c_{j,V}(U_{ij}, v; \boldsymbol{\theta}_j)\mathrm{d}v\right). \tag{2.14}$$

As the latent factor V is unobservable, one has to resort to numerical integration techniques to compute the integral in (2.14). Krupskii and Joe (2013) propose to use Gauss-Legendre quadrature to numerically approximate the integral in (2.14). See Ziegel (1987) for more details on Gauss-Legendre quadrature.

As the dimension d increases, numerical integration becomes more computationally cumbersome and numerical issues can arise. Krupskii and Joe (2022) propose to create a proxy $\bar{U}_d = \sum_{j=1}^{d} U_j/d$ for the unobservable factor V by using the cross-dimensional average of the observable variables $U_1, \ldots . U_d$. The likelihood for the one-factor copula density can be approximated by plugging $\hat{V}_{id} = \mathrm{rank}(\bar{U}_{id})/(n+1)$, with $\bar{U}_{id} = \sum_{j=1}^{d} U_{ij}/d$, into the expression for the likelihood in (2.14):

$$L_n\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{n}\sum_{j=1}^{d} \ln\left(c_{j,V}(U_{ij}, \hat{V}_{id}; \boldsymbol{\theta}_j)\right),$$

which can be maximised to obtain estimates of $\boldsymbol{\theta}_j$, for $j \in \{1, \ldots, d\}$. However, for $\bar{U}_d$ to be a good enough proxy for V, it is required that d is set sufficiently large (Krupskii and Joe, 2022).

2.4 Nonparametric inference for copula models

In the previous section, parametric inference methods were introduced for copula models, with the margins being modelled either parametrically or nonparametrically. In this section, nonparametric inference methods for the copula itself are introduced using concepts from empirical process theory, for which a more exhaustive overview can be found in Shorack and Wellner (1986), van der Vaart (1998), and Kosorok (2008).

Recall the empirical df introduced in Definition 2.3.1. For fixed x, $nF_{1:n}(x)$ is a binomial random variable with $\mathbb{E}\{F_{1:n}(x)\} = F(x)$. It follows from the Central Limit Theorem (CLT) that for every x,

$$\sqrt{n}\left(F_{1:n}(x) - F(x)\right) \rightsquigarrow N\left(0, F(x)\left(1 - F(x)\right)\right),$$

where '$\rightsquigarrow$' denotes weak convergence.

Example 2.4.1 (Weak convergence of the empirical df). Let F be the standard normal distribution. A sample of size $n = 10000$ is drawn independently 10000 times and the empirical df $F_{1:n}(x)$ is computed at $x = 0$. Weak convergence of $\sqrt{n}\left(F_{1:n}(x) - F(x)\right)$ at $x = 0$ is illustrated in the right-hand panel of Figure 2.7 which depicts the histogram of $\sqrt{n}\left(F_{1:n}(0) - F(0)\right)$, as well as the density of $N\left(0, F(0)\left(1 - F(0)\right)\right)$.

Let T be a set, then $\ell^\infty(T) = \{z : T \to \mathbb{R} \big| \sup_{t\in T} |z(t)| < \infty\}$ is the space of bounded functions equipped with the uniform norm $\|z\|_\infty = \sup_{t\in T} |z(t)|$. A different perspective for weak convergence is to consider $x \mapsto F_{1:n}(x)$ as a random function, and the sequence of *empirical processes* $\sqrt{n}\left(F_{1:n} - F\right)$ as random functions. In the space of bounded functions $\ell^\infty\left(\mathbb{R}^d\right)$ equipped with the uniform metric,

$$\sqrt{n}\left(F_{1:n} - F\right) \rightsquigarrow \mathbb{G}_F,$$

where $\mathbb{G}_F$ is an F-Brownian bridge in $\mathbb{R}^d$, that is, a continuous Gaussian process with mean zero and covariance function $F\left(\boldsymbol{x} \wedge \boldsymbol{y}\right) - F(\boldsymbol{x})F(\boldsymbol{y})$, with $\boldsymbol{x} = (x_1, \ldots, x_d)$, and $\boldsymbol{y} = (y_1, \ldots, y_d)$.

Example 2.4.2 (Weak convergence of the empirical process). Let F be the standard uniform distribution. Convergence for increasing sample size towards a Brownian bridge tied down at zero and one is depicted in the upper, middle, and lower panel of Figure 2.8, which depicts the processes $\sqrt{n}\left(F_{1:n}(x) - F(x)\right)$ for $n = 50$, $n = 500$, and $n = 5000$, respectively.

Let $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{id})$, with $U_{ij} = F_j(X_{ij})$. The sample $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ forms a true sample from the unknown copula C, which can be estimated using the empirical df:

$$\tilde{C}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\boldsymbol{U}_i \leq \boldsymbol{u}\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(F_1(X_{i1}) \leq u_1, \ldots, F_d(X_{i1}) \leq u_d\right).$$

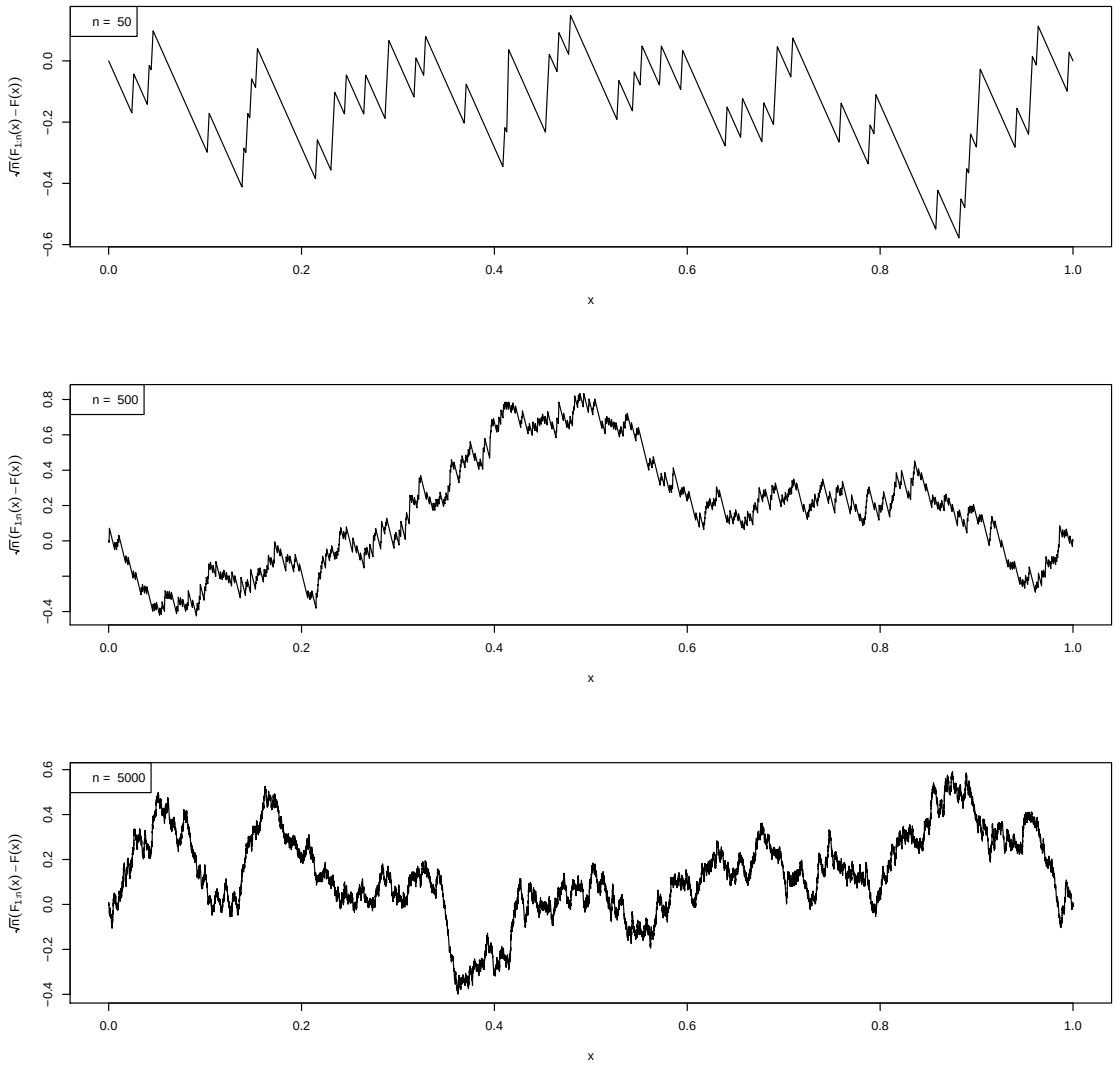


Figure 2.8: The empirical process $\sqrt{n}\left(F_{1:n}(x) - F(x)\right)$, where $F_{1:n}$ is the empirical df and F is a standard uniform df.

By the CLT, for each $\boldsymbol{u} = (u_1, \ldots, u_d)$,

$$\mathbb{G}_n(\boldsymbol{u}) = \sqrt{n}\left\{\tilde{C}_n\left(\boldsymbol{u}\right) - C\left(\boldsymbol{u}\right)\right\} \rightsquigarrow N\left(0, C\left(\boldsymbol{u}\right)\left(1 - C\left(\boldsymbol{u}\right)\right)\right). \tag{2.15}$$

Furthermore, it follows that in the space of bounded functions $\ell^\infty\left([0,1]^d\right)$ equipped with the uniform metric, the empirical process $\mathbb{G}_n \rightsquigarrow \mathbb{G}_C$, where $\mathbb{G}_C$ is a C-Brownian bridge with covariance function $C(\boldsymbol{u} \wedge \boldsymbol{v}) - C(\boldsymbol{u})C(\boldsymbol{v})$, where $\boldsymbol{u} = (u_1, \ldots, u_d)$, and $\boldsymbol{v} = (v_1, \ldots, v_d)$, which coincides with the covariance $C\left(\boldsymbol{u}\right)\left(1 - C\left(\boldsymbol{u}\right)\right)$ when considering pointwise convergence in (2.15).

In practice, the true margins are often unknown, and need to be estimated using the rescaled ranks in (2.13). The empirical copula was first defined in Rüschendorf (1976) and Deheuvels (1979, 1981) as the empirical df of the rescaled ranks. Thus,

a nonparametric estimator of C in (2.6) is

$$C_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{\boldsymbol{U}}_i^{1:n} \leq \boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} \mathbf{1}(\hat{U}_{ij}^{1:n} \leq u_j), \tag{2.16}$$

where $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$, and $\hat{\boldsymbol{U}}_i^{1:n}$ is defined in (2.13). Furthermore, the empirical copula process $\mathbb{C}_n$ is

$$\mathbb{C}_n(\boldsymbol{u}) = \sqrt{n}\left\{C_n(\boldsymbol{u}) - C(\boldsymbol{u})\right\}, \tag{2.17}$$

where C_n and C are given in 2.16 and in Definition 2.2.5, respectively.

To establish convergence of the empirical copula process, we use the following non-restrictive condition from Segers (2012).

Condition 2.4.3 (Smooth partial derivative of the copula)**.** *For each $j \in \{1, \ldots, d\}$, the jth first-order partial derivative $\dot{C}_j = \partial C/\partial u_j$ exists and is continuous on the set $V_{d,j} = \{\boldsymbol{u} \in [0,1]^d : u_j \in (0,1)\}$.*

Furthermore, we arbitrarily define $\dot{C}_j$ to be zero on the set $\{\boldsymbol{u} \in [0,1]^d : u_j \in \{0,1\}\}$, which implies that $\dot{C}_j$ is defined on the entire interval $[0,1]^d$.

Let

$$\tilde{\mathbb{C}}_n(\boldsymbol{u}) = \mathbb{G}_n(\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u})\mathbb{G}_n\left(\boldsymbol{u}^{(j)}\right), \tag{2.18}$$

where $\mathbb{G}_n(\boldsymbol{u})$ is defined in (2.15), and $\mathbb{G}_n \rightsquigarrow \mathbb{G}_C$, with $\mathbb{G}_C$ is a C-Brownian bridge. Let $g(\boldsymbol{u}) = f(\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u}) f\left(\boldsymbol{u}^{(j)}\right)$. Because the map $f \mapsto g$ from $\ell^\infty\left([0,1]^d\right)$ onto $\ell^\infty\left([0,1]^d\right)$ is linear and bounded, it follows from the continuous mapping theorem (CMT) that $\tilde{\mathbb{C}}_n \rightsquigarrow \mathbb{C}_C$, where

$$\mathbb{C}_C(\boldsymbol{u}) = \mathbb{G}_C(\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u})\mathbb{G}_C\left(\boldsymbol{u}^{(j)}\right), \tag{2.19}$$

with $\mathbb{G}_C$ a C-Brownian bridge.

Weak convergence of the empirical copula process was established by Segers (2012).

Proposition 2.4.4 (Convergence of the empirical copula process)**.** *If Condition 2.4.3 holds, then*

$$\sup_{\boldsymbol{u}\in[0,1]^d} |\mathbb{C}_n(\boldsymbol{u}) - \tilde{\mathbb{C}}_n(\boldsymbol{u})| \xrightarrow{\mathbb{P}} 0,$$

where $\mathbb{C}_n$ and $\tilde{\mathbb{C}}_n$ are defined in (2.17) *and* (2.18)*, respectively. As a consequence, in $\ell^\infty\left([0,1]^d\right)$,*

$$\mathbb{C}_n \rightsquigarrow \mathbb{C}_C.$$

The proof of Proposition 2.4.4 can be found in the proof of Proposition 3.1 in Segers (2012).

A different approach to nonparametric estimation of the copula is the empirical beta copula. The empirical beta copula of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is

$$C_n^{\beta}(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} F_{n,R_{ij}^{1:n}}(u_j),$$

where $\boldsymbol{u} \in [0,1]^d$, $F_{n,r}$ is the beta distribution with parameters r and $n+1-r$, and $R_{ij}^{1:n}$ is the maximal rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$. The empirical beta copula of Segers et al. (2017) imposes smoothness on C by replacing the indicator function in the definition of the empirical copula by a product of beta dfs, and has standard uniform univariate margins if there are no ties in the component samples of the observations. Therefore, it is a genuine copula (Hofert et al., 2018). The advantage of the empirical beta copula is that, in small samples, it outperforms the empirical copula in terms of bias and variance, while its asymptotic distribution is the same as that of the empirical copula (Segers et al., 2017).

2.5 Statistical process control and change-point detection

In the previous sections we introduced models to capture dependence in multivariate distributions. However, when one has at hand a multivariate time series, it is highly likely that the assumption of stationarity in the underlying time series is violated. If the parameters of a multivariate model are estimated, the statistician might want to know whether the model can be used for forecasting. *Change-point detection* methods can be used to detect deviations from stationarity in a possibly multivariate time series, and has a long history with applications in a wide range of research fields such as quality control (Page, 1955; Hawkins and Olwell, 2012), economics (Perron et al., 2006), and finance (Andreou and Ghysels, 2009; Aue et al., 2012). We refer the interested reader to Hawkins et al. (2003), Montgomery (2007), Aue and Horváth (2013), and Gösmann et al. (2022) for a more exhaustive overview of the literature.

2.5.1 Statistical process control

Change-point detection originates from quality control in manufacturing. The traditional Statistical Process Control (SPC) approach uses Shewhart charts (Shewhart, 1931), also known as SPC charts, and exponentially weighted moving average (EMWA) control charts (Roberts, 2000), and tries to control the average run length (ARL) of the control charts, where the ARL is the average number of observations from the start of the monitoring until a false alarm is raised (that is, until the process is declared to be out-of-control), under the null hypothesis of no change (Montgomery, 2007). The first phase of the SPC approach assumes that the true statistical parameters are exactly known from a long-lasting stable period of observations, and the following monitoring procedure is implemented using these parameters. Such an approach might be realistic in a quality control setting within

the manufacturing industry, but seems highly unrealistic when analysing financial or economic time series, where initial parameters are more often than not unknown, and have to be estimated using an initial stretch of time series observations. In addition to the aforementioned limitation, the original Shewhart charts only use the last available observation to make a decision on whether or not to reject the null hypothesis, while Page (1954) proposed an SPC approach using cumulative sums (CUSUM), see Page (1954, 1955), and Montgomery (2007) for more details. When the cost associated with a false alarm is too high, it can be more important to control the type I error, the probability of raising a false alarm. In general the SPC approach leads to the rejection of the null hypothesis of stationarity with probability one. Chu et al. (1996) propose a change point detection model such that the probability of a false alarm can be controlled asymptotically. The approach by Chu et al. (1996) guarantees that the type I error can be controlled as the size of the initial stable learning sample, m, tends to infinity, and allows for the estimation of parameters such as the long run variance.

2.5.2 Change point detection

The main idea of change point detection is that one has at hand an initial stable, possibly multivariate, sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ of size m, called the *learning sample*, and starts monitoring new data observations. There are two approaches to detect change points when monitoring new data: *closed-end* and *open-end* change-point detection. The first, closed-end approach, monitors the data until the arrival of an observation $\boldsymbol{X}_n$, with $n > m$ a fixed endpoint, and if there is no evidence of nonstationarity in the underlying data, the monitoring stops after the arrival of this observation $\boldsymbol{X}_n$. Thus, at most $n - m$ observations will be considered before monitoring stops. The corresponding null hypothesis will be

$$H_0 : \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \ldots, \boldsymbol{X}_n \text{ is a stretch from a stationary time series,}$$

against the alternative that there is a change point at position $k^\star$, where $m \leq k^\star \leq n$. One approach to test the null hypothesis of stationarity in a closed-end framework is to use a likelihood ratio approach. In this setting, it is assumed that a change point is located at position $k^\star$, and the data is split into two subsamples with different parameters. Then, a likelihood function is constructed that takes into account that these subsamples have different parameters, and the null hypothesis of stationarity will be rejected when the likelihood ratio statistic is very large (Aue and Horváth, 2013).

The second change point detection framework is an *open-end* approach, in which observations $\boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots$ arrive sequentially, and monitoring can in theory continue indefinitely. The corresponding null hypothesis will be

$$H_0 : \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots \text{ is a stretch from a stationary time series,}$$

against the alternative that there is a change point at position $k^\star$, where $m \leq k^\star$. While conceptually the closed-end and open-end framework are closely related, from a mathematical point of view there are substantial differences, as the asymptotic theory of closed-end procedures can be derived using functional limit theorems, while the asymptotic theory in the open-end change point detection framework requires additional assumptions, as will become clear in Chapter 4.

Furthermore, one can distinguish between two types of data collection for monitoring data. On the one hand, in *retrospective* or *offline* change point detection, the entire sample $\boldsymbol{X}_{m+1}, \ldots, \boldsymbol{X}_n$ is already available before monitoring starts. On the other hand, in *sequential* or *online* analysis, monitoring is considered as the data $\boldsymbol{X}_{m+1}$, $\boldsymbol{X}_{m+2}, \ldots$ arrives sequentially. For both closed-end and open-end approaches, the goal is to detect changes as quickly as possible while controlling the probability of a type I error. In this thesis, we focus on nonparametric sequential open-end change-point detection procedures to monitor for changes in the empirical df of possibly multivariate time series data.

The open-end nonparametric monitoring procedure consists of a detector and a threshold function and compares, after the arrival of the kth observation $\boldsymbol{X}_k$, a positive statistic, the detector T_m, to a threshold. If after the arrival of observation $\boldsymbol{X}_k$, with $k > m$, the detector exceeds the threshold, the hypothesis of stationarity is rejected. If not, a new observation $\boldsymbol{X}_{k+1}$ is collected and the monitoring procedure is repeated from the start. For a given significance level $\alpha \in (0, 1/2)$ the goal is to find a threshold $w(k/m)$ such that, under the null hypothesis of stationarity, the following holds:

$$\mathbb{P}\left\{T_m(k) \leq w(k/m) \text{ for all } k > m\right\} = \mathbb{P}\left\{\sup_{k>m} \frac{T_m(k)}{w(k/m)} \leq 1\right\} = 1 - \alpha. \tag{2.20}$$

One can consider a threshold function

$$w(t) = q_{\boldsymbol{y}}^{(1-\alpha)} v(t), \quad t \in [1, \infty),$$

where $v(t)$ is a weighting function such that (2.20) holds, $\boldsymbol{y}$ is a (possibly empty) vector of parameters, and $q_{\boldsymbol{y}}^{(1-\alpha)}$ is the $(1-\alpha)$-quantile of the limiting random variable $\mathcal{I}_{\boldsymbol{y}}$, the weak limit of $\sup_{k>m} T_m(k)/v(k/m)$. By the Portmanteau lemma,

$$\begin{aligned}\lim_{m\to\infty} \mathbb{P}\left\{\sup_{k>m} \frac{T_m(k)}{w(k/m)} \leq 1\right\} &= \lim_{m\to\infty} \mathbb{P}\left\{\sup_{k>m} \frac{T_m(k)}{v(k/m)} \leq q_{\boldsymbol{y}}^{(1-\alpha)}\right\} \\ &= \mathbb{P}\left\{\mathcal{I}_{\boldsymbol{y}} \leq q_{\boldsymbol{y}}^{(1-\alpha)}\right\} = 1 - \alpha.\end{aligned}$$

When having at hand a learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ and a sample $\boldsymbol{X}_{m+1}, \ldots, \boldsymbol{X}_k$, $k \geq m + 1$, a first approach used by Horváth et al. (2004) and Aue et al. (2006) for linear models, is to build a detector that evaluates the differences between $\theta_{1:m}$ computed from the observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, and $\theta_{m+1:k}$ computed from the observations $\boldsymbol{X}_{m+1}, \ldots, \boldsymbol{X}_k$:

$$\theta_{1:m} - \theta_{m+1:k}.$$

Another approach by Fremdt (2015) and Kirch and Weber (2018) uses a so-called Page-CUSUM approach. For this approach, a detector is constructed that considers the differences

$$\theta_{1:m} - \theta_{j+1:k},$$

where $j \in \{m + 1, \ldots, k - 1\}$. Dette and Gösmann (2020) and Gösmann et al. (2021) propose to use a detector that is motivated by the likelihood ratio test, in

the closed-end and open-end case, respectively. This detector takes into account all the differences

$$\theta_{1:j} - \theta_{j+1:k},$$

where $j \in \{m+1, \ldots, k-1\}$. The advantage of the last approach is that it screens for all potential change points at each $j \in \{m+1, \ldots, k-1\}$.

Change point detection procedures have been widely studied, and the corresponding monitoring procedures have been used in a wide range of data applications. Horváth et al. (1999) in a closed-end setting and Horváth et al. (2004) in an open-end setting setting screen for changes in the parameters of linear models with a statistic based on the the residuals, and illustrate their approach by monitoring for changes in monthly temperature data. Similarly, Aue et al. (2006) develop a statistic based on the residuals of linear models to monitor for changes in the parameters of linear models, but their approach allows for the presence of heteroscedasticity in the error terms, and is used to monitor for changes in financial data.

Fremdt (2015) studies changes in linear models in time series regression and uses the procedure to monitor a set of capital asset pricing data in the context of the Fama–French extension of the Capital Asset Pricing Model (CAPM). Another application of change point detection in the CAPM framework is proposed by Aue et al. (2012), who monitor parameters in univariate GARCH models and screen for change points in high frequency portfolio betas in the CAPM.

A monitoring procedure for structural change in multivariate time series is proposed by Groen et al. (2013), and focuses on parameters in linear regression models, while Pape et al. (2016) monitor for structural changes in variances in a series of random vectors of small dimension, which is illustrated using log-returns of stocks listed on the DAX. Kojadinovic and Verdier (2021) propose a procedure to monitor for changes in the empirical distribution function in a closed-end setting and illustrate using the NASDAQ composite index. Dette and Gösmann (2020) work in the closed-end setting and propose a detector to monitor for changes in the mean, variance, correlation and quantiles of time series, and use their approach to screen for change points in stock prices during the dot-com bubble. Gösmann et al. (2021) and Holmes and Kojadinovic (2021) design a procedure to monitor for changes in the mean of a univariate time series. Gösmann et al. (2021) use their monitoring procedure to screen for changes exchange rates, while Holmes and Kojadinovic (2021) apply their change point detection scheme to monitor for changes in global temperature anomalies. Gösmann et al. (2022) develop a sequential procedure to detect for changes in a high-dimensional time series and illustrate using a data example from hydrology.

2.6 Overview of the remaining chapters

The remainder of the thesis is organised as follows. As presented in Section 2.2.4, one-factor copula models can be used to model high-dimensional multivariate data. In Chapter 3, we expand upon existing methods by proposing tractable fast inference methods for very high-dimensional one-factor copula models that combine the results from the proxy approach for one-factor copula models in Section 2.3.3 with the characteristics of Gaussian factor models introduced in Section 2.1.

Multivariate models presented in Section 2.1 and Section 2.2 are a straightforward method to capture dependence in the data. Nevertheless, it is highly unlikely that the underlying time series remains stable over time. Chapter 4 proposes a new sequential change-point detection scheme to monitor for changes in the empirical df of possibly multivariate data.

Chapter 5 concludes by summarising the results obtained in this thesis, and provides suggestions for future research. All proofs have been deferred to Appendices A and B to improve the flow of the main text.

Chapter 3

Fast inference methods for high-dimensional factor copulas

This chapter and the corresponding Appendix A.1 and Appendix A.2 have been published in *Dependence Modeling* (Verhoijsen and Krupskiy, 2022). Additional material related to the computation of the gradient for the Generalised Method of Moments approach and a fast matrix inversion scheme for the variance-covariance matrix of the model are presented in Appendix A.3 and Appendix A.4, respectively.

Abstract

Gaussian factor models allow the statistician to capture multivariate dependence between variables. However, they are computationally cumbersome in high-dimensions and are not able to capture multivariate skewness in the data. We propose a copula model that allows for arbitrary margins, and multivariate skewness in the data by including a non-Gaussian factor whose dependence structure is the result of a one-factor copula model. Estimation is carried out using a two-step procedure: margins are modelled separately and transformed to the normal scale, after which the dependence structure is estimated. We develop an estimation procedure that allows

for fast estimation of the model parameters in a high-dimensional setting. We first prove the theoretical results of the model with up to three Gaussian factors. Then, simulation results confirm that the model works as the sample size and dimensionality grow larger. Finally, we apply the model to a selection of stocks of the SP500 which demonstrates that our model is able to capture cross-sectional skewness in the stock market data.

3.1 Introduction

In the literature, Gaussian factor models are used as a dimensionality reduction technique to describe a vector of d variables $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ as a linear combination of a vector of p Gaussian latent factors $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$, where for meaningful cases $p < d$. Thus, variables form groups based on their correlation, such that correlation within one particular group of variables can be high, while the correlation between variables assigned to different groups can be low (Kent et al., 1979; Johnson and Wichern, 2002). Because of the minimal assumptions underlying factor models, this technique is amply used in a wide range of applications. This includes econometrics, where factor models are either used to identify the factor structure itself (Bai and Ng, 2002, 2008; Stock and Watson, 2002), or to control for endogenous common factors in factor-augmented panel data regressions (Pesaran, 2006). Other applications include the finance literature, see Chapter 5 in McNeil et al. (2005) or Chapter 9 in Tsay (2005), and hydrology (Subbarao et al., 1996).

Nevertheless, the assumption of multivariate Gaussianity implies that these factor models cannot capture cross-sectional skewness or kurtosis in the underlying multivariate distribution. Copula models, on the other hand, provide a straightforward technique to capture the entire dependence structure between random variables, and allow for non-Gaussian skewness and kurtosis. A d-dimensional copula $C : [0,1]^d \to [0,1]$ is the cumulative distribution function (cdf) of a random vector $\boldsymbol{U} = (U_1, \ldots, U_d)$, where $U_j \sim U(0,1)$. Let $F_1, \ldots, F_d$ be univariate cdf's, and let $\boldsymbol{x} = (x_1, \ldots, x_d)$. Then by Sklar (1959), the function $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ is a d-dimensional cdf with margins $F_1, \ldots, F_d$. Vice-versa, if F is a d-dimensional cdf with univariate cdf's $F_1, \ldots, F_d$, then there exists a copula such that $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ holds (Sklar, 1959). If the margins are continuous, then one speaks of *the* copula, which is unique and equal to $C(\boldsymbol{u}) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$, where $F_j^{\leftarrow}$ is the generalised inverse function of F_j. Note, however, that the copula model will only be unique in the case of continuous random variables. If the random variables are discrete, this is not necessarily the case, as illustrated by Example 5.5 in McNeil et al. (2005).

While copula models provide an elegant technique to model dependence between random variables, they come with certain restrictions. On the one hand, they can be computationally demanding in a high-dimensional setting, while on the other hand they might not be flexible enough to model dependence between a large number of random variables. A solution was provided by Krupskii and Joe (2013), who propose a class of factor copula models that are both parsimonious and flexible. The idea is to assume p latent factors $V_1, \ldots, V_p$ that cause dependence between d observable variables $U_1, \ldots, U_d$ via bivariate linking copulas. These types of factor copula models can be extended to allow for a more structured approach (Krupskii

and Joe, 2015). A different type of factor copula model was proposed by Oh and Patton (2017). The latter model uses a simulated method of moments (SMM) estimation method described in Oh and Patton (2013), and is extended to allow for time-varying parameters in Oh and Patton (2018). However, estimation for this type of factor copula model is particularly slow in higher dimensions, as SMM is based on a resampling scheme. Bayesian estimation methods for factor copula models are explored by Creal and Tsay (2015), Schamberger et al. (2017), Tan et al. (2019), Nguyen et al. (2020), and Kreuzer and Czado (2021), but are generally also computationally demanding. A novel method to decrease the computational time to estimate factor copula models is proposed by Krupskii and Joe (2021), who provide a procedure that allows for fast estimation of the factor copula parameters in non-overlapping factor copula models. The latter include the one-factor copula model in Krupskii and Joe (2013) or the nested factor copula model in Krupskii and Joe (2015).

The main contribution of this article is to provide a fast and computationally tractable procedure to model very high-dimensional data characterised by non-Gaussian skewness in the multivariate distribution. This is done in a two-step approach. First, marginals are modelled separately and transformed to normal scores. Second, dependence between the marginals is modelled by means of $p + 1$ common factors: one factor resulting from a one-factor copula model as described in Krupskii and Joe (2013), and p Gaussian factors.

Factor loadings and the residual correlation parameter are estimated using a Generalised Method of Moments (GMM) approach, which provides consistent and asymptotically normal estimates as the sample size n tends to infinity. In a subsequent step, the unobserved Gaussian factors are estimated using a weighted combination of the observed variables. Weights are obtained by maximising the correlation between the estimated and the true Gaussian factors, while convergence is shown by letting the sample size n tend to infinity, and setting the dimension d sufficiently large. To allow for fast inference of the one-factor copula parameter, we use the proxy factor idea proposed by Krupskii and Joe (2021). We show that as n and d tend to infinity, the proxy converges towards the true factor.

In sum, the proposed model exploits both the versatility of Gaussian factor models and the fast inference scheme of non-overlapping factor copula models. The remainder of the article is organised as follows. The new model is proposed in Section 3.2, its main properties are discussed, and the statistical procedure is outlined. The theoretical results are outlined in Section 3.3, and the estimation strategy is illustrated with a simulation experiment in Section 3.4, where we propose a closed-form expression to quickly invert the variance-covariance matrix of the observed data. We illustrate the model by applying it to a real-world data example in Section 3.5. Section 3.6 wraps up the article with a discussion and ideas for future research.

3.2 Model outline and inference

In Section 3.2.1, a new factor model is presented that combines the characteristics of Gaussian factor models and of factor copulas, while still being computationally tractable in high dimensions. The estimation procedure is subsequently discussed in Section 3.2.2.

3.2.1 Model outline

Assume one has at hand data from arbitrary margins which in a first step are transformed to $N(0,1)$ marginals. Then $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ is a d-dimensional vector of observable variables so that $X_j \sim N(0,1)$, for $j = 1, \ldots, d$. In a p-factor model, the observable variables X_j are a linear combination of p independent latent Gaussian factors and an error term:

$$X_j = \sum_{k=1}^{p} \lambda_{jk} Z_k + \gamma_j \varepsilon_j, \quad Z_1, \ldots, Z_k, \varepsilon_j \sim_{\text{i.i.d.}} N(0,1), \tag{3.1}$$

where $\lambda_{jk} = \text{Corr}(X_j, Z_k)$ and $\gamma_j = \sqrt{1 - \sum_{k=1}^{p} \lambda_{jk}^2}$, $j = 1, \ldots, d$. If the errors $\varepsilon_1, \ldots, \varepsilon_p$ are independent, then the model (3.1) is a classical Gaussian p-factor model. In this work, we assume that $\mathbb{E}(\varepsilon_{j_1} \varepsilon_{j_2}) = \eta$ for $j_1 \neq j_2$, and the errors are independent conditional on some unobserved factor $V \sim U(0,1)$ and the respective copula

$$C(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1}^{d} C_{U|V}(u_j \mid v)\, \mathrm{d}v, \tag{3.2}$$

where $C_{U|V}(u_j \mid v) = \partial C_{U,V}(u_j, v) / \partial v$, and $C_{U,V}$ is the copula linking (ε_j, V). In other words, C is a one-factor copula; see Krupskii and Joe (2013) for more details. Casting the model in (3.1) into matrix notation gives:

$$\boldsymbol{X} = \boldsymbol{\Lambda} \boldsymbol{Z} + \boldsymbol{\Gamma}^{1/2} \boldsymbol{\varepsilon}, \tag{3.3}$$

where $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ is a column vector of d observable variables, and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$ is a $d \times p$ matrix of p column vectors $\boldsymbol{\lambda}_k = (\lambda_{1k}, \ldots, \lambda_{dk})^\top$ consisting of d factor loadings, where λ_{jk} captures the impact of factor Z_k on the j-th variable X_j. The p Gaussian factors are elements of the column vector $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top$, while the loadings on the d residuals $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)^\top$ are given in the $d \times d$ diagonal matrix $\boldsymbol{\Gamma}^{1/2} = \text{diag}(\gamma_1, \ldots, \gamma_d)$. Lastly, dependence between the residuals is modelled by the homogeneous one-factor copula model in (3.2).

Denote

$$\mathbf{H} = \begin{pmatrix} 1 & \eta & \cdots & \eta \\ \eta & 1 & \cdots & \eta \\ \vdots & \vdots & \ddots & \vdots \\ \eta & \eta & \cdots & 1 \end{pmatrix},$$

$\mathbf{H}$ is a $d \times d$ matrix, with $\eta = \mathbb{E}(\varepsilon_{j_1} \varepsilon_{j_2})$, where $j_1 \neq j_2$ and $j_1, j_2 = 1, \ldots, d$. Then the correlation matrix of $\boldsymbol{X}$ is

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top + \boldsymbol{\Gamma}^{1/2} \mathbf{H} \boldsymbol{\Gamma}^{1/2}. \tag{3.4}$$

The difference with a standard Gaussian factor model lies in the assumption that $\mathbb{E}\left(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top\right) = \mathbf{H} \neq \mathbf{I}_d$, which implies the presence of correlation between the residual terms.

The following example shows that the model 3.1 can capture asymmetric dependence.

Example 3.2.1. Let $p = 2$, $d = 4$ and assume that the copula linking (ε_j, V) is the Clayton copula with parameter 2 which corresponds to moderate lower tail dependence. In this case, $\eta \approx 0.52$. Consider the model:

$$\begin{aligned} X_j &= 0.6Z_1 + 0.8\varepsilon_j, \quad j = 1, 2, \\ X_j &= 0.6Z_2 + 0.8\varepsilon_j, \quad j = 3, 4. \end{aligned}$$

Figure 3.1 shows scatterplots of different pairs of the observed variables. It is seen that pairs (X_1, X_2) and (X_3, X_4) have stronger dependence in the lower tail, while (X_1, X_3) has much weaker dependence and it does not exhibit strong asymmetry.

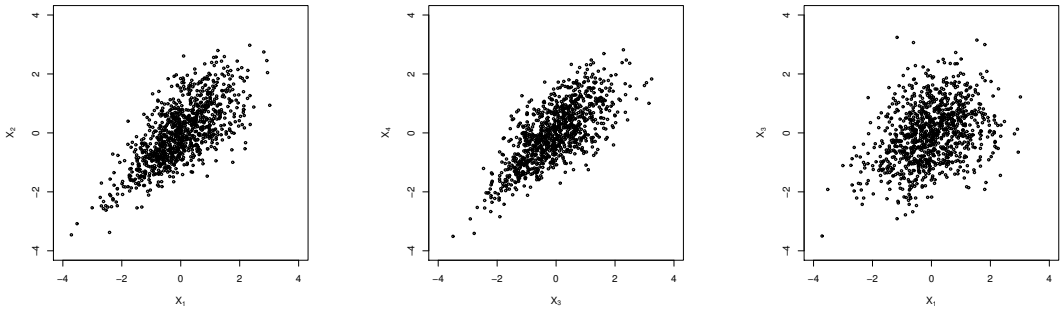


Figure 3.1: Scatterplots of different pairs of observed variables in Example 3.2.1

3.2.2 Inference

In this section, we present the statistical procedure that ensures we can estimate the model parameters, the Gaussian factors, and the latent factor of the one-factor copula model. The corresponding theoretical results are outlined in Section 3.3.

We begin by estimating the parameters in the model, for which we implement a GMM approach. Let $\mathbf{X}_n = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be a sample of size n, where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$, for $i = 1, \ldots, n$. Moreover, let $\hat{\boldsymbol{\Sigma}}_n = \hat{\boldsymbol{\Sigma}}_{\mathbf{X}_n}$ and define

$$Q_n(\boldsymbol{\Lambda}, \eta) = \sum_{j=1}^{d} \sum_{k=1}^{d} \left\{ \left(\boldsymbol{\Sigma}_{\boldsymbol{X}} - \hat{\boldsymbol{\Sigma}}_n \right)^2 \right\}_{j,k=1}^{d},$$

then the model parameters can be obtained by

$$(\hat{\boldsymbol{\Lambda}}, \hat{\eta}) = \operatorname{argmin}_{(\boldsymbol{\Lambda}, \eta)} Q_n(\boldsymbol{\Lambda}, \eta).$$

After obtaining the parameters, we want to estimate the Gaussian factors. This can be done by creating a proxy $\hat{Z}^{\star}_{knd} = \hat{\boldsymbol{w}}^{\star\top}_{knd} \mathbf{X}_n$ of the true Gaussian factors Z_k (for $k = 1, \ldots, p$) by taking a weighted combination of the observable variables. Moreover, it is possible to obtain unique optimal weights $\hat{\boldsymbol{w}}^{\star}_{knd} = (\hat{\boldsymbol{\lambda}}_k^{\top} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1} \hat{\boldsymbol{\lambda}}_k)^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1} \hat{\boldsymbol{\lambda}}_k$ by minimising the $\mathscr{L}_2$ norm between the proxy factor and the true factor. In Section 3.4.1, we propose a closed-form expression to quickly invert the $d \times d$ variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$.

The next step is to recover the latent factor in the one-factor copula model. To this end, the estimated residuals can be recovered as:

$$\hat{\varepsilon}_{jnd} = \frac{1}{\hat{\gamma}_j}\left(X_j - \sum_{k=1}^{p} \hat{\lambda}_{jk}\hat{Z}^{\star}_{knd}\right),$$

using the estimated values of the factor loadings. We can then use a proxy $\bar{U}_{nd} = \sum_{j=1}^{d} \hat{\varepsilon}_{jnd}/d$ to approximate the unknown factor V in the one-factor copula model. The density of the one-factor copula model is obtained by differentiating (3.2), that is

$$c(u_1, \ldots, u_d) = \frac{C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} = \int_0^1 \prod_{j=1}^{d} c_{U,V}\left(u_j, v\right) \mathrm{d}v.$$

Thus, by plugging in $\bar{U}_{nd}$ as a proxy for V, and using $\hat{V}_{ind} = \mathrm{rank}(\bar{U}_{ind})/(n+1)$, we can obtain the copula parameter θ by maximising the log-likelihood

$$\sum_{j=1}^{d}\sum_{i=1}^{n} \ln c_{U,V}\left(U_{ij}, \hat{V}_{ind}; \theta\right).$$

3.3 Theoretical results

We prove these results for the model with $p \leq 3$ Gaussian factors. All proofs are deferred to Appendix A.2.

We assume the following conditions hold.

Condition 3.3.1. *Let $\delta_{jk} = \lambda_{jk}/\gamma_j$, the vector $\Delta_k = \{(\delta_{lk} - \delta_{mk})/d\}_{l,m=1}^{d}$ and ϕ_{k_1,k_2} is the angle between Δ_{k_1} and Δ_{k_2}. There exist $\xi_0, \xi_L, \xi_U \in (0, \infty)$, and a constant integer $d_0 \in \mathbb{N}$ such that the following hold for any $d > d_0$ and all $j = 1, \ldots, d$, and for all $k, k_1, k_2, k_3 = 1, \ldots, p$:*

(i) $\xi_L < \frac{1}{d}\sum_{l=1}^{d} \delta_{lk}^2 < \xi_U < \infty$

(ii) $\boldsymbol{\Sigma}_X$ *is invertible and* $\xi_0 < |(\boldsymbol{\Sigma}_X)_{j,k}| < 1 - \xi_0$ *for* $j \neq k$*;*

(iii) $||\Delta_k|| > \xi_0$, $\phi_{k_1,k_2} + \phi_{k_1,k_3} - \phi_{k_2,k_3} > \xi_0$ *for* $k_1 \neq k_2 \neq k_3$ *and* $\phi_{k_1,k_2} + \phi_{k_1,k_3} + \phi_{k_2,k_3} < 2\pi - \xi_0$, *where* $\|\Delta_k\| = \frac{1}{d}\sqrt{\sum_{l=1}^{d}\sum_{m=1}^{d}(\delta_{lk} - \delta_{mk})^2}$ *is the Eucledian norm.*

These conditions imply that vectors $\Delta_1, \Delta_2, \Delta_3$ are linearly independent. In Appendix A.1, we show that these are not restrictive conditions if factor loadings are not linearly dependent.

The next theorem states that GMM estimation yields parameter estimates that converge in probability to their true values, given some regularity conditions. Moreover, the resulting estimators are asymptotically normal, as stated in the subsequent theorem.

Theorem 3.3.2 (Consistency of the GMM estimators)**.** *Let* $f(\mathbf{X}_n, \boldsymbol{\Lambda}, \eta) = \boldsymbol{\Sigma}_{\boldsymbol{X}} - \hat{\boldsymbol{\Sigma}}_n$*. The GMM estimators converge in probability to their true values as* $n \to \infty$*:*

$$\left(\hat{\boldsymbol{\Lambda}}, \hat{\eta}\right) \xrightarrow{\mathbb{P}} (\boldsymbol{\Lambda}, \eta), \tag{3.5}$$

if the following conditions are satisfied:

(i) $\mathbb{E}[f(\mathbf{X}, \boldsymbol{\Lambda}, \eta)] = 0$ *only for* $(\boldsymbol{\Lambda}, \eta) = (\boldsymbol{\Lambda}_0, \eta_0)$*;*

(ii) The parameter space $\Theta \subset \mathbb{R}^{pd+1}$ *is compact;*

(iii) With probability one, $f(\mathbf{X}, \boldsymbol{\Lambda}, \eta)$ *is continuous at each* $(\boldsymbol{\Lambda}, \eta)$*;*

(iv) $\mathbb{E}\left[\sup_{(\boldsymbol{\Lambda},\eta)\in\Theta} \|f(\mathbf{X}, \boldsymbol{\Lambda}, \eta)\|\right] < \infty$.

Theorem 3.3.3 (Asymptotic normality of GMM estimators)**.** *Let* $f(\mathbf{X}_n, \boldsymbol{\Lambda}, \eta) = \boldsymbol{\Sigma}_{\boldsymbol{X}} - \hat{\boldsymbol{\Sigma}}_n$*. The GMM estimators are asymptotically normal if the following conditions are satisfied:*

(i) $\left(\hat{\boldsymbol{\Lambda}}, \hat{\eta}\right)$ *is consistent;*

(ii) The parameter space $\Theta \subset \mathbb{R}^{pd+1}$ *is compact;*

(iii) $f(\mathbf{X}, \boldsymbol{\Lambda}, \eta)$ *is continuously differentiable in some neighborhood of* $(\boldsymbol{\Lambda}_0, \eta_0)$ *with probability one;*

(iv) $\mathbb{E}\left[\|f(\mathbf{X}_i, \boldsymbol{\Lambda}, \eta)\|^2\right] < \infty$*;*

(v) $\sup_{(\boldsymbol{\Lambda},\eta)\in\Theta} \left\|\nabla_{(\boldsymbol{\Lambda},\eta)} \boldsymbol{\Sigma}_{\boldsymbol{X}}\right\| < \infty$*;*

(vi) The matrix $\left(\nabla_{(\boldsymbol{\Lambda},\eta)} \boldsymbol{\Sigma}_{\boldsymbol{X}}\right)^\top \nabla_{(\boldsymbol{\Lambda},\eta)} \boldsymbol{\Sigma}_{\boldsymbol{X}}$ *is non-singular.*

The following lemma states that we can obtain a proxy of the true Gaussian factors Z_k (for $k = 1, \ldots, p$) by taking a weighted combination of the observable variables $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$.

Lemma 3.3.4 (Optimal weights)**.** *Let* $\boldsymbol{w}_{kd} = (w_{k1}, \ldots, w_{kd})^\top$ *be a weights vector of length* d *(for* $k = 1, \ldots, p$*), such that the weighted sum of the observed variables* $\hat{Z}_{kd} = \boldsymbol{w}_{kd}^\top \boldsymbol{X}$ *is a proxy of the Gaussian factor* Z_k*. Then one can obtain the unique optimal weight vector* $\boldsymbol{w}_{kd}^\star = (\boldsymbol{\lambda}_k^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} \boldsymbol{\lambda}_k)^{-1/2} \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} \boldsymbol{\lambda}_k$ *by minimising* $\mathbb{E}\{(\hat{Z}_{kd} - Z_k)^2\}$ *with respect to* $\boldsymbol{w}_{kd}$ *such that* $\mathrm{Var}(\hat{Z}_{kd}) = 1$.

Theorem 3.3.5 states that $\hat{Z}_{kd}^\star = \boldsymbol{w}_{kd}^{\star\top} \boldsymbol{X}$ will converge to its true value if the dimension is sufficiently large.

Theorem 3.3.5 (Convergence in $\mathscr{L}_2$ of Gaussian factor proxies)**.** *Assume Condition 3.3.1 is satisfied. For* $k = 1, \ldots, p$*, let* $\hat{Z}_{kd}^\star = \boldsymbol{w}_{kd}^{\star\top} \boldsymbol{X}$*, with* $\boldsymbol{w}_{kd}^\star = (\boldsymbol{\lambda}_k^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} \boldsymbol{\lambda}_k)^{-1/2} \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} \boldsymbol{\lambda}_k$*. Then the following holds as* $d \to \infty$*,*

$$\lim_{d\to\infty} \mathbb{E}\{(\hat{Z}_{kd}^\star - Z_k)^2\} = 0.$$

Theorem 3.3.6 states that $\hat{Z}_{knd}^\star$ will also converge to its true value.

Theorem 3.3.6 (Convergence in $\mathscr{L}_2$ of estimated Gaussian factor proxies)**.** *Let* $\hat{Z}^{\star}_{knd} = \hat{\boldsymbol{w}}^{\star\top}_{knd}\mathbf{X}_n$*, with* $\hat{\boldsymbol{w}}^{\star}_{knd} = (\hat{\boldsymbol{\lambda}}_k^{\top}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1}\hat{\boldsymbol{\lambda}}_k)^{-1/2}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1}\hat{\boldsymbol{\lambda}}_k$*. Let* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1}$ *and* $\hat{\boldsymbol{\lambda}}_k$ *be the estimated counterparts of* $\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$ *and* $\boldsymbol{\lambda}_k$*, respectively. Then, for* $k = 1, \ldots, p$

$$\lim_{d\to\infty}\lim_{n\to\infty}\mathbb{E}\{(\hat{Z}^{\star}_{knd} - Z_k)^2\} = 0.$$

Notice that in Theorem 3.3.6, the order of the limits is crucial to ensure convergence takes place.

Denote

$$\hat{\varepsilon}_{jd} = \frac{1}{\gamma_j}\left(X_j - \sum_{k=1}^{p}\lambda_{jk}\hat{Z}^{\star}_{kd}\right).$$

A first step is to show in Lemma 3.3.7 that the estimated proxy for the latent factor $\bar{U}_{nd} = \sum_{j=1}^{d}\hat{\varepsilon}_{jnd}/d$ converges towards the proxy for the latent factor $\bar{U}_d = \sum_{j=1}^{d}\hat{\varepsilon}_{jd}/d$. Subsequently, in Theorem 3.3.8 we show that $\bar{U}_d$ converges in probability to a monotone function of the true latent factor, if the dimension is sufficiently large.

Lemma 3.3.7. *Let* $\bar{U}_{nd} = \sum_{j=1}^{d}\hat{\varepsilon}_{jnd}/d$*, and let* $\bar{U}_d = \sum_{j=1}^{d}\varepsilon_{jd}/d$*. Then*

$$\bar{U}_{nd} \xrightarrow{\mathbb{P}} \bar{U}_d, \qquad \text{as } n \to \infty.$$

Theorem 3.3.8 (Convergence in probability to a monotonic function of the factor V as $d \to \infty$)**.** *Let* $\bar{U}_{nd} = \sum_{j=1}^{d}\hat{\varepsilon}_{jnd}/d$*, and let* $\bar{U}_d = \sum_{j=1}^{d}\varepsilon_{jd}/d$*. Under the assumptions of Theorem 5 in Krupskii and Joe (2021) and Theorem 3.3.5,*

$$\bar{U}_{nd} \xrightarrow{\mathbb{P}} m\left(V\right),$$

as $n \to \infty$ *and* $d \to \infty$*, where* $m\left(\cdot\right)$ *is a monotone function.*

The next result shows that the observed variables in (3.1) are asymptotically independent so this model is not suitable for modeling data with strong tail dependence.

Theorem 3.3.9 (Asymptotic independence of $\boldsymbol{X}$)**.** *If* $0 < \gamma_l < 1$*,* $0 < \gamma_m < 1$*, and* $\gamma_l(1-\gamma_m^2)^{1/2} \neq \gamma_m(1-\gamma_l^2)^{1/2}$ *for* $1 \leq l < m \leq d$*, then*

$$\lim_{z\to-\infty}\frac{P(X_l < z, X_m < z)}{\Phi(z)} = 0,$$

where $\Phi(\cdot)$ *is the cdf of a standard normal distribution. This result implies that the copula linking* (X_l, X_m) *has no lower tail dependence. Similar result holds for the upper tail.*

3.4 Simulation study

In this section, we perform a simulation study with the goal to evaluate the performance of the estimation methods for the model (3.1). The numerical implementation techniques that can be used to efficiently compute the presented estimators of the model are discussed in Section 3.4.1, Section 3.4.2 sets up the simulation experiment for which the results are presented in Section 3.4.3.

3.4.1 Numerical implementation

The main hurdle in applying the calculations in the previous section is in obtaining the inverse of the correlation matrix, $\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$. Such an inversion becomes computationally infeasible whenever the dimension d is too large. As mentioned in the proof of Theorem 3.3.5, the following expression can be obtained using Woodbury's formula (Chapter 2 in Van Loan and Golub, 1996):

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} = &\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} \\ &- \left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}^{\top}\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\Lambda}+\mathbf{I}_p\right)^{-1}\boldsymbol{\Lambda}^{\top}\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}.\end{aligned}$$

As one can see, the bottleneck lies in the inversion of the symmetric $d \times d$ matrix $\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}$. However, we can find that $(\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}\boldsymbol{\Gamma}^{\frac{1}{2}})^{-1} = \{(y_{j,k})\}_{j,k=1}^{d}$, where

$$\begin{aligned}y_{jj} &= \frac{1+(d-2)\,\eta}{\gamma_j^2\,(1+(d-2)\,\eta-(d-1)\,\eta^2)} \\ y_{jk} &= \frac{-\eta}{\gamma_j\gamma_k\,(1+(d-2)\,\eta-(d-1)\,\eta^2)},\end{aligned}$$

where $j \neq k$ for $j, k = 1, \ldots, d$. Using this closed-form expression for the inverse greatly improves computational efficiency, as we no longer have to deal with the inversion of a $d \times d$ matrix. Indeed, the remaining computations only require the inversion of the $p \times p$ matrix $\boldsymbol{\Lambda}^{\top}(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2})^{-1}\boldsymbol{\Lambda} + \mathbf{I}_p$, for which in practical applications $p \ll d$.

3.4.2 Simulation setup

We assume that we have at hand a sample of size n and dimension d which admits the factor model in (3.1) with two Gaussian factors and residual one-factor copula dependence:

$$X_{ij} = \lambda_{j1}Z_{i1} + \lambda_{j2}Z_{i2} + \gamma_j\varepsilon_{ij}, \tag{3.6}$$

for $i = 1, \ldots, n$, and $j = 1, \ldots, d$. In addition, let $\lambda_{j1} \sim U(0.4, 0.8)$, and $\lambda_{j2} \sim U(0.2, 0.6)$. The following simulation procedure is used to generate the Gaussian factors and the residuals ε_{ij}:

1. Draw a sample of size n for $Z_{i1} \sim N(0,1)$, $Z_{i2} \sim N(0,1)$, and $W_{i1}, \ldots, W_{id}, V_i \sim U(0,1)$ with $i = 1, \ldots, n$.
2. For each $j = 1, \ldots, d$, let $U_{ij} = C_{U|V}^{-1}(W_{ij}\,|V_i)$.
3. Set $\varepsilon_{ij} = \Phi^{-1}(U_{ij})$, where $\Phi^{-1}(\cdot)$ is the quantile function of the Gaussian cdf.

In a subsequent step, the factor model parameters are estimated using GMM, with parameter λ_{11} fixed at its true value to make the model parameters identifiable. Then, the factors are estimated using the weighted variables approach, where the weights are computed as in Theorem 3.3.6. Lastly, the factor copula parameter is estimated using standard maximum likelihood.

This experiment is repeated for 1000 independent simulations. For every experiment, one of the following settings is used for the sample size n, the dimension d, the copula parameter θ, and the copula family C:

- The sample size is set to $n \in \{250, 500, 750, 1000, 2500, 5000\}$.
- The dimension is set to $d \in \{10, 50, 100, 250, 500\}$.
- The copula parameter is calculated using Kendall's tau, $\tau_K = 0.50$.
- The family of bivariate copulas are Clayton (C_{Cl}), and t_4 (C_{t_4}).

Next, over all simulation results, we calculate the mean and variance of the correlation between the true and the estimated Gaussian factors, and the Pearson correlation coefficient between the estimated copula factor $\hat{V}$, and the true copula factor V. It is easy to see that minimising the $\mathscr{L}_2$ norm between the true and the estimated factor is equivalent to maximising the correlation between the true and the estimated factor. Lastly, we calculate the average bias of the copula parameter and its variance over all simulations. In the next section, we report these estimation results.

3.4.3 Simulation results

The results of the simulation experiments are presented in Tables 3.1 and 3.2. As one can see in both tables, for fixed n, the correlation between the estimated and the true factors increases as d becomes larger. However, for $n = 250$ and $n = 500$, there is a small drop in the correlation as d increases from 500 to 1000. Nevertheless, when n increases to 750, this is no longer the case. Similarly, for the bias of the estimated copula parameter reported in the last column of the tables, we can see that, for each n, there is a decrease in the bias as the dimension d becomes larger. However, as d increases from 500 to 1000, the bias starts to increase again. On the other hand, as the sample size n increases, this decrease becomes less important.

3.5 Data application

In this section, we apply the model outlined in the paper to a real-world dataset. We consider three different sectors in the S&P500 between January 3rd (2017) and December 30th (2020), which accounts for 1006 observations. The sectors considered are Industrials (71 stocks), Information Technology (74 stocks), and Financials (64 stocks) as classified by the Global Industry Classification Standard. Three stocks (CARR, IR, and OTIS) are excluded from the Industrials sector, as they were only included in the S&P500 in 2020.

To model the univariate time series, we use an AR(1)-GARCH(1,1) process for the log-returns $s_{j,t} = \log(S_{j,t}/S_{j,t-1})$ of stock price $S_{j,t}$, for each individual stock $j = 1, \ldots, d$ and each time period $t = 1, \ldots, T$. The univariate time series model can be written as

$$s_{j,t} = \mu_j + \phi_j s_{j,t-1} + \kappa_{j,t}\xi_{j,t}$$

Table 3.1: Simulation results for 1000 independent runs from a two-factor model with residual one-factor copula dependence with Clayton linking copulas. Loadings on the first factors are uniformly chosen between 0.4 and 0.8, while loadings on the second factor are uniformly chosen between 0.2 and 0.6.

Clayton copula with $\tau_K = 0.5$					
n	d	$\rho(Z_1, \hat{Z}_{1nd})$ (s.d)	$\rho(Z_2, \hat{Z}_{2nd})$ (s.d)	$\rho(V, \hat{V}_{nd})$ (s.d)	Bias($\hat{\theta}$) (s.d.)
250	10	0.84 (0.05)	0.68 (0.10)	0.82 (0.09)	0.13 (0.38)
	50	0.95 (0.01)	0.92 (0.01)	0.92 (0.01)	-0.07 (0.20)
	100	0.97 (0.01)	0.95 (0.01)	0.95 (0.01)	-0.10 (0.19)
	250	0.98 (0.01)	0.98 (0.01)	0.97 (0.01)	-0.07 (0.19)
	500	0.98 (0.03)	0.97 (0.03)	0.98 (0.01)	-0.07 (0.19)
500	10	0.86 (0.04)	0.74 (0.06)	0.85 (0.04)	0.25 (0.22)
	50	0.96 (0.01)	0.93 (0.01)	0.92 (0.01)	-0.07 (0.15)
	100	0.97 (< 0.01)	0.95 (0.01)	0.95 (0.01)	-0.05 (0.14)
	250	0.99 (< 0.01)	0.98 (0.01)	0.98 (< 0.01)	-0.04 (0.14)
	500	0.99 (0.02)	0.98 (0.02)	0.98 (< 0.01)	-0.05 (0.14)
750	10	0.86 (0.03)	0.75 (0.04)	0.86 (0.01)	0.29 (0.16)
	50	0.96 (< 0.01)	0.93 (0.02)	0.93 (< 0.01)	-0.06 (0.12)
	100	0.97 (< 0.01)	0.95 (< 0.01)	0.95 (< 0.01)	-0.04 (0.12)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.98 (< 0.01)	-0.03 (0.12)
	500	0.99 (0.02)	0.98 (0.02)	0.99 (< 0.01)	-0.05 (0.12)
1000	10	0.87 (0.03)	0.76 (0.03)	0.86 (0.01)	0.29 (0.14)
	50	0.96 (< 0.01)	0.93 (< 0.01)	0.93 (< 0.01)	-0.06 (0.10)
	100	0.97 (< 0.01)	0.95 (< 0.01)	0.96 (< 0.01)	-0.04 (0.10)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.98 (< 0.01)	-0.02 (0.10)
	500	0.99 (0.01)	0.98 (0.01)	0.99 (< 0.01)	-0.05 (0.10)
2500	10	0.88 (0.02)	0.78 (0.02)	0.87 (0.01)	0.31 (0.09)
	50	0.96 (< 0.01)	0.93 (< 0.01)	0.93 (< 0.01)	-0.04 (0.07)
	100	0.98 (< 0.01)	0.96 (< 0.01)	0.96 (< 0.01)	-0.03 (0.06)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.98 (< 0.01)	-0.02 (0.06)
	500	0.99 (0.01)	0.98 (0.01)	0.99 (< 0.01)	-0.04 (0.07)
5000	10	0.88 (0.01)	0.78 (0.01)	0.87 (< 0.01)	0.31 (0.06)
	50	0.96 (< 0.01)	0.93 (< 0.01)	0.93 (< 0.01)	-0.04 (0.05)
	100	0.98 (< 0.01)	0.96 (< 0.01)	0.96 (< 0.01)	-0.03 (0.05)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.98 (< 0.01)	-0.01 (0.05)
	500	0.99 (0.01)	0.98 (0.01)	0.99 (< 0.01)	-0.04 (0.05)

$$\kappa_{j,t}^2 = \omega_j + \alpha_j s_{j,t-1}^2 + \beta_j \kappa_{j,t-1}^2,$$

where μ_j is the unconditional mean, and ϕ_j gives, for stock j, the dynamic impact of a change in the log-returns in period $t-1$. Furthermore, $\kappa_{j,t}$ introduces, for stock j, non-constant volatility to the process. The latter is a function of the average volatility ω_j, the dynamic impact α_j of a change in the previous period of the squared series, and β_j captures the impact of a change in the volatility at time $t-1$.

Table 3.2: Simulation results for 1000 independent runs from a two-factor model with residual one-factor copula dependence with t_4 linking copulas. Loadings on the first factors are uniformly chosen between 0.4 and 0.8, while loadings on the second factor are uniformly chosen between 0.2 and 0.6.

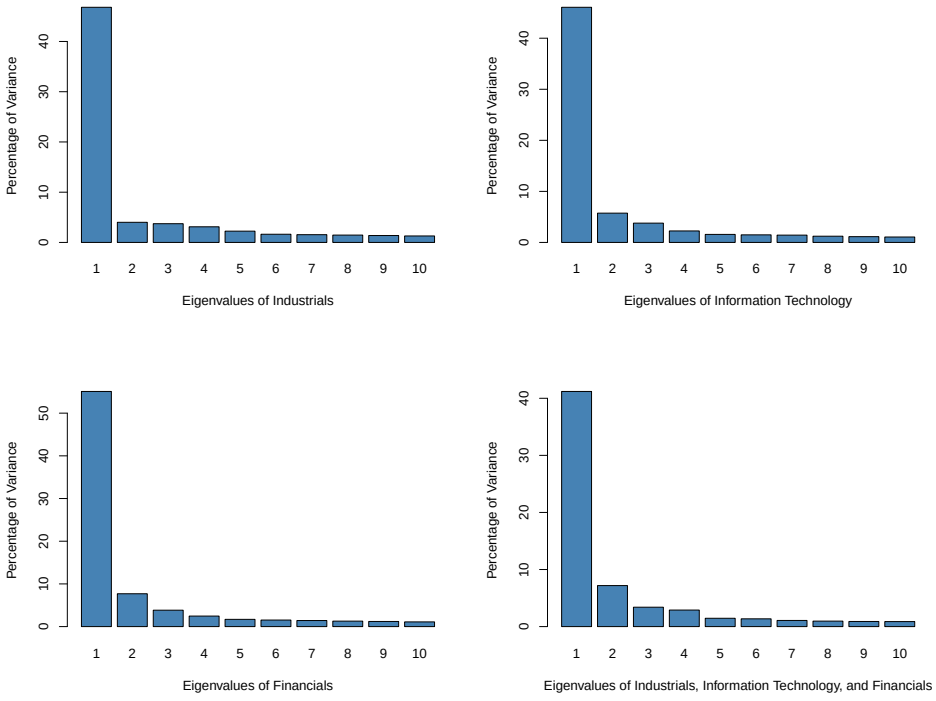
t_4 copula with $\tau_K = 0.5$					
n	d	$\rho(Z_1, \hat{Z}_{1nd})$ (s.d)	$\rho(Z_2, \hat{Z}_{2nd})$ (s.d)	$\rho(V, \hat{V}_{nd})$ (s.d)	Bias($\hat{\theta}$) (s.d.)
250	10	0.83 (0.05)	0.63 (0.11)	0.79 (0.13)	-0.05 (0.02)
	50	0.94 (0.02)	0.90 (0.03)	0.93 (0.02)	-0.02 (0.03)
	100	0.96 (0.02)	0.93 (0.03)	0.96 (0.02)	-0.02 (0.03)
	250	0.98 (0.02)	0.97 (0.02)	0.98 (0.01)	-0.02 (0.02)
	500	0.96 (0.05)	0.95 (0.06)	0.98 (0.02)	-0.02 (0.03)
500	10	0.84 (0.04)	0.70 (0.08)	0.86 (0.07)	0.02 (0.12)
	50	0.95 (0.01)	0.92 (0.01)	0.94 (0.01)	-0.01 (0.02)
	100	0.97 (0.01)	0.95 (0.01)	0.97 (< 0.01)	-0.01 (0.02)
	250	0.99 (0.01)	0.98 (0.01)	0.99 (< 0.01)	-0.01 (0.02)
	500	0.98 (0.03)	0.97 (0.03)	0.99 (< 0.01)	-0.01 (0.02)
750	10	0.85 (0.04)	0.73 (0.06)	0.88 (0.04)	0.04 (0.06)
	50	0.95 (< 0.01)	0.92 (< 0.01)	0.95 (0.01)	-0.01 (0.02)
	100	0.97 (< 0.01)	0.95 (< 0.01)	0.97 (< 0.01)	-0.01 (0.01)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.99 (< 0.01)	-0.01 (0.01)
	500	0.99 (0.02)	0.98 (0.02)	0.99 (< 0.01)	-0.01 (0.01)
1000	10	0.86 (0.04)	0.74 (0.04)	0.88 (0.02)	0.05 (0.02)
	50	0.95 (< 0.01)	0.92 (< 0.01)	0.95 (< 0.01)	-0.01 (0.01)
	100	0.97 (< 0.01)	0.95 (< 0.01)	0.97 (< 0.01)	-0.01 (0.01)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.99 (< 0.01)	-0.01 (0.01)
	500	0.99 (0.01)	0.98 (0.01)	0.99 (< 0.01)	-0.01 (0.01)
2500	10	0.87 (0.02)	0.77 (0.02)	0.89 (0.02)	0.05 (0.01)
	50	0.96 (< 0.01)	0.93 (< 0.01)	0.95 (< 0.01)	-0.01 (0.01)
	100	0.97 (< 0.01)	0.95 (< 0.01)	0.97 (< 0.01)	0.00 (0.01)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.99 (< 0.01)	0.00 (0.01)
	500	0.99 (0.01)	0.98 (0.01)	0.99 (< 0.01)	-0.01 (0.01)
5000	10	0.88 (< 0.01)	0.77 (< 0.01)	0.89 (< 0.01)	0.05 (< 0.01)
	50	0.96 (< 0.01)	0.93 (< 0.01)	0.95 (< 0.01)	0.00 (< 0.01)
	100	0.97 (< 0.01)	0.95 (< 0.01)	0.97 (< 0.01)	0.00 (< 0.01)
	250	0.99 (< 0.01)	0.98 (< 0.01)	0.99 (< 0.01)	-0.01 (< 0.01)
	500	0.99 (< 0.01)	0.98 (< 0.01)	0.99 (< 0.01)	-0.01 (< 0.01)

Lastly, error terms $\xi_{j,t}$ are i.i.d. Skew-t distributed with ν degrees of freedom and skewness parameter γ. After estimating the univariate parameters, we transform the estimated residuals to a uniform scale.

Next, we want to identify the number of underlying factors. In Section 2.4 of Oh and Patton (2017) it is shown that, under regularity conditions, one can use a scree plot based on the eigenvalues of the rank correlation matrix to determine the

appropriate number of factors. The scree plot for the AR(1)-GARCH(1,1) filtered data transformed to uniform ranks is given in Figure 3.2. One can see that for the Industrials sector, the majority of the variance in the data is explained by one factor, while for the Information Technology and Financials sectors, two common factors appear to be more appropriate to capture the variance in the data. Thus, when we look at the number of factors required to simultaneously model the Industrials, Information Technology and Financials sectors, it is no surprise that the number of factors appears to be between 2 and 4.

Figure 3.2: Scree plot of the 10 largest eigenvalues of the rank correlation matrix for the AR(1)-GARCH(1,1) filtered observations in the Industrials, Information Technology, and Financials stocks of the S&P500.



Next, we model the multivariate dependence between the different log-returns of stock prices. First, we transform the data to normal scores and look at the empirical semi-correlation of the data. Semi-correlation can be defined as the correlation in the lower tail ($\rho_{<0}$) or upper tail ($\rho_{>0}$) of the data. We can see in Table 3.3 that the average absolute semi-correlation indicates that the data in each of the three sectors is skewed to the left. The absolute semi-correlation is calculated to account for the seven negative values of upper semi-correlation in the Information Technology sector. However, these are small enough such that they do not affect the average up to two decimals. This is consistent with stylized facts in the financial literature that show stronger correlation during downturns. We therefore expect that a copula with stronger dependence in the lower tail is more appropriate to model these data. For comparison, we use different copula models to see which one fits these characteristics best.

First, we model dependence within each sector separately using only one Gaus-

Table 3.3: Sector-wise average absolute lower semi-correlation ($\rho_{<0}$) and upper semi-correlation ($\rho_{>0}$) using three different sectors in the S&P500.

Sector	$\lvert\rho_{<0}\rvert$	$\lvert\rho_{>0}\rvert$
Industrials	0.38	0.25
Information Technology	0.39	0.21
Financials	0.44	0.34

sian factor, and a Clayton, Frank, and normal one-factor copula model to capture the residual dependence. Because the lower tail dependence does not appear to be very strong, we also use a reflected Gumbel copula model to capture the residual dependence. Then, we subsequently use two and three Gaussian factors while using the same one-factor copula models. These estimation results for each sector are presented in Table 3.4. As one can see in the model with one Gaussian factor, all three sectors establish a moderate degree of dependence between its constituent stocks. This is not surprising, as linear dependence is already captured by the Gaussian factor. Adding a second Gaussian factor to the model does not drastically change the parameter estimates. Similarly, when pairing industry sectors together in a model with two Gaussian factors, we can see that there is a moderate degree of dependence between the stocks in each of the pairs of industry sectors. Lastly, when using three Gaussian factors, we can see that the estimated parameter for each of the three copula models is somewhat higher than the estimated dependence parameters in the model with two Gaussian factors.

To evaluate the performance of our model, we use the estimated parameters to simulate a sample of size 1,000,000. We then calculate the semi-correlations of this simulated sample, and calculate the average (absolute) bias with respect to the empirical semi-correlations in Table 3.3. These results are reported in Table 3.5 and 3.6. As we can see, the average absolute bias of the lower semi-correlation, $\mathrm{b}(\lvert\rho_{<0}\rvert)$, is smallest for the model using reflected Gumbel linking copulas in the model with one Gaussian factor. This shows that our model is capable of capturing multivariate skewness in high-dimensional data. The average absolute bias of the upper semi-correlation is somewhat lower for the Frank and Normal copula, but the difference with that of the reflected Gumbel copula is relatively small, except for the model with one Gaussian factor with the Financials sector. Lastly, we also modelled the data using a multivariate Student t_ν copula with a three-factor covariance structure. However, the estimated degrees of freedom parameter is close to 20, which makes it indistinguishable from a normal copula model.

3.6 Discussion

Applying Gaussian factor models to high-dimensional data requires extensive computational power. Moreover, Gaussian factor models do not capture multivariate skewness in the data. In this paper, we provided a method that exploits the high-dimensionality of the data to accurately estimate the unobserved factors and simultaneously allows us to model multivariate skewness in the data. The proposed model is flexible, as it allows to separately model the margins from the dependence structure. We proved that our model works for a factor model with up to three

Table 3.4: Estimated copula parameters for three different sectors in the S&P500. Top: one-factor model, middle: two-factor model, bottom: three-factor model.

Sectors	r-Gumbel	Clayton	Frank	Normal	d
Industrials	1.46	0.73	3.51	0.50	71
Information Technology	1.50	0.79	3.75	0.52	74
Financials	1.74	1.12	4.98	0.63	64
Industrials-IT	1.35	0.56	2.86	0.43	145
Industrials-Financials	1.49	0.77	3.61	0.51	135
IT-Financials	1.46	0.73	3.48	0.50	138
Industrials-IT-Financials	1.61	0.91	4.32	0.57	209

Gaussian factors. Simulation studies showed that the accuracy of our estimations improves as the sample size and dimension increase. To illustrate the model, we applied it to a dataset of three constituent sectors of the S&P500 stock data and found evidence that the stocks in the three sectors are skewed, consistent with findings in the finance literature.

There are different avenues for future research. One direction is to include fast inference models that allow for tail dependence in factor models, or for a nested structure in the factor copula dependence. Another option is to include multiple variables that capture dependence between the data using different one-factor copula models.

Table 3.5: Absolute average bias for simulated semi-correlations, with $n = 1000000$, and dimension matching that of each sector. Parameters used in the simulation are the estimated parameters for each sector.

	r-Gumbel		Clayton		Frank		Normal	
Sectors	b($\lvert\rho_{<0}\rvert$)	b($\lvert\rho_{>0}\rvert$)	b($\lvert\rho_{<0}\rvert$)	b($\lvert\rho_{>0}\rvert$)	b($\lvert\rho_{<0}\rvert$)	b($\lvert\rho_{>0}\rvert$)	b($\lvert\rho_{<0}\rvert$)	b($\lvert\rho_{>0}\rvert$)
Ind	0.06	0.07	0.07	0.10	0.16	0.05	0.14	0.05
IT	0.07	0.07	0.07	0.09	0.18	0.06	0.16	0.06
Fin	0.06	0.07	0.06	0.12	0.14	0.06	0.11	0.05
Ind-IT	0.08	0.06	0.09	0.07	0.16	0.05	0.15	0.05
Ind-Fin	0.11	0.09	0.12	0.10	0.18	0.07	0.17	0.06
IT-Fin	0.10	0.06	0.10	0.08	0.17	0.06	0.17	0.05
Ind-IT-Fin	0.06	0.06	0.06	0.09	0.15	0.05	0.13	0.06

Table 3.6: Average bias for simulated semi-correlations, with $n = 1000000$, and dimension matching that of each sector. Parameters used in the simulation are the estimated parameters for each sector.

	r-Gumbel		Clayton		Frank		Normal	
Sectors	b($\rho_{<0}$)	b($\rho_{>0}$)	b($\rho_{<0}$)	b($\rho_{>0}$)	b($\rho_{<0}$)	b($\rho_{>0}$)	b($\rho_{<0}$)	b($\rho_{>0}$)
Ind	-0.04	-0.09	-0.04	-0.09	-0.16	-0.02	-0.14	-0.01
IT	-0.04	-0.07	-0.04	-0.07	-0.18	0.01	-0.16	0.02
Fin	-0.01	-0.12	-0.01	-0.12	-0.14	-0.03	-0.11	-0.01
Ind-IT	-0.07	-0.03	-0.08	-0.05	-0.16	0.00	-0.15	0.01
Ind-Fin	-0.11	-0.08	-0.11	-0.10	-0.18	-0.06	-0.17	-0.05
IT-Fin	-0.09	-0.04	-0.09	-0.07	-0.17	-0.02	-0.17	-0.01
Ind-IT-Fin	-0.01	-0.04	-0.01	-0.08	-0.15	-0.00	-0.13	0.02

Chapter 4

Multi-purpose open-end monitoring procedures for multivariate observations based on the empirical distribution function

This chapter and the corresponding Appendix B have been published in the *Journal of Time Series Analysis* (Holmes et al., 2023). The corresponding R code has been included in the package (Kojadinovic and Verhoijsen, 2022) available on CRAN.

Abstract

We propose nonparametric open-end sequential testing procedures that can detect all types of changes in the contemporary distribution function of possibly multivariate observations. Their asymptotic properties are theoretically investigated under stationarity and under alternatives to stationarity. Monte Carlo experiments reveal their good finite-sample behavior in the case of continuous univariate, bivariate and trivariate observations. A short data example concludes the work.

4.1 Introduction

From an historical perspective, monitoring is often associated with *control charts* also known as *Shewart charts*. Such graphical tools central to *statistical process control* (see, e.g., Lai, 2001; Montgomery, 2007, for an overview) are usually calibrated in terms of the so-called *average run length* (ARL) controlling how many monitoring steps are necessary on average before the data generating process is declared *out of control*. The fact that this conclusion (that is, that the probabilistic properties of the monitored observations have changed) is reached with probability one could be regarded as a drawback of this type of procedure, in particular if *false alarms* are very costly. To remedy this situation, Chu et al. (1996) have proposed to treat the issue of monitoring from the point of view of statistical testing. The main advantage is that, when observations arise from a stationary time series, monitoring procedures *à la* Chu et al. (1996) will lead to the conclusion that a change has occurred in the data generating process only with a small probability α controlled by the user. For a recent nicely written literature review comparing monitoring as carried out in statistical process control to approaches based on statistical tests *à la* Chu et al. (1996), we refer the reader to the introduction of Gösmann et al. (2022).

In addition to being statistical tests, the monitoring procedures investigated in this work are nonparametric and deal with d-dimensional observations, $d \geq 1$. In that respect, as we continue, we use the superscript $^{[\ell]}$ to denote the ℓth coordinate of a vector (for instance, $\boldsymbol{x} = (x^{[1]}, \ldots, x^{[d]}) \in \mathbb{R}^d$). As is customary in the sequential testing literature, we assume that we have at hand $m \geq 1$ observations from the initial data generating process. Monitoring starts immediately thereafter. To be more precise, we assume that we have at our disposal a stretch $\boldsymbol{X}_i = (X_i^{[1]}, \ldots, X_i^{[d]})$, $i \in \{1, \ldots, m\}$, from a d-dimensional stationary time series with unknown contemporary distribution function (d.f.) F given by $F(\boldsymbol{x}) = \mathbb{P}(X_1^{[1]} \leq x^{[1]}, \ldots, X_1^{[d]} \leq x^{[d]}) = \mathbb{P}(\boldsymbol{X}_1 \leq \boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$. These available observations will be referred to as the *learning sample* as we continue. Once the monitoring starts, new observations $\boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots$ arrive sequentially and the aim is to issue an alarm as soon as possible if there is evidence that the contemporary distribution of the most recent observations is no longer equal to F.

Most approaches in the literature are of a *closed-end* nature: the monitoring eventually stops if stationarity is not rejected after the arrival of a final observation $\boldsymbol{X}_n$, $n > m$. Our focus in this work is on the more difficult scenario in which the monitoring can in principle continue indefinitely: this is called the *open-end* setting. From a practical perspective, it can be argued that the fact that the *monitoring horizon* n does not need to be specified is a great advantage of open-end procedures. The price to pay for open-endness is however a significantly more complicated theoretical setting. Indeed, as discussed for instance in Remark 2.2 of Gösmann et al.

(2021), while the asymptotics of closed-end procedures can usually be derived using functional central limit theorems, such results are insufficient in the open-end case and need to be either combined with Háyék-Réyni type inequalities (see, e.g., Kirch and Weber, 2018, and the references therein) or replaced by approximations of the form of forthcoming Condition 4.2.5 (see Aue and Horváth, 2004; Aue et al., 2006).

The null hypothesis of the sequential testing procedures studied in this work is

$$H_0 : \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots, \text{ is a stretch from a stationary time series.} \tag{4.1}$$

When $d = 1$, starting from the work of Gösmann et al. (2021), Holmes and Kojadinovic (2021) have recently introduced a detection procedure that is particularly sensitive to changes in the mean. Because it uses the *retrospective cumulative sum (CUSUM) statistic* as *detector*, it turns out to be more powerful then existing procedures as long as changes do not occur at the very beginning of the monitoring. As noted in Section 6 of the latter reference, this approach can be adapted to obtain alternative procedures that are particularly sensitive for instance to changes in the variance or some other moments. The goal of this work is to generalize the method of Holmes and Kojadinovic (2021) in order to obtain open-end monitoring procedures that can be sensitive simultaneously to all types of changes in the d.f. F. Although such procedures already exist in a closed-end setting (see, e.g., Kojadinovic and Verdier, 2021), to the best of our knowledge, they are unavailable in the open-end setting.

This paper is organized as follows. In the second section, starting from the work of Gösmann et al. (2021) and Holmes and Kojadinovic (2021), we propose a detector that can be sensitive to all types of changes in the contemporary d.f. of multivariate observations. We additionally introduce a suitable threshold function and study the asymptotics of the resulting monitoring procedure under H_0 in (4.1) and under sequences of alternatives to H_0. In the third section, we focus on the case of continuous observations and provide additional asymptotic results under the null. In the fourth section, using *asymptotic regression models*, we address the estimation of high quantiles of the distributions appearing in the asymptotic results under H_0 which are necessary in practice to carry out the sequential tests. The fifth section summarizes the results of numerous Monte Carlo experiments for $d \in \{1, 2, 3\}$ whose aim is to study the finite-sample behavior of the monitoring procedures under H_0 and under alternatives to H_0. A data example and concluding remarks are gathered in the last section.

Unless mentioned otherwise, all convergences are as $m \to \infty$. Also, as there are many discrete intervals appearing in this work, we will conveniently use the notation $[\![j, k]\!]$, $[\![j, k[\![$, $]\!]j, k]\!]$, and $]\!]j, k[\![$ for the sets of integers $\{j, \ldots, k\}$, $\{j, \ldots, k-1\}$, $\{j-1, \ldots, k\}$, and $\{j-1, \ldots, k-1\}$, respectively. Note that all mathematical proofs are gathered in a series of appendices and a non-optimized implementation of the monitoring procedures studied in this work is available in the package `npcp` (Kojadinovic and Verhoijsen, 2022) for the R statistical environment (R Core Team, 2022).

4.2 A first detector, the threshold function and related asymptotics

4.2.1 A first detector and the threshold function

One of the two main ingredients of a monitoring procedure *à la* Chu et al. (1996) is a statistic, also called a *detector*, that is potentially computed after the arrival of every new observation $\boldsymbol{X}_k$, $k > m$. This statistic is typically positive and quantifies some type of departure from stationarity. Once computed, it is compared to a positive threshold, possibly also depending on k. If greater, evidence against H_0 in (4.1) is deemed significant and the monitoring stops. Otherwise, a new observation is collected.

For sensitivity to changes in the mean of univariate ($d = 1$) observations, Holmes and Kojadinovic (2021) used among others the detector

$$R_m(k) = \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} |\bar{X}^{[1]}_{1:j} - \bar{X}^{[1]}_{j+1:k}|, \qquad k \geq m+1, \tag{4.2}$$

where $\bar{X}^{[1]}_{j:k} = \frac{1}{k-j+1} \sum_{i=j}^{k} X^{[1]}_i$, $1 \leq j \leq k$. Some thought reveals that, for any fixed $k \geq m+1$, $R_m(k)$ is akin to the so-called *retrospective CUSUM statistic* frequently used in offline change-point detection tests (see, e.g., Csörgő and Horváth, 1997; Aue and Horváth, 2013).

When $d \geq 1$, to be sensitive to changes in the d.f. at a fixed point $\boldsymbol{x} = (x^{[1]}, \ldots, x^{[d]}) \in \mathbb{R}^d$, a straightforward adaptation of the previous approach would be to compute (4.2) from the stretch of univariate observations $\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}), \ldots, \mathbf{1}(\boldsymbol{X}_m \leq \boldsymbol{x}), \mathbf{1}(\boldsymbol{X}_{m+1} \leq \boldsymbol{x}), \ldots, \mathbf{1}(\boldsymbol{X}_k \leq \boldsymbol{x})$, where inequalities between vectors are to be understood componentwise. The detector R_m can then be equivalently expressed as

$$E^{\boldsymbol{x}}_m(k) = \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} |F_{1:j}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x})|, \qquad k \geq m+1, \tag{4.3}$$

where, for any integers $j, k \geq 1$,

$$F_{j:k}(\boldsymbol{x}) = \begin{cases} \dfrac{1}{k-j+1} \displaystyle\sum_{i=j}^{k} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}), & \text{if } j \leq k, \\ 0, & \text{otherwise,} \end{cases} \tag{4.4}$$

is the empirical d.f. of $\boldsymbol{X}_j, \ldots, \boldsymbol{X}_k$ evaluated at $\boldsymbol{x}$. Our aim is to extend the previous approach using $p \geq 1$ points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ in $\mathbb{R}^d$, where the integer p and the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ are chosen by the user. Let $\mathcal{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ and, for any $i \in \mathbb{N}$, let $\boldsymbol{Y}^{\mathcal{P}}_i = \big(\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_1), \ldots, \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_p)\big)$, which is a p-dimensional random vector. Combining the approach of Gösmann et al. (2021) with the one of Holmes and Kojadinovic (2021), the first detector considered in this work is defined by

$$D^{\mathcal{P}}_m(k) = \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}^{\mathcal{P}}_{1:j} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k}\|_{(\Sigma^{\mathcal{P}}_m)^{-1}}, \qquad k \geq m+1, \tag{4.5}$$

where, for any integers $j, k \geq 1$, $\boldsymbol{F}^{\mathcal{P}}_{j:k} = \big(F_{j:k}(\boldsymbol{x}_1), \ldots, F_{j:k}(\boldsymbol{x}_p)\big) \in \mathbb{R}^p$, $\Sigma^{\mathcal{P}}_m$ is an estimator (based on $\boldsymbol{Y}^{\mathcal{P}}_1, \ldots, \boldsymbol{Y}^{\mathcal{P}}_m$) of the long-run $p \times p$ covariance matrix

$$\Sigma^{\mathcal{P}} = \mathrm{Cov}(\boldsymbol{Y}^{\mathcal{P}}_1, \boldsymbol{Y}^{\mathcal{P}}_1) + \sum_{i=2}^{\infty} \{\mathrm{Cov}(\boldsymbol{Y}^{\mathcal{P}}_1, \boldsymbol{Y}^{\mathcal{P}}_i) + \mathrm{Cov}(\boldsymbol{Y}^{\mathcal{P}}_i, \boldsymbol{Y}^{\mathcal{P}}_1)\} \tag{4.6}$$

of the p-dimensional time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$ and, for any $\boldsymbol{y}\in\mathbb{R}^p$, $\|\boldsymbol{y}\|_M=\sqrt{(\boldsymbol{y}^\top M\boldsymbol{y})/p}$ denotes a weighted norm of $\boldsymbol{y}$ induced by a $p\times p$ positive-definite matrix M and the integer p.

Note that (4.5) can be equivalently rewritten as

$$D_m^{\mathcal{P}}(k)=\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}}-\bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma_m^{\mathcal{P}})^{-1}},\qquad k\geq m+1,\tag{4.7}$$

where, for any integers $1\leq j\leq k$, $\bar{\boldsymbol{Y}}_{j:k}^{\mathcal{P}}=\frac{1}{k-j+1}\sum_{i=j}^k\boldsymbol{Y}_i^{\mathcal{P}}$.

Remark 4.2.1. The role of the matrix $(\Sigma_m^{\mathcal{P}})^{-1}$ when using the Mahalanobis-like norm $\|\cdot\|_{(\Sigma_m^{\mathcal{P}})^{-1}}$ in (4.5) is, roughly speaking, to standardize and decorrelate under the null vectors of the form $\boldsymbol{F}_{1:j}^{\mathcal{P}}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}}$ before computing their L_2 norm (scaled by $1/\sqrt{p}$). As shall become clearer from Theorem 4.2.6 below, a consequence of that step is that a key limiting null distribution playing a central role in the testing procedure will not depend on the characteristics of the underlying time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ but only on the number of points p chosen by the user. The latter desirable property from a practical perspective is also the reason why we did not consider, instead of $D_m^{\mathcal{P}}$ in (4.5), alternative detectors that evaluate differences of empirical d.f.s at all the points in $\mathbb{R}^d$. One natural such alternative detector is

$$D_m^{\sup}(k)=\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\sup_{\boldsymbol{x}\in\mathbb{R}^d}|F_{1:j}(\boldsymbol{x})-F_{j+1:k}(\boldsymbol{x})|,\qquad k\geq m+1.\tag{4.8}$$

Such a detector was actually considered in a closed-end setting by Kojadinovic and Verdier (2021) and required in practice the use of bootstrapping when monitoring serially dependent observations. The major practical obstacle related to its use in an open-end setting will be discussed in Remark 4.2.13.

The second key ingredient of a monitoring procedure is a *threshold function*. In the considered open-end setting, given a significance level $\alpha\in(0,\frac{1}{2})$, the aim is to define a deterministic function $w:[1,\infty)\to(0,\infty)$ such that (ideally) under H_0 in (4.1),

$$\mathbb{P}\Big(D_m^{\mathcal{P}}(k)\leq w(k/m)\text{ for all }k>m\Big)=\mathbb{P}\left(\sup_{k>m}\frac{D_m^{\mathcal{P}}(k)}{w(k/m)}\leq 1\right)=1-\alpha,\tag{4.9}$$

where the supremum is over integers $k>m$. Because the detector $D_m^{\mathcal{P}}$ in (4.5) or in (4.7) can be regarded as a multivariate generalization of the detector R_m in (4.2), one can use the same reasoning as in Section 2 of Holmes and Kojadinovic (2021) to suggest that a meaningful threshold function in the considered case is

$$w(t)=q_{p,\eta}^{(1-\alpha)}t^{\frac{3}{2}+\eta},\qquad t\in[1,\infty),\tag{4.10}$$

where η is a positive real parameter and $q_{p,\eta}^{(1-\alpha)}>0$ is the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$, the weak limit of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}D_m^{\mathcal{P}}(k)$ under H_0, assuming that $\mathcal{L}_{p,\eta}$ is continuous. In that case, under H_0, by the Portmanteau theorem,

$$\lim_{m\to\infty}\mathbb{P}\left(\sup_{k>m}\frac{D_m^{\mathcal{P}}(k)}{w(k/m)}\leq 1\right)=\lim_{m\to\infty}\mathbb{P}\left(\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}D_m^{\mathcal{P}}(k)\leq q_{p,\eta}^{(1-\alpha)}\right)$$

$$= \mathbb{P}(\mathcal{L}_{p,\eta} \leq q_{p,\eta}^{(1-\alpha)}) = 1 - \alpha, \tag{4.11}$$

which can be regarded as an "asymptotic version" of (4.9). As far as η is concerned, as a consequence of Proposition 4.2.10 below, it needs to be chosen strictly positive so that $\mathcal{L}_{p,\eta}$ is almost surely finite. In practice, we follow the recommendation made in Holmes and Kojadinovic (2021) and set η to 0.001.

Remark 4.2.2. Proceeding for instance along the lines of Horváth et al. (2004), Fremdt (2015), Kirch and Weber (2018), Gösmann et al. (2021) or Holmes and Kojadinovic (2021), instead of w in (4.10), one could alternatively consider as a threshold function $\tilde{w}$ defined by $\tilde{w}(t) = \bar{w}_\gamma(t)w(t)$, $t \in [1, \infty)$, where

$$\bar{w}_\gamma(t) = \max\left\{\left(\frac{t-1}{t}\right)^\gamma, \epsilon\right\}, \qquad t \in [1, \infty),$$

with $\gamma \geq 0$ a real parameter and $\epsilon > 0$ a technical constant that can be taken very small in practice. The multiplication of a candidate threshold function by $\bar{w}_\gamma$ was initially considered in Horváth et al. (2004) and Aue and Horváth (2004) for the so-called *ordinary CUSUM* detector in order to study, under suitable alternatives, the limiting distribution of the detection delay (the time delay after which the detector exceeds the threshold function). From a practical perspective, as discussed in Holmes and Kojadinovic (2021), an appropriate choice of $\gamma \geq 0$ may improve the finite-sample performance of the sequential test at the beginning of the monitoring. The multiplication of a candidate threshold function by $\bar{w}_\gamma$ does not however affect the asymptotics of the underlying monitoring procedure.

4.2.2 Asymptotics under the null

One of the first assumptions required to be able to study the asymptotics (as $m \to \infty$, of the monitoring procedure based on $D_m^{\mathcal{P}}$ in (4.5) and w in (4.10)) concerns the long-run covariance matrix $\Sigma^{\mathcal{P}}$ in (4.6) of the p-dimensional time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$, under the null. As we shall see later in this section, it will be necessary to consider both its inverse $(\Sigma^{\mathcal{P}})^{-1}$ and its square root $(\Sigma^{\mathcal{P}})^{\frac{1}{2}}$. The following assumption guarantees that these two matrices exist (and are unique).

Condition 4.2.3 (On the long-run covariance matrix $\Sigma^{\mathcal{P}}$)**.** *Under H_0 in* (4.1)*, the long-run covariance matrix $\Sigma^{\mathcal{P}}$ in* (4.6) *of the p-dimensional (stationary) time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$ exists and is positive-definite.*

Remark 4.2.4. Under H_0 and when the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ consists of independent observations, the $p \times p$ elements of $\Sigma^{\mathcal{P}}$ in (4.6) are simply

$$\mathrm{Cov}\{\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}_i), \mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}_j)\} = F\big(\min(\boldsymbol{x}_i, \boldsymbol{x}_j)\big) - F(\boldsymbol{x}_i)F(\boldsymbol{x}_j), \qquad i, j \in [\![1, p]\!],$$

where min denotes the element-wise minimum operator. Hence, in the case of serially independent observations, by definition of positive-definiteness, Condition 4.2.3 will hold if the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ appearing in $\mathcal{P}$ are chosen such that any linear combination of the $\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}_i)$, $i \in [\![1, p]\!]$, has a strictly positive variance. A necessary condition for this is that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ all belong to the support of $\boldsymbol{X}_1$ and are all distinct. Since the law of $\boldsymbol{X}_1$ is unknown, the user could in practice rely on the

learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ to choose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$. If the learning sample seems to be a stretch from a discrete time series, a natural possibility consists of choosing $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ from a subset of frequently occurring observations. The choice of $\mathcal{P}$ when the observations in the learning sample seem to arise from a continuous time series will be discussed in Section 4.3.

As shall become clearer in the forthcoming paragraphs, studying the asymptotics under the null (of the monitoring procedure as $m \to \infty$) actually amounts to establishing the weak limit of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ under H_0 in (4.1). The following assumption on the time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$ (a type of "strong approximation" condition) is typical of the kinds of assumptions in the sequential change-point literature; see, e.g., Assumption 2.3 in Gösmann et al. (2021), Condition 3.1 in Holmes and Kojadinovic (2021) and the corresponding discussions in these references. Let $\|\cdot\|_2$ denote the Euclidean norm.

Condition 4.2.5 (Approximation)**.** *There exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *on which:*

- $(\boldsymbol{Y}_i^{\mathcal{P}})_{i\in\mathbb{N}}$ *is a* p*-dimensional stationary time series satisfying Condition 4.2.3,*
- *for each* $m \in \mathbb{N}$, $\boldsymbol{W}_{1,m}$ *and* $\boldsymbol{W}_{2,m}$ *are two independent* p*-dimensional standard Brownian motions,*

such that, for some $0 < \xi < \frac{1}{2}$,

$$\sup_{k>m} \frac{1}{(k-m)^{\xi}} \left\| \sum_{i=m+1}^{k} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(k-m) \right\|_2 = O_{\mathbb{P}}(1) \tag{4.12}$$

and

$$\frac{1}{m^{\xi}} \left\| \sum_{i=1}^{m} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{2,m}(m) \right\|_2 = O_{\mathbb{P}}(1). \tag{4.13}$$

We use the notation '$\rightsquigarrow$' to denote convergence in distribution (weak convergence) and I_p to denote the $p \times p$ identity matrix. The following result, proven in Appendix B.1, can be regarded as a multivariate extension of Theorem 3.3 of Holmes and Kojadinovic (2021).

Theorem 4.2.6. *Fix* $\eta > 0$*. Under Condition 4.2.5, if* $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ *then*

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p},$$

where $D_m^{\mathcal{P}}$ *is defined in* (4.5) *and* $\boldsymbol{W}$ *is a* p*-dimensional standard Brownian motion. In addition, the limiting random variable* $\mathcal{L}_{p,\eta}$ *is almost surely finite.*

Note that in Theorem 4.2.6 the supremum on the left is over integers k while the supremum on the right is over real numbers s, t.

Remark 4.2.7. It is important to note that the limiting random variable $\mathcal{L}_{p,\eta}$ depends neither on the characteristics of the underlying time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ (such as its dimension d, its serial dependence properties or the unknown d.f. F), nor on the user-chosen points $\mathcal{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ involved in the definition of $D_m^{\mathcal{P}}$. It only depends on the integer p and on the real η. The latter is due to the use of the Mahalanobis-like norm $\|\cdot\|_{(\Sigma_m^{\mathcal{P}})^{-1}}$ in (4.5) as hinted at in Remark 4.2.1. As shall become clearer below, an important practical consequence of this is that the monitoring procedure can be used as soon as it is possible to compute or estimate quantiles of $\mathcal{L}_{p,\eta}$ for the chosen parameters p and η. This important aspect will be investigated in Section 4.4 in more detail.

For a given serial dependence scenario under H_0 in (4.1), it is hoped that Condition 4.2.5 will hold for many different vectors of points $\mathcal{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$. The following proposition shows that this is for instance the case when the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is *strongly mixing* under H_0. Given a time series $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ and for any $j, k \in \mathbb{N}\cup\{+\infty\}$, denote by $\mathcal{M}_j^k$ the σ-field generated by $(\boldsymbol{Z}_i)_{j\le i\le k}$ and recall that the strong mixing coefficients corresponding to $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ are defined by

$$\alpha_r^{\boldsymbol{Z}} = \sup_{k\in\mathbb{N}} \sup_{A\in\mathcal{M}_1^k, B\in\mathcal{M}_{k+r}^{+\infty}} \big|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\big|, \qquad r\in\mathbb{N}.$$

The sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ is then said to be *strongly mixing* if $\alpha_r^{\boldsymbol{Z}} \to 0$ as $r\to\infty$.

The following result, proven in Appendix B.2, is a consequence of Theorem 4 of Kuelbs and Philipp (1980).

Proposition 4.2.8. *Assume that the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is stationary and strongly mixing, and that its strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{X}} = O(r^{-a})$ as $r\to\infty$ with $a > 3$. Then, Condition 4.2.5 holds for all vectors of points $\mathcal{P}$ such that Condition 4.2.3 holds.*

The previous proposition leads to the following immediate corollary of Theorem 4.2.6.

Corollary 4.2.9. *Assume that the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is stationary and strongly mixing, and that its strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{X}} = O(r^{-a})$ as $r\to\infty$ with $a > 3$. Then, for any fixed $\eta > 0$ and any vector of points $\mathcal{P}$ such that Condition 4.2.3 holds,*

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

The strong mixing conditions in the previous corollary are for instance satisfied (with much to spare) when $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is a stationary vector ARMA process with absolutely continuous innovations (see Mokkadem, 1988).

The following result, proven in Appendix B.2, can be regarded as a multivariate extension of Proposition 3.4 of Holmes and Kojadinovic (2021). It shows that imposing that η is strictly positive in Theorem 4.2.6 and Corollary 4.2.9 is necessary and sufficient for ensuring that the limiting random variable $\mathcal{L}_{p,\eta}$ is almost surely finite.

Proposition 4.2.10. *For any fixed $M > 0$,*

$$\mathbb{P}\Big(\sup_{1\le s\le t<\infty} t^{-\frac{3}{2}}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \ge M\Big) = 1.$$

Remark 4.2.11. In relation to the previous result, note that t^η in (4.10) could actually be replaced by $h(t)$, where $h(t) = \sqrt{\log\log t}$ when $t > e^e$ and $h(t) = 1$ when $t \leq e^e$. Indeed, as explained in Remark 3.5 of Holmes and Kojadinovic (2021), by the law of the iterated logarithm for Brownian motion, all the results stated before Proposition 4.2.10 should continue to hold with such a modification which could be considered optimal in the sense that, as $t \to \infty$, h diverges slower to infinity than $t \mapsto t^\eta$ for any $\eta > 0$. We did not however consider such a change as it is unwieldy from a practical perspective as shall become clearer from Section 4.4.

The next proposition, also proven in Appendix B.2, shows that the weak limit appearing in Theorem 4.2.6 and Corollary 4.2.9 is absolutely continuous. The proof is an application of Theorem 7.1 of Davydov and Lifshits (1984) together with an argument allowing us to reduce the problem to compact sets.

Proposition 4.2.12. *For any $\eta > 0$ and $p \in \mathbb{N}$, $\mathcal{L}_{p,\eta}$ is an absolutely continuous random variable.*

Let us finally explain how Theorem 4.2.6 can be used to carry out the monitoring in practice for a chosen vector of points $\mathcal{P}$ for which Condition 4.2.3 is assumed to hold. Given a significance level $\alpha \in (0, \frac{1}{2})$, suppose that we are able to compute $q_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$. Then, under H_0 in (4.1) and Condition 4.2.5, from the Portmanteau theorem, (4.11) holds. Hence, for large m, we can expect that, under H_0 and Condition 4.2.5,

$$\mathbb{P}\Big(D_m^{\mathcal{P}}(k) > q_{p,\eta}^{(1-\alpha)}(k/m)^{\frac{3}{2}+\eta} \text{ for some } k \geq m+1\Big) \simeq \alpha.$$

In practice, after the arrival of observation $\boldsymbol{X}_k$, $k > m$, $D_m^{\mathcal{P}}(k)$ is computed from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ and compared to the threshold $q_{p,\eta}^{(1-\alpha)}(k/m)^{\frac{3}{2}+\eta}$ (or, equivalently, $(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ is computed and compared to $q_{p,\eta}^{(1-\alpha)}$). If greater, the null hypothesis is rejected and the monitoring stops. Otherwise, $\boldsymbol{X}_{k+1}$ is collected and the previous iteration is repeated using the $k+1$ available observations.

Remark 4.2.13. Under a suitable transformation of Condition 4.2.5, we suspect that it is possible to obtain an analogue of Theorem 4.2.6 for the detector $D_m^{\sup}$ in (4.8). From the closed-end results obtained in Proposition 2.5 of Kojadinovic and Verdier (2021), we can actually guess the form of the corresponding weak limit. This leads us to believe that, under H_0 in (4.1) and a suitable version of Condition 4.2.5,

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\sup}(k) \rightsquigarrow \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}-\eta} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |tK(s,\boldsymbol{x}) - sK(t,\boldsymbol{x})|, \tag{4.14}$$

where the limit is almost surely finite and K is a *Kiefer process*, that is, a two-parameter centered Gaussian process whose covariance function is given, for any $s, t \in [0,\infty)$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, by

$$\begin{aligned}\Gamma(s,t,\boldsymbol{x},\boldsymbol{y}) = \min(s,t)\Big(&\mathrm{Cov}\{1(\boldsymbol{X}_1 \leq \boldsymbol{x}), 1(\boldsymbol{X}_1 \leq \boldsymbol{y})\} \\ &+ \sum_{i=2}^{\infty}\big[\mathrm{Cov}\{1(\boldsymbol{X}_1 \leq \boldsymbol{x}), 1(\boldsymbol{X}_i \leq \boldsymbol{y})\} + \mathrm{Cov}\{1(\boldsymbol{X}_i \leq \boldsymbol{x}), 1(\boldsymbol{X}_1 \leq \boldsymbol{y})\}\big]\Big).\end{aligned} \tag{4.15}$$

Thus, it appears that in general, the weak limit in (4.14) depends on the characteristics of the underlying time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$. This implies that in general, to carry

out monitoring based on the detector $D_m^{\sup}$, one would need to be able to estimate high quantiles of the weak limit in (4.14) prior to every execution of the procedure. Given the unwieldy form of the weak limit, this seems to be a major obstacle to the use of the detector $D_m^{\sup}$. One exception may be when the monitored observations are univariate ($d = 1$), continuous and serially independent. In that case, using a change of variable $F(x) \mapsto u$ (as is classically done for instance when dealing with a Brownian bridge), it can be verified that the weak limit in (4.14) no longer depends on the characteristics of the underlying time series $(X_i^{[1]})_{i\in\mathbb{N}}$. To estimate high quantiles of the resulting weak limit, one could then proceed as in forthcoming Section 4.4 where the estimation of high quantiles of the weak limit $\mathcal{L}_{p,\eta}$ appearing in Theorem 4.2.6 is addressed. Due to the apparently rather limited scope of application of an open-end monitoring procedure based on $D_m^{\sup}$, we do not pursue the investigation of such a sequential test in this work and leave this for future research.

4.2.3 Asymptotics under alternatives

To complement the previously stated asymptotic results, it is necessary to study the asymptotics of the monitoring procedure under sequences of alternatives to H_0 in (4.1). Because it is based on the detector $D_m^{\mathcal{P}}$ in (4.5), the studied monitoring procedure is expected to be particularly sensitive to alternative hypotheses of the form

$$H_1 : \exists\, k^\star \geq m \text{ and } \ell \in [\![1, p]\!] \text{ such that } \mathbb{P}(\boldsymbol{X}_1 \leq \boldsymbol{x}_\ell) = \cdots = \mathbb{P}(\boldsymbol{X}_{k^\star} \leq \boldsymbol{x}_\ell) \neq \mathbb{P}(\boldsymbol{X}_{k^\star+1} \leq \boldsymbol{x}_\ell) = \mathbb{P}(\boldsymbol{X}_{k^\star+2} \leq \boldsymbol{x}_\ell) = \dots$$

corresponding to a change in the d.f. at one or more of the chosen evaluation points. Note that this can be interpreted as a change in mean since by rewriting in terms of the univariate time series $\big(Y_i^{\mathcal{P},[\ell]}\big)_{i\in\mathbb{N}} = \big(\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_\ell)\big)_{i\in\mathbb{N}}$, we get the following equivalent statement

$$H_1 : \exists\, k^\star \geq m \text{ and } \ell \in [\![1, p]\!] \text{ such that } \mathbb{E}(Y_1^{\mathcal{P},[\ell]}) = \cdots = \mathbb{E}(Y_{k^\star}^{\mathcal{P},[\ell]}) \neq \mathbb{E}(Y_{k^\star+1}^{\mathcal{P},[\ell]}) = \mathbb{E}(Y_{k^\star+2}^{\mathcal{P},[\ell]}) = \dots . \quad (4.16)$$

As already mentioned at the beginning of Section 4.2, monitoring procedures designed to be particularly sensitive to changes in the mean were studied in Holmes and Kojadinovic (2021). Theorem 3.7 and Condition 3.7 in the latter reference specifically provide conditions under which, for a sequence of alternatives to H_0 related to H_1 in (4.16), $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} E_m^{\boldsymbol{x}_\ell}(k) \xrightarrow{\mathbb{P}} \infty$, where $E_m^{\boldsymbol{x}_\ell}$ is defined as in (4.3) with $\boldsymbol{x} = \boldsymbol{x}_\ell$. For the sake of brevity, we do not restate these conditions with the notation used in this work as they are lengthy to write. Very roughly speaking, they imply that for "early" or "late" changes in the d.f. of the observations at $\boldsymbol{x}_\ell$, the scaled detector $k \mapsto (m/k)^{\frac{3}{2}+\eta} E_m^{\boldsymbol{x}_\ell}(k)$ will end up exceeding any fixed threshold provided m is sufficiently large. The following result, proven in Appendix B.3, shows that, as $\boldsymbol{x}_\ell \in \mathcal{P}$, the same will hold for the scaled detector $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$, where $D_m^{\mathcal{P}}$ is defined in (4.5).

Proposition 4.2.14. *Let $\eta > 0$ and assume that for some $\boldsymbol{x}_\ell \in \mathcal{P}$, $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} E_m^{\boldsymbol{x}_\ell}(k) \xrightarrow{\mathbb{P}} \infty$. Then, if $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$, where $\Sigma^{\mathcal{P}}$ is positive-definite and*

$\Sigma_m^{\mathcal{P}}$ *is positive-definite almost surely for all* $m \in \mathbb{N}$,

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \xrightarrow{\mathbb{P}} \infty.$$

4.3 The case of continuous observations: practical implementation and additional asymptotic results under the null

Prior to using the monitoring procedure based on the detector $D_m^{\mathcal{P}}$ in (4.5), the user needs to choose the points $\mathcal{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$. Following the discussion initiated in Remark 4.2.4, using the learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ to do so seems meaningful. As mentioned in the latter remark, when the observations are discrete, a natural possibility consists of choosing $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ from a subset of frequently occurring observations. We focus in this section on the more complicated situation when the learning sample seems to be a stretch from a continuous time series.

Fix $\eta > 0$ and assume that m is large. Having Theorem 4.2.6 as well as Remark 4.2.7 and Corollary 4.2.9 in mind, one can hope that, under H_0 in (4.1), $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ has roughly the same distribution as the random variable $\mathcal{L}_{p,\eta}$ for all vectors of points $\mathcal{P}$ such that Condition 4.2.3 holds. A user who is interested in very specific changes in the d.f. may choose $\mathcal{P}$ accordingly. Otherwise, one natural possibility is to select the vector of points $\mathcal{P}$ such that the coordinates of each of the p points are empirical quantiles computed from the coordinate samples of the learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$. As we continue, for any $1 \leq j \leq k$, let $F_{j:k}^{[1]}, \ldots, F_{j:k}^{[d]}$ be the d univariate margins of $F_{j:k}$ defined in (4.4). Also, for any univariate d.f. H, let H^{-1} denote its associated quantile function (generalized inverse) defined by $H^{-1}(y) = \inf\{x \in \mathbb{R} : H(x) \geq y\}$, $y \in [0,1]$, with the convention that $\inf \emptyset = \infty$. Finally, let $\boldsymbol{\mathcal{X}}_{1,m}, \ldots, \boldsymbol{\mathcal{X}}_{p,m}$ denote the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ when chosen automatically from the learning sample and let $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \ldots, \boldsymbol{\mathcal{X}}_{p,m})$.

4.3.1 The univariate case

When $d = 1$, a natural instantiation of the previous generic strategy for choosing $\mathcal{P}_m$ consists of setting $\boldsymbol{\mathcal{X}}_{j,m}^{[1]} = F_{1:m}^{[1],-1}\big(j/(p+1)\big)$, $j \in [\![1, p]\!]$, that is, the $\boldsymbol{\mathcal{X}}_{j,m}^{[1]}$'s are merely taken as the $j/(p+1)$-empirical quantiles of the learning sample $X_1^{[1]}, \ldots, X_m^{[1]}$. As we will see in Section 4.5, this strategy seems to lead to powerful multi-purpose open-end monitoring procedures in the case of univariate observations.

4.3.2 The multivariate case

A natural first idea when $d > 1$ is simply to apply the univariate strategy above to each component sample of the learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, yielding sets $\mathcal{P}_m^{[i]} = \{\boldsymbol{\mathcal{X}}_{j,m}^{[i]} : j \in [\![1, p]\!]\}$ of size p for all $i \in [\![1, d]\!]$. For each dimension $i \in [\![1, d]\!]$, each selected coordinate $\boldsymbol{\mathcal{X}}_{j,m}^{[i]}$, $j \in [\![1, p]\!]$, typically corresponds to a unique d-dimensional vector of the learning sample. One could then define $\mathcal{P}_m$ to be the union of the corresponding d sets of d-dimensional points, which implies that $p \leq |\mathcal{P}_m| \leq dp$.

A preliminary implementation of this strategy showed however that (among other things) this approach can sometimes lead to the selection of points in the learning sample that are too close to the "border" of the point cloud $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, resulting in numerical difficulties when computing the inverse or the square root of $\Sigma_m^{\mathcal{P}_m}$ (see also Remark 4.2.4).

Another natural adaption of the strategy considered in the univariate case would be to choose an integer $r \geq 1$, consider the uniformly-spaced grid (containing r^d points)

$$\Pi = \{(j_1/(r+1), \ldots, j_d/(r+1)) : j_1, \ldots, j_d \in [\![1, r]\!]\} \subseteq (0,1)^d \tag{4.17}$$

and define $\mathcal{P}_m$ as consisting of the r^d points

$$\left\{ \left(F_{1:m}^{[1],-1}(\pi^{[1]}), \ldots, F_{1:m}^{[d],-1}(\pi^{[d]}) \right) : \boldsymbol{\pi} \in \Pi \right\}. \tag{4.18}$$

This strategy needs however to be refined because some of the above points might not belong to the support of F which, as hinted at in Remark 4.2.4, is a necessary condition for Condition 4.2.3 to hold. Let $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$ be the unobservable sample obtained from the learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ by probability integral transformations, that is, let

$$\boldsymbol{U}_i = (U_i^{[1]}, \ldots, U_i^{[d]}) = \left(F^{[1]}(X_i^{[1]}), \ldots, F^{[d]}(X_i^{[d]}) \right), \tag{4.19}$$

where $F^{[1]}, \ldots, F^{[d]}$ are the d unknown univariate margins of F. Note in passing that $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$ can be regarded as a stretch from a d-dimensional time series of continuous random vectors with contemporary d.f. C, where C is the (unique) copula of F (see, e.g., Sklar, 1959) satisfying

$$C(\boldsymbol{u}) = F\left(F^{[1],-1}(u^{[1]}), \ldots, F^{[d],-1}(u^{[d]}) \right), \qquad \boldsymbol{u} \in [0,1]^d,$$

and

$$F(\boldsymbol{x}) = C\left(F^{[1]}(x^{[1]}), \ldots, F^{[d]}(x^{[d]}) \right), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

Adapting the approach briefly described in Section 4.2 of Li and Genton (2013), we propose to keep in $\mathcal{P}_m$ only those points in (4.18) constructed from grid points in (4.17) whose "neighborhood" contains a sufficiently large proportion of the $\boldsymbol{U}_i$'s. As $F^{[1]}, \ldots, F^{[d]}$ are unknown, we follow one of the classical approaches used in the copula literature (see, e.g., Hofert et al., 2018, and the references therein) and use $\hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_m$ as a proxy for $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$, where

$$\hat{\boldsymbol{U}}_i = \frac{m}{m+1} \left(F_{1:m}^{[1]}(X_i^{[1]}), \ldots, F_{1:m}^{[d]}(X_i^{[d]}) \right). \tag{4.20}$$

For any $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$ such that $\boldsymbol{a} < \boldsymbol{b}$, let $(\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{u} \in [0,1]^d : \boldsymbol{a} < \boldsymbol{u} \leq \boldsymbol{b}\}$. Furthermore, let ν_m be the empirical measure of $\hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_m$ and let $\boldsymbol{s} = \big(1/(r+1), \ldots, 1/(r+1)\big) \in \mathbb{R}^d$. Given $\boldsymbol{\pi} \in \Pi$ and if the d components of $\boldsymbol{U}_1$ are independent, it is expected that the proportion of $\boldsymbol{U}_i$'s in $(\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]$ be approximately equal to $1/(r+1)^d$. This motivates the following strategy: we choose to retain in $\mathcal{P}_m$ only the points in (4.18) constructed from grid points $\boldsymbol{\pi} \in \Pi_m$, where

$$\Pi_m = \left\{ \boldsymbol{\pi} \in \Pi : \nu_m\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) > \frac{1}{\kappa (r+1)^d} \right\}, \tag{4.21}$$

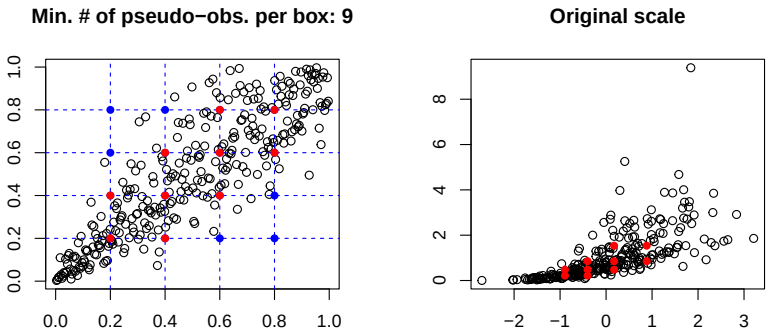


Figure 4.1: Automatic choice of $\mathcal{P}_m$ in the bivariate case for $m = 300$, $r = 4$ and $\kappa = 1.5$. Left: scatterplot of the "pseudo-observations" $\hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_m$ obtained from the learning sample, initial uniformly-spaced grid Π in (4.17) in blue and selected points $\Pi_m \subseteq \Pi$ in red. Right: scatterplot of the learning sample in black and points in $\mathcal{P}_m$ in red.

Π is defined in (4.17) and $\kappa > 1$ is a user-chosen parameter. The number of automatically chosen points $p = |\Pi_m|$ depends on m. Figure 4.1 illustrates the automatic choice of $\mathcal{P}_m$ in the bivariate case for $r = 4$ and $\kappa = 1.5$. Values for κ and r appearing to lead to powerful multi-purpose open-end monitoring procedures will be recommended in the case $d \in \{2, 3\}$ in Section 4.5.

We end this section by stating an asymptotic property of the proposed selection procedure. Let ν_C be the measure on $[0,1]^d$ associated with the copula C of F and let

$$\Pi_C = \left\{ \boldsymbol{\pi} \in \Pi : \nu_C\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) > \frac{1}{\kappa(r+1)^d} \right\}. \tag{4.22}$$

Also, recall the definition of Π_m in (4.21) and, as classically done in the literature (see, e.g., Hofert et al., 2018, Chapter 4 and the references therein), let the empirical copula C_m of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be defined as the empirical d.f. of the "pseudo-observations" $\hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_m$ defined in (4.20). Of course, one hopes that Π_m is close to the deterministic (but unknown) set Π_C when m is large. Proposition 4.3.2 below (proven in Appendix B.4) makes this statement rigorous under the following mild condition.

Condition 4.3.1.

(i) *For each* $\boldsymbol{\pi} \in \Pi$, $\nu_C\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) \neq 1/(\kappa(r+1)^d)$, *and*

(ii) $\sup_{\boldsymbol{u} \in [0,1]^d} |C_m(\boldsymbol{u}) - C(\boldsymbol{u})| \xrightarrow{a.s.} 0$.

Proposition 4.3.2. *Assume that Condition 4.3.1 holds. Then, almost surely, for all* m *sufficiently large,* $\Pi_m = \Pi_C$.

In other words, under Condition 4.3.1 and provided m is sufficiently large, we can regard $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \ldots, \boldsymbol{\mathcal{X}}_{p,m})$ as being formed of the $p = |\Pi_C|$ points

$$\left\{ \left(F_{1:m}^{[1],-1}(\pi^{[1]}), \ldots, F_{1:m}^{[d],-d}(\pi^{[d]}) \right) : \boldsymbol{\pi} \in \Pi_C \right\},$$

where Π_C is defined in (4.22).

4.3.3 Additional asymptotic results under the null

The asymptotic results stated in Sections 4.2.2 and 4.2.3 concern the monitoring procedure based on the detector $D_m^{\mathcal{P}}$ in (4.5) where the chosen evaluation points $\mathcal{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ are fixed (they are not allowed to change in the asymptotics with m). To fully asymptotically justify the use of the detector $D_m^{\mathcal{P}_m}$ resulting from the automatic choices of points $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \ldots, \boldsymbol{\mathcal{X}}_{p,m})$ considered in Sections 4.3.1 or 4.3.2, one needs an analogue of Theorem 4.2.6 in which the points at which the empirical d.f.s are evaluated are allowed to change with m.

In the rest of this section, we assume that the automatically chosen points in $\mathcal{P}_m$ are of the form

$$\boldsymbol{\mathcal{X}}_{i,m} = \left(F_{1:m}^{[1],-1}(\pi_i^{[1]}), \ldots, F_{1:m}^{[d],-1}(\pi_i^{[d]})\right), \tag{4.23}$$

for some p vectors of probabilities $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_p \in (0,1)^d$ not depending on m. This is clearly the case for the univariate selection strategy proposed in Section 4.3.1. From Proposition 4.3.2, it is also the case for the multivariate strategy proposed in Section 4.3.2 upon additionally assuming that Condition 4.3.1 holds and that m is sufficiently large.

Next, set $\boldsymbol{x}_i = \left(F^{[1],-1}(\pi_i^{[1]}), \ldots, F^{[d],-1}(\pi_i^{[d]})\right)$, write $\mathcal{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ and notice that the points in $\mathcal{P}$ are unobservable since $F^{[1]}, \ldots, F^{[d]}$ are unknown. The p-dimensional random vectors $\boldsymbol{Y}_i^{\mathcal{P}} = \left(\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_1), \ldots, \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_p)\right)$ are also unobservable. Since $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \ldots, \boldsymbol{\mathcal{X}}_{p,m})$ with $\boldsymbol{\mathcal{X}}_{i,m}$ given by (4.23) is an estimator of $\mathcal{P}$, the long-run covariance matrix $\Sigma^{\mathcal{P}}$ in (4.6) of the unobservable p-dimensional time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$ may still be estimated from the sample $\boldsymbol{Y}_1^{\mathcal{P}_m}, \ldots, \boldsymbol{Y}_m^{\mathcal{P}_m}$ which is a proxy for the sample $\boldsymbol{Y}_1^{\mathcal{P}}, \ldots, \boldsymbol{Y}_m^{\mathcal{P}}$.

We first state a condition under which the monitoring procedure based on $D_m^{\mathcal{P}_m}$ and the unobservable monitoring procedure based on $D_m^{\mathcal{P}}$ in (4.5) are asymptotically equivalent.

Condition 4.3.3 (For the asymptotic equivalence of $D_m^{\mathcal{P}_m}$ and $D_m^{\mathcal{P}}$)**.** *For any* $i \in [\![1, p]\!]$,

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} j|F_{1:j}(\boldsymbol{\mathcal{X}}_{i,m}) - F(\boldsymbol{\mathcal{X}}_{i,m}) - F_{1:j}(\boldsymbol{x}_i) + F(\boldsymbol{x}_i)| = o_{\mathbb{P}}(1). \tag{4.24}$$

The following result, proven in Appendix B.4, shows that the previous condition can be satisfied under the null and *absolute regularity*. Given a time series $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$, recall that, for $j, k \in \mathbb{N} \cup \{+\infty\}$, $\mathcal{M}_j^k$ denotes the σ-field generated by $(\boldsymbol{Z}_i)_{j\leq i\leq k}$, that the absolute regularity coefficients corresponding to $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ are defined by

$$\beta_r^{\boldsymbol{Z}} = \mathbb{E}\left\{\sup_{k\in\mathbb{N}} \sup_{B\in\mathcal{M}_{k+r}^{+\infty}} \left|\mathbb{P}(B \mid \mathcal{M}_1^k) - \mathbb{P}(B)\right|\right\}, \qquad r \in \mathbb{N}, \tag{4.25}$$

and that the sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ is said to be *absolutely regular* if $\beta_r^{\boldsymbol{Z}} \to 0$ as $r \to \infty$. Also, note that absolute regularity is known to imply strong mixing (see, e.g., Dehling and Philipp, 2002) and that an independent sequence is clearly absolutely regular since in this case for every k and B as in (4.25), $\mathbb{P}(B|\mathcal{M}_1^k) = \mathbb{P}(B)$ almost surely.

Proposition 4.3.4 (Condition 4.3.3 can hold under the null)**.** *Assume that the underlying time series* $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ *is stationary and absolutely regular, and that its*

absolute regularity coefficients satisfy $\beta_r^{\boldsymbol{X}} = O(r^{-a})$ *as* $r \to \infty$ *with* $a > 1$*. Then, if the* d *univariate margins* $F^{[1]}, \ldots, F^{[d]}$ *of* F *are continuous and if, for each* $i \in [\![1, p]\!]$, $\boldsymbol{\mathcal{X}}_{i,m} \xrightarrow{\mathbb{P}} \boldsymbol{x}_i$*, Condition 4.3.3 holds.*

Remark 4.3.5. Assumptions related to Condition 4.3.3 appear in Dette and Gösmann (2020) in the context of the study of the asymptotics of closed-end sequential tests designed to be sensitive to changes in the mean, the variance or certain quantiles. In an open-end setting, related conditions are stated in Assumption 2.5 of Gösmann et al. (2021) and in Condition 6.1 of Holmes and Kojadinovic (2021) under an "almost sure" form. As an inspection of the proof of Proposition 4.3.4 reveals, Condition 4.3.3 is essentially a consequence of the continuity of the margins of F and Theorem 3.1 of Dedecker et al. (2014) which provides an adequate strong approximation result for the empirical process under absolute regularity (but not under strong mixing).

Remark 4.3.6. In the statement of Proposition 4.3.4, it is assumed that, for each $i \in [\![1, p]\!]$, $\boldsymbol{\mathcal{X}}_{i,m} \xrightarrow{\mathbb{P}} \boldsymbol{x}_i$. This condition can actually be dispensed with provided additional conditions of the true unobservable quantile functions $F^{[1],-1}, \ldots, F^{[d],-1}$ are assumed instead. Indeed, from Rio (1998), we know that the condition on the absolute regularity coefficients in Proposition 4.3.4 implies that, for any $\ell \in [\![1, d]\!]$, $F_{1:m}^{[\ell]} \xrightarrow{\mathbb{P}} F^{[\ell]}$ in $\ell^\infty(\mathbb{R})$, where $\ell^\infty(\mathbb{R})$ denotes the space of bounded functions on $\mathbb{R}$ equipped with the uniform metric. From Lemma 21.2 in van der Vaart (1998), this is then equivalent to the fact that $F_{1:m}^{[\ell],-1}(\pi) \xrightarrow{\mathbb{P}} F^{[\ell],-1}(\pi)$ at every $\pi \in (0, 1)$ at which $F^{[\ell],-1}$ is continuous. Consequently, the condition that, for any $i \in [\![1, p]\!]$, $\boldsymbol{\mathcal{X}}_{i,m} \xrightarrow{\mathbb{P}} \boldsymbol{x}_i$ could be replaced by the condition that, for any $\ell \in [\![1, d]\!]$, $F^{[\ell],-1}$ is continuous at $\pi_i^{[\ell]}$, for all $i \in [\![1, p]\!]$.

The next proposition, also proven in Appendix B.4, states that, under Conditions 4.2.5 and 4.3.3, the monitoring procedures based on $D_m^{\mathcal{P}_m}$ and $D_m^{\mathcal{P}}$ are asymptotically equivalent.

Proposition 4.3.7. *Under Conditions 4.2.5 and 4.3.3, and if* $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ *and* $\Sigma_m^{\mathcal{P}_m} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$*, for any* $\eta > 0$*,*

$$\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} |D_m^{\mathcal{P}_m}(k) - D_m^{\mathcal{P}}(k)| = o_{\mathbb{P}}(1),$$

and, consequently,

$$\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}_m}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1 \le s \le t < \infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

The last claim of the previous proposition suggests to carry out the monitoring procedure based on the detector $D_m^{\mathcal{P}_m}$ exactly as the procedure based on the detector $D_m^{\mathcal{P}}$ when the points $\mathcal{P}$ are hand-picked by the user (see the last paragraph of Section 4.2.2).

4.3.4 The monitoring procedure based on $D_m^{\mathcal{P}_m}$ is margin-free under the null

We end this section by verifying that, under the considered assumption that the true unknown marginal d.f.s $F^{[1]}, \ldots, F^{[d]}$ are continuous, the procedure based on

the detector $D_m^{\mathcal{P}_m}$ is margin-free under H_0 in (4.1), that is, it does not depend on $F^{[1]}, \ldots, F^{[d]}$ under the null. To see this, let $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_k$ be the unobservable sample obtained from the available observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ using (4.19) and let $G_{1:m}$ be the empirical d.f. of $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$. Notice that we can recover the $\boldsymbol{X}_i$ from the $\boldsymbol{U}_i$ by marginal quantile transformations, that is, $\boldsymbol{X}_i = (F_1^{[1],-1}(U_i^{[1]}), \ldots, F_d^{[d],-1}(U_i^{[d]}))$. Furthermore, for any $j \in [\![1, d]\!]$, by (right) continuity of $F^{[j]}$, we have that $\mathbf{1}\{F^{[j],-1}(u) \leq x\} = \mathbf{1}\{u \leq F^{[j]}(x)\}$ for all $u \in [0, 1]$ and $x \in \mathbb{R}$; see, e.g., Proposition 1 (5) in Embrechts and Hofert (2013). Then, it can be verified that, for every $\boldsymbol{\pi} \in (0, 1)^d$, $i \in [\![1, k]\!]$ and $j \in [\![1, d]\!]$,

$$1\{X_i^{[j]} \leq F_{1:m}^{[j],-1}(\pi^{[j]})\} = 1\{U_i^{[j]} \leq F^{[j]}(F_{1:m}^{[j],-1}(\pi^{[j]}))\} = 1\{U_i^{[j]} \leq G_{1:m}^{[j],-1}(\pi^{[j]})\},$$

which implies that, under the null and the current setting, the detector at k can be rewritten to depend only on $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_k$.

4.4 Estimation of high quantiles of the limiting distribution

From the two previous sections, we know that, to carry out the studied monitoring procedures, it is necessary to be able to accurately estimate high quantiles of the random variable $\mathcal{L}_{p,\eta}$, $p \geq 1$, $\eta > 0$, appearing first in the statement of Theorem 4.2.6. The underlying estimation problem was empirically solved in Holmes and Kojadinovic (2021) for $p = 1$ using *asymptotic regression modeling*. We choose to use the same approach when $p > 1$ and refer the reader to Section 4 of the aforementioned reference where the motivation for this way of proceeding is explained in detail.

Fix $\eta > 0$, $p \geq 1$, $\alpha \in (0, \frac{1}{2})$ and let $q_{p,\eta}^{(1-\alpha)}$ be the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$. Furthermore, let $d = 1$, let $\big(X_i^{[1]}\big)_{i \in \mathbb{N}}$ be an infinite sequence of independent standard normals and let $\mathcal{P} = (x_1^{[1]}, \ldots, x_p^{[1]})$ where $x_i^{[1]} = \Phi^{-1}\big(i/(p+1)\big)$, $i \in [\![1, p]\!]$ and Φ is the d.f. of the standard normal. According to Theorem 4.2.6, for large m the distribution of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ should be close to that of $\mathcal{L}_{p,\eta}$. If for a given realization of $\big(X_i^{[1]}\big)_{i \in \mathbb{N}}$ we could compute the corresponding realization of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$, then $q_{p,\eta}^{(1-\alpha)}$ could be estimated by $\hat{q}_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-empirical quantile of a large sample of realizations of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$. As this is not possible because of the supremum over $k > m$, the idea taken from Holmes and Kojadinovic (2021) is to model the relationship between $r \in [\![9, 16]\!]$ and $\hat{q}_{p,\eta,r}^{(1-\alpha)}$, the $(1-\alpha)$-empirical quantile of $\sup_{k \in]\!]m, m+2^r]\!]}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$, using an asymptotic regression model.

To begin with, using a computer grid, we computed the empirical quantiles $\hat{q}_{p,\eta,r}^{(1-\alpha)}$, $r \in [\![9, 16]\!]$, from 10,000 simulated trajectories of the scaled detector $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ for $k \in]\!]m, m + 2^{16}]\!]$ and $m = 500$. In a next step, an asymptotic regression model was fitted to the points $(r, \hat{q}_{p,\eta,r}^{(1-\alpha)})$, $r \in [\![9, 16]\!]$. The considered model is a three-parameter model with mean function

$$f(x) = \beta_1 + (\beta_2 - \beta_1)\{1 - \exp(-x/\beta_3)\},$$

where $y = \beta_2$ is the equation of the upper horizontal asymptote of f. Its fitting was carried out using the R package `drc` (Ritz et al., 2015). A candidate estimate $\hat{q}_{p,\eta}^{(1-\alpha)}$ of

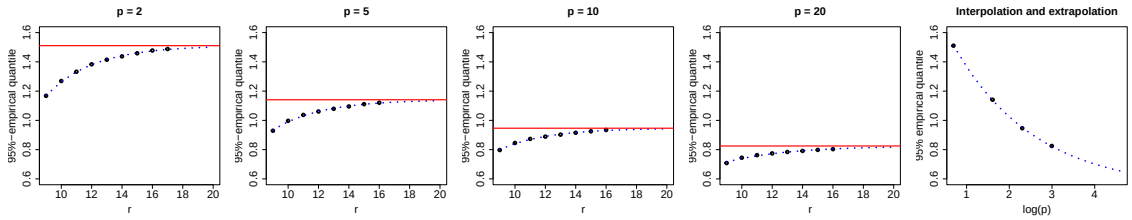


Figure 4.2: First four panels: for $\alpha = 0.05$, $\eta = 0.001$ and $p \in \{2, 5, 10, 20\}$, scatter plots of $\{(r, \hat{q}_{p,\eta,r}^{(1-\alpha)})\}_{r\in[\![9,16]\!]}$, corresponding fitted asymptotic regression models (dotted blue curves) and estimates of the upper horizontal asymptotes (solid red lines) which are candidate estimates of $q_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$. Fifth panel: scatter plot of $\{(\log(p), \hat{q}_{p,\eta}^{(1-\alpha)})\}_{p\in\{2,5,10,20\}}$ and corresponding transformed fitted asymptotic regression model (dotted blue curve) that could be used to interpolate (resp. extrapolate) the value of $\hat{q}_{p,\eta}^{(1-\alpha)}$ for $p \in [2, 20]$ (resp. for p slightly larger than 20).

Table 4.1: Columns 2 to 5: for $\eta = 0.001$, $p \in \{2, 5, 10, 20\}$ and $\alpha \in \{0.01, 0.05, 0.1\}$, estimates $\hat{q}_{p,\eta}^{(1-\alpha)}$ of the $(1-\alpha)$-quantiles $q_{p,\eta}^{(1-\alpha)}$ of the distribution of $\mathcal{L}_{p,\eta}$. Columns 6 to 8: corresponding estimates of the parameters of the transformed asymptotic regression models that can be used to interpolate (resp. extrapolate) the value of $\hat{q}_{p,\eta}^{(1-\alpha)}$ for $p \in [2, 20]$ (resp. for p slightly larger than 20).

	p				$p \notin \{2, 5, 10, 20\}$		
$1-\alpha$	2	5	10	20	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.99	1.654	1.234	1.010	0.860	-0.126	1.535	2.080
0.95	1.511	1.141	0.946	0.825	0.060	1.475	1.921
0.90	1.450	1.099	0.921	0.806	0.140	1.462	1.870

$q_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$, is then the resulting estimate of the parameter β_2. The previous steps were carried out for $p \in \{2, 5, 10, 20\}$, $\alpha \in \{0.01, 0.05, 0.1\}$ and $\eta = 0.001$ (following the practical recommendation made in Holmes and Kojadinovic (2021)), and can be visualized in the first four panels of Figure 4.2 for $\alpha = 0.05$. The corresponding estimates of the quantiles $q_{p,\eta}^{(1-\alpha)}$ are given in columns two to five of Table 4.1.

In a last step, to be able to carry out the monitoring procedures for some values of p different than those in $\{2, 5, 10, 20\}$, we fitted asymptotic regression models to the points $(\log(p), 2 - \hat{q}_{p,\eta}^{(1-\alpha)})$, $p \in \{2, 5, 10, 20\}$, for $\alpha \in \{0.01, 0.05, 0.1\}$ and $\eta = 0.001$. The estimates of the parameters β_1, β_2 and β_3 are reported in the last three columns of Table 4.1. We make no claim regarding the theoretical adequacy of this type of model. The aim is only to be able to interpolate (resp. extrapolate) the value of $\hat{q}_{p,\eta}^{(1-\alpha)}$ for $p \in [2, 20]$ (resp. for p slightly larger than 20). Note that, since $\mathcal{L}_{p,\eta} \overset{a.s.}{\to} 0$ as $p \to \infty$ (as a consequence of the definition of the norm $\|\cdot\|_{I_p}$), a fitted two-parameter submodel with β_2 fixed to 2 is expected to behave better for large p. We have nonetheless decided to keep the fitted three parameter model because its accuracy for values of p slightly larger than 20 was found to be better in our Monte Carlo experiments summarized in the forthcoming section.

4.5 Monte Carlo experiments

We carried out rather extensive numerical experiments in the case of low-dimensional ($d \in \{1, 2, 3\}$) continuous observations to investigate the finite-sample behavior of the monitoring procedure based on the detector $D_m^{\mathcal{P}_m}$ introduced in Section 4.3. Specifically, for $d = 1$, recall from Section 4.3.1 that the p real points at which the (univariate) empirical d.f.s are evaluated are chosen as empirical quantiles of order $i/(p+1)$, $i \in [\![1, p]\!]$, computed from the learning sample $X_1^{[1]}, \ldots, X_m^{[1]}$. For $d \in \{2, 3\}$, the approach is slightly more involved: as explained in Section 4.3.2, the evaluation points are chosen using a point selection procedure which relies on two parameters: an integer $r \geq 1$ specifying the maximal value r^d of p and the constant κ controlling how many of the initial r^d grid points will actually be retained.

From the definition of the detector $D_m^{\mathcal{P}_m}$, we see that an underlying unknown long-run covariance matrix needs to be estimated from the sample $\boldsymbol{Y}_1^{\mathcal{P}_m}, \ldots, \boldsymbol{Y}_m^{\mathcal{P}_m}$. In practice, for $\Sigma_m^{\mathcal{P}_m}$, we used the estimator of Andrews (1991) based on the quadratic spectral kernel with automatic bandwidth selection as implemented in the function `lrvar()` of the R package `sandwich` (Zeileis, 2004; Zeileis et al., 2020). Note that we did not however use prewhitening as suggested in Andrews and Monahan (1992). The fact that the monitoring procedure studied in this work is available in the R package `npcp` (Kojadinovic and Verhoijsen, 2022) not only makes all our experiments fully reproducible but also allows a user to change the long run covariance estimator (for instance to that of Newey and West, 1987) by passing parameters to the main function which will be passed to the function `lrvar()`.

Before we present our empirical findings, it is important to keep in mind that, in the case of open-end approaches, numerical experiments only provide a biased view of their behavior as finite computing resources impose that monitoring has to be stopped eventually. From the point of view of statistical testing, the main consequence of this is that all rejection percentages are underestimated.

The full details of our simulations are available in Appendix B.5. We provide hereafter a summary of our findings:

- For monitoring univariate data, taking $p \in \{5, \ldots, 10\}$ evaluation points and choosing them as suggested in Section 4.3.1 seems a good choice in general. Unless serial dependence is very strong, a learning sample of size $m \geq 800$ seems to lead to a procedure that holds its level well and displays good power against several alternatives involving a change in the mean or the variance, or such that the d.f. changes with the mean and variance remaining constant. In the case of very strong serial dependence (such as for an AR(1) model with autoregressive parameter 0.7), a larger learning sample (for instance $m = 1600$) seems necessary for the monitoring procedure to hold its level reasonably well.

- For monitoring low-dimensional data ($d \in \{2, 3\}$), using $r = 4$ if $d = 2$, $r = 3$ if $d = 3$ and $\kappa = 1.5$ in the point selection procedure of Section 4.3.2 seems to be a reasonable choice in general. As for $d = 1$, in the case of mild serial dependence, taking $m = 800$ seems sufficient to obtain a sequential test that holds its level reasonably well. With such settings, the procedure seems powerful against various alternatives involving changes in one margin or in the copula of the contemporary d.f.

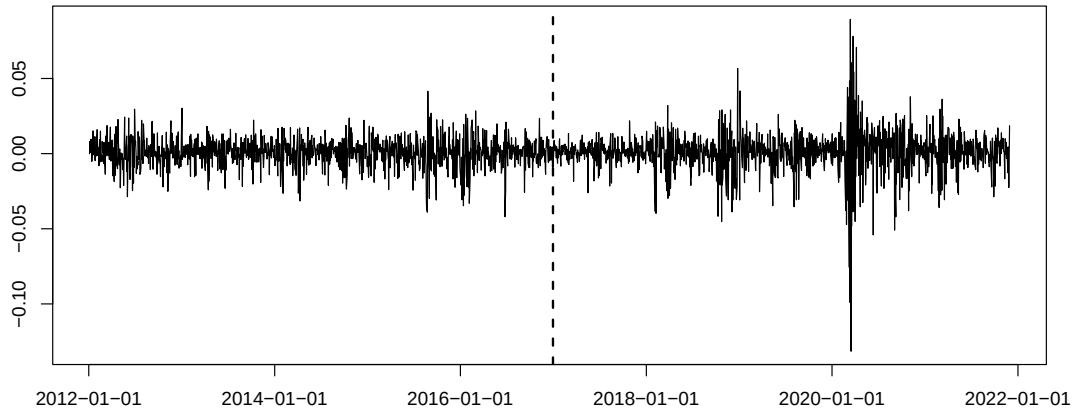


Figure 4.3: Daily log-returns computed from closing quotes of the NASDAQ composite index from 2012 to 2021. The learning sample in our fictitious data example corresponds to the period 2012 – 2016. Monitoring starts on the first trading day of 2017, which is represented by a dashed vertical line.

4.6 Data example and concluding remarks

Let us briefly illustrate how the procedure based on the detector $D_m^{\mathcal{P}_m}$ considered in Section 4.3 could be used for monitoring changes in the contemporary distribution of daily log-returns of a financial index. In our fictitious example, the learning sample (whose stationarity would need to be tested) corresponds to 1257 trading days of the NASDAQ for the period January 3rd 2012 – December 30th 2016. Monitoring starts on the first trading day of 2017. The corresponding daily log-returns are represented in Figure 4.3, where the dashed vertical line represents the start of the monitoring.

The sample paths of $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}_m}(k)$ are represented in Figure 4.4 for $p \in \{5, 10\}$. The horizontal dashed line in each panel represent one of the 95%-quantiles given in the second row of Table 4.1. The dashed vertical lines mark the first time that the threshold is exceeded (corresponding to May 2020). A change in the data generating process slightly prior to this date seems likely as an inspection of Figure 4.3 reveals: the first months of 2020 are indeed characterized by a period of very high volatility.

We end this section by stating a few remarks:

- One important practical advantage of the monitoring procedure proposed in this work comes from its open-end nature and is that the monitoring horizon does not need to be specified. As a consequence, monitoring could theoretically run forever. This is however not possible from a practical perspective because of the form of the detector in (4.5): it is indeed clear that the cost of computing the detector at time k increases with k. The latter implies that its computation will become impossible for k large enough. Some other detectors, such as the ordinary CUSUM, will not be affected by such a practical

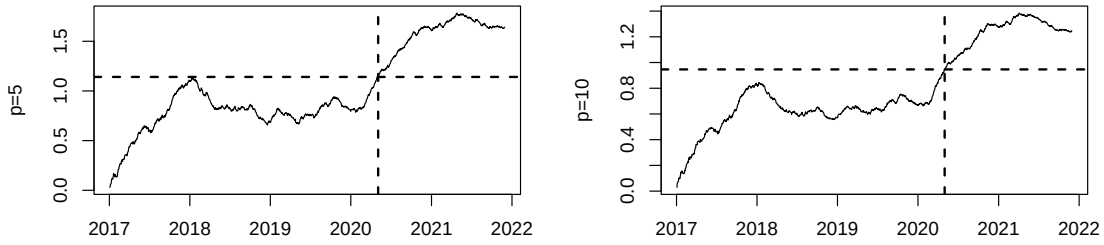


Figure 4.4: Sample paths of $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}_m}(k)$ for $p = 5$ and 10. The horizontal dashed lines represent the 95%-quantiles given in the second row of Table 4.1. The vertical dashed lines mark the first time the scaled versions of the detectors exceed the corresponding quantiles.

issue. Yet, as empirically observed in Gösmann et al. (2021) and in Holmes and Kojadinovic (2021), the ordinary CUSUM is substantially less powerful than more computationally costly detectors similar to the ones considered in this work. In a related way, let us mention that the implementation of the studied monitoring procedure available in the R package `npcp` is merely a proof of concept and is not optimized for very long-term monitoring.

- The price to pay for open-end monitoring is that the detection power decreases as time elapses. For the studied class of procedures, this is due to the parameter η as discussed in Section 4 of Holmes and Kojadinovic (2021). On one hand, the smaller the value of η, the weaker the power decrease. On the other hand, the smaller η, the more conceptually difficult and computationally costly it is to estimate high quantiles of the distribution $\mathcal{L}_{p,\eta}$ appearing in Theorem 4.2.6. The latter suggests to devote more research to the estimation of high quantiles of $\mathcal{L}_{p,\eta}$.

- The type of monitoring procedure used in this work could also be used to detect changes in the serial dependence. For instance, to detect such changes "at lag 1" from an initial sequence of univariate observations $Z_1, \ldots, Z_m, Z_{m+1}, \ldots$, the $\boldsymbol{X}_i$ could be formed as $\boldsymbol{X}_i = (Z_i, Z_{i+1})$.

Chapter 5

Ongoing investigations, conclusions, and directions for future research

5.1 Ongoing investigations: a nonparametric sequential test for change point detection in the copula

The work in Chapter 4 is concerned with monitoring possibly multivariate observations to detect changes in the d.f., the latter being estimated using the empirical d.f. The nonparametric monitoring procedure was shown to be sensitive to changes in the margins and in the multivariate d.f. We investigate a new nonparametric monitoring procedure for multivariate continuous observations that is adapted to specifically detect changes in the cross-sectional dependence by monitoring the contemporary copula, which can be estimated using the empirical (beta) copula.

Recall from Chapter 4 that we have at hand a learning sample $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$ with contemporary d.f. F, $i \in \{1, \ldots, m\}$, from a d-dimensional stationary time series. In what follows, it is assumed that the d unknown univariate margins $F_1, \ldots, F_d$ of F are continuous. As a consequence of Sklar's theorem (see Theorem 2.2.7), the multivariate d.f. F can be expressed as

$$F(\boldsymbol{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \qquad \boldsymbol{x} \in \mathbb{R}^d,$$

in terms of a unique *copula* C, that is, a unique d-dimensional d.f. with continuous standard uniform margins.

The nonparametric open-ended monitoring procedure under investigation is similar to the one introduced in Section 4.1. Assume that after monitoring starts, the collected d-dimensional observations $\boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots$ all have contemporary d.f.s with continuous univariate margins, which implies that their copulas are uniquely defined. We want to detect as quickly as possible when these copulas change.

Let $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots$ be a stretch from the unobservable time series $(\boldsymbol{U}_i)_{i \in \mathbb{N}}$, where $\boldsymbol{U}_i = (F_{i1}(X_{i1}), \ldots, F_{id}(X_{id}))$. Fix $p \geq 1$ points $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ in $[0,1]^d$, where the integer p is chosen by the user such that, for each $j \in [\![1, d]\!]$, the points $u_{j1}, \ldots, u_{jp}$

are uniformly spread out over the unit interval $[0, 1]$. The goal is to derive open-end nonparametric monitoring procedures that are particularly sensitive to alternatives of the form

$$H_1 : \exists k^\star \geq m \text{ and } \ell \in [\![1, p]\!] \text{ such that } \mathbb{P}\left(\boldsymbol{U}_1 \leq \boldsymbol{u}_\ell\right) = \cdots = \mathbb{P}\left(\boldsymbol{U}_{k^\star} \leq \boldsymbol{u}_\ell\right) \neq \mathbb{P}\left(\boldsymbol{U}_{k^\star+1} \leq \boldsymbol{u}_\ell\right) = \mathbb{P}\left(\boldsymbol{U}_{k^\star+2} \leq \boldsymbol{u}_\ell\right) = \cdots, \tag{5.1}$$

where inequalities between vectors are to be understood componentwise. For $d \geq 2$, we concern ourselves with testing for changes in the contemporary copula C. The null hypothesis of the procedure is still

$$H_0 : \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots, \quad \text{is a stretch from a stationary time series,} \tag{5.2}$$

which is exactly the same null hypothesis as the one in (4.1). In other words, we wish to develop a test that is specifically sensitive to H_1 in (5.1).

We briefly summarise the monitoring procedure which is similar to the one outlined in Chapter 4. The nonparametric monitoring procedure consists of a detector and a threshold function and compares, after the arrival of the kth observation $\boldsymbol{X}_k$, a positive statistic, the detector T_m, to a threshold. If after the arrival of observation $\boldsymbol{X}_k$, with $k > m$, the detector exceeds the threshold, the hypothesis of stationarity is rejected. If not, a new observation $\boldsymbol{X}_{k+1}$ is collected and the monitoring procedure is repeated from the start. Recall that for a significance level $\alpha \in (0, 1/2)$ we want to find a threshold $w(k/m)$ such that, under H_0 in (5.2),

$$\mathbb{P}\left\{T_m(k) \leq w(k/m) \text{ for all } k > m\right\} = \mathbb{P}\left\{\sup_{k>m} \frac{T_m(k)}{w(k/m)} \leq 1\right\} = 1 - \alpha. \tag{5.3}$$

Concerning the above, there is no difference between what was done in Chapter 4 and what is proposed here. Following the results in Chapter 4, we consider the threshold function

$$w(t) = q_{\boldsymbol{y}}^{(1-\alpha)} v(t), \quad t \in [1, \infty), \tag{5.4}$$

where $v(t)$ is a weighting function such that (5.3) holds, $\boldsymbol{y}$ is a (possibly empty) vector of parameters, and $q_{\boldsymbol{y}}^{(1-\alpha)}$ is the $(1-\alpha)$-quantile of the limiting random variable $\mathcal{I}_{\boldsymbol{y}}$, the weak limit of $\sup_{k>m} T_m(k)/v(k/m)$. By the Portmanteau lemma,

$$\begin{aligned}\lim_{m\to\infty} \mathbb{P}\left\{\sup_{k>m} \frac{T_m(k)}{w(k/m)} \leq 1\right\} &= \lim_{m\to\infty} \mathbb{P}\left\{\sup_{k>m} \frac{T_m(k)}{v(k/m)} \leq q_{\boldsymbol{y}}^{(1-\alpha)}\right\} \\ &= \mathbb{P}\left\{\mathcal{I}_{\boldsymbol{y}} \leq q_{\boldsymbol{y}}^{(1-\alpha)}\right\} = 1 - \alpha.\end{aligned}$$

In what follows we propose three different detectors. We first introduce the ordinary CUSUM detector of Horváth et al. (2004) to monitor for changes in the d.f. Then, we adapt the same approach to monitor for changes in the contemporary copula, estimated using the empirical beta copula. Lastly, the monitoring procedure in Chapter 4 is adapted to monitor for changes in the contemporary copula estimated by the empirical beta copula.

Let $k, \ell \in \mathbb{N}$ with $k \leq \ell$ and consider the stretch $\boldsymbol{\mathcal{X}}_{k:\ell} = (\boldsymbol{X}_k, \ldots, \boldsymbol{X}_\ell)$ of d-dimensional observations. For $j \in [\![1, d]\!]$, let $F_{k:\ell,j}$ be the empirical d.f. computed

from the jth component subsample $X_{kj}, \ldots, X_{\ell j}$ as in (4.4). Recall from Chapter 4 that we fixed $p \geq 1$ points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ in $\mathbb{R}^d$, where the integer p and the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ are chosen by the user. Let $\mathcal{P}_{\boldsymbol{x}} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ and, for any $i \in \mathbb{N}$, let $\boldsymbol{Y}_i^{\mathcal{P}_{\boldsymbol{x}}} = \big(\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_1), \ldots, \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_p)\big)$, which is a p-dimensional random vector. Recall the long-run covariance matrix $\Sigma^{\mathcal{P}_{\boldsymbol{x}}}$ in (4.6) and the estimator $\Sigma_m^{\mathcal{P}_{\boldsymbol{x}}}$ based on $\boldsymbol{Y}_1^{\mathcal{P}_{\boldsymbol{x}}}, \ldots, \boldsymbol{Y}_m^{\mathcal{P}_{\boldsymbol{x}}}$. The ordinary CUSUM statistic for the empirical d.f. is

$$Q_m^{F,\mathcal{P}_{\boldsymbol{x}}}(k) = \frac{k-m}{m^{\frac{1}{2}}} \|\boldsymbol{F}_{1:m}^{\mathcal{P}_{\boldsymbol{x}}} - \boldsymbol{F}_{m+1:k}^{\mathcal{P}_{\boldsymbol{x}}}\|_{(\Sigma_m^{\mathcal{P}_{\boldsymbol{x}}})^{-1}}, \qquad k \geq m+1. \tag{5.5}$$

Recall from (2.16), that the empirical copula is the empirical d.f. of the pseudo-observations. Hence, it is straightforward to adapt the detector in (5.5) to monitor for changes in the contemporary copula C using the empirical copula.

Let $R_{ij}^{k:\ell} = (\ell-k+1)F_{k:\ell,j}(X_{ij})$ be the (maximal) rank of X_{ij} among $X_{kj}, \ldots, X_{\ell j}$. Furthermore, let

$$\boldsymbol{R}_i^{k:\ell} = \left(R_{i1}^{k:\ell}, \ldots, R_{id}^{k:\ell}\right),\ i \in [\![k, \ell]\!] \qquad \text{and} \qquad \hat{\boldsymbol{U}}_i^{k:\ell} = \frac{\boldsymbol{R}_i^{k:\ell}}{\ell - k + 1},\ i \in [\![k, \ell]\!],$$

be the multivariate ranks and the multivariate scaled ranks (or pseudo-observations) respectively, obtained from $\boldsymbol{\mathcal{X}}_{k:\ell}$. The empirical copula $C_{k:\ell}$ of $\boldsymbol{\mathcal{X}}_{k:\ell}$ at $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$ as defined in Rüschendorf (1976) is

$$C_{k:\ell}(\boldsymbol{u}) = \frac{1}{\ell - k + 1} \sum_{i=k}^{\ell} \prod_{j=1}^{d} \mathbf{1}\left(\frac{R_{ij}^{k:\ell}}{\ell - k + 1} \leq u_j\right) = \frac{1}{\ell - k + 1} \sum_{i=k}^{\ell} \mathbf{1}(\hat{\boldsymbol{U}}_i^{k:\ell} \leq \boldsymbol{u}), \tag{5.6}$$

where the last inequality between vectors is to be understood componentwise. In other words, the empirical copula in (5.6) is the empirical d.f. applied to the pseudo-observations, and a detector for the contemporary copula will thus be similar to the detector for the d.f. in (5.5). However, the empirical copula is rough and, when the sample size is small, characterised by large jumps resulting in a large bias. The empirical beta copula of $\boldsymbol{\mathcal{X}}_{k:\ell}$ is a smooth copula given by

$$C_{k:\ell}^{\beta}(\boldsymbol{u}) = \frac{1}{\ell - k + 1} \sum_{i=k}^{\ell} \prod_{j=1}^{d} F_{k:\ell, R_{ij}^{k:\ell}}(u_j), \tag{5.7}$$

where for any $r \in \{1, \ldots, n\}$, $F_{k:\ell,r}$ is the beta d.f. with parameters r and $\ell-k+2-r$. By replacing the indicator functions in the empirical copula in (5.6) by beta d.f.s, the beta copula $C_{k:\ell}^{\beta}$ in (5.7) becomes a smooth copula. In addition, the empirical beta copula has standard uniform univariate margins as long as there are not ties in the component samples of $\boldsymbol{\mathcal{X}}_{k:\ell}$, and is thus a genuine copula in that case, unlike the empirical copula (Hofert et al., 2018). The empirical beta copula can be generalised to a smooth copula model as proposed in Kojadinovic and Yi (2022).

First, let us define the long-run covariance matrix that will be used in the detector for the contemporary copula C. Let $\mathcal{P} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$. For any $i \in \mathbb{N}$, let

$$\boldsymbol{Z}_i^{\mathcal{P}} = \left(Z_i^{\boldsymbol{u}_1}, \ldots, Z_i^{\boldsymbol{u}_p}\right) \tag{5.8}$$

where, for any $\boldsymbol{u} \in [0,1]^d$,

$$Z_i^{\boldsymbol{u}} = \mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u}) \mathbf{1}(U_{ij} \leq u_j), \tag{5.9}$$

and $\Sigma^{\mathcal{P}}$ is the long-run covariance matrix of the unobservable p-dimensional time series $(\boldsymbol{Z}_i^{\mathcal{P}})_{i\in\mathbb{N}}$, that is,

$$\Sigma^{\mathcal{P}} = \mathrm{Cov}(\boldsymbol{Z}_1^{\mathcal{P}}, \boldsymbol{Z}_1^{\mathcal{P}}) + \sum_{i=2}^{\infty}\{\mathrm{Cov}(\boldsymbol{Z}_1^{\mathcal{P}}, \boldsymbol{Z}_i^{\mathcal{P}}) + \mathrm{Cov}(\boldsymbol{Z}_i^{\mathcal{P}}, \boldsymbol{Z}_1^{\mathcal{P}})\}. \tag{5.10}$$

Then, the ordinary CUSUM that is sensitive to changes in the contemporary copula is

$$Q_m^{\beta,\mathcal{P}}(k) = \frac{k-m}{m^{\frac{1}{2}}}\|\boldsymbol{C}_{1:m}^{\beta,\mathcal{P}} - \boldsymbol{C}_{m+1:k}^{\beta,\mathcal{P}}\|_{(\Sigma_m^{\beta,\mathcal{P}})^{-1}}, \qquad k \geq m+1, \tag{5.11}$$

where

- for any integers $1 \leq k \leq \ell$, $\boldsymbol{C}_{k:\ell}^{\beta,\mathcal{P}} = (C_{k:\ell}^{\beta}(\boldsymbol{u}_1), \ldots, C_{k:\ell}^{\beta}(\boldsymbol{u}_p)) \in [0,1]^p$, with the convention that $C_{k:\ell}^{\beta} = 0$ whenever $k > \ell$,
- for any $\boldsymbol{y} \in \mathbb{R}^p$, $\|\boldsymbol{y}\|_M = \sqrt{(\boldsymbol{y}^\top M \boldsymbol{y})/p}$ denotes a weighted norm of $\boldsymbol{y}$ induced by a $p \times p$ positive-definite matrix M and the integer p,
- $\Sigma_m^{\beta,\mathcal{P}}$ is an estimator of $\Sigma^{\mathcal{P}}$ based on $\boldsymbol{Z}_1^{\mathcal{P}}, \ldots, \boldsymbol{Z}_m^{\mathcal{P}}$, and we assume that $\Sigma_m^{\beta,\mathcal{P}}$ is almost surely positive-definite for all $m \in \mathbb{N}$.

The following condition is required as we continue.

Condition 5.1.1 (Smooth partial derivatives)**.** *For any $j \in [\![1,d]\!]$, the partial derivative $\dot{C}_j = \partial C/\partial u_j$ exists and is continuous on the set $V_{d,j} = \{\boldsymbol{u} \in [0,1]^d : u_j \in (0,1)\}$.*

For any $j \in [\![1,d]\!]$, we assume that $\dot{C}_j = 0$ on the set $\{\boldsymbol{u} \in [0,1]^d : u_j \in \{0,1\}\}$.

Let '$\rightsquigarrow$' denote convergence in distribution (weak convergence), and let $\|\cdot\|_2$ be the Euclidean norm. We expect that the following holds.

Conjecture 5.1.2 (Weak convergence of $Q_m^{\mathcal{P}}$)**.** *Under H_0 in* (5.2) *and conditions similar to 4.2.5 and 5.1.1, and if $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$,*

$$\sup_{k>m} \frac{m}{k} Q_m^{\beta,\mathcal{P}}(k) \rightsquigarrow \mathcal{K} = \sup_{t\in[0,1]} \|\boldsymbol{W}(t)\|_2,$$

where $Q_m^{\beta,\mathcal{P}}$ is defined in (5.11)*, and $\boldsymbol{W}$ is a p-dimensional standard Brownian motion.*

The supremum on the left is over integers k, and the limiting variable $\mathcal{K}$ depends only on the choice of the number of points p, not on the dimension d, nor on the choice of the set of points $\mathcal{P}$.

We also propose to use the approach in Chapter 4 for the empirical beta copula.

$$D_m^{\beta,\mathcal{P}}(k) = \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}}\|\boldsymbol{C}_{1:j}^{\beta,\mathcal{P}} - \boldsymbol{C}_{j+1:k}^{\beta,\mathcal{P}}\|_{(\Sigma_m^{\beta,\mathcal{P}})^{-1}}, \qquad k \geq m+1, \tag{5.12}$$

and expect weak convergence of the detector in (5.12) to hold under regularity conditions.

Conjecture 5.1.3 (Weak convergence of $D_m^{\beta,\mathcal{P}}$)**.** *Under H_0 in* (5.2)*, Condition 5.1.1, Condition 4.2.5 and conditions similar to Dette and Gösmann (2020, Assumption 3.2) and Gösmann et al. (2021, Assumption 2.5), and if $\Sigma_m^{\beta,\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$, for any $\eta > 0$,*

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\beta,\mathcal{P}}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

From Proposition 4.2.10 in Chapter 4, we know imposing $\eta > 0$ is sufficient to ensure that the limiting variable is almost surely finite. The limiting random variables do not depend on the dimension d of the initial distribution nor on the choice of the evaluation points $\mathcal{P} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$.

It remains to be proven that Conjectures 5.1.2 and 5.1.3 hold. Furthermore, simulation studies should confirm that the proposed tests hold their level in finite samples, and are able to detect departures from stationarity resulting from changes in the contemporary copula.

5.2 Conclusions and directions for future research

The aim of this thesis was to develop new methods for sequential change-point detection and fast-inference methods to estimate dependence structures in high-dimensional data. In Chapter 2 we introduced the necessary tools that were used in the rest of this thesis. We presented Gaussian factor models and (factor) copula models, and their respective estimation procedures. We also provided an overview of the different approaches to sequential change point detection.

In Chapter 3 we presented the first original work of this thesis. We outlined a fast inference method for very high-dimensional one-factor copula models. In a first step, margins are modelled separately, and transformed to the normal scale. Then, the factor loadings and the correlation between the error terms were estimated using a GMM approach, and for which we had to impose constraints on the factor loading to ensure their unique identifiability. Next, we estimated the Gaussian factors by minimising the $\mathscr{L}_2$ norm between the weighted observations and the true Gaussian factors, where it is of paramount importance that the sample size tends to infinity before setting the dimension sufficiently large. In the last step, we estimated the one-factor copula parameter using a maximum likelihood approach in combination with a proxy for the common factor as in Krupskii and Joe (2022). We established the asymptotic properties of our approach and showed that the estimation procedure works for a model with up to three Gaussian factors. Lastly, a practical application looked at the ability of the model to capture dependence between a selection of stocks in different industry sectors. The results have been published in Verhoijsen and Krupskiy (2022). A current limitation of the contribution is the the model's robustness and performance under misspecification. We see several possible avenues for future research contributions. First, future research could investigate the theoretical and practical implications of the model when it is not correctly specified. Furthermore, another direction of future research could try to generalise the estimation procedure to allow for an arbitrary number of Gaussian factors. In addition, one could explore the validity of the model for datasets whose margins are discrete, by using one-factor copula models for discrete data as in Nikoloulopoulos and Joe (2015). A last extension could focus on generalising the approach to allow for a

residual dependence structure that includes non-overlapping groups in the factor copula model.

In Chapter 4 we presented a sequential open-end change point detection mechanism to monitor for changes in the d.f. of possibly multivariate data. We started by establishing a detector and showed its asymptotic properties under the null hypothesis and alternatives to the null. We showed that, by using a Mahalanobis-like norm in the definition of our detector, its limiting distribution does not depend on the characteristics of the initial sample, nor on the choice of the evaluation points, but only on the number of evaluation points. We discussed how evaluation points can be chosen in the univariate case, and proposed a procedure for the selection of evaluation points in the multivariate case. Extensive Monte Carlo simulations illustrated the finite-sample behaviour of our approach, and its usefulness in detecting changes in the mean, variance and distribution of low-dimensional datasets. Lastly, the monitoring scheme was applied to the stock returns of the NASDAQ composite index. Note that for this data example, we make no claim regarding the stationarity of the learning sample, which would need to be tested in practice. Lastly, a future research project could compare the change-point detection procedure developed in this thesis to a monitoring procedure based on the likelihood ratio principle.

The results of this work have been published in Holmes et al. (2023), and corresponding code has been included in the R package `npcp` (Kojadinovic and Verhoijsen, 2022). The current implementation in the R package `npcp` is not optimised for longer term monitoring, and future research could focus on improving this. Furthermore, while our method is capable of testing for possible change-points, we did not focus on the estimation of the actual instant that the change occurs. Lastly, estimation of the high quantiles of the limiting random variable is currently done using a practical but heuristic approach, and further investigation into the validity of this approach is warranted.

Bibliography

Andreou, E. and Ghysels, E. (2009). Structural breaks in financial time series. *Handbook of financial time series*, pages 839–870.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Andrews, D. and Monahan, J. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.*, 34(1):1–16.

Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics and Probability Letters*, 67(3):221 – 231.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal*, 9(3):373–403.

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, 28(4):804–837.

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika*, 77:669–687.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

Cherubini, G., Vecchiato, W., and Luciano, E. (2004). *Copula models in finance*. Wiley, New-York.

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

Creal, D. D. and Tsay, R. S. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189(2):335–345.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley and Sons, Chichester, UK.

Davydov, Y. A. and Lifshits, M. A. (1984). The fibering method in some probability problems. In *Probability theory. Mathematical statistics. Theoretical cybernetics, Vol. 22*, Itogi Nauki i Tekhniki, pages 61–157, 204. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow.

Dedecker, J., Merlevède, F., and Rio, E. (2014). Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electron. J. Probab.*, 19:no. 9, 56.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. 5th Ser.*, 65:274–292.

Deheuvels, P. (1981). A non parametric test for independence. *Publications de l'Institut de Statistique de l'Université de Paris*, 26:29–50.

Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In Dehling, H., Mikosch, T., and Sorensen, M., editors, *Empirical process techniques for dependent data*, pages 1–113. Birkhäuser, Boston.

Dette, H. and Gösmann, J. (2020). A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*, 115(531):1361–1377.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, 77:423–432.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk management: Value at risk and beyond*, pages 176–223. Cambridge University Press.

Fermanian, J.-D. and Scaillet, O. (2005). Some statistical pitfalls in copula modelling for financial applications. In Klein, E., editor, *Capital formation, governance and banking*, pages 59–74. Nova Science.

Fremdt, S. (2015). Page's sequential procedure for change-point detection in time series regression. *Statistics*, 49(1):128–155.

Genest, C. (2021). A tribute to abe sklar. *Dependence Modeling*, 9(1):200–224.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of copulas in finance. *European Journal of Finance*, 15:609–618.

Gösmann, J., Kley, T., and Dette, H. (2021). A new approach for open-end sequential change point monitoring. *Journal of the Time Series Analysis*, 42(1):63–84.

Gösmann, J., Stoehr, C., Heiny, J., and Dette, H. (2022). Sequential change point detection in high dimensional time series. *Electronic Journal of Statistics*, 16(1):3608 – 3671.

Groen, J. J., Kapetanios, G., and Price, S. (2013). Multivariate methods for monitoring structural change. *Journal of Applied Econometrics*, 28(2):250–274.

Härdle, W., Simar, L., et al. (2007). *Applied multivariate statistical analysis*, volume 22007. Springer.

Hawkins, D. M. and Olwell, D. H. (2012). *Cumulative sum charts and charting for quality improvement*. Springer Science & Business Media.

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of quality technology*, 35(4):355–366.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2018). *Elements of copula modeling with R*. Springer.

Holmes, M. and Kojadinovic, I. (2021). Open-end nonparametric sequential change-point detection based on the retrospective cusum statistic. *Electron. J. Statist.*

Holmes, M., Kojadinovic, I., and Verhoijsen, A. (2023). Multi-purpose open-end monitoring procedures for multivariate observations based on the empirical distribution function. *Journal of Time Series Analysis*.

Horváth, L., Kokoszka, P., and Steinebach, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis*, 68(1):96–119.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126(1):225 – 251.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall, London.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419.

Joe, H. (2015). *Dependence Modeling with Copulas*. CRC Press.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. Technical report, Department of Statistics, University of British Columbia.

Johnson, R. A. and Wichern, D. W. (2002). *Applied multivariate statistical analysis*, volume 5. Prentice Hall Upper Saddle River, NJ.

Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Springer, London.

Kent, J., Bibby, J., and Mardia, K. (1979). *Multivariate analysis*. Academic Press Amsterdam.

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electron. J. Statist.*, 12(1):1579–1613.

Kojadinovic, I. and Verdier, G. (2021). Nonparametric sequential change-point detection for multivariate time series based on empirical distribution functions. *Electron. J. Statist.*, 15(1):773–829.

Kojadinovic, I. and Verhoijsen, A. (2022). *npcp: Some Nonparametric Tests for Change-Point Detection in Possibly Multivariate Observations*. R package version 0.2-4.

Kojadinovic, I. and Yi, B. (2022). A class of smooth, possibly data-adaptive nonparametric copula estimators containing the empirical beta copula. *Journal of Multivariate Analysis (in press)*.

Kosorok, M. (2008). *Introduction to empirical processes and semiparametric inference*. Springer, New York.

Kreuzer, A. and Czado, C. (2021). Bayesian inference for a single factor copula stochastic volatility model using Hamiltonian Monte Carlo. *Econometrics and Statistics*, page In Press.

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

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

Krupskii, P. and Joe, H. (2021). Approximate likelihood with proxy variables for parameter estimation in high-dimensional factor copula models. *Statistical Papers*, pages 1–27.

Krupskii, P. and Joe, H. (2022). Approximate likelihood with proxy variables for parameter estimation in high-dimensional factor copula models. *Statistical Papers*, 63(2):543–569.

Kuelbs, J. and Philipp, W. (1980). Almost Sure Invariance Principles for Partial Sums of Mixing B-Valued Random Variables. *The Annals of Probability*, 8(6):1003 – 1036.

Lai, T. (2001). Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, 11:303–351.

Li, B. and Genton, M. (2013). Nonparametric identification of copula structures. *Journal of the American Statistical Association*, 108(502):666–675.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton University Press, New Jersey.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and Applications*, 29(2):309–315.

Montgomery, D. (2007). *Introduction to statistical quality control*. John Wiley & Sons.

Nelsen, R. (2006). *An introduction to copulas*. Springer, New-York. Second edition.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Nguyen, H., Ausín, M. C., and Galeano, P. (2020). Variational inference for high dimensional structured factor copulas. *Computational Statistics & Data Analysis*, 151:107012.

Nikoloulopoulos, A. K. and Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, 80(1):126–150.

Oh, D. H. and Patton, A. J. (2013). Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502):689–700.

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

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

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527.

Paparoditis, E. and Politis, D. (2001). Tapered block bootstrap. *Biometrika*, 88(4):1105–1119.

Pape, K., Wied, D., and Galeano, P. (2016). Monitoring multivariate variance changes. *Journal of Empirical Finance*, 39:54–68.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Perron, P. et al. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics*, 1(2):278–352.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rio, E. (1998). Processus empiriques absolument réguliers et entropie universelle. *Probability Theory and Related Fields*, 111:585–608.

Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-response analysis using R. *PLOS ONE*, 10(e0146021).

Roberts, S. (2000). Control chart tests based on geometric moving averages. *Technometrics*, 42(1):97–101.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics*, 4:912–923.

Schamberger, B., Gruber, L., and Czado, C. (2017). Bayesian inference for latent factor copulas and application to financial risk forecasting. *Econometrics*, 5(2):21.

Schreyer, M., Paulin, R., and Trutschnig, W. (2017). On the exact region determined by kendall's τ and spearman's ρ. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(2):613–633.

Segers, J. (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18:764–782.

Segers, J., Sibuya, M., and Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis*, 155:35–51.

Shewhart, W. A. (1931). Statistical method from an engineering viewpoint. *Journal of the American Statistical Association*, 26(175):262–269.

Shorack, G. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Subbarao, C., Subbarao, N., and Chandu, S. (1996). Characterization of groundwater contamination using factor analysis. *Environmental Geology*, 28(4):175–180.

Tan, B. K., Panagiotelis, A., and Athanasopoulos, G. (2019). Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics*, 28(1):155–173.

Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press.

Van Loan, C. F. and Golub, G. (1996). Matrix computations (Johns Hopkins studies in mathematical sciences). *Matrix Computations*.

Verhoijsen, A. and Krupskiy, P. (2022). Fast inference methods for high-dimensional factor copulas. *Dependence Modeling*, 10(1):270–289.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

Zeileis, A., Köll, S., and Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software*, 95(1):1–36.

Ziegel, E. (1987). *Numerical recipes: The art of scientific computing*. Taylor & Francis.

Appendix A

Appendix for Chapter 3

A.1 Validity of Condition 3.3.1(iii)

We can illustrate Condition 3.3.1(iii) in the case where the factor loadings are uniformly spread out in the interval $[a_k, b_k]$, with $-1 < a_k < b_k < 1$,

$$\lambda_{lk} = a_k + \frac{\sigma_k(l)}{d}(b_k - a_k),$$

where $\sigma_k(\cdot)$ is a permutation of indexes $\{1, 2, \ldots, d\}$. Recall that $\delta_{lk} = \lambda_{lk}/\gamma_l$.

We first consider the case $p = 1$ and we can assume without loss of generality $\sigma_1(j) = j$ so that

$$\delta_{l1} = \frac{a_1 + (l/d)(b_1 - a_1)}{\left[1 - \{a_1 + (l/d)(b_1 - a_1)\}^2\right]^{1/2}}.$$

Let $g_k(x) = a_k + x(b_k - a_k)$. We find that

$$\begin{aligned}\lim_{d\to\infty} ||\Delta_1||^2 &= \lim_{d\to\infty} \frac{1}{d^2}\sum_{l=1}^{d}\sum_{m=1}^{d}(\delta_{l1} - \delta_{m1})^2 = 2\lim_{d\to\infty}\frac{1}{d}\sum_{l=1}^{d}\delta_{l1}^2 - 2\left(\lim_{d\to\infty}\frac{1}{d}\sum_{l=1}^{d}\delta_{l1}\right)^2 \\ &= 2\int_0^1 \frac{g_1(x)^2}{1 - g_1(x)^2}dx - 2\left(\int_0^1 \frac{g_1(x)}{\{1 - g_1(x)^2\}^{1/2}}dx\right)^2. \qquad \text{(A.1)}\end{aligned}$$

We illustrate the validity of Condition 3.3.1(iii) in Table A.1 and A.2 by plugging a range of values for a_1 and b_1 into (A.1).

Table A.1: Values for the integral in (A.1) for different values of (a_1, b_1).

a_1	-0.99	-0.80	-0.50	0.00	0.50	0.80
b_1	0.99	0.99	0.99	0.99	0.99	0.99
$\lim_{d\to\infty} \|\|\Delta_1\|\|^2$	3.347	2.053	1.816	1.841	2.183	2.627

Table A.2: Values for the integral in (A.1) for different values of (a_1, b_1).

a_1	-0.20	-0.15	-0.10	0.00	0.10	0.15
b_1	0.20	0.20	0.20	0.20	0.20	0.20
$\lim_{d\to\infty} \|\|\Delta_1\|\|^2$	0.0273	0.0208	0.0153	0.0069	0.0018	0.0005

The limiting norm of the vector Δ_1 is smaller if a_1 and b_1 are close. In practical applications, it indicates that more accurate parameter estimates can be obtained if the values of the loadings λ_{l1} are more spread.

Now we consider the case $p = 2$ with

$$\delta_{lk} = \frac{a_k + (\sigma_k(l)/d)(b_k - a_k)}{\left[1 - \{a_1 + (\sigma_1(l)/d)(b_1 - a_1)\}^2 - \{a_2 + (\sigma_2(l)/d)(b_2 - a_2)\}^2\right]^{1/2}}, \quad k = 1, 2.$$

Further, we consider the case when the loadings $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are monotonically dependent with $\sigma_1(l) = \sigma_2(l)$ (case 1), and when they are independent (case 2).

In the first case, for $k = 1, 2$,

$$\lim_{d\to\infty} ||\Delta_k||^2 = 2\int_0^1 \frac{g_k(x)^2}{1 - g_1(x)^2 - g_2(x)^2}dx - 2\left(\int_0^1 \frac{g_k(x)}{\{1 - g_1(x)^2 - g_2(x)^2\}^{1/2}}dx\right)^2,$$

and in the second case, for $k = 1, 2$,

$$\begin{aligned}\lim_{d\to\infty} ||\Delta_k||^2 = 2\int_0^1\int_0^1 \frac{g_k(x_k)^2}{1 - g_1(x_1)^2 - g_2(x_2)^2}dx_1dx_2 \\ - 2\left(\int_0^1\int_0^1 \frac{g_k(x_k)}{\{1 - g_1(x_1)^2 - g_2(x_2)^2\}^{1/2}}dx_1dx_2\right)^2.\end{aligned}$$

Furthermore, we need to check if $\cos^2(\phi_{1,2}) = \langle\Delta_1, \Delta_2\rangle^2/(||\Delta_1||^2||\Delta_2||^2)$ is bounded away from one as $d \to \infty$ where $\langle\Delta_1, \Delta_2\rangle$ is the inner product. We find:

$$\begin{aligned}\langle\Delta_1, \Delta_2\rangle &= \frac{1}{d^2}\left\{\sum_{l=1}^d\sum_{m=1}^d(\delta_{l1} - \delta_{m1})(\delta_{l2} - \delta_{m2})\right\} \\ &= \frac{2}{d}\sum_{l=1}^d \delta_{l1}\delta_{l2} - 2\left(\frac{1}{d}\sum_{l=1}^d \delta_{l1}\right)\left(\frac{1}{d}\sum_{l=1}^d \delta_{l2}\right).\end{aligned}$$

It implies that, in case 1,

$$\begin{aligned}\lim_{d\to\infty}\langle\Delta_1, \Delta_2\rangle = 2\int_0^1 \frac{g_1(x)g_2(x)}{1 - g_1(x)^2 - g_2(x)^2}dx \\ - 2\prod_{k=1}^2\left(\int_0^1 \frac{g_k(x)}{\{1 - g_1(x)^2 - g_2(x)^2\}^{1/2}}dx\right).\end{aligned}$$

and in case 2,

$$\begin{aligned}\lim_{d\to\infty}\langle\Delta_1, \Delta_2\rangle = 2\int_0^1\int_0^1 \frac{g_1(x_1)g_2(x_2)}{1 - g_1(x_1)^2 - g_2(x_2)^2}dx_1dx_2 \\ - 2\prod_{k=1}^2\left(\int_0^1\int_0^1 \frac{g_k(x_k)}{\{1 - g_1(x_1)^2 - g_2(x_2)^2\}^{1/2}}dx_1dx_2\right).\end{aligned}$$

Tables A.3 and A.4 show the limiting values for some pairs (a_1, b_1) and (a_2, b_2) for case 1 and 2, respectively.

Table A.3: Limiting values of $||\Delta_k||^2$, $k = 1, 2$ and $\cos^2(\phi_{1,2})$ for different values of (a_1, b_1) and (a_2, b_2); case 1.

(a_1, b_1)	$(-0.5, 0.6)$	(0.2, 0.6)	(0.2, 0.6)	(0.2, 0.4)	(0.2, 0.4)
(a_2, b_2)	$(-0.3, 0.7)$	(0.2, 0.6)	(0.0, 0.4)	(0.1, 0.9)	(0.4, 0.8)
$\lim_{d\to\infty} \|\|\Delta_1\|\|^2$	0.427	0.119	0.067	0.219	0.061
$\lim_{d\to\infty} \|\|\Delta_2\|\|^2$	0.419	0.119	0.051	1.359	0.243
$\lim_{d\to\infty} \cos^2(\phi_{1,2})$	0.997	1.000	0.999	0.997	1.000

Table A.4: Limiting values of $||\Delta_k||^2$, $k = 1, 2$ and $\cos^2(\phi_{1,2})$ for different values of (a_1, b_1) and (a_2, b_2); case 2.

(a_1, b_1)	$(-0.5, 0.6)$	(0.2, 0.6)	(0.2, 0.6)	(0.2, 0.4)	(0.2, 0.4)
(a_2, b_2)	$(-0.3, 0.7)$	(0.2, 0.6)	(0.0, 0.4)	(0.1, 0.9)	(0.4, 0.8)
$\lim_{d\to\infty} \|\|\Delta_1\|\|^2$	0.319	0.075	0.054	0.071	0.028
$\lim_{d\to\infty} \|\|\Delta_2\|\|^2$	0.296	0.075	0.040	0.796	0.180
$\lim_{d\to\infty} \cos^2(\phi_{1,2})$	0.004	0.189	0.041	0.703	0.416

Note that in case 1, $\cos^2(\phi_{1,2})$ is very close to one, so the convergence of the parameter estimates to their true values may be very slow. In fact, $\mathbf{\Lambda_1}$ and $\mathbf{\Lambda_2}$ are linearly dependent with $\mathbf{\Lambda_1} = \mathbf{\Lambda_2}$ when $(a_1, b_1) = (a_2, b_2) = (0.2, 0.6)$ and $\mathbf{\Lambda_1} = 0.5\mathbf{\Lambda_2}$ with $(a_1, b_1) = (0.2, 0.4)$ and $(a_2, b_2) = (0.4, 0.8)$. Condition 3.3.1(iii) fails in this case.

In case 2, $\cos^2(\phi_{1,2})$ is far from one in all cases, and accurate parameter estimates can be obtained in this case.

Similar results can be obtained when $p = 3$, so they are not reported here.

A.2 Proofs

In this section, proofs are provided for theoretical results presented in Section 3.3.

Proof of Theorem 3.3.2. (i) In order for the the parameter vector $(\mathbf{\Lambda}_0, \eta_0)$ to be unique, one has to impose $p(p-1)/2$ restrictions on the factor loadings $\mathbf{\Lambda}_0$. Otherwise, one can always find a rotation $\mathbf{\Lambda}_{\text{rot}} = \mathbf{\Lambda}\mathbf{A}$, with $\mathbf{A}$ a $p \times p$ orthogonal matrix, such that the variance-covariance matrix remains the same.

(ii) This is trivially satisfied, as the values of the correlation matrix by definition are restricted to the interval $[-1, 1]$.

(iii) This is trivially satisfied, as the correlation matrix is continuous at each $(\mathbf{\Lambda}, \eta)$ in the parameter space.

(iv) Since the correlation parameter $\sigma_{ij} = \text{Corr}\,(X_i, X_j)$ is bounded by the interval $[-1, 1]$, one can show that

$$\mathbb{E}\left[\sup_{(\mathbf{\Lambda},\eta)\in\Theta} \|f\,(\mathbf{X}, \mathbf{\Lambda}, \eta)\|\right] \leq d.$$

■

Proof of Theorem 3.3.3. (i) The GMM estimator is consistent if the conditions from Theorem 3.3.2 are satisfied.

(ii) This condition is satisfied, as the parameter space for the candidate loadings λ_{jk} and correlation between the residuals, η, is restricted to the interval $[-1, 1]$.

(iii) We know that $\nabla_{(\mathbf{\Lambda},\eta)} f(\mathbf{X}_i, \mathbf{\Lambda}, \eta) = \nabla_{(\mathbf{\Lambda},\eta)} \mathbf{\Sigma}_{\boldsymbol{X}}$. So for every element $\sigma_{jl} = \sum_{k=1}^{p} \lambda_{jk}\lambda_{lk} + \eta\gamma_j\gamma_l$, we have

$$\frac{\partial}{\partial \lambda_{jk}}\sigma_{jl} = \gamma_l(\delta_{lk} - \eta\delta_{jk}), \quad \frac{\partial}{\partial \eta}\sigma_{jl} = \gamma_j\gamma_l,$$

The derivatives exist and are finite under Condition 3.3.1(i).

(iv) It is easy to see that $\mathbb{E}\left[\|f(\mathbf{X}_i, \mathbf{\Lambda}, \eta)\|^2\right] \leq d^2$.

(v) Note that

$$\left(\frac{\partial}{\partial \lambda_{jk}}\sigma_{jl}\right)^2 \leq 2\delta_{lk}^2 + 2\delta_{jk}^2, \quad \left(\frac{\partial}{\partial \lambda_{lk}}\sigma_{jl}\right)^2 \leq 2\delta_{lk}^2 + 2\delta_{jk}^2, \quad \left(\frac{\partial}{\partial \eta}\sigma_{jl}\right)^2 \leq 1.$$

Under Condition 3.3.1(i),

$$\sup_{(\mathbf{\Lambda},\eta)\in\Theta} \left\|\nabla_{(\mathbf{\Lambda},\eta)} f(\mathbf{X}_i, \mathbf{\Lambda}, \eta)\right\|^2 \leq 8\sum_{k=1}^{p}\sum_{l=1}^{d}\delta_{lk}^2 + pd = pd(8\xi_U + 1).$$

(vi) To check this condition, we have to show that the matrix

$$G^\top G = \left(\nabla_{(\mathbf{\Lambda},\eta)} \mathbf{\Sigma}_{\boldsymbol{X}}\right)^\top \nabla_{(\mathbf{\Lambda},\eta)} \mathbf{\Sigma}_{\boldsymbol{X}}.$$

is invertible. The condition is illustrated for a model with one Gaussian factor, and the proof is similar in the general case. The model with one factor has the following derivatives:

$$\frac{\partial \sigma_{ij}}{\partial \lambda_i} = \lambda_j - \eta\frac{\gamma_j}{\gamma_i}\lambda_i, \qquad \frac{\partial \sigma_{ij}}{\partial \lambda_j} = \lambda_i - \eta\frac{\gamma_i}{\gamma_j}\lambda_j, \qquad \frac{\partial \sigma_{ij}}{\partial \eta} = \gamma_i\gamma_j,$$

for $i, j = 1, \ldots, d$ and $i \neq j$. Then it it suffices to show that the following does not hold

$$a_i\left(\lambda_j - \eta\frac{\gamma_j}{\gamma_i}\lambda_i\right) + a_j\left(\lambda_i - \eta\frac{\gamma_i}{\gamma_j}\lambda_j\right) = \gamma_i\gamma_j$$
$$\tilde{a}_i(\delta_j - \eta\delta_i) + \tilde{a}_j(\delta_i - \eta\delta_j) = 1$$

where $\tilde{a}_i = a_i/\gamma_i$, and $\delta_j = \lambda_j/\gamma_j$. Then for $i \neq j$, and $i \neq k$ we have the following system of equations:

$$\begin{cases} \tilde{a}_i(\delta_j - \eta\delta_i) + \tilde{a}_j(\delta_i - \eta\delta_j) = 1 \\ \tilde{a}_i(\delta_k - \eta\delta_i) + \tilde{a}_k(\delta_i - \eta\delta_k) = 1 \end{cases}$$

subtracting the two equations gives another set of equations, where $j \neq l$ and $k \neq l$

$$\begin{cases} \tilde{a}_i(\delta_j - \delta_k) + \delta_i(\tilde{a}_j - \tilde{a}_k) + \eta(\tilde{a}_k\delta_k - \tilde{a}_j\delta_j) = 0 \\ \tilde{a}_l(\delta_j - \delta_k) + \delta_l(\tilde{a}_j - \tilde{a}_k) + \eta(\tilde{a}_k\delta_k - \tilde{a}_j\delta_j) = 0 \end{cases}$$

again subtracting the above equations gives

$$(\tilde{a}_i - \tilde{a}_l)(\delta_j - \delta_k) = -(\tilde{a}_j - \tilde{a}_k)(\delta_i - \delta_l). \tag{A.2}$$

Hence, there are two cases. In the first case, $\tilde{a}_i - \tilde{a}_l = 0$, which implies that $\delta_i - \delta_l = 0$. Otherwise, all differences $\tilde{a}_i - \tilde{a}_l$ and $\tilde{a}_j - \tilde{a}_k$ would be zero, for every possible combination of the pairs (i, l) and (j, k). In the second case, $\delta_i - \delta_l$ and $\delta_j - \delta_k$ are different from zero, and we can rewrite the display in (A.2) as

$$\frac{\tilde{a}_i - \tilde{a}_l}{\delta_i - \delta_l} = -\frac{\tilde{a}_j - \tilde{a}_k}{\delta_j - \delta_k}.$$

The numerator and denominator on both sides of the above display are nonzero, even though the ratios are the same. However, looping through every possible combination of the pairs (i, l) and (j, k), the only valid result is that all differences $\tilde{a}_i - \tilde{a}_l$ are zero, which implies that all δ's are equal. However, this contradicts Condition 3.3.1(iii), which states that there can only be up to $O(d^2)$ terms that are zero, which concludes the proof.

■

Proof of Lemma 3.3.4. The optimal weight vector $\boldsymbol{w}_{kd}^{\star}$, can be derived by differentiating

$$\mathbb{E}\{(\hat{Z}_{kd} - Z_k)^2\} = \boldsymbol{w}_{kd}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{w}_{kd} - 2\boldsymbol{w}_{kd}^{\top}\boldsymbol{\lambda}_k + 1 \tag{A.3}$$

with respect to $\boldsymbol{w}_{kd}$ for each $k = 1, \ldots, p$. Recall that $\mathbb{E}(\hat{Z}_{kd}) = \mathbb{E}(Z_k) = 0$. Then assuming that $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ is invertible, one can obtain the optimal weight vector $\boldsymbol{w}_{kd}^{\star}$ by optimising (A.3) with respect to $\boldsymbol{w}_{kd}$ subject to constraint $\mathrm{Var}(\hat{Z}_{kd}) = \boldsymbol{w}_{kd}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{w}_{kd} = 1$. ■

Proof of Theorem 3.3.5. After substituting $\boldsymbol{w}_{kd}^{\star} = (\boldsymbol{\lambda}_k^{\top}\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\lambda})^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_k$ back into (A.3), one gets

$$\lim_{d\to\infty} \mathbb{E}\{(\hat{Z}_{kd} - Z_k)^2\} = 2 - 2\lim_{d\to\infty} \left(\boldsymbol{\lambda}_k^{\top}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_k\right)^{1/2}.$$

Therefore we need to show that $\lim_{d\to\infty} \boldsymbol{\lambda}_k^{\top}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_k = 1$. Using Woodbury's formula, one can show that

$$\begin{aligned} \boldsymbol{\lambda}_k^{\top}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_k^{\top}&\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\lambda}_k \\ &- \boldsymbol{\lambda}_k^{\top}\left(\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}^{\top}\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\Lambda} + \mathbf{I}_p\right)^{-1} \\ &\times \boldsymbol{\Lambda}^{\top}\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\lambda}_k. \end{aligned}$$

Without loss of generality, convergence of the factor estimates is illustrated for $k = 1$. However, the results hold for any $k = 1, \ldots, p$. Furthermore define $b_{qr} = \boldsymbol{\lambda}_q^\top \left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} \boldsymbol{\lambda}_r$ and $a_k = b_{kk}$ for $k, r, q = 1, \ldots, p$. We find:

$$a_k = \frac{1 + (d-2)\,\eta}{1 + (d-2)\,\eta - (d-1)\,\eta^2} \sum_{l=1}^{d} \delta_{lk}^2 - \frac{\eta}{(1 + (d-2)\,\eta - (d-1)\,\eta^2)} \sum_{\substack{l=1 \\ l \neq m}}^{d} \delta_{lk}\delta_{mk}$$

$$b_{qr} = \frac{1 + (d-2)\,\eta}{1 + (d-2)\,\eta - (d-1)\,\eta^2} \sum_{l=1}^{d} \delta_{lq}\delta_{lr} - \frac{\eta}{(1 + (d-2)\,\eta - (d-1)\,\eta^2)} \sum_{\substack{l=1 \\ l \neq m}}^{d} \delta_{lq}\delta_{mr}$$

where $\delta_{lk} = \lambda_{lk}/\gamma_l$. Then $a_1 = \boldsymbol{\lambda}_1^\top \left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} \boldsymbol{\lambda}_1$ is a scalar.

Proof for a one-factor model ($p = 1$, and $k = 1$). In the case where $p = 1$, it follows that $\boldsymbol{\Lambda} = \boldsymbol{\lambda}_1$. Hence

$$\boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} \boldsymbol{\lambda}_1 = \frac{a_1}{1 + a_1}.$$

It remains to be shown that $a_1 \to \infty$ as $d \to \infty$. If Condition 3.3.1(i) and Condition 3.3.1(ii) are satisfied, then

$$\begin{aligned}
a_1/d &= \frac{1}{1-\eta} \left\{ \frac{1}{d} \sum_{l=1}^{d} \delta_{l1}^2 - \frac{1}{d^2} \left(\sum_{l=1}^{d} \sum_{m=1}^{d} \delta_{l1}\delta_{m1} - \sum_{l=1}^{d} \delta_{l1}^2 \right) \right\} + O(1/d) \\
&= \frac{1}{1-\eta} \left\{ \frac{1}{d} \sum_{l=1}^{d} \delta_{l1}^2 - \frac{1}{d^2} \left(\sum_{l=1}^{d} \delta_{l1} \right)^2 \right\} + O(1/d) \\
&= \frac{1}{1-\eta} \frac{1}{2d^2} \sum_{l=1}^{d} \sum_{m=1}^{d} \left(\delta_{l1} - \delta_{m1}\right)^2 + O(1/d) = \frac{0.5||\Delta_1||^2}{1-\eta} + O(1/d) \\
&> \frac{0.5\xi_0^2}{1-\eta} + O(1/d),
\end{aligned}$$

where Condition 3.3.1(iii) is used in the last step. It follows that $\lim_{d\to\infty} \boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} \boldsymbol{\lambda}_1 = 1$.

Proof for a two-factor model ($p = 2$, and $k = 1$). In the case of a two-factor model,

$$\boldsymbol{\Lambda}^\top \left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} \boldsymbol{\Lambda} + \mathbf{I}_2 = \begin{pmatrix} 1 + a_1 & b_{12} \\ b_{12} & 1 + a_2 \end{pmatrix},$$

Computing the inverse, we get

$$\left(\boldsymbol{\Lambda}^\top \left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} \boldsymbol{\Lambda} + \mathbf{I}_2\right)^{-1} = \frac{1}{(1 + a_1)\,(1 + a_2) - b_{12}^2} \begin{pmatrix} 1 + a_2 & -b_{12} \\ -b_{12} & 1 + a_1 \end{pmatrix}.$$

Lastly,

$$\boldsymbol{\lambda}_1^\top \left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1}\boldsymbol{\Lambda} = \begin{pmatrix} a_1 & b_{12}\end{pmatrix}.$$

Putting everything together, we get

$$\boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_1 = \frac{a_1\,(1+a_2) - b_{12}^2}{(1+a_1)\,(1+a_2) - b_{12}^2} = 1 - \frac{1}{1+a_1}\left\{1 - \frac{b_{12}^2}{(1+a_1)\,(1+a_2)}\right\}^{-1}.$$

It is required to show that $\alpha_{12} = 1 - b_{12}^2/[(1+a_1)\,(1+a_2)]$ is bounded from below from zero. To start with,

$$\begin{aligned} b_{12}/d &= \frac{1}{1-\eta}\left(\frac{1}{d}\sum_{l=1}^{d}\delta_{l1}\delta_{l2} - \frac{1}{d^2}\sum_{\substack{l=1\\m\neq l}}^{d}\delta_{l1}\delta_{m2}\right) + O(1/d) \\ &= \frac{1}{1-\eta}\left\{\frac{1}{d}\sum_{l=1}^{d}\delta_{l1}\delta_{l2} - \left(\frac{1}{d}\sum_{l=1}^{d}\delta_{l1}\right)\left(\frac{1}{d}\sum_{m=1}^{d}\delta_{m2}\right)\right\} + O(1/d) \\ &= \frac{0.5\langle\Delta_1,\Delta_2\rangle}{1-\eta} + O(1/d) = \frac{0.5||\Delta_1||||\Delta_2||\cos(\phi_{1,2})}{1-\eta} + O(1/d). \end{aligned}$$

Then Condition 3.3.1(iii) implies that $\alpha_{1,2} = 1 - \cos^2(\phi_{1,2}) + O(1/d) = \sin^2(\phi_{1,2}) + O(1/d) > \sin^2(\xi_0/2) + O(1/d)$.

Proof for a three-factor model ($p = 3$, and $k = 1$). For a three-factor model, we get the following expression:

$$\boldsymbol{\Lambda}^\top \left(\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}\boldsymbol{\Gamma}^{\frac{1}{2}}\right)^{-1}\boldsymbol{\Lambda} + \mathbf{I}_p = \begin{pmatrix} 1+a_1 & b_{12} & b_{13} \\ b_{12} & 1+a_2 & b_{23} \\ b_{13} & b_{23} & 1+a_3 \end{pmatrix},$$

After some calculations, we find:

$$\boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_1 = 1 - \frac{b_{123}^{-1}}{a_1+1}\left\{1 - \frac{b_{23}^2}{(1+a_2)(1+a_3)}\right\}, \tag{A.4}$$

where

$$\begin{aligned} b_{123} &= 1 + 2\frac{b_{12}b_{13}b_{23}}{(1+a_1)\,(1+a_2)\,(1+a_3)} \\ &\quad - \frac{b_{23}^2}{(1+a_2)\,(1+a_3)} - \frac{b_{13}^2}{(1+a_1)\,(1+a_3)} - \frac{b_{12}^2}{(1+a_1)\,(1+a_2)} \\ &= 1 + 2\cos(\phi_{1,2})\cos(\phi_{1,3})\cos(\phi_{2,3}) - \cos^2(\phi_{1,2}) - \cos^2(\phi_{1,3}) - \cos^2(\phi_{2,3}) \\ &\quad + O(1/d) \\ &= 4\sin(0.5(\phi_{1,2}+\phi_{1,3}-\phi_{2,3}))\sin(0.5(\phi_{1,2}+\phi_{2,3}-\phi_{1,3})) \\ &\quad \times \sin(0.5(\phi_{1,3}+\phi_{2,3}-\phi_{1,2}))\sin(0.5(\phi_{1,2}+\phi_{1,3}+\phi_{2,3})) + O(1/d) \\ &\geq 4\sin^4(\xi_0/2) + O(1/d) \end{aligned}$$

by Condition 3.3.1(iii). It implies that $\boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\lambda}_1 \geq 1 - \sin^{-4}(\xi_0/2)/(a_1+1) + O(1/d) \to 1$ as $d \to \infty$. ■

Proof of Theorem 3.3.6. Using the properties of the GMM estimator, we know that $\hat{\boldsymbol{\lambda}}_k \xrightarrow{\mathbb{P}} \boldsymbol{\lambda}_k$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}} \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_{\boldsymbol{X}}$. Under Condition 3.3.1(ii) $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ is invertible. Thus by the Continuous Mapping Theorem (CMT) it follows that $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$. As a result $\hat{\boldsymbol{w}}_{knd}^{\star\top} = (\hat{\boldsymbol{\lambda}}_k^\top \hat{\boldsymbol{\Sigma}}_X^{-1} \hat{\boldsymbol{\lambda}}_k)^{-1/2} \hat{\boldsymbol{\lambda}}_k^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{w}_{kd}^{\star\top} = (\boldsymbol{\lambda}_k^\top \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\lambda}_k)^{-1/2} \boldsymbol{\lambda}_k^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$. By expanding, we get

$$\begin{aligned}\lim_{d\to\infty}\lim_{n\to\infty} \mathbb{E}\{(\hat{Z}_{knd}^\star - Z_k)^2\} = \lim_{d\to\infty}\lim_{n\to\infty} \mathbb{E}\left\{(\boldsymbol{X}^\top \hat{\boldsymbol{w}}_{knd}^\star)^2\right\} \\ - 2\lim_{d\to\infty}\lim_{n\to\infty} \mathbb{E}\left(\hat{\boldsymbol{w}}_{knd}^{\star\top} \boldsymbol{X} Z_k\right) + 1. \end{aligned} \tag{A.5}$$

The first and second term in (A.5) can be rewritten as

$$\mathbb{E}\left\{(\boldsymbol{X}^\top \hat{\boldsymbol{w}}_{knd}^\star)^2\right\} = \mathbb{E}\{\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star + \hat{\boldsymbol{w}}_{kd}^\star)\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)\} + \mathbb{E}\left\{(\boldsymbol{X}^\top \hat{\boldsymbol{w}}_{kd}^\star)^2\right\}, \tag{A.6}$$

$$\mathbb{E}\left(\hat{\boldsymbol{w}}_{knd}^{\star\top} \boldsymbol{X} Z_k\right) = \mathbb{E}\left\{(\hat{\boldsymbol{w}}_{knd}^{\star\top} - \hat{\boldsymbol{w}}_{kd}^{\star\top})\boldsymbol{X} Z_k\right\} + \mathbb{E}\left(\boldsymbol{w}_{kd}^{\star\top} \boldsymbol{X} Z_k\right), \tag{A.7}$$

respectively. From Theorem 3.3.5, we know that

$$\lim_{d\to\infty} \mathbb{E}\left\{(\boldsymbol{X}^\top \boldsymbol{w}_{kd}^\star)^2\right\} - 2\lim_{d\to\infty} \mathbb{E}\left(\boldsymbol{w}_{kd}^{\star\top} \boldsymbol{X} Z_k\right) + 1 = 0.$$

Thus, it remains to be shown that the first term in (A.6) and in (A.7) converge to zero as $n \to \infty$. Using Cauchy's inequality, the first term in (A.6) is bounded by

$$\left(\mathbb{E}\left[\left\{\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star + \hat{\boldsymbol{w}}_{kd}^\star)\right\}^2\right] \cdot \mathbb{E}\left[\left\{\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)\right\}^2\right]\right)^{1/2}$$

Again using Cauchy's inequality, one gets

$$\begin{aligned}\mathbb{E}\left[\left\{\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star + \hat{\boldsymbol{w}}_{kd}^\star)\right\}^2\right] &\leq \left[\mathbb{E}(\boldsymbol{X^2}^\top \boldsymbol{X^2}) \cdot \mathbb{E}\{(\hat{\boldsymbol{w}}_{knd}^\star + \hat{\boldsymbol{w}}_{kd}^\star)^{2\top}(\hat{\boldsymbol{w}}_{knd}^\star + \hat{\boldsymbol{w}}_{kd}^\star)^2\}\right]^{1/2}, \\ \mathbb{E}\left[\left\{\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)\right\}^2\right] &\leq \left[\mathbb{E}(\boldsymbol{X^2}^\top \boldsymbol{X^2}) \cdot \mathbb{E}\{(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)^{2\top}(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)^2\}\right]^{1/2}.\end{aligned}$$

Note that $\mathbb{E}(\boldsymbol{X^2}^\top \boldsymbol{X^2}) = 3d$, and $\hat{\boldsymbol{w}}_{kd}^{\star\top} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}} \hat{\boldsymbol{w}}_{kd}^\star = 1$. From Condition 3.3.1(ii) we can conclude that $\hat{\boldsymbol{w}}_{kd}^\star$ and $\hat{\boldsymbol{w}}_{knd}^\star$ are bounded.

Since we know that $\hat{\boldsymbol{w}}_{knd}^{\star\top} \xrightarrow{\mathbb{P}} \boldsymbol{w}_{kd}^{\star\top}$ and $(\hat{\boldsymbol{w}}_{knd}^{\star\top} - \hat{\boldsymbol{w}}_{kd}^{\star\top})^4$ is uniformly integrable, we conclude that $\mathbb{E}\{(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)^{2\top}(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)^2\} \to 0$ as $n \to \infty$ by Vitali's convergence theorem. It implies that $\mathbb{E}\{\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star + \hat{\boldsymbol{w}}_{kd}^\star)\boldsymbol{X}^\top(\hat{\boldsymbol{w}}_{knd}^\star - \hat{\boldsymbol{w}}_{kd}^\star)\} \to 0$ as $n \to \infty$. Similarly, $\mathbb{E}\left\{(\hat{\boldsymbol{w}}_{knd}^{\star\top} - \hat{\boldsymbol{w}}_{kd}^{\star\top})\boldsymbol{X} Z_k\right\} \to 0$ as $n \to \infty$ which concludes the proof.

■

Proof of Lemma 3.3.7. This follows immediately from the fact that the GMM estimators converge in probability combined with the continuous mapping theorem, and from Theorem 3.3.6. ■

Proof of Theorem 3.3.8. From Lemma 3.3.7, we know that $\bar{U}_{nd} \xrightarrow{\mathbb{P}} \bar{U}_d$ as $n \to \infty$. Moreover, from Theorem 3.3.5, it follows that $\hat{Z}_{kd}^\star \xrightarrow{\mathscr{L}_2} Z_k$, for each $k = 1, \ldots, p$. Rewriting, we get

$$\hat{\varepsilon}_{jd} = \frac{1}{\gamma_j}\left(X_j - \sum_{k=1}^p \lambda_{jk} \hat{Z}_{kd}^\star\right) = \varepsilon_j - \frac{1}{\gamma_j}\sum_{k=1}^p \lambda_{jk}\left(\hat{Z}_{kd}^\star - Z_k\right),$$

and

$$\bar{U}_d = \bar{\varepsilon}_d - \frac{1}{d}\sum_{j=1}^{d}\sum_{k=1}^{p}\delta_{jk}\left(\hat{Z}^{\star}_{kd} - Z_k\right)$$

where $\bar{\varepsilon}_d = (1/d)\sum_{i=1}^{d}\varepsilon_{jd} \xrightarrow{\mathbb{P}} m(V)$ by Theorem 5 in Krupskii and Joe (2021) and $\mathbb{E}\{(\hat{Z}^{*}_{kd} - Z_k)^2\} < \xi^*/d$ for some constant $\xi^* > 0$ as follows from the proof of Theorem 3.3.5 so that

$$\begin{aligned}\mathbb{E}\left[\left\{\frac{1}{d}\sum_{j=1}^{d}\sum_{k=1}^{p}\delta_{jk}\left(\hat{Z}^{\star}_{kd} - Z_k\right)\right\}^2\right] &\leq \frac{p}{d}\sum_{j=1}^{d}\sum_{k=1}^{p}\delta_{jk}^2\mathbb{E}\{(\hat{Z}^{\star}_{kd} - Z_k)^2\} \\ &\leq \frac{\xi^* p}{d}\sum_{k=1}^{p}\left(\frac{1}{d}\sum_{j=1}^{d}\delta_{jk}^2\right) \to 0 \quad \text{as } d \to \infty.\end{aligned}$$

It implies that $\bar{U}_d \xrightarrow{\mathbb{P}} m(V)$, as $n \to \infty$ and $d \to \infty$.

■

Proof of Theorem 3.3.9. Let $\tilde{\gamma}_l = (1-\gamma_l^2)^{1/2}$ and $\tilde{\gamma}_m = (1-\gamma_m^2)^{1/2}$. Denote $W_l = \tilde{\gamma}_l^{-1}\sum_{k=1}^{p}\lambda_{lk}Z_k$ and $W_m = \tilde{\gamma}_m^{-1}\sum_{k=1}^{p}\lambda_{mk}Z_k$. Note that (W_m, W_l) have the bivariate standard normal distribution. Let $\phi_{l,m}$ be the joint density of (W_l, W_m) and $F_{l,m}$ be the joint cdf of (ϵ_l, ϵ_m). We find:

$$\begin{aligned}P(X_l < z, X_m < z) &= \int_{\mathbb{R}^2} F_{l,m}\left(\frac{z-\tilde{\gamma}_l w_l}{\gamma_l}, \frac{z-\tilde{\gamma}_m w_m}{\gamma_m}\right)\phi_{l,m}(w_l, w_m)dw_l dw_m \\ &\leq \int_{\mathbb{R}^2}\min\left\{\Phi\left(\frac{z-\tilde{\gamma}_l w_l}{\gamma_l}\right), \Phi\left(\frac{z-\tilde{\gamma}_m w_m}{\gamma_m}\right)\right\}\phi_{l,m}(w_l, w_m)dw_l dw_m \\ &= P(\tilde{\gamma}_l W_l + \gamma_l V < z, \tilde{\gamma}_m W_m + \gamma_m V < z) = \Phi_{\rho_{l,m}}(z, z),\end{aligned}$$

where $\Phi_{\rho_{l,m}}$ is the cdf of a standard bivariate normal distribution with the correlation $\rho_{l,m} = \tilde{\gamma}_l\tilde{\gamma}_m\mathrm{Cor}(W_l, W_m) + \gamma_l\gamma_m < 1$, so that

$$\lim_{z\to-\infty}\frac{P(X_l < z, X_m < z)}{\Phi(z)} \leq \lim_{z\to-\infty}\frac{\Phi_{\rho_{l,m}}(z,z)}{\Phi(z)} = 0.$$

■

A.3 Gradients of the GMM objective function

The gradient of the expression

$$Q_n(\boldsymbol{\Lambda}, \eta) = \sum_{i=1}^{d}\sum_{j=1}^{d}\left\{\left(\boldsymbol{\Sigma}_{\boldsymbol{X}} - \hat{\boldsymbol{\Sigma}}_n\right)^2\right\}_{i,j=1}^{d},$$

is the vector $\nabla Q_n(\boldsymbol{\Lambda}, \eta)$ of length $pd+1$, where the first pd elements are

$$4\left(\boldsymbol{\sigma}_j - \hat{\boldsymbol{\sigma}}_j\right)\cdot\frac{\partial\boldsymbol{\sigma}_j}{\partial\lambda_{jk}},$$

for $j = 1, \ldots, d$, $k = 1, \ldots, p$, and where $\boldsymbol{\sigma}_j$ (resp. $\hat{\boldsymbol{\sigma}}_j$) is the j-th column of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ (resp. $\hat{\boldsymbol{\Sigma}}_n$). Let $\odot$ denote element-wise multiplication for matrices. The last element in the vector $\nabla Q_n(\boldsymbol{\Lambda}, \eta)$ is

$$2 \sum_{i=1}^{d} \sum_{j=1}^{d} \left\{ \left(\boldsymbol{\Sigma}_{\boldsymbol{X}} - \hat{\boldsymbol{\Sigma}}_n \right) \odot \partial_\eta \boldsymbol{\Sigma}_{\boldsymbol{X}} \right\}_{i,j=1}^{d},$$

where $\partial_\eta \boldsymbol{\Sigma}_{\boldsymbol{X}}$ is a matrix with off-diagonal elements $\gamma_i \gamma_j$ and zeroes on the diagonal, for $i, j = 1, \ldots, d$. To fix notation, let $\boldsymbol{a}_{[j]k} = \left(a_{1k}, \ldots, a_{(j-1)k}, 0, a_{(j+1)k}, \ldots, a_{dk}\right)^\top$ be the vector $\boldsymbol{a}_k$ with the jth element equal to zero.

One-factor model. Let $k = 1$, then

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^\top + \eta \boldsymbol{\gamma} \boldsymbol{\gamma}^\top = \begin{cases} \left(\lambda_{i1} \lambda_{j1} + \eta \gamma_i \gamma_j\right)_{i,j=1}^{d} & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

and

$$\frac{\partial \boldsymbol{\sigma}_j}{\partial \lambda_{j1}} = \boldsymbol{\lambda}_{[j]1} - \eta \frac{\lambda_j}{\gamma_j} \boldsymbol{\gamma}_{[j]},$$

for $i, j = 1, \ldots, d$.

Two-factor model. Let $k = 2$, then

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^\top + \boldsymbol{\lambda}_2^\star \boldsymbol{\lambda}_2^{\star\top} + \eta \boldsymbol{\gamma} \boldsymbol{\gamma}^\top = \begin{cases} \left(\lambda_{i1} \lambda_{j1} + \lambda_{i2}^\star \lambda_{j2}^\star + \eta \gamma_i \gamma_j\right)_{i,j=1}^{d} & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

where $\lambda_{j2}^\star = \lambda_{j2} \sqrt{1 - \lambda_{j1}^2}$, and $\gamma_j = \sqrt{1 - \lambda_{j1}^2 - \lambda_{j2}^{\star 2}}$. Then

$$\frac{\partial \boldsymbol{\sigma}_j}{\partial \lambda_{j1}} = \boldsymbol{\lambda}_{[j]1} - \lambda_{j1} \left\{ \frac{\lambda_{j2}^\star}{1 - \lambda_{j1}^2} \boldsymbol{\lambda}_{[j]2}^\star + \eta \frac{\left(1 - \lambda_{j2}^2\right)}{\gamma_j} \boldsymbol{\gamma}_{[j]} \right\}.$$

Similarly,

$$\frac{\partial \boldsymbol{\sigma}_j}{\partial \lambda_{j2}} = \sqrt{1 - \lambda_{j1}^2} \left(\boldsymbol{\lambda}_{[j]2}^\star - \eta \frac{\lambda_{j2}^\star}{\gamma_j} \boldsymbol{\gamma}_{[j]} \right).$$

Three-factor model. Let $k = 3$, then

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{X}} &= \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^\top + \boldsymbol{\lambda}_2^\star \boldsymbol{\lambda}_2^{\star\top} + \boldsymbol{\lambda}_3^\star \boldsymbol{\lambda}_3^{\star\top} + \eta \boldsymbol{\gamma} \boldsymbol{\gamma}^\top \\ &= \begin{cases} \left(\lambda_{i1} \lambda_{j1} + \lambda_{i2}^\star \lambda_{j2}^\star + \lambda_{i3}^\star \lambda_{j3}^\star + \eta \gamma_i \gamma_j\right)_{i,j=1}^{d} & \text{if } i \neq j \\ 1 & \text{if } i = j, \end{cases} \end{aligned}$$

where $\lambda_{j2}^\star = \lambda_{j2} \sqrt{1 - \lambda_{j1}^2}$, $\lambda_{j3}^\star = \lambda_{j3} \sqrt{1 - \lambda_{j1}^2 - \lambda_{j2}^{\star 2}}$, and $\gamma_j = \sqrt{1 - \lambda_{j1}^2 - \lambda_{j2}^{\star 2} - \lambda_{j3}^{\star 2}}$. Then

$$\frac{\partial \boldsymbol{\sigma}_j}{\partial \lambda_{j1}} = \boldsymbol{\lambda}_{[1]1} - \lambda_{j1} \left\{ \frac{\lambda_{j2}^\star}{1 - \lambda_{j1}^2} \boldsymbol{\lambda}_{[j]2}^\star + \frac{\lambda_{j3}^\star \left(1 - \lambda_{j2}^2\right)}{1 - \lambda_{j1}^2 - \lambda_{j2}^{\star 2}} \boldsymbol{\lambda}_{[j]3}^\star + \eta \left(1 - \lambda_{j3}^2\right) \left(1 - \lambda_{j2}^2\right) \frac{\boldsymbol{\gamma}_{[j]}}{\gamma_j} \right\}.$$

Similarly, we find

$$\frac{\partial \boldsymbol{\sigma}_j}{\partial \lambda_{j2}} = \sqrt{1-\lambda_{j1}^2}\left\{\boldsymbol{\lambda}^\star_{[j]2} - \frac{\lambda^\star_{j3}\lambda^\star_{j2}}{1-\lambda_{j1}^2-\lambda_{j2}^{\star 2}}\boldsymbol{\lambda}^\star_{[j]3} - \eta\lambda^\star_{j2}\frac{1-\lambda_{j3}^2}{\gamma_j}\boldsymbol{\gamma}_{[j]}\right\}.$$

Lastly,

$$\frac{\partial \boldsymbol{\sigma}_j}{\partial \lambda_{j3}} = \sqrt{1-\lambda_{j1}^2-\lambda_{j2}^{\star 2}}\left(\boldsymbol{\lambda}^\star_{[j]3} - \eta\frac{\lambda^\star_{j3}}{\gamma_j}\boldsymbol{\gamma}_{[j]}\right).$$

A.4 Matrix inversion

Let

$$\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2} = \begin{pmatrix} \gamma_1^2 & \eta\gamma_1\gamma_2 & \cdots & \eta\gamma_1\gamma_d \\ \eta\gamma_1\gamma_2 & \gamma_2^2 & \cdots & \eta\gamma_2\gamma_d \\ \vdots & \vdots & \ddots & \vdots \\ \eta\gamma_1\gamma_d & \eta\gamma_2\gamma_d & \cdots & \gamma_d^2 \end{pmatrix}$$

The proof follows by induction. For $d = 2$,

$$\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} = \begin{pmatrix} \frac{1}{\gamma_1^2(1-\eta^2)} & \frac{-\eta}{\gamma_1\gamma_2(1-\eta^2)} \\ \frac{-\eta}{\gamma_1\gamma_2(1-\eta^2)} & \frac{1}{\gamma_2^2(1-\eta^2)} \end{pmatrix}.$$

For $d = 3$,

$$\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} = \begin{pmatrix} \frac{1+\eta}{\gamma_1^2(1+\eta-2\eta^2)} & \frac{-\eta}{\gamma_1\gamma_2(1+\eta-2\eta^2)} & \frac{-\eta}{\gamma_1\gamma_3(1+\eta-2\eta^2)} \\ \frac{-\eta}{\gamma_1\gamma_2(1+\eta-2\eta^2)} & \frac{1+\eta}{\gamma_2^2(1+\eta-2\eta^2)} & \frac{-\eta}{\gamma_2\gamma_3(1+\eta-2\eta^2)} \\ \frac{-\eta}{\gamma_1\gamma_3(1+\eta-2\eta^2)} & \frac{-\eta}{\gamma_2\gamma_3(1+\eta-2\eta^2)} & \frac{1+\eta}{\gamma_3^2(1+\eta-2\eta^2)} \end{pmatrix}.$$

For $d = 4$,

$$\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} = \begin{pmatrix} \frac{1+2\eta}{\gamma_1^2(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_1\gamma_2(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_1\gamma_3(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_1\gamma_4(1+2\eta-3\eta^2)} \\ \frac{-\eta}{\gamma_1\gamma_2(1+2\eta-3\eta^2)} & \frac{1+2\eta}{\gamma_2^2(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_2\gamma_3(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_2\gamma_4(1+2\eta-3\eta^2)} \\ \frac{-\eta}{\gamma_1\gamma_3(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_2\gamma_3(1+2\eta-3\eta^2)} & \frac{1+2\eta}{\gamma_3^2(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_3\gamma_4(1+2\eta-3\eta^2)} \\ \frac{-\eta}{\gamma_1\gamma_4(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_2\gamma_4(1+2\eta-3\eta^2)} & \frac{-\eta}{\gamma_3\gamma_4(1+2\eta-3\eta^2)} & \frac{1+2\eta}{\gamma_3^2(1+2\eta-3\eta^2)} \end{pmatrix}.$$

Thus for arbitrary d we can show that,

$$\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} = \begin{pmatrix} \frac{1+(d-2)\eta}{\gamma_1^2(1+(d-2)\eta-(d-1)\eta^2)} & \cdots & \frac{-\eta}{\gamma_1\gamma_d(1+(d-2)\eta-(d-1)\eta^2)} \\ \vdots & \ddots & \vdots \\ \frac{-\eta}{\gamma_1\gamma_d(1+(d-2)\eta-(d-1)\eta^2)} & \cdots & \frac{1+(d-2)\eta}{\gamma_d^2(1+(d-2)\eta-(d-1)\eta^2)} \end{pmatrix}.$$

It is easy to verify that $\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\left(\boldsymbol{\Gamma}^{1/2}\mathbf{H}\boldsymbol{\Gamma}^{1/2}\right)^{-1} = I_d$, where I_d is the $d \times d$ identity matrix.

Appendix B

Appendix for Chapter 4

B.1 Proof of Theorem 4.2.6

The proof of Theorem 4.2.6 is based on four lemmas which we prove first. Throughout the remainder of the proof, we let

$$\tilde{D}_m^{\mathcal{P}}(k) = \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}, \qquad k \geq m+1, \tag{B.1}$$

which is the version of the detector in (4.7), in which the estimated long-run variance $\Sigma_m^{\mathcal{P}}$ is replaced by the true long-run variance $\Sigma^{\mathcal{P}}$.

On the probability space of Condition 4.2.5 (assuming that this condition holds), we may define

$$\begin{aligned}\bar{D}_m^{\mathcal{P}}(k) = \frac{1}{\sqrt{m}} \max_{j \in [\![m,k[\![} \Big\| & \frac{k}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(j-m)\} \\ & - \frac{j}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(k-m)\} \Big\|_{(\Sigma^{\mathcal{P}})^{-1}},\end{aligned} \tag{B.2}$$

where we recall that for each $m \in \mathbb{N}$, $\boldsymbol{W}_{1,m}$ and $\boldsymbol{W}_{2,m}$ are independent p-dimensional standard Brownian motions.

Lemma B.1.1. *Assume that Condition 4.2.5 holds. Then, for any $\eta > 0$,*

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \left|\tilde{D}_m^{\mathcal{P}}(k) - \bar{D}_m^{\mathcal{P}}(k)\right| = o_{\mathbb{P}}(1).$$

Proof. Fix $\eta > 0$. Applying the reverse triangle inequality for suprema

$$|\sup_{x \in A} f(x) - \sup_{x \in A} g(x)| \leq \sup_{x \in A} |f(x) - g(x)| \tag{B.3}$$

to the maximum over $j \in [\![m, k[\![$ below gives

$$\begin{aligned}&\left|\tilde{D}_m^{\mathcal{P}}(k) - \bar{D}_m^{\mathcal{P}}(k)\right| \\ &\leq \frac{1}{\sqrt{m}} \max_{j \in [\![m,k[\![} \left| \frac{j(k-j)}{m} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} - \left\| \frac{k}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(j-m)\} \right.\right.\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{j}{m}(\Sigma^{\mathcal{P}})^{\frac{1}{2}}\{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(k-m)\}\Big\|_{(\Sigma^{\mathcal{P}})^{-1}}\Big| \\
&\leq \frac{1}{\sqrt{m}} \max_{j\in[\![m,k[\![} \Big\| \frac{j(k-j)}{m}\{\bar{\boldsymbol{Y}}^{\mathcal{P}}_{1:j} - \bar{\boldsymbol{Y}}^{\mathcal{P}}_{j+1:k}\} - \frac{k}{m}(\Sigma^{\mathcal{P}})^{\frac{1}{2}}\{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(j-m)\} \\
&\qquad + \frac{j}{m}(\Sigma^{\mathcal{P}})^{\frac{1}{2}}\{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(k-m)\}\Big\|_{(\Sigma^{\mathcal{P}})^{-1}} = U^{\mathcal{P}}_m(k).
\end{aligned}
$$

Next, we rewrite $j(k-j)\{\bar{\boldsymbol{Y}}^{\mathcal{P}}_{1:j} - \bar{\boldsymbol{Y}}^{\mathcal{P}}_{j+1:k}\}$ as

$$
\begin{aligned}
&j(k-j)\left\{\frac{1}{j}\sum_{i=1}^{j}\boldsymbol{Y}^{\mathcal{P}}_i - \frac{1}{k-j}\sum_{i=j+1}^{k}\boldsymbol{Y}^{\mathcal{P}}_i\right\} \\
&= (k-j)\sum_{i=1}^{j}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - j\sum_{i=j+1}^{k}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} \\
&= k\sum_{i=1}^{j}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - j\sum_{i=1}^{k}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} \\
&= k\sum_{i=1}^{m}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} + k\sum_{i=m+1}^{j}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} \\
&\quad - j\sum_{i=1}^{m}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - j\sum_{i=m+1}^{k}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\}.
\end{aligned}
$$

Hence, using the triangle inequality, we have that

$$
\begin{aligned}
\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\left|\tilde{D}^{\mathcal{P}}_m(k) - \bar{D}^{\mathcal{P}}_m(k)\right| &\leq \sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}U^{\mathcal{P}}_m(k) \\
&\leq I^{\mathcal{P}}_m + (I^{\mathcal{P}}_m)' + J^{\mathcal{P}}_m + (J^{\mathcal{P}}_m)',
\end{aligned}
$$

where (using the convention that empty sums are equal to 0)

$$
\begin{aligned}
I^{\mathcal{P}}_m &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{1}{2}+\eta}\max_{j\in[\![m,k[\![}\left\|\sum_{i=1}^{m}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{2,m}(m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}, \\
(I^{\mathcal{P}}_m)' &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{1}{2}+\eta} \\
&\quad \times \max_{j\in[\![m,k[\![}\left\|\sum_{i=m+1}^{j}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{1,m}(j-m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}, \\
J^{\mathcal{P}}_m &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j}{m}\left\|\sum_{i=1}^{m}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{2,m}(m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}, \\
(J^{\mathcal{P}}_m)' &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \\
&\quad \times \max_{j\in[\![m,k[\![}\frac{j}{m}\left\|\sum_{i=m+1}^{k}\{\boldsymbol{Y}^{\mathcal{P}}_i - \mathbb{E}(\boldsymbol{Y}^{\mathcal{P}}_1)\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{1,m}(k-m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}.
\end{aligned}
$$

It suffices to show that $I_m^{\mathcal{P}}$, $(I_m^{\mathcal{P}})'$, $J_m^{\mathcal{P}}$, and $(J_m^{\mathcal{P}})'$ are $o_{\mathbb{P}}(1)$. Since $(m/k) \times (j/m) = j/k \leq 1$ when $m \leq j < k$, we have that $J_m^{\mathcal{P}} \leq I_m^{\mathcal{P}}$ and $(J_m^{\mathcal{P}})' \leq (I_m^{\mathcal{P}})'$. So it suffices to consider $I_m^{\mathcal{P}}$ and $(I_m^{\mathcal{P}})'$. Let $0 < \xi < \frac{1}{2}$ be as in Condition 4.2.5. Then,

$$\begin{aligned} I_m^{\mathcal{P}} &= \frac{1}{m^{\xi}} \left\| \sum_{i=1}^{m} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{2,m}(m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}} m^{\xi - \frac{1}{2}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}+\eta} \\ &\leq \frac{1}{m^{\xi}} \left\| \sum_{i=1}^{m} \left\{\boldsymbol{Y}_{i,m}^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\right\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{2,m}(m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}} m^{\xi - \frac{1}{2}}. \end{aligned}$$

Hence, by equivalence of norms on $\mathbb{R}^p$, (4.13) and the fact that $\xi < \frac{1}{2}$, $I_m^{\mathcal{P}}$ converges in probability to zero as $m \to \infty$. Next, since the norm in $(I_m^{\mathcal{P}})'$ is zero when $j = m$ and using the fact that $j - m < k - m < k$, we see that $(I_m^{\mathcal{P}})'$ is equal to

$$\begin{aligned} &\frac{1}{\sqrt{m}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}+\eta} \max_{j \in]\!]m,k[\!]} \left\| \sum_{i=m+1}^{j} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(j-m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}} \\ &\leq \frac{1}{m^{\frac{1}{2}-\xi}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}-\xi+\eta} \\ &\quad \times \max_{j \in]\!]m,k[\!]} \frac{1}{(j-m)^{\xi}} \left\| \sum_{i=m+1}^{j} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(j-m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}} \\ &\leq \frac{1}{m^{\frac{1}{2}-\xi}} \sup_{j>m} \frac{1}{(j-m)^{\xi}} \left\| \sum_{i=m+1}^{j} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(j-m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}} \\ &\quad \times \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}-\xi+\eta}. \end{aligned}$$

The supremum over k above is less than 1 since $\xi < \frac{1}{2}$. The supremum over j is bounded in probability by (4.12) and equivalence of norms. Since $m^{-\frac{1}{2}+\xi} \to 0$, this proves that $(I_m^{\mathcal{P}})'$ converges to 0 in probability. ■

Given p-dimensional independent standard Brownian motions $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ and $\eta > 0$, define the random function D_η by

$$D_\eta(s,t) = t^{-\frac{3}{2}-\eta} \left\| (t-s)\boldsymbol{W}_2(1) + t\boldsymbol{W}_1(s-1) - s\boldsymbol{W}_1(t-1) \right\|_{I_p}, \quad 1 \leq s \leq t < \infty. \tag{B.4}$$

Lemma B.1.2. *For any $\eta > 0$, the function D_η is almost surely bounded and uniformly continuous.*

Proof. Fix $\eta > 0$. For $i \in [\![1,p]\!]$ and $1 \leq s \leq t < \infty$, define $A^{[i]}(s,t)$ to be the i-th coordinate of

$$\boldsymbol{A}(s,t) = t^{-\frac{3}{2}-\eta} \left\{ (t-s)\boldsymbol{W}_2(1) + t\boldsymbol{W}_1(s-1) - s\boldsymbol{W}_1(t-1) \right\}, \quad 1 \leq s \leq t < \infty,$$

and note that $D_\eta(s,t) = \|\boldsymbol{A}(s,t)\|_{I_p}$. From Lemma A.2 in Holmes and Kojadinovic (2021), we know that $|A^{[i]}|$ is almost surely bounded and uniformly continuous for

each i. It follows that each $A^{[i]}$ is also bounded and uniformly continuous, almost surely. Now note that

$$\begin{aligned}|D_\eta(s,t) - D_\eta(s',t')| &= \Big| \|\boldsymbol{A}(s,t)\|_{I_p} - \|\boldsymbol{A}(s',t')\|_{I_p} \Big| \\ &\leq \|\boldsymbol{A}(s,t) - \boldsymbol{A}(s',t')\|_{I_p} \\ &= \sqrt{\frac{1}{p}\sum_{i=1}^{p} \{A^{[i]}(s,t) - A^{[i]}(s',t')\}^2}.\end{aligned}$$

So D_η is almost surely bounded and uniformly continuous since each $A^{[i]}$ is. ■

Lemma B.1.3. *For any $\eta > 0$,*

$$\sup_{k>m} \Big(\frac{m}{k}\Big)^{\frac{3}{2}+\eta} \bar{D}_m^{\mathcal{P}}(k) \rightsquigarrow \sup_{1\leq s\leq t<\infty} D_\eta(s,t),$$

where $\bar{D}_m^{\mathcal{P}}$ is defined in (B.2) *and D_η is defined in* (B.4)*.*

Proof. Fix $\eta > 0$. The random variable $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}\bar{D}_m^{\mathcal{P}}(k)$ is equal in distribution to

$$\frac{1}{\sqrt{m}} \sup_{k\geq m} \Big(\frac{m}{k}\Big)^{\frac{3}{2}+\eta} \max_{j\in[\![m,k]\!]} \Big\| \frac{k}{m}\{\boldsymbol{W}_2(m) + \boldsymbol{W}_1(j-m)\} - \frac{j}{m}\{\boldsymbol{W}_2(m) + \boldsymbol{W}_1(k-m)\} \Big\|_{I_p},$$

where we note that the norm is equal to zero when $j = k$, which has allowed us to include the cases $j = k$ and $k = m$ in the maximum and supremum, respectively. Using Brownian scaling, this is equal in distribution to

$$\sup_{k\geq m} \Big(\frac{m}{k}\Big)^{\frac{3}{2}+\eta} \max_{j\in[\![m,k]\!]} \Big\| \frac{k}{m}\Big\{\boldsymbol{W}_2(1) + \boldsymbol{W}_1\Big(\frac{j}{m} - 1\Big)\Big\} - \frac{j}{m}\Big\{\boldsymbol{W}_2(1) + \boldsymbol{W}_1\Big(\frac{k}{m} - 1\Big)\Big\} \Big\|_{I_p}.$$

In the above, j and k are integers. Letting $k = \lfloor mt \rfloor$ and $j = \lfloor ms \rfloor$ (where $s, t \in \mathbb{R}$ and $1 \leq s \leq t < \infty$), the above display becomes

$$\begin{aligned}&\sup_{t\geq 1} \Big(\frac{m}{\lfloor mt \rfloor}\Big)^{\frac{3}{2}+\eta} \\ &\times \sup_{s\in[1,t]} \Big\| \frac{\lfloor mt \rfloor}{m}\Big\{\boldsymbol{W}_2(1) + \boldsymbol{W}_1\Big(\frac{\lfloor ms \rfloor}{m} - 1\Big)\Big\} - \frac{\lfloor ms \rfloor}{m}\Big\{\boldsymbol{W}_2(1) + \boldsymbol{W}_1\Big(\frac{\lfloor mt \rfloor}{m} - 1\Big)\Big\} \Big\|_{I_p} \\ &= \sup_{1\leq s\leq t<\infty} D_\eta\left(\frac{\lfloor ms \rfloor}{m}, \frac{\lfloor mt \rfloor}{m}\right).\end{aligned}$$

Using (B.3) with the supremum over (s,t), we have

$$\begin{aligned}&\left| \sup_{1\leq s\leq t<\infty} D_\eta\left(\frac{\lfloor ms \rfloor}{m}, \frac{\lfloor mt \rfloor}{m}\right) - \sup_{1\leq s\leq t<\infty} D_\eta(s,t) \right| \\ &\leq \sup_{1\leq s\leq t<\infty} \left| D_\eta\left(\frac{\lfloor ms \rfloor}{m}, \frac{\lfloor mt \rfloor}{m}\right) - D_\eta(s,t) \right|,\end{aligned}$$

which converges to zero almost surely as $m \to \infty$ since D_η is almost surely uniformly continuous and $t - 1/m < \frac{\lfloor mt \rfloor}{m} \leq t$. ■

For a $p \times p$ matrix A, denote the operator norm of A by

$$\|A\|_{\mathrm{op}} = \inf\{c \geq 0 : \|A\boldsymbol{v}\|_2 \leq c\|\boldsymbol{v}\|_2, \text{ for all } \boldsymbol{v} \in \mathbb{R}^p\}. \tag{B.5}$$

Lemma B.1.4. *Assume that Condition 4.2.5 holds and that $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$. Then, for any $\eta > 0$,*

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |\tilde{D}_m^{\mathcal{P}}(k) - D_m^{\mathcal{P}}(k)| = o_{\mathbb{P}}(1),$$

where $\tilde{D}_m^{\mathcal{P}}$ is defined in (B.1) *and $D_m^{\mathcal{P}}$ is defined in* (4.5).

Proof. Fix $\eta > 0$. Applying (B.3) to the maximum over $j \in [\![m, k[\![$ and using the inequality $|a^{\frac{1}{2}} - b^{\frac{1}{2}}| \leq |a-b|^{\frac{1}{2}}$ for $a, b \geq 0$, we obtain that

$$\begin{aligned}
&\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |\tilde{D}_m^{\mathcal{P}}(k) - D_m^{\mathcal{P}}(k)| \\
&\quad \leq \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \left| \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} - \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma_m^{\mathcal{P}})^{-1}} \right| \\
&\quad \leq \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \qquad (B.6) \\
&\qquad \times \max_{j \in [\![m,k[\![} \frac{j(k-j)}{p^{\frac{1}{2}} m^{\frac{3}{2}}} \left| \left(\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right)^{\top} \left((\Sigma^{\mathcal{P}})^{-1} - (\Sigma_m^{\mathcal{P}})^{-1}\right) \left(\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right) \right|^{\frac{1}{2}}. \qquad (B.7)
\end{aligned}$$

For $A \in \mathbb{R}^{p \times p}$ and $\boldsymbol{v} \in \mathbb{R}^p$, we have that $|\boldsymbol{v}^{\top} A \boldsymbol{v}| \leq \|A\boldsymbol{v}\|_2 \|\boldsymbol{v}\|_2$ by the Cauchy-Schwarz inequality. By (B.5), $\|A\boldsymbol{v}\|_2 \leq \|A\|_{\mathrm{op}} \|\boldsymbol{v}\|_2$, so that $|\boldsymbol{v}^{\top} A \boldsymbol{v}|^{\frac{1}{2}} \leq \|A\|_{\mathrm{op}}^{\frac{1}{2}} \|\boldsymbol{v}\|_2$. From (B.7), we therefore obtain

$$\begin{aligned}
&\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |\tilde{D}_m^{\mathcal{P}}(k) - D_m^{\mathcal{P}}(k)| \\
&\qquad \leq \frac{1}{p^{\frac{1}{2}}} \left\|(\Sigma^{\mathcal{P}})^{-1} - (\Sigma_m^{\mathcal{P}})^{-1}\right\|_{\mathrm{op}}^{\frac{1}{2}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \left\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right\|_2.
\end{aligned}$$

It is well-known that the mapping that maps an invertible square matrix to its inverse is continuous. Since $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ and Condition 4.2.3 holds, the continuous mapping theorem immediately implies that $(\Sigma_m^{\mathcal{P}})^{-1} \xrightarrow{\mathbb{P}} (\Sigma^{\mathcal{P}})^{-1}$, which in turn implies that $\|(\Sigma^{\mathcal{P}})^{-1} - (\Sigma_m^{\mathcal{P}})^{-1}\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$ by equivalence of norms. Furthermore, from Lemmas B.1.1 and B.1.3, we have that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} \tilde{D}_m^{\mathcal{P}}(k)$ converges weakly, which implies that it is bounded in probability. By equivalence of norms on $\mathbb{R}^p$, we immediately obtain that

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \left\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right\|_2 = O_{\mathbb{P}}(1)$$

and therefore the desired result. ∎

Proof of Theorem 4.2.6. From Lemmas B.1.1–B.1.4, we have that

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \rightsquigarrow \sup_{1 \leq s \leq t < \infty} D_\eta(s,t),$$

where D_η is defined in (B.4). It remains to be shown that

$$\sup_{1\le s\le t<\infty} D_\eta(s,t) \stackrel{d}{=} \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{W}(s)-s\boldsymbol{W}(t)\|_{I_p}. \tag{B.8}$$

Let $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ be two p-dimensional Gaussian processes defined, for any $1\le s\le t$, by

$$\begin{aligned}\boldsymbol{U}_p(s,t) &= (t-s)\,\boldsymbol{W}_2(1)+t\boldsymbol{W}_1(s-1)-s\boldsymbol{W}_1(t-1),\\ \boldsymbol{V}_p(s,t) &= t\boldsymbol{W}(s)-s\boldsymbol{W}(t).\end{aligned}$$

Since $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, and $\boldsymbol{W}$ are p-dimensional standard Brownian motions, it follows that the coordinates of the Gaussian processes $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ are centered. Thus, to show that the random functions $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ are equal in distribution (which will immediately imply (B.8)), it suffices to establish equality of their covariance functions at any (s,t,s',t') with $1\le s\le t$ and $1\le s'\le t'$. On one hand, the covariance function of the Gaussian process $\boldsymbol{U}_p$ at (s,t,s',t') is

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{U}_p(s,t)\boldsymbol{U}_p(s',t')^\top\}\\ &= \mathbb{E}\big[\big((t-s)\,\boldsymbol{W}_2(1)+t\boldsymbol{W}_1(s-1)-s\boldsymbol{W}_1(t-1)\big)\\ &\quad\times\big((t'-s')\boldsymbol{W}_2(1)+t'\boldsymbol{W}_1(s'-1)-s'\boldsymbol{W}_1(t'-1)\big)^\top\big]\\ &= \big[(t-s)(t'-s')+t't\{\min(s,s')-1\}-s't\{\min(s,t')-1\}\\ &\quad-t's\{\min(t,s')-1\}+s's\{\min(t,t')-1\}\big]I_p\\ &= \{t't\min(s,s')-s't\min(s,t')-t's\min(t,s')+s's\min(t,t')\}\,I_p,\end{aligned}$$

while, on the other hand, the covariance function of the Gaussian process $\boldsymbol{V}_p$ at (s,t,s',t') is

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{V}(s,t)\boldsymbol{V}(s',t')^\top\}\\ &= \mathbb{E}\big[\big(t\boldsymbol{W}(s)-s\boldsymbol{W}(t)\big)\big(t'\boldsymbol{W}(s')-s'\boldsymbol{W}(t')\big)^\top\big]\\ &= \{tt'\min(s,s')-s't\min(s,t')-st'\min(t,s')+ss'\min(t,t')\}I_p,\end{aligned}$$

which concludes the proof. ■

B.2 Proofs of Propositions 4.2.8, 4.2.10 and 4.2.12

Proof of Proposition 4.2.8. Let $\mathcal{P}\in(\mathbb{R}^d)^p$ be such that Condition 4.2.3 holds. Since $\boldsymbol{Y}_i^{\mathcal{P}}=\big(\mathbf{1}(\boldsymbol{X}_i\le\boldsymbol{x}_1),\ldots,\mathbf{1}(\boldsymbol{X}_i\le\boldsymbol{x}_p)\big)$, it is immediate that $\alpha_r^{\boldsymbol{Y}}\le\alpha_r^{\boldsymbol{X}}=O(r^{-a})$ as $r\to\infty$. Furthermore, as all the components of the $\boldsymbol{Y}_i^{\mathcal{P}}$ are bounded in absolute value by 1, from Theorem 4 of Kuelbs and Philipp (1980), we can redefine the sequence $(\boldsymbol{Y}_i^{\mathcal{P}})_{i\in\mathbb{N}}$ on a new probability space together with a p-dimensional standard Brownian motion $\boldsymbol{W}$ such that, almost surely,

$$\left\|\sum_{i=1}^m\{\boldsymbol{Y}_i^{\mathcal{P}}-\mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\}-(\Sigma^{\mathcal{P}})^{\frac12}\boldsymbol{W}(m)\right\|_2=O(m^{\frac12-\lambda}),$$

for some $\lambda\in(0,\frac12)$ that depends on a, p and $\mathcal{P}$. Let $\xi\in(\frac12-\lambda,\frac12)$. Then, almost surely,

$$\lim_{m\to\infty}\frac{1}{m^\xi}\left\|\sum_{i=1}^m\{\boldsymbol{Y}_i^{\mathcal{P}}-\mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\}-(\Sigma^{\mathcal{P}})^{\frac12}\boldsymbol{W}(m)\right\|_2=0. \tag{B.9}$$

Let $\boldsymbol{W}'$ be another p-dimensional standard Brownian motion independent of $\boldsymbol{W}$ and define $\boldsymbol{W}_2^{(m)}$ as

$$\boldsymbol{W}_2^{(m)}(s) = \begin{cases} \boldsymbol{W}(s) \text{ if } s \in [0, m], \\ \boldsymbol{W}'(s-m) + \boldsymbol{W}(m) \text{ otherwise,} \end{cases}$$

and $\boldsymbol{W}_1^{(m)}$ as $\boldsymbol{W}_1^{(m)}(s) = \boldsymbol{W}(m+s) - \boldsymbol{W}(m)$, $s \geq 0$. Then, for each $m \in \mathbb{N}$, $\boldsymbol{W}_1^{(m)}$ and $\boldsymbol{W}_2^{(m)}$ are independent p-dimensional standard Brownian motions.

For each $m \geq 0$, let

$$V_m = \sup_{k>m} \frac{1}{(k-m)^\xi} \left\| \sum_{i=m+1}^{k} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_1^{(m)}(k-m) \right\|_2$$

and note that the sequence $(V_m)_{m\geq 0}$ consists of identically distributed random variables. Therefore, to show that (4.12) holds, it is sufficient to show that $V_0 = O_{\mathbb{P}}(1)$. From the previous definition, V_0 can be rewritten as

$$V_0 = \sup_{k>0} \frac{1}{k^\xi} \left\| \sum_{i=1}^{k} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}(k) \right\|_2,$$

and, from (B.9), it is the supremum of an almost surely convergent sequence. It follows that V_0 is an almost surely finite random variable, so $V_0 = O_{\mathbb{P}}(1)$ and therefore (4.12) holds. Finally, (B.9) and the definition of $\boldsymbol{W}_2^{(m)}$ immediately imply (4.13). ■

Proof of Proposition 4.2.10. It is immediate that

$$\sqrt{p} \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \geq \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}} |tW^{[1]}(s) - sW^{[1]}(t)|,$$

where $W^{[1]}$ is the first component of the p-dimensional Brownian motion $\boldsymbol{W}$. Hence, for any fixed $M > 0$,

$$\mathbb{P}\Big(\sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \geq M \Big) \geq \mathbb{P}\Big(p^{-1/2} \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}} |tW^{[1]}(s) - sW^{[1]}(t)| \geq M \Big) = 1,$$

where the last equality is a consequence of Proposition 3.4 of Holmes and Kojadinovic (2021). ■

Proof of Proposition 4.2.12. Fix $\eta > 0$ and $p \in \mathbb{N}$. Note first that $\boldsymbol{W}$ is a p-variate continuous Gaussian process, which implies that $\boldsymbol{W} \in C([0,\infty), \mathbb{R}^p)$ almost surely. For $v \geq 2$, let $f_v : C([0, v], \mathbb{R}^p) \to [0, \infty)$ be defined by

$$f_v(\boldsymbol{w}) = \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p}, \qquad \boldsymbol{w} \in C([0, v], \mathbb{R}^p),$$

and, similarly, let $f : C([0,\infty), \mathbb{R}^p) \to [0, \infty)$ be defined by

$$f(\boldsymbol{w}) = \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p}, \qquad \boldsymbol{w} \in C([0,\infty), \mathbb{R}^p).$$

Notice that $\mathcal{L}_{p,\eta} = f(\boldsymbol{W})$ and that $f_v(\boldsymbol{w}) \leq f(\boldsymbol{w})$ for every $v \geq 2$.

Next, for any $v \geq 2$, we equip $C([0,v],\mathbb{R}^p)$ with the uniform distance, which is then a separable (and locally convex) metric space. Furthermore, as we shall verify below, f_v is continuous and convex for any $v \geq 2$. We can then apply Theorem 7.1 of Davydov and Lifshits (1984) to obtain that, for any $v \geq 2$, the distribution of $f_v(\boldsymbol{W})$ is concentrated on $[0,\infty)$ and absolutely continuous on $(0,\infty)$. In addition, some thought reveals that, for any $v \geq 2$,

$$\mathbb{P}(f_v(\boldsymbol{W}) = 0) \leq \mathbb{P}\big(f_2(\boldsymbol{W}) = 0\big) \leq \mathbb{P}(\|2\boldsymbol{W}(1) - 1\boldsymbol{W}(2)\|_{I_p} = 0\big) = 0$$

since $2\boldsymbol{W}(1) - 1\boldsymbol{W}(2)$ is a centered multivariate normal random vector with covariance matrix $2I_p$. Hence, for any $v \geq 2$, the distribution of $f_v(\boldsymbol{W})$ has no atom at 0 and is therefore absolutely continuous.

Proof of the continuity of f_v: To show (uniform) continuity on $C([0,v],\mathbb{R}^p)$ (equipped with the uniform distance), let $\varepsilon > 0$ be given and let $\delta = \varepsilon/3$. Let $\boldsymbol{w}, \boldsymbol{w}' \in C([0,v],\mathbb{R}^p)$ be such that $\sup_{0\leq t\leq v} \|\boldsymbol{w}(t) - \boldsymbol{w}'(t)\|_{I_p} < \delta$. Then,

$$\begin{aligned}
f_v(\boldsymbol{w}') &= \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{w}'(s) - s\boldsymbol{w}'(t)\|_{I_p} \\
&= \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta}\|\{t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\} + t\{\boldsymbol{w}'(s) - \boldsymbol{w}(s)\} - s\{\boldsymbol{w}'(t) - \boldsymbol{w}(t)\}\|_{I_p} \\
&\leq \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta}\Big\{\|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p} + t\|\boldsymbol{w}'(s) - \boldsymbol{w}(s)\|_{I_p} + s\|\boldsymbol{w}'(t) - \boldsymbol{w}(t)\|_{I_p}\Big\} \\
&\leq \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p} + 2\delta.
\end{aligned}$$

This shows that $f_v(\boldsymbol{w}') - f_v(\boldsymbol{w}) \leq 2\delta < \varepsilon$. Similarly (or just by symmetry of the above argument), $f_v(\boldsymbol{w}) - f_v(\boldsymbol{w}') < \varepsilon$. Thus, $|f_v(\boldsymbol{w}) - f_v(\boldsymbol{w}')| < \varepsilon$ as required.

Proof of the convexity of f_v: To show convexity, let $\lambda \in (0,1)$ and $\boldsymbol{w}, \boldsymbol{w}' \in C([0,v],\mathbb{R}^p)$. Then $\lambda\boldsymbol{w} + (1-\lambda)\boldsymbol{w}' \in C([0,v],\mathbb{R}^p)$ and

$$\begin{aligned}
&f_v(\lambda\boldsymbol{w} + (1-\lambda)\boldsymbol{w}') \\
&= \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta}\Big\|t\{\lambda\boldsymbol{w}(s) + (1-\lambda)\boldsymbol{w}'(s)\} - s\{\lambda\boldsymbol{w}(t) + (1-\lambda)\boldsymbol{w}'(t)\}\Big\|_{I_p} \\
&= \sup_{1\leq s\leq t\leq v} t^{-\frac{3}{2}-\eta}\Big\|\lambda\{t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\} + (1-\lambda)\{t\boldsymbol{w}'(s) - s\boldsymbol{w}'(t)\}\Big\|_{I_p} \\
&\leq \lambda f_v(\boldsymbol{w}) + (1-\lambda) f_v(\boldsymbol{w}'),
\end{aligned}$$

as required.

To complete the proof of the proposition, it suffices to fix $r, \varepsilon > 0$ and show that $\mathbb{P}(f(\boldsymbol{W}) = r) < \varepsilon$. Notice first that, for any $u \geq 1$, almost surely,

$$\sup_{(s,t):1\leq s\leq t, u\leq t<\infty} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \leq u^{-\frac{\eta}{2}} \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}-\frac{\eta}{2}}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

Since according to Theorem 4.2.6, $\mathcal{L}_{p,\frac{\eta}{2}}$ is almost surely finite, we see that

$$\sup_{(s,t):1\leq s\leq t, u\leq t<\infty} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \overset{a.s.}{\to} 0, \qquad \text{as } u \to \infty.$$

Hence, there exists $u_0 < \infty$ (depending on r, ε) such that

$$\mathbb{P}\left(\sup_{(s,t):1\leq s\leq t, u_0\leq t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \geq r/2 \right) < \varepsilon. \tag{B.10}$$

Now, if $f(\boldsymbol{W}) = r$, then either $f_{u_0}(\boldsymbol{W}) = r$ or the supremum (r) in the definition of f is attained for some s, t with $t > u_0$. Absolute continuity of $f_{u_0}(\boldsymbol{W})$ shows that the former has probability 0, while (B.10) shows that the latter has probability at most ε. ■

B.3 Proof of Proposition 4.2.14

Proof of Proposition 4.2.14. Let A be a $p \times p$ symmetric positive-definite matrix. Then, there exits a diagonal matrix Δ with $\lambda_1 > 0, \ldots, \lambda_p > 0$ on its diagonal and a $p \times p$ orthogonal matrix P such that the columns of P are the eigenvectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p \in \mathbb{R}^p$ of A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_p$ and $A = P\Delta P^\top$. Starting from the definition of $\|\cdot\|_{A^{-1}}$ given below (4.6), for any $\boldsymbol{v} \in \mathbb{R}^p$,

$$\begin{aligned}
\sqrt{p}\|\boldsymbol{v}\|_{A^{-1}} &= |\boldsymbol{v}^\top A^{-1}\boldsymbol{v}|^{\frac{1}{2}} = |\boldsymbol{v}^\top (P\Delta P^\top)^{-1}\boldsymbol{v}|^{\frac{1}{2}} = |\boldsymbol{v}^\top P\Delta^{-1}P^\top \boldsymbol{v}|^{\frac{1}{2}} \\
&\geq \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \|P^\top \boldsymbol{v}\|_2 = \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \sqrt{(P^\top \boldsymbol{v})\cdot(P^\top \boldsymbol{v})} \\
&= \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \|\boldsymbol{v}\|_2 \\
&\geq \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \|\boldsymbol{v}\|_\infty,
\end{aligned}$$

where we have used the fact that orthogonal matrices preserve the dot product.

Let ψ be the map from the set $\mathcal{S}_p$ of $p \times p$ symmetric positive-definite matrices to $(0, \infty)$ such that, for any $A \in \mathcal{S}_p$, $\psi(A) = \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}}$. Since eigenvalue decomposition is a continuous operation and since the minimum is a continuous function, the map ψ is continuous. Hence, $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ and the continuous mapping theorem imply that $\psi(\Sigma_m^{\mathcal{P}}) \xrightarrow{\mathbb{P}} \psi(\Sigma^{\mathcal{P}}) > 0$.

From the previous derivations and the assumptions of the proposition, we thus have that, for all $m \in \mathbb{N}$, $\|\cdot\|_{(\Sigma_m^{\mathcal{P}})^{-1}} \geq p^{-\frac{1}{2}}\psi(\Sigma_m^{\mathcal{P}})\|\cdot\|_\infty$ almost surely. Therefore, for any $m \in \mathbb{N}$ and $k \geq m+1$, almost surely,

$$\begin{aligned}
D_m^{\mathcal{P}}(k) &= \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma_m^{\mathcal{P}})^{-1}} \\
&\geq p^{-\frac{1}{2}}\psi(\Sigma_m^{\mathcal{P}}) \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_\infty \geq p^{-\frac{1}{2}}\psi(\Sigma_m^{\mathcal{P}}) E_m^{\boldsymbol{x}_\ell}(k),
\end{aligned}$$

which immediately implies that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \xrightarrow{\mathbb{P}} \infty$ since $\psi(\Sigma_m^{\mathcal{P}}) \xrightarrow{\mathbb{P}} \psi(\Sigma^{\mathcal{P}}) > 0$ and $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} E_m^{\boldsymbol{x}_\ell}(k) \xrightarrow{\mathbb{P}} \infty$. ■

B.4 Proofs of Propositions 4.3.2, 4.3.4, and 4.3.7

Proof of Proposition 4.3.2. Let $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$ such that $\boldsymbol{a} < \boldsymbol{b}$. Note that

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

and

$$\nu_m\big((\boldsymbol{a}, \boldsymbol{b}]\big) = \sum_{\boldsymbol{i} \in \{0,1\}^d} (-1)^{\sum_{j=1}^d i_j} C_m\big(a_1^{i_1} b_1^{1-i_1}, \ldots, a_d^{i_d} b_d^{1-i_d}\big).$$

By Condition 4.3.1 (ii), we immediately see that, for each $\boldsymbol{\pi} \in \Pi$ (where Π is defined in (4.17)), $\nu_m\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) \overset{a.s.}{\to} \nu_C\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big)$ as $m \to \infty$. By Condition 4.3.1 (i), for each $\boldsymbol{\pi} \in \Pi$, $\nu_C\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) \neq 1/(\kappa(r+1)^d)$. We conclude that, almost surely, for each $\boldsymbol{\pi} \in \Pi$,

$$\begin{aligned} &\nu_m\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) > 1/(\kappa(r+1)^d) \text{ for all } m \text{ sufficiently large} \\ &\quad \iff \nu_C\big((\boldsymbol{\pi} - \boldsymbol{s}, \boldsymbol{\pi}]\big) > 1/(\kappa(r+1)^d). \end{aligned}$$

This completes the proof. ■

Proof of Proposition 4.3.4. Let

$$R(n, \boldsymbol{x}) = \sum_{i=1}^n \{\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}) - F(\boldsymbol{x})\}, \qquad n \in \mathbb{N}, \boldsymbol{x} \in \mathbb{R}^d.$$

From Theorem 3.1 part 2(b) in Dedecker et al. (2014), without changing its distribution, the empirical process R can be redefined on a richer probability space on which there exists a *Kiefer process*, that is, a two-parameter centered continuous Gaussian process K with covariance function given by (4.15), and a random variable $C > 0$ such that, almost surely (a.s.),

$$\sup_{t \in [0,1]} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |R(\lfloor nt \rfloor, \boldsymbol{x}) - K(\lfloor nt \rfloor, \boldsymbol{x})| \leq C n^{\frac{1}{2} - \lambda}, \qquad \text{for all } n \in \mathbb{N} \tag{B.11}$$

for some $\lambda \in (0, 1/2)$ only depending on d and a. Let $i \in [\![1, p]\!]$. Then,

$$\begin{aligned} &\sup_{k > m} k^{-\frac{1}{2}} \max_{j \in [\![1,k]\!]} j |F_{1:j}(\boldsymbol{\mathcal{X}}_{i,m}) - F(\boldsymbol{\mathcal{X}}_{i,m}) - F_{1:j}(\boldsymbol{x}_i) + F(\boldsymbol{x}_i)| \\ &\quad = \sup_{k > m} k^{-\frac{1}{2}} \max_{j \in [\![1,k]\!]} |R(j, \boldsymbol{\mathcal{X}}_{i,m}) - R(j, \boldsymbol{x}_i)| \\ &\quad \leq \sup_{k > m} k^{-\frac{1}{2}} \max_{j \in [\![1,k]\!]} |R(j, \boldsymbol{\mathcal{X}}_{i,m}) - K(j, \boldsymbol{\mathcal{X}}_{i,m})| + \sup_{k > m} k^{-\frac{1}{2}} \max_{j \in [\![1,k]\!]} |R(j, \boldsymbol{x}_i) - K(j, \boldsymbol{x}_i)| \\ &\qquad + \sup_{k > m} k^{-\frac{1}{2}} \max_{j \in [\![1,k]\!]} |K(j, \boldsymbol{\mathcal{X}}_{i,m}) - K(j, \boldsymbol{x}_i)|. \end{aligned}$$

From (B.11), with probability one, the first two terms on the right-hand side of the last display are smaller than

$$2 \sup_{k > m} k^{-\frac{1}{2}} \sup_{t \in [0,1]} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |R(\lfloor kt \rfloor, \boldsymbol{x}) - K(\lfloor kt \rfloor, \boldsymbol{x})| \leq 2 \sup_{k > m} k^{-\frac{1}{2}} C k^{\frac{1}{2} - \lambda} \leq 2C m^{-\lambda} \overset{a.s.}{\to} 0.$$

The third term is smaller than

$$I_m = \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \|\boldsymbol{x}-\boldsymbol{y}\|_\infty\le\Delta_m}} |K(j,\boldsymbol{x}) - K(j,\boldsymbol{y})|, \tag{B.12}$$

where $\Delta_m = \|\boldsymbol{\mathcal{X}}_{i,m} - \boldsymbol{x}_i\|_\infty$. To complete the proof, it thus suffices to show that $I_m \xrightarrow{\mathbb{P}} 0$.

From Theorem 3.1 part 2(a) in Dedecker et al. (2014), we know that the sample paths of the Gaussian process K are almost surely uniformly continuous with respect to the pseudo-metric ρ on $[0,\infty)\times\mathbb{R}^d$ defined by

$$\rho\big((s,\boldsymbol{x}),(t,\boldsymbol{y})\big) = |s-t| + \sum_{\ell=1}^{d} |F^{[\ell]}(x^{[\ell]}) - F^{[\ell]}(y^{[\ell]})|, \qquad s,t\in[0,\infty), \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d.$$

We will use this fact to show that $I_m \xrightarrow{\mathbb{P}} 0$. To this end, define $\Delta_m^* = \psi(\Delta_m)$, where

$$\psi(\delta) = \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \|\boldsymbol{x}-\boldsymbol{y}\|_\infty\le\delta}} \sum_{\ell=1}^{d} |F^{[\ell]}(x^{[\ell]}) - F^{[\ell]}(y^{[\ell]})|, \qquad \delta\ge 0.$$

Since $\|\boldsymbol{x}-\boldsymbol{y}\|_\infty \le \Delta_m$ implies that $\rho\big((t,\boldsymbol{x}),(t,\boldsymbol{y})\big) \le \Delta_m^*$, we have that

$$I_m \le \sup_{k>m} k^{-\frac{1}{2}} \sup_{\substack{s,t\in[0,k] \\ s=t}} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \|\boldsymbol{x}-\boldsymbol{y}\|_\infty\le\Delta_m}} |K(s,\boldsymbol{x}) - K(t,\boldsymbol{y})| \le \sup_{k>m} k^{-\frac{1}{2}} J_m \le m^{-\frac{1}{2}} J_m,$$

where $J_m = \phi(\Delta_m^*)$ with

$$\phi(\delta) = \sup_{\substack{s,t\in[0,\infty),\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \rho((s,\boldsymbol{x}),(t,\boldsymbol{y}))\le\delta}} |K(s,\boldsymbol{x}) - K(t,\boldsymbol{y})|, \qquad \delta\ge 0.$$

Let $\varepsilon > 0$. By almost sure uniform continuity of the sample paths of the process K, there exists $\delta_1 = \delta_1(\varepsilon) > 0$ such that, for all $0 \le \delta \le \delta_1$, $\phi(\delta) < \varepsilon$ almost surely. Since the d univariate margins $F^{[1]},\dots,F^{[d]}$ of F are (uniformly) continuous, there exists $\delta_0 = \delta_0(\delta_1) > 0$ such that, for all $0 \le \delta \le \delta_0$, $\psi(\delta) < \delta_1$. Therefore

$$\begin{aligned}
\mathbb{P}(J_m > \varepsilon) = \mathbb{P}\big(\phi(\Delta_m^*) > \varepsilon\big) &\le \mathbb{P}\big(\phi(\Delta_m^*) > \varepsilon, \Delta_m^* \le \delta_1\big) + \mathbb{P}(\Delta_m^* > \delta_1) \\
&= \mathbb{P}(\Delta_m^* > \delta_1) \\
&= \mathbb{P}\big(\psi(\Delta_m) > \delta_1\big) \\
&\le \mathbb{P}\big(\psi(\Delta_m) > \delta_1, \Delta_m \le \delta_0\big) + \mathbb{P}(\Delta_m > \delta_0) \\
&= \mathbb{P}(\Delta_m > \delta_0).
\end{aligned}$$

Since $\Delta_m = \|\boldsymbol{\mathcal{X}}_{i,m} - \boldsymbol{x}_i\|_\infty \xrightarrow{\mathbb{P}} 0$ by assumption, this shows that $J_m \xrightarrow{\mathbb{P}} 0$ which completes the proof since $I_m \le m^{-\frac{1}{2}} J_m$. ■

Proof of Proposition 4.3.7. The second claim is immediate from the first claim and Theorem 4.2.6, so we need only prove the first claim.

First, we assert that, for any $\ell \in [\![1,p]\!]$,

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{1\le i<j\le k} (j-i+1)|F_{i:j}(\boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - F_{i:j}(\boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)| = o_{\mathbb{P}}(1). \tag{B.13}$$

Indeed, the left-hand side of (B.13) is equal to

$$\begin{aligned}
&\sup_{k>m} k^{-\frac{1}{2}} \max_{1\le i<j\le k} \Big| \sum_{r=i}^{j} \{1(\boldsymbol{X}_r \le \boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - 1(\boldsymbol{X}_r \le \boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)\} \Big| \\
&\quad = \sup_{k>m} k^{-\frac{1}{2}} \max_{1\le i<j\le k} \Big| \sum_{r=1}^{j} \{1(\boldsymbol{X}_r \le \boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - 1(\boldsymbol{X}_r \le \boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)\} \\
&\qquad - \sum_{r=1}^{i-1} \{1(\boldsymbol{X}_r \le \boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - 1(\boldsymbol{X}_r \le \boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)\} \Big| \\
&\quad \le 2 \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} j |F_{1:j}(\boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - F_{1:j}(\boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)|,
\end{aligned}$$

so the assertion holds by (4.24). Using the fact that all norms on $\mathbb{R}^p$ are equivalent, (B.13) implies that

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{1\le i<j\le k} (j-i+1) \|\boldsymbol{F}^{\mathcal{P}_m}_{i:j} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{i:j} - \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} = o_{\mathbb{P}}(1), \tag{B.14}$$

where $\boldsymbol{F}^{\mathcal{P}_m} = \big(F(\boldsymbol{\mathcal{X}}_{1,m}), \dots, F(\boldsymbol{\mathcal{X}}_{p,m})\big)$ and $\boldsymbol{F}^{\mathcal{P}} = \big(F(\boldsymbol{x}_1), \dots, F(\boldsymbol{x}_p)\big)$ (similarly for $\boldsymbol{F}^{\cdot}_{i:j}$).

Recall the definition of $\tilde{D}^{\mathcal{P}}_m$ in (B.1) which can be rewritten as

$$\tilde{D}^{\mathcal{P}}_m(k) = \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}^{\mathcal{P}}_{1:j} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k}\|_{(\Sigma^{\mathcal{P}})^{-1}}, \qquad k \ge m+1,$$

and define the unobservable detector $\tilde{D}^{\mathcal{P}_m}_m$ by changing the norm in the definition of $D^{\mathcal{P}_m}_m$ as

$$\tilde{D}^{\mathcal{P}_m}_m(k) = \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\|_{(\Sigma^{\mathcal{P}})^{-1}}, \qquad k \ge m+1.$$

Using the reverse triangle inequality for the maximum norm and the norm $\|\cdot\|_{(\Sigma^{\mathcal{P}})^{-1}}$, we then obtain that, for any $k \ge m+1$,

$$\begin{aligned}
&|\tilde{D}^{\mathcal{P}_m}_m(k) - \tilde{D}^{\mathcal{P}}_m(k)| \\
&\le \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \\
&\quad \times \Big| \|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k} + \boldsymbol{F}^{\mathcal{P}_m}\|_{(\Sigma^{\mathcal{P}})^{-1}} - \|\boldsymbol{F}^{\mathcal{P}}_{1:j} - \boldsymbol{F}^{\mathcal{P}} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \Big| \\
&\le \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{1:j} + \boldsymbol{F}^{\mathcal{P}} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k} + \boldsymbol{F}^{\mathcal{P}_m} + \boldsymbol{F}^{\mathcal{P}}_{j+1:k} - \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \\
&\le \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \\
&\quad \times \Big\{ \|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{1:j} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} + \|\boldsymbol{F}^{\mathcal{P}_m}_{j+1:k} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \Big\}.
\end{aligned}$$

Therefore,

$$\sup_{k>m} \Big(\frac{m}{k}\Big)^{\frac{3}{2}+\eta} |\tilde{D}^{\mathcal{P}_m}_m(k) - \tilde{D}^{\mathcal{P}}_m(k)|$$

$$\leq \sup_{k>m} \left(\frac{m}{k}\right)^{\eta} k^{-\frac{3}{2}} \max_{j\in[\![m,k[\![} j(k-j)\|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{1:j} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}$$
$$+ \sup_{k>m} \left(\frac{m}{k}\right)^{\eta} k^{-\frac{3}{2}} \max_{j\in[\![m,k[\![} j(k-j)\|\boldsymbol{F}^{\mathcal{P}_m}_{j+1:k} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}.$$

We claim that both terms on the right hand side converge to 0 in probability. For example, since $j \leq k$ and $k > m$, the second term is at most

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![m,k[\![} (k-j)\|\boldsymbol{F}^{\mathcal{P}_m}_{j+1:k} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}$$
$$\leq \sup_{k>m} k^{-\frac{1}{2}} \max_{1\leq i<j\leq k} (j-i+1)\|\boldsymbol{F}^{\mathcal{P}_m}_{i:j} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}}_{i:j} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}},$$

which converges to 0 in probability by (B.14). The first term is similar. Hence,

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |\tilde{D}^{\mathcal{P}_m}_m(k) - \tilde{D}^{\mathcal{P}}_m(k)| = o_{\mathbb{P}}(1). \tag{B.15}$$

From Lemma B.1.4, we have that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}|\tilde{D}^{\mathcal{P}}_m(k) - D^{\mathcal{P}}_m(k)| = o_{\mathbb{P}}(1)$. Therefore, it remains to prove that

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |D^{\mathcal{P}_m}_m(k) - \tilde{D}^{\mathcal{P}_m}_m(k)| = o_{\mathbb{P}}(1).$$

Proceeding as in the proof of Lemma B.1.4, we have that

$$\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}|\tilde{D}^{\mathcal{P}_m}_m(k) - D^{\mathcal{P}_m}_m(k)|$$
$$\leq \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \left| \|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\|_{(\Sigma^{\mathcal{P}})^{-1}} - \|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\|_{(\Sigma^{\mathcal{P}_m}_m)^{-1}} \right|$$
$$\leq \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}$$
$$\times \max_{j\in[\![m,k[\![} \frac{j(k-j)}{p^{\frac{1}{2}}m^{\frac{3}{2}}} \left| \left(\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\right)^{\top} \left((\Sigma^{\mathcal{P}})^{-1} - (\Sigma^{\mathcal{P}_m}_m)^{-1}\right) \left(\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\right) \right|^{\frac{1}{2}}$$
$$\leq \frac{1}{p^{\frac{1}{2}}} \left\|(\Sigma^{\mathcal{P}})^{-1} - (\Sigma^{\mathcal{P}_m}_m)^{-1}\right\|^{\frac{1}{2}}_{\mathrm{op}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \left\|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\right\|_2.$$

We claim that this converges to zero in probability (which completes the proof). By Condition 4.2.3 and the fact that $\Sigma^{\mathcal{P}_m}_m \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$, we have that $\|(\Sigma^{\mathcal{P}})^{-1} - (\Sigma^{\mathcal{P}}_m)^{-1}\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$. To prove this final claim, it therefore suffices to show that

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \left\|\boldsymbol{F}^{\mathcal{P}_m}_{1:j} - \boldsymbol{F}^{\mathcal{P}_m}_{j+1:k}\right\|_2 = O_{\mathbb{P}}(1).$$

Indeed, by equivalence of norms on $\mathbb{R}^p$, this follows from the fact that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}\tilde{D}^{\mathcal{P}_m}_m(k) = O_{\mathbb{P}}(1)$, itself a consequence of (B.15), Lemma B.1.4, and Theorem 4.2.6. ■

B.5 Details of Monte Carlo experiments

In this section, we provide the implementation details of the Monte Carlo experiments that we carried out in the case of low-dimensional ($d \in \{1, 2, 3\}$) continuous observations and whose main findings are summarized in Section 4.5. In all experiments, the sequential tests were carried out at the $\alpha = 5\%$ nominal level.

B.5.1 Univariate experiments for the procedure based on $D_m^{\mathcal{P}_m}$ under the null

To investigate the empirical levels of the sequential test based on $D_m^{\mathcal{P}_m}$ when $d = 1$, we considered 9 data generating models, denoted M1, ..., M9. Models M1, ..., M6 are AR(1) models with independent standard normal innovations whose autoregressive parameter is equal to 0, 0.1, 0.3, 0.5, 0.7 and -0.7, respectively. Model M7 is a GARCH(1,1) model with independent standard normal innovations and parameters $\omega = 0.012$, $\beta = 0.919$ and $\alpha = 0.072$ to mimic SP500 daily log-returns following Jondeau et al. (2007). Models M8 and M9 are the nonlinear autoregressive model used in Paparoditis and Politis (2001, Section 3.3) and the exponential autoregressive model considered in Auestad and Tjøstheim (1990) and Paparoditis and Politis (2001, Section 3.3), respectively. The underlying generating equations are

$$X_i = 0.6 \sin(X_{i-1}) + \epsilon_i$$

and

$$X_i = \{0.8 - 1.1 \exp(-50X_{i-1}^2)\}X_{i-1} + 0.1\epsilon_i,$$

respectively, where the ϵ_i are independent standard normal innovations. Note that, for all time series models, a burn-out sample of 100 observations was used.

For each of the nine models, the probability of rejection of H_0 in (4.1) was estimated from 1000 samples of size $n = m + 5000$ with $m \in \{200, 400, 800, 1600\}$ and for $p \in \{2, 5, 10, 20\}$. The empirical levels are reported in Table B.1. As one can see, for any fixed p, reassuringly, they decrease as m increases. For $m \geq 800$, it is mostly for the models with strong serial dependence such as M5, M8 and M9 that the empirical levels are not below the 5% nominal level. The latter is not so surprising and highlights the difficulty of the estimation of the long-run covariance matrix $\Sigma^{\mathcal{P}}$ using the estimator $\Sigma_m^{\mathcal{P}_m}$ in the case of strong serial dependence. It may be slightly more surprising for the GARCH(1,1) model M7 for which $m = 1600$ seems necessary to obtain a reasonably good estimate of $\Sigma^{\mathcal{P}}$. The fact that for most other models, the empirical levels are all below the 5% nominal level when $m \geq 800$ is a consequence of the fact that they are underestimated in all settings. Indeed, the monitoring was stopped after 5000 steps whereas it would theoretically be necessary to monitor "indefinitely" to compute empirical levels accurately. For any fixed m, we see that increasing p tends in general to increase the empirical level. This is again a consequence of the difficulty of the estimation of the long-run covariance matrix $\Sigma^{\mathcal{P}}$ which is a $p \times p$ matrix. Note that, since the monitoring procedure based on $D_m^{\mathcal{P}_m}$ is margin-free (as verified in Section 4.3.4), it is not necessary to empirically study the influence of the contemporary distribution F on the empirical levels.

In a second experiment, we briefly investigated the effect of the estimation of $\Sigma^{\mathcal{P}}$ on the empirical levels in the case of independent observations. Instead of estimating $\Sigma^{\mathcal{P}}$ from the learning sample, we used its true value whose elements, in the considered setting (see Remark 4.2.4), are given by

$$\mathrm{Cov}\{\mathbf{1}(X_1^{[1]} \leq F^{-1}(i/(p+1)), \mathbf{1}(X_1^{[1]} \leq F^{-1}(j/(p+1))\} = \min(i,j)/(p+1) - ij/(p+1)^2,$$

for $i, j \in [\![1, p]\!]$. The empirical levels were then estimated from 1000 random samples of size $n = m+5000$ from the standard normal distribution. The results are reported in Table B.2. By comparing the results with the first horizontal block of Table B.1,

Table B.1: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.1. The rejection percentages are computed from 1000 samples of size $n = m + 5000$ generated from the time series models M1, ..., M9.

Model	m	$p = 2$	$p = 5$	$p = 10$	$p = 20$
M1	200	3.4	3	3.3	5.4
	400	1.9	2.1	3.2	2.8
	800	1.5	1.3	2	1.3
	1600	1	0.9	0.9	1
M2	200	4.4	4.3	8.1	13.8
	400	3.2	2.7	3.6	5.6
	800	1.8	2.8	2.1	3.2
	1600	1.3	1	1.7	1
M3	200	6.9	10.6	21.5	53.9
	400	3.4	4.1	7.2	16.7
	800	3	2.5	3	5.6
	1600	1.5	1.3	1.5	1.4
M4	200	9.5	18.1	44.6	93.2
	400	5.3	7.1	16.5	42.9
	800	3	3.6	5.3	12.3
	1600	2.2	2.3	2.8	2.2
M5	200	18.9	39.4	82.1	100
	400	9.6	17.5	40.3	87
	800	5.7	6.7	14.4	34.5
	1600	2.4	2.7	4.4	7.5
M6	200	7.3	8.6	14.1	26.7
	400	3.7	4.6	6.4	11
	800	2.6	2.8	3	3.6
	1600	1.2	1.5	1.3	1.2
M7	200	9.1	14.8	17.2	16.8
	400	8	12	12.6	9.9
	800	7.5	9.8	9.2	5.2
	1600	4.6	6.1	4.8	2.9
M8	200	7.8	13	30.6	74.3
	400	4.7	5.5	10.8	30.8
	800	2.8	3.4	4.7	7.9
	1600	1.8	1.8	1.9	2.1
M9	200	15	29	61.1	95.6
	400	8.6	12.9	25.5	57.9
	800	5.5	6.7	10.1	17.3
	1600	2.1	2.4	3.4	5.3

we see, as could have been expected, that for the same value of m, the use of the true long-run covariance matrix leads to lower empirical levels than when it is estimated.

As a last experiment under the null, we investigated the quality of the model fitted at the end of Section 4.4 to extrapolate the values of the quantiles of the distribution of $\mathcal{L}_{p,\eta}$ for $p > 20$ and $\eta = 0.001$. Using $m = 1600$ and 1000 random samples of size $n = m + 5000$ from the standard normal distribution, we estimated rejection percentages for $p \in \{30, 40, 50\}$. These are given in Table B.3 and suggest that the quality of the model for extrapolating the values of the quantiles may

Table B.2: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.1 when the true underlying long-run covariance matrix is used instead of its estimate. The rejection percentages are computed from 1000 random samples of size $n = m + 5000$ generated from the normal distribution.

m	$p = 2$	$p = 5$	$p = 10$	$p = 20$
200	2.4	2.4	2.3	3.1
400	1.7	1.9	3.1	2.3
800	1.3	1.3	1.6	0.7
1600	0.8	0.9	0.9	0.6

Table B.3: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.1 when, for $p > 20$ and $\eta = 0.001$, estimates of the 0.95-quantiles of the distribution $\mathcal{L}_{p,\eta}$ are extrapolated using the model fitted at the end of Section 4.4. The rejection percentages are computed from 1000 samples of size $n = m + 5000$ generated from the standard normal distribution with $m = 1600$.

$p = 2$	$p = 5$	$p = 10$	$p = 20$	$p = 30$	$p = 40$	$p = 50$
1.0	0.9	0.9	1.0	0.4	0.3	0.5

be acceptable when $20 < p \leq 50$ (although it may lead to some slightly more conservative tests).

B.5.2 Univariate experiments for the procedure based on $D_m^{\mathcal{P}_m}$ under alternatives

In order to understand the behavior of the monitoring procedure based on the detector $D_m^{\mathcal{P}_m}$ considered in Section 4.3.1 under alternatives to H_0 in (4.1), we considered successively changes in the expectation of the $X_i^{[1]}$'s, in their variance and in their d.f. (while keeping their expectation and variance constant). As a first experiment, we studied the finite-sample behavior of the sequential test under a change in the expectation of an AR(1) model with autoregressive parameter equal to 0.3 (Model M3). Specifically, to estimate rejection percentages, we generated 1000 samples of size $n = m + 5000$ from Model M3 with $m = 800$ and, for each sample, added a positive offset of δ to all observations after position $m + k$ with $k \in \{0, 500, 1000, 2000\}$. The results are reported in Figure B.1. Notice that only the exceedences (of the detectors with respect to their thresholds) after position $m + k$ are taken into account when calculating the rejection percentages.

As one can see from the top row of graphs in Figure B.1, as expected, the power of all procedures increases as δ increases. Furthermore, for the procedure based on $D_m^{\mathcal{P}_m}$ and a fixed value of the offset δ, increasing the number p of evaluation points slightly lowers the power of the test. We also see that the procedure based on R_m in (4.2) is always the most powerful. This was to be expected as the latter was specifically designed to be sensitive to changes in the mean. Similarly, from the second row of plots in Figure B.1, we see that mean detection delays are smallest for the procedure based on R_m and increase for the procedure based on $D_m^{\mathcal{P}_m}$ as p increases. Note finally that the power of every procedure becomes larger as the time k at which the offset δ is added becomes closer to half of $n = m + 5000$. This is a

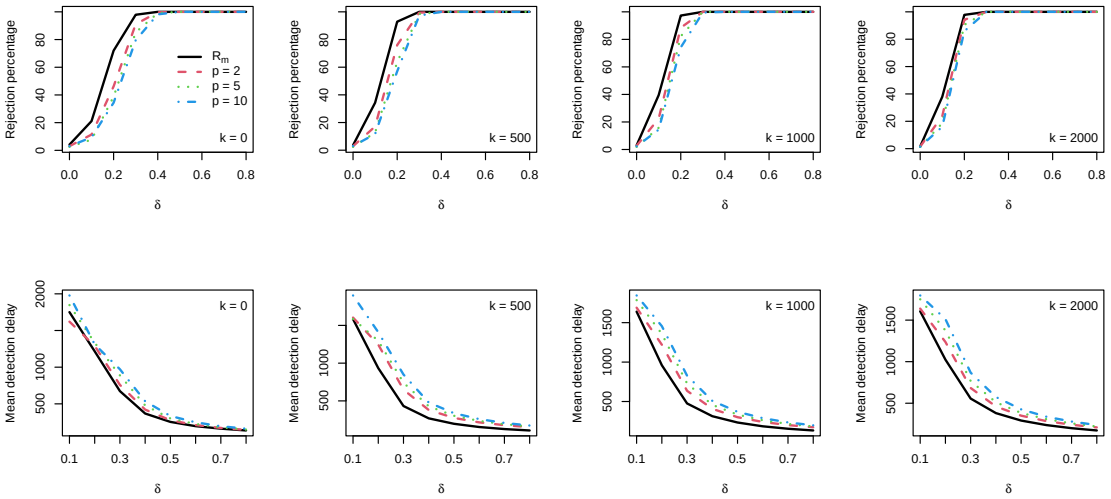


Figure B.1: Rejection percentages of H_0 in (4.1) (first row) and corresponding mean detection delays (second row) for the procedure based on R_m in (4.2) (solid line) and for the procedure $D_m^{\mathcal{P}_m}$ considered in Section 4.3.1 with $p \in \{2, 5, 10\}$ (dash, dotted and dash-dotted lines) estimated from 1000 samples of size $n = m + 5000$ from Model M3 with $m = 800$ such that, for each sample, a positive offset of δ was added to all observations after position $m + k$.

consequence of using of CUSUM statistics to define the detectors.

As a second experiment, we considered a change in the variance of independent centered observations. To estimate the power of the sequential test, we generated 1000 samples of size $n = m + 5000$ with $m = 800$ such that, observations up to position $m + k$ with $k \in \{0, 500, 1000, 2000\}$ are from the standard normal distribution while observations after position $m + k$ are from the $N(0, \sigma^2)$ distribution. The results are reported in Figure B.2. As expected, the power of the procedure based on $D_m^{\mathcal{P}_m}$ increases as σ deviates further away from one. In contrast to the first experiment however, the rejection percentages (resp. mean detection delays) increase (resp. decrease) as the number of evaluation points p increases. Notice that the improvement as p increases from 5 to 10 appears to be rather small.

As a final experiment, we considered a change in the contemporary distribution of independent observations that keeps the expectation and the variance constant. To estimate the rejection percentages, we generated 1000 samples of size $n = m + 5000$ with $m = 800$ such that observations up to position $m + k$ with $k \in \{0, 500, 1000, 2000\}$ are from the scaled Student t distribution with 3 degrees of freedom (where the scaling is performed so that the variance is equal to one) while observations after position $m + k$ are from the scaled Student t distribution with $\nu \in [\![3, 10]\!]$ degrees of freedom. The results are reported in Figure B.3. As expected, the power of the procedure based on $D_m^{\mathcal{P}_m}$ increases as ν increases. Furthermore, as in the previous experiment, the rejection percentages (resp. mean detection delays) are larger (resp. smaller) when $p \in \{5, 10\}$. Somehow surprisingly however, the results seem slightly better when $p = 5$.

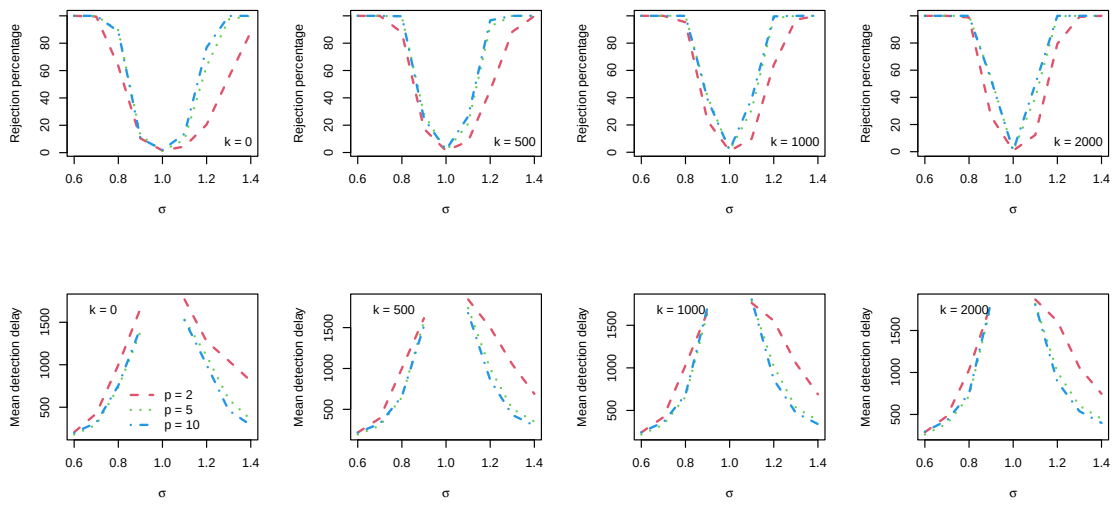


Figure B.2: Rejection percentages of H_0 in (4.1) and corresponding mean detection delays for the procedure based on $D_m^{\mathcal{P}m}$ with $p \in \{2, 5, 10\}$ estimated from 1000 random samples of size $n = m + 5000$ with $m = 800$ such that observations up to position $m + k$ with $k \in \{0, 500, 1000, 2000\}$ are from the standard normal distribution while observations after position $m + k$ are for the $N(0, \sigma^2)$ distribution.

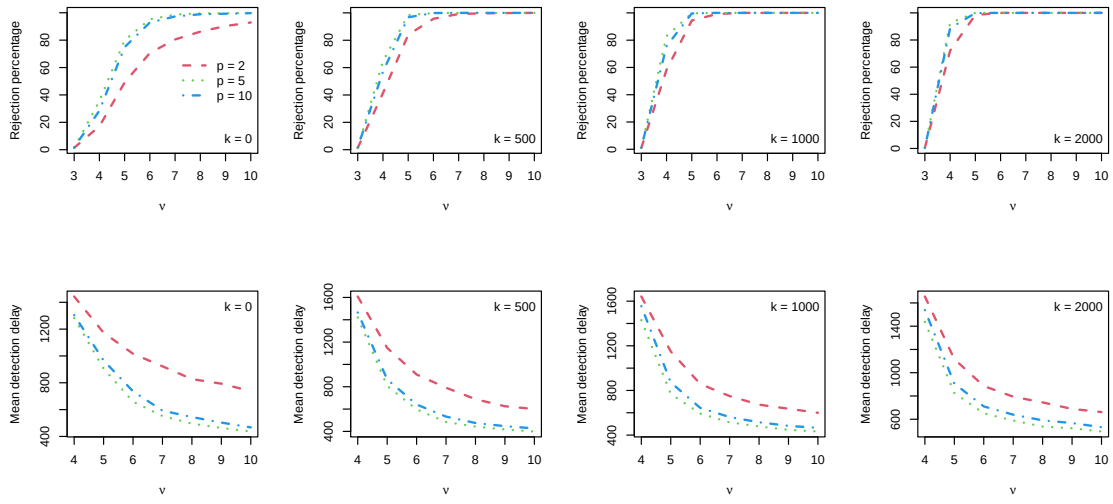


Figure B.3: Rejection percentages of H_0 in (4.1) and corresponding mean detection delays for the procedure based on $D_m^{\mathcal{P}m}$ with $p \in \{2, 5, 10\}$ estimated from 1000 random samples of size $n = m + 5000$ with $m = 800$ such that observations up to position $m+k$ with $k \in \{0, 500, 1000, 2000\}$ are from the scaled Student t distribution with $\nu = 3$ degrees of freedom while observations after position $m + k$ are from the scaled Student t distribution with $\nu \in [\![3, 10]\!]$ degrees of freedom.

Table B.4: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.2 with $r \in \{3, 4\}$ and $\kappa \in \{1.5, 2, 3\}$. The rejection percentages are computed from 1000 bivariate samples of size $n = m + 5000$ generated from the time series model (B.16) with $\beta = 0.3$ and C the bivariate Gumbel–Hougaard copula with a Kendall's tau of $\tau \in \{0, 0.3, 0.6, 0.9\}$. The column $\bar{p}$ reports the average number of grid points retained by the point selection procedure.

		$\kappa = 1.5$				$\kappa = 2$				$\kappa = 3$			
		$r = 3$		$r = 4$		$r = 3$		$r = 4$		$r = 3$		$r = 4$	
m	τ	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$
400	0.00	8.9	9.1	15.4	14.8	9.0	10.4	15.9	18.4	9.0	10.4	16.0	21.6
	0.30	8.6	7.4	14.3	13.2	8.9	8.4	15.3	16.6	9.0	9.4	15.9	20.5
	0.60	7.0	7.7	10.7	10.0	7.1	8.0	11.7	11.3	7.7	8.1	13.4	13.6
	0.90	3.0	4.0	5.8	3.7	3.4	3.6	7.8	4.6	5.9	4.5	9.8	12.9
800	0.00	9.0	4.5	15.9	6.6	9.0	4.7	16.0	7.6	9.0	4.7	16.0	7.6
	0.30	8.7	4.6	14.5	3.7	9.0	4.9	15.6	4.9	9.0	5.0	16.0	5.6
	0.60	7.0	3.6	10.3	4.4	7.0	3.7	11.9	3.5	7.8	3.5	13.8	5.9
	0.90	3.0	2.1	5.2	1.8	3.2	2.2	8.4	3.1	6.3	3.0	10.0	7.5
1600	0.00	9.0	2.2	16.0	2.7	9.0	2.2	16.0	2.7	9.0	2.2	16.0	2.7
	0.30	8.9	1.8	14.7	2.8	9.0	2.2	15.8	2.6	9.0	2.2	16.0	2.7
	0.60	7.0	1.2	10.1	1.1	7.0	1.2	12.0	1.1	7.6	1.3	14.0	1.8
	0.90	3.0	1.2	4.9	1.3	3.0	1.3	8.8	2.0	6.5	0.9	10.0	3.3

B.5.3 Multivariate experiments for the procedure based on $D_m^{\mathcal{P}_m}$ under the null

Given $d \in \{2, 3\}$ and a d-dimensional copula C, we used a multivariate AR(1) model to generate potentially serially dependent observations under H_0 in (4.1). Let $\boldsymbol{U}_i$, $i \in [\![-100, n]\!]$, be a d-dimensional i.i.d. sample from a copula C. Then, set $\boldsymbol{\epsilon}_i = (\Phi^{-1}(U_i^{[1]}), \ldots, \Phi^{-1}(U_i^{[d]}))$, where Φ is the d.f. of the standard normal distribution, and $\boldsymbol{X}_{-100} = \boldsymbol{\epsilon}_{-100}$. Finally, for any $j \in [\![1, d]\!]$ and $i \in [\![-99, n]\!]$, compute recursively

$$X_i^{[j]} = \beta X_{i-1}^{[j]} + \epsilon_i^{[j]}. \tag{B.16}$$

Recall that, when $d > 1$, the evaluation points of the monitoring procedure based on $D_m^{\mathcal{P}_m}$ are chosen from the learning sample using the point selection procedure described in Section 4.3.2. To evaluate the behavior of the procedure when $d = 2$ with $r \in \{3, 4\}$ and $\kappa \in \{1.5, 2, 3\}$ under the null, in a first experiment, we computed its rejection percentages from 1000 bivariate samples of size $n = m + 5000$ generated from the time series model (B.16) with $\beta = 0.3$ and C the bivariate Gumbel–Hougaard copula with a Kendall's tau of $\tau \in \{0, 0.33, 0.66\}$. The empirical levels are reported in the columns $D_m^{\mathcal{P}_m}$ of Table B.4. The columns $\bar{p}$ report the average number of grid points retained by the point selection procedure of Section 4.3.2. As one can see, reassuringly, the empirical levels improve in all settings as m increases. Unsurprisingly, they are higher for $r = 4$ than for $r = 3$ since a larger value of r tends to result in a larger number of selected points p and thus in a more difficult estimation of the underlying long run covariance matrix. Also unsurprisingly, the number of selected points p tends to increase as κ increases and to decrease as τ increases, that is, as the cross-sectional dependence in the underlying time series

Table B.5: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.2 with $r \in \{2, 3\}$ and $\kappa \in \{1.5, 2, 3\}$. The rejection percentages are computed from 1000 trivariate samples of size $n = m + 5000$ generated from the time series model (B.16) with $\beta = 0.3$ and C the trivariate Clayton copula whose bivariate margins have a Kendall's tau of $\tau \in \{0, 0.3, 0.6, 0.9\}$. The column $\bar{p}$ reports the average number of grid points retained by the point selection procedure.

		$\kappa = 1.5$				$\kappa = 2$				$\kappa = 3$			
		$r = 2$		$r = 3$		$r = 2$		$r = 3$		$r = 2$		$r = 3$	
m	τ	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$	$\bar{p}$	$D_m^{\mathcal{P}_m}$
400	0.00	7.6	12.5	20.8	32.0	7.9	14.6	24.1	43.6	8.0	16.0	26.0	55.6
	0.30	7.4	10.4	18.3	22.4	7.8	12.2	21.4	33.7	8.0	15.4	24.0	49.2
	0.60	6.5	6.5	14.6	18.4	7.5	10.9	15.4	22.3	8.0	16.5	16.5	25.4
	0.90	2.0	3.9	6.4	4.4	2.0	3.8	8.5	6.7	3.2	3.4	11.2	14.9
800	0.00	7.9	5.6	24.1	11.7	8.0	6.6	26.3	18.1	8.0	6.8	26.9	22.6
	0.30	7.6	4.1	19.9	9.8	8.0	5.7	23.1	14.3	8.0	6.1	25.6	19.8
	0.60	6.7	2.8	15.0	7.8	7.9	5.2	15.3	9.1	8.0	6.7	16.5	8.8
	0.90	2.0	2.7	6.6	3.0	2.0	2.7	9.2	4.7	2.5	1.9	12.2	6.7
1600	0.00	8.0	2.4	26.2	3.8	8.0	2.4	27.0	5.7	8.0	2.4	27.0	6.0
	0.30	7.8	1.9	20.7	2.5	8.0	2.6	24.3	4.2	8.0	2.6	26.4	6.3
	0.60	6.9	1.0	15.0	2.2	8.0	1.8	15.1	2.3	8.0	2.2	16.2	2.4
	0.90	2.0	1.6	6.8	1.4	2.0	1.6	9.5	2.3	2.1	1.4	13.0	4.3

changes from independence to stronger positive association.

We additionally considered a trivariate version of the previous experiment under the null based on the Clayton copula. The results, reported in Table B.5, are qualitatively the same.

B.5.4 Multivariate experiments for the procedure based on $D_m^{\mathcal{P}_m}$ under alternatives

In a last series of bivariate and trivariate experiments, we investigated the power of the procedure based on $D_m^{\mathcal{P}_m}$.

We first estimated its rejection percentages and corresponding mean detection delays for $r \in \{3, 4\}$ and $\kappa = 1.5$ from 1000 samples of size $n = m + 5000$ with $m = 800$ generated from the time series model (B.16) with $\beta = 0.3$ and C the bivariate Frank copula with a Kendall's tau of $\tau = 0.5$ such that, for each sample, a positive offset of δ was added to the first component of all bivariate observations after position $m + k$. The results are represented in Figure B.4. As one can see, the value of $r \in \{3, 4\}$ has hardly any influence on the power or on the mean detection delay.

We next considered a similar experiment where the change affects only the first margin which changes from the scaled Student t with 3 degrees of freedom to the scaled Student t distribution with $\nu \in [\![3, 10]\!]$ degrees of freedom. The copula (the bivariate Frank with a Kendall's tau of 0.5) and the second margin (the Student t with $\nu = 3$ degrees of freedom) remain constant. The results are displayed in Figure B.5. As one can see, using $r = 4$ rather than $r = 3$ leads to a slightly more

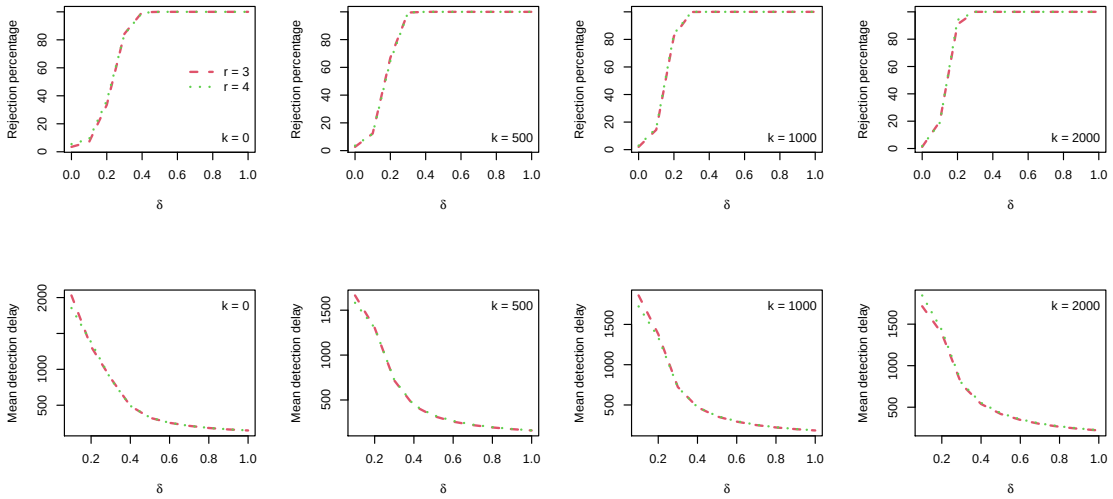


Figure B.4: Rejection percentages of H_0 in (4.1) and corresponding mean detection delays for the procedure based on $D_m^{\mathcal{P}_m}$ with $r \in \{3, 4\}$ and $\kappa = 1.5$ estimated from 1000 bivariate samples of size $n = m + 5000$ with $m = 800$ generated from the time series model (B.16) with $\beta = 0.3$ and C the bivariate Frank copula with a Kendall's tau of 0.5 such that, for each sample, a positive offset of δ was added to the first component of all bivariate observations after position $m + k$. The average number of selected points is approximately 7 for $r = 3$ and 10.9 for $r = 4$.

powerful procedure which detects the change faster on average.

In a third experiment, we focused on the effect of a change of the dependence parameter of the copula in the case of serially independent data. Before the change, observations are generated from the bivariate Normal copula with a Kendall's tau of 0.5, while after the change they come from the bivariate Normal copula with a Kendall's tau of $\tau \in \{0.1, \ldots, 0.9\}$. The rejection percentages and corresponding mean detection delays are represented in Figure B.6. As in the previous experiment, the results for $r = 4$ are slightly better than for $r = 3$.

In a fourth bivariate experiment, we considered the situation where the copula changes while the strength of association measured in terms of Kendall's tau remains constant. Specifically, before the change, observations are generated from the bivariate Clayton copula with a Kendall's tau of 0.5 (which is lower tail dependent), while after the change they arise from the bivariate Gumbel–Hougaard copula with a Kendall's tau of 0.5 (which is upper tail dependent). The results are reported Table B.6. The procedure with $r = 4$ is again slightly more powerful and detects the change faster than the procedure with $r = 3$.

We concluded our multivariate simulations under alternatives by considering trivariate versions of the previous bivariate experiments. We used $r = 3$ and $\kappa = 1.5$. The results are reported in Figure B.7 and Table B.7 and are qualitatively the same as in the bivariate case. Notice however the rather low estimated rejection percentages for the trivariate version of the second bivariate experiment reported in the second row of graphs of Figure B.7.

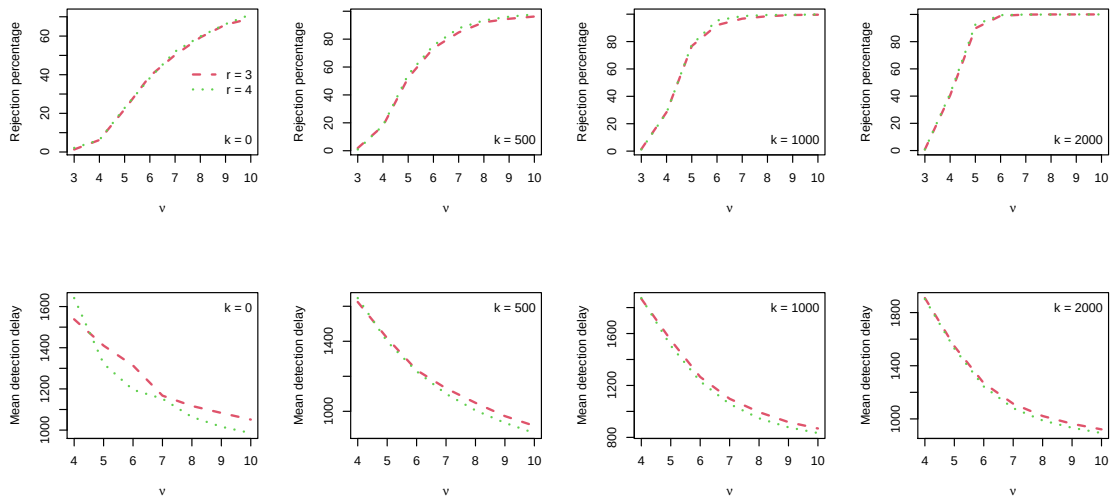


Figure B.5: Rejection percentages of H_0 in (4.1) and corresponding mean detection delays for the procedure based on $D_m^{\mathcal{P}_m}$ with $r \in \{3, 4\}$ and $\kappa = 1.5$ estimated from 1000 bivariate random samples of size $n = m + 5000$ with $m = 800$ such that observations up to position $m+k$ are from a d.f. whose copula is the bivariate Frank with a Kendall's tau of 0.5 and whose margins are scaled Student t with $\nu = 3$ degrees of freedom, while observations after position $m+k$ are still from a d.f. with the same copula and same second margin but with first margin the scaled Student t with $\nu \in [\![3, 10]\!]$ degrees of freedom. The average number of selected points is approximately 7 for $r = 3$ and 10.6 for $r = 4$.

Table B.6: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.2 with $r \in \{3, 4\}$ and $\kappa = 1.5$. The rejection percentages are computed from 1000 bivariate samples of size $n = m + 5000$ with $m = 800$ such that, up to time $m + k$, observations come from a bivariate Clayton copula with a Kendall's tau of 0.5, while observations after time $m + k$ are generated from a Gumbel–Hougaard with a Kendall's tau of 0.5. The abreviation "m.d.d." stands for "mean detection delay".

	$r = 3$			$r = 4$		
k	$\bar{p}$	$D_m^{\mathcal{P}_m}$	m.d.d.	$\bar{p}$	$D_m^{\mathcal{P}_m}$	m.d.d.
0	6.8	72.3	1063.9	10.7	74.4	1007.7
500	6.8	92.9	967.9	10.8	94.8	936.4
1000	6.8	99.0	990.1	10.7	98.9	905.6
2000	6.9	99.8	1002.3	10.7	99.8	965.1

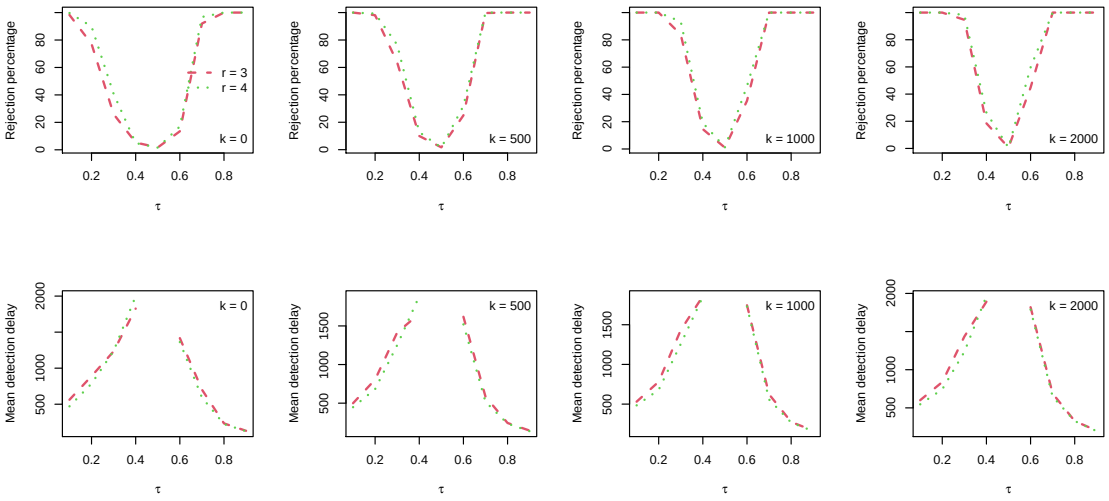


Figure B.6: Rejection percentages of H_0 in (4.1) and corresponding mean detection delays for the procedure based on $D_m^{\mathcal{P}_m}$ with $r \in \{3, 4\}$ and $\kappa = 1.5$ estimated from 1000 bivariate random samples of size $n = m + 5000$ with $m = 800$ such that observations up to position $m + k$ are from the bivariate normal copula with a Kendall's tau of 0.5 while observations after position $m + k$ are from the bivariate normal copula with a Kendall's tau of $\tau \in \{0.1, \ldots, 0.9\}$. The average number of selected points is approximately 7 for $r = 3$ and 11.8 for $r = 4$.

Table B.7: Percentages of rejection of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 4.3.2 with $r = 3$ and $\kappa = 1.5$. The rejection percentages are computed from 1000 trivariate samples of size $n = m + 5000$ with $m = 800$ such that, up to time $m + k$, observations come from a trivariate Clayton copula whose bivariate margins have a Kendall's tau of 0.5, while observations after time $m + k$ are generated from a Gumbel–Hougaard copula whose bivariate margins have a Kendall's tau of 0.5.

k	$\bar{p}$	$D_m^{\mathcal{P}_m}$
0	15.3	95.2
500	15.2	99.8
1000	15.2	100.0
2000	15.2	100.0

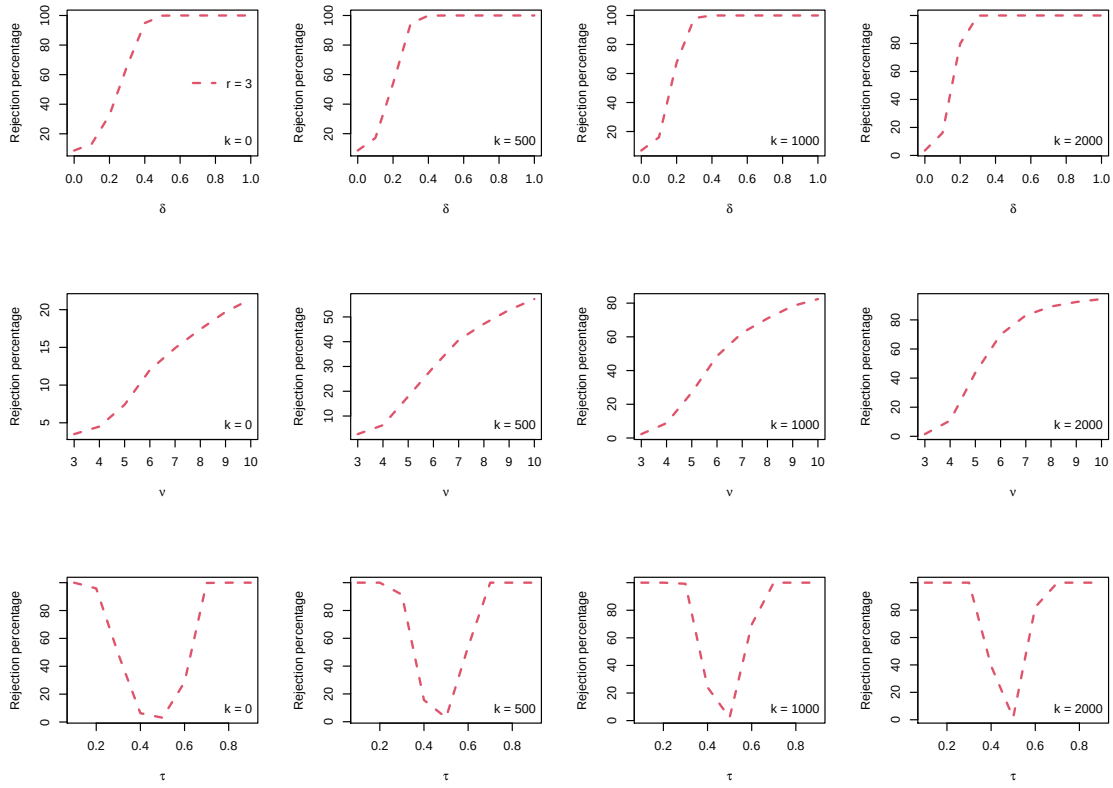


Figure B.7: Rejection percentages of H_0 in (4.1) for the procedure based on $D_m^{\mathcal{P}_m}$ in the trivariate case with $r = 3$ and $\kappa = 1.5$. The graphs in the first, second and third row correspond to experiments which are the trivariate analogs of those reported in Figures B.4, B.5 and B.6, respectively. The corresponding average numbers of selected points are 17.7, 17.3, and 18.3, respectively.