Weighted least-squares inference based on dependence coefficients for multivariate copulas

Gildas Mazo, Stéphane Girard, Florence Forbes

Weighted least-squares inference based on dependence coefficients for multivariate copulas

Gildas Mazo, Stéphane Girard and Florence Forbes

MISTIS, Inria - Laboratoire Jean Kuntzmann, France

Abstract

In this paper, we address the issue of estimating the parameters of general multivariate copulas, that is, copulas whose partial derivatives may not exist. To this aim, we consider a weighted least-squares estimator based on dependence coefficients, and establish its consistency and asymptotic normality. The estimator's performance on finite samples is illustrated on simulations and a real dataset.

Keywords: partial derivatives, singular component, least-squares, method-of-moments, dependence coefficients, parametric inference, copulas, multivariate.

1 Introduction

The concept of copulas is useful to model multivariate distributions. Given a multivariate random vector of interest, copulas allow to separate the analysis of the margins from the dependence structure. Standard books covering this subject include [9, 24, 28]. See also [13] for an introduction to this topic.

Some copulas possess a singular component, meaning that they are not absolutely continuous (with respect to the Lebesgue measure). For instance, take the copula given below, introduced in [8]:

$$C(u_1,u_2,u_3,u_4)=\prod_{i=1}^{4}u_i^{1-\sum_{j\neq i}\theta_{ij}}\prod_{i<j}\min(u_i,u_j)^{\theta_{ij}}, \tag{1}$$

$$\sum_{j\neq i}\theta_{ij}\leq 1,\ i=1,\dots,4,\ \theta_{ij}=\theta_{ji}\in[0,1].$$

One can see that, on the diagonal of the unit hypercube, the partial derivatives do not exist. Yet, most inference methods for multivariate copulas make the assumption that these derivatives exist, and even be continuous. This is the case, for example, of the minimum-distance estimator [33], the simulated method of moments [30], and, of course, likelihood-based methods (Section 10.1 [24], [14]). When one does not make this assumption, some methods can still be applied but only in specific situations. For example, when there are only two dimensions, one can rely on the inversion of the Kendall's tau, see [18] and [13]. When they are an arbitrary number of dimensions but only one parameter to estimate, an extension of this method can also be found in [15]. Also, if the copulas of interest

are elliptical copulas, one can use the analysis of covariance structures [25]. This issue, that the partial derivatives need to exist and be continuous on the unit hypercube in order to properly apply most of the inference methods, was raised in [3, 32]. In these papers, the authors weaken the differentiability assumptions in empirical copula process theory, which is often used to establish asymptotic results of the methods. Nevertheless, they still need the partial derivatives to exist and be continuous on the interior of the unit hypercube. But, as shown in (1), these derivatives may not even exist on this space.

In order to estimate the parameters of general multivariate copulas, we consider a weighted least-squares (WLS) estimator based on dependence coefficients. The consistency and asymptotic normality of the estimator are derived without assuming that the copulas of interest have partial derivatives at all. This method is therefore broadly applicable and allows to estimate the parameters of any kind of copulas, provided that one can calculate their dependence coefficients.

In Section 2 of this paper, the consistency and asymptotic normality of the WLS estimator are established. The theoretical results are illustrated on simulated and real datasets in Section 3. The proofs are postponed to the Appendix.

2 Asymptotic properties of the WLS estimator based on dependence coefficients

In this section, we derive the consistency and asymptotic normality of a generic WLS estimator in Section 2.1 and give three examples based on the Spearman's rho, the Kendall's tau, and the extremal dependence coefficients in Section 2.2.

Let $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ with $\mathbf{X}^{(k)} = (X_1^{(k)}, \dots, X_d^{(k)})$, $k = 1, \dots, n$, be independent and identically distributed copies of a vector $\mathbf{X} = (X_1, \dots, X_d)$ with distribution F and copula C. The marginal distributions $F_1, \dots, F_d$ are assumed to be continuous. The copula C is assumed to belong to the family $(C_{\boldsymbol{\theta}})$ for $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$. The true parameter vector is denoted by $\boldsymbol{\theta}_0$, that is, $C = C_{\boldsymbol{\theta}_0}$. Let $p = d(d-1)/2$ be the number of variable pairs (X_i, X_j), for $i = 1, \dots, d-1$, $j = 2, \dots, d$, $i < j$. Let us define the vector map

$$\begin{aligned} \boldsymbol{\mathcal{D}} : \Theta &\to \boldsymbol{\mathcal{D}}(\Theta) \subset \mathbb{R}^p \\ \boldsymbol{\theta} &\mapsto (\mathcal{D}_{1,2}(\boldsymbol{\theta}), \dots, \mathcal{D}_{d-1,d}(\boldsymbol{\theta}))\,, \end{aligned} \tag{2}$$

where $\mathcal{D}_{i,j}(\cdot)$ can represent, but is not limited to, a well chosen dependence coefficient between the variables X_i and X_j (see Section 2.2 for examples). The space $\boldsymbol{\mathcal{D}}(\Theta)$ stands for the image of Θ by the multivariate map $\boldsymbol{\mathcal{D}}$. The coordinates of $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta})$ are the $\mathcal{D}_{i,j}(\boldsymbol{\theta})$ sorted in the lexicographical order. When the map $\boldsymbol{\mathcal{D}}$ is differentiable, its Jacobian matrix at $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ is denoted by

$$\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \mathcal{D}_{1,2}(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \mathcal{D}_{1,2}(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial \mathcal{D}_{1,2}(\boldsymbol{\theta})}{\partial \theta_q} \\ \vdots & & & \vdots \\ \frac{\partial \mathcal{D}_{d-1,d}(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \mathcal{D}_{d-1,d}(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial \mathcal{D}_{d-1,d}(\boldsymbol{\theta})}{\partial \theta_q} \end{pmatrix}.$$

Besides, let $\widehat{\boldsymbol{\mathcal{D}}} = (\widehat{\mathcal{D}}_{1,2}, \ldots, \widehat{\mathcal{D}}_{d-1,d})$ be an empirical (nonparametric) estimator of $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$. To simplify the notations, we shall write $\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}_0) = \dot{\boldsymbol{\mathcal{D}}}$, $\mathcal{D}_{i,j}(\boldsymbol{\theta}_0) = \mathcal{D}_{i,j}$ and $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$. Vectors are assumed to be column vectors and ^{T} denotes the transpose symbol.

The WLS estimator of $\boldsymbol{\theta}_0$ studied in this paper is defined as

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta})\right)^T \widehat{\mathbf{W}} \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta})\right), \tag{3}$$

where $\widehat{\mathbf{W}} = \widehat{\mathbf{W}}_n$ is a sequence ($n = 1, 2, \ldots$) of symmetric and positive definite matrices with full rank. Let us note $\hat{\ell}(\boldsymbol{\theta})$ the loss function to be minimized in (3). In general, the minimizer $\hat{\boldsymbol{\theta}}$ of $\hat{\ell}(\cdot)$ may not exist, or may not be unique. However, it will be seen in Section 2.1 that the existence and uniqueness of $\hat{\boldsymbol{\theta}}$ hold with probability tending to one as the sample size increases. Since $\widehat{\mathbf{W}}$ is positive definite, the loss function $\hat{\ell}$ is such that $\hat{\ell}(\boldsymbol{\theta}) \geq 0$ for all $\boldsymbol{\theta} \in \Theta$ and vanishes at $\hat{\boldsymbol{\theta}}$ if and only if $\hat{\boldsymbol{\theta}} \in \boldsymbol{\mathcal{D}}^{-1}(\{\widehat{\boldsymbol{\mathcal{D}}}\})$, where $\boldsymbol{\mathcal{D}}^{-1}(\{\widehat{\boldsymbol{\mathcal{D}}}\})$ denotes the set of all $\boldsymbol{\theta}$ in Θ such that $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{\mathcal{D}}}$. In this case, the WLS estimator does not depend on the weights and $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\mathcal{D}}}$. Moreover, if the multivariate map $\boldsymbol{\mathcal{D}}$ is one-to-one, then the WLS estimator takes the form $\hat{\boldsymbol{\theta}} = \boldsymbol{\mathcal{D}}^{-1}(\widehat{\boldsymbol{\mathcal{D}}})$.

2.1 Asymptotic properties of the generic WLS estimator

The assumptions needed to derive the asymptotic properties of the WLS estimator are given below. The symbol $\|\cdot\|$ denotes the Euclidean norm.

Assumptions. *(A1) The true parameter vector $\boldsymbol{\theta}_0$ lies in the interior of Θ. Moreover, there exists $\varepsilon_0 > 0$ such that the set $\{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \varepsilon_0\}$ is closed (and thus compact).*

(A2) As $n \to \infty$, the sequence of weight matrices $\widehat{\mathbf{W}}$ converges in probability to a symmetric and positive definite matrix $\mathbf{W}$ with full rank.

(A3) The map $\boldsymbol{\mathcal{D}}$ defined in (2) is a twice continuously differentiable homeomorphism such that $\dot{\boldsymbol{\mathcal{D}}}$ is of full rank.

(A4) As $n \to \infty$, the empirical estimator $\widehat{\boldsymbol{\mathcal{D}}}$ is such that

$$\widehat{\boldsymbol{\mathcal{D}}} \xrightarrow{P} \boldsymbol{\mathcal{D}}, \text{ and, } \sqrt{n}\left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}\right) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is some symmetric, positive definite $p \times p$ matrix noted as follows

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{1,2;1,2} & \Sigma_{1,2;1,3} & \cdots & \Sigma_{1,2;d-1,d} \\ \Sigma_{1,3;1,2} & \Sigma_{1,3;1,3} & \cdots & \Sigma_{1,3;d-1,d} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_{d-1,d;1,2} & \Sigma_{d-1,d;1,3} & \cdots & \Sigma_{d-1,d;d-1,d} \end{pmatrix}. \tag{4}$$

Assumption (A1), which is rather standard, see, e.g. [10], is not too restrictive for most copula models. Indeed, a parameter lying in the parameter space boundaries often means that the copula of interest is in fact the independence copula or one of the Fréchet-Hoeffding bounds, that is, a copula where the

dependence is "perfect", see for instance [28] chapter 2. This is not an issue because one does not encounter perfect dependence in practice. As for independence, one might carry out a statistical test as in [16], and, based on the results, decide wether independence holds or not. If not, then one can safely assume that the parameters lie in the interior of the parameter space. A sequence of weight matrices verifying Assumption (A2) can always be constructed. A trivial example is $\widehat{\mathbf{W}} = \mathbf{I}_p$, where $\mathbf{I}_p$ is the identity matrix of size p. The construction of optimal weights is addressed in Proposition 1 below. The estimation of the copula parameter vector is performed by matching the theoretical and empirical dependence coefficients. Hence, a successful match should ensure that the resulting parameter vector estimate is close to the true value. This identifyability condition, also made in [10] in order to estimate extreme-value copulas with a singular component, is the essence of Assumption (A3). The last assumption, (A4), naturally states that one should have convergence of the dependence coefficient empirical estimator to ensure convergence of the WLS estimator.

Theorem 1. *Assume that (A1)–(A4) hold. Then, as $n \to \infty$ and with probability tending to one, the WLS estimator defined by* (3) *exists and is unique. Moreover, it is consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \boldsymbol{\Xi}\right), \text{ where} \tag{5}$$
$$\boldsymbol{\Xi} = \left(\dot{\boldsymbol{\mathcal{D}}}^T \mathbf{W} \dot{\boldsymbol{\mathcal{D}}}\right)^{-1} \dot{\boldsymbol{\mathcal{D}}}^T \mathbf{W} \boldsymbol{\Sigma} \mathbf{W} \dot{\boldsymbol{\mathcal{D}}} \left(\dot{\boldsymbol{\mathcal{D}}}^T \mathbf{W} \dot{\boldsymbol{\mathcal{D}}}\right)^{-1}.$$

As usual, the results of Theorem 1 allow to derive the asymptotic distribution of quadratic forms in $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}})$. These asymptotics serve to build confidence regions and statistical tests for the parameters and the dependence coefficients. Let χ^2_q denote the Chi square distribution with q degrees of freedom. Let us write $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ to emphasize that in general these matrices depend on $\boldsymbol{\theta}$. The continuity of matrices with respect to the parameter vector $\boldsymbol{\theta}$ is meant elementwise. Corollary 1, given below, may serve to build confidence regions around $\hat{\boldsymbol{\theta}}$ or $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}})$.

Corollary 1. *Suppose that the assumptions of Theorem 1 hold.*

(i) If $\boldsymbol{\Xi}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta}$ in Θ and $\boldsymbol{\Sigma}(\cdot)$ is continuous at $\boldsymbol{\theta}_0$, then, as $n \to \infty$,

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2_q.$$

(ii) Define $\widehat{\boldsymbol{\Sigma}}$ such that $\widehat{\boldsymbol{\Sigma}}$ is invertible and converges to $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ in probability as $n \to \infty$. Then as $n \to \infty$,

$$n(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0))^T \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)) \xrightarrow{d} \chi^2_p.$$

For a particular value $\boldsymbol{\theta}_1^\star \in \mathbb{R}^r$, $r \leq q-1$, the test $H_0 : \boldsymbol{\theta}_{01} = \boldsymbol{\theta}_1^\star$ against $H_1 : \boldsymbol{\theta}_{01} \neq \boldsymbol{\theta}_1^\star$, where $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{01}, \boldsymbol{\theta}_{02}) \in \mathbb{R}^r \times \mathbb{R}^{q-r}$, may be carried out using the asymptotic approximation suggested by Corollary 2, given next. In general, write $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^r \times \mathbb{R}^{q-r}$ for $\boldsymbol{\theta} \in \Theta$, and, likewise, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$. Let $\boldsymbol{\Xi}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the asymptotic covariance $r \times r$ matrix corresponding to $\boldsymbol{\theta}_1$, that is, the upper left part of $\boldsymbol{\Xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

Corollary 2. *Under the assumptions of Corollary 1 (i), as* $n \to \infty$,

$$n(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^\star)^T \boldsymbol{\Xi}_1(\boldsymbol{\theta}_1^\star, \hat{\boldsymbol{\theta}}_2)^{-1}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^\star) \xrightarrow{d} \chi_r^2.$$

The test H_0: "the chosen parametric model is the true model of the underlying copula" against H_1 "the chosen parametric model is false" may be carried out by using the asymptotic approximation suggested by Corollary 3 below, adapted from [22].

Corollary 3. *Suppose that the assumptions of Theorem 1 and Corollary 1 (ii) hold. For* $\boldsymbol{\theta} \in \Theta$, *define*

$$\mathbf{A}(\boldsymbol{\theta}) := \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \right)^{-1} \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T,$$

$\widehat{\mathbf{A}} = \mathbf{A}(\hat{\boldsymbol{\theta}})$, *and note* k *the rank of* $\mathbf{I}_p - \mathbf{A}(\boldsymbol{\theta}_0)$. *If* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *is invertible for all* $\boldsymbol{\theta}$ *in* Θ *and* $\boldsymbol{\Sigma}(\cdot)$ *is continuous at* $\boldsymbol{\theta}_0$, *then as* $n \to \infty$,

$$n \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right)^T (\mathbf{I}_p - \widehat{\mathbf{A}})[(\mathbf{I}_p - \widehat{\mathbf{A}})\widehat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \widehat{\mathbf{A}}) + \widehat{\mathbf{A}}]^{-1}(\mathbf{I}_p - \widehat{\mathbf{A}}) \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) \to \chi_k^2.$$

The asymptotic covariance matrix $\boldsymbol{\Xi}$ in (5) depends on the weight matrix $\mathbf{W}$. The optimal weight matrix $\mathbf{W}^\star$, in the sense that it allows to minimize the asymptotic covariance matrix $\boldsymbol{\Xi}$, is given in Proposition 1 below (due to [22]). The above mentionned ordering of covariance matrices is to be understood in the following sense. The notation $\mathbf{A} \geq 0$ means that the matrix $\mathbf{A}$ is nonnegative definite. For two nonnegative definite matrices $\mathbf{A}$ and $\mathbf{B}$, define $\mathbf{A}$ to be less or equal than $\mathbf{B}$ if $\mathbf{B} - \mathbf{A} \geq 0$. It is easily checked that $\mathbf{A} \leq \mathbf{B}$ implies $\mathrm{tr}(\mathbf{A}) \leq \mathrm{tr}(\mathbf{B})$, where $\mathrm{tr}(\cdot)$ stands for the trace operator of matrices. Thus, the distribution with the smallest covariance matrix is the one for which the sum of the variances is minimum. In view of (6), an optimal estimator, that is, an estimator that leads to the smallest asymptotic covariance matrix, can be constructed by letting the sequence of weight matrices converge to $\boldsymbol{\Sigma}^{-1}$.

Proposition 1. *Suppose that* $\boldsymbol{\Sigma}$ *defined in (A4) is invertible. Then the asymptotic covariance matrix* $\boldsymbol{\Xi}$ *is minimum for* $\mathbf{W}^\star$ *such that*

$$\mathbf{W}^\star \dot{\boldsymbol{\mathcal{D}}} \propto \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\mathcal{D}}}, \tag{6}$$

where the symbol $\propto$ *denotes proportionality.*

An estimate of the optimal weight matrix $\boldsymbol{\Sigma}^{-1}$ can be based on empirical data or constructed as follows. Define the *zero-step estimator* $\hat{\boldsymbol{\theta}}^0$ to be the WLS estimator (3) with $\widehat{\boldsymbol{W}} = \mathbf{I}_p$. Define the *one-step estimator* $\hat{\boldsymbol{\theta}}^1$ to be the WLS estimator with $\widehat{\boldsymbol{W}} = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}^0)$, where $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}^0)$ is an estimate of $\boldsymbol{\Sigma}$ based on the zero-step estimator. For instance, one may simulate data according to $C = C(\hat{\boldsymbol{\theta}}^0)$ and use them to construct $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}^0)$. This one-step estimator is then an optimal estimator. The performances of the zero-step and the optimal estimators will be compared in Section 3.1.

When there are as many pairs as parameters, the WLS estimator does not depend on the weights, as stated in the next proposition.

Proposition 2. *Suppose that the assumptions of Theorem 1 hold. If $p = q$ then, as $n \to \infty$ and with probability tending to one,*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mathcal{D}}^{-1}\left(\widehat{\boldsymbol{\mathcal{D}}}\right), \tag{7}$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \left(\dot{\boldsymbol{\mathcal{D}}}^T\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\right). \tag{8}$$

2.2 Examples of three dependence coefficients verifying Assumption (A4)

Three examples of a dependence coefficient for which the pair of vectors $(\boldsymbol{\mathcal{D}}, \widehat{\boldsymbol{\mathcal{D}}})$ satisfies Assumption (A4) are provided. These coefficients are the Spearman's rho, the Kendall's tau, and the extremal dependence coefficient. They are widely used in practice, and that is why we illustrate our methodology on them. But others can be used, as long as (A4) holds. See [24,28] for more about these coefficients and [23] for their asymptotic properties. Recall that F_i is the distribution of X_i and let

$$\widehat{F}_i(x) = \frac{1}{n+1}\sum_{k=1}^{n}\mathbf{1}(X_i^{(k)} \leq x), \quad x \in \mathbb{R}.$$

Put $U_i = F_i(X_i)$ and $\widehat{U}_i^{(k)} = \widehat{F}_i(X_i^{(k)})$. Recall that $F_{i,j}$ is the distribution function of (X_i, X_j) and that $C_{i,j}$ denotes its copula.

Example 1 (Spearman's rho)**.** *The Spearman's rho dependence coefficient of the pair (X_i, X_j) is given by*

$$\mathcal{D}_{i,j} = 12\int_{[0,1]^2} C_{i,j}(u,v)\,du\,dv - 3. \tag{9}$$

Its empirical counterpart is defined as

$$\widehat{\mathcal{D}}_{i,j} = \frac{\sum_{k=1}^{n}\left(\widehat{U}_i^{(k)} - \overline{\widehat{U}}_i\right)\left(\widehat{U}_j^{(k)} - \overline{\widehat{U}}_j\right)}{\left[\sum_{k=1}^{n}\left(\widehat{U}_i^{(k)} - \overline{\widehat{U}}_i\right)^2\sum_{k=1}^{n}\left(\widehat{U}_j^{(k)} - \overline{\widehat{U}}_j\right)^2\right]^{1/2}},$$

where $\overline{\widehat{U}_i} = \sum_{k=1}^{n}\widehat{U}_i^{(k)}/n$. From [23] Theorem 7.1, Assumption (A4) holds with

$$\begin{aligned}\Sigma_{i,j;k,l} = 9\int_{[0,1]^2} &[3(4C_{i,j}(u_i,u_j) + 1 - 2u_i - 2u_j) - \mathcal{D}_{i,j}] \\ &\times [3(4C_{k,l}(u_k,u_l) + 1 - 2u_k - 2u_l) - \mathcal{D}_{k,l}]\,dC(u_1,\ldots,u_d).\end{aligned}$$

Example 2 (Kendall's tau)**.** *The Kendall's tau dependence coefficient of the pair (X_i, X_j) is given by*

$$\mathcal{D}_{i,j} = 4\int_{[0,1]^2} C_{i,j}(u,v)\,dC_{i,j}(u,v) - 1. \tag{10}$$

Its empirical counterpart is defined as

$$\widehat{\mathcal{D}}_{i,j} = \binom{n}{2}^{-1} \sum_{k<l} sign\left((X_i^{(k)} - X_i^{(l)})(X_j^{(k)} - X_j^{(l)})\right), \tag{11}$$

where $sign(x) = 1$ *if* $x > 0$, -1 *if* $x < 0$ *and* 0 *if* $x = 0$. *From [23] Theorem 7.1, Assumption (A4) holds with*

$$\begin{aligned}\Sigma_{i,j;k,l} = 4 \int_{[0,1]^2} & [4C_{i,j}(u_i, u_j) + 1 - \mathcal{D}_{i,j} - 2u_i - 2u_j] \\ & \times [4C_{k,l}(u_k, u_l) + 1 - \mathcal{D}_{k,l} - 2u_k - 2u_l]\, dC(u_1, \ldots, u_d).\end{aligned} \tag{12}$$

The third example deals with extreme-value copulas, which are theoretically well grounded for performing a statistical analysis of extreme values, such as maxima of samples. Recall that a copula $C_\#$ is an extreme-value copula if there exists a copula $\tilde{C}$ such that

$$C_\#(u_1, \ldots, u_d) = \lim_{n\uparrow\infty} \tilde{C}^n(u_1^{1/n}, \ldots, u_d^{1/n}), \quad (u_1, \ldots, u_d) \in [0,1]^d,$$

see, e.g. [20]. The class of extreme-value copulas corresponds exactly to the class of max-stable copulas, that is, the copulas $C_\#$ such that

$$C_\#^n(u_1^{1/n}, \ldots, u_d^{1/n}) = C_\#(u_1, \ldots, u_d), \quad n \geq 1,\ (u_1, \ldots, u_d) \in [0,1]^d.$$

The extremal dependence coefficient is implicitely defined by the following representation of bivariate extreme-value copulas on the diagonal of the unit square:

$$C_\#(u, u) = u^{2-\lambda}, \quad \lambda \in [0,1]. \tag{13}$$

If $\lambda = 0$ then $C_\#(u,u) = \Pi(u,u) = u^2$, where Π stands for the independence copula. If $\lambda = 1$ then $C_\#(u,u) = M(u,u) = \min(u,u) = u$, where M stands for the Fréchet-Hoeffding upper bound for copulas, that is, the case of perfect dependence. In the case of extreme-value copulas, the extremal dependence coefficient corresponds to the well known upper tail dependence coefficient

$$\lambda = \lim_{u\uparrow 1} \frac{1 - 2u + C_\#(u,u)}{1-u},$$

which measures the dependence in the tails. Nonetheless, for extreme-value copulas, the interpolation between Π and M on the diagonal of the unit square (13) makes the extremal dependence coefficient a natural coefficient of general dependence, and not just a coefficient that measures dependence in the tails. For further information about extreme-value statistics, see, e.g. [5]. An account about extreme-value copulas can be found in [20].

Estimators of the extremal dependence coefficient for which the asymptotic properties are derived under unknwon margins can be found in [2, 19]. However, in order to obtain the results, the existence of partial derivatives for the underlying copulas was assumed. Hence, these estimators cannot be used since we aim at estimating the parameters of copulas for which these derivatives may not exist.

If the marginal distributions are assumed to be known, however, various estimators of the extremal dependence coefficient and their asymptotic properties can be found in the literature [4,7,11,21,31]. A review can be found in [20]. Our choice of the estimator presented in Example 3 below, that of [11], is arbitrary. One can choose an other estimator in the literature and adapt the results.

Example 3 (Extremal dependence coefficient)**.** *Assume that the copula of interest C is an extreme-value copula and let $\mathcal{D}_{i,j}$ be the extremal dependence coefficient of the pair (X_i, X_j), implicitely defined in* (13)*, and given by*

$$\mathcal{D}_{i,j} = 2 + \log C_{i,j}(e^{-1}, e^{-1}). \tag{14}$$

Its empirical counterpart, as defined in [11], is given by

$$\widehat{\mathcal{D}}_{i,j} = 3 - \frac{1}{1 - \sum_{k=1}^{n} \max(U_i^{(k)}, U_j^{(k)})/n}.$$

By adapting [11] to the multivariate case, Assumption (A4) holds with

$$\Sigma_{i,j;k,l} = (3 - \mathcal{D}_{i,j})^2 (3 - \mathcal{D}_{k,l})^2 \; Cov\left(\max(U_i, U_j), \max(U_k, U_l)\right). \tag{15}$$

In practice, one usually does not know the margins. However, assuming that F is an extreme-value distribution, the margins should be Generalized Extreme-Value (GEV) distributions, see [5]. Hence, one can fit a GEV to the margins and act as if the marginal distributions were known, provided that this approximation has been carefully checked.

3 Illustrations on simulated and real datasets

In order to assess the WLS estimator's performance on finite samples, numerical experiments are undertaken in Section 3.1 and a real dataset application is presented in Section 3.2. In both the experiments and the application, we aim at estimating the parameters of multivariate copulas possessing a singular component.

3.1 Estimating the parameters of multivariate copulas possessing a singular component

By substituting the Fréchet copulas [12]

$$C_{0k}(u_0, u_k) = \theta_k \min(u_0, u_k) + (1 - \theta_k) u_0 u_k, \quad \theta_k \in [0, 1]$$

into the one-factor copula [27]

$$C(u_1, \ldots, u_d) = \int_0^1 \prod_{k=1}^{d} \frac{\partial C_{0k}(u_0, u_k)}{\partial u_0} \, du_0,$$

one obtains a copula C with a singular component and whose bivariate margins are given by the following Fréchet copulas

$$C_{ij}(u_i, u_j) = \theta_i \theta_j \min(u_i, u_j) + (1 - \theta_i \theta_j) u_i u_j, \quad \theta_i, \theta_j \in [0, 1]. \tag{16}$$

The Spearman's rho and Kendall's tau coefficients of (16) are respectively equal to $\theta_i\theta_j$ and $\theta_i\theta_j(\theta_i\theta_j+2)/3$. The extreme-value copula $C_\#$ associated to C can be derived by calculating the limit

$$C_\#(u_1,\dots,u_d)=\lim_{n\uparrow\infty} C^n(u_1^{1/n},\dots,u_d^{1/n}).$$

It appears that the bivariate margins of $C_\#$ are Cuadras-Augé copulas [6]

$$C_{\#,ij}(u_i,u_j)=\min(u_i,u_j)\max(u_i,u_j)^{1-\theta_i\theta_j},\quad \theta_i,\theta_j\in[0,1]$$

with extremal dependence coefficient given by $\theta_i\theta_j$. As C, $C_\#$ possess a singular component.

		$d=4$		$d=10$	
		zero-step	one-step	zero-step	one-step
$n=50$	(S1)	0.11	0.11	0.10	0.12
	(S2)	0.10	0.10	0.09	0.10
	(S3)	0.18	0.18	0.17	0.20
$n=200$	(S1)	0.06	0.06	0.05	0.05
	(S2)	0.05	0.05	0.04	0.04
	(S3)	0.10	0.10	0.09	0.09
$n=500$	(S1)	0.04	0.04	0.03	0.03
	(S2)	0.03	0.03	0.03	0.03
	(S3)	0.06	0.06	0.06	0.05

Table 1: Averaged MAEs for the three studied situations with respect to the dataset sample size n and dimension d.

The two copulas C and $C_\#$ are considered in the following numerical experiment. For each combination (d,n) with $d=4,10$ and $n=50,200,500$, we generated 200 datasets according to these copulas. The true parameter vector coordinates θ_{0k}, $k=1,\dots,d$, were chosen to be regularly spaced between 0.3 and 0.9. Three situations were studied:

(S1) the parameters of C are estimated with the Spearman's rho (see Example 1),

(S2) the parameters of C are estimated with the Kendall's tau (see Example 2), and

(S3) the parameters of $C_\#$ are estimated with the extremal dependence coefficient (see Example 3).

For each situation (Si) above, the zero-step and one-step WLS estimators of Section 2.1 were tested (recall that the one-step estimator is optimal, see Proposition 1). For each dataset and each situation (Si), the mean absolute error, defined as

$$\mathrm{MAE}=\frac{1}{d}\sum_{k=1}^{d}|\hat\theta_k-\theta_{0k}|$$

was computed and averaged over the replications. These criteria are reported in Table 1. From this Table, we see that there is almost no difference between the zero-step and one-step estimators. This lack of weighting effect was also mentionned in [30] Section 3. This suggests that the zero-step estimator is already near optimal. The comparison of the rows (S1) and (S2) shows that the choice between the Spearman's rho and Kendall's tau in the WLS estimator has very little impact on its performance. Estimating the parameters of an extreme-value copula with the extremal dependence coefficient, however, appears to be less accurate–see the (S3) row of the table. Finally, the comparison of the two columns $d = 4$ and $d = 50$ shows that the dimension of the inference problem does not seem to affect the estimator's performance. This property makes it very attractive to deal with high-dimensional applications. To complete the study of the estimator's abilities, its asymptotic distribution derived in Theorem 1 is tested. Since this distribution is multivariate, we checked the Chi-square approximation of Corollary 1 instead. The values $n(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi} (\hat{\boldsymbol{\theta}}^{(k)})^{-1} (\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)$, $k = 1, \ldots, 200$, should be approximately χ^2_d distributed, where $\hat{\boldsymbol{\theta}}^{(k)}$ denotes the parameter vector estimated on the k-th dataset replication. This approximation, shown in Figure 1, seems rather satisfactory.

3.2 Measuring uncertainty for multivariate return periods in hydrology

In hydrology, the severity and frequency of extreme events must be quantified. Such potentially dangerous events are underlain by the behavior of a random vector $(X_1, \ldots, X_d)$ distributed according to a certain distribution F with continuous margins $F_1, \ldots, F_d$ and copula C. Suppose that C is determined by a parameter vector $\boldsymbol{\theta}$ in Θ. For a certain potentially dangerous event, define the *return period* T and the *critical level* p through the relationship

$$T = \frac{1}{1 - K_{\boldsymbol{\theta}}(p)}, \tag{17}$$

where $K_{\boldsymbol{\theta}}(t) = P(C(F_1(X_1), \ldots, F_d(X_d)) \leq t)$, $t \in [0, 1]$, is called the Kendall's distribution function associated to C, see [29]. The return period can be interpreted as the average time elapsing between two dangerous events. For instance, $T = 30$ years means that the event happens once every 30 years in average. The critical level can be viewed as a measure of how dangerous the underlying event is. The following question naturally arises: given a certain return period, what is the critical level of the underlying event? To answer this question, it suffices to invert (17) to get p as a function of T:

$$p_T(\boldsymbol{\theta}) = K_{\boldsymbol{\theta}}^{-1}(1 - 1/T).$$

Let $\boldsymbol{\theta}_0$ denote the true parameter vector and let $p_T = p_T(\boldsymbol{\theta}_0)$. The estimation of p_T, or, in other words, the estimation of $\boldsymbol{\theta}_0$, was performed in [8] for all the pairs of $d = 3$ sites in Italy (Airole, Merelli and Poggi). The parametric model proposed for C was the extreme-value copula

$$C(u_1, \ldots, u_d) = \left(\prod_{i=1}^{d} u_i^{1-\theta_i} \right) \min_{i=1,\ldots,d} (u_i^{\theta_i}), \quad \theta_i \in [0, 1],\ i = 1, \ldots, d. \tag{18}$$

Figure 1: Histograms of $n(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi} (\hat{\boldsymbol{\theta}}^{(k)})^{-1} (\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)$, $k = 1, \ldots, 200$ together with the density of a χ^2_d distribution. The considered experiment parameters were $n = 500$ and $d = 4$. Upper left: (S2). Upper right: (S1). Bottom: (S3).

As it is seen from (18), this copula has a singular component. The authors chose to base the inference on the Kendall's tau (see Example 2). For $\boldsymbol{\theta}$ in $[0,1]^d$, the Kendall's tau coefficients are given by

$$\tau_{i,j}(\boldsymbol{\theta}) = \frac{\theta_i \theta_j}{\theta_i + \theta_j - \theta_i \theta_j}, \quad i < j. \tag{19}$$

By inverting (19), one obtains

$$\hat{\theta}_i = \frac{1}{2} \left(1 + \frac{1}{\hat{\tau}_{i,j}} + \frac{1}{\hat{\tau}_{i,k}} - \frac{1}{\hat{\tau}_{j,k}} \right), \tag{20}$$

where i, j, k denote the indexes of the three sites and $\hat{\tau}_{i,j}$ is given by (11). Observe that this is the solution of the equation (7), and, under the light of Proposition 2 (since $p = q = d = 3$), we see that this estimator has the smallest asymptotic variance within the class (3). However, in [8], the asymptotic behavior of $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ was not derived. This is done next, and we shall see that it allows to quantify the uncertainties around the critical levels.

The asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is established by applying Theorem 1. It suffices to verify that assumption (A3) holds, which is easily checked from (19). Hence, as $n \to \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Xi}), \tag{21}$$

where $\boldsymbol{\Xi}$ is given by (8) and (12). Now, the derivation of the asymptotic behavior of the critical levels is straightforward. From (21), we get by the delta-method that, as $n \to \infty$

$$\sqrt{n}\left(p_T(\hat{\boldsymbol{\theta}}) - p_T\right) \xrightarrow{d} N(0, s_T^2), \tag{22}$$

with $s_T^2 = \dot{\boldsymbol{p}}_T \boldsymbol{\Xi} \dot{\boldsymbol{p}}_T^T$, and where $\dot{\boldsymbol{p}}_T$ is the Jacobian of $p_T(\cdot)$ at the true parameter value. It follows that confidence intervals can be computed from the finite-sample approximation of (22), provided that the sample size is large enough. In [8], the critical levels in terms of return periods were reported for the three pair of sites (Airole-Merelli, Airole-Poggi and Merelli-Poggi). We added to their figure 95% confidence intervals for the critical levels (Figure 2).

The test based on Corollary 3 has no power to detect a wrong model in this situation. Indeed, since $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\mathcal{D}}}$, the test statistic is always zero. Other tests can be performed to achieve such a task, see the original paper [8].

When studying extreme events, it is common to have only a limited amount of data. For instance, in [8], only $n = 34$ (multivariate) observations were available. With such a small sample size, the approximation of the distribution of $\sqrt{n}(p_T(\hat{\boldsymbol{\theta}}) - p_T)$ to a normal distribution may be questionable. To assess the goodness of this approximation for small and moderate sample sizes, we carried out the following numerical experiment. $N = 500$ dataset of size n were generated according to (18) with $\boldsymbol{\theta}_0 = (0.6, 0.7, 0.2)$. For the m-th dataset ($m = 1, \ldots, N$), the parameter vector estimate $\hat{\boldsymbol{\theta}}^{(m)}$ was computed. Let $s_T(\hat{\boldsymbol{\theta}}^{(m)})$ be the asymptotic standard deviation in (22) at $\hat{\boldsymbol{\theta}}^{(m)}$ where $s_T(\boldsymbol{\theta})$ is regarded as a function of $\boldsymbol{\theta}$. The critical levels $p_T(\hat{\boldsymbol{\theta}}^{(m)})$ together with the 95% confidence bands $p_T(\hat{\boldsymbol{\theta}}^{(m)}) \pm 1.96 s_T(\hat{\boldsymbol{\theta}}^{(m)})/\sqrt{n}$ were computed for $T = 10, 20, 30$. Some of the $\hat{\boldsymbol{\theta}}^{(m)}$ did not lie in their theoretical bounds $[0, 1]$, which led to numerical difficulties for computing $s_T(\hat{\boldsymbol{\theta}}^{(m)})$. Therefore, these were dropped from the experiment. The results reported Table 2 show that the finite sample approximation is rather good for $n = 100$. Even for $n = 34$, this approximation appears to be good for the pair Airole-Merelli. Despite these encouraging results for moderate and small samples, we finish by stressing that the number of missing outputs (recall that this happens when $\hat{\boldsymbol{\theta}}^{(m)}$ do not belong to $[0, 1]$) were quite high: 354 and 298 over the 500 dataset replications for $n = 34$ and $n = 100$ respectively. Consequently, it would be of interest to improve the estimator (20) to reduce this vexing effect.

One can observe from Figure 2 that the curves for the pairs (Airole,Poggi) and (Merelli,Poggi) are similar comparing to that of the pair (Airole,Merelli). Hence to illustrate the use of Corollary 2, we performed the test $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. The change of parameters $\mu_1 := \theta_1 - \theta_2$, $\mu_2 := \theta_1 + \theta_2$ and $\mu_3 := \theta_3$ was applied to the copula model (18). By Corollary 2, the test statistics

$n\hat{\mu}_1^2/\Xi_1(0, \hat{\mu}_2, \hat{\mu}_3)$ converges in distribution to a χ_1^2 variable. We obtained a p-value of 95%, indicating that there is no statistical arguments against the null hypothesis. This high p-value also suggests that this test has little power for $n = 34$ data. The p-value for testing $\theta_2 = \theta_3$ and $\theta_1 = \theta_3$ were 83% and 84% respectively. The search of powerful tests for copulas is still an active area of research [1, 17, 26].

	pair	(Airole,Merelli)			(Airole,Poggi)			(Merelli,Poggi)		
n	T	10	20	30	10	20	30	10	20	30
34		0.95	0.95	0.93	0.89	0.84	0.82	0.90	0.87	0.82
100		0.95	0.94	0.94	0.96	0.94	0.93	0.96	0.94	0.93

Table 2: Proportion of inclusions within the 95% confidence intervals for the true value p_T.

4 Discussion

In this paper, we considered a weighted least-squares (WLS) estimator in order to estimate the parameters of general multivariate copulas, that is, copulas for which the partial derivatives may not exist. We established its asymptotic properties and studied its performance on finite samples. In particular, the numerical experiments revealed that the weights may have little impact on the accuracy. Moreover, and this is interesting for practical purposes, the accuracy of the WLS estimator does not seem to depend on the dimension of the statistical problems being addressed. In our work, we provided three dependence coefficients which can be used to form the WLS estimator: the Spearman's rho, the Kendall's tau, and the extremal dependence coefficient. We chose popular dependence coefficients, but others can be used. Even combinations of them may be considered, as long as the formed vector $\widehat{\boldsymbol{\mathcal{D}}}$ verifies Assumption (A4). In the hydrological application of Section 3.2, this may help to make the system of equations (20) more robust numerically.

Acknowledgment. The authors thank Fabrizio Durante and Gianfausto Salvadori for sharing the dataset used in Section 3.2.

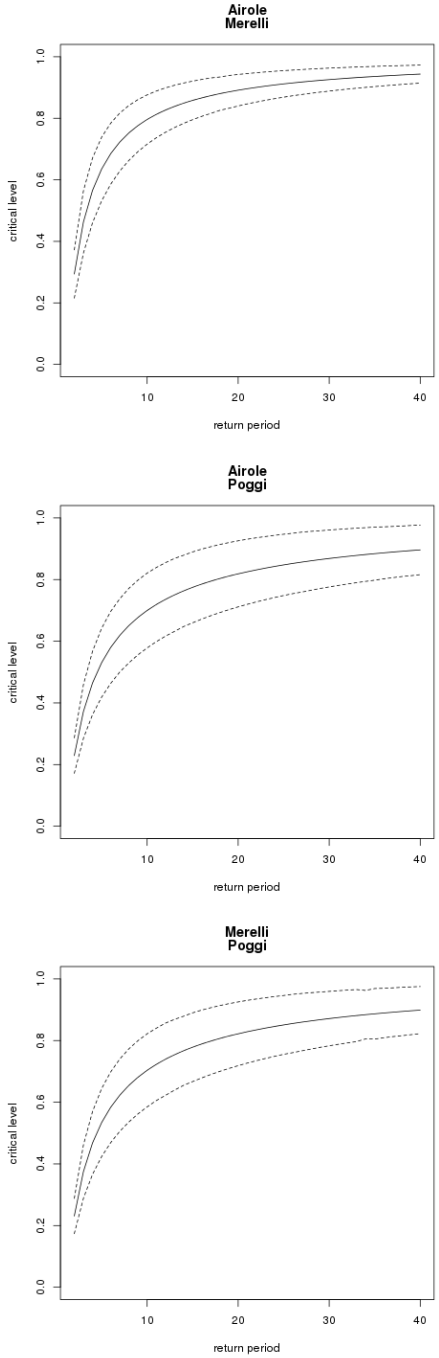

Figure 2: Critical levels $p_T(\hat{\boldsymbol{\theta}})$ for $T = 2, \ldots, 40$ together with 95% confidence intervals.

Appendix: proofs

In order to prove Theorem 1, we first establish two lemmas. These lemmas, as well as their proofs, are adapted from [10]. It will appear that the proof of the theorem is a straightforward application of these lemmas.

Let Θ and ε_0 as in assumption (A1). Define the vector map

$$\begin{aligned} \boldsymbol{\varphi} : \Theta \subset \mathbb{R}^q &\to \boldsymbol{\varphi}(\Theta) \subset \mathbb{R}^p \\ \boldsymbol{\varphi}(\boldsymbol{\theta}) &\mapsto (\varphi_1(\boldsymbol{\theta}), \ldots, \varphi_p(\boldsymbol{\theta}))^T, \end{aligned} \tag{23}$$

and assume that $\boldsymbol{\varphi}$ is twice continuously differentiable. Denote by $\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta})$ the $p \times q$ Jacobian matrix of $\boldsymbol{\varphi}$ at $\boldsymbol{\theta}$ and define $\dot{\boldsymbol{\varphi}} := \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)$. Let

$$\mathbf{Y}_n = (Y_{n,1}, \ldots, Y_{n,p})^T$$

be a random vector in $\mathbb{R}^p$ depending on an integer n and assume that $\mathbf{Y}_n \xrightarrow{P} \boldsymbol{\varphi}(\boldsymbol{\theta}_0)$ as $n \to \infty$. Let $\widehat{\mathbf{W}} = \widehat{\mathbf{W}}_n$ be a $p \times p$ symmetric and positive definite matrix with full rank and suppose that $\widehat{\mathbf{W}}$ converges in probability to a symmetric and positive definite matrix $\mathbf{W}$ with full rank as $n \to \infty$. Then the Cholesky decomposition entails that $\widehat{\mathbf{W}} = \widehat{\mathbf{V}}^T \widehat{\mathbf{V}}$ for some $p \times p$ matrix $\widehat{\mathbf{V}}$. Denote by $\hat{\Theta}_n$ the set of all minimizers of the loss function

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= (\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}))^T \widehat{\mathbf{W}} (\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta})) \\ &= \left\| \widehat{\mathbf{V}} (\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta})) \right\|^2, \quad \boldsymbol{\theta} \in \Theta, \end{aligned} \tag{24}$$

where $\|\cdot\|$ stands for the Euclidean norm. Observe that this set may contain several or no elements. Let $\mathbf{H}_n(\boldsymbol{\theta})$ be the Hessian matrix of ℓ_n at $\boldsymbol{\theta}$, that is, the matrix whose (k,l) element is given by

$$H_{n,kl}(\boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l}.$$

Let $\mathbf{Q}(\boldsymbol{\theta})$ be the $d \times d$ matrix whose (k,l) element writes

$$Q_{kl}(\boldsymbol{\theta}) = \left(\frac{\partial^2 \varphi_1(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l}, \ldots, \frac{\partial^2 \varphi_p(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l} \right) \widehat{\mathbf{W}}^T (\boldsymbol{\varphi}(\boldsymbol{\theta}) - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)),$$

and $\mathbf{H}(\boldsymbol{\theta})$ be the $d \times d$ matrix defined by

$$\mathbf{H}(\boldsymbol{\theta}) = 2 \left(\mathbf{Q}(\boldsymbol{\theta}) + \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta})^T \mathbf{W}^T \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}) \right).$$

Finally write $\overline{B}_\varepsilon(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \Theta : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq \varepsilon\}$ the closed ball around $\boldsymbol{\theta}_0$ with radius $\varepsilon > 0$ and assume that there exists $\varepsilon_0 > 0$ such that $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$ is closed. Then $\overline{B}_\varepsilon(\boldsymbol{\theta}_0)$ is compact for all $0 < \varepsilon \leq \varepsilon_0$.

Lemma 1. *(i) The elementwise convergence $\mathbf{H}_n(\boldsymbol{\theta}) \xrightarrow{P} \mathbf{H}(\boldsymbol{\theta})$ holds uniformly for all $\boldsymbol{\theta}$ in $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$.*

(ii) If $\dot{\boldsymbol{\varphi}}$ is of full rank then, with probability tending to 1, $\mathbf{H}_n(\boldsymbol{\theta})$ is positive definite for all $\boldsymbol{\theta}$ in some closed neighborhood of $\boldsymbol{\theta}_0$.

Proof. (i) It is easily seen that $\mathbf{H}_n(\boldsymbol{\theta}) = 2\left(\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta})^T\widehat{\mathbf{W}}^T\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}) + \mathbf{Q}_n(\boldsymbol{\theta})\right)$ where $\mathbf{Q}_n(\boldsymbol{\theta})$ is a $[d \times d]$ matrix such that its (k,l) element is given by

$$Q_{n,kl}(\boldsymbol{\theta}) = \left(\frac{\partial^2\varphi_1(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}, \ldots, \frac{\partial^2\varphi_p(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}\right)\widehat{\mathbf{W}}^T\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \mathbf{Y}_n\right).$$

Let $\widehat{W}_{ji}$ denote the element of $\widehat{\mathbf{W}}$ in the j-th row and i-th column. For all $\boldsymbol{\theta}$ in $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$,

$$\begin{aligned}
|H_{n,kl}(\boldsymbol{\theta}) - H_{kl}(\boldsymbol{\theta})| &= 2\left|\sum_{i,j=1}^{p}\frac{\partial^2\varphi_i(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}\widehat{W}_{ji}\left(\varphi_j(\boldsymbol{\theta}_0) - Y_{n,j}\right)\right| \\
&\leq \sum_{i,j=1}^{p}\left|\frac{\partial^2\varphi_i(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}\right|\left|\widehat{W}_{ji}\right|\left|\varphi_j(\boldsymbol{\theta}_0) - Y_{n,j}\right| \\
&\leq \text{constant} \times \sum_{i,j=1}^{p}\left|\widehat{W}_{ji}\right|\left|\varphi_j(\boldsymbol{\theta}_0) - Y_{n,j}\right|,
\end{aligned}$$

the last inequality holding because, since the second order derivatives of the φ_i's are continuous on the closed and thus compact set $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$, they are uniformly bounded by some constant on this set. Therefore, as $n \to \infty$,

$$\sup_{\boldsymbol{\theta}\in\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)}|H_{n,kl}(\boldsymbol{\theta}) - H_{kl}(\boldsymbol{\theta})| \leq \text{constant} \times \sum_{i,j=1}^{p}\left|\widehat{W}_{ji}\right|\left|\varphi_j(\boldsymbol{\theta}_0) - Y_{n,j}\right| \xrightarrow{P} 0,$$

which follows from the weak consistency of $\mathbf{Y}_n$ and $\widehat{\mathbf{W}}$.

(ii) Notice that since $\dot{\boldsymbol{\varphi}}$ is of full rank, $\mathbf{H}(\boldsymbol{\theta}_0)$ is positive definite. Hence for every $\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^q$, the map $\boldsymbol{\theta} \mapsto \mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}$ is continuous and one can choose a sufficiently small $\varepsilon(\mathbf{x}) > 0$ such that there exists $\delta(\mathbf{x}) > 0$ for which $\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x} \geq \mathbf{x}^T\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{x} - \varepsilon(\mathbf{x}) > 0$. In other words, $\forall\mathbf{x} \in \mathbb{R}^q$, $\exists\delta(\mathbf{x}) > 0$: $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta(\mathbf{x}) \implies \mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x} > 0$. Define $0 \leq \delta := \inf_{\mathbf{x}\in\mathbb{R}^q}\{\delta(\mathbf{x})\}$. Then for all $\boldsymbol{\theta}$ in Θ, $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$ implies $\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. We have shown that $\mathbf{H}(\boldsymbol{\theta})$ is positive definite on $\overline{B}_\delta(\boldsymbol{\theta}_0)$. Now define

$$A_{ij} = \left\{\sup_{\boldsymbol{\theta}\in\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)}|H_{n,ij}(\boldsymbol{\theta}) - H_{ij}(\boldsymbol{\theta})| \leq \inf_{x\in\mathbb{R}^q, x\neq 0,\, \boldsymbol{\theta}\in\overline{B}_\delta(\boldsymbol{\theta}_0)}\frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2\sum_{i,j=1}^{q}|x_ix_j|}\right\}$$

and put $A = \bigcap_{i,j} A_{ij}$. On the event A, for all $\mathbf{x} \neq \mathbf{0}$ and for all $\boldsymbol{\theta}$ in $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$, we have

$$\begin{aligned}
\left|\mathbf{x}^T\left(\mathbf{H}(\boldsymbol{\theta}) - \mathbf{H}_n(\boldsymbol{\theta})\right)\mathbf{x}\right| &\leq \sum_{i,j=1}^{q}|x_ix_j| \inf_{x\in\mathbb{R}^q, x\neq 0,\, \boldsymbol{\theta}\in\overline{B}_\delta(\boldsymbol{\theta}_0)}\frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2\sum_{i,j=1}^{q}|x_ix_j|} \\
&\leq \inf_{\boldsymbol{\theta}\in\overline{B}_\delta(\boldsymbol{\theta}_0)}\frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2}.
\end{aligned}$$

If, moreover, $\boldsymbol{\theta} \in \overline{B}_\delta(\boldsymbol{\theta}_0)$, then

$$\mathbf{x}^T\mathbf{H}_n(\boldsymbol{\theta})\mathbf{x} \geq \frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2} > 0$$

because $\mathbf{H}(\boldsymbol{\theta})$ is positive definite on $\overline{B}_\delta(\boldsymbol{\theta}_0)$. Hence on A and for all $\boldsymbol{\theta}$ in $\overline{B}_\delta(\boldsymbol{\theta}_0) \bigcap \overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$, the matrix $\mathbf{H}_n(\boldsymbol{\theta})$ is positive definite. By (i), $P(A) \to 1$ as $n \to \infty$, which concludes the proof.

Lemma 2. *(i) If $\boldsymbol{\varphi}$ in (23) is an homeomorphism, then for all ε such that $0 < \varepsilon \leq \varepsilon_0$, as $n \to \infty$,*

$$P\left[\hat{\Theta}_n \neq \emptyset \text{ and } \hat{\Theta}_n \subset \overline{B}_\varepsilon(\boldsymbol{\theta}_0)\right] \to 1.$$

(ii) If, moreover, $\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)$ is of full rank then as $n \to \infty$,

$$P\left[\text{ card } \hat{\Theta} = 1\right] \to 1,$$

where card denotes the cardinal of a set. Define $\hat{\boldsymbol{\theta}}$ to be the unique element of $\widehat{\Theta}$ if card $\hat{\Theta} = 1$, and any arbitrary point otherwise. Then $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ as $n \to \infty$.

(iii) If in addition to the assumptions of (i) and (ii)

$$\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \left(\dot{\boldsymbol{\varphi}}^T \mathbf{W} \dot{\boldsymbol{\varphi}}\right)^{-1} \dot{\boldsymbol{\varphi}}^T \mathbf{W} \boldsymbol{\Sigma} \mathbf{W} \dot{\boldsymbol{\varphi}} \left(\dot{\boldsymbol{\varphi}}^T \mathbf{W} \dot{\boldsymbol{\varphi}}\right)^{-1}\right)$$

Proof. (i) Let $0 < \varepsilon < \varepsilon_0$. Since $\boldsymbol{\varphi}$ is a homeomorphism and $\widehat{\mathbf{W}}$ has full rank, $\widehat{\mathbf{V}}\boldsymbol{\varphi}$ is also homeomorphism. Hence there exists $\delta > 0$ such that $\boldsymbol{\theta} \in \Theta$ and $\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\| \leq \delta$ imply $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \varepsilon$. Thus for every $\boldsymbol{\theta} \in \Theta$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \varepsilon$ we have $\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\| > \delta$. On the event $A_n = \{\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}_0) - \mathbf{Y}_n\right)\| \leq \delta/2\}$ and for $\boldsymbol{\theta}$ outside $\boldsymbol{\theta} \in \overline{B}_\varepsilon(\boldsymbol{\theta}_0)$, the inequality

$$\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\| \leq \|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \mathbf{Y}_n\right)\| + \|\widehat{\mathbf{V}}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|$$

implies

$$\begin{aligned}
\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \mathbf{Y}_n\right)\| &\geq \|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\| - \|\widehat{\mathbf{V}}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\| \\
&> \delta - \delta/2 \\
&= \delta/2 \\
&\geq \|\widehat{\mathbf{V}}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|.
\end{aligned}$$

Therefore

$$\min_{\boldsymbol{\theta} \in \overline{B}_\varepsilon(\boldsymbol{\theta}_0)} \|\widehat{\mathbf{V}}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta})\right)\| \leq \inf_{\boldsymbol{\theta} \notin \overline{B}_\varepsilon(\boldsymbol{\theta}_0)} \|\widehat{\mathbf{V}}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta})\right)\|,$$

where in the left hand side the minimum is attained because $\overline{B}_\varepsilon(\boldsymbol{\theta}_0)$ is compact. By consistency of $\mathbf{Y}_n$ and $\widehat{\mathbf{W}}$, we have $P(A_n) \to 1$. It follows that the event $\left\{\hat{\Theta}_n \neq \emptyset \text{ and } \hat{\Theta}_n \subset \overline{B}_\varepsilon(\boldsymbol{\theta}_0)\right\}$ has probability tending to 1.

(ii) Without loss of generality denote by $\overline{B}_\eta(\boldsymbol{\theta}_0)$, $\eta < \varepsilon_0$, the closed neighborhood of Lemma 1 (ii). Assume that the event

$$\left\{\hat{\Theta} \neq \emptyset,\ \hat{\Theta} \subset \overline{B}_\eta(\boldsymbol{\theta}_0) \text{ and } \mathbf{H}_n(\boldsymbol{\theta}) \text{ is positive definite for all } \boldsymbol{\theta} \text{ in } \overline{B}_\eta(\boldsymbol{\theta}_0)\right\} \quad (25)$$

happens. Let $\boldsymbol{\theta} \in \overline{B}_\eta(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}^\star$ be a vector in $\hat{\Theta}$. A Taylor expansion of ℓ_n in (24) at $\boldsymbol{\theta}^\star$ gives

$$\ell_n(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}^\star) + (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \nabla \ell_n(\boldsymbol{\theta}^\star) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\theta}^\star),$$

where $\tilde{\boldsymbol{\theta}} = t\boldsymbol{\theta} + (1-t)\boldsymbol{\theta}^\star$, $t \in (0,1)$ and $\nabla \ell_n$ denotes the gradient of ℓ_n. In view of Lemma 2 (i), $\boldsymbol{\theta}^\star$ is in some open neighborhood of $\boldsymbol{\theta}_0$ and thus $\nabla \ell_n(\boldsymbol{\theta}^\star) = 0$. The fact that $\tilde{\boldsymbol{\theta}} \in \overline{B}_\eta(\boldsymbol{\theta}_0)$ entails that $\mathbf{H}_n(\tilde{\boldsymbol{\theta}})$ is positive definite. Therefore, we have shown that $\ell_n(\boldsymbol{\theta}) > \ell_n(\boldsymbol{\theta}^\star)$ for all $\boldsymbol{\theta}$ in $\overline{B}_\eta(\boldsymbol{\theta}_0)$. This implies that the cardinal of $\hat{\Theta}$ is 1 when (25) holds. By Lemma 1 (ii) and Lemma 2 (i), the event (25) has probability tending to 1, hence, $P[\text{ card } \hat{\Theta} = 1] \to 1$. Now let $\hat{\boldsymbol{\theta}}$ be as in Lemma 2 (ii) and let $\varepsilon > 0$. Without loss of generality, assume that $\varepsilon \leq \varepsilon_0$. Then

$$\lim_{n\to\infty} P\left[\hat{\boldsymbol{\theta}} \in \overline{B}_\varepsilon(\boldsymbol{\theta}_0)\right] = \lim_{n\to\infty} P\left[\hat{\boldsymbol{\theta}} \in \overline{B}_\varepsilon(\boldsymbol{\theta}_0) \text{ and } \text{ card } \hat{\Theta} = 1\right] = 1,$$

the last equality holding because of Lemma 2 (i). Thus the consistency of $\hat{\boldsymbol{\theta}}$ is proved.

(iii) A Taylor expansion for the gradient $\nabla \ell_n$ of ℓ_n in equation (24) around $\boldsymbol{\theta}_0$ entails

$$\nabla \ell_n(\hat{\boldsymbol{\theta}}) = \nabla \ell_n(\boldsymbol{\theta}_0) + \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\tilde{\boldsymbol{\theta}} = t\hat{\boldsymbol{\theta}} + (1-t)\boldsymbol{\theta}_0$, $t \in (0,1)$. By the same arguments as in the proof of Lemma 2 (ii), $\nabla \ell_n(\hat{\boldsymbol{\theta}}) = 0$, hence,

$$\begin{aligned}
\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \sqrt{n}\left(\nabla \ell_n(\hat{\boldsymbol{\theta}}) - \nabla \ell_n(\boldsymbol{\theta}_0)\right) \\
&= -\sqrt{n}\nabla \ell_n(\boldsymbol{\theta}_0) \\
&= 2\dot{\boldsymbol{\varphi}}^T \widehat{\mathbf{W}} \sqrt{n}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right).
\end{aligned}$$

For $\mathbf{x}$ in $\mathbb{R}^q$, we have

$$\begin{aligned}
P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x}\right] =& P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{ card } \hat{\Theta} = 1\right] \\
&+ P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{ card } \hat{\Theta} \neq 1\right].
\end{aligned} \quad (26)$$

Since the second term in the sum in the right hand side of (26) tends to 0, we have that

$$\begin{aligned}
&\lim_{n\to\infty} P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{ card } \hat{\Theta} = 1\right] \\
&= \lim_{n\to\infty} P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x}\right] \\
&= \lim_{n\to\infty} P\left[2\dot{\boldsymbol{\varphi}}^T \widehat{\mathbf{W}} \sqrt{n}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right) \leq \mathbf{x}\right].
\end{aligned}$$

By the assumptions of Lemma 2 (iii) and by consistency of $\widehat{\mathbf{W}}$, we have

$$2\dot{\boldsymbol{\varphi}}^T\widehat{\mathbf{W}}\sqrt{n}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N_q\left(0, 4\dot{\boldsymbol{\varphi}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T\dot{\boldsymbol{\varphi}}\right).$$

If $\mathbf{H}_n(\tilde{\boldsymbol{\theta}})$ converges in probability to $\mathbf{H}(\boldsymbol{\theta}_0) = 2\dot{\boldsymbol{\varphi}}\mathbf{W}\dot{\boldsymbol{\varphi}}$, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \left(\dot{\boldsymbol{\varphi}}^T\mathbf{W}^T\dot{\boldsymbol{\varphi}}\right)^{-1}\dot{\boldsymbol{\varphi}}^T\mathbf{W}\Sigma\mathbf{W}^T\dot{\boldsymbol{\varphi}}\left[\left(\dot{\boldsymbol{\varphi}}^T\mathbf{W}^T\dot{\boldsymbol{\varphi}}\right)^{-1}\right]^T\right).$$

Therefore, to conclude the proof, it suffices to prove that $\mathbf{H}_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{P} \mathbf{H}(\boldsymbol{\theta}_0)$.

Let $\varepsilon > 0$. Assume that

$$\sup_{\boldsymbol{\theta}\in\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)} |H_{n,ij}(\boldsymbol{\theta}) - H_{ij}(\boldsymbol{\theta})| < \frac{\varepsilon}{2}.$$

The map $\boldsymbol{\theta} \mapsto H_{n,ij}(\boldsymbol{\theta})$ is continuous, hence, there exists $\delta > 0$ such that $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \delta$ implies $|H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{n,ij}(\boldsymbol{\theta}_0)| < \varepsilon/2$. Assume that $\hat{\boldsymbol{\theta}} \in \overline{B}_\delta(\boldsymbol{\theta}_0)$ and suppose without loss of generality that $\delta \leq \varepsilon_0$. Then it holds that

$$\begin{aligned}|H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{ij}(\boldsymbol{\theta}_0)| &\leq |H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{n,ij}(\boldsymbol{\theta}_0)| + |H_{n,ij}(\boldsymbol{\theta}_0) - H_{ij}(\boldsymbol{\theta}_0)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

By Lemma 1 (i) and Lemma 2 (i) we have shown that for all $\varepsilon > 0$, the event $\left\{|H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{ij}(\boldsymbol{\theta}_0)| \leq \varepsilon\right\}$ has probability tending to 1. Hence the proof is finished.

Proof of Theorem 1

The proof of Theorem 1 is a direct application of Lemma 2 with $\boldsymbol{\varphi} = \boldsymbol{\mathcal{D}}$ and $\mathbf{Y}_n = \hat{\boldsymbol{\mathcal{D}}}$.

Proof of Corollary 1

(i) The limiting covariance matrix of $\hat{\boldsymbol{\theta}}$, viewed as a function of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\Xi}(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\mathbf{W}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})\left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})\right)^{-1}.$$

By assumption, $\dot{\boldsymbol{\mathcal{D}}}(\cdot)$ and $\boldsymbol{\Sigma}(\cdot)$ are continuous at $\boldsymbol{\theta}_0$, hence so is $\boldsymbol{\Xi}(\cdot)$. Therefore, since $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$, we also have that $\boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})$ converges in probability to $\boldsymbol{\Xi}(\boldsymbol{\theta}_0)$. Moreover, since $\boldsymbol{\Xi}(\boldsymbol{\theta})$ is invertible and nonnegative definite for all $\boldsymbol{\theta}$ in Θ, we have $\boldsymbol{\Xi}(\boldsymbol{\theta}) = \boldsymbol{\Xi}^{1/2}(\boldsymbol{\theta})\boldsymbol{\Xi}^{1/2}(\boldsymbol{\theta})$ where $\boldsymbol{\Xi}^{1/2}(\boldsymbol{\theta})$ is also invertible. Therefore, by Theorem 1, as $n \to \infty$,

$$\sqrt{n}\boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_q),$$

leading to the desired result.

(ii) By Assumption (A4),

$$\sqrt{n}\left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$$

as $n \to \infty$. The arguments in the proof of (i) can be easily adapted to prove (ii).

Proof of Corollary 2

The proof of Corollary 2 is similar to that of Corollary 1 (i).

Proof of Corollary 3

Note $\boldsymbol{\mathcal{D}}_0 := \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$ and write

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} = \boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 + \boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}}. \tag{27}$$

A Taylor expansion yields

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 = \widetilde{\dot{\boldsymbol{\mathcal{D}}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{28}$$

where $\widetilde{\dot{\boldsymbol{\mathcal{D}}}} := \dot{\boldsymbol{\mathcal{D}}}(\tilde{\boldsymbol{\theta}})$ with $\tilde{\boldsymbol{\theta}}$ being a vector between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Substitute (28) into (27) to get

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} = \widetilde{\dot{\boldsymbol{\mathcal{D}}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}}. \tag{29}$$

From (28), we have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}})^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T (\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0). \tag{30}$$

Substitute (30) into (29) to obtain

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} = \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \left(\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \right)^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 \right) + \left(\boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}} \right).$$

Since

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 = \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) + \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}_0 \right),$$

we have

$$\left(\mathbf{I}_p - \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \left(\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \right)^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \right) \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) = \left(\mathbf{I}_p - \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \left(\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \right)^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \right) \left(\boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}} \right).$$

Take $\boldsymbol{\theta} \in \Theta$ and define $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) := \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \right)^{-1} \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T$. Likewise, write $\widetilde{\mathbf{A}} := \mathbf{A}(\widetilde{\boldsymbol{\theta}})$. By Assumption (A4) and because $\boldsymbol{\mathcal{D}}$ is continuously differentiable, as $n \to \infty$,

$$(\mathbf{I}_p - \widetilde{\mathbf{A}})\sqrt{n} \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) \xrightarrow{d} N\left(\mathbf{0}, (\mathbf{I}_p - \mathbf{A}_0)\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{A}_0)\right) \tag{31}$$

where $\mathbf{A}_0 := \dot{\boldsymbol{\mathcal{D}}}_0 \left(\dot{\boldsymbol{\mathcal{D}}}_0^T \dot{\boldsymbol{\mathcal{D}}}_0 \right)^{-1} \dot{\boldsymbol{\mathcal{D}}}_0^T$ and $\dot{\boldsymbol{\mathcal{D}}}_0 := \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}_0)$. Now write $\mathbf{I}_p - \mathbf{A}_0 = \mathbf{Q}\boldsymbol{\Delta}\mathbf{Q}^T$, where $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_p$, and $\boldsymbol{\Delta} = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$ with the number of ones being equal to k. Pre-multiply the left member of (31) by

$$\mathbf{Q}^T \left[(\mathbf{I}_p - \mathbf{A}_0)\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{A}_0) + \mathbf{A}_0 \right]^{-1/2} = \left[\boldsymbol{\Delta}\mathbf{Q}^T\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Delta} + \mathbf{I}_p - \boldsymbol{\Delta} \right]^{-1/2} \mathbf{Q}^T,$$

Note that the matrix between the brackets in the right-hand side is bloc-diagonal. It then can be verified that the limit normal distribution in the right member will have covariance matrix $\boldsymbol{\Delta}$, entailing

$$n \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right)^T (\mathbf{I}_p - \widetilde{\mathbf{A}})[(\mathbf{I}_p - \mathbf{A}_0)\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{A}_0) + \mathbf{A}_0]^{-1}(\mathbf{I}_p - \widetilde{\mathbf{A}}) \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) \to \chi^2_k.$$

Put $\widehat{\mathbf{A}} := \mathbf{A}(\hat{\boldsymbol{\theta}})$. Since $\widetilde{\mathbf{A}} \to \mathbf{A}_0$ in probability, we can replace $\widetilde{\mathbf{A}}$ and $\mathbf{A}_0$ by $\widehat{\mathbf{A}}$ to get the desired result.

Proof of Proposition 1

Without loss of generality, assume that $\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}} = \alpha\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mathcal{D}}}$ for some scalar α. Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{W})$ and note $\hat{\boldsymbol{\theta}}(\mathbf{W}^\star)$ the estimator for which $\mathbf{W} = \mathbf{W}^\star$. Denote by $\boldsymbol{\Xi}(\mathbf{W})$ and $\boldsymbol{\Xi}(\mathbf{W}^\star)$ the associated limiting covariance matrices of Theorem 1. We have

$$
\begin{aligned}
&\boldsymbol{\Xi}(\mathbf{W}) - \boldsymbol{\Xi}(\mathbf{W}^\star) \\
&= \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1} - \alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1} \\
&= \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}} - \dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1} \\
&= \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}^{1/2}\left(\mathbf{I}_p - \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2}\right)\boldsymbol{\Sigma}^{1/2}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}} \\
&\quad \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1},
\end{aligned}
$$

where $\boldsymbol{\Sigma}^{1/2}$ is the symmetric and invertible matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$. Write $\mathbf{A} = \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2}$. Note that $\mathbf{A}$ is idempotent, that is, $\mathbf{A}^2 = \mathbf{A}$. Indeed,

$$
\begin{aligned}
\mathbf{A}^2 =& \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2} \\
=& \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2} \\
=& \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2} \\
=& \mathbf{A}.
\end{aligned}
$$

Hence $\mathbf{I}_p - \mathbf{A}$ is idempotent as well and therefore

$$
\boldsymbol{\Xi}(\mathbf{W}) - \boldsymbol{\Xi}(\mathbf{W}^\star) = \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}^{1/2}(\mathbf{I}_p - \mathbf{A})(\mathbf{I}_p - \mathbf{A})\boldsymbol{\Sigma}^{1/2}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}
$$

which is easily seen to be nonnegative definite.

Proof of Proposition 2

The gradient of the loss function (3) is equal to $\mathbf{0}$ if and only if

$$
\dot{\boldsymbol{\mathcal{D}}}^T\widehat{\mathbf{W}}\left(\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\mathcal{D}}}\right) = 0.
$$

But since $\dot{\boldsymbol{\mathcal{D}}}$ is of full rank and $p = q$, the kernel of $\dot{\boldsymbol{\mathcal{D}}}^T$ is null, hence

$$
\widehat{\mathbf{W}}\left(\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\mathcal{D}}}\right) = 0.
$$

The fact that $\widehat{\mathbf{W}}$ is of full rank concludes the proof.

References

[1] D. Berg. Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701, 2009.

[2] A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics*, 39(4):1963–2006, 2011.

[3] A. Bücher, J. Segers, and S. Volgushev. When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *The Annals of Statistics*, 42(4):1598–1634, 08 2014.

[4] P. Capéraà, A.L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 1997.

[5] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[6] C. M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353, 1981.

[7] P. Deheuvels. On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5):429–439, 1991.

[8] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.

[9] F. Durante and C. Sempi. Copula theory: An introduction. In *Copula Theory and Its Applications*, pages 3–31. Springer, 2010.

[10] J. Einmahl, A. Krajina, and J. Segers. An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics*, 40(3):1764–1793, 2012.

[11] M. Ferreira. Nonparametric estimation of the tail-dependence coefficient. *REVSTAT–Statistical Journal*, 11(1):1–16, 2013.

[12] M. Fréchet. Remarques au sujet de la note précédente. *CR Acad. Sci. Paris Sér. I Math*, 246:2719–2720, 1958.

[13] C. Genest and A. C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

[14] C. Genest, K. Ghoudi, and L. P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.

[15] C. Genest, J. Nešlehová, and N. Ben Ghorbal. Estimators based on Kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2):157–177, 2011.

[16] C. Genest and B. Rémillard. Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369, 2004.

[17] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213, 2009.

[18] C. Genest and L. P. Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043, 1993.

[19] C. Genest and J. Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022, 2009.

[20] G. Gudendorf and J. Segers. Extreme-value copulas. In *Copula Theory and Its Applications*, pages 127–145. Springer, 2010.

[21] P. Hall and N. Tajvidi. Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(5):835–844, 2000.

[22] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[23] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

[24] H. Joe. *Multivariate models and dependence concepts*. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[25] C. Klüppelberg and G. Kuhn. Copula structure analysis. *Journal of the Royal Statistical Society: Series B*, 71(3):737–753, 2009.

[26] I. Kojadinovic and J. Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21(1):17–30, 2011.

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

[28] R. B. Nelsen. *An introduction to copulas*. Springer, 2006.

[29] R. B. Nelsen, J. J. Quesada-Molina, J.A. Rodríguez-Lallena, and M. Úbeda-Flores. Kendall distribution functions. *Statistics & Probability Letters*, 65(3):263–268, 2003.

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

[31] J. Pickands. Multivariate extreme value distributions. *Proceedings of the 43rd Session of the International Statistical Institute*, 2:859–878, 1981.

[32] J. Segers. Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782, 2012.

[33] H. Tsukahara. Semiparametric estimation in copula models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 33(3):357–375, 2005.