Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXIII ciclo

Inference on copula–based correlation structures

Enrico Foscolo

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Marzo 2011

Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXIII ciclo

Inference on copula–based correlation structures

Enrico Foscolo

Coordinatore: Prof.ssa Daniela Cocchi
Tutor: Dott.ssa Alessandra Luati
Co-Tutor: Dott.ssa Daniela Giovanna Calò

Settore Disciplinare: SECS-S/01

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Marzo 2011

Ringraziamenti

Ringrazio Alessandra e Daniela per avermi introdotto alla problematica oggetto di questa tesi di dottorato e per aver sostenuto il progetto con discussioni costruttive.

Infine, sebbene sia impossibile farne un elenco, voglio qui rimarcare la mia gratitudine verso coloro che, a vario titolo ed in momenti diversi, hanno contribuito a rendere il manoscritto un'opera migliore.

A tutte queste persone va un caloroso abbraccio.

Abstract

We propose an extension of the approach provided by Klüppelberg and Kuhn (2009) for inference on second–order structure moments. As in Klüppelberg and Kuhn (2009) we adopt a copula–based approach instead of assuming normal distribution for the variables, thus relaxing the equality in distribution assumption. A new copula–based estimator for structure moments is investigated. The methodology provided by Klüppelberg and Kuhn (2009) is also extended considering the copulas associated with the family of Eyraud–Farlie–Gumbel–Morgenstern distribution functions (Kotz, Balakrishnan, and Johnson, 2000, Equation 44.73). Finally, a comprehensive simulation study and an application to real financial data are performed in order to compare the different approaches.

"*. . . fatti non foste a viver come bruti,*
ma per seguir virtute e canoscenza."
(Dante, *Commedia*, *Inferno*, XXVI, 119 – 120)

Contents

Chapter 1

Dependence concepts, copulas, and latent variable models: a new challenge

Modern data analysis calls for an understanding of stochastic dependence going beyond simple linear correlation and gaussianity. Literature has been shown a growing interest in modeling multivariate observations using flexible functional forms for distribution functions and in estimating parameters that capture the dependence among different random variables. One of the main reasons for such interest is that the traditional approach based on linear correlation and multivariate normal distribution is not flexible enough for representing a wide range of distribution shapes.

The need of overwhelming linear correlation–based measures and normal distribution assumptions goes well with a typical problems for researchers that are interested in studying the dependence structure between observed variables aiming at a reduction in dimension. Dimension reduction means the possibility of isolating a lower set of underlying, explanatory, not immediately observable, information sources that describe the dependence structure between the observed variables.

Typically, a linear combination of these so–called *latent variables* is considered for a multivariate dataset. In other words, we say that the manifest variables are equally distributed to a linear combination of a few number of latent variables. Thus, this relationship generates what we call a *structure* and it explains the strength of dependence of the data. In what follows, we shall refer to the latent variable model as a *linear structure model* for the observations.

The linear structural model immediately implies a parametric structure for the moments and product–moments of the observed variables. The moments thus present a specific pattern and they can be estimated in reference to the parameters that characterized the latent variable model.

Estimating and testing the model usually involve the *moment structure* representations and normality. In practice, the literature on structural models has concentrated on the moment structure of only the first two product moments, specifically means and covariances or correlations. Nevertheless, it is entirely possible to generate structural models for higher–order moments (see Bentler, 1983). This neglect of higher–order moments almost surely has been aided by the historical dominance of the multivariate normal distribution assumption. Under such a assumption, the two lower–order moments are sufficient statistics and higher–order central moments are indeed zero or simple functions of the second–order moment. The specification of the covariance or correlation matrix of the observed variables as a function of the structure model parameters is known as *covariance* or *correlation structure*, respectively. Covariance or correlation structures, sometimes with associated mean structures, occur in psychology (Bentler, 1980), econometrics (Newey and McFadden, 1994), education (Bell et al., 1990), sociology (Huba, Wingard, and Bentler, 1981) among others.

Linear correlation is a natural dependence measure for multivariate normally and, more generally, elliptically distributed variables. Nevertheless, other dependence concepts like comonotonicity and rank correlation should also be understood by the practitioners. The fallacies of linear correlation arise from the *naive* assumption that dependence properties of the elliptical world also hold in the non–elliptical world. However, empirical researches in finance, psychology, education show that the distributions of the real world are seldom in this class.

Embrechts, McNeil, and Straumann (1999) highlight a number of important fallacies concerning correlation which arise when we work with non–normal models. Firstly, the linear correlation is a measure of linear dependence and it requires that the variances are finite. Secondly, linear correlation has the serious deficiency that it is not invariant by increasing transformations of the observed variables. As a simple illustration, we suppose to have a probability model for dependent insurance losses. If we decide that our interest now lies in modeling the logarithm of these losses, the value of correlation coefficients will change. Similarly, if we change from a model of percentage returns on several financial assets to a model of logarithmic returns, we will obtain a similar result.

Moreover, only in the case of the multivariate normal is it permissible to interpret uncorrelatedness as implying independence. This implication is no longer valid when the joint distribution function is non–normal. Spherical distributions model uncorrelated random variables but are not, except in the case of the multivariate normal, the distributions of independent random variables.

In socio–economic theory the notion of correlation anyway remains central, even though there is a general reject of normal assumption and, as a consequence, there are doubts about the usefulness of the linear–dependence measure. In this doctoral dissertation, we are interested in developing inferential methods for latent variable models (i.e., covariance or correlation structures) that combine second–order structure moments with less restrictive distribution assumptions than equality of marginal distributions, normality, and linearity. We want to assume flexible probability models for the latent variables that guarantee the presence of correlation–like dependence parameters. We show how to reach a no–moment–based correlation matrix, without a supposed linear or normal dependence, and to estimate and test the correlation structure with this unusual dependence measure. Our approach is based on copula functions, which can be useful in defining inferential methods on second–order structure models, as recently shown by Klüppelberg and Kuhn (2009). .

In our opinion the copula–based approach affords the best understanding of the general concept of dependence. From a practical point of view, copulas are attractive because of their flexibility in model specification. By Sklar's theorem (Sklar, 1959), the distribution function of each multivariate random variable can be indeed described through its margins and a suitable dependence structure represented by a copula, separately. Many multivariate models for dependence can be generated by parametric families of copulas, typically indexed by a real– or vector–valued parameter, named *copula parameter*. Examples of such systems are given in Joe (1997) and Nelsen (2006), among others. Hoeffding (1940, 1994) also had the basic idea of summarizing the dependence properties of a multivariate distribution by its corresponding copula, but he chose to define the corresponding function on $\left[-\frac{1}{2}, \frac{1}{2}\right]^p$ rather than on $[0, 1]^p$ (Sklar, 1959), where p stands for the number of the observed variables. Copulas are a less well known approach to describing dependence than correlation, but the dependence structure as summarized by a copula is invariant by increasing transformations of the variables.

Motivated by Klüppelberg and Kuhn (2009), which base their proposal on copulas of elliptical distributions, we extend their methodology to other families that can be profitably assumed in moment structure models. Firstly, we are aware that this research involves copulas, whose parameters must be interpreted as a correlation–like measure. Secondly, we note that a necessary condition here consists in handling multivariate distribution functions where each bivariate margin may be governed by an exclusive parameter. One difficulty with most families of multivariate copulas is the paucity of parameters (generally, only 1 or 2). Moreover, in the multivariate one (or two)–parameter case, *exchangeability* is a key assumption. As a consequence, all the bivariate margins are the same and the correlation structure is identical for all pairs of variables. On the contrary, for each bivariate margin an one–to–one analytic relation between its parameters and the corresponding bivariate Pearson's linear correlation coefficient has to exist for the moment structure analysis purpose. If it is, we are able to estimate in a consistent way the correlation structure model through copula parameters estimates, as an alternative to the moment–based estimation procedure used in the linear correlation approach.

In order to overwhelm useless sophistications, we suggest to adopt the Eyraud–Farlie–Gumbel–Morgenstern (shortly, EFGM) family of multivariate copulas, consisting of the copulas associated with the family of Eyraud–Farlie–Gumbel–Morgenstern distribution functions (Kotz, Balakrishnan, and Johnson, 2000, Equation 44.73). It is attractive due to its simplicity, and Prieger (2002) advocates its use as a proper model in health insurance plan analysis. EFGM copula is ideally suited for various models with small or moderate dependence and it represents an alternative to the copula proposed by Klüppelberg and Kuhn (2009). There are several examples in which it is essential to consider weak dependent structures instead of simple independence. It is in particular related to some of the most popular conditions used by econometricians to transcribe the notion of fading memory. Various generalizations of independence have been introduced to tackle to this problem. The *martingale* setting was the first extension of the independence framework (Hall and Heyde, 1980). Another point of view is given by the *mixing* properties of stationary sequences in the sense of ergodic theory (Doukhan, 1994). Nevertheless, in some situations classical tools of weak dependence such as mixing are useless. For instance, when bootstrap techniques are used, no mixing conditions can be expected. Weakening martingale conditions yields *mixingales* (Andrews, 1988; McLeish, 1975). A more general concept is the *near epoch dependence* (shortly, NED) on a mixing process. Its definition can be found in the work by Billingsley (1968), who considered functions of uniform mixing processes (Ibragimov, 1962).

Since our attention is focused on inferential methods for covariance or correlation structure models, we use different estimators for copula parameters and we test the consequent benefits to the asymptotic distribution of test statistic in correlation structure model selection.

By summarizing, in this doctoral dissertation we propose an extension of the approach provided by Klüppelberg and Kuhn (2009) for inference on second–order structure moments. As in Klüppelberg and Kuhn (2009) we adopt a copula–based approach instead of assuming normal distribution for the variables, thus relaxing the equality in distribution assumption. Then, we assume that the manifest variables have the same copula of the linear combination of latent variables.

We estimate and test the latent variable models through moment structure representations by assuming copula functions. Unlike the classical methods, we do not use a moment–based estimator of covariance or correlation matrix. We rather exploit the copula assumption and we obtain a second–order moment estimator based on the estimates of copula parameters. This procedure underlines the importance of copulas as a tool to capture general and not necessarily linear dependence structures between variables. Our contribution is here twofold. Firstly, we assume a non–elliptical copula for moderate dependence systems; i.e., the EFGM copula. We also provide a discussion about conditions for extending linear structure model to other families of copulas. Secondly, we propose an alternative estimator of copula parameters in correlation structure analysis; i.e., the maximum pseudo–likelihood estimator. We supply detailed computational explanation for inference on second–order moments, also valid for the methodology in Klüppelberg and Kuhn (2009). Finally, a comprehensive simulation study and an application to real financial data are performed. We will not deal with higher–order moments since our interest is here focused only on second–order moment structure models. Moreover, we only deal with the static (non–time–dependent) case. There are various other problems concerning the modeling and interpretation of serial correlation in stochastic processes and cross–correlation between processes; e.g., see Boyer, Gibson, and Loretan (1999).

The doctoral dissertation is organized as follows. We start with definitions and preliminary results on moment structure analysis in Chapter 2. In Chapter 3 we introduce the new copula structure model proposed by Klüppelberg and Kuhn (2009) and show which classical inferential methods can be used for structure analysis and model selection. In Section 3.3 we provide a detailed computational procedure for estimating and testing purposes.

In Chapter 4 we present our main contributions. In Section 4.1 we revise the properties of EFGM class and we show that the dependence properties of this family are closely related with linear correlation concept. By analogy with Klüppelberg and Kuhn (2009) we assume EFGM copulas for the observed variables and we investigate in a simulation study the performance of the estimator of the correlation structure in case of a well known latent variable model, the *exploratory factor analysis*.

Supported by the simulation studies carried out by Genest, Ghoudi, and Rivest (1995), Fermanian and Scaillet (2004), Tsukahara (2005), and recently by Kojadinovic and Yan (2010b), in Section 4.2 we propose to adopt the maximum pseudo–likelihood estimator for copula parameters (Genest, Ghoudi, and Rivest, 1995), instead of the estimator provided by Klüppelberg and Kuhn (2009).

In Section 4.3 we compare the sample distribution of test statistic via the maximum pseudo–likelihood estimator of copula parameters and the estimator provided by Klüppelberg and Kuhn (2009), respectively, with the corresponding asymptotic distribution by QQ–plots and kernel densities. Moreover, we investigate the influence of sample size and correct specification of copula on the performance of the above mentioned test statistics. Finally, we show our method at work on a financial dataset and explain differences between our copula–based and the classical normal–based approach.

Final remarks about the use of copulas in moment structure analysis and conditions in order to extend the methodology to a wider class of non–normal distributions are provided in last chapter.

Additional tools for moment structure analysis are provided in Appendices A and B.

Chapter 2

Inference on moment structure models

Structural models can be defined at various levels or orders of parametric complexity. In linear structural models, common practice involves specification of a structural representation for the random vector of observable variables $\boldsymbol{x} \in \mathbb{R}^p$; i.e.,

$$\boldsymbol{x} \stackrel{d}{=} \boldsymbol{A}\left(\boldsymbol{\vartheta}_{\mathbf{0}}\right)\boldsymbol{\zeta}\,, \tag{2.1}$$

where $\boldsymbol{A}\left(\boldsymbol{\vartheta}_0\right)$ is a matrix–valued function with respect to a vector of population parameters $\boldsymbol{\vartheta}_0$. We standardly use $\stackrel{d}{=}$ to denote equality in distribution. The underlying generating random variables $\boldsymbol{\zeta} \in \mathbb{R}^z$, for $z \geq p$, may represent latent (or unobservable) variables and errors of measurement. General introduction to the field as well as more advanced treatments can be found in Jöreskog (1978) and Browne (1982). Discussions on key developments of these topics are provided by Steiger (1994) and Bentler and Dudgeon (1996).

Examples of such models include path analysis (Wright, 1918, 1934), principal component analysis (Hotelling, 1933; Pearson, 1901), exploratory and confirmatory factor analysis (Spearman, 1904, 1926), simultaneous equations (Anderson, 1976; Haavelmo, 1944), errors–in–variables models (Dolby and Freeman, 1975; Gleser, 1981), and especially the generalized linear structural equations models (Jöreskog, 1973, 1977) made popular in the social and behavioral sciences by such computer programs as `LISREL` (Jöreskog and Sörbom, 1983) and `EQS` (Bentler and Wu, 1995a,b).

Statistical methods for structural models are concerned with estimating the parameters of model (2.1) in asymptotically efficient ways, as well as with testing the goodness–of–fit of (2.1). That is, the parameters of the model can be estimated, and the model null hypothesis tested, without using the $\boldsymbol{\zeta}$ variables by relying on sample estimators as $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ of the population mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$ of the variables $\boldsymbol{x}$, respectively. In fact, any linear structural model implies a more basic set of parameters $\boldsymbol{\theta}_0 = (\theta_{0,1}, \ldots, \theta_{0,q})$, so that $\boldsymbol{\mu}_0 = \boldsymbol{\mu}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. The q parameters in $\boldsymbol{\theta}_0$ represent elements of $\boldsymbol{\vartheta}_0$, like mean vectors, loadings, variances and covariances or correlations of the variables $\boldsymbol{\zeta}$. Here the representation as well as the estimation and testing in model (2.1) will be restricted to a small subset of structural models, namely, those that involve continuous observable and unobservable variables whose essential characteristics can be investigated via covariance or correlation matrices.

In general, inference on covariance or correlation structure models is a straightforward matter when the model is linear and the latent variables, and hence the observed variables, are presumed to be multivariate normally distributed. Since the only unknown parameters for a multivariate normal distribution are elements of mean vectors and covariance matrices, linear structural model generates structures for population mean vectors and covariance matrices alone. Normal–theory–based methods such as maximum likelihood (Jöreskog, 1967; Lawley and Maxwell, 1971) and generalized least squares (Browne, 1974; Jöreskog and Goldberger, 1972) are frequently applied. The sample mean vector and covariance matrix are jointly sufficient statistics, and maximum likelihood estimation reduces to fitting structural models to sample mean vectors and covariance matrices. Nevertheless, most social, behavioral, and economic data are seldom normal, so normal–based methods can yield very distorted results. For example, in one distribution condition of a simulation with a confirmatory factor analysis model, Hu, Bentler, and Kano (1992) find that the likelihood ratio test based on normal–theory maximum likelihood estimator rejected the true model in 1194 out of 1200 samples at sample sizes that ranged from $n = 150$ to $n = 5000$. Possible deviations of the distribution function from normality have led researchers to develop asymptotically distribution free (shortly, ADF) estimation methods for covariance structures in which $\boldsymbol{\mu}_0$ is unstructured using the minimum modified chi–squared principle by Ferguson (1958) (Browne, 1982, 1984; Chamberlain, 1982). Although the ADF method attains reasonable asymptotically good performance on sets of few variables, in large systems with small– to medium–sized samples it can be extremely misleading; i.e., it leads to inaccurate decisions regarding the adequacy of models (Hu, Bentler, and Kano, 1992). A computationally intensive improvement on ADF statistics has been made (Yung and Bentler, 1994), but the ADF theory remains inadequate to evaluate covariance structure models in such situations (Bentler and Dudgeon, 1996; Steiger, 1994).

Increasingly relaxing the normal assumption of classical moment structure analysis, one assumption still remains, namely $\boldsymbol{x} \in \mathbb{R}^p$ can be described as a linear combination of some (latent) random variables $\boldsymbol{\zeta}$ with existing second moments (and existing fourth moments to ensure asymptotic distributional limits of sample covariance estimators). A wider class of distributions including the multivariate normal distribution but also containing platykurtic and leptokurtic distributions is the elliptical one. Consequently the assumption of a distribution from the elliptical class is substantially less restrictive than the usual assumption of multivariate normality. Browne (1982, 1984) introduces elliptical distribution theory for covariance structure analysis. Under the assumption of a distribution belonging to the elliptical class, a correction for kurtosis of normal–theory–based methods for the estimators of covariance matrix and test statistics is provided. Nevertheless, as Kano, Berkane, and Bentler (1990) point out, most empirical data have heterogeneous values of marginal kurtosis, whereas elliptical distributions require homogeneous ones. Therefore, the results based on elliptical theory may not be robust to violation of ellipticity. Starting from the elliptical class, Kano, Berkane, and Bentler (1990) discuss the analysis of covariance structures in a wider class of distributions whose marginal distributions may have heterogeneous kurtosis parameters. An attractive feature of the heterogeneous kurtosis (shortly, HK) method is that the fourth–order moments of $\boldsymbol{x}$ do not need to be computed as Browne (1984) does, because these moments are just a function of the variances and covariances between variables $\boldsymbol{x}$ and of the kurtosis parameters. As a result, HK method can be used on models that are based on a substantially large number of observed variables. Unfortunately, Kano, Berkane, and Bentler (1990, Section 4) do not give necessary and sufficient conditions in order to verify the existence of elliptical distributions with distinct marginal kurtosis coefficients and provide just a simple example in two dimensions.

Finally, in order to completely relax the equality in distribution assumption and manage flexible probability models one possible choice may be represented by copulas. Nevertheless, before reviewing their use in moment structure analysis, in Section 2.1 we start with classical theoretical backgrounds concerning the estimation of $\boldsymbol{\theta}_0$ in model (2.1) by weighted least squares. The asymptotic distribution of the estimator is discussed in the ADF context and it is also considered under a more general elliptical distribution assumption. In Section 2.2 we present an example of structural model; i.e., the *factor analysis model.* We also talk about problems of identification and estimation when $\boldsymbol{x}$ are assumed to be multivariate normally distributed.

2.1 Weighted least squares estimates in the analysis of covariance structures

Let $\boldsymbol{X}$ represent a $(n+1) \times p$ data matrix whose rows are drawn by a random vector of independent and identically distributed variables with population mean $\boldsymbol{\mu}_0$ and population covariance matrix $\boldsymbol{\Sigma}_0$. A covariance structure is a model where the elements of $\boldsymbol{\Sigma}_0$ are regarded as functions of a q–dimensional parameter $\boldsymbol{\theta}_0 \in \Theta \subseteq \mathbb{R}^q$. Thus $\boldsymbol{\Sigma}_0$ is a matrix–valued function with respect to $\boldsymbol{\theta}_0$. The model is said to hold if there exists a $\boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$.

Let $\hat{\boldsymbol{\Sigma}}$, the sample covariance matrix based on a sample of size $n+1$, be an unbiased estimator of $\boldsymbol{\Sigma}_0$ and consider a discrepancy function $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$

which gives an indication of discrepancy between $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ (Browne, 1982). This scalar valued function has the following properties:

(P.1) $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} \geq 0$;

(P.2) $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} = 0$ if and only if $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$

(P.3) $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$ is a twice continuously differentiable function of $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

A discrepancy function $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$ does not need to be symmetric in $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, that is $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$ does not need to be equal to $D\left\{\boldsymbol{\Sigma}(\boldsymbol{\theta}), \hat{\boldsymbol{\Sigma}}\right\}$.

If the estimate of $\boldsymbol{\theta}_0$ is obtained by minimizing some discrepancy function $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$, then

$$D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)\right\} = \min_{\boldsymbol{\theta}\in\Theta} D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} .$$

The reproduced covariance matrix will be denoted by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)$. Therefore, an estimator $\hat{\boldsymbol{\theta}}$, taken to minimize $D\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$, is called a *minimum discrepancy function estimator*. We call $n\,D\left(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\right)$ the *associated minimum discrepancy function test statistic*. Since $\boldsymbol{\Sigma}_0$ is supposed to be equal to $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ according to (2.1), we shall regard $\boldsymbol{\theta}_0$ as the value of $\boldsymbol{\theta}$ which minimizes $D\{\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta})\}$; i.e.,

$$\min_{\boldsymbol{\theta}\in\Theta} D\{\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta})\} = D\{\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\} .$$

The asymptotic distribution of the estimator $\hat{\boldsymbol{\theta}}$ will depend on the particular discrepancy function minimized. Examples of discrepancy functions are the likelihood function under the normality assumption for $\boldsymbol{x}$,

$$D_L\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \log\left|\hat{\boldsymbol{\Sigma}}\right| + tr\left[\hat{\boldsymbol{\Sigma}}\{\boldsymbol{\Sigma}(\boldsymbol{\theta})\}^{-1}\right] - p\,, \qquad (2.2)$$

which leads to the maximum likelihood estimator (Jöreskog, 1967; Lawley and Maxwell, 1971), and the quadratic (or weighted least squares) discrepancy function

$$D_{QD}\left\{\hat{\Sigma}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} = \{\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}^{\top} \boldsymbol{W}^{-1} \{\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}\ , \tag{2.3}$$

where $\boldsymbol{\sigma}(\boldsymbol{\theta}) = vech\{\boldsymbol{\Sigma}(\boldsymbol{\theta})\}$, $\hat{\boldsymbol{\sigma}} = vech\left(\hat{\boldsymbol{\Sigma}}\right)$, and $\boldsymbol{W}$ is a $p^{\star} \times p^{\star}$ weight matrix converging in probability to some positive definite matrix $\boldsymbol{W}_0$ as $n \to \infty$, with $p^{\star} = p(p+1)/2$ (Browne, 1982, 1984). See Appendix A for a definition of *vec* and *vech* operators. Typically $\boldsymbol{W}$ is considered to be a fixed, possible estimated, positive definite matrix, although the theory can be extended to random weight matrices (Bentler and Dijkstra, 1983). If $\boldsymbol{W}$ is represented by $2\,\boldsymbol{G}_p^{\top}(\boldsymbol{V} \otimes \boldsymbol{V})\,\boldsymbol{G}_p$, where $\boldsymbol{V}$ is a $p \times p$ positive definite stochastic matrix which converges in probability to a positive definite matrix $\boldsymbol{V}_0$ as $n \to \infty$ and $\boldsymbol{G}_p$ represents the transition or duplication matrix (see Appendix A for a formal definition), then the function in (2.3) is reduced to

$$D_{GLS}\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} = \frac{1}{2}\,tr\left[\left\{\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}\boldsymbol{V}^{-1}\right]^2 , \tag{2.4}$$

which is the normal–theory–based generalized least squares discrepancy function (Browne, 1974; Jöreskog and Goldberger, 1972). One possible choice for $\boldsymbol{V}$ is $\boldsymbol{V} = \hat{\boldsymbol{\Sigma}}$, so that $\boldsymbol{V}_0 = \boldsymbol{\Sigma}_0$. An other possible choice for $\boldsymbol{V}$ is $\boldsymbol{V} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}_{ML}\right)$, where $\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}_{ML}\right)$ is the estimator which maximizes the Wishart likelihood function for $\hat{\boldsymbol{\Sigma}}$ when $\boldsymbol{x}$ has a multivariate normal distribution (Browne, 1974).

The following usual regularity assumptions are imposed to guarantee suitable asymptotic properties of the estimators $\hat{\boldsymbol{\theta}}$ via quadratic discrepancy function and the associated test statistics (Browne, 1984).

(A.0) As $n \to \infty$, $n^{1/2}\{\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}(\boldsymbol{\theta}_0)\}$ converges in law to a multivariate normal with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}$, a $p^{\star} \times p^{\star}$ positive definite matrix.

Remark. When $\boldsymbol{x}$ is normally distributed with covariance matrix $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}$ is represented in the form $\boldsymbol{\Sigma}_{\boldsymbol{\sigma}} = 2\,\boldsymbol{G}_p^{\top}(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\,\boldsymbol{G}_p$.

(A.1) $\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}$ is positive definite.

Remark. If $\boldsymbol{x}$ is normally distributed, **(A.1)** is equivalent to the condition that $\boldsymbol{\Sigma}_0$ be positive definite.

(A.2) $D\left\{\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$ has an unique minimum on Θ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; i.e., $\boldsymbol{\Sigma}(\boldsymbol{\theta}^\star) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, $\boldsymbol{\theta}^\star \in \Theta$, implies that $\boldsymbol{\theta}^\star = \boldsymbol{\theta}_0$.

(A.3) $\boldsymbol{\theta}_0$ is an interior point of the parameter space Θ.

(A.4) The $p^\star \times q$ Jacobian matrix $\mathbf{J}_{\boldsymbol{\theta}_0} = \mathbf{J}(\boldsymbol{\theta}_0) := \left[\partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^\top\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is of full rank q.

(A.5) $\|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\|$ is $O\left(n^{-1/2}\right)$.

Remark. This condition assumes that systematic errors due to lack of fit of the model to the population covariance matrix are not large relative to random sampling errors in $\hat{\boldsymbol{\Sigma}}$. Clearly **(A.5)** is always satisfied if the structural model hold; i.e., $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$.

(A.6) The parameter set Θ is closed and bounded.

(A.7) $\mathbf{J}_{\boldsymbol{\theta}}$ and, consequently, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are continuous function of $\boldsymbol{\theta}$.

Under the assumptions **(A.0)**–**(A.7)**, Browne (1984, Corollary 2.1) and Chamberlain (1982) showed that the estimator $\hat{\boldsymbol{\theta}}$ associated with the discrepancy function (2.3) is consistent and asymptotically normal and that the Cramèr–Rao lower bound of the asymptotic covariance matrix is

$$\left(\mathbf{J}_{\boldsymbol{\theta}_0}^\top \boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-1} \mathbf{J}_{\boldsymbol{\theta}_0}\right)^{-1}, \tag{2.5}$$

attained when $\boldsymbol{W} = \boldsymbol{\Sigma}_{\boldsymbol{\sigma}}$. An estimator is said to be asymptotically efficient within the class of all minimum discrepancy function estimators if its asymptotic covariance matrix is equal to (2.5). In this case, the associated minimum discrepancy function test statistic, $n\, D_{QD}\left(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\right)$, was shown to be asymptotically chi–squared with $p^\star - q$ degrees of freedom (Browne, 1984, Corollary 4.1).

Inference based on the discrepancy function (2.3) by excluding assumption **(A.0)** is called the *asymptotically distribution–free method.* However, weighted least squares estimation can easily become distribution specific. This is accomplished by specializing the optimal weight matrix $\boldsymbol{W}$ into the form that it must have if the variables have a specified distribution. In other words, the ADF method is a weighted least squares procedure in which the weight matrix has to be properly specified in order to guarantee that the asymptotic properties of standard normal theory estimators and test statistics are obtained. Asymptotically this method has good properties, however one needs a very large sample for the asymptotics to be appropriate (Hu, Bentler, and Kano, 1992), and sometimes it could be computationally difficult to invert the $p^\star \times p^\star$ weight matrix $\boldsymbol{W}$ for moderate values of p.

When a p–variate random vector $\boldsymbol{x}$ is elliptical distributed, the weighted least squares method can easily specialized to ellipticity. In this case, $\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}$ can be represented as

$$\boldsymbol{\Sigma}_{\boldsymbol{\sigma}} = 2\,\eta\,\boldsymbol{G}_p^\top\left(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0\right)\boldsymbol{G}_p + \boldsymbol{G}_p^\top \boldsymbol{\sigma}_0\left(\eta - 1\right)\boldsymbol{\sigma}_0^\top \boldsymbol{G}_p\,,$$

where $\boldsymbol{\sigma}_0 = vec\left(\boldsymbol{\Sigma}_0\right)$ and $\eta = E\left\{\left(\boldsymbol{x} - \boldsymbol{\mu}_0\right)^\top\,\boldsymbol{\Sigma}_0^{-1}\,\left(\boldsymbol{x} - \boldsymbol{\mu}_0\right)\right\}^2 / \left\{p\left(p+2\right)\right\}$ is the relative Mardia (1970)'s multivariate kurtosis parameter of $\boldsymbol{x}$.

Browne (1984, Section 4) proposed a rescaled test statistic

$$\hat{\eta}^{-1}\,n\,D_{QD}\left(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\right)\,, \tag{2.6}$$

where

$$\hat{\eta} = \frac{n+2}{n\left(n+1\right)}\sum_{a=1}^{n+1}\left\{\left(\boldsymbol{x}_a - \hat{\boldsymbol{\mu}}\right)^\top\,\hat{\boldsymbol{\Sigma}}^{-1}\,\left(\boldsymbol{x}_a - \hat{\boldsymbol{\mu}}\right)\right\}^2 / \left\{p\left(p+2\right)\right\}\,, \quad \boldsymbol{x}_a \in \mathbb{R}^p\,.$$

Test statistic (2.6) is asymptotically chi–squared with $p^\star - q$ degrees of freedom if the structural model for covariance matrix is invariant under a constant scaling factor. This condition is satisfied if, given any $\boldsymbol{\theta} \in \Theta$ and any positive scalar c^2, there exists another parameter $\boldsymbol{\theta}^\star \in \Theta$ such that $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}^\star\right) = c^2\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$ (Browne, 1982, 1984). An important consequence of this adaptation is that the normal–theory–based weighted least squares method is robust against non–normality among elliptical distributions after a correction for kurtosis.

2.2 Factor analysis models

Originally developed by Spearman (1904) for the case of one common factor, and later generalized by Thurstone (1947) and others to the case of multiple factors, factor analysis is probably the most frequently used psychometric procedure. The analysis of moment structures originated with the factor analysis model and with some simple pattern hypothesis concerning equality of elements of mean vectors and covariance matrices. Most models involving covariance structures that are in current use are related with factor analysis in some way, either by being special cases with restrictions on parameters or, more commonly, extensions incorporating additional assumptions; see, e.g., the generalized linear structural equations models (Jöreskog, 1973, 1977).

The aim of factor analysis is to account for the covariances of the observed variates in terms of a much smaller number of hypothetical variates or factors. The question is: if there is correlation, is there a random variate ϕ_1 such that all partial correlations coefficients between variables in $\boldsymbol{x}$ after eliminating the effect of ϕ_1 are zero? If not, are there two random variates ϕ_1 and ϕ_2 such that all partial correlation coefficients between variables in $\boldsymbol{x}$ after eliminating the effects of ϕ_1 and ϕ_2 are zero? The process continues until all partial correlation coefficients between variables in $\boldsymbol{x}$ are zero. Therefore, the factor analysis model partitions the covariance or correlation matrix into that which is due to common factors, and that which is unique.

To introduce the factor analysis model, let $\boldsymbol{A}(\boldsymbol{\vartheta_0}) = \{diag(\boldsymbol{\mu}), \boldsymbol{\Lambda}, \boldsymbol{I}_p\}$ and $\boldsymbol{\zeta} = \left(\mathbf{1}_p^\top, \boldsymbol{\phi}^\top, \boldsymbol{\upsilon}^\top\right)^\top$ in (2.1), where $\boldsymbol{I}_p$ stands for the identity matrix of order p and $\mathbf{1}_p$ denotes the p–dimensional vector whose elements are all equal to 1. The linear latent variable structure becomes

$$\boldsymbol{x} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\phi} + \boldsymbol{\upsilon}\,, \tag{2.7}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location parameter, $\boldsymbol{\phi} \in \Phi \subseteq \mathbb{R}^m$ for $m \ll p$ is a vector of non–observable and uncorrelated factors and $\boldsymbol{\upsilon} \in \Upsilon \subseteq \mathbb{R}^p$ is a vector of noise variables υ_j representing sources of variation affecting only the variate x_j. Without loss of generality, we suppose that the means of all variates are zero; i.e., $E(\boldsymbol{x}) = \mathbf{0}$, $E(\boldsymbol{\phi}) = \mathbf{0}$, and $E(\boldsymbol{\upsilon}) = \mathbf{0}$. In the case of uncorrelated factors and of rescaled variances to unit, $E\left(\boldsymbol{\phi}\boldsymbol{\phi}^\top\right) = \boldsymbol{I}_m$. The coefficient $\lambda_{j,k}$ for $k = 1, \ldots, m$ is known as the *loading* of x_j on ϕ_m or, alternatively, as the loading of ϕ_m in x_j. The p random variates υ_j are assumed to be uncorrelated between each others and the m factors; i.e., $E\left(\boldsymbol{\upsilon}\boldsymbol{\upsilon}^\top\right) = \boldsymbol{\Psi} = diag\left(\psi_1, \ldots, \psi_p\right)$ and $E\left(\boldsymbol{\phi}\boldsymbol{\upsilon}^\top\right) = \mathbf{0}$. The variance of υ_j is termed *residual variance* or *unique variance* of x_j and denoted by ψ_j. Then, describing the dependence structure of $\boldsymbol{x}$ through its covariance matrix yields the covariance structure,

$$var\left(\boldsymbol{x}\right) = \boldsymbol{\Sigma}_0 = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}\,, \tag{2.8}$$

namely, the dependence of $\boldsymbol{x}$ is described through the entries of $\boldsymbol{\Lambda}$.

Thus (2.7) corresponds to (2.1) and the parameter vector $\boldsymbol{\theta}_0$ consists of $q = pm + p$ elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$.

2.2.1 Uniqueness of the parameters

Given a sample covariance matrix $\hat{\boldsymbol{\Sigma}}$, we want to obtain an estimator of the parameter vector $\boldsymbol{\theta}_0$. First of all, we ask whether for a specified value of m, less than p, it is possible to define a unique $\boldsymbol{\Psi}$ with positive diagonal elements and a unique $\boldsymbol{\Lambda}$ satisfying (2.8). Since only arbitrary constraints will be imposed upon the parameters to define them uniquely, the model will be termed *unrestricted.*

Let us first suppose that there is a unique $\boldsymbol{\Psi}$. The matrix $\boldsymbol{\Sigma}_0 - \boldsymbol{\Psi}$ must be of rank m: this quantity is equal to the covariance matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ in which each diagonal element represents not the total variance of the corresponding variate in $\boldsymbol{x}$ but only the part due to the m common factors. This is known as *communality* of the variate.

If $m = 1$, then $\boldsymbol{\Lambda}$ reduces to a column vector of p elements. It is unique, apart from a possible change of sign of all its elements, which corresponds merely to changing the sign of the factor.

For $m > 1$ there is an infinity of choices for $\boldsymbol{\Lambda}$. (2.7) and (2.8) are still satisfied if we replace $\boldsymbol{\phi}$ by $\boldsymbol{M\phi}$ and $\boldsymbol{\Lambda}$ by $\boldsymbol{\Lambda M}^\top$, where $\boldsymbol{M}$ is any orthogonal matrix of order m. In the terminology of factor analysis this corresponds to a factor rotation.

Suppose that each variate is rescaled in such a way that its residual variance is unity. Then

$$\boldsymbol{\Sigma}_0^\star = \boldsymbol{\Psi}^{-1/2}\,\boldsymbol{\Sigma}_0\,\boldsymbol{\Psi}^{-1/2} = \boldsymbol{\Psi}^{-1/2}\,\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top\,\boldsymbol{\Psi}^{-1/2} + \boldsymbol{I}_p$$

and

$$\boldsymbol{\Sigma}_0^\star - \boldsymbol{I}_p = \boldsymbol{\Psi}^{-1/2}\,\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top\,\boldsymbol{\Psi}^{-1/2} = \boldsymbol{\Psi}^{-1/2}\,\left(\boldsymbol{\Sigma}_0 - \boldsymbol{\Psi}\right)\,\boldsymbol{\Psi}^{-1/2}\,.$$

The matrix $\boldsymbol{\Sigma}_0^\star - \boldsymbol{I}_p$ is symmetric and of rank m and it may be expressed in the form $\boldsymbol{\Omega}\,\boldsymbol{\Xi}\,\boldsymbol{\Omega}^\top$, where $\boldsymbol{\Xi}$ is a diagonal matrix of order m, where the elements are the m non zero eigenvalues of $\boldsymbol{\Sigma}_0^\star - \boldsymbol{I}_p$, and $\boldsymbol{\Omega}$ is a $p \times m$ matrix satisfying $\boldsymbol{\Omega}\boldsymbol{\Omega}^\top = \boldsymbol{I}_p$, where the columns are the corresponding eigenvectors. Note that $\boldsymbol{\Sigma}_0^\star$ has the same eigenvectors as $\boldsymbol{\Sigma}_0^\star - \boldsymbol{I}_p$, and that its p eigenvalues are those of $\boldsymbol{\Sigma}_0^\star - \boldsymbol{I}_p$ increased by unit.

We may define $\boldsymbol{\Lambda}$ uniquely as

$$\boldsymbol{\Lambda} = \boldsymbol{\Psi}^{1/2}\,\boldsymbol{\Omega}\,\boldsymbol{\Xi}^{1/2}\,. \tag{2.9}$$

Since $\boldsymbol{\Psi}^{-1/2}\,\boldsymbol{\Lambda} = \boldsymbol{\Psi}^{-1/2}\,\boldsymbol{\Psi}^{1/2}\,\boldsymbol{\Omega}\,\boldsymbol{\Xi}^{1/2} = \boldsymbol{\Omega}\,\boldsymbol{\Xi}^{1/2}$,

$$\left(\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Lambda}\right)^\top\,\boldsymbol{\Psi}^{-1/2}\,\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^\top\,\boldsymbol{\Psi}^{-1}\,\boldsymbol{\Lambda} = \boldsymbol{\Xi}^{1/2}\,\boldsymbol{\Omega}^\top\,\boldsymbol{\Omega}\,\boldsymbol{\Xi}^{1/2} = \boldsymbol{\Xi}\,.$$

Thus, from (2.9), we have chosen $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Lambda}^\top\,\boldsymbol{\Psi}^{-1}\,\boldsymbol{\Lambda}$ is a diagonal matrix whose positive and distinct elements are arranged in descending order of magnitude. Then $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are uniquely determined.

For $m > 1$, the fact that $\boldsymbol{\Lambda}^\top\,\boldsymbol{\Psi}^{-1}\,\boldsymbol{\Lambda}$ should be diagonal has the effect of imposing $m\,(m-1)\,/2$ constraints upon the parameters. Hence the number of free (unknown) parameters in $\boldsymbol{\theta}_0$ becomes

$$pm + p - \frac{1}{2}m\,(m-1) = q - \frac{1}{2}m\,(m-1)\ .$$

If we equate corresponding elements of the matrices on both sides of (2.8), we obtain $p^\star$ distinct equations. The degrees of freedom of the model are

$$p^\star - q + \frac{1}{2}m\,(m-1) = \frac{1}{2}\left\{(p-m)^2 - (p+m)\right\}\ .$$

If the result of subtracting from $p^\star$ the number of free parameters is equal to zero, we have as many equations as free parameters, so that $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are uniquely determined. If it is less than zero, there are fewer equations than free parameters, so that we have an infinity of choices for $\mathbf{\Lambda}$ and $\mathbf{\Psi}$. Finally, if it is grater than zero, we have more equations than free parameters and the solutions are not trivial.

2.2.2 Factor Analysis by generalized least squares

We suppose that there is a unique $\mathbf{\Psi}$, with positive diagonal elements, and a unique $\mathbf{\Lambda}$ such that $\mathbf{\Lambda}^\top\,\mathbf{\Psi}^{-1}\,\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are positive, distinct and arranged in decreasing order of magnitude.

Following Jöreskog and Goldberger (1972), we assume that $\boldsymbol{x}$ is multivariate normal distributed, that is $\hat{\mathbf{\Sigma}}$ has the Wishart distribution with expectation $\mathbf{\Sigma}_0$ and covariance matrix $2\,n^{-1}\,(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0)$. Therefore, a straightforward application of generalized least squares principle would choose parameter estimates to minimize the quantity (2.4). Using the estimate $\hat{\mathbf{\Sigma}}$ in place of $\boldsymbol{V}$ in (2.4) gives

$$D_{GLS}\left\{\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}\,(\boldsymbol{\theta})\right\} = \frac{1}{2}\,tr\left\{\boldsymbol{I}_p - \mathbf{\Sigma}\,(\boldsymbol{\theta})\,\hat{\mathbf{\Sigma}}^{-1}\right\}^2 = \frac{1}{2}\,tr\left\{\boldsymbol{I}_p - \hat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\,(\boldsymbol{\theta})\right\}^2, \tag{2.10}$$

which is the criterion to be minimized in the generalized least squares procedure. It is also possible to show that the maximum likelihood criterion (2.2) can be viewed as an approximation of (2.10) under the normal distribution assumption for $\boldsymbol{x}$.

Equation (2.10) is now regarded as a function of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ and it has to be minimized with respect to these matrices. The minimization is done in two steps. We first find the conditional minimum of (2.10) for a given $\mathbf{\Psi}$ and then we find the overall minimum.

To begin we shall assume that $\mathbf{\Psi}$ is nonsingular. We set equal to zero the partial derivative of (2.10) with respect to $\mathbf{\Lambda}$ and premultiplying by $\hat{\mathbf{\Sigma}}$ we obtain

$$\left(\mathbf{\Psi}^{1/2}\,\hat{\mathbf{\Sigma}}^{-1}\,\mathbf{\Psi}^{1/2}\right)\,\mathbf{\Psi}^{-1/2}\,\mathbf{\Lambda} = \mathbf{\Psi}^{-1/2}\,\mathbf{\Lambda}\,\left(\boldsymbol{I}_m + \mathbf{\Lambda}^\top\,\mathbf{\Psi}^{-1}\,\mathbf{\Lambda}\right)^{-1}\,, \qquad (2.11)$$

where we use the (ordinary) inverse of matrices of the form $\mathbf{\Psi}+\mathbf{\Lambda}\boldsymbol{I}_m\mathbf{\Lambda}^\top$ for $\mathbf{\Sigma}\,(\boldsymbol{\theta})$. The matrix $\mathbf{\Lambda}^\top\,\mathbf{\Psi}^{-1}\,\mathbf{\Lambda}$ may be assumed to be diagonal. The columns of the matrix on the right side of (2.11) then become proportional to the columns of $\mathbf{\Psi}^{-1/2}\,\mathbf{\Lambda}$. Thus the columns of $\mathbf{\Psi}^{-1/2}\,\mathbf{\Lambda}$ are characteristic vectors of $\mathbf{\Psi}^{1/2}\,\hat{\mathbf{\Sigma}}^{-1}\,\mathbf{\Psi}^{1/2}$ and the diagonal elements of $\left(\boldsymbol{I}_m + \mathbf{\Lambda}^\top\,\mathbf{\Psi}^{-1}\,\mathbf{\Lambda}\right)^{-1}$ are the corresponding roots. Let $\xi_1 \leq \ldots \leq \xi_p$ be the characteristic roots of $\mathbf{\Psi}^{1/2}\,\hat{\mathbf{\Sigma}}^{-1}\,\mathbf{\Psi}^{1/2}$ and let $\omega_1 \leq \ldots \leq \omega_p$ be an orthonormal set of corresponding characteristic vectors. Let $\Xi = diag\,(\xi_1,\ldots,\xi_p)$ be partitioned as $\Xi = diag\,(\Xi_1,\Xi_2)$, where $\Xi_1 = diag\,(\xi_1,\ldots,\xi_m)$ and $\Xi_2 = diag\,(\xi_{m+1},\ldots,\xi_p)$. Let $\mathbf{\Omega} = [\omega_1\,\ldots\,\omega_p]$ be partitioned as $\mathbf{\Omega} = [\mathbf{\Omega}_1\,\mathbf{\Omega}_2]$, where $\mathbf{\Omega}_1$ consists of the first m vectors and $\mathbf{\Omega}_2$ of the last $p-m$ vectors. Then $\mathbf{\Psi}^{1/2}\,\hat{\mathbf{\Sigma}}^{-1}\,\mathbf{\Psi}^{1/2} = \mathbf{\Omega}_1\Xi_1\mathbf{\Omega}_1^\top + \mathbf{\Omega}_2\Xi_2\mathbf{\Omega}_2^\top$ and the conditional solution $\hat{\mathbf{\Lambda}}$ is given by

$$\hat{\mathbf{\Lambda}} = \mathbf{\Psi}^{1/2}\,\mathbf{\Omega}_1\,\left(\Xi_1 - \boldsymbol{I}_m\right)^{1/2}\,. \qquad (2.12)$$

Defining $\tilde{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}^\top + \mathbf{\Psi}$, it can be verified that $\mathbf{\Psi}^{-1/2}\,\tilde{\mathbf{\Sigma}}\,\mathbf{\Psi}^{-1/2} = \mathbf{\Omega}_1\,\Xi_1\,\mathbf{\Omega}_1^\top + \mathbf{\Omega}_2\,\mathbf{\Omega}_2^\top$ and $\boldsymbol{I}_p - \hat{\mathbf{\Sigma}}^{-1}\,\tilde{\mathbf{\Sigma}} = \mathbf{\Psi}^{-1/2}\left\{\mathbf{\Omega}_2\left(\boldsymbol{I}_{p-m} - \Xi_2\right)\mathbf{\Omega}_2^\top\right\}\mathbf{\Psi}^{1/2}$ so that

$$tr\left(\boldsymbol{I}_p - \hat{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Sigma}}\right)^2 = tr\left(\boldsymbol{I}_{p-m} - \Xi_2\right)^2 = \sum_{j=m+1}^{p}\left(\xi_j - 1\right)^2\,.$$

Therefore the conditional minimum of (2.10), with respect $\mathbf{\Lambda}$ for a given $\mathbf{\Psi}$ is the function defined by

$$D_{GLS}\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} = \frac{1}{2}\sum_{j=m+1}^{p}(\xi_j - 1)^2 \ . \tag{2.13}$$

Any other set of roots will give a larger $D_{GLS}\left\{\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right\}$.

To start the two–steps procedure we require an initial estimate for $\boldsymbol{\Psi}$. We could take $\boldsymbol{\Psi}^{(0)} = \boldsymbol{I}_p$. A better choice for $\boldsymbol{\Psi}^{(0)}$ is however given by

$$\hat{\psi}_{i,i}^{(0)} = \left(1 - \frac{1}{2}m/p\right)\left(1/\hat{\sigma}^{i,i}\right) \ , \quad i = 1, \ldots, p \, , \tag{2.14}$$

where $\hat{\sigma}^{i,j}$ denotes the elements in the i–th row and j–th column of $\hat{\boldsymbol{\Sigma}}^{-1}$. This choice has been justified by Jöreskog (1963) and appears to work reasonably well in practice.

Chapter 3

Analysis of correlation structures: the Copula Structure Analysis

The theory for structural model analysis has been mostly developed for covariance matrices. This contrasts with common practice in which correlations are most often emphasized in data analysis. Correlation structures are of primary interest in situations when the different variables under consideration have arbitrary scales. Applying a covariance structure model to a correlation matrix will produce different test statistics, unbiased standard errors or parameter estimates and may alter the model being studied, unless the model under examination is appropriate for scale changes. The reason for this problem is not difficult to understand. If a correlation matrix is input, the elements on the main diagonal are no longer random variables: they are always equal to 1. Clearly, then, when a covariance matrix is replaced by a correlation matrix, a random vector containing $p^\star = p\,(p+1)\,/2$ random variables is replaced by a random vector with only $p^{\star\star} = p\,(p-1)\,/2$ elements free to vary.

Scale invariance is an essential property in order to apply covariance structure models to correlation matrices, but it is only minimally restrictive. The covariance structure $\boldsymbol{\Sigma}\,(\boldsymbol{\theta})$ is said to be invariant under a constant scaling factor if for any positive scalar c^2 and $\boldsymbol{\theta} \in \Theta$, there exists $\boldsymbol{\theta}^\star \in \Theta$ such that $c^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}^\star)$. The covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is said to be fully scale invariant if for any positive definite diagonal matrix $\boldsymbol{C}$ and any $\boldsymbol{\theta} \in \Theta$, there exists $\boldsymbol{\theta}^\star \in \Theta$ such that $\boldsymbol{C}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{C} = \boldsymbol{\Sigma}(\boldsymbol{\theta}^\star)$ (Browne, 1982). For instance, exploratory factor analysis and most of confirmatory factor analysis, `LISREL`, and `EQS` models satisfy this latter assumption. $\boldsymbol{C}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{C}$ means a change of units of measurement, therefore, if some model in multivariate analysis does not satisfy the scale invariance assumption, the model will depend on the units of measurement, which is usually unreasonable. An example of a model which is not fully scale invariant is the confirmatory factor analysis model with some factor loadings fixed at non–zero values or the confirmatory factor analysis model with some factor loadings fixed at zero and with restrictions on factor inter–correlations (Cudeck, 1989). However, in general, by transforming a model on covariances to a model on correlations, the model will be fully scale invariant. A careful discussion of the difficulties associated with the analysis of correlation matrices as covariance matrices and related problems are provided by Cudeck (1989). Moreover, Shapiro and Browne (1990) investigate conditions under which methods intended for the analysis of covariance structures result in correct statistical conclusions when employed for the analysis of correlation structures.

Linear correlation structure analysis is concerned with the representation of the linear dependence structure aiming at a reduction in dimension. Let us consider a random vector $\boldsymbol{x} \in \mathbb{R}^p$ such that (2.1) holds. Correlation structure analysis is now based on the assumption that the population correlation matrix of the variables, $\boldsymbol{R}_0$, satisfies the equation $\boldsymbol{R}_0 = \boldsymbol{R}(\boldsymbol{\theta}_0)$, where $\boldsymbol{R}(\boldsymbol{\theta}_0)$ is the correlation matrix according to the model (2.1).

Let $\boldsymbol{\theta}_0 \in \Theta \subseteq \mathbb{R}$ be a q–dimensional parameter. A correlation structure model is then a matrix–valued function with respect to $\boldsymbol{\theta}_0$,

$$\boldsymbol{R} : \Theta \to \mathbb{R}^{p \times p}, \quad \boldsymbol{\theta}_0 \mapsto \boldsymbol{R}(\boldsymbol{\theta}_0), \tag{3.1}$$

such that $\boldsymbol{R}(\boldsymbol{\theta}_0)$ is a correlation matrix.

Provided that the data are normally distributed, the approach of decomposing the correlation structure analogously to (2.8) is justified, since dependence in normal data is uniquely determined by correlation. However, many data sets exhibit properties contradicting the normality assumption. *Copula structure analysis* is a statistical method for correlation structures introduced by Klüppelberg and Kuhn (2009) to tackle non–normality, problems of non-existing moments (second and fourth moments that ensure asymptotic distributional limits of sample covariance or correlation estimator) or different marginal distributions by using copula models. Klüppelberg and Kuhn (2009) focus on elliptical copulas: as the correlation matrix is the parameter of an elliptical copulas, correlation structure analysis can be extended to such copulas. They only need independent and identically distributed data to ensure consistency and asymptotic normality of the estimated parameter $\hat{\boldsymbol{\theta}}$ as well as the asymptotic χ^2–distribution of the test statistic for model selection, that is for the estimation of the number of latent variables.

Next sections are completely devoted to briefly review the theory of copulas and its use in correlation structure analysis.

3.1 Copula theory: an introduction

The history of copulas may be said to begin with Fréchet (1951). He studied the following problem, which is stated here in a bi–dimensional context: given the distribution functions F_1 and F_2 of two random variables x_1 and x_2 defined on the same probability space $(\mathbb{R}, \mathscr{B}, pr)$, what can be said about the set $\mathcal{C}$ of the bivariate distribution functions whose marginals are F_1 and F_2? It is immediate to note that the set $\mathcal{C}$, now called the Fréchet class of F_1 and F_2, is not empty since, if x_1 and x_2 are independent, then the distribution function $(x_1, x_2) \mapsto F(x_1, x_2) = F_1(x_1) F_2(x_2)$ always belongs to $\mathcal{C}$. But, it was not clear which the other elements of $\mathcal{C}$ were. In 1959, Sklar obtained the deepest result in this respect, by introducing the notion, and the name, of copula.

Definition 3.1 *For every $p \geq 2$, a p–dimensional copula C is a p–variate distribution function on $[0,1]^p$ whose univariate marginals are uniformly dis-*

tributed on $[0,1]$.

Thus, each p–dimensional copula may be associated with a random variable $\boldsymbol{u} = (u_1, \ldots, u_p)^\top$ such that $u_j \sim Unif\,(0,1)$ for every $j = 1, \ldots, p$ and $\boldsymbol{u} \sim C$. Conversely, any random vector whose components are uniformly distributed on $[0,1]$ is distributed according to some copula. The notation $u_j \sim Unif\,(0,1)$ means that the random variable u_j has uniform distribution function on $[0,1]$. The notation $:=$ will be also used for representing the equality by definition later.

Sklar's theorem is the building block of the theory of copulas; without it, the concept of copula would be one in a rich set of joint distribution functions.

Theorem 3.1 (Sklar, 1959) *Let F be a p–dimensional distribution function with univariate margins $F_1, \ldots, F_p$. Let Ran_j denote the range of F_j, $\mathsf{Ran}_j := F_j(\mathbb{R})$ $(j = 1, \ldots, p)$. Then, there exists a copula C such that for all $(x_1, \ldots, x_p)^\top \in \mathbb{R}^p$,*

$$F(x_1, \ldots, x_p) = C\{F_1(x_1), \ldots, F_p(x_p)\} \,. \tag{3.2}$$

Such a C is uniquely determined on $\mathsf{Ran}_1 \times \ldots \times \mathsf{Ran}_p$ and, hence, it is unique when $F_1, \ldots, F_p$ are all continuous.

Theorem 3.1 also admits the following converse implication, usually very important when one wants to construct statistical models by considering, separately, the univariate behavior of the components of a random vector and their dependence properties as captured by some copula.

Theorem 3.2 *If $F_1, \ldots, F_p$ are univariate distribution functions, and if C is any p–dimensional copula, then the function $F : \mathbb{R}^p \to [0,1]$ defined by (3.2) is a p–dimensional distribution function with margins $F_1, \ldots, F_p$.*

The joint distribution function C of $\{F_1(x_1), \ldots, F_p(x_p)\}^\top$ is then called the copula of the random vector $(x_1, \ldots, x_p)^\top$ or the multivariate distribution F. If $F_1, \ldots, F_n$ are not all continuous it can still be shown (see Schweizer and Sklar, 1983, Chapter 6) that the joint distribution function can always be expressed as in (3.2), although in this case C is no longer unique and we refer to it as a possible copula of F.

The proof of Sklar's theorem was not given in Sklar (1959). A sketch of it was provided in Sklar (1973) (see also Schweizer and Sklar, 1974), so that for a few years practitioners in the field had to reconstruct it relying on the hand–written notes by Sklar himself. It should be also mentioned that some "indirect" proofs of Sklar's theorem (without mentioning copula) were later discovered by Moore and Spruill (1975). More recent proofs are also provided by Sklar (1996), Burchard and Hajaiej (2006), and Rüschendorf (2009).

Since copulas are multivariate distribution functions, they can be characterized in the following equivalent way.

Theorem 3.3 *A function $C : [0,1]^p \rightarrow [0,1]$ is a copula if, and only if, the following properties hold:*

(P.1) *for every $j = 1, \ldots, p$, $C(\boldsymbol{u}) = u_j$ when all the components of $\boldsymbol{u}$ are equal to 1 with the exception of the j–th one that is equal to $u_j \in [0,1]$;*

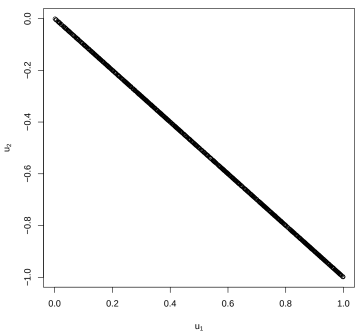
(P.2) *C is isotonic; i.e., $C(\boldsymbol{u}) \leq C(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^p$, $\boldsymbol{u} \leq \boldsymbol{v}$;*

(P.3) *C is p–increasing.*

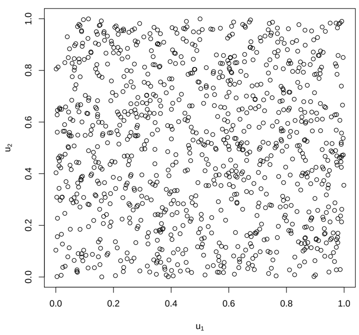
As a consequence, we can prove also that $C(\boldsymbol{u}) = 0$ for every $\boldsymbol{u} \in [0,1]^p$ having at least one of its components equal to 0.

Basic class of copulas are:

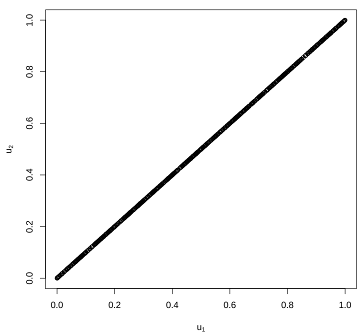
- the *independence copula* $\Pi_p(\boldsymbol{u}) = u_1 \ldots u_p$ associated with a random vector $\boldsymbol{u} = (u_1, \ldots, u_p)^\top$ whose components are independent and uniformly distributed on $[0,1]^p$;
- the *comonotonicity copula* $M_p(\boldsymbol{u}) = \min(u_1, \ldots, u_p)$ associated with a vector $\boldsymbol{u} = (u_1, \ldots, u_p)^\top$ of random variables uniformly distributed on $[0,1]^p$ and such that $u_1 = \ldots = u_p$ almost surely;



(a)



(b)



(c)

Figure 3.1: *Independent realizations from bivariate countermonotonicity (a), independence (b), comonotonicity (c) copulas, respectively.*

- the *countermonotonicity copula* $W_2(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$ associated with a vector $\boldsymbol{u} = (u_1, u_2)^\top$ of random variables uniformly distributed on $[0,1]^2$ and such that $u_1 = 1 - u_2$ almost surely.

By summarizing, from any p–variate distribution function F one can derive a copula C via (3.2). Specifically, when F_j is continuous for every $j = 1, \ldots, p$, C can be obtained by means of the formula

$$C(u_1, \ldots, u_p) = F\left\{F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)\right\},$$

where $F_j^{-1}(u) := \inf\{x \in \mathbb{R} \,|\, F_j(x) \geq u\,,\ u \in [0,1]\}$ denotes the pseudo–inverse of F_j. Thus, copulas are essentially a way for transforming the random variables $(x_1, \ldots, x_p)^\top$ into another random variable $(u_1, \ldots, u_p)^\top$, $u_j = F_j(x_j)$, having the margins uniform on $[0, 1]$ and preserving the dependence among the components. Alternatively, one could transform $\boldsymbol{x}$ to any other distribution, but $Unif(0, 1)$ is particularly easy.

On the other hand, any copula can be combined with different univariate distribution functions in order to obtain a p–variate distribution function by using (3.2). In particular, copulas can serve for modeling situations where a different distribution is needed for each marginal, providing a valid alternative to several classical multivariate distribution functions such Gaussian, Student's t, Pareto, etc., as Durante and Sempi (2010) point out.

In what follows, we deal with semi–parametric copula models $\mathcal{P}$, which are defined as follows. Let $\mathcal{C} = \left\{C_{\boldsymbol{x}}(\cdot\,;\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \mathbb{A} \subset \mathbb{R}^d\right\}$ be a parametric family of copulas on $[0, 1]^p$ with density $c_{\boldsymbol{x}}(\cdot\,;\boldsymbol{\alpha})$ with respect to Lebesgue measure on $[0, 1]^p$, indexed by a d–dimensional real parameter vector $\boldsymbol{\alpha}$.

For $\boldsymbol{\alpha} \in \mathbb{A}$ and arbitrary distribution functions $F_1, \ldots, F_p$ on $\mathbb{R}$, let $F_{\boldsymbol{\alpha},F_1,\ldots,F_p}$ be the distribution function on $\mathbb{R}^p$ defined by

$$F_{\boldsymbol{\alpha},F_1,\ldots,F_p}(x_1, \ldots, x_p) = C_{\boldsymbol{x}}\{F_1(x_1), \ldots, F_p(x_p)\,;\boldsymbol{\alpha}\}$$

for $(x_1, \ldots, x_p) \in \mathbb{R}^p$. Then with $pr(\cdot;\boldsymbol{\alpha}, F_1, \ldots, F_p)$ denoting the corresponding probability measure on $(\mathbb{R}^p, \mathscr{B}^p)$, where $\mathbb{R}^p$ is the p–dimensional real Euclidean space and $\mathscr{B}^p$ its Borel σ–field, and $\mathscr{F}$ denoting the collection of all distribution function on $\mathbb{R}$,

$$\mathcal{P} = \{pr(\cdot;\boldsymbol{\alpha}, F_1, \ldots, F_p) : \boldsymbol{\alpha} \in \mathbb{A}, F_j \in \mathscr{F}, j = 1, \ldots, p\}$$

is a semi–parametric copula model.

One simple example, which is widely exploited by Klüppelberg and Kuhn (2009), is provided by the family of elliptical copulas being the copulas of elliptical distributions. These copulas are very flexible and easy to handle also in high dimensions. For instance, let us consider the copula resulting from the multivariate normal distribution on $\mathbb{R}^p$, $F_{\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0}$, having mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. Let F_j denote the one–dimensional standard normal distribution function with mean 0 and variance 1. Then, $C_{\boldsymbol{\alpha}}$ satisfies

$$C_{\boldsymbol{x}}\left(u_1,\ldots,u_p\,;\,\boldsymbol{\alpha}\right)=F_{\boldsymbol{x}}\left\{F_1^{-1}\left(u_1\right),\ldots,F_p^{-1}\left(u_p\right)\,;\,\boldsymbol{\alpha}\right\}$$

where, in this case, $\boldsymbol{\alpha}$ consists of the population linear correlation coefficients between variables $\boldsymbol{x}$.

3.1.1 The elliptical and meta–elliptical copulas

As underlined in the previous section, copulas play an important role in the construction of multivariate distribution function. As a consequence, having at one's disposal a variety of copulas can be very useful for building stochastic models with different properties, sometimes indispensable in practice (e.g., heavy tails, asymmetries, etc.). Therefore, several investigations have been carried out concerning the construction of different families of copulas and their properties. In this work we deal just two of them, by focusing in this chapter on the family that Klüppelberg and Kuhn (2009) use in their work, namely, elliptical copulas. Different families (or construction methods) are discussed in the books of Joe (1997) and Nelsen (2006).

Elliptical copulas describe the dependence structure in elliptical distributions as well as in their extensions, the meta–elliptical distributions, which have been originally introduced in Fang, Fang, and Kotz (2002). Their properties are examined by Frahm, Junker, and Szimayer (2003) and Abdous, Genest, and Rèmillard (2005). These dependence structures are popular in actuarial science and in finance; see Malevergne and Sornette (2003), Cherubini, Luciano, and Vecchiato (2004), McNeil, Frey, and Embrechts (2005) and references therein. We start by recalling the definition of an elliptical distribution and we refer to Fang, Kotz, and Ng (1990) for a comprehensive overview.

A random vector $\boldsymbol{x}\in\mathbb{R}^p$ has an elliptical distribution with parameters $\boldsymbol{\mu}_0\in\mathbb{R}^p$ and a positive (semi) definite matrix $\boldsymbol{\Sigma}_0\in\mathbb{R}^{p\times p}$, if $\boldsymbol{x}$ has the stochastic representation

$$\boldsymbol{x}\stackrel{d}{=}\boldsymbol{\mu}_0+\mathsf{r}\boldsymbol{A}\boldsymbol{u}\,,\quad \boldsymbol{A}\in\mathbb{R}^{p\times p}\,, \tag{3.3}$$

where $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Sigma}_0$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_0$, $\mathsf{r} \geq 0$ is a random variable, $\boldsymbol{u}$ is uniformly distributed on the unit sphere in $\mathbb{R}^p$ and is independent of r.

We write $\boldsymbol{x} \sim \mathcal{E}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, h\right)$, where $h\left(\cdot\right)$ is a scale function uniquely determined by the distribution of r. The random variable r is called the generating variable. Further, if the first moment exists, then $E\left(\boldsymbol{x}\right) = \boldsymbol{\mu}_0$ and, if the second moment exists, then r can be chosen such that $cov\left(\boldsymbol{x}\right) = \boldsymbol{\Sigma}_0$. We define the correlation matrix $\boldsymbol{R}_0$ of $\boldsymbol{x}$ as $\boldsymbol{R}_0 := diag\left(\boldsymbol{\Sigma}_0\right)^{-1/2} \boldsymbol{\Sigma}_0\, diag\left(\boldsymbol{\Sigma}_0\right)^{-1/2}$. If $\boldsymbol{x}$ has finite second moment, then $cor\left(\boldsymbol{x}\right) = \boldsymbol{R}_0$.

The representation (3.3) is such that, when r has a density, the multivariate density of $\boldsymbol{x}$ is given by

$$f\left(\boldsymbol{t}\right) = \left|\boldsymbol{\Sigma}_0\right|^{-1/2} h\left\{\left(\boldsymbol{t} - \boldsymbol{\mu}_0\right)^\top \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{t} - \boldsymbol{\mu}_0\right)\right\}, \quad \boldsymbol{t} \in \mathbb{R}^p .$$

When $h\left(t\right) = e^{-t/2}$, for instance, $\boldsymbol{x}$ is multivariate normal. Similarly, $h\left(t\right) = c\left(1 + t/\nu\right)^{-(p+\nu)/2}$, for a suitable constant c, generates the multivariate Student's t distribution with ν degrees of freedom.

We define an elliptical copula as the copula of $\boldsymbol{x} \sim \mathcal{E}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, h\right)$, denoted by $\mathcal{E}C\left(\boldsymbol{R}_0, h\right)$. We call $\boldsymbol{R}_0$ the copula correlation matrix. The notion $\mathcal{E}C\left(\boldsymbol{R}_0, h\right)$ for an elliptical copula makes sense, since it is characterized by the generating variable r (which is unique up to a multiplicative constant) and the copula correlation matrix $\boldsymbol{R}_0$.

One inconvenient limitation of elliptical distributions is that the scaled variables (with respect to the standard deviation) are identically distributed according to a distribution function F. However, models based on the unique meta–elliptical distribution associated with $\boldsymbol{x}$ do not suffer from this defect. We regain the flexibility of modeling the margins separately, while keeping the dependence structure of an elliptical distribution, by considering meta–elliptical distribution functions. The dependence structure in a meta–elliptical distribution is hence described by the corresponding elliptical copula.

3.2 Copula Structure Analysis assuming elliptical copulas

For a linear correlation structure model with elliptical latent variables, function (3.1) corresponds to the following situation. Let $\boldsymbol{\zeta} \sim \mathcal{E}(\mathbf{0}, \boldsymbol{I}_z, h)$ be a z–dimensional elliptical random vector, let $\boldsymbol{A} : \Theta \to \mathbb{R}^{p \times z}, \boldsymbol{\theta}_0 \mapsto \boldsymbol{A}(\boldsymbol{\theta}_0)$, be some matrix–valued function with argument $\boldsymbol{\theta}_0$ and define

$$\boldsymbol{\Sigma} : \Theta \to \mathbb{R}^{p \times p}, \quad \boldsymbol{\theta}_0 \mapsto \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) := \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{A}(\boldsymbol{\theta}_0)^\top . \tag{3.4}$$

Then expression (3.1) can be written as

$$\boldsymbol{R}(\boldsymbol{\theta}_0) = diag\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\, diag\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}^{-1/2} . \tag{3.5}$$

As a correlation matrix is a parameter of an elliptical copula, we can extend the usual correlation structure model to elliptical copulas. Denote by $C_{\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\zeta}}$ the copula of $\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\zeta} \in \mathbb{R}^p$. Klüppelberg and Kuhn (2009) state that the random vector $\boldsymbol{x} \in \mathbb{R}^p$ with copula $C_{\boldsymbol{x}}$ satisfies a copula structure model, if

$$C_{\boldsymbol{x}} = C_{\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\zeta}} \in \mathcal{E}C\{\boldsymbol{R}(\boldsymbol{\theta}_0), h\} , \tag{3.6}$$

where $\boldsymbol{R}(\boldsymbol{\theta}_0)$ is defined in (3.5).

Define $\boldsymbol{F}^{-1}(\boldsymbol{u}) := \left\{F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)\right\}^\top$ as the vector of the pseudo–inverses of the marginal distribution functions of $\boldsymbol{x}$ and $\boldsymbol{H}(\boldsymbol{x}) := \{H_1(x_1), \ldots, H_p(x_p)\}^\top$ as the vector of the marginal distribution functions of $\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\zeta}$. Then condition (3.6) is equivalent to $\boldsymbol{x} \sim \boldsymbol{F}^{-1}[\boldsymbol{H}\{\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\zeta}\}]$, where all operations are component–wise. Hence, the copula model can also be seen as an extension of a correlation structure model for elliptical data, where the equality in distribution assumption for the variables in $\boldsymbol{x}$ is relaxed. If not only $C_{\boldsymbol{x}} = C_{\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\zeta}}$ holds but also $\boldsymbol{H} = \boldsymbol{F}$ with existing second moment, then this is a classical correlation or covariance structure model. For normal $\boldsymbol{\zeta}$ it gives back the classical normal model.

The standard correlation structure model assumes some (functional) structure for the correlation matrix of the observed data. The only difference lies in the interpretation of the correlation matrix. In the classical model it represents the linear correlation between the data. In the copula model it represents a more general dependence parameter which can be interpreted as a correlation–like measure.

Now, let's turn to the problem of estimating a copula structure model. It means to estimate the parameter $\boldsymbol{\theta}_0$ that characterizes the correlation structure. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be an *IID* sequence of random vectors in $\mathbb{R}^p$ and denote by $\hat{\boldsymbol{R}} := \hat{\boldsymbol{R}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ an arbitrary estimator of the correlation matrix $\boldsymbol{R}_0$ of $\boldsymbol{x}$ as for instance the empirical correlation or a copula correlation estimator. Given the estimator $\hat{\boldsymbol{R}}$, Klüppelberg and Kuhn (2009) want to find some parameter vector $\boldsymbol{\theta}$ which fits the assumed structure $\boldsymbol{R}(\boldsymbol{\theta})$ to $\hat{\boldsymbol{R}}$ as well as possible. They define $\hat{\boldsymbol{r}} := vecp\left(\hat{\boldsymbol{R}}\right)$ and $\boldsymbol{r}(\boldsymbol{\theta}) := vecp\{\boldsymbol{R}(\boldsymbol{\theta})\}$, the vectors of patterned matrices $\hat{\boldsymbol{R}}$ and $\boldsymbol{R}(\boldsymbol{\theta})$ (see Appendix A), and they estimate $\boldsymbol{\theta}_0$ by minimizing the discrepancy function (2.3) defined by

$$D_{QD}\{\hat{\boldsymbol{r}}, \boldsymbol{r}(\boldsymbol{\theta}) | \boldsymbol{W}\} = \{\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{\theta})\}^\top \boldsymbol{W}^{-1} \{\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{\theta})\}, \tag{3.7}$$

where $\boldsymbol{W}$ is a positive definite matrix or a consistent estimator of some positive definite matrix.

We now review some results due to Browne (1984), which Klüppelberg and Kuhn (2009) exploit for the estimation of the copula structure model. Given a discrepancy function D and some estimator $\hat{\boldsymbol{R}}$ of the correlation matrix $\boldsymbol{R}_0$, Klüppelberg and Kuhn (2009) can firstly define a consistent estimator of $\boldsymbol{\theta}_0$.

Proposition 3.1 (Browne, 1984, Proposition 1) *Let $\boldsymbol{R}_0$ be the population correlation matrix, and $\boldsymbol{r}_0 := vecp(\boldsymbol{R}_0) \in \mathbb{R}^{p^{\star\star}}$, $p^{\star\star} = p(p-1)/2$. Assume that $\hat{\boldsymbol{r}}$ is an estimator of $\boldsymbol{r}_0$ based on an* IID *sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and that $\hat{\boldsymbol{r}} \xrightarrow{p} \boldsymbol{r}_0$ as $n \to \infty$. Further suppose that D is a discrepancy function satisfying properties **(P.1)**, **(P.2)** and **(P.3)** and that regularity conditions **(A.2)**, **(A.6)** and **(A.7)** hold, as specified in Section 2.1. Define the estimator*

$$\hat{\boldsymbol{\theta}} := arg \min_{\boldsymbol{\theta} \in \Theta} D \left\{ \hat{\boldsymbol{r}}, \boldsymbol{r}\left(\boldsymbol{\theta}\right) | \boldsymbol{W} \right\} . \tag{3.8}$$

Then

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0 \quad as\ n \to \infty.$$

Given the estimator of $\boldsymbol{\theta}_0$, Klüppelberg and Kuhn (2009) show how to test the assumed correlation structure. Under the assumption of Proposition 3.1, let $T_{\boldsymbol{W}}$ be the test statistic,

$$T_{\boldsymbol{W}} := n \min_{\boldsymbol{\theta} \in \Theta} D_{QD} \left\{ \hat{\boldsymbol{r}}, \boldsymbol{r}\left(\boldsymbol{\theta}\right) | \boldsymbol{W} \right\} , \tag{3.9}$$

for some matrix $\boldsymbol{W}$. The null hypothesis is that the true correlation vector $\boldsymbol{r}_0$ satisfies a prespecified correlation structure model; i.e.,

$$H_0 : \boldsymbol{r}_0 = \boldsymbol{r}\left(\boldsymbol{\theta}_0\right) \tag{3.10}$$

for some $\boldsymbol{\theta}_0 \in \Theta$.

To obtain the limit distribution of $T_{\boldsymbol{W}}$ for the quadratic discrepancy function (3.7), Klüppelberg and Kuhn (2009) apply the following result due to Browne (1984).

Theorem 3.4 (Browne, 1984, Corollary 4.1) *Assume that the conditions of Proposition 3.1 and **(A.3)** and **(A.4)** hold, as specified in Section 2.1. Furthermore, assume that* $n^{1/2}\left(\hat{\boldsymbol{r}} - \boldsymbol{r}_0\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\boldsymbol{0}, \boldsymbol{W}_0\right)$ *and that* $\hat{\boldsymbol{W}}$ *is a consistent estimator of* $\boldsymbol{W}_0$. *Then, under the null hypothesis* (3.10),

$$T_{\hat{\boldsymbol{W}}} := n \min_{\boldsymbol{\theta} \in \Theta} D \left\{ \hat{\boldsymbol{r}}, \boldsymbol{r}\left(\boldsymbol{\theta}\right) \middle| \hat{\boldsymbol{W}} \right\} \xrightarrow{\mathscr{L}} \chi^2 ,$$

as $n \to \infty$, *where the degrees of freedom are* $p^{\star\star} - q$, *with* q *being the dimension of* $\boldsymbol{\theta}$.

To select an appropriate correlation structure model, that is to correctly estimate the number of latent variables, Klüppelberg and Kuhn (2009) take a set of g nested models (such that all satisfy the assumptions of Theorem

3.4) and define the null hypotheses $H_0^{(s)} : \boldsymbol{r}_0 = \boldsymbol{r}\left\{\boldsymbol{\theta}_0^{(s)}\right\}$ for some $\boldsymbol{\theta}_0^{(s)} \in \Theta^{(s)}$, $1 \leq s \leq g$. Assume that at least one of these null hypotheses holds true; i.e., there is some s' such that $H_0^{(s)}$ does not hold for $1 \leq s < s'$ and does hold for $s' \leq s \leq g$. As Klüppelberg and Kuhn (2009) are interested in a structure model, which is likely to explain the observed dependence structure and is as simple as possible, the smallest index s' where the null hypothesis is not rejected must be estimated. By Theorem 3.4 the corresponding test statistics

$$T_{\hat{\boldsymbol{W}}}^{(s)} := n \min_{\boldsymbol{\theta} \in \Theta^{(s)}} D_{QD}\left\{\hat{\boldsymbol{r}}, \boldsymbol{r}\left(\boldsymbol{\theta}_0^{(s)}\right) \middle| \hat{\boldsymbol{W}}\right\}$$

are not χ^2 distributed for $1 \leq s < s'$ and are χ^2 distributed for $s' \leq s \leq g$. Consequently, Klüppelberg and Kuhn (2009) reject a null hypothesis $H_0^{(s)}$, if the corresponding test statistic $T_{\hat{\boldsymbol{W}}}^{(s)}$ is larger than some χ^2 quantile. Hence, s' represents the smallest number of latent variables where $H_0^{(s')}$ cannot be rejected.

As Klüppelberg and Kuhn (2009) consider a copula structure model, according to Theorem 3.4 they need an estimator $\hat{\boldsymbol{R}}$ of the copula correlation matrix $\boldsymbol{R}_0$, such that the vector of its patterned version is asymptotically distributed as a multivariate normal with mean $\boldsymbol{R}_0$ and covariance matrix $\boldsymbol{W}_0$, a $p^{\star\star} \times p^{\star\star}$ positive definite matrix. Moreover, they need a consistent estimator for $\boldsymbol{W}_0$ to be included as weight matrix $\boldsymbol{W}$ in (3.7).

Concerning elliptical copulas $\mathcal{EC}(\boldsymbol{R}_0, h)$ with absolute continuous generating variable $\mathsf{r} > 0$, Fang, Fang, and Kotz (2002) (originally, Kruskal, 1958) provide a functional relationship between correlation matrix $\boldsymbol{R}_0$ and Kendall's τ–matrix $\boldsymbol{T} := [\tau_{i,j}]_{1 \leq i,j \leq p}$.

Theorem 3.5 (Fang, Fang, and Kotz, 2002, Theorem 3.1) *Let $\boldsymbol{x}$ be a vector of random variables with elliptical copula $\mathcal{EC}(\boldsymbol{R}_0, h)$ and absolutely continuous generating variable $\mathsf{r} > 0$; then*

$$\rho_{i,j} = sin\left(\frac{\pi}{2}\tau_{i,j}\right). \tag{3.11}$$

Since Klüppelberg and Kuhn (2009) consider an elliptical copula, they invoke the relationship (3.11) for the estimation of $\boldsymbol{R}_0$. Estimating the copula correlation matrix via Kendall's τ yields a general useful result in order to provide conditions for Theorem 3.4. This *naive* method of estimation for copula parameters, which is in the spirit of Pearson's method of moments, is typical for some copula families. A rough–and–ready strategy thus might be to estimate the copula correlation coefficients by replacing in (3.11) the population Kendall's tau with its sample value. The main idea then involves computing the matrix of sample Kendall's taus, and then inverting the resulting matrix element–wise using (3.11).

The copula moment–based estimation of $\boldsymbol{R}_0$ can then be seen as a robust extension of the usual correlation structure analysis, where it is not required the existence of moments.

Theorem 3.6 (Klüppelberg and Kuhn, 2009, Theorem 3) *Let* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *be an* IID *sequence in* $\mathbb{R}^p$ *with elliptical copula* $\mathcal{EC}(\boldsymbol{R}_0, h)$ *and absolutely continuous generating variable* $\mathsf{r} > 0$. *Let* $\hat{\boldsymbol{T}} := [\hat{\tau}_{i,j}]_{1\le i,j\le p}$ *be the estimated Kendall's* τ*–matrix. Further, define the estimated correlation matrix as*

$$\hat{\boldsymbol{R}}_\tau := \sin\left(\frac{\pi}{2}\hat{\boldsymbol{T}}\right) , \tag{3.12}$$

where the sine function is used componentwise, and define $\hat{\boldsymbol{r}}_\tau := vecp\left(\hat{\boldsymbol{R}}_\tau\right)$ *and* $\boldsymbol{r}_0 := vecp\left(\boldsymbol{R}_0\right)$, *the vectors of patterned matrices* $\hat{\boldsymbol{R}}_\tau$ *and* $\boldsymbol{R}_0$, *respectively. Then, as* $n \to \infty$,

$$n^{1/2}\left(\hat{\boldsymbol{r}}_\tau - \boldsymbol{r}_0\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_\tau\right) ,$$

where $\boldsymbol{\Sigma}_\tau := \left[\sigma^\tau_{ij,kl}\right]_{1\le i\ne j, k\ne l\le p}$ *and*

$$\sigma^\tau_{ij,kl} = \pi^2 \cos\left(\frac{\pi}{2}\tau_{i,j}\right)\cos\left(\frac{\pi}{2}\tau_{k,l}\right)\left(\tau_{ij,kl} - \tau_{i,j}\tau_{k,l}\right) ,$$

$$\tau_{i,j} = E\left[sgn\left\{\left(x_{1,i} - x_{2,i}\right)\left(x_{1,j} - x_{2,j}\right)\right\}\right] , \tag{3.13}$$

$$\tau_{ij,kl} = E\left(E\left[sgn\left\{\left(x_{1,i} - x_{2,i}\right)\left(x_{1,j} - x_{2,j}\right)\right\} | \boldsymbol{x}_1\right]\right.$$

$$E\left[sgn\left\{\left(x_{1,k}-x_{3,k}\right)\left(x_{1,l}-x_{3,l}\right)\right\}|\boldsymbol{x}_1\right]). \tag{3.14}$$

The following result provides a consistent estimator for the asymptotic covariance matrix $\boldsymbol{\Sigma}_\tau$.

Theorem 3.7 (Klüppelberg and Kuhn, 2009, Theorem 4) *Under the assumptions of Theorem 3.6, let us define the estimator of* $\boldsymbol{\Sigma}_\tau$ *as*

$$\hat{\boldsymbol{\Sigma}}_\tau := \left[\pi^2 \cos\left(\frac{\pi}{2}\hat{\tau}_{i,j}\right)\cos\left(\frac{\pi}{2}\hat{\tau}_{k,l}\right)\left(\hat{\tau}_{ij,kl}-\hat{\tau}_{i,j}\hat{\tau}_{k,l}\right)\right]_{1\le i\neq j,k\neq l\le p}, \tag{3.15}$$

where

$$\hat{\tau}_{i,j} = \binom{n}{2}^{-1}\sum_{1\le a<b\le n} sgn\left\{\left(X_{a,i}-X_{b,i}\right)\left(X_{a,j}-X_{b,j}\right)\right\}$$

and

$$\hat{\tau}_{ij,kl} = \frac{1}{n\left(n-1\right)^2}\sum_{a=1}^{n}\left(\left[\sum_{b=1,b\neq a}^{n} sgn\left\{\left(X_{a,i}-X_{b,i}\right)\left(X_{a,j}-X_{b,j}\right)\right\}\right]\times\right.$$
$$\left.\times\left[\sum_{c=1,c\neq a}^{n} sgn\left\{\left(X_{a,k}-X_{c,k}\right)\left(X_{a,l}-X_{c,l}\right)\right\}\right]\right).$$

Then, $vech\left(\hat{\boldsymbol{\Sigma}}_\tau\right)$ *is consistent and asymptotically normal.*

Unfortunately, both the Kendall's τ–based estimated correlation matrix (3.12) as well as its estimated asymptotic covariance matrix (3.15) may sometimes not be positive definite. In such a case, Klüppelberg and Kuhn (2009) suggest to replace them by its projection into the class of correlation or covariance matrices, respectively. An algorithm for the computation of the projection $\hat{\boldsymbol{R}}_\tau^\star$ of $\hat{\boldsymbol{R}}_\tau$ into the class of correlation matrices iteratively replaces negative eigenvalues by 0 and then replaces the diagonal of the resulting matrix by 1; see Rousseeuw and Molenberghs (1993) or Higham (2002). It can be shown that the projection $\hat{\boldsymbol{\Sigma}}_\tau^\star$ of $\hat{\boldsymbol{\Sigma}}_\tau$ into the class of covariance matrices is obtained by replacing the negative eigenvalues of $\hat{\boldsymbol{\Sigma}}_\tau$ by 0; also see Rousseeuw and Molenberghs (1993) or Higham (2002).

By exploiting the results of Theorem 3.6 and Theorem 3.7, Klüppelberg and Kuhn (2009) can now apply the test statistic (3.9) in order to test a specified structural model $\boldsymbol{r}_0 = \boldsymbol{r}(\boldsymbol{\theta}_0)$ for some $\boldsymbol{\theta}_0 \in \Theta$. Since the asymptotic χ^2–distribution of the test statistic (3.9) depends on some analytic regularity conditions, which may not be satisfied, a robust test statistic has been suggested in Browne (1984, Proposition 4) (also see Yuan and Bentler, 1999, and Satorra and Bentler, 2001). Instead of using $\hat{\boldsymbol{\Sigma}}_\tau^{-1}$ as weight matrix in the test statistic (3.9), the corrected version

$$\hat{\boldsymbol{\Sigma}}_\tau^{-1} - \hat{\boldsymbol{\Sigma}}_\tau^{-1}\hat{\mathbf{J}}\left(\hat{\mathbf{J}}^\top\hat{\boldsymbol{\Sigma}}_\tau^{-1}\hat{\mathbf{J}}\right)^{-1}\hat{\mathbf{J}}^\top\hat{\boldsymbol{\Sigma}}_\tau^{-1}$$

is taken, where $\hat{\mathbf{J}}$ is an estimator of the Jacobian matrix $\mathbf{J}_{\boldsymbol{\theta}_0} = \mathbf{J}(\boldsymbol{\theta}_0) := \left[\partial\boldsymbol{r}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

3.3 The copula factor model

Klüppelberg and Kuhn (2009) state that the random vector $\boldsymbol{x} \in \mathbb{R}^p$ with copula $C_{\boldsymbol{x}}$ satisfies an *elliptical copula factor model* if condition (3.6) hold, that is if there exists $\boldsymbol{\zeta} \sim \mathcal{E}(\mathbf{0}, \boldsymbol{I}_z, h)$ with $z = m + p$ such that

$$C_{\boldsymbol{x}} = C_{(\boldsymbol{\Lambda},\boldsymbol{I}_p)\boldsymbol{\zeta}}\,, \tag{3.16}$$

where $\boldsymbol{\theta}_0 = vecp\,(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$, and the correlation matrix is assumed to be of the form $\boldsymbol{R}_0 = \boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}_0\right) = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^\top + \tilde{\boldsymbol{\Psi}}$ for some $m \ll p$, $\tilde{\boldsymbol{\Lambda}} = diag\,\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}^{-1/2}\,\boldsymbol{\Lambda} \in \mathbb{R}^{p\times m}$ and $\tilde{\boldsymbol{\Psi}} = diag\,\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}^{-1/2}\,\boldsymbol{\Psi}\,diag\,\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}^{-1/2} \in \mathbb{R}^{p\times p}$, with $\tilde{\boldsymbol{\theta}}_0 = vecp\left(\tilde{\boldsymbol{\Lambda}}, \tilde{\boldsymbol{\Psi}}\right)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}$. Using the estimators (3.12) and (3.15) together with the quadratic discrepancy function (3.7), Klüppelberg and Kuhn (2009) can estimate $\tilde{\boldsymbol{\theta}}_0$ and test the elliptical copula factor model.

In general, a unique estimated parameter vector $\hat{\tilde{\boldsymbol{\theta}}}$ does not exist. As revised in Section 2.2.1, in the classical factor model, $\tilde{\boldsymbol{\Lambda}}$ can always be replaced by $\tilde{\mathbf{\Lambda}}\boldsymbol{M}^\top$, where $\boldsymbol{M}$ is any orthogonal matrix of order m. By a minor adaptation of the parameter space Θ (i.e., $\tilde{\mathbf{\Lambda}}^\top \tilde{\mathbf{\Psi}}^{-1} \tilde{\mathbf{\Lambda}}$ must be diagonal), $\hat{\tilde{\boldsymbol{\theta}}}$ can be forced to be unique and Proposition 3.1 applies. By Lee and Bentler (1980) the degrees of freedom in (3.9) are then increased by the number of additional constraints.

In the case of the copula factor model Klüppelberg and Kuhn (2009) need to estimate only the loading matrix $\tilde{\mathbf{\Lambda}}$, since $diag\left(\tilde{\mathbf{\Psi}}\right) = \mathbf{1}_p - diag\left(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top\right)$. Therefore the number of free parameters is pm minus the number of additional constraints to ensure that $\tilde{\mathbf{\Lambda}}^\top \tilde{\mathbf{\Psi}}^{-1} \tilde{\mathbf{\Lambda}}$ is diagonal; i.e., the degrees of freedom of the limiting χ^2–distribution of test statistic are $p^{\star\star} - pm + m\left(m-1\right)/2$.

For the computation of $\hat{\tilde{\boldsymbol{\theta}}}$ and the test statistic as defined in (3.8) and in (3.9), respectively, Klüppelberg and Kuhn (2009) used the statistical software package `R` and the optimization routine `optim` with the `Nelder-Mead` method therein. By adding appropriate penalty terms to the discrepancy functions, they take both side–conditions into account; i.e., $\tilde{\mathbf{\Lambda}}^\top \tilde{\mathbf{\Psi}}^{-1} \tilde{\mathbf{\Lambda}}$ is diagonal and $diag\left(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top + \tilde{\mathbf{\Psi}}\right) = \mathbf{1}$. As starting values for the optimization algorithm, they take the loadings that are derived from the standard factor analysis routine `factanal`, which uses the normal maximum likelihood discrepancy function (2.2).

Since Klüppelberg and Kuhn (2009) do not provide detailed steps for estimating $\tilde{\boldsymbol{\theta}}_0$ and testing the correlation structure, we supply a procedure that can be adapted to any estimator of correlation matrix satisfying conditions of Theorem (3.4). We limit ourself to describe the computational algorithm for copula factor models. Unlike Klüppelberg and Kuhn (2009) we show the analytic partial derivatives obtained by using linear algebra.

Our aim is to minimize the discrepancy function (2.3) defined by

$$\begin{aligned} D_{QD}\left\{\hat{\boldsymbol{r}}, \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) \middle| \hat{\mathbf{\Sigma}}_{\hat{r}}\right\} &= \left\{\hat{\boldsymbol{r}} - \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right)\right\}^\top \hat{\mathbf{\Sigma}}_{\hat{r}}^{-1} \left\{\hat{\boldsymbol{r}} - \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right)\right\} \\ &\approx \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right)^\top \hat{\mathbf{\Sigma}}_{\hat{r}}^{-1} \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) - 2\hat{\boldsymbol{r}}^\top \hat{\mathbf{\Sigma}}_{\hat{r}}^{-1} \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) \end{aligned} \quad (3.17)$$

such that $\tilde{\mathbf{\Lambda}}^\top \tilde{\mathbf{\Psi}}^{-1} \tilde{\mathbf{\Lambda}}$ is diagonal.

The main point of this procedure consists in considering a suitable transformation of the vectorized patterned correlation matrix as specified by copula factor model; i.e.,

$$\begin{aligned} \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) &:= vecp\left\{\boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}\right)\right\} = vecp\left(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top\right) = \\ &= \boldsymbol{P}_p\left(\boldsymbol{I}_p \otimes \tilde{\mathbf{\Lambda}}\right)\boldsymbol{K}_{pm}\, vec\left(\tilde{\mathbf{\Lambda}}\right), \end{aligned}$$

where $\boldsymbol{P}_p$ represents the left inverse of the transition or duplication matrix $\boldsymbol{Q}_p$ for patterned matrices and $\boldsymbol{K}_{pm}$ is referred to as a vec–permutation matrix or, more commonly, commutation matrix (see Appendix A for a formal definition of both). The diagonal elements of $\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top$ are here regarded redundant.

For the purpose of minimizing the function (3.17) we require its partial derivatives with respect to the elements of $\tilde{\mathbf{\Lambda}}$. Firstly, we provide the partial derivatives of the patterned correlation matrix and the constraint on loadings respect to the loadings that we will use later.

$$\begin{aligned} \frac{\partial}{\partial\left\{vec\left(\tilde{\mathbf{\Lambda}}\right)\right\}^\top}\boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) &= \frac{\partial}{\partial\left\{vec\left(\tilde{\mathbf{\Lambda}}\right)\right\}^\top} vecp\left(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top\right) = \\ &= \boldsymbol{P}_p \frac{\partial}{\partial\left\{vec\left(\tilde{\mathbf{\Lambda}}\right)\right\}^\top} vec\left(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top\right) = \\ &= \boldsymbol{P}_p\left[\left(\boldsymbol{I}_p \otimes \tilde{\mathbf{\Lambda}}\right)\frac{\partial}{\partial\left\{vec\left(\tilde{\mathbf{\Lambda}}\right)\right\}^\top} vec\left(\tilde{\mathbf{\Lambda}}^\top\right) + \right. \\ &\left. + \left(\tilde{\mathbf{\Lambda}} \otimes \boldsymbol{I}_p\right)\frac{\partial}{\partial\left\{vec\left(\tilde{\mathbf{\Lambda}}\right)\right\}^\top} vec\left(\tilde{\mathbf{\Lambda}}\right)\right] = \\ &= \boldsymbol{P}_p\left\{\left(\boldsymbol{I}_p \otimes \tilde{\mathbf{\Lambda}}\right)\boldsymbol{K}_{pm} + \left(\tilde{\mathbf{\Lambda}} \otimes \boldsymbol{I}_p\right)\right\}, \end{aligned}$$

$$
\begin{aligned}
&\frac{\partial}{\partial\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}} vech\left(\tilde{\boldsymbol{\Lambda}}^{\top}\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right) = \boldsymbol{H}_m \frac{\partial}{\partial\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}} vec\left(\tilde{\boldsymbol{\Lambda}}^{\top}\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right) = \\
&= \boldsymbol{H}_m \left[\left\{\left(\tilde{\boldsymbol{\Lambda}}^{\top}\tilde{\boldsymbol{\Psi}}^{-1}\right)\otimes \boldsymbol{I}_m\right\} \boldsymbol{K}_{pm} \frac{\partial}{\partial\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}} vec\left(\tilde{\boldsymbol{\Lambda}}\right) + \right. \\
&\left. + \left(\boldsymbol{I}_m \otimes \tilde{\boldsymbol{\Lambda}}^{\top}\right) \frac{\partial}{\partial\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}} vec\left(\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\right] = \\
&= \boldsymbol{H}_m \left[\left\{\left(\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\otimes \boldsymbol{I}_m\right\}^{\top} \boldsymbol{K}_{pm} + \right. \\
&\left. + \left(\boldsymbol{I}_m \otimes \tilde{\boldsymbol{\Lambda}}\right)^{\top} \left\{\left(\boldsymbol{I}_m \otimes \tilde{\boldsymbol{\Psi}}^{-1}\right) \frac{\partial}{\partial\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}} vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}\right] = \\
&= \boldsymbol{H}_m \left[\left\{\left(\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\otimes \boldsymbol{I}_m\right\}^{\top} \boldsymbol{K}_{pm} + \left(\boldsymbol{I}_m \otimes \tilde{\boldsymbol{\Lambda}}\right)^{\top} \left(\boldsymbol{I}_m \otimes \tilde{\boldsymbol{\Psi}}^{-1}\right)\right].
\end{aligned}
$$

Let $\boldsymbol{l}_1$ represent a arbitrary $m^{\star} \times 1$, $m^{\star} = m\,(m+1)\,/2$ real vector. The Lagrangian function for our constrained minimization problem can be expressed as

$$
D_{QD}\left\{\hat{\boldsymbol{r}}, \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) \middle| \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right\} - 2\,\boldsymbol{l}_1^{\top}\left\{vech\left(\boldsymbol{\Xi}\right) - vech\left(\tilde{\boldsymbol{\Lambda}}^{\top}\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\right\}, \tag{3.18}
$$

where $\boldsymbol{\Xi}$ is the diagonal matrix containing the m eigenvalues of $\hat{\boldsymbol{r}}$. The $m^{\star}$ scalars $-2\,l_1, \ldots, -2\,l_{m^{\star}}$ are the Lagrange multipliers. The partial derivative of (3.18) respect to the loadings is hence given by

$$\frac{\partial}{\partial\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}}\left[D_{QD}\left\{\hat{\boldsymbol{r}},\boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right)\Big|\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right\}-2\,\boldsymbol{l}_1^{\top}\left\{vech\left(\boldsymbol{\Xi}\right)-vech\left(\tilde{\boldsymbol{\Lambda}}^{\top}\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\right\}\right]=$$

$$=2\left[\left\{vecp\left(\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^{\top}\right)\right\}^{\top}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\boldsymbol{P}_p\left\{\left(\boldsymbol{I}_p\otimes\tilde{\boldsymbol{\Lambda}}\right)\boldsymbol{K}_{pm}+\left(\tilde{\boldsymbol{\Lambda}}\otimes\boldsymbol{I}_p\right)\right\}+\right.$$

$$\left.-\,\hat{\boldsymbol{r}}^{\top}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\boldsymbol{P}_p\left\{\left(\boldsymbol{I}_p\otimes\tilde{\boldsymbol{\Lambda}}\right)\boldsymbol{K}_{pm}+\left(\tilde{\boldsymbol{\Lambda}}\otimes\boldsymbol{I}_p\right)\right\}\right]+$$

$$+\,2\,\boldsymbol{l}_1^{\top}\boldsymbol{H}_m\left[\left\{\left(\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\otimes\boldsymbol{I}_m\right\}^{\top}\boldsymbol{K}_{pm}+\left(\boldsymbol{I}_m\otimes\tilde{\boldsymbol{\Lambda}}\right)^{\top}\left(\boldsymbol{I}_m\otimes\tilde{\boldsymbol{\Psi}}^{-1}\right)\right]=$$

$$=\left[\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}\boldsymbol{K}_{mp}\left(\boldsymbol{I}_p\otimes\tilde{\boldsymbol{\Lambda}}^{\top}\right)\boldsymbol{P}_p^{\top}-\hat{\boldsymbol{r}}^{\top}\right]\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\boldsymbol{P}_p\left\{\left(\boldsymbol{I}_p\otimes\tilde{\boldsymbol{\Lambda}}\right)\boldsymbol{K}_{pm}+\left(\tilde{\boldsymbol{\Lambda}}\otimes\boldsymbol{I}_p\right)\right\}+$$

$$+\,\boldsymbol{l}_1^{\top}\boldsymbol{H}_m\left[\left\{\left(\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\otimes\boldsymbol{I}_m\right\}^{\top}\boldsymbol{K}_{pm}+\left(\boldsymbol{I}_m\otimes\tilde{\boldsymbol{\Lambda}}\right)^{\top}\left(\boldsymbol{I}_m\otimes\tilde{\boldsymbol{\Psi}}^{-1}\right)\right].$$

$\hat{\tilde{\boldsymbol{\Lambda}}}$ and $\hat{\boldsymbol{l}}_1$ are, respectively, parts of a solution to the system comprising the two equations

$$\left[\left\{vec\left(\tilde{\boldsymbol{\Lambda}}\right)\right\}^{\top}\boldsymbol{K}_{mp}\left(\boldsymbol{I}_p\otimes\tilde{\boldsymbol{\Lambda}}^{\top}\right)\boldsymbol{P}_p^{\top}-\hat{\boldsymbol{r}}^{\top}\right]\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\boldsymbol{P}_p\left\{\left(\boldsymbol{I}_p\otimes\tilde{\boldsymbol{\Lambda}}\right)\boldsymbol{K}_{pm}+\left(\tilde{\boldsymbol{\Lambda}}\otimes\boldsymbol{I}_p\right)\right\}+$$

$$+\,\boldsymbol{l}_1^{\top}\boldsymbol{H}_m\left[\left\{\left(\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)\otimes\boldsymbol{I}_m\right\}^{\top}\boldsymbol{K}_{pm}+\left(\boldsymbol{I}_m\otimes\tilde{\boldsymbol{\Lambda}}\right)^{\top}\left(\boldsymbol{I}_m\otimes\tilde{\boldsymbol{\Psi}}^{-1}\right)\right]=\boldsymbol{0}^{\top}$$

and

$$vech\left(\boldsymbol{\Xi}\right)-vech\left(\tilde{\boldsymbol{\Lambda}}^{\top}\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\Lambda}}\right)=\boldsymbol{0}\,.$$

Finally, as Klüppelberg and Kuhn (2009) we obtain the residual correlations by difference. Unfortunately there is no guarantee that such a procedure will converge. It must be kept in mind that elements of $\tilde{\boldsymbol{\Lambda}}$ are function of elements of $\tilde{\boldsymbol{\Psi}}$. An iteration procedure must be carried out. The essence of the method is that in each iteration a minimum point of (3.18) is found. This results in a sequence of matrices $\hat{\tilde{\boldsymbol{\Psi}}}^{(1)}$, $\hat{\tilde{\boldsymbol{\Psi}}}^{(2)}, \ldots$, such that

$$D_{QD}\left[\hat{\boldsymbol{r}}, \boldsymbol{r}\left\{\hat{\tilde{\boldsymbol{\theta}}}^{(i)}\right\}\bigg|\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right] < D_{QD}\left[\hat{\boldsymbol{r}}, \boldsymbol{r}\left\{\hat{\tilde{\boldsymbol{\theta}}}^{(i-1)}\right\}\bigg|\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right] .$$

The sequence converges rapidly to a final matrix of estimated $\hat{\tilde{\boldsymbol{\Lambda}}}$ and $\hat{\tilde{\boldsymbol{\Psi}}}$ in few iterations. To start the iterative procedure we choose the initial estimates by ordinary least squares. The iterations run out when

$$D_{QD}\left[\hat{\boldsymbol{r}}, \boldsymbol{r}\left\{\hat{\tilde{\boldsymbol{\theta}}}^{(i-1)}\right\}\bigg|\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right] - D_{QD}\left[\hat{\boldsymbol{r}}, \boldsymbol{r}\left\{\hat{\tilde{\boldsymbol{\theta}}}^{(i)}\right\}\bigg|\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right] < \varepsilon ,$$

where ε is an arbitrarily small positive value, typically 10^{-3}.

To conclude this section we clarify the nonlinear optimization technique used in order to computationally minimize (3.18). The Barzilai and Borwein (1988) gradient method for large scale minimization problems is considered. This method requires few storage locations and very inexpensive computations. Raydan (1993) established global convergence for the strictly convex quadratic case with any number of variables. This result has been extended to the (not necessarily strictly) convex quadratic case by Friedlander, Martínez, and Raydan (1995) to incorporate the method in a box constrained optimization technique. Here, we have chosen Barzilai and Borwein (1988) gradient method instead of `Nelder-Mead`, since it generally performed better than the other scheme in our numerical experiments. Nevertheless, the relative simplicity and superiority of this algorithm does not seem to exclude Heywood cases Heywood (1931); i.e., situations where correlation estimates greater than one are obtained during the estimating process.

Algorithm 1 Estimate $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Psi}}$ and test the copula factor model.

Require:

$\hat{\boldsymbol{r}}$ // estimated correlation matrix
$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}$ // estimated covariance matrix associated with $\hat{\boldsymbol{r}}$
$\boldsymbol{\Xi}$ // characteristic roots of $\hat{\boldsymbol{r}}$ (diagonal matrix)
$p\,,\,m$ // number of manifest and latent variables, respectively
$\hat{\tilde{\boldsymbol{\Lambda}}}^{(0)}\,,\,\hat{\tilde{\boldsymbol{\Psi}}}^{(0)}$ // initial values
n // number of observations
$max.iter$ // number of maximum iterations during the estimation
ε // constant

Ensure:

$$\boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right) := vecp\left(\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^\top\right) = \boldsymbol{P}_p\left(\boldsymbol{I}_p \otimes \tilde{\boldsymbol{\Lambda}}\right)\boldsymbol{K}_{pm}\, vec\left(\tilde{\boldsymbol{\Lambda}}\right)$$

while $abs\left\{D_{QD}^{(i-1)} - D_{QD}^{(i)}\right\} > \varepsilon$ **and** $max.iter$ **do**

$$\min_{vec\left(\tilde{\boldsymbol{\Lambda}}\right)^\top, \boldsymbol{l}_1^\top} \quad D_{QD}\left\{\hat{\boldsymbol{r}}, \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}\right)\Big|\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}\right\} + \\ -2\boldsymbol{l}_1^\top\left(vech\left(\boldsymbol{\Xi}\right) - vech\left[\tilde{\boldsymbol{\Lambda}}^\top\left\{\hat{\tilde{\boldsymbol{\Psi}}}^{(i)}\right\}^{-1}\tilde{\boldsymbol{\Lambda}}\right]\right)$$

$$\hat{\tilde{\boldsymbol{\Psi}}}^{(i+1)} = \mathbf{1}_p - diag\left[\hat{\tilde{\boldsymbol{\Lambda}}}^{(i+1)}\left\{\hat{\tilde{\boldsymbol{\Lambda}}}^{(i+1)}\right\}^\top\right]$$

end while

$H_0 : \boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}_0\right) = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^\top + \tilde{\boldsymbol{\Psi}}$ // testing the structure with m latent variables

$df = p\,(p-1)\,/2 - pm + m\,(m-1)\,/2$ // degrees of freedom

$$T_{\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}} := n\left[\hat{\boldsymbol{r}} - \boldsymbol{r}\left\{\hat{\tilde{\boldsymbol{\theta}}}^{(i+1)}\right\}\right]^\top\left\{\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1} - \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\hat{\mathbf{J}}\left(\hat{\mathbf{J}}^\top\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\hat{\mathbf{J}}\right)^{-1}\hat{\mathbf{J}}^\top\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}^{-1}\right\}\times \\ \times\left[\hat{\boldsymbol{r}} - \boldsymbol{r}\left\{\hat{\tilde{\boldsymbol{\theta}}}^{(i+1)}\right\}\right]$$

if $T_{\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{r}}}} > \chi^2_{df;0.05}$ **then**
    Reject H_0
else
    return $\hat{\tilde{\boldsymbol{\Lambda}}}^{(i+1)}\,,\,\hat{\tilde{\boldsymbol{\Psi}}}^{(i+1)}$
end if

Chapter 4

Extending Copula Structure Analysis: EFGM copulas and maximum pseudo–likelihood estimates

In this chapter, after introducing the EFGM families of copulas, we apply copula structure analysis assuming such models and we derive some theoretical results by analogy with Theorems 3.6 and 3.7. In order to avoid some drawbacks of the moment–based estimation procedure for copula parameters, on which the approach of Klüppelberg and Kuhn (2009) is built, we suggest the use of the celebrated pseudo–maximum likelihood estimator investigated by Genest, Ghoudi, and Rivest (1995). We show that the conditions for exploiting Proposition 3.1 and Theorem 3.4 also hold under these different distribution assumptions and estimator of correlation matrix $\boldsymbol{R}_0$.

The chapter closes with a large Monte Carlo experiment designed to assess the effects of the strength of dependence of the data, despite the sample size, on the power of χ^2 goodness–of–fit test statistic (3.9) via Kendall's τ–based and maximum pseudo–likelihood method, respectively. Finally, an application to real data is performed.

4.1 Copula Structure Analysis assuming EFGM copulas

The so–called EFGM distributions have been considered by Morgenstern (1956) and Gumbel (1958, 1960), further developed by Farlie (1960). However, as Durante and Sempi (2010) point out, the idea of considering such distributions originates in an earlier and, for many years, forgotten work by Eyraud (1936). On account of the fact that EFGM copulas do not allow to model large dependence among the random variables involved, several extensions have been proposed in the literature designed to increase the maximal value of the dependence measures, starting with the works by Farlie (1960). EFGM copulas and their generalizations are ideally suited for various models with small or moderate dependence and do not depend on a particular physical model which may or may not be appropriate in a given situation. A complete survey about these generalized EFGM models of dependence is given in Drouet-Mari and Kotz (2001), where a list of several other references can be also found.

EFGM family of multivariate copulas is constituted by the polynomial copulas associated with the family of EFGM distribution functions (see Kotz, Balakrishnan, and Johnson, 2000, Equation 44.73) and is given by

$$C_{\boldsymbol{x}}\left\{F_1\left(x_1\right),\ldots,F_p\left(x_p\right)\,;\,\boldsymbol{\alpha}\right\}=C_{\boldsymbol{x}}\left(u_1,\ldots,u_p\,;\,\boldsymbol{\alpha}\right)=\prod_{j=1}^{p}u_j\times$$
$$\times\left\{1+\sum_{j=2}^{p}\sum_{1\le i_1<i_2<\ldots<i_j\le p}\alpha_{i_1 i_2\ldots i_j}\prod_{k=1}^{j}\left(1-u_{i_k}\right)\right\},$$

where the total number of the suitable parameters $\alpha_{i_1 i_2 \ldots i_j}$ is 2^p-p-1.

It can be shown that any EFGM copula is absolutely continuous with density given by

$$c_{\boldsymbol{x}}\left(u_1, \ldots, u_p\,;\,\boldsymbol{\alpha}\right) = 1 + \sum_{j=2}^{p} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le p} \alpha_{i_1 i_2 \ldots i_j} \prod_{k=1}^{j} \left(1 - 2\,u_{i_k}\right)\,.$$

As a consequence, the parameters $\alpha_{i_1 i_2 \ldots i_j}$ have to satisfy the following restrictions,

$$1 + \sum_{j=2}^{p} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le p} \alpha_{i_1 i_2 \ldots i_j} \prod_{k=1}^{j} \left(1 - 2\,u_{i_k}\right) \ge 0\,.$$

Generally, however, each parameter must meet the condition $\left|\alpha_{i_1 i_2 \ldots i_j}\right| \le 1$ (Cambanis, 1977).

For the bivariate and trivariate cases, respectively, EFGM copulas have the following explicit expressions:

$$C_{\boldsymbol{x}}\left(u_1, u_2\,;\,\alpha_{1,2}\right) = u_1 u_2 \left\{1 + \alpha_{1,2}\left(1 - u_1\right)\left(1 - u_2\right)\right\} \tag{4.1}$$

and

$$\begin{aligned} C_{\boldsymbol{x}}\left(u_1, u_2, u_3\,;\,\boldsymbol{\alpha}\right) = u_1 u_2 u_3 \{1 &+ \alpha_{1,2}\left(1 - u_1\right)\left(1 - u_2\right) + \\ &+ \alpha_{1,3}\left(1 - u_1\right)\left(1 - u_3\right) + \\ &+ \alpha_{2,3}\left(1 - u_2\right)\left(1 - u_3\right) + \\ &+ \alpha_{1,2,3}\left(1 - u_1\right)\left(1 - u_2\right)\left(1 - u_3\right)\}\,. \end{aligned} \tag{4.2}$$

More recent investigations for extending EFGM copulas based on the construction of copulas that is quadratic in one variable are provided by Quesada-Molina and Rodríguez-Lallena (1995) and Rodriguez-Lallena and Úbeda-Flores (2009). If we see the 3–dimensional copula given by (4.2) as a copula of the form proposed in Rodriguez-Lallena and Úbeda-Flores (2009), then

$$C_{\boldsymbol{x}}\left(u_1, u_2, u_3\,;\,\boldsymbol{\alpha}\right) = u_3\, D_{\boldsymbol{x}}\left(u_1, u_2\,;\,\alpha_{1,2}\right) + u_3\left(1-u_3\right)\kappa\left(u_1, u_2; \alpha_{1,3}, \alpha_{2,3}, \alpha_{1,2,3}\right)\,,$$

where $D_{\boldsymbol{x}}\left(u_1, u_2\,;\,\alpha_{1,2}\right)$ is a 2–dimensional EFGM copula of the form (4.1) and

$$\begin{aligned}\kappa\left(u_1, u_2; \alpha_{1,3}, \alpha_{2,3}, \alpha_{1,2,3}\right) = u_1 u_2 \{&\left(\alpha_{1,3}+\alpha_{1,2,3}\right)\left(1-u_1\right)+\\ &+\left(\alpha_{2,3}+\alpha_{1,2,3}\right)\left(1-u_2\right)+\\ &-\alpha_{1,2,3}\left(1-u_1 u_2\right)\}\end{aligned}$$

satisfies the condition

$$\kappa\left(u_1, 0; \alpha_{1,3}, \alpha_{2,3}, \alpha_{1,2,3}\right) = \kappa\left(0, u_2; \alpha_{1,3}, \alpha_{2,3}, \alpha_{1,2,3}\right) = \kappa\left(1, 1; \alpha_{1,3}, \alpha_{2,3}, \alpha_{1,2,3}\right) = 0\,.$$

In order to estimate and test a copula structure model with EFGM copulas, recalling the methodology by Klüppelberg and Kuhn (2009), we firstly need to identify a moment–based estimator of copula correlation matrix $\boldsymbol{R}_0$ of $\boldsymbol{x} \in \mathbb{R}^p$ via Kendall's τ–matrix, denoted by $\boldsymbol{T}$. Besides elliptical copulas we do not recognize other families of copulas whose dependence parameter vector coincides with Pearson's linear correlation matrix. Therefore, we begin with the research of a relationship between correlation matrix and copula dependence parameters. The latter step concerns the typical role that copulas play in concordance and measure of association as Kendall's τ. Finally, we just have to merge this two relationships that yield a direct link between correlation matrix and Kendall's τ. EFGM copulas indeed represent a suitable situation where we can extend the work of Klüppelberg and Kuhn (2009) to other families of copulas. Dependence properties of this family are closely related with linear correlation coefficients although *a priori* the pivotal parameter of this bivariate family is not obviously associated with this concept. Without lack of generality we henceforth consider the simplest among the multivariate EFGM distributions with univariate absolutely continuous marginals $F_1(x_1), \ldots, F_p(x_p)$ discussed by Cambanis (1991); i.e., those of the form

$$C_{\boldsymbol{x}}(u_1, \ldots, u_p\,; \boldsymbol{\alpha}) = \prod_{j=1}^{p} u_j \left\{ 1 + \sum_{1 \leq i_1 < i_2 \leq p} \alpha_{i_1 i_2} \prod_{k=1}^{2} (1 - u_{i_k}) \right\}, \tag{4.3}$$

where $\alpha_{i_1 i_2 i_3} = 0$, $\alpha_{i_1 i_2 i_3 i_4} = 0$, and so on.

The parameters of (4.3) are the $p^{\star\star}$ constants $\alpha_{i_1 i_2}$, $1 \leq i_1 < i_2 \leq p$, whose admissible values are determined by the 2^p inequalities

$$1 + \sum_{1 \leq i_1 < i_2 \leq p} \alpha_{i_1 i_2} \prod_{k=1}^{2} (1 - 2\,u_{i_k}) \geq 0\,.$$

Multivariate distributions of the form (4.3) are uniquely determined by the bivariate margin of the form (4.1). Also all their marginals (of order $p-1, \ldots, 2$) are of the same type.

Here we assume that the bivariate random variable $(x_i, x_j)^\top$, $i \neq j$, has joint distribution function (4.1), finite mean $(\mu_i, \mu_j)^\top$ and positive and finite variance $(\sigma_{i,i}, \sigma_{j,j})^\top$. Following Johnson and Kotz (1977) and Schucany, Parr, and Boyer (1978), we are immediately able to handle the relationship between Pearson's bivariate linear correlation coefficient and copula dependence parameter concerning the i, j–margin of EFGM copulas; i.e.,

$$\rho_{i,j} = \frac{\alpha_{i,j}\delta_{2;i}\delta_{2;j}}{\sqrt{\sigma_{i,i}\sigma_{j,j}}}\,, \quad i \neq j\,,$$

where

$$\begin{aligned}\delta_{2;j} &= \int_0^1 F_j^{-1}(u)\,(2u-1)\,du = \\ &= \int_0^1 uF_j^{-1}(u)\,du - \int_0^1 F_j^{-1}(u)\,(1-u)\,du =\end{aligned}$$

$$= \frac{1}{2} \left\{ E\left(X_{j;2:2}\right) - E\left(X_{j;1:2}\right) \right\}$$

denotes the second L–moment of x_j and $x_{j;k:r}$ the k–th order statistic (k–th smallest value) in an independent sample of size r drawn from the distribution of x_j (see Appendix B).

The case $\alpha_{i,j} = 1$ and $\alpha_{i,j} = -1$ represent the maximal degrees of positive and negative dependence, respectively, allowed in the bivariate EFGM family of copulas. Cambanis (1991, Proposition 1) proves that among all bivariate distributions (4.1) with absolutely continuous marginals, the ones with uniform margins over $(0, 1)$ have the broadest range of correlation values; i.e., $|\rho_{i,j}| \leq 1/3$. Similar conclusions are provided by Schucany, Parr, and Boyer (1978, Theorem 1).

So far we have considered the relationship between the bivariate copula parameter $\alpha_{i,j}$ and the related bivariate Pearson's linear correlation coefficient $\rho_{i,j}$. Since $\tau_{i,j} = 4\,E\left\{C_{\boldsymbol{x}}\left(u_i, u_j\,;\,\alpha_{i,j}\right)\right\} - 1 = 2/9\,\alpha_{i,j}$, where $C_{\boldsymbol{x}}$ is of the form (4.1), we obtain the relationship we were talking about; i.e.,

$$\rho_{i,j} = \frac{9}{2} \frac{\tau_{i,j}\delta_{2;i}\delta_{2;j}}{\sqrt{\sigma_{i,i}\sigma_{j,j}}}\,, \quad i \neq j\,, \tag{4.4}$$

We are now ready to adapt Theorem 3.6 in order to define a moment–based estimator of population linear correlation matrix $\boldsymbol{R}_0$ when $\boldsymbol{x}$ is a vector of random variables with EFGM copulas.

Theorem 4.1 *Let $\boldsymbol{x} \in \mathbb{R}^p$ be a vector of random variables with finite first and second moments. Let the EFGM copula of the form (4.3) be the distribution function associated with $\boldsymbol{x}$.*

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ an independent sequence in $\mathbb{R}^p$ identically distributed according to $\boldsymbol{x}$.

Define $\boldsymbol{r}_0 := vecp\left(\boldsymbol{R}_0\right)$, where $\boldsymbol{R}_0$ denotes the population correlation matrix of $\boldsymbol{x}$.

Let $\hat{\boldsymbol{r}}_\tau$ be the estimated vector of non–duplicated and non–fixed elements of $\boldsymbol{R}_0$ via estimated Kendall's τ–matrix $\hat{\boldsymbol{T}}$,

$$\hat{\boldsymbol{r}}_\tau := \frac{9}{2}\, vecp \left\{ \hat{\boldsymbol{\delta}}\, diag\left(\hat{\boldsymbol{\Sigma}}\right)^{-1/2} \hat{\boldsymbol{T}}\, diag\left(\hat{\boldsymbol{\Sigma}}\right)^{-1/2} \hat{\boldsymbol{\delta}} \right\} ,$$

where $\hat{\boldsymbol{\Sigma}}$ is an estimator of the covariance matrix $\boldsymbol{\Sigma}_0$ of $\boldsymbol{x}$ and $\hat{\boldsymbol{\delta}}$ is a diagonal matrix with elements (B.4) *defined in Appendix B.*

Then, as $n \to \infty$,

$$n^{1/2}\left(\hat{\boldsymbol{r}}_\tau - \boldsymbol{r}_0\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_\tau\right) , \tag{4.5}$$

where $\boldsymbol{\Sigma}_\tau := \left[\sigma^\tau_{ij,kl}\right]_{1 \le i \ne j, k \ne l \le p}$ and

$$\sigma^\tau_{ij,kl} = \frac{81 \delta_{2;i}\delta_{2;j}\delta_{2;k}\delta_{2;l}\left(\tau_{ij,kl} - \tau_{i,j}\tau_{k,l}\right)}{\sqrt{\sigma_{i,i}\sigma_{j,j}\sigma_{k,k}\sigma_{l,l}}} ,$$

with $\tau_{i,j}$ and $\tau_{ij,kl}$ given in (3.13) *and* (3.14)*, respectively.*

Proof

The proof follows the arguments used by Klüppelberg and Kuhn (2009) to prove their Theorem 3.

Define $\hat{\boldsymbol{t}} := vecp\left(\hat{\boldsymbol{T}}\right)$ and $\boldsymbol{t} := vecp\left(\boldsymbol{T}\right)$, the vectors of patterned matrix $\hat{\boldsymbol{T}}$ and $\boldsymbol{T}$, respectively. Since $\hat{\boldsymbol{t}}$ is a vector of U–statistics estimators with a non–zero first component in the H–decomposition having expectations $\boldsymbol{t}$ and kernels $t_{i,j}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) := sgn\left(x_{1,i} - x_{2,i}\right)\left(x_{1,j} - x_{2,j}\right)$ of degree equal to 2, Lee (1990, Theorem 2, Section 3.2.1) applies:

$$n^{1/2}\left(\hat{\boldsymbol{t}} - \boldsymbol{t}\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\boldsymbol{0}, 4\boldsymbol{\Sigma}_{\boldsymbol{t}}\right) ,$$

where $\boldsymbol{\Sigma}_{\boldsymbol{t}} := \left[\sigma^{\boldsymbol{t}}_{ij,kl}\right]_{1 \le i \ne j, k \ne l \le p}$, and constant 4 comes from the squared kernel's degree. The covariance structure is stated in Lee (1990, Theorem 1, Section 1.4) together with the remark about the consequence of this result,

$$\begin{aligned} \sigma^{\boldsymbol{t}}_{ij,kl} &= cov\left\{t_{i,j}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right), t_{k,l}\left(\boldsymbol{x}_1, \boldsymbol{x}_3\right)\right\} = \\ &= E\left\{t_{i,j}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) t_{k,l}\left(\boldsymbol{x}_1, \boldsymbol{x}_3\right)\right\} - E\left\{t_{i,j}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)\right\} E\left\{t_{k,l}\left(\boldsymbol{x}_1, \boldsymbol{x}_3\right)\right\} = \\ &= E\left\{t_{i,j;1}\left(\boldsymbol{x}_1\right) t_{k,l;1}\left(\boldsymbol{x}_1\right)\right\} - \tau_{i,j}\tau_{k,l} = \end{aligned}$$

$$= E\left[E\left\{t_{i,j}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)|\boldsymbol{x}_1\right\} E\left\{t_{k,l}\left(\boldsymbol{x}_1, \boldsymbol{x}_3\right)|\boldsymbol{x}_1\right\}\right] - \tau_{i,j}\tau_{k,l} =$$
$$= \tau_{ij,kl} - \tau_{i,j}\tau_{k,l} \,,$$

where $t_{i,j;1}\left(\boldsymbol{x}_1\right) := E\left\{t_{i,j}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)|\boldsymbol{x}_1\right\}$ (Lee, 1990, Equation 1, Section 1.3).

Since the correlation matrix $\boldsymbol{R}_0$ can be seen as a function of Kendall's τ–matrix $\boldsymbol{T}$,

$$\begin{aligned}
\hat{\boldsymbol{r}}_\tau &= \frac{9}{2}\, vecp\left\{\hat{\boldsymbol{\delta}}\, diag\left(\hat{\boldsymbol{\Sigma}}\right)^{-1/2} \hat{\boldsymbol{T}}\, diag\left(\hat{\boldsymbol{\Sigma}}\right)^{-1/2} \hat{\boldsymbol{\delta}}\right\} = \\
&= \frac{9}{2}\boldsymbol{P}_p\left[\left\{diag\left(\hat{\boldsymbol{\Sigma}}\right)^{-1/2}\hat{\boldsymbol{\delta}}\right\} \otimes \left\{diag\left(\hat{\boldsymbol{\Sigma}}\right)^{-1/2}\hat{\boldsymbol{\delta}}\right\}\right]\boldsymbol{Q}_p\, vecp\left(\hat{\boldsymbol{T}}\right) = \\
&= \frac{9}{2}\boldsymbol{P}_p\,\hat{\boldsymbol{\Delta}}\,\boldsymbol{Q}_p\,\hat{\boldsymbol{t}}\,,
\end{aligned}$$

where $\hat{\boldsymbol{\Delta}}$ is a $p^2 \times p^2$ diagonal matrix with estimated elements of $\boldsymbol{\Delta}$,

$$\begin{aligned}
\hat{\boldsymbol{\Delta}} = diag\Bigg(&\frac{\hat{\delta}^2_{2;1}}{\hat{\sigma}_{1,1}}, \frac{\hat{\delta}_{2;1}\hat{\delta}_{2;2}}{\sqrt{\hat{\sigma}_{1,1}\hat{\sigma}_{2,2}}}, \ldots, \frac{\hat{\delta}_{2;1}\hat{\delta}_{2;p}}{\sqrt{\hat{\sigma}_{1,1}\hat{\sigma}_{p,p}}}, \\
&\frac{\hat{\delta}_{2;1}\hat{\delta}_{2;2}}{\sqrt{\hat{\sigma}_{1,1}\hat{\sigma}_{2,2}}}, \frac{\hat{\delta}^2_{2;2}}{\hat{\sigma}_{2,2}}, \ldots, \frac{\hat{\delta}_{2;2}\hat{\delta}_{2;p}}{\sqrt{\hat{\sigma}_{2,2}\hat{\sigma}_{p,p}}}, \\
&\ldots \\
&\frac{\hat{\delta}_{2;1}\hat{\delta}_{2;p}}{\sqrt{\hat{\sigma}_{1,1}\hat{\sigma}_{p,p}}}, \frac{\hat{\delta}_{2;2}\hat{\delta}_{2;p}}{\sqrt{\hat{\sigma}_{2,2}\hat{\sigma}_{p,p}}}, \ldots, \frac{\hat{\delta}^2_{2;p}}{\hat{\sigma}_{p,p}}\Bigg)\,,
\end{aligned}$$

the transition or duplication matrix $\boldsymbol{Q}_p$ and its left inverse $\boldsymbol{P}_p$ are defined in Appendix A, and the Jacobian matrix is given by

$$\mathsf{J} = \frac{9}{2}\frac{\partial}{\partial \boldsymbol{t}^\top}\boldsymbol{P}_p\,\boldsymbol{\Delta}\,\boldsymbol{Q}_p\,\boldsymbol{t} = \frac{9}{2}\boldsymbol{P}_p\,\boldsymbol{\Delta}\,\boldsymbol{Q}_p\,,$$

by the multivariate delta method (see Lehmann and Casella, 1998, Theorem 8.22, Section 1.8)

$$n^{1/2}\left(\hat{\boldsymbol{r}}_\tau - \boldsymbol{r}_0\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\mathbf{0}, 4\,\mathbf{J}\,\boldsymbol{\Sigma}_t\,\mathbf{J}^\top\right) .$$

■

We remark the fact that in this case we can not assume the non–existence of the moments in order to satisfy a copula structure model. As (4.4) shows, we need finite first and second moments of observed variables $\boldsymbol{x}$. Without this conditions we are not able to define a relationship between Pearson's linear correlation coefficients and Kendall's tau and L–moments of $\boldsymbol{x}$ would not exist (see Theorem B.1 in Appendix B). This is in contrast with the purpose to avoid problems of non–existing moments declared by Klüppelberg and Kuhn (2009). However, higher order moments do not need to be computed here as Browne (1984) does.

To ensure conditions for Theorem 3.4 we finally have to prove the consistency of the estimator of the asymptotic covariance matrix $\boldsymbol{\Sigma}_\tau$ in (4.5). Before discussing the consistency of the estimator of the asymptotic covariance matrix, we recall one result due to Slutsky.

Lemma 4.1 (Ferguson, 1996, Theorem 6', Section 6) *Assume $\boldsymbol{Z}_n \in \mathbb{R}^p$, $\boldsymbol{Z}_n \xrightarrow{p} \boldsymbol{Z}$, and $\boldsymbol{\gamma} : \mathbb{R}^p \to \mathbb{R}^d$ is such that pr $\{\boldsymbol{Z} \in \mathscr{C}(\boldsymbol{\gamma})\} = 1$, where $\mathscr{C}(\boldsymbol{\gamma})$ is the continuity set of $\boldsymbol{\gamma}$. Then,*

(a) $\boldsymbol{\gamma}(\boldsymbol{Z}_n) \xrightarrow{p} \boldsymbol{\gamma}(\boldsymbol{Z})$;

(b) *if* $\boldsymbol{Z}_n - \boldsymbol{Y}_n \xrightarrow{p} \mathbf{0}$, $\boldsymbol{Y}_n \xrightarrow{p} \boldsymbol{Z}$;

(c) *if* $\boldsymbol{Y}_n \xrightarrow{p} \boldsymbol{Y}$, $\binom{\boldsymbol{Z}_n}{\boldsymbol{Y}_n} \xrightarrow{p} \binom{\boldsymbol{Z}}{\boldsymbol{Y}}$.

Theorem 4.2 *Under the assumptions of Theorem 4.1,*

$$vech\left(\hat{\boldsymbol{\Sigma}}_\tau\right) = 81\,vech\left(\boldsymbol{P}_p\,\hat{\boldsymbol{\Delta}}\,\boldsymbol{Q}_p\,\hat{\boldsymbol{\Sigma}}_t\,\boldsymbol{Q}_p^\top\,\hat{\boldsymbol{\Delta}}\,\boldsymbol{P}_p^\top\right)$$

is a consistent estimator of the asymptotic covariance matrix in (4.5).

Proof

Let $\boldsymbol{H}_p$ represent the left inverse of the transition or duplication matrix $\boldsymbol{G}_p$.

$$vech\left(\hat{\boldsymbol{\Sigma}}_\tau\right) = 81\, \boldsymbol{H}_{p^{\star\star}} \left\{\left(\boldsymbol{P}_p\, \hat{\boldsymbol{\Delta}}\, \boldsymbol{Q}_p\right) \otimes \left(\boldsymbol{P}_p\, \hat{\boldsymbol{\Delta}}\, \boldsymbol{Q}_p\right)\right\} \boldsymbol{G}_{p^{\star\star}}\, vech\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}}\right).$$

From the part **(a)** of Lemma 4.1, $vech\left(\hat{\boldsymbol{\Sigma}}_\tau\right)$ is a consistent estimator of $vech\left(\boldsymbol{\Sigma}_\tau\right)$ if $vech\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}}\right)$ converges in probability to $vech\left(\boldsymbol{\Sigma}_{\boldsymbol{t}}\right)$.

We rewrite $vech\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}}\right)$ as $vech\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t};ij,kl}\right) - vech\left(\hat{\boldsymbol{t}}\,\hat{\boldsymbol{t}}^\top\right)$, where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t};ij,kl} := \left[\hat{\tau}_{ij,kl}\right]_{1\leq i\neq j,k\neq l\leq p}$. Klüppelberg and Kuhn (2009) establish the consistency of the estimator $vech\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t};ij,kl}\right)$ since elements $\hat{\tau}_{ij,kl}$ can be regarded as a linear combination of U–statistics. On the other hand, $\hat{\boldsymbol{t}} \xrightarrow{p} \boldsymbol{t}$ for the Law of Large Numbers. Then, from the combination of parts **(c)** and **(a)** of Lemma 4.1 $vech\left(\hat{\boldsymbol{t}}\,\hat{\boldsymbol{t}}^\top\right) \xrightarrow{p} vech\left(\boldsymbol{t}\,\boldsymbol{t}^\top\right)$.

Therefore, $vech\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t};ij,kl}\right) - vech\left(\hat{\boldsymbol{t}}\,\hat{\boldsymbol{t}}^\top\right) \xrightarrow{p} vech\left(\boldsymbol{\Sigma}_{\boldsymbol{t};ij,kl}\right) - vech\left(\boldsymbol{t}\,\boldsymbol{t}^\top\right)$ directly follows from the application of parts **(c)** and **(a)** of Lemma 4.1.

■

4.2 Copula Structure Analysis by maximizing the pseudo–likelihood

The copula structure analysis proposed by Klüppelberg and Kuhn (2009) represent a simple and convincing method to avoid typical problems and limits of traditional approach to covariance or correlation structure analysis due to Browne (1984). Nevertheless, we note that the copula–based method requires a closed–form relationship between Kendall's τ and copula parameters and, consequently, Pearson's linear correlation coefficients. There exist some families of copulas for which this condition does not hold; e.g., the so-called Joe family of copulas (Joe, 1993), Galambos family of copulas (Galambos, 1975), and Hüsler–Reiss family of copulas (Hüsler and Reiss, 1989). Moreover, Kendall's τ does not depend on the magnitude of the data and it neglects large and small values (Mikosch, 2006). Unless r is the generator of the multivariate normal distribution, $\tau_{i,j} = \rho_{i,j} = 0$ never corresponds to independence. Therefore, we propose to adopt the maximum pseudo–likelihood estimator of copula parameters instead of the moment–based estimator in order to only focus the attention on the direct relationship between the copula parameters and Pearson's linear correlation coefficients.

Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$ be a d–dimensional dependence parameter from a joint distribution function $F_{\boldsymbol{x}}(x_1, \ldots, x_p\,; \boldsymbol{\alpha})$ with associated copula $C_{\boldsymbol{x}}(u_1, \ldots, u_p\,; \boldsymbol{\alpha})$ and density $c_{\boldsymbol{x}}(\cdot\,; \boldsymbol{\alpha})$. If the true copula is assumed to belong to a parametric family $\mathcal{C} = \left\{C_{\boldsymbol{x}}(\cdot\,; \boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \mathbb{A} \subseteq \mathbb{R}^d\right\}$, consistent and asymptotically normally distributed estimates of the parameter $\boldsymbol{\alpha}$ can be obtained through maximum likelihood methods. There are mainly two ways to achieve this: a fully parametric method and a semi–parametric method. The first method relies on the assumption of parametric univariate marginal distributions. Each parametric margin is then plugged in the full likelihood and this objective function is maximized with respect to the parameter $\boldsymbol{\alpha}$. The resulting estimate for $\boldsymbol{\alpha}$ would then be margin–dependent. Alternatively and without any parametric assumptions for margins, the univariate empirical distribution functions can be plugged in the likelihood to yield a semi–parametric method. When nonparametric estimates are contemplated for the marginals, inference about the dependence parameter $\boldsymbol{\alpha}$ will be margin–free. These two commonly used methods are detailed in Genest, Ghoudi, and Rivest (1995) and Shih and Louis (1995). Since the result of the first method depends on the right specification of all margins, this may induce too severe constraints, and this aspect lessens the interest of working with copulas. The semi–parametric estimation procedure where margins are left unspecified does not suffer from this inconvenient feature, but suffers from a loss of efficiency (see Genest, Ghoudi, and Rivest, 1995, Equation 3). In what follows, we shall refer to this as the maximum pseudo–likelihood estimation method. As Fermanian and Scaillet (2004) point out commenting their simulation studies designed to asses the potential impact of misspecified margins on the estimation of copula parameters, "*if the researcher has any doubt about the correct modeling of the margins, there is probably little to loose but lots* *to gain from shifting towards a semi–parametric approach*". In practice, it is however impossible to be certain that the marginal distribution functions have been correctly specified, even if appropriate univariate goodness–of–fit tests are used. As it may be argued that the copula parameters estimates should not be affected by the choice of the marginal distribution functions, many authors advocate the use of the maximum pseudo–likelihood estimator. Note further that there is no guarantee in both cases that the specified copula is indeed the true one. If not the asymptotic variance should be modified adequately (for inference under misspecified copulas see Cebrian, Denuit, and Scaillet, 2004). For the rest of this work we shall assume that the selected copula corresponds to the true one.

Maximum pseudo–likelihood estimation is not the only semi–parametric approach used in practice. In the one–parameter case, two maybe even more popular methods for estimating the copula parameter are based on the inversion of Kendall's tau and Spearman's rho (see e.g. Genest and Rivest, 1993, and the references therein). They are frequently referred to as methods of moments. Two other semiparametric approaches were investigated by Tsukahara (2005), namely rank approximate Z–estimation and minimum–distance estimation. In his simulation study, both methods were found overall to lead to a higher estimated mean square error than the maximum pseudo–likelihood estimator. More recently, Chen, Fan, and Tsyrennikov (2006) have studied a version of the maximum pseudo–likelihood estimator in which the unknown marginal densities are approximated by linear combinations of finite–dimensional known basis functions with increasing complexity, called sieves. They showed that the resulting estimator is asymptotically semiparametrically efficient for copula parameter provided that additional smoothness conditions are satisfied. The study of the finite–sample performance of the method for sample size $n = 400$ revealed that this approach performs significantly better than the standard maximum pseudo–likelihood estimator when one of the marginal distribution functions is known. This advantage does not seem to hold anymore when all the marginals are unknown.

An extensive Monte Carlo study carried out by Kojadinovic and Yan (2010b) however shows that the maximum pseudo–likelihood estimator appears as the best choice in terms of mean square error in all situations except for small and weakly dependent samples. Among the two method–of–moment estimators, the one based on Kendall's tau appears overall significantly better than that based on Spearman's rho. From a computational perspective, estimation based on Kendall's tau is generally faster than maximum pseudo–likelihood estimation, except maybe for very large samples. As a consequence, we focus our attention on asymptotic properties of the maximum pseudo–likelihood estimator and we exploit them for giving an alternative to the kendall's tau–based estimator of $\boldsymbol{R}_0$ performed by Klüppelberg and Kuhn (2009).

Let $\{(x_{a1}, \ldots, x_{ap}) : a = 1, \ldots, n\}$ represent a random *IID* sample from $F_{\boldsymbol{x}}(x_1, \ldots, x_p\,; \boldsymbol{\alpha}) = C_{\boldsymbol{x}}\{F_1(x_1), \ldots, F_p(x_p)\,; \boldsymbol{\alpha}\}$. The estimator $\hat{\boldsymbol{\alpha}}_{MPL}$ of $\boldsymbol{\alpha}$ is obtained as a solution of the system

$$\sum_{a=1}^{n} \frac{\partial}{\partial \alpha_r} \log\{c_{\boldsymbol{x}}(\hat{u}_{a,1}, \ldots, \hat{u}_{a,p}\,; \boldsymbol{\alpha})\} = 0 \quad (1 \leq r \leq d)\ ,$$

where $(\hat{u}_{a,1}, \ldots, \hat{u}_{a,p})^\top$ are pseudo–observations computed from the $(x_{a,1}, \ldots, x_{a,p})^\top$ by $\hat{u}_{a,j} = R_{a,j}/(n+1)$, with $R_{a,j}$ being the rank of $x_{a,j}$ among $(x_{1,j}, \ldots, x_{n,j})^\top$. Notice that the maximum pseudo–likelihood method could be seen as a version of the inference function for margins method by Joe (1997, Chapter 10) in which the marginal distribution functions are estimated nonparametrically. Indeed, it can be checked that $\hat{u}_{a,j} = n\,\hat{F}_j(x_{a,j})/(n+1)$, where $\hat{F}_j$ is the empirical distribution function computed from $(x_{1,j}, \ldots, x_{n,j})^\top$. Note that the scaling factor $n/(n+1)$ is classically introduced to avoid numerical problems at the boundary of $[0,1]^p$. The proposed semi–parametric estimator is then obtained as a solution of a pseudo log–likelihood equations system. In what follows, we shall refer to $\hat{\boldsymbol{\alpha}}_{MPL}$ as the maximum pseudo–likelihood estimator of copula parameters.

Under the standard regularity conditions for consistency of multidimensional maximum likelihood estimators (see, for instance, Lehmann, 1983, Section 6.4) and regularity conditions for multivariate rank statistics proposed by Ruymgaart, Shorack, and Zwet (1972), Ruymgaart (1974), and Rüschendorf (1976), Genest, Ghoudi, and Rivest (1995, Section 4) show that the estimator $\hat{\boldsymbol{\alpha}}_{MPL}$ is consistent and asymptotically normal,

$$n^{1/2}\left(\hat{\boldsymbol{\alpha}}_{MPL}-\boldsymbol{\alpha}\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\mathbf{0}, \boldsymbol{I}\left(\boldsymbol{\alpha}\right)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\, \boldsymbol{I}\left(\boldsymbol{\alpha}\right)^{-1}\right) \quad \text{as } n \to \infty\,, \tag{4.6}$$

where $\boldsymbol{I}\left(\boldsymbol{\alpha}\right)$ is the Fisher information matrix associated with $c_{\boldsymbol{\alpha}}$,

$$\boldsymbol{I}\left(\boldsymbol{\alpha}\right) = -E\left(\frac{\partial^2}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} \log\left[c_{\boldsymbol{\alpha}}\left\{F_1\left(x_1\right), \ldots, F_p\left(x_p\right)\right\}\right]\right)\,,$$

and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ is the covariance matrix of the d–dimensional random vector whose r–th component is given by

$$\frac{\partial}{\partial \alpha_r} \log\left[c_{\boldsymbol{x}}\left\{F_1\left(x_{a,1}\right), \ldots, F_p\left(x_{a,p}\right)\,;\, \boldsymbol{\alpha}\right\}\right] + \sum_{j=1}^{p} \mathcal{W}_{j,r}\left(x_j\right) \quad \left(1 \leq r \leq d\right)\,,$$

with

$$\mathcal{W}_{j,r}\left(x_j\right) = \int \mathbb{I}\left\{F_j\left(x_j\right) \leq u_j\right\} \frac{\partial^2}{\partial \alpha_r \partial u_j} \log\left\{c_{\boldsymbol{x}}\left(u_1, \ldots, u_p\,;\, \boldsymbol{\alpha}\right)\right\} dC_{\boldsymbol{x}}\left(u_1, \ldots, u_p\,;\, \boldsymbol{\alpha}\right)\,.$$

Following Genest and Favre (2007), let us assume that the original sample $\left(x_{1,1}, \ldots, x_{1,p}\right), \ldots, \left(x_{n,1}, \ldots, x_{n,p}\right)$ have been relabeled so that $x_{1,1} < x_{2,1} < \ldots < x_{n,1}$. As a consequence one then has $R_{1,1} = 1, \ldots, R_{n,1} = n$. Moreover, let us denote by $L\left(\boldsymbol{\alpha}, u_1, \ldots, u_p\right)$ the log–likelihood $\log\left\{c_{\boldsymbol{x}}\left(u_1, \ldots, u_p\,;\, \boldsymbol{\alpha}\right)\right\}$ and by L_{α_r}, L_{u_j} and $L_{\alpha_r \alpha_{r'}}$ the derivatives of L with respect to α_r, u_j and both α_r and $\alpha_{r'}$, respectively. An efficient way of estimating the information matrix $\boldsymbol{I}\left(\boldsymbol{\alpha}\right)$ is given by the Hessian matrix associated with $L\left(\boldsymbol{\alpha}, u_1, \ldots, u_p\right)$ at $\hat{\boldsymbol{\alpha}}_{MPL}$, namely, the $d \times d$ matrix whose $\left(r, r'\right)$ entry is given by

$$-\frac{1}{n} \sum_{a=1}^{n} L_{\alpha_r \alpha_{r'}}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{a}{n+1}, \hat{u}_{a,2}, \ldots, \hat{u}_{a,p}\right)\,. \tag{4.7}$$

The estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ is represented by the sample covariance matrix of the variables $\left(\boldsymbol{\mathcal{M}}_1, \ldots, \boldsymbol{\mathcal{M}}_d\right)^\top$, for which the pseudo–observations are

$$
\begin{aligned}
\hat{\mathcal{M}}_{a,r} = & \, L_{\alpha_r}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{a}{n+1}, \hat{u}_{a,2}, \ldots, \hat{u}_{a,p}\right) + \\
& - \frac{1}{n}\sum_{b=1}^{n} L_{\alpha_r}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{b}{n+1}, \hat{u}_{b,2}, \ldots, \hat{u}_{b,p}\right) \times \\
& \quad \times L_{u_1}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{b}{n+1}, \hat{u}_{b,2}, \ldots, \hat{u}_{b,p}\right) + \\
& - \frac{1}{n}\sum_{R_{b,2}\geq R_{a,2}} L_{\alpha_r}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{b}{n+1}, \hat{u}_{b,2}, \ldots, \hat{u}_{b,p}\right) \times \\
& \quad \times L_{u_2}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{b}{n+1}, \hat{u}_{b,2}, \ldots, \hat{u}_{b,p}\right) + \\
& \ldots \\
& - \frac{1}{n}\sum_{R_{b,p}\geq R_{a,p}} L_{\alpha_r}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{b}{n+1}, \hat{u}_{b,2}, \ldots, \hat{u}_{b,p}\right) \times \\
& \quad \times L_{u_p}\left(\hat{\boldsymbol{\alpha}}_{MPL}, \frac{b}{n+1}, \hat{u}_{b,2}, \ldots, \hat{u}_{b,p}\right) ,
\end{aligned}
$$

for $a \in (1, \ldots, n)$ and $r \in (1, \ldots, d)$.

Unlike Klüppelberg and Kuhn (2009), our approach to correlation structure analysis is based on the maximum pseudo–likelihood estimator of $\boldsymbol{\alpha}$.

Theorem 4.3 *Let $\boldsymbol{x} \in \mathbb{R}^p$ be a vector of random variables with absolutely continuous copula C indexed by a d–dimensional parameter $\boldsymbol{\alpha}_0$. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ an independent sequence in $\mathbb{R}^p$ identically distributed according to $\boldsymbol{x}$.*

Let $\boldsymbol{\gamma}(\boldsymbol{\alpha}_0)$ be a vector of real–valued, invertible, and continuously differentiable in a neighborhood $\mathcal{N}_{\boldsymbol{\alpha}_0}$ of the parameter vector $\boldsymbol{\alpha}_0$ functions such that

$$\boldsymbol{\gamma}(\boldsymbol{\alpha}_0) = \boldsymbol{r}_0 := vecp(\boldsymbol{R}_0) \, , \tag{4.8}$$

where $\boldsymbol{R}_0$ is the population correlation matrix of $\boldsymbol{x}$ and $\boldsymbol{r}_0$ its patterned vectorized version.

Let the Jacobian matrix $\mathbf{J}_{\boldsymbol{\alpha}_0} = \mathbf{J}(\boldsymbol{\alpha}_0) = \left[\partial\boldsymbol{\gamma}(\boldsymbol{\alpha}) / \partial\boldsymbol{\alpha}^\top\right]_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ be nonsingular in $\mathcal{N}_{\boldsymbol{\alpha}_0}$.

Then, as $n \to \infty$,

$$n^{1/2}\left(\hat{\boldsymbol{r}}_{MPL} - \boldsymbol{r}_0\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\mathbf{0}, \mathbf{J}_{\boldsymbol{\alpha}_0}\, \boldsymbol{I}\left(\boldsymbol{\alpha}_0\right)^{-1}\, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0}\, \boldsymbol{I}\left(\boldsymbol{\alpha}_0\right)^{-1}\, \mathbf{J}_{\boldsymbol{\alpha}_0}^{\top}\right), \tag{4.9}$$

where $\hat{\boldsymbol{r}}_{MPL} = \boldsymbol{\gamma}\left(\hat{\boldsymbol{\alpha}}_{MPL}\right)$ represents the column vector of estimated correlation matrix via maximum pseudo–likelihood estimates of $\boldsymbol{\alpha}_0$.

Proof

(4.9) immediately follows by applying the multivariate delta method (Lehmann and Casella, 1998, Theorem 8.22, Section 1.8) and invoking the asymptotic distribution of the estimator $\hat{\boldsymbol{\alpha}}_{MPL}$, given by (4.6).

■

Although the result in Theorem 4.3 is valid in general, to the best of our knowledge we only recognize two families of copulas that satisfy condition (4.8), namely, elliptical copulas and EFGM copulas of the form (4.3), respectively, where $d = p^{\star\star}$. With these families the copula parameter is characterized by a number of elements equal to non–duplicated and non–fixed elements of Pearson's linear correlation matrix $\boldsymbol{R}_0$. We remember that the meaning of condition (4.8) is to establish a direct link between copula parameters, relating to the measure of association between variables $\boldsymbol{x}$, and Pearson's correlation coefficients, reserved for a measure of the linear dependence between random variates.

Corollary 4.1 *Let $\boldsymbol{x} \in \mathbb{R}^p$ be a vector of random variables with elliptical copula $\mathcal{EC}\left(\boldsymbol{R}_0, h\right)$ and absolutely continuous generating variable $\mathsf{r} > 0$. Then, $\hat{\boldsymbol{r}}_{MPL} \equiv \hat{\boldsymbol{\alpha}}_{MPL}$.*

Proof

It is immediate to notice that in case of elliptical copulas $\boldsymbol{\gamma}\left(\boldsymbol{\alpha}_0\right)$ in Theorem 4.3 corresponds to a vector of functions where each component is the identity function, so that $\mathbf{J}_{\boldsymbol{\alpha}_0} = \boldsymbol{I}_{p^{\star\star}}$, the identity matrix of order $p^{\star\star}$. Therefore, the asymptotic covariance matrix in (4.9) is given by

$$\boldsymbol{\Sigma}_{MPL} \equiv \boldsymbol{I}\left(\boldsymbol{\alpha}_0\right)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0} \boldsymbol{I}\left(\boldsymbol{\alpha}_0\right)^{-1}$$

■

One classical example of elliptical distribution and its related copula is represented by t Student distribution. Let $\boldsymbol{x} = (x_1, \ldots, x_p)^\top$ be $\mathbf{t}$–distributed, denoted by $\mathbf{t}_\nu(\mathbf{0}, \boldsymbol{\Sigma}_0)$, and let t_ν be the univariate t–distribution function in $\mathbb{R}$ with ν degree of freedom; then $\{\mathsf{t}_\nu(x_1), \ldots, \mathsf{t}_\nu(x_p)\}^\top$ is a $\mathbf{t}_\nu$ Student copula. Notice that, for technical reasons discussed for instance in Demarta and McNeil (2005) or Kojadinovic and Yan (2010a), the number of degrees of freedom of the $\mathbf{t}_\nu$–copula has to be fixed (or previously estimated) and it will therefore not any more be considered as a parameter to be estimated.

Corollary 4.2 *Let $\boldsymbol{x} \in \mathbb{R}^p$ be a vector of random variables with finite first and second moments. Let the EFGM copula of the form (4.3) be the distribution function associated with $\boldsymbol{x}$. Then, as $n \to \infty$,*

$$n^{1/2}\left(\hat{\boldsymbol{r}}_{MPL} - \boldsymbol{r}_0\right) \xrightarrow{\mathscr{L}} \mathsf{N}\left(\mathbf{0}, \mathbf{J}_{\boldsymbol{\alpha}_0} \boldsymbol{I}\left(\boldsymbol{\alpha}_0\right)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0} \boldsymbol{I}\left(\boldsymbol{\alpha}_0\right)^{-1} \mathbf{J}_{\boldsymbol{\alpha}_0}\right) ,$$

where $\hat{\boldsymbol{r}}_{MPL}$ is a $p^{\star\star}$–dimensional vector with elements

$$\hat{r}_{MPL;i,j} = \hat{\gamma}_{i,j}\left(\hat{\alpha}_{MPL;i,j}\right) = \frac{\hat{\alpha}_{MPL;i,j}\,\hat{\delta}_{2;i}\hat{\delta}_{2;j}}{\sqrt{\hat{\sigma}_{i,i}\hat{\sigma}_{j,j}}} ,$$

and

$$\mathbf{J}_{\boldsymbol{\alpha}_0} = diag\left(\frac{\delta_{2;2}\delta_{2;1}}{\sqrt{\sigma_{2,2}\sigma_{1,1}}}, \ldots, \frac{\delta_{2;p}\delta_{2;1}}{\sqrt{\sigma_{p,p}\sigma_{1,1}}}, \frac{\delta_{2;3}\delta_{2;2}}{\sqrt{\sigma_{3,3}\sigma_{2,2}}}, \ldots, \frac{\delta_{2;p}\delta_{2;2}}{\sqrt{\sigma_{p,p}\sigma_{2,2}}}, \ldots, \frac{\delta_{2;p}\delta_{2;p-1}}{\sqrt{\sigma_{p,p}\sigma_{p-1,p-1}}}\right) .$$

We underline the importance of the condition about the existence of first and second moments with EFGM copula once again. As discussed in Section 4.1, we remember that the relationship between Pearson's linear correlation coefficients and Kendall's tau needs existing second moment. Moreover, L–moments of $\boldsymbol{x}$ would not exist with non finite first moment (see Theorem B.1 in Appendix B).

In order to provide conditions for Theorem 3.4, we study the properties of the estimator of $\mathbf{J}_{\boldsymbol{\alpha}_0} \boldsymbol{I}(\boldsymbol{\alpha}_0)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0} \boldsymbol{I}(\boldsymbol{\alpha}_0)^{-1} \mathbf{J}_{\boldsymbol{\alpha}_0}^\top$. Given an *IID* sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, we define the estimator of the asymptotic covariance matrix in (4.9) as $\hat{\boldsymbol{\Sigma}}_{MPL} := \hat{\mathbf{J}}^\top \boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{MPL}} \boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1} \hat{\mathbf{J}}$, where $\hat{\mathbf{J}}$ is a suitable estimator for $\mathbf{J}_{\boldsymbol{\alpha}_0}$, $\boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})$ is the estimator of Fisher information matrix given by (4.7), and $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{MPL}}$ is the sample covariance matrix of the variables $(\boldsymbol{\mathcal{M}}_1, \ldots, \boldsymbol{\mathcal{M}}_{p^{\star\star}})^\top$.

In order to prove the consistency of $\hat{\boldsymbol{\Sigma}}_{MPL}$ we provide the following result, by analogy with Theorem 3.7.

Theorem 4.4 *Under the assumptions of Theorem 4.3,*

$$vech\left(\hat{\boldsymbol{\Sigma}}_{MPL}\right) = vech\left\{\hat{\mathbf{J}}\, \boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{MPL}} \boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1} \hat{\mathbf{J}}^\top\right\}$$

is a consistent estimator of the asymptotic covariance matrix in (4.9).

Proof

Let $\boldsymbol{H}_p$ represent the left inverse of the transition or duplication matrix $\boldsymbol{G}_p$.

$$\begin{aligned} vech\left(\hat{\boldsymbol{\Sigma}}_{MPL}\right) = \boldsymbol{H}_{p^{\star\star}} \left(\hat{\mathbf{J}} \otimes \hat{\mathbf{J}}\right) \boldsymbol{G}_{p^{\star\star}} \times \\ \times\, vech\left\{\boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{MPL}} \boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1}\right\}. \end{aligned}$$

Since $vech\left\{\boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{MPL}} \boldsymbol{I}(\hat{\boldsymbol{\alpha}}_{MPL})^{-1}\right\}$ is a consistent estimator of $vech\left\{\boldsymbol{I}(\boldsymbol{\alpha}_0)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0} \boldsymbol{I}(\boldsymbol{\alpha}_0)^{-1}\right\}$ as Genest, Ghoudi, and Rivest (1995, Section 4) point out, the result then follows by using the part **(a)** of Lemma 4.1.

■

x_j	$\tilde{\mathbf{\Lambda}}_{\cdot,1}$	$\tilde{\mathbf{\Lambda}}_{\cdot,2}$	$\tilde{\psi}_j$
x_1	0.90	0.00	0.19
x_2	0.90	0.00	0.19
x_3	0.90	0.00	0.19
x_4	0.90	0.00	0.19
x_5	0.90	0.00	0.19
x_6	0.00	0.90	0.19
x_7	0.00	0.90	0.19
x_8	0.00	0.90	0.19

Table 4.1: *Factor loadings and residual variances of the simulation study.*

4.3 A comprehensive empirical study

In order to see at work the test statistic (3.9) via Kendall's tau and the one by maximum pseudo likelihood, respectively, we firstly perform a simulation study where different sample sizes and also misspecified copula functions are considered. We stress the fact that the usefulness of the test statistic can be affected by different elements. Secondly, we apply a correlation structure analysis to real data.

In what follows we only consider the copula factor model as defined in Section 3.3. For the sake of simplicity and practicality, our attention will be focused on elliptical copulas, in particular Normal and $\mathbf{t}_\nu$ Student copulas. In this manner we can make a comparison with Klüppelberg and Kuhn (2009)'s results.

We start with the simulation study. We choose a $p = 8$ dimensional setting with $m = 2$ factors. Loadings and residual correlations are given in Table 4.1. The structured correlation matrix by factor model is hence given by $\boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}_0\right) = \tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top + \tilde{\mathbf{\Psi}}$.

The simulations were run according to a balanced experimental design involving the following components. Three sample sizes are considered; i.e., $n = 100, 250, 1000$, representing the case of small, medium, and large sized samples, respectively. For each of these three cases we carried out $N = 1000$ simulations consisting in drawing data from a copula with structured copula parameter (i.e., $\boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}_0\right) = \tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^\top + \tilde{\mathbf{\Psi}}$) and estimating and testing the copula factor model $H_0 : \boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}_0\right) = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^\top + \tilde{\boldsymbol{\Psi}}$. Three couples of copula models are considered in each simulation. An hypothesized copula under H_0 and a copula model from which the data were generated. We only considered Normal and $\mathbf{t}_3$ Student copulas for playing these roles. In each of these $3 \times 3 \times 1000$ repetitions we monitored the behavior of test statistic (3.9) by using QQ–plots and kernel densities representations of its sample distribution.

All the procedures used to carry out the computations are written in the language of the statistical software `R`. They are based on the pseudo–code described in Section 3.3. We refer to the `copula R` package (Kojadinovic and Yan, 2010c) available on the Comprehensive `R` Archive Network for computational issues about copulas.

In the case of a two–factor setting, to ensure uniqueness of the loadings, we use the restriction that $\tilde{\boldsymbol{\Lambda}}^\top \tilde{\boldsymbol{\Psi}}^{-1} \tilde{\boldsymbol{\Lambda}}$ is diagonal; see Section 2.2.1. Hence, we have $m\,(m-1)\,/2 = 1$ additional constraints. Using this restriction and the two–factor setting, test statistic (3.9) via Kendall's tau and maximum pseudo–likelihood should be asymptotically χ^2–distributed with $df = p\,(p-1)\,/2 - pm + m\,(m-1)\,/2 = 13$ degrees of freedom; see Theorem 3.4. Figures 4.2, 4.3, and 4.4 show the situation in case of Kendall's tau–based test statistic and sample size are equal to $n = 100, 250, 1000$, respectively. Similarly, Figures 4.5, 4.6, and 4.7 report the same features for maximum pseudo–likelihood–based choice.

First of all, we can observe physiological Heywood cases (see Table 4.2) that affect both of the two methodologies with the small sample size ($n = 100$). They totally disappear when the number of observations increases ($n = 250, 1000$). We believe that small samples can lack in information in order to recognize the correlation structure from which they were generated.

Comments on the power of test statistic and its approximation to χ^2–distribution must be separately led for the case of correct specification and misspecification of copula under H_0. Each line of Table 4.3 shows the number of rejections of $H_0 : C_{\boldsymbol{x}}\left\{\cdot; \boldsymbol{r}\left(\tilde{\boldsymbol{\theta}}_0\right) = vecp\left(\tilde{\boldsymbol{\Lambda}}^\top\tilde{\boldsymbol{\Lambda}} + \tilde{\boldsymbol{\Psi}}\right)\right\}$ associated with the different tests, given a choice of $C_{\boldsymbol{x}}$ and a true underlying copula C_0. As Klüppelberg and Kuhn (2009) point out, in case of right choice of the dependence model the empirical distribution of (3.9) fits the χ^2_{13}–distribution quite

	Copula under H_0	True Copula	$n = 100$	$n = 250$	$n = 1000$
τ	Student 3 df	Student 3 df	81	0	0
	Student 3 df	Normal	21	0	0
	Normal	Student 3 df	35	0	0
MPL	Student 3 df	Student 3 df	50	0	0
	Student 3 df	Normal	16	0	0
	Normal	Student 3 df	61	4	0

Table 4.2: *Number of Heywood cases in* $N = 1000$ *simulated samples.*

well for Kendall's tau–based methodology with $n = 100$ (Figures 4.2 (a)–(b), respectively). Figures 4.3 (a)–(b) and 4.4 (a)–(b) show an almost perfect fit to the χ^2_{13}–distribution for large samples. In case of maximum pseudo–likelihood–based test statistic quite good fit is only obtained when $n = 1000$ (Figures 4.7 (a)–(b), respectively). On the other hand, this careful behavior bring into a few number of rejections of the true correlation structure in contrast with the results of Kendall's tau–based test statistic. Table 4.3 indeed shows that maximum pseudo–likelihood–based test statistic performs better than the Kendall's tau–based counterpart.

	Copula under H_0	True Copula	$n = 100$	$n = 250$	$n = 1000$
τ	Student 3 df	Student 3 df	22	28	38
	Student 3 df	Normal	20	28	45
	Normal	Student 3 df	12	28	45
MPL	Student 3 df	Student 3 df	0	3	21
	Student 3 df	Normal	0	0	0
	Normal	Student 3 df	62	498	802

Table 4.3: *Number of rejections of* H_0 *(5% significance level) assuming different copula models in* $N = 1000$ *simulated samples (every time decreased by the Heywood cases, respectively).*

It is a classical fact of statistics that the power of a test increases with sample size. As Table 4.3 clearly shows, the present case is no exception.

Nevertheless, we must clarify the reasons of the different results obtained with the two versions of test statistic when the true copula is misspecified. To illustrate the difficulties associated with the proper identification of a dependence structure, Figure 4.1 portrays typical contour plots for the bivariate Normal and $\mathbf{t}_3$ Student copula densities considered in the study, respectively. When data are generated from Normal copula, the distinctive features of the two models are hardly distinguishable. Therefore, the misspecification error can be rarely detected both by Kendall's tau–based and by maximum pseudo–likelihood–based test statistic and its sample distribution fits the χ^2_{13}–distribution as well as in case of correct specification. In contrast, if $\mathbf{t}_3$ Student copula represents the data generating process, the characteristics of the two different models are then much easier to pick out. For instance, their lower– and upper–tail dependences translate into greater densities of points in the lower–left and upper–right corners of the unit square, respectively. Since Kendall's tau–based statistic test use the same seminal relationship (3.11) with Pearson's linear correlation coefficients for all elliptical copulas, it can not be able to make distinctions between Normal and $\mathbf{t}_3$ Student copulas. The number of rejections of H_0 remains low also when the sample size increases and the test statistic sample distribution anyway approximates the χ^2_{13}–distribution. On the contrary, with the maximum pseudo–likelihood–based test statistic we use much more informations provided by copula densities and we can distinguish the two models. The number of rejections of H_0 is almost equal to the $N = 1000$ simulations when the sample size is $n = 1000$. Moreover, maximum pseudo–likelihood test statistics are never χ^2_{13}–distributed.

Now we change our perspective and we turn over real data. A dataset is taken into account from `Datastream`, not totally identical to that one used by Klüppelberg and Kuhn (2009). The daily values for financial indices Standar & Poor 500 Composite, Dow Jones Industrials, and NIKKEI 225 Stock Average, for Crude Oil–Brent US$/BBL, and for exchange rates Great Britain Pound to Euro, US Dollar to Euro, Swiss Franc to Euro, and Japanese Yen to Euro are considered. We will shortly indicate them with the labels SP500, DJ, NIK, OIL, GBP, USD, SWISS, YEN, respectively. All indices, exchange

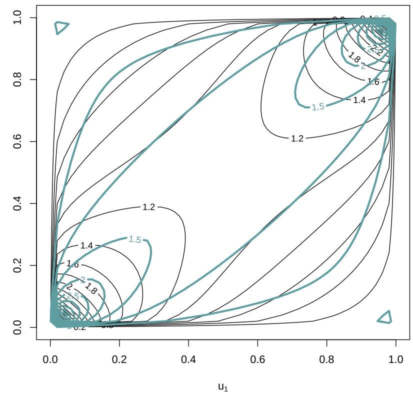


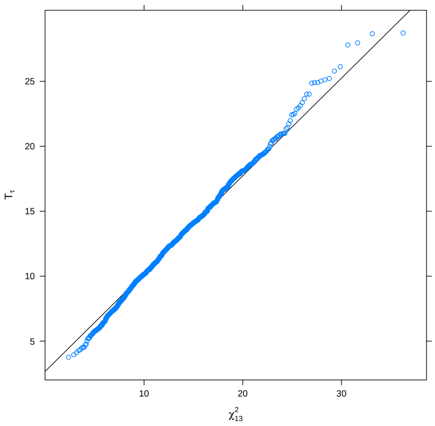
Figure 4.1: *Contour plot of bivariate Normal (black line) and* $\mathbf{t}_3$ *Student (blue line) copula densities, respectively.*

rates and oil price are obtained for the period [06/12/1990;06/12/2010], resulting in $n = 5218$ observations. Motivated by the interest on the Great Financial Recession of 2007/2009, we apply a copula factor model to the data to better understand the presence of common latent risk factors in this period. In what follows we will refer to the period December 1990 to December 2010 as the "full sample" and January 2007 to December 2010 as the "financial crisis", resulting in $n_{fc} = 1026$ observations.

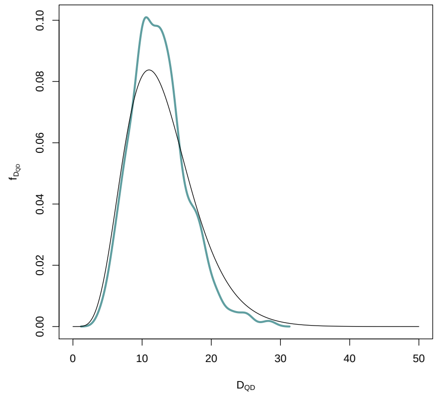
Instead of analyzing the daily values themselves, we calculated and considered (percentual) continuosly compounded returns (log–returns) $r_{t,j} = 100\,(\log p_{t,j}/\log p_{t-1,j}) \quad (t = 2, \ldots, n_{fc}\,;\, j = 1, \ldots, 8)$. Log–returns are displayed in Figures 4.8.

Table 4.4 summarizes descriptive statistics. All series feature negative skewness (except one; i.e., NIK) and high kurtosis. Moreover, there is empirical evidence for serial correlation and GARCH effects as the Ljung–Box statistic and Engle's Lagrange Multiplier statistic indicate (see Table 4.5).

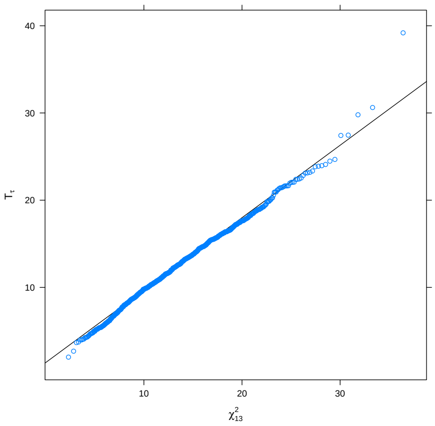
Because individual risk series in finance are typically serially dependent, Chen and Fan (2006) introduced a class of semiparametric copula–based multivariate dynamic models, in which the conditional mean and conditional variance of individual risk series are parametrically specified. On the converse the joint distribution of the standardized innovations is a parametric



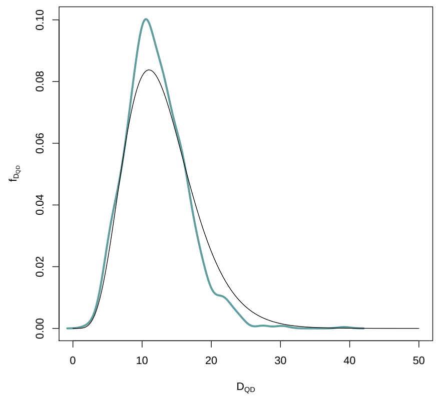
(a)



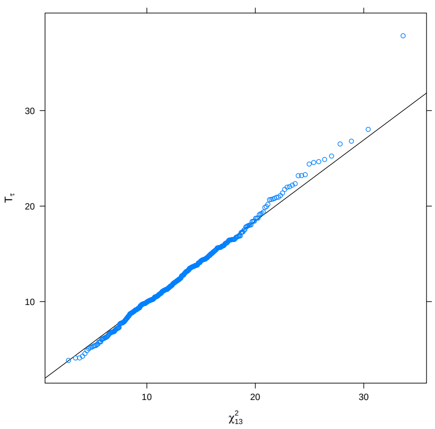
(b)



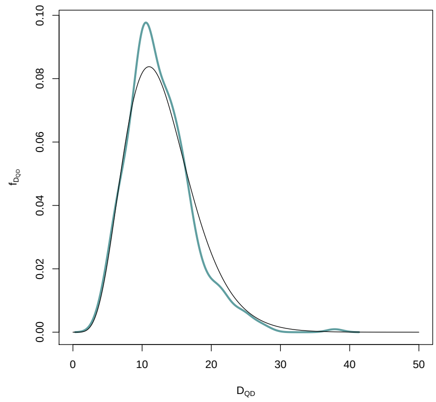
(c)



(d)



(e)



(f)

Figure 4.2: *QQ–plot and kernel density of ordered Kendall's tau–based test statistic estimates against the* χ^2_{13}*–quantiles and* χ^2_{13}*–density (sample size* $n = 100$*). (a) and (b) represents the case of correct specification of the true underlying copula (*$\mathbf{t}_3$ *Student copula). (c) and (d) and (e) and (f) represent two cases of copula misspecification (*$\mathbf{t}_3$ *Student and Normal copula assumed, respectively).*

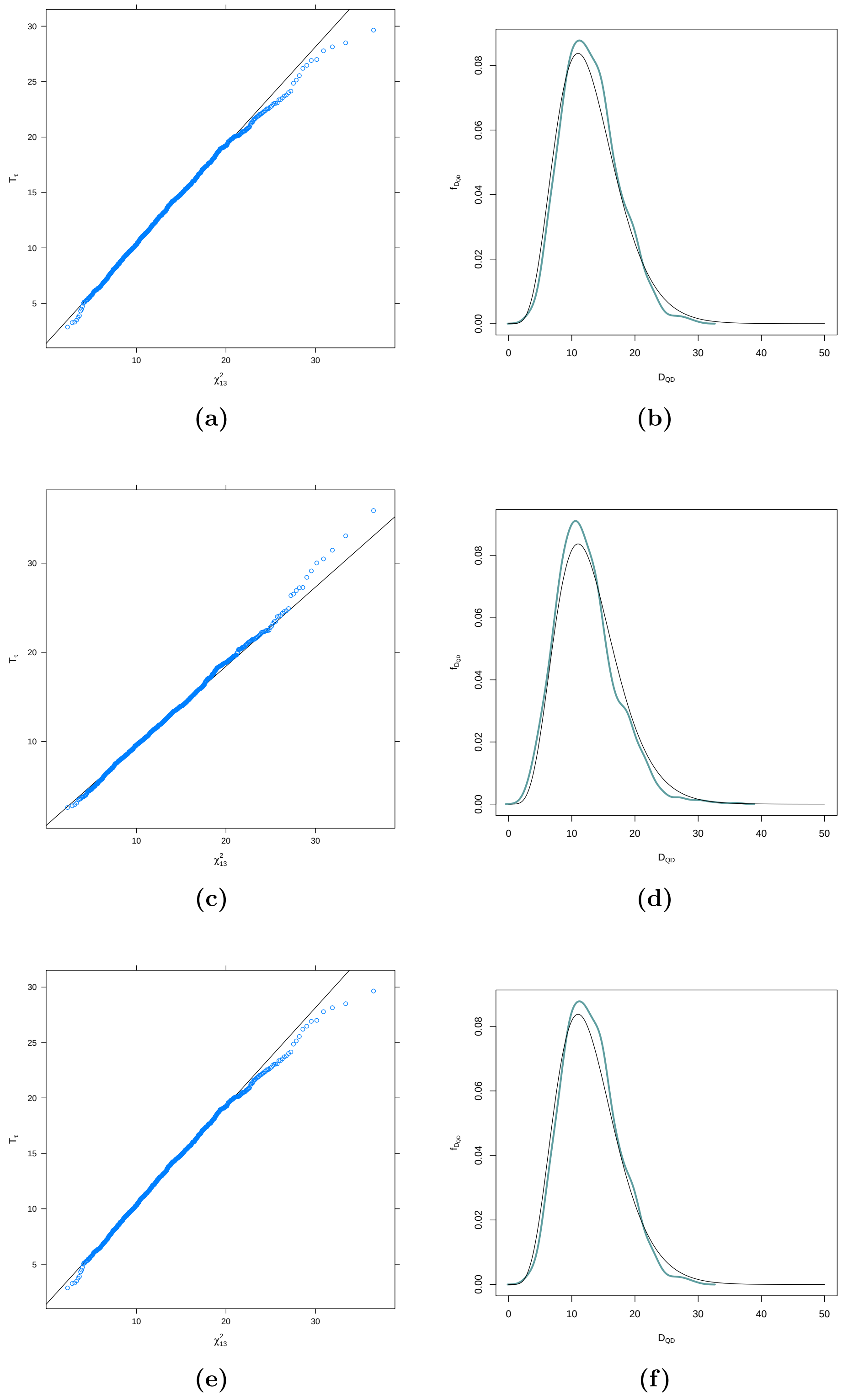


Figure 4.3: *QQ–plot and kernel density of ordered Kendall's tau–based test statistic estimates against the* χ^2_{13}*–quantiles and* χ^2_{13}*–density (sample size* $n = 250$*). (a) and (b) represents the case of correct specification of the true underlying copula (*$\mathbf{t}_3$ *Student copula). (c) and (d) and (e) and (f) represent two cases of copula misspecification (*$\mathbf{t}_3$ *Student and Normal copula assumed, respectively).*

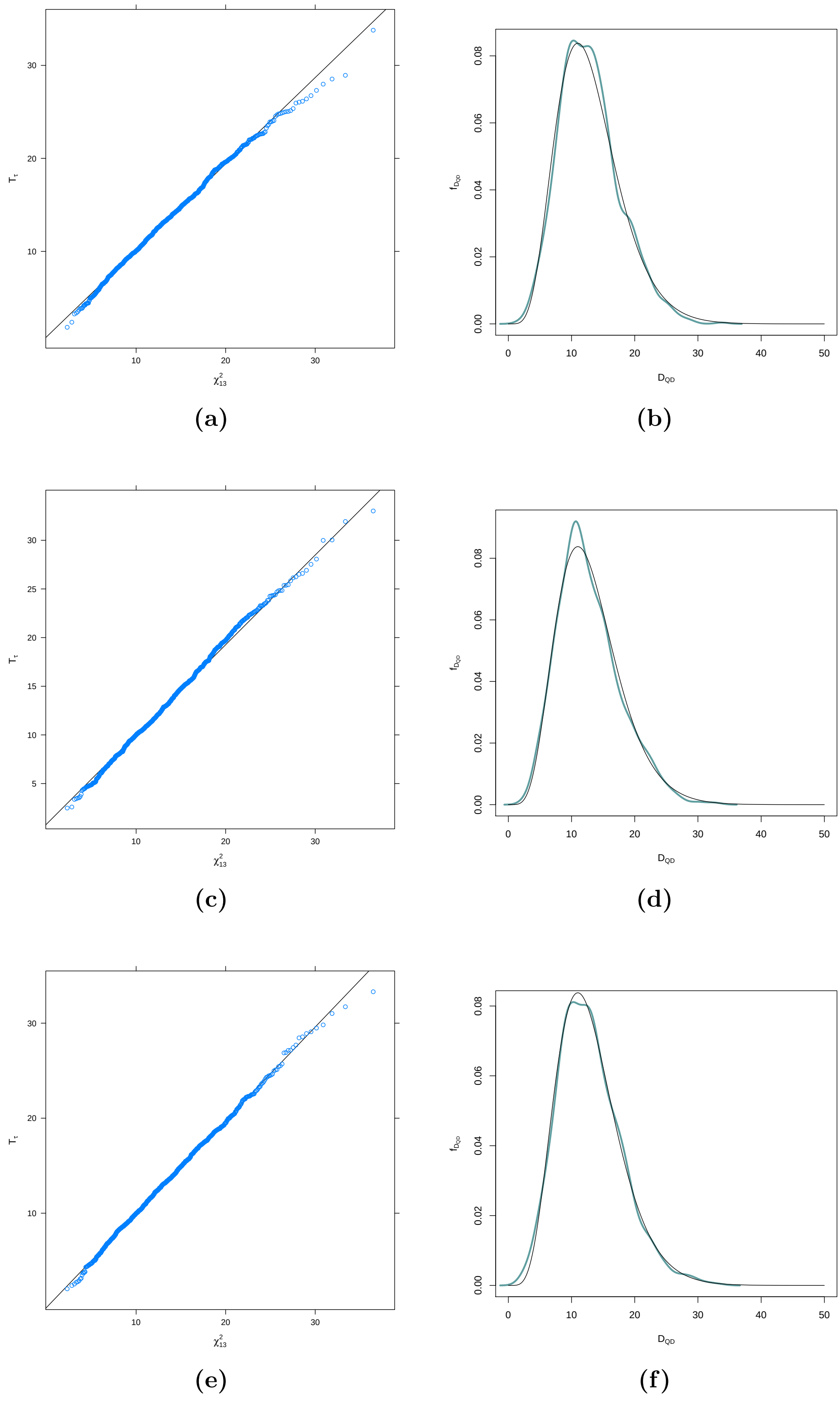


Figure 4.4: *QQ–plot and kernel density of ordered Kendall's tau–based test statistic estimates against the* χ^2_{13}*–quantiles and* χ^2_{13}*–density (sample size* $n = 1000$*). (a) and (b) represents the case of correct specification of the true underlying copula (*$\mathbf{t}_3$ *Student copula). (c) and (d) and (e) and (f) represent two cases of copula misspecification (*$\mathbf{t}_3$ *Student and Normal copula assumed, respectively).*

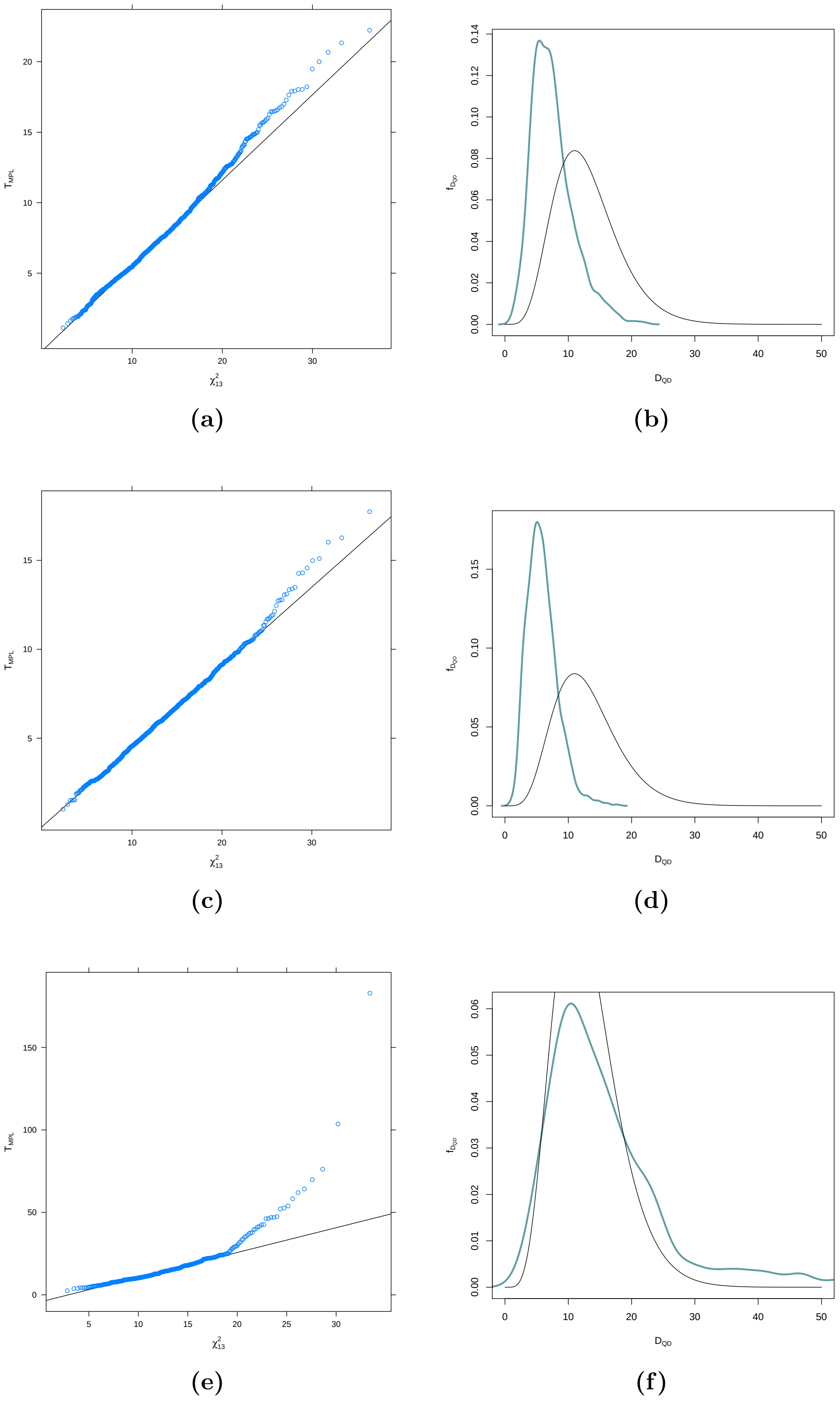


Figure 4.5: *QQ–plot and kernel density of ordered maximum pseudo–likelihood–based test statistic estimates against the* χ^2_{13}*–quantiles and* χ^2_{13}*–density (sample size* $n = 100$*). (a) and (b) represents the case of correct specification of the true underlying copula (*$\mathbf{t}_3$ *Student copula). (c) and (d) and (e) and (f) represent two cases of copula misspecification (*$\mathbf{t}_3$ *Student and Normal copula assumed, respectively).*

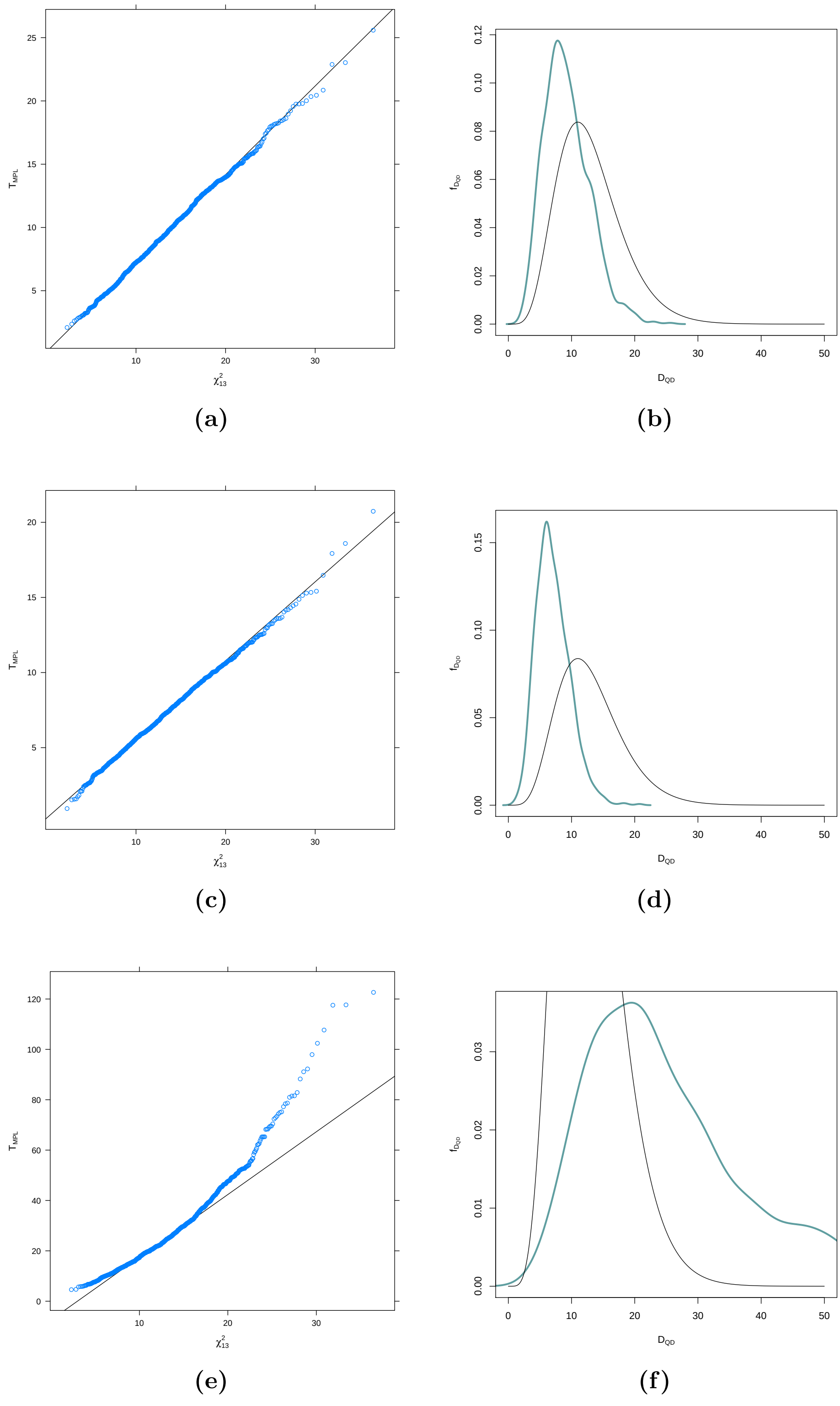


Figure 4.6: *QQ–plot and kernel density of ordered maximum pseudo–likelihood–based test statistic estimates against the χ^2_{13}–quantiles and χ^2_{13}–density (sample size $n = 250$). (a) and (b) represents the case of correct specification of the true underlying copula ($\mathbf{t}_3$ Student copula). (c) and (d) and (e) and (f) represent two cases of copula misspecification ($\mathbf{t}_3$ Student and Normal copula assumed, respectively).*

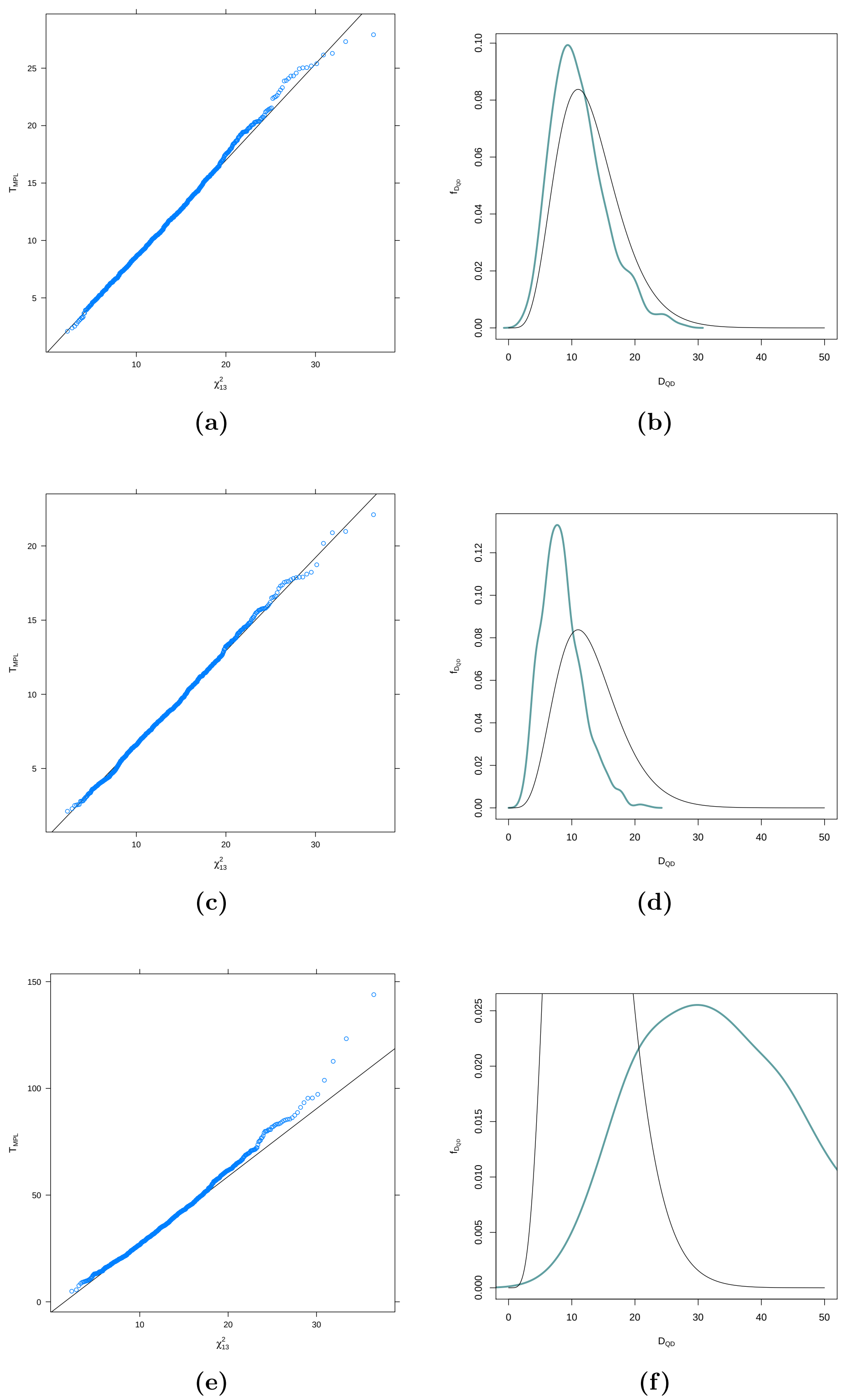


Figure 4.7: *QQ–plot and kernel density of ordered maximum pseudo–likelihood–based test statistic estimates against the* χ^2_{13}*–quantiles and* χ^2_{13}*–density (sample size* $n = 1000$*). (a) and (b) represents the case of correct specification of the true underlying copula (*$\mathbf{t}_3$ *Student copula). (c) and (d) and (e) and (f) represent two cases of copula misspecification (*$\mathbf{t}_3$ *Student and Normal copula assumed, respectively).*

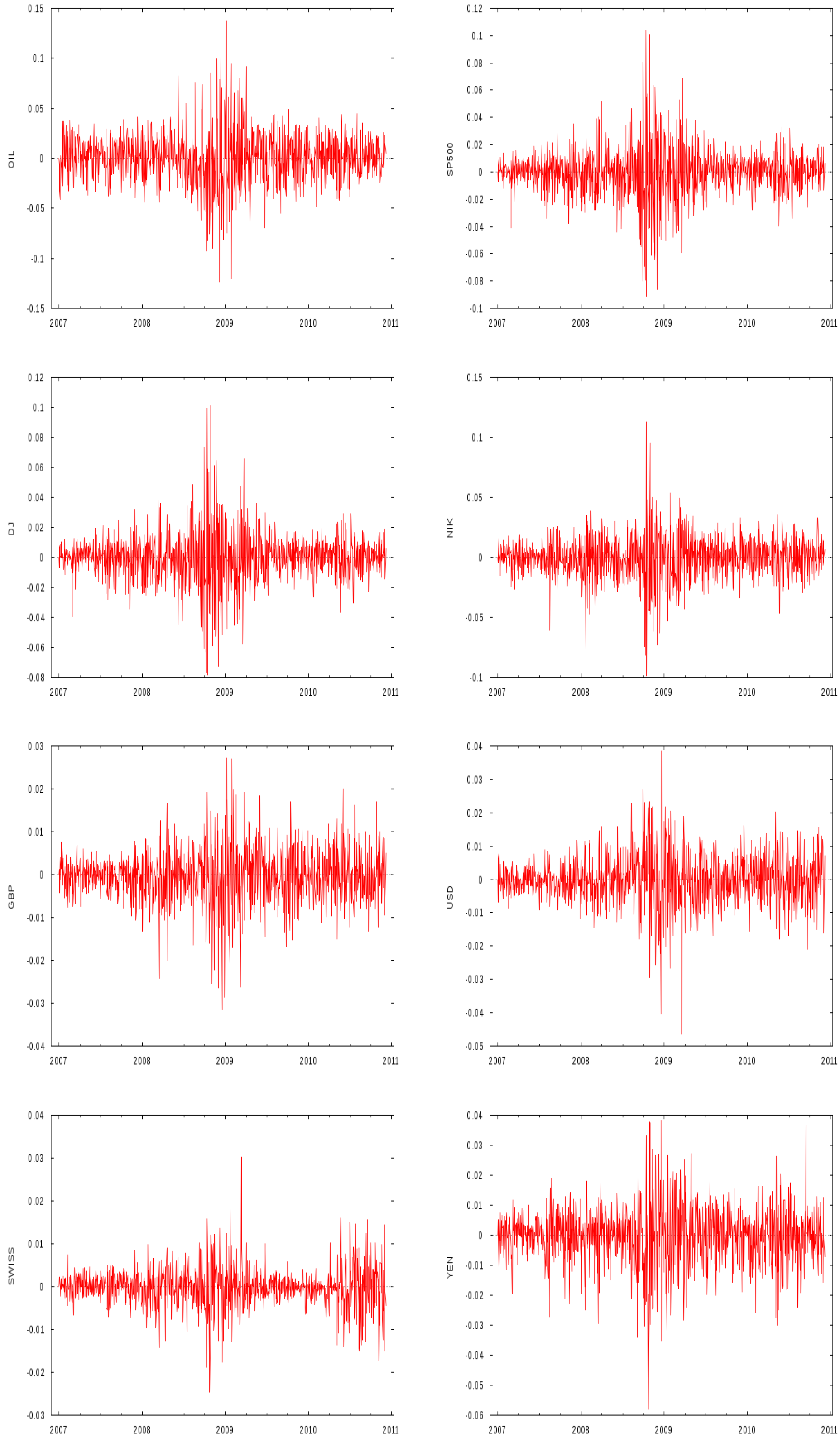


Figure 4.8: Log–returns for the oil price, indices, and exchange rates.

Variable	Mean	Median	Minimum	Maximum
OIL	0.0002	0.0000	−0.4420	0.1516
SP500	0.0003	0.0004	−0.0911	0.1037
DJ	0.0003	0.0005	−0.0845	0.1011
NIK	−4.4520e–05	−6.2006e–05	−0.0989	0.1131
GBP	−5.6342e–05	0.0000	−0.0314	0.0272
USD	9.5560e–06	0.0000	−0.0463	0.0383
SWISS	−5.4964e–05	0.0000	−0.0245	0.0302
YEN	−9.6349e–05	0.0000	−0.0560	0.0540
Variable	**Std. Dev.**	**C.V.**	**Skewness**	**Ex. Kurtosis**
OIL	0.0232	102.6330	−1.3536	27.8941
SP500	0.0132	49.0982	−0.1334	5.3350
DJ	0.0128	42.7366	−0.0971	4.9701
NIK	0.0162	363.1860	0.0356	3.0851
GBP	0.0050	89.5909	−0.2380	3.3613
USD	0.0064	667.2330	−0.0830	2.8416
SWISS	0.0031	57.2594	−0.2740	7.3547
YEN	0.0076	79.1326	−0.3624	4.5221

Table 4.4: *Summary statistics of the dataset, using the "full sample".*

copula evaluated using nonparametric marginal estimates. That is, a scalar GARCH(p_j, q_j) model is used to capture volatility of individual risk series and a parametric copula is used to model the contemporaneous dependence between different risks. The conventional approach is to assume independence and normality for the standardized innovations, while Chen and Fan (2006)'s approach is to assume a copula for them. The main contribution of *GARCH–copula* model is that it permits modeling the conditional correlation and dependence structure, separately and simultaneously. For a general survey on multivariate GARCH models, see Bauwens, Laurent, and Rombouts (2006).

Suppose the observations $\left\{\boldsymbol{r}_t = (r_{t,1}, \ldots, r_{t,8})^\top\right\}_{t=1}^{n_{fc}}$ satisfy

$$r_{t,j} = \mu_{t,j} + \sigma_{t,j}\epsilon_{t,j}\,, \quad \sigma^2_{t,j} = c_{t,j} + \sum_{i=1}^{p_j} a_{i,j}\,\sigma^2_{t-i,j}\,\epsilon^2_{t-i,j} + \sum_{i=1}^{q_j} b_{i,j}\,\sigma^2_{t-i,j}\,,$$

Variable	$\mathcal{LB}\,(37)$	$\mathcal{LM}\,(2)$
OIL	82.9753*	1051.0000*
SP500	103.6649*	1490.2200*
DJ	88.2872*	7908.7900*
NIK	58.9312*	813.1480*
GBP	76.0937*	632.2620*
USD	50.4626	559.1650*
SWISS	73.1215*	1502.8400*
YEN	71.4617*	1243.7600*

Table 4.5: *Ljung–Box ($\mathcal{LB}$) statistic and Engle's Lagrange Multiplier ($\mathcal{LM}$) statistic (significant statistics at the 5% level are marked with an asterisk).*

$j = 1, \ldots, 8$, where $\left\{ \tilde{\boldsymbol{\epsilon}}_t = (\epsilon_{t,1}/\sigma_{t,1}, \ldots, \epsilon_{t,8}/\sigma_{t,8})^\top \right\}_{t=1}^{n_{fc}}$ is a sequence of *IID* random vectors with $E\left(\tilde{\boldsymbol{\epsilon}}_t\right) = \mathbf{0}$, $E\left(\tilde{\boldsymbol{\epsilon}}_t \tilde{\boldsymbol{\epsilon}}_t^\top\right) = \boldsymbol{I}_8$. The joint distribution function $F_{\tilde{\boldsymbol{\epsilon}}}$ of the standardized $\tilde{\boldsymbol{\epsilon}}_t$ is assumed to take the semiparametric form $F_{\tilde{\boldsymbol{\epsilon}}}\left(\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_8\right) = C_{\tilde{\boldsymbol{\epsilon}}}\left\{F_{\tilde{\boldsymbol{\epsilon}},1}\left(\tilde{\epsilon}_1\right), \ldots, F_{\tilde{\boldsymbol{\epsilon}},8}\left(\tilde{\epsilon}_8\right); \boldsymbol{\alpha}_0\right\}$. Here $C_{\tilde{\boldsymbol{\epsilon}}}\left(\cdot\,; \boldsymbol{\alpha}_0\right)$ is a copula function parametrized up to the unknown parameter $\boldsymbol{\alpha}_0 \in \mathbb{A} \subset \mathbb{R}^d$, and for $j = 1, \ldots, 8$, $F_{\tilde{\boldsymbol{\epsilon}},j}\left(\tilde{\epsilon}_j\right)$ is the marginal distribution function of $\tilde{\epsilon}_{t,j}$, assumed to be continuous. Let $C_{\tilde{\boldsymbol{\epsilon}}}$ denote the unique copula corresponding to the true joint distribution $F_{\tilde{\boldsymbol{\epsilon}}}$ of the GARCH residual vector $\tilde{\boldsymbol{\epsilon}}_t$. We call $C_{\tilde{\boldsymbol{\epsilon}}}$ the *residual copula* according to Chen and Fan (2006).

Crucial to the validity of the Chen and Fan (2006)'s model estimation and selection test is the result that the asymptotic distribution of the estimator $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}_0$ is not affected by the initial step estimation of the GARCH parameters. The limit distribution of $\hat{\boldsymbol{\alpha}}$ and goodness–of–fit test statistic are independent of the GARCH filtering.

We apply the procedure by Chen and Fan (2006) to our dataset and we estimate and test a copula factor model for the standardized innovations. In other words, we assume there exists an underlying generating random vector $\boldsymbol{\zeta} \in \mathbb{R}^z$ such that $C_{\tilde{\boldsymbol{\epsilon}}} = C_{(\boldsymbol{\Lambda}, \boldsymbol{I}_p)\boldsymbol{\zeta}}$. Here, we fit the Normal and $\mathbf{t}_\nu$ Student copulas to the standardized residuals from filtering a GARCH$(1,1)$ for each series; see Tables 4.6 for GARCH parameter estimates. The copula correlation matrix is thus assumed to be of the form $\boldsymbol{R}\left(\tilde{\boldsymbol{\theta}}_0\right) = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^\top + \tilde{\boldsymbol{\Psi}}$ for some $m < 8$, $\tilde{\mathbf{\Lambda}} \in \mathbb{R}^{8\times m}$ and $\tilde{\mathbf{\Psi}} \in \mathbb{R}^{8\times 8}$, where $\tilde{\boldsymbol{\theta}}_0 = vecp\left(\tilde{\mathbf{\Lambda}}, \tilde{\mathbf{\Psi}}\right)$.

Goodness–of–fit tests for copulas are summarized in Table 4.7. The test statistic is the Cramér–von Mises functional defined in equation (2) of Genest, Rémillard, and Beaudoin (2009) and it is based on the empirical process comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis. Approximate p–values for the test statistic has been obtained using parametric bootstrap. Both the Normal and the $\mathbf{t}_{14.87}$ Student copula are not rejected. For Student copula we previously estimated the degree of freedom as remarked in Section 4.2. Therefore, we fit a Normal and $\mathbf{t}_{14.87}$ Student copula factor model to our standardized innovations using the test statistic (3.9) via Kendall's tau and maximum pseudo–likelihood, respectively.

In order to estimate the number of latent factors, we use a 95% confidence test; i.e., we reject the null hypothesis of having a copula m–factor model, if the test statistic (3.9) is larger than the 95%–quantile of the χ^2_{df}–distribution. This suggests a four factor model under both the Kendalls tau–based and the maximum pseudo–likelihood–based test statistics (see Table 4.8).

We present the corresponding factor loadings in Figure 4.9 for both elliptical copula factor models. In Figures 4.9 (a)–(c)–(e)–(g) the Normal copula–based estimated loadings are plotted, meanwhile in Figures 4.9 (b)–(d)–(f)–(h) the $\mathbf{t}_{14.87}$ Student copula–based estimates are drawn. Firstly, we emphasize that, although we have plotted the factors in the same figures, they are obtained by the two different estimation methods (i.e., Kendall's tau and maximum pseudo–likelihood, respectively) and may have different interpretations. Secondly, since the estimated Student degrees of freedom are quite high, the distinctive characteristics of the two copula models are hardly distinguishable. Anyway, the use of classical normal–based correlation structure analysis can here be substituted by $\mathbf{t}$ Student copula factor analysis, which has completely different features from those of the Normal copula as the generating variable $\mathtt{r}$ is heavy tailed.

For the first factor the loadings of the different correlation estimators behave very similarly under the two elliptical copulas assumed. The first factor has weights that is close to 1 for SP500 and DJ. Hence, factor 1 can

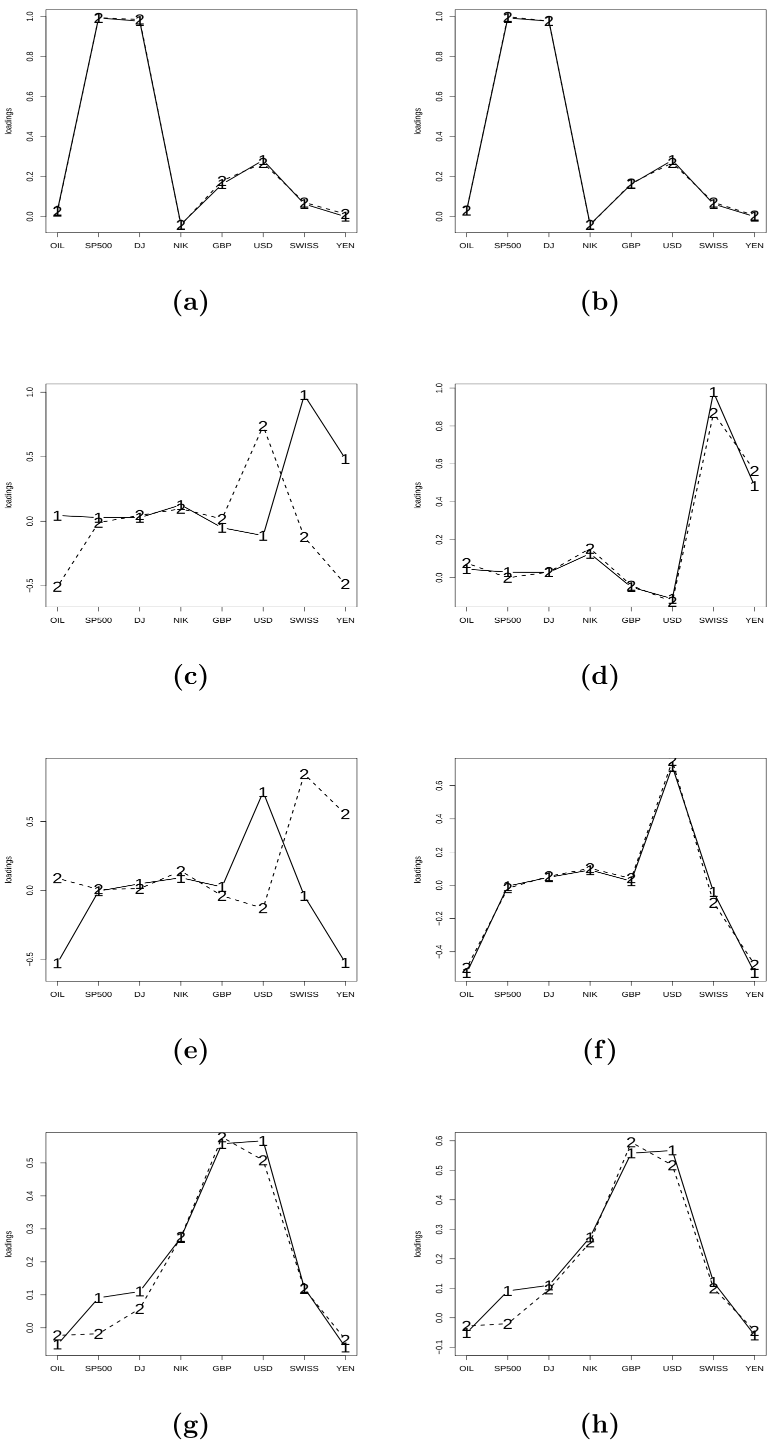


Figure 4.9: *Loadings of the four factors, where plotted points* 1 *and* 2 *represent the estimated loadings via Kendall's tau– and maximum–pseudo likelihood–based procedures, respectively; in (a), (c), (e), (g) the Normal copula is assumed; in (b), (d), (f), (h) the* $\mathbf{t}_{14.87}$ *Student copula is assumed.*

be interpreted as the United States–risk–factor. It also can be seen that this factor has a positive weight for all components, but not for the Nikkei index, which is very small negative (-0.039 and -0.042 by using Kendall's tau– and maximum pseudo–likelihood–based procedure, respectively, under both Normal and Student's t copula).

For factor 2 we observe for both correlation estimators in case of **t** Student copula a large weight on Swiss Franc and Japanese Yen, so we call it the Swiss/Nippon–risk–factor. We note that this factor has a very small negative weight on USD, maybe as a consequence of the previously United States–risk–factor. In case of Normal copula, the two estimators give different interpretations. Maximum–pseudo likelihood–based estimated loadings attaches a lot of importance to the exchange rate USD and negative weights to OIL and to the couple SWISS–YEN. We can name this factor US Dollar–risk–factor. Kendall's tau estimates seem to have the same interpretation of **t** Student copula case.

Considering factor 3, we are present at an inversion of the comments made for factor 2. For the case of **t** Student copula we interpret the factor as US Dollar–risk–factor. On the contrary, in case of Normal copula maximum–pseudo likelihood–based estimated loadings bring to the Swiss/Nippon–risk–factor, while Kendall's tau estimates to US Dollar–risk–factor again.

Finally, as for the first factor, the loadings of the different correlation estimators behave very similarly under the two elliptical copulas assumed in case of factor 4. We interpret it as Anglo/American Currencies–risk–factor.

We conclude that these interpretations of latent factors can be reasonable since the recession was originated in United States and it involved the American country and Great Britain with a lot numbers of companies bankruptcies.

	Coefficient	Estimate	Std. Error	p–value
OIL	$\hat{\mu}_{OIL}$	0.0005	0.0003	0.0612
	$\hat{c}_{OIL}$	4.5069e–06	1.1397e–06	0.0001
	$\hat{a}_{1,OIL}$	0.0688	0.00588908	0.0000
	$\hat{b}_{1,OIL}$	0.9260	0.00626105	0.0000
SP500	$\hat{\mu}_{SP500}$	0.0004	0.0001	0.0017
	$\hat{c}_{SP500}$	1.3692e–06	3.0678e–07	8.0800e–06
	$\hat{a}_{1,SP500}$	0.0546	0.0060	1.8700e–19
	$\hat{b}_{1,SP500}$	0.9369	0.0071	0.0000
DJ	$\hat{\mu}_{DJ}$	0.0005	0.0001	0.0007
	$\hat{c}_{DJ}$	1.5660e–06	3.4227e–07	4.7600e–06
	$\hat{a}_{1,DJ}$	0.0573	0.0064	2.6800e–19
	$\hat{b}_{1,DJ}$	0.9325	0.0076	0.0000
NIK	$\hat{\mu}_{NIK}$	0.0001	0.0002	0.5152
	$\hat{c}_{NIK}$	3.9403e–06	8.0001e–07	8.4200e–07
	$\hat{a}_{1,NIK}$	0.0745	0.0074	8.3900e–24
	$\hat{b}_{1,NIK}$	0.9118	0.0085	0.0000
GBP	$\hat{\mu}_{GBP}$	−4.4262e–05	7.4072e–05	0.5501
	$\hat{c}_{GBP}$	8.1116e–08	3.4906e–08	0.0201
	$\hat{a}_{1,GBP}$	0.0342	0.0049	2.1300e–12
	$\hat{b}_{1,GBP}$	0.9630	0.0053	0.0000
USD	$\hat{\mu}_{USD}$	−6.9617e–05	7.8666e–05	0.3762
	$\hat{c}_{USD}$	1.7582e–07	5.1917e–08	0.0007
	$\hat{a}_{1,USD}$	0.0296	0.0031	7.2700e–22
	$\hat{b}_{1,USD}$	0.9664	0.0035	0.0000
SWISS	$\hat{\mu}_{SWISS}$	1.3110e–05	3.2723e–05	0.6887
	$\hat{c}_{SWISS}$	6.3510e–08	1.5545e–08	4.3900e–05
	$\hat{a}_{1,SWISS}$	0.0671	0.0069	2.1800e–22
	$\hat{b}_{1,SWISS}$	0.9300	0.0070	0.0000
YEN	$\hat{\mu}_{YEN}$	0.0001	8.2440e–05	0.1597
	$\hat{c}_{YEN}$	4.6302e–07	1.1226e–07	3.7100e–05
	$\hat{a}_{1,YEN}$	0.0726	0.0076	1.7600e–21
	$\hat{b}_{1,YEN}$	0.9217	0.0080	0.0000

Table 4.6: *GARCH*$(1,1)$ *estimates for each series in the dataset.*

Copula under H_0	Test statistic	p–value
Student 14.87 **df**	0.0121	0.0974
Normal	0.0129	0.0794

Table 4.7: *Goodness–of–fit tests under the null hypothesis that the residual copula is Normal and* $\mathbf{t}_{14.87}$ *Student, respectively (significant statistics at the 5% level are marked with an asterisk).*

Copula under H_0	Number of factors	df	T_τ	T_{MPL}	$\chi^2_{df;0.95}$
Normal	1	20	1536.3500	7592.3700	31.4104
	2	13	248.6181	254.8052	22.3620
	3	7	25.6227	20.4397	14.0671
	4	2	2.3506	0.3972	5.9915
Student 14.87 **df**	1	20	1536.3500	2311.9470	31.4104
	2	13	248.6181	238.5667	22.3620
	3	7	25.6227	19.7557	14.0671
	4	2	2.3506	0.3115	5.9915

Table 4.8: *Test statistics* (3.9) *via Kendall's tau* (T_τ) *and maximum pseudo–likelihood* (T_{MPL}) *applied to the standardized residuals under various numbers of factors and assuming Normal and* $\mathbf{t}_{14.87}$ *Student copula, respectively.*

Chapter 5

Concluding remarks and discussions

Along this doctoral dissertation we devote ourselves to the study of moment structure models as one of the oldest field of applications of statistics. These models are typically characterized by the assumption of linearity and normality for observed variables. Our aim was to relax the underlying conditions in order to catch the widest class of non necessarily linear dependence structures. For doing that, we firstly exploited the inferential tools provided by the seminal paper of Browne (1984). Secondly, we based our approach to estimate and test the models on copulas as Klüppelberg and Kuhn (2009) have recently investigated. We were carried away by the persuasion that copula functions can play an important role in statistical modeling and the huge number of contributions in the last ten years are an overwhelming evidence.

Klüppelberg and Kuhn (2009) only focused their attention to elliptical copulas. In this dissertation we extended the methodology to other copula families. We found that the EFGM copula discussed by Cambanis (1991) can be profitably used in correlation structure analysis. An extension of Klüppelberg and Kuhn (2009)'s approach is not trivial and requires some restrictive conditions on copula parameters. Dependence properties of copulas are rarely closely related with linear correlation coefficients because they mainly represent nonlinear dependence. Besides elliptical copulas we do not recognize other families whose dependence parameter vector coincides with Pearson's linear correlation coefficients. Therefore, an analytical one–to–one relation between correlation matrix and copula parameters has to be detected for any different copula models. For instance for elliptical copulas the link provided by Fang, Fang, and Kotz (2002, Theorem 3.1) is at our disposal. In case of EFGM, Schucany, Parr, and Boyer (1978) suggest a way that we exploited in order to obtain a similar result. On the contrary the well known Archimedean copulas as for instance Clayton and Frank (Clayton, 1978; Frank, 1979) can not be used here because of the exchangeability and the paucity of parameters (generally, 1 or 2).

Klüppelberg and Kuhn (2009) proposed to obtain a copula–based correlation matrix for correlation structure model by using Kendall's tau matrix. That is, a *bridge* between correlation matrix and Kendall's tau matrix to carry out with copula parameters. We recognize some possible drawbacks with this choice. Firstly, the required analytic relation between Kendall's τ and copula parameters does not exist for all copulas; e.g., the so-called Joe family of copulas (Joe, 1993), Galambos family of copulas (Galambos, 1975), and Hüsler–Reiss family of copulas (Hüsler and Reiss, 1989). Secondly, Kendall's τ does not depend on the magnitude of the data and it neglects large and small values (Mikosch, 2006). Moreover, unless r is the generator of the multivariate normal distribution, $\tau_{i,j} = \rho_{i,j} = 0$ never corresponds to independence; see Section 3.1.1. We also noted from the simulation study in Section 4.3 that the use of the same relation in case of elliptical copulas can not be able to distinguish different elements belonging to the same family. In order to make this research less expensive, we provide a correlation structure analysis through the maximum pseudo–likelihood–based copula parameters estimates. Hence, we suggest to give up the so–called moment–based procedure involving Kendall's tau matrix and to focus the attention on the direct link between correlation matrix and copula parameters, by using maximum pseudo–likelihood–based estimates. The need of an analytic one–to–one relation between copula parameters and correlation coefficients clearly remains, but there is no more the need of a link with the concordance measure.

We carried out a comprehensive simulation experiment in Section 4.3 in order to assess the performances of the proposed maximum pseudo–likelihood–based test statistic for testing an elliptical copula factor model and to compare the results with those obtained by Klüppelberg and Kuhn (2009). We pointed out that our test statistic resulted more conservative and powerful than the Kendall's tau–based counterpart. Moreover, an application to real data has shown that the interpretation of latent factors via maximum pseudo–likelihood–based inferential procedure is reasonable as well as that supplied by Klüppelberg and Kuhn (2009).

We conclude with a summary about the improvements provided by copulas in moment structure analysis in connection with Browne (1984). We have just mentioned the opportunity given by copulas to capture a wider range of dependence structures. An other important benefit is to avoid heavy calculations for higher–order moments planned by Browne (1984); i.e., the fourth–order moment estimation. In case of elliptical copulas we are able to provide a correlation–like matrix without assumptions about the existence of moments. Nevertheless, in case of EFGM copula we require existing first and second moments; see Theorem 4.1 and Corollary 4.2. Anyway, we recognize a clear decreased computational effort in comparison with Browne (1984)'s contribution.

Appendix A

Kronecker products and the Vec, Vech, and patterned Vec operators

We denote by $\otimes$ the right Kronecker product. The Kronecker product of two matrices, say a $m \times n$ matrix $\boldsymbol{B}$ and a $p \times q$ matrix $\boldsymbol{C}$, is denoted by the symbol $(\boldsymbol{B} \otimes \boldsymbol{C})$ and is defined to be a $mp \times nq$ matrix obtained by replacing each element $[\boldsymbol{B}]_{i,j}$ of $\boldsymbol{B}$ with the $p \times q$ matrix $[\boldsymbol{B}]_{i,j}\,\boldsymbol{C}$. Thus, the Kronecker product of $\boldsymbol{B}$ and $\boldsymbol{C}$ is a partitioned matrix, comprising m rows and n columns of $p \times q$ dimensional blocks, the i, j–th of which is $[\boldsymbol{B}]_{i,j}\,\boldsymbol{C}$.

Let $\boldsymbol{A}$ represent an $n \times n$ symmetric matrix (but we could also consider non–symmetric matrix). We denote by $vec\,(\boldsymbol{A})$ the $n^2 \times 1$ column vector obtained by stacking the columns $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, …, $\boldsymbol{a}_n$ of $\boldsymbol{A}$, and by positioning them one under the other. The following theorem provides a general result widely exploited in this work.

Theorem A.1 (Harville, 1997, Theorem 16.2.1) *For any $m \times n$ matrix $\boldsymbol{D}$, $n \times p$ matrix $\boldsymbol{E}$, and $p \times q$ matrix $\boldsymbol{F}$,*

$$vec\,(\boldsymbol{DEF}) = \left(\boldsymbol{F}^{\top} \otimes \boldsymbol{D}\right) vec\,(\boldsymbol{E})\ . \tag{A.1}$$

Generally speaking, let $\boldsymbol{A}$ represent an $m \times n$ matrix, and denote the first, …, n–th columns of $\boldsymbol{A}$ by $\boldsymbol{a}_1$, …, $\boldsymbol{a}_n$, respectively, and the first, …,

m–th rows of $\boldsymbol{A}$ by $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m$, respectively. Both $vec\left(\boldsymbol{A}^\top\right)$ and $vec\left(\boldsymbol{A}\right)$ are obtained by rearranging the elements of $\boldsymbol{A}$ in the form of an mn–dimensional column vector. However, they are arranged row by row in $vec\left(\boldsymbol{A}^\top\right)$ instead of column by column, as in $vec\left(\boldsymbol{A}\right)$. Clearly, $vec\left(\boldsymbol{A}^\top\right)$ can be obtained by permuting the elements of $vec\left(\boldsymbol{A}\right)$. Accordingly, there exists an $mn \times mn$ permutation matrix, to be denoted by the symbol $\boldsymbol{K}_{mn}$, such that

$$vec\left(\boldsymbol{A}^\top\right) = \boldsymbol{K}_{mn}\, vec\left(\boldsymbol{A}\right) .$$

The matrix $\boldsymbol{K}_{mn}$ is referred to as a *vec–permutation matrix* (e.g., Henderson and Searle, 1979) or, more commonly, as a *commutation matrix* (e.g., Magnus and Neudecker, 1979). Note that, since the transpose $\boldsymbol{A}^\top$ of the $m \times n$ matrix $\boldsymbol{A}$ is of dimensions $n \times m$, it follows that

$$vec\left(\boldsymbol{A}\right) = vec\left\{\left(\boldsymbol{A}^\top\right)^\top\right\} = \boldsymbol{K}_{nm}\, vec\left(\boldsymbol{A}^\top\right) = \boldsymbol{K}_{nm}\, \boldsymbol{K}_{mn}\, vec\left(\boldsymbol{A}\right) ,$$

implying that $\boldsymbol{I}_{mn} = \boldsymbol{K}_{nm}\, \boldsymbol{K}_{mn}$. Thus $\boldsymbol{K}_{mn}$ is nonsingular and $\boldsymbol{K}_{mn}^{-1} = \boldsymbol{K}_{nm}$.

Let $vech\left(\boldsymbol{A}\right)$ represent the $n^\star \times 1$ column vector formed from the non–duplicated elements of the symmetric $\boldsymbol{A}$, where $n^\star = n\left(n+1\right)/2$. Let $\boldsymbol{G}_n$ be the *transition* or *duplication matrix* of order $n^2 \times n^\star$ such that $vec\left(\boldsymbol{A}\right) = \boldsymbol{G}_n\, vech\left(\boldsymbol{A}\right)$ and $vech\left(\boldsymbol{A}\right) = \boldsymbol{H}_n\, vec\left(\boldsymbol{A}\right)$ for every symmetric matrix $\boldsymbol{A}$, where one choice for $\boldsymbol{H}_n$ is $\boldsymbol{H}_n = \left(\boldsymbol{G}_n^\top \boldsymbol{G}_n\right)^{-1} \boldsymbol{G}_n^\top$ (since $\boldsymbol{G}_n$ is of full column rank, $\boldsymbol{G}_n^\top \boldsymbol{G}_n$ is nonsingular).

For instance, $\boldsymbol{G}_1 = (1)$, $\boldsymbol{G}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and $\boldsymbol{G}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$, and $\boldsymbol{H}_1 = (1)$, $\boldsymbol{H}_2 = \begin{pmatrix} 1 & 0.0 & 0.0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.0 & 0.0 & 1 \end{pmatrix}$, and $\boldsymbol{G}_3 = \begin{pmatrix} 1 & 0.0 & 0.0 & 0.0 & 0 & 0.0 & 0.0 & 0.0 & 0 \\ 0 & 0.5 & 0.0 & 0.5 & 0 & 0.0 & 0.0 & 0.0 & 0 \\ 0 & 0.0 & 0.5 & 0.0 & 0 & 0.0 & 0.5 & 0.0 & 0 \\ 0 & 0.0 & 0.0 & 0.0 & 1 & 0.0 & 0.0 & 0.0 & 0 \\ 0 & 0.0 & 0.0 & 0.0 & 0 & 0.5 & 0.0 & 0.5 & 0 \\ 0 & 0.0 & 0.0 & 0.0 & 0 & 0.0 & 0.0 & 0.0 & 1 \end{pmatrix}$.

Let $\boldsymbol{L}$ represent a $n \times n$ matrix and $\boldsymbol{X}$ a $n \times n$ symmetric matrix. Observe (in light of Theorem A.1) that

$$vec\left(\boldsymbol{L}\boldsymbol{X}\boldsymbol{L}^{\top}\right) = (\boldsymbol{L}\otimes\boldsymbol{L})\, vec\,(\boldsymbol{X}) = (\boldsymbol{L}\otimes\boldsymbol{L})\, \boldsymbol{G}_n\, vech\,(\boldsymbol{X})\ . \tag{A.2}$$

Observe also that (since $\boldsymbol{L}\boldsymbol{X}\boldsymbol{L}^{\top}$ is symmetric)

$$vech\left(\boldsymbol{L}\boldsymbol{X}\boldsymbol{L}^{\top}\right) = \boldsymbol{H}_n\, vec\left(\boldsymbol{L}\boldsymbol{X}\boldsymbol{L}^{\top}\right)\ . \tag{A.3}$$

Together, equations (A.2) and (A.3) imply that

$$vech\left(\boldsymbol{L}\boldsymbol{X}\boldsymbol{L}^{\top}\right) = \boldsymbol{H}_n\ (\boldsymbol{L}\otimes\boldsymbol{L})\, \boldsymbol{G}_n\, vech\,(\boldsymbol{X})\ . \tag{A.4}$$

Result (A.4) can be regarded as the *vech* counterpart of formula (A.1) for the *vec* of a product of matrices.

Consider now a modification of the stacking columns process in which (before or after the stacking) the $n\,(n+1)\,/2$ "supra–diagonal" elements of $\boldsymbol{A}$, included diagonal elements, are eliminated from $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, …, $\boldsymbol{a}_n$. The result is the $\{n\,(n-1)\,/2\}$–dimensional vector

$$\begin{bmatrix} \boldsymbol{a}_1^{\star} \\ \boldsymbol{a}_2^{\star} \\ \vdots \\ \boldsymbol{a}_n^{\star} \end{bmatrix}, \tag{A.5}$$

where $\boldsymbol{a}_i^{\star} = \left([\boldsymbol{A}]_{i+1,i}\,, [\boldsymbol{A}]_{i+2,i}\,, \ldots, [\boldsymbol{A}]_{n,i}\right)^{\top}$ for $i = 1, 2, \ldots, n$ is the subvector of $\boldsymbol{a}_i$ obtained by striking out its first i elements. Thus, by definition, the vector (A.5) is a subvector of $vec\,(\boldsymbol{A})$ obtained by striking out a particular set of duplicate or redundant elements (in the special case where $\boldsymbol{A}$ is symmetric).

Following Henderson and Searle (1979) and Nel (1985), let us refer to the vector (A.5) as the *vecp* of $\boldsymbol{A}$ – think of *vecp* as being an abbreviation for vector of patterned matrix. Denote this vector by the symbol $vecp\,(\boldsymbol{A})$. Like $vec\,(\boldsymbol{A})$, $vecp\,(\boldsymbol{A})$ can be regarded as the value assigned to $\boldsymbol{A}$ by a vector–valued function or operator, whose domain is $\mathbb{R}^{n\times n}$. A symmetric matrix $\boldsymbol{A}$ is called patterned if $\boldsymbol{A}$ has $n^2 - r - c$ mathematically independent and variable elements, where r denotes the number of elements which are repeated, even with a negative sign, and c denotes the number of constant elements of $\boldsymbol{A}$. If we refer to the pattern of $\boldsymbol{A}$, we refer to the positions and signs of the mathematically independent and variable elements of $\boldsymbol{A}$. Correlation matrices are typical examples of squared patterned matrices. Other examples are given in Henderson and Searle (1979).

If a patterned matrix $\boldsymbol{A}$ is stacked column–wise in order of appearance into a single column vector, we use the convention that constant elements in $\boldsymbol{A}$ are replaced with zeros in $vec\,(\boldsymbol{A})$. Thus we are only interested in the positions of the mathematically independent and variable elements in $\boldsymbol{A}$ and $vec\,(\boldsymbol{A})$, and not in the numerical values of the constant elements.

Let us consider the pattern of a correlation matrix. Observe that the total number of elements in the j vectors $\boldsymbol{a}_1^\star$, $\boldsymbol{a}_2^\star$, …, $\boldsymbol{a}_j^\star$ is

$$(n-1)+(n-1)-1+\ldots+(n-1)-(j-1)= \\ = nj-j-(0+1+\ldots+j-1)=nj-\frac{j\,(j+1)}{2}$$

and that, of the $(n-1)-(j-1)=n-j$ elements of $\boldsymbol{a}_j^\star$, there are (for $i>j$) $n-i$ elements that come after $a_{i,j}$. Since $nj-j\,(j+1)\,/2-(n-i)=n\,(j-1)-j\,(j+1)\,/2+i$, it follows that (for $i>j$) the i,j–th element of $\boldsymbol{A}$ is the $\{(j-1)\,n-j\,(j+1)\,/2+i\}$–th element of $vecp\,(\boldsymbol{A})$. By way of comparison, the i,j–th element of $\boldsymbol{A}$ is the $\{(j-1)\,n+i\}$–th element of $vec\,(\boldsymbol{A})$, so that (for $i>j$) the $\{(j-1)\,n+i\}$–th element of $vec\,(\boldsymbol{A})$ is the $\{(j-1)\,n-j\,(j+1)\,/2+i\}$–th element of $vecp\,(\boldsymbol{A})$.

Since non-fixed elements of a $n\times n$ symmetric matrix $\boldsymbol{A}$, and hence every non–fixed elements of $vec\,(\boldsymbol{A})$, are either elements of $vecp\,(\boldsymbol{A})$, there exists a unique $n^2\times n^{\star\star}$ matrix, where $n^{\star\star}=n\,(n-1)\,/2$, to be denoted by the symbol $\boldsymbol{Q}_n$, such that $vec\,(\boldsymbol{A})=\boldsymbol{Q}_n\,vecp\,(\boldsymbol{A})$. Since the duplication matrix $\boldsymbol{Q}_n$ is of full column rank, it has a left inverse. Thus, by definition, $\boldsymbol{P}_n$ is a $n^{\star\star}\times n^2$ matrix such that $vecp\,(\boldsymbol{A})=\boldsymbol{P}_n\,vec\,(\boldsymbol{A})$ for every symmetric matrix $\boldsymbol{A}$, where one choice for $\boldsymbol{P}_n$ is $\boldsymbol{P}_n=\left(\boldsymbol{P}_n^\top\boldsymbol{P}_n\right)^{-1}\boldsymbol{P}_n^\top$ (since $\boldsymbol{P}_n$ is of full column rank, $\boldsymbol{P}_n^\top \boldsymbol{P}_n$ is nonsingular). Finally, *mutatis mutandis*, formula (A.4) still holds for *vecp* operator.

Appendix B

L–moments

L–moments are analogous to the conventional moments but they can be estimated by linear combinations of order statistics; i.e., L–statistics. The L in L–moments emphasizes the fact that L–moments are linear functions of the expected order statistics. These moments have the theoretical advantages over conventional moments of being able to characterize at wider range of distributions and, when estimated from a sample, of being more robust to the presence of outliers in the data. Moreover, they approximate their asymptotic normal distribution more closely in finite sample. For a detailed review on the L–moment see Hosking (1990) and David and Nagaraja (2003).

Let x be a real random variable with distribution function F, and let $x_{1:n} \leq \ldots \leq x_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of x. Define the L–moments of x as the quantity

$$\delta_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E\left(x_{r-k;r}\right) \quad r = 1, 2, \ldots . \tag{B.1}$$

The expectation of an order statistic may be written as

$$E\left(x_{j;r}\right) = \frac{r!}{(j-1)!\,(r-j)!} \int x^{j-1} \left\{1 - F\left(x\right)\right\}^{r-j} dF\left(x\right) . \tag{B.2}$$

Substituting (B.2) in (B.1), expanding the binomials in $F(x)$ and summing the coefficients of each power of $F(x)$ gives

$$\delta_r = \int x\, P^\star_{r-1}\left\{F(x)\right\}\, dF(x)\ ,$$

where

$$P^\star_r\left\{F(x)\right\} = \sum_{k=0}^{r} p^\star_{r,k}\ \left\{F(x)\right\}^k \tag{B.3}$$

and

$$p^\star_{r,k} = (-1)^{r-k} \binom{r}{k}\binom{r+k}{k}.$$

$\delta_1 = \int x\, dF(x)$, the mean, is a measure of location. To interpret

$$\delta_2 = \frac{1}{2}\, E\left(x_{2:2} - x_{1:2}\right) = \int F^{-1}(x)\left[2F(x) - 1\right]\, dF(x)$$

consider the typical configuration of a sample of size 2. If the two values tend to be close together, then δ_2 will be smaller than if they are far apart. Thus, δ_2 can be thought of as measuring the scale or dispersion of the distribution. To compare δ_2 with the more familiar scale measure σ, the standard deviation, write

$$\delta_2 = \frac{1}{2}E\left(x_{2:2} - x_{1:2}\right)\ , \quad \sigma^2 = \frac{1}{2}E\left(x_{2:2} - x_{1:2}\right)^2\ .$$

Both quantities measure the difference between two randomly drawn elements of a distribution, but σ^2 gives relatively more weight to the largest differences.

The use of L–moments to describe probability distributions is justified by the following theorem.

Theorem B.1 (Hosking, 1990, Theorem 1) *Let x be a real random variable.*

(a) *The L–moments δ_r, $r = 1, 2, \ldots$, of x exist if and only if x has finite mean.*

(b) *A distribution whose mean exists is characterized by its L–moments $\{\delta_r; r = 1, 2, \ldots\}$.*

Thus a distribution may be specified by its L–moments even if some of its conventional moments do not exist. Moreover, under a linear transformation of the data, the sample L–moments are transformed isomorphically with the corresponding population L–moments. If $x_i \to A\, x_i + B \;\; \forall\, i = 1, \ldots, n$, then $\delta_1 \to A\, \delta_1 + B$ and $\delta_r \to (signA)^r\, A\, \delta_r \;\; \forall\, r \geq 2$.

The natural estimator of δ_r based on an observed sample of data is a linear combination of the ordered data values; i.e., the L–statistics. Because δ_r is a function of the expected order statistics of a sample of size r, it is natural to estimate it by a U–statistic (Lee, 1990); i.e., the corresponding function of the sample order statistics averaged over all subsamples of size r which can be constructed from the observed sample of size n. Let $x_1, \ldots, x_n$ be the sample and $x_{1:n} \leq \ldots \leq x_{n:n}$ the ordered sample, and define the r–th sample L–moment as

$$\hat{\delta}_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1 < \ldots < i_r \leq n} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_{r-k}:n} \qquad r = 1, \ldots, n\,.$$

In particular,

$$\hat{\delta}_2 = \frac{1}{2} \binom{n}{2}^{-1} \sum_{i>j} (x_{i:n} - x_{j:n})\;.$$

It is now clear that δ_2 is a scalar multiple of Gini's mean difference statistic $G = \binom{n}{2}^{-1} \sum_{i>j} (x_{i:n} - x_{j:n})$. Nevertheless, it is not necessary to iterate over all subsamples of size r. The statistics can be expressed explicitly as a linear combination of order statistics of a sample of size n. Wang (1996) suggests to estimate L–moments following closely their definition; for instance, the second sample L–moment can be defined as follows,

$$\hat{\delta}_2 = \frac{1}{2} \binom{n}{2}^{-1} \sum_{i=1}^{n} [2i - (n+1)]\, x_{i:n}\,. \tag{B.4}$$

Regarding the shape of the sampling distributions of L–moments, exact sampling distributions are difficult to obtain. The most practically useful results come from asymptotic distribution theory.

Theorem B.2 (Hosking, 1990, Theorem 3) *Let x be a real continuous random variable with distribution function F, L–moments δ_r and finite variance. Let $\hat{\delta}_r$, $r = 1, \ldots, m$, be the sample L–moments calculated from a random sample of size n drawn from the distribution of x. Let $\beta_r = \delta_r / \delta_2$ and $b_r = \hat{\delta}_r / \hat{\delta}_2$, $r = 3, \ldots, m$. Then, as $n \to \infty$,*

$$n^{1/2} \left(\hat{\boldsymbol{l}} - \boldsymbol{l} \right) \xrightarrow{\mathscr{L}} \mathsf{N} \left(\mathbf{0}, \boldsymbol{\Sigma}_L \right) ,$$

where $\hat{\boldsymbol{l}} = \left(\hat{\delta}_1, \hat{\delta}_2, b_3, \ldots, b_m \right)^\top$, $\boldsymbol{l} = (\delta_1, \delta_2, \beta_3, \ldots, \beta_m)^\top$, and the elements of $\boldsymbol{\Sigma}_L = \left[\sigma^L_{r,s} \right]_{1 \le r,s \le m}$ are equal to

$$\sigma^L_{r,s} = \begin{cases} \varsigma_{rs} & r \le 2, s \le 2, \\ \left(\varsigma_{rs} - \beta_r \varsigma_{2s} \right) / \delta_2 & r \ge 3, s \le 2, \\ \left(\varsigma_{rs} - \beta_r \varsigma_{2s} - \beta_s \varsigma_{2r} + \beta_r \beta_s \varsigma_{22} \right) / \delta_2^2 & r \ge 3, s \ge 3, \end{cases}$$

with

$$\varsigma_{rs} = \int_0^1 \int_0^1 \left\{ P^\star_{r-1}(u) \quad P^\star_{s-1}(v) + P^\star_{s-1}(u) \quad P^\star_{r-1}(v) \right\} \times \\ \times u (1 - v) \left. \frac{\partial}{\partial x} F^{-1}(x) \right|_{x=u} \left. \frac{\partial}{\partial x} F^{-1}(x) \right|_{x=v} du dv \,,$$

$P^\star_r(u)$ is given in (B.3), and $u = F(x)$.

Bibliography

Abdous, B., C. Genest, and B. Rèmillard (2005). “Dependence properties of metaelliptical copulas”. In: *Statistical modeling and analysis for complex data problems.* Vol. 1. GERARD 25th Anniv. Ser. New York: Springer, pp. 1–15.

Anderson, T. W. (1976). “Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics (with discussion)”. In: *J. Roy. Stat. Soc. B* 38, pp. 1–36.

Andrews, D. W. K. (1988). “Laws of large numbers for dependent non-identically distributed random variables”. In: *Economet. Theor.* 4, pp. 458–467.

Barzilai, J. and J. M. Borwein (1988). “Two–point step size gradient methods”. In: *IMA J. Numer. Anal.* 8, pp. 141–148.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). “Multivariate GARCH models: a survey”. In: *J. Appl. Econom.* 21, pp. 79–109.

Bell, R. C. et al. (1990). *Teacher stress in Victoria.* Melbourne: Victorian Minist. Educ.

Bentler, P. M. (1980). “Multivariate analysis with latent variables: causal modeling”. In: *Annu. Rev. Psychol.* 31, pp. 419–456.

— (1983). “Some contributions to efficient statistics for structural models: specification and estimation of moment structures”. In: *Psychometrika* 48, pp. 493–517.

Bentler, P. M. and T. Dijkstra (1983). “Efficient estimation via linearization in structural models”. In: ed. by P. R. Krishnaiah. Multivariate Analysis-VI. New York: Elsevier, pp. 9–42.

Bentler, P. M. and P. Dudgeon (1996). "Covariance structure analysis: statistical practice, theory, and directions". In: *Annu. Rev. Psychol.* 47, pp. 541–570.

Bentler, P. M. and E. J. C. Wu (1995a). *EQS for Machintosh user's guide.* Encinno, CA: Multivariate Software.

— (1995b). *EQS for Windows user's guide.* Encinno, CA: Multivariate Software.

Billingsley, P. (1968). *Convergence of probability measures.* New York: Wiley.

Boyer, H. B., M. S. Gibson, and M. Loretan (1999). "Pitfalls in tests for changes in correlations". In: *Federal Reserve Boars, IFS Discussion Paper No. 597R*, p. 597.

Browne, M. W. (1974). "Generalized least squares estimators in the analysis of covariance structures". In: *S. Afr. Stat. J.* 8, pp. 1–24.

— (1982). "Covariance structures". In: *Topics in Applied Multivariate Analysis.* Ed. by D. M. Hawkins. Cambridge: Cambridge University Press, pp. 72–141.

— (1984). "Asymptotically distribution-free methods for the analysis of covariance structures". In: *Brit. J. Math. Stat. Psy.* 37, pp. 62–83.

Burchard, A. and H. Hajaiej (2006). "Rearrangement inequalities for functionals with monotone integrands". In: *J. Funct. Anal.* 233, pp. 561–582.

Cambanis, S. (1977). "Some properties and generalizations of multivariate Eyraud–Gumbel–Morgenstern distributions". In: *J. Multivariate Anal.* 7, pp. 551–559.

— (1991). "On Eyraud–Farlie–Gumbel–Morgenstern random processes". In: *Advances in probability distributions with given marginals: beyond the copulas.* Ed. by G. Dall'Aglio, S. Kotz, and G. Salinetti, pp. 207–222.

Cebrian, A., M. Denuit, and O. Scaillet (2004). "Testing for concordance ordering". In: *Astin Bull.* 34, pp. 151–173.

Chamberlain, G. (1982). "Multivariate regression models for panel data". In: *J. Econometrics* 18, pp. 5–46.

Chen, X. and Y. Fan (2006). "Estimation and model selection of semiparametric copula–based multivariate dynamic models under copula misspecification". In: *J. Econom.* 135, pp. 125–154.

Chen, X., Y. Fan, and V. Tsyrennikov (2006). "Efficient estimation of semiparametric multivariate copula models". In: *J. Am. Stat. Assoc.* 101, pp. 1228–1240.

Cherubini, U., E. Luciano, and W. Vecchiato (2004). *Copula methods in finance*. The Wiley Finance Series. Chichester: John Wiley & Sons.

Clayton, D. G. (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence." In: *Biometrika* 65, pp. 141–151.

Cudeck, R. (1989). "Analysis of correlation matrices using covariance structure models". In: *Psychol. Bull.* 105, pp. 317–327.

David, H. A. and H. N. Nagaraja (2003). *Order Statistics, 3rd edn.* Wiley Series in Probability and Statistics. Hoboken: Wiley.

Demarta, S. and A. J. McNeil (2005). "The t copula and related copulas". In: *Int. Stat. Rev.* 73, pp. 111–129.

Dolby, G. R. and T. Freeman (1975). "Functional reletionships having many independent variates and errors with multivariate normal distribution". In: *J. Multivariate Anal.* 5, pp. 466–479.

Doukhan, P. (1994). *Mixing: properties and examples.* Vol. 85. Lecture Notes in Statistics. New York: Springer–Verlag.

Drouet-Mari, D. and S. Kotz (2001). *Correlation and dependence.* London: Imperial College Press.

Durante, F. and C. Sempi (2010). "Copula Theory: an Introduction". In: *Workshop on Copula Theory and its Applications.* Ed. by P. Jaworski et al. Lecture Notes in Statistics. Berlin/Heidelberg: Springer.

Embrechts, P., A. McNeil, and D. Straumann (1999). *Correlation and dependence in risk management: properties and pitfalls.* Tech. rep.

Eyraud, H. (1936). "Les principes de la mesure des correlations". In: *Ann. Univ. Lyon, III. Ser., Sect. A* 1, pp. 30–47.

Fang, H. B., K. T. Fang, and S. Kotz (2002). "The meta–elliptical distributions with given marginals". In: *J. Multivariate Anal.* 82, pp. 1–16.

Fang, K. T., S. Kotz, and K. W. Ng (1990). *Symmetric Multivariate and Related Distributions.* London: Chapman and Hall.

Farlie, D. J. G. (1960). "The performance of some correlation coefficients for a general bivariate distribution". In: *Biometrika* 47, pp. 307–323.

Ferguson, T. S. (1958). "A method for generating best asymptotically normal estimates with applications to the estimation of bacterial densities". In: *Ann. Math. Stat.* 29, pp. 1046–1062.

— (1996). *A course in large sample theory.* London: Chapman and Hall.

Fermanian, J.-D. and O. Scaillet (2004). "Some statistical pitfalls in copula modeling for financial applications". In: *HEC Genève.*

Frahm, G., M. Junker, and A. Szimayer (2003). "Elliptical copulas: Applicability and limitations". In: *Stat. Probabil. Lett.* 63, pp. 275–286.

Frank, M. J. (1979). "On the simultaneous associativity of F(x,y) and x+y-F(x,y)." In: *Aequationes Math.* 19, pp. 194–226.

Fréchet, M. (1951). "Sur les tableaux de corrélation dont les marges sont données." In: *Ann. Univ. Lyon. Sect. A* 14, pp. 53–77.

Friedlander, A., J. M. Martínez, and M. Raydan (1995). "A new method for large–scale box constrained convex quadratic minimization problems". In: *Optim. Method. Softw.* 5, pp. 57–74.

Galambos, J. (1975). "Order statistics of samples from multivariate distributions". In: *J. Am. Stat. Assoc.* 70, pp. 674–680.

Genest, C. and A.-C. Favre (2007). "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask". In: *Journal of Hydrologic Engineering* 12, pp. 347–368.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995). "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions." In: *Biometrika* 82, pp. 543–552.

Genest, C., B. Rémillard, and D. Beaudoin (2009). "Goodness-of-fit tests for copulas: a review and a power study". In: *Insur. Math. Econ.* 44, pp. 199–213.

Genest, C. and L.-P. Rivest (1993). "Statistical inference procedures for bivariate Archimedean copulas". In: *J. Am. Stat. Assoc.* 88, pp. 1034–1043.

Gleser, L. J. (1981). "Estimation in a multivariate "errors in variables" regression model: large sample results". In: *Ann. Stat.* 9, pp. 24–44.

Gumbel, E J. (1958). "Distributions des valeurs extrêmes en plusieurs dimensions." In: *Publ. Inst. Stat. Univ. Paris* 9, pp. 171–173.

— (1960). "Bivariate exponential distributions". In: *J. Am. Stat. Assoc.* 55, pp. 698–707.

Haavelmo, T. (1944). "The probability approach in econometrics". In: *Econometrica* 12, pp. 1–118.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Applications.* New York: Academic Press.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective.* New York: Springer–Verlag.

Henderson, H. V. and S. R. Searle (1979). "Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics". In: *Can. J. Stat.* 7, pp. 65–81.

Heywood, H. B. (1931). "On finite sequences of real numbers". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character.* Vol. 134, pp. 486–501.

Higham, N. (2002). "Computing the nearest correlation matrix – a problem from finance". In: *IMA J. Numer. Anal.* 22, pp. 329–343.

Hoeffding, W. (1940). "Massstabinvariante korrelationstheorie". In: *Schriften Math. Inst. Angew. Math. Univ. Berlin* 5, pp. 179–233.

— (1994). *The collected works of Wassily Hoeffding.* Ed. by N. I. Fisher and P. K. Sen. New York: Springer-Verlag.

Hosking, J. R. M. (1990). "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics". In: *J. Roy. Stat. Soc. B* 52, pp. 105–124.

Hotelling, H. (1933). "Analysis of a Complex of Statistical Variables into Principal Components". In: *J. Educ. Psychol.* 24, pp. 417–441, 498–520.

Hu, L., P. M. Bentler, and Y. Kano (1992). "Can test statistics in covariance structure analysis be trusted?" In: *Psychol. Bull.* 112, pp. 351–362.

Huba, G. J., J. A. Wingard, and P. M. Bentler (1981). "A comparison of two latent variable causal models for adolescent drug use". In: *J. Pers. Soc. Psychol.* 40, pp. 180–193.

Hüsler, J. and R.-D. Reiss (1989). "Maxima of normal random vectors: between independence and complete dependence". In: *Stat. Probabil. Lett.* 7, pp. 283–286.

Ibragimov, I. A. (1962). "Some limit theorems for stationary processes". In: *Theor. Prob. Appl.* 7, pp. 349–382.

Joe, H. (1993). "Parametric families of multivariate distributions with given margins". In: *J. Multivariate Anal.* 46, pp. 262–282.

— (1997). *Multivariate models and dependence concepts*. Monographs on Statistics, Applied Probability. 73. London: Chapman, and Hall.

Johnson, N. L. and S. Kotz (1977). "On some generalized Farlie–Gumbel–Morgenstern distributions - II regression, correlation and further generalizations". In: *Commun. Stat. Theory* 6, pp. 485–496.

Jöreskog, K. G. (1963). *Statistical estimation in factor analysis: a new technique and its foundation*. Stockholm: Almqvist & Wiksell.

— (1967). "Some contributions to maximum likelihood factor analysis". In: *Psychometrika* 32, pp. 443–482.

— (1973). "A general method for estimating a linear structural equation system". In: *Structural equation models in the social sciences*. Ed. by A. S. Goldberger and O. D. Duncan. New York: Academic Press, pp. 85–112.

— (1977). "Structural equation models in the social sciences: specification estimation and testing". In: *Applications of statistics*. Ed. by P. R. Krishnaiah. Amsterdam: North Holland, pp. 265–287.

— (1978). "Structural analysis of covariance and correlation matrices". In: *Psychometrika* 43, pp. 443–477.

Jöreskog, K. G. and A. S. Goldberger (1972). "Factor analysis by generalized least squares". In: *Psychometrika* 37, pp. 243–260.

Jöreskog, K. G. and D. Sörbom (1983). *LISREL 8 user's reference guide*. Chicago: Scientifc Software International.

Kano, Y., M. Berkane, and P. M. Bentler (1990). "Covariance structure analysis with heterogeneous kurtosis parameters". In: *Biometrika* 77, pp. 575–585.

Klüppelberg, C. and G. Kuhn (2009). "Copula structure analysis". In: *J. Roy. Stat. Soc. B* 71, pp. 737–753.

Kojadinovic, I. and J. Yan (2010a). “A goodness–of–fit test for multivariate multiparaemter copulas based on multiplier central limit theorems”. In: *Stat. Comput.* in press (2).

— (2010b). “Comparison of three semiparametric methods for estimating dependence parameters in copula models”. In: *Insur. Math. Econ.* 47 (2), pp. 52–63.

— (2010c). “Modeling multivariate distributions with continuous margins using the copula `R` package”. In: *J. Stat. Soft.* 34 (2), pp. 1–20.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2000). *Continuous multivariate distributions, models and applications.* 2nd. Vol. 1. Wiley, New York.

Kruskal, W. H. (1958). “Ordinal measures of association.” In: *J. Am. Stat. Assoc.* 53, pp. 814–861.

Lawley, D. N. and A. E. Maxwell (1971). *Factor analysis as a statistical model, 2nd edn.* New York: Elsevier.

Lee, A. J. (1990). *U–statistics.* Dekker, New York.

Lee, S. Y. and P. M. Bentler (1980). “Some asymptotic properties of constrained generalized least squares estimation in covariance structure models”. In: *S. Afr. Stat. J.* 14, pp. 121–136.

Lehmann, E. L. (1983). *Theory of point estimation.* New York: Wiley.

Lehmann, E. L. and G. Casella (1998). *Theory of point estimation.* New York: Springer.

Magnus, J. R. and H. Neudecker (1979). “The commutation matrix: some properties and applications”. In: *Ann. Stat.* 7, pp. 381–394.

Malevergne, Y. and D. Sornette (2003). “Testing the Gaussian copula hypothesis for financial assets dependences”. In: *Quant. Financ.* 3, pp. 231–250.

Mardia, K. V. (1970). “Measures of multivariate skewness and kurtosis with applications”. In: *Biometrika* 57, pp. 519–530.

McLeish, D. L. (1975). “A maximal inequality and dependent strong laws”. In: *Ann. Prob.* 3, pp. 829–839.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative risk management.* Princeton Univeristy Press, Princeton, NJ.

Mikosch, T. (2006). “Copulas: Tales and facts”. In: *Extremes* 9 (1), pp. 3–20.

Moore, D. S. and M. C. Spruill (1975). “Unified large–sample theory of general chi–squared statistics for tests of fit”. In: *Ann. Stat.* 3, pp. 599–616.

Morgenstern, D. (1956). “Einfache Beispiele zweidimensionaler Verteilungen”. In: *Mitteilungsbl. Math. Statist.* 8, pp. 234–235.

Nel, D. G. (1985). “A matrix derivation of the asymptotic covariance matrix of sample correlation coefficients”. In: *Linear Algebra Appl.* 67, pp. 137–145.

Nelsen, R. B. (2006). *An introduction to copulas, 2nd edn.* New York: Springer.

Newey, W. K. and D. McFadden (1994). “Large sample estimation and hypothesis testing”. In: *Handbook of econometrics.* Ed. by R. F. Engle and D. McFadden. Amsterdam: Elsevier Sci., pp. 2112–2245.

Pearson, K. (1901). “On lines and planes of closest fit to systems of points in space”. In: *Phil. Mag.* 2, pp. 559–572.

Prieger, J. E. (2002). “A flexible parametric selection model for non–normal data with application to health care usage”. In: *J. Appl. Econom.* 17, pp. 367–392.

Quesada-Molina, J. J. and J. A. Rodríguez-Lallena (1995). “Bivariate copulas with quadratic sections”. In: *J. Nonparametr. Stat.* 5, pp. 323–337.

Raydan, M. (1993). “On the Barzilai and Borwein choice of steplength for the gradient method”. In: *IMA J. Numer. Anal.* 13, pp. 321–326.

Rodriguez-Lallena, J. A. and M. Úbeda-Flores (2009). “Multivariate copulas with quadratic sections in one variable.” In: *Metrika* (*in press*).

Rousseeuw, P. J. and G. Molenberghs (1993). “Transformation of non positive semidefinite correlation matrices”. In: *Commun. Stat. Theory* 22, pp. 965–984.

Rüschendorf, L. (1976). “Asymptotic Distributions of Multivariate Rank Order Statistics”. In: *Ann. Stat.* 4, pp. 912–923.

Rüschendorf, L. (2009). “On the distributional transform, Sklar’s theorem, and the empirical copula process”. In: *J. Stat. Plan. Infer.* 139, pp. 3921–3927.

Ruymgaart, F. H. (1974). “Asymptotic Normality of Nonparametric Tests for Independence”. In: *Ann. Stat.* 2, pp. 892–910.

Ruymgaart, F. H., G. R. Shorack, and W. R. van Zwet (1972). "Asymptotic Normality of Nonparametric Tests for Independence". In: *Ann. Math. Stat.* 43, pp. 1122–1135.

Satorra, A. and P. M. Bentler (2001). "A scaled difference chi–squared test statistic for moment structure analysis". In: *Psychometrika* 66, pp. 507–514.

Schucany, W. R., W. C. Parr, and J. E. Boyer (1978). "Correlation structure in Farlie–Gumbel–Morgenstern distributions". In: *Biometrika* 65, pp. 650–653.

Schweizer, B. and A. Sklar (1974). "Operations on distribution functions not derivable from operations on random variables". In: *Stud. Math.* 52, pp. 43–52.

— (1983). *Probabilistic metric spaces.* New York: North Holland / Elsevier.

Shapiro, A. and M. W. Browne (1990). "On the treatment of correlation structures as covariance structures". In: *Linear Algebra Appl.* 127, pp. 567–587.

Shih, J. H. and T. A. Louis (1995). "Inferences on the association parameter in copula models for bivariate survival data". In: *Biometrics* 51, pp. 1384–1399.

Sklar, A. (1959). "Fonctions de répartition á n dimensions et leurs marges". In: *Publ. Inst. Stat. Univ. Paris* 8, pp. 229–231.

— (1973). "Random variables, joint distribution functions, and copulas". In: *Kybernetika* 9, pp. 449–460.

— (1996). "Random variables, distribution functions, and copulas – a personal look backward and forward". In: *Distributions with fixed marginals and related topics.* Ed. by L. Rüschendorf, B. Schweizer, and M. D. Taylor. Vol. 28. IMS Lecture Notes–Monograph Series. Hayward: Institute of Mathematical Statistics, pp. 1–14.

Spearman, C. (1904). "General intelligence objectively determined and measured". In: *Am. J. Psychol.* 15, pp. 201–293.

— (1926). *The Abilities of Man.* London: MacMillan.

Steiger, J. H. (1994). "Factor analysis in the 1980's and the 1990's: some old debates and some new developments". In: *Trends and perspective in* *empirical social research*. Ed. by I. Borg and P. P. Mohler. Berlin: Walter de Gruyter.

Thurstone, L. L. (1947). *Multiple–factor analysis: a development and expansion of "The vectors of the mind"*. Chicago: The University of Chicago Press.

Tsukahara, H. (2005). "Semiparametric estimation in copula models." In: *Can. J. Stat.* 33, pp. 357–375.

Wang, Q. J. (1996). "Direct sample estimators of L moments". In: *Water Resour. Res.* 32, pp. 3617–3619.

Wright, S. (1918). "On the nature of size factors". In: *Genetics* 3, pp. 367–374.

— (1934). "The method of path coefficients". In: *Ann. Math. Stat.* 5, pp. 161–215.

Yuan, K.-H. and P. M. Bentler (1999). "On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions". In: *Stat. Sinica* 9, pp. 831–853.

Yung, Y. F. and P. M. Bentler (1994). "Bootstrap–corrected ADF test statistics in covariance structure analysis". In: *Brit. J. Math. Stat. Psy.* 47, pp. 63–84.