



**HAL**  
open science

## Weighted least square inference based on dependence coefficients for multivariate copulas

Gildas Mazo, Stéphane Girard, Florence Forbes

► **To cite this version:**

Gildas Mazo, Stéphane Girard, Florence Forbes. Weighted least square inference based on dependence coefficients for multivariate copulas. 2014. hal-00979151v1

**HAL Id: hal-00979151**

**<https://hal.science/hal-00979151v1>**

Preprint submitted on 15 Apr 2014 (v1), last revised 22 Oct 2015 (v6)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Weighted least square inference based on dependence coefficients for multivariate copulas

Gildas Mazo, Stéphane Girard and Florence Forbes

April 15, 2014

## Abstract

The concept of copulas is useful to model multivariate distributions. Given a parametric family of copulas, the inference of the parameter vector commonly relies on likelihood-based methods. However, for some copula families, the likelihood may not exist, or may lead to slow or complex numerical optimization procedures. Therefore, it is desirable to consider alternative estimation strategies. A natural approach is to build the inference on bivariate dependence coefficients, where the weighted sum of the squared residuals between the dependence coefficients under the model and their empirical counterparts is minimized. This method has already been used in some applications but in a rather heuristic way. The asymptotic properties of the resulting estimator have not been investigated yet. In this paper, we derive the consistency and asymptotic normality of the weighted least square estimator based on three standard dependence coefficients. Finally we illustrate how our results can be used to address three statistical questions.

## 1 Introduction

The concept of copulas is useful to model multivariate distributions. Given a multivariate random vector of interest, copula allow to separate the analysis of the margins from the dependence structure. Standard books covering this subject include [19, 23]. See also [10] for an introduction to this topic.

When a parametric family of copulas has been chosen, the inference of the parameter vector must be considered. It is then common to rely on likelihood-based methods [1, 3, 17, 20]. However, for some copula families, the likelihood may not exist [5, 6, 8, 18, 21, 28], or may lead to slow [29] or complex [1] numerical optimization procedures. Therefore, it is desirable to consider alternative estimation strategies. An heuristic approach is to build the inference on bivariate dependence coefficients, where the weighted sum of the squared residuals between the dependence coefficients under the model and their empirical counterparts is minimized. This idea has been applied in several applications. See, for instance, [15, 29] for the estimation of the parameter of multivariate Archimedean copulas, or [6, 28] for the estimation of parameter vectors of non-differentiable copulas. We also refer to [30] for the estimation of the parameter vector of a spatial extreme-value model. The inference of elliptical copulas (see [4] and [22] Example 5.53) can be viewed as a particular case of the method

considered here. Finally, see [2] for an application of this estimation strategy to goodness-of-fit testing.

The asymptotic properties of the weighted least square estimator based on dependence coefficients have not been investigated yet. This work aims at bridging this gap.

The remaining of this paper is as follows. In Section 2, we derive the consistency and asymptotic normality of the weighted least square estimator based on three standard dependence coefficients, namely, the Spearman's rho, the Kendall's tau, and the upper tail dependence coefficient. In Section 3, we illustrate the theoretical results of Section 2 with three statistical questions arising in three scientific fields. All the proofs are postponed to the Appendix.

## 2 Asymptotic properties of the weighted least square estimator based on dependence coefficients

In this section, we derive the consistency and asymptotic normality of the weighted least square estimator based on the Spearman's rho, the Kendall's tau, and the upper tail dependence coefficients. In practice, the different dependence coefficients define different estimators. However, since the methods to derive the properties of these estimators are similar, we shall define and study one generic estimator embedding all the three cases. In Section 2.1 the asymptotic properties of the generic estimator are established. In Section 2.2 we deal with the practical estimators based on the actual dependence coefficients.

Let  $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(n)}$  with  $\mathbf{U}^{(k)} = (U_1^{(k)}, \dots, U_d^{(k)})$ ,  $k = 1, \dots, n$  be independent and identically distributed copies of  $\mathbf{U} = (U_1, \dots, U_d)$  whose distribution is a copula  $C$  with parameter vector  $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^q$ . Let  $p = d(d-1)/2$  be the number of variable pairs  $(U_i, U_j)$ , for  $i = 1, \dots, d-1$ ,  $j = 2, \dots, d$ ,  $i < j$ . Let us define the vector map

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

where  $\mathcal{D}_{i,j}(\cdot)$  represents a well chosen dependence coefficient between the variables  $U_i$  and  $U_j$ .  $\mathcal{D}(\Theta)$  stands for the image of  $\Theta$  by the multivariate map  $\mathcal{D}$ . The coordinates of  $\mathcal{D}(\boldsymbol{\theta})$  are the  $\mathcal{D}_{i,j}(\boldsymbol{\theta})$  sorted in the lexicographical order. When the map  $\mathcal{D}$  is differentiable, its Jacobian matrix at  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$  is denoted by

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

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

The estimator of  $\theta_0$  studied in this paper is defined as

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

where  $\widehat{\mathbf{W}} = \widehat{\mathbf{W}}_n$  is a sequence ( $n = 1, 2, \dots$ ) of symmetric and positive definite matrices with full rank. Let us note  $\hat{\ell}(\theta)$  the loss function to be minimized in (3). For instance, if  $\widehat{\mathbf{W}}$  is a diagonal matrix with  $(\hat{w}_{1,2}, \dots, \hat{w}_{d-1,d})$  on its diagonal, then  $\hat{\ell}(\theta)$  reduces to the weighted sum of squared residuals

$$\sum_{i < j} \hat{w}_{i,j} (\widehat{\mathcal{D}}_{i,j} - \mathcal{D}_{i,j}(\theta))^2.$$

In theory, the minimizer  $\hat{\theta}$  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  $\hat{\theta}$  hold with probability tending to one as the sample size increases. Since  $\widehat{\mathbf{W}}$  is nonnegative definite, the loss function  $\hat{\ell}$  is such that  $\hat{\ell}(\theta) \geq 0$  for all  $\theta \in \Theta$  and vanishes at  $\hat{\theta}$  if and only if  $\hat{\theta} \in \mathcal{D}^{-1}(\{\widehat{\mathcal{D}}\})$ , where  $\mathcal{D}^{-1}(\{\widehat{\mathcal{D}}\})$  denotes the set of all  $\theta$  in  $\Theta$  such that  $\mathcal{D}(\theta) = \widehat{\mathcal{D}}$ . In this case, the estimator does not depend on the weights and  $\mathcal{D}(\hat{\theta}) = \widehat{\mathcal{D}}$ . Moreover, if the multivariate map  $\mathcal{D}$  is one-to-one, then  $\hat{\theta} = \mathcal{D}^{-1}(\widehat{\mathcal{D}})$ . If  $\widehat{\mathcal{D}}$  takes the form of a  $p$ -dimensional empirical mean,  $\hat{\theta}$  can be regarded as a generalized method-of-moment estimator [12]. In this paper, however, more general functionals of the sample are considered and therefore the generalized method-of-moment framework does not apply. This remark was also made in [25] where the authors used a simulated method-of-moment to estimate the parameters of copulas.

In Section 2.2 and after,  $\mathcal{D}_{i,j}(\theta)$  will be replaced by a dependence coefficient between  $U_i$  et  $U_j$ , such as the Spearman's rho, the Kendall's tau, or the upper tail dependence coefficient (their definitions are postponed to Section 2.2).

## 2.1 Asymptotic properties of the generic estimator

The assumptions needed to derive the asymptotic properties of  $\hat{\theta}$  defined by the minimization of the loss function  $\hat{\ell}$  (3) are given below. The symbol  $\|\cdot\|$  denotes the Euclidean norm.

**Assumption.** (A1) *The true parameter vector  $\theta_0$  lies in the interior of  $\Theta$ . Moreover, there exists  $\varepsilon_0 > 0$  such that the set  $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \varepsilon_0\}$  is closed (and thus compact).*

(A2) *As  $n \rightarrow \infty$ , 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  $\mathcal{D}$  defined in (1) is a twice continuously differentiable homeomorphism such that  $\widehat{\mathcal{D}}$  is of full rank.*

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

$$\begin{aligned} \widehat{\mathcal{D}} &\xrightarrow{P} \mathcal{D}, \text{ and} \\ \sqrt{n} \left( \widehat{\mathcal{D}} - \mathcal{D} \right) &\xrightarrow{d} N_p(\mathbf{0}, \Sigma), \end{aligned} \quad (4)$$

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

$$\Sigma = \begin{pmatrix} \Sigma_{1,2;1,2} & \Sigma_{1,2;1,3} & \cdots & \Sigma_{1,2;d-1,d} \\ \Sigma_{1,3;1,2} & \Sigma_{1,3;1,3} & \cdots & \Sigma_{1,3;d-1,d} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_{d-1,d;1,2} & \Sigma_{d-1,d;1,3} & \cdots & \Sigma_{d-1,d;d-1,d} \end{pmatrix}. \quad (5)$$

Assumption (A1) is quite standard and will be taken for granted throughout this paper. A sequence of weight matrices verifying (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  $p$ . Needless to say, better alternatives can be chosen, as we shall see afterwards. This assumption was also made in [12] in the context of the generalized method-of-moment. Assumption (A3) is natural. Indeed, the estimation of the copula parameter vector is performed by trying to make the theoretical and empirical dependence coefficients match, hence a success in this match should ensure that the resulting parameter vector estimate is close to the true value. In other words, the model should be identifiable. The homeomorphism condition ensures this. This assumption was also made in [8] where the authors estimate the parameter vector of an extreme-value model. Assumption (A4) simply states that one should have reasonable empirical estimators to ensure good convergence properties for  $\hat{\boldsymbol{\theta}}$ .

**Theorem 1.** *Assume that (A1), (A2), (A3) and (A4) hold. Then with probability tending to one and as  $n \rightarrow \infty$ , the estimator  $\boldsymbol{\theta}$  defined by (3) exists, is unique, and is consistent for  $\boldsymbol{\theta}_0$ . Moreover, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q(\mathbf{0}, \Xi), \quad (6)$$

where

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

It is possible to build confidence regions for the true parameter vector and the dependence coefficients, using the results of the following corollary. Let  $\chi_q^2$  denote the chi square distribution with  $q$  degrees of freedom. Let us write  $\Xi = \Xi(\boldsymbol{\theta})$  and  $\Sigma = \Sigma(\boldsymbol{\theta})$  to emphasize that in general these matrices depend on  $\boldsymbol{\theta}$ . In the following, when matrices are said to be continuous with respect to the parameter vector  $\boldsymbol{\theta}$ , the continuity is meant elementwise.

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

- (i) *If  $\Xi(\boldsymbol{\theta})$  is invertible for all  $\boldsymbol{\theta}$  in  $\Theta$  and  $\boldsymbol{\theta} \mapsto \Sigma(\boldsymbol{\theta})$  is continuous at  $\boldsymbol{\theta}_0$ , then, as  $n \rightarrow \infty$ ,*

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

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

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

Observe that, if  $\Sigma(\cdot)$  is invertible for all  $\theta$  in  $\Theta$  and is continuous at  $\theta_0$ , then putting  $\widehat{\Sigma}^{-1} = \Sigma(\widehat{\theta})$  allows to satisfy the assumption of Corollary 1 (ii).

It is possible to test particular values for a subset of the parameter vector based on Corollary 2 below. For  $\theta \in \Theta$ , write  $\theta = (\theta_1, \theta_2) \in \mathbb{R}^r \times \mathbb{R}^{q-r}$ . Then the estimator and the true parameter vector write  $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)$  and  $\theta_0 = (\theta_{01}, \theta_{02})$  respectively. Let  $\Xi_1(\theta_1, \theta_2)$  denote the asymptotic covariance matrix of size  $r \times r$  corresponding to  $\theta_1$ , that is

$$\Xi(\theta_1, \theta_2) = \begin{pmatrix} \Xi_1(\theta_1, \theta_2) & | & \dots \\ \hline \vdots & & | & \dots \end{pmatrix}.$$

The test  $H0 : \theta_{01} = \theta_1^*$  against  $H1 : \theta_{01} \neq \theta_1^*$  for a particular value  $\theta_1^*$  can be performed based on the following corollary.

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

$$n(\widehat{\theta}_1 - \theta_1^*)^T \Xi_2(\theta_1^*, \widehat{\theta}_2)^{-1} (\widehat{\theta}_1 - \theta_1^*) \xrightarrow{d} \chi_r^2.$$

How to decide if the parametric model chosen for the dependence coefficients is reasonable? Let  $H0$  be the null hypothesis “the chosen parametric model  $\theta \mapsto \mathcal{D}(\theta)$  is the true map of the underlying copula” against  $H1$  “the chosen parametric model  $\theta \mapsto \mathcal{D}(\theta)$  is *not* the true map of the underlying copula”. If  $H0$  is true, then  $\mathcal{D}(\widehat{\theta}) - \widehat{\mathcal{D}}$  should be close to zero. To carry out this test, one may use the result of the following corollary.

**Corollary 3.** *Suppose that the assumptions of Theorem 1 hold. Define*

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

*If  $\mathbf{I}_p - \mathbf{A}(\theta)$  and  $\Sigma(\theta)$  are invertible for all  $\theta$  in  $\Theta$  and  $\Sigma(\cdot)$  is continuous at  $\theta_0$ , then*

$$n \left( \mathcal{D}(\widehat{\theta}) - \widehat{\mathcal{D}} \right) \Sigma^{-1}(\widehat{\theta}) \left( \mathbf{I}_p - \mathbf{A}(\widehat{\theta}) \right)^{-1} \Sigma^{-1}(\widehat{\theta}) \left( \mathcal{D}(\widehat{\theta}) - \widehat{\mathcal{D}} \right) \xrightarrow{d} \chi_p^2$$

*as  $n \rightarrow \infty$ .*

The asymptotic covariance matrix  $\Xi$  for the limiting distribution (6) depends on the limiting weight matrix  $\mathbf{W}$ . Hence it is natural to wonder what is the “best” estimator  $\widehat{\theta}$  one can construct in the sense of the “best” limiting distribution. This problem of finding optimal weights was studied in the context of the generalized method of moments [12], and here we derive similar results. To define what “best” could mean in this context, the mathematical concept of nonnegativeness is appropriate. Let us introduce the notation  $\mathbf{A} \geq 0$  for a nonnegative matrix  $\mathbf{A}$ . Also, for two nonnegative matrices  $\mathbf{A}$  and  $\mathbf{B}$ , we note  $\mathbf{A} \leq \mathbf{B}$  if  $\mathbf{B} - \mathbf{A} \geq 0$ . Hence, given two limiting distributions  $N(\mathbf{0}, \mathbf{A})$  and  $N(\mathbf{0}, \mathbf{B})$ , the best limiting distribution is defined as the one which has the smallest covariance matrix. In particular, it is easily checked that  $\mathbf{A} \leq \mathbf{B}$  implies  $\text{tr}(\mathbf{A}) \leq \text{tr}(\mathbf{B})$ , where  $\text{tr}(\cdot)$  stands for the trace operator of matrices. Thus, within the class of all possible limiting distribution, the best one is such that the sum of the variances is minimum. The next proposition gives the optimal choice for the limiting weight matrix  $\mathbf{W}$ , leading to the best limiting distribution.

**Proposition 1.** *Suppose that  $\Sigma$  defined in (A4) is invertible. Then the asymptotic covariance matrix  $\Xi$  (6) is minimum for  $\mathbf{W}^*$  such that*

$$\mathbf{W}^* \dot{\mathcal{D}} \propto \Sigma^{-1} \dot{\mathcal{D}} \quad (7)$$

where the symbol  $\propto$  denotes proportionality.

The result of Proposition 1 indicates that one can choose a weight matrix  $\widehat{\mathbf{W}}$  in the construction of  $\widehat{\boldsymbol{\theta}}$  to lead to an optimal estimator in the sense that  $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$  converges to the most efficient asymptotic distribution, that is, the one which has the smallest asymptotic covariance matrix. Let us write  $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\widehat{\mathbf{W}})$  to emphasize the dependence on  $\widehat{\mathbf{W}}$ . In order to construct such an optimal estimator, it suffices to let the sequence of weight matrices to converge to  $\Sigma^{-1}$ . This is summarized in the next corollary (left without proof).

**Corollary 4.** *If, as  $n \rightarrow \infty$ ,  $\widehat{\Sigma}$  is a consistent estimator for  $\Sigma$  (elementwise), then  $\boldsymbol{\theta}(\widehat{\Sigma}^{-1})$  is optimal and*

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

The estimator  $\widehat{\Sigma}$  could be an empirical estimator of  $\Sigma$ , but, since  $\Sigma$  depends on  $\boldsymbol{\theta}_0$ , it also could be constructed from an estimator of  $\boldsymbol{\theta}_0$ . This estimator could be chosen in turn as the inverse of a weight matrix to estimate the parameter, and so on. Algorithm 1 constructs an updating sequence of estimates  $\widehat{\Sigma}^{(t)}$ ,  $t = 0, 1, 2, \dots, T$  such that each of the elements of the sequence for  $t = 1, 2, \dots$  converges in probability to  $\Sigma(\boldsymbol{\theta}_0)$  as  $n \rightarrow \infty$ . The performance of

---

**Algorithm 1**

---

Put  $\widehat{\Sigma}^{(0)} = \mathbf{I}_p$ .  
**for**  $t = 0, 1, \dots, T$  **do**  
     $\widehat{\boldsymbol{\theta}}^{(t)} = \boldsymbol{\theta}(\widehat{\Sigma}^{(t)})$   
     $\widehat{\Sigma}^{(t+1)} = \Sigma[\widehat{\boldsymbol{\theta}}^{(t)}]$ .  
**end for**

---

these estimators will be assessed numerically in Section 3.2.

The benefit of weighting can vanish if the distribution of  $(U_1, \dots, U_d)$  is exchangeable, that is, if  $C(u_1, \dots, u_d) = C(u_{\pi(1)}, \dots, u_{\pi(d)})$  for any permutation  $\pi$  of  $(1, \dots, d)$ , as the next proposition shows.

**Proposition 2.** *If the distribution of  $(\widehat{\mathcal{D}}_{i,j}, \widehat{\mathcal{D}}_{k,l})$  only depends on that of*

$$(U_i, U_j, U_k, U_l) \quad (\forall i < j, k < l),$$

and the underlying copula  $C$  is exchangeable, then  $\mathbf{W}^* = \mathbf{I}_p$  verifies (7).

Proposition 2 holds for the popular class of Archimedean copulas. A copula  $C$  is an Archimedean copula if

$$C(u_1, \dots, u_d) = \varphi^{-1}\left(\sum_{i=1}^d \varphi(u_i)\right) \quad (8)$$

where  $\varphi : [0, 1] \rightarrow [0, \infty]$  is a continuous strictly decreasing function such that  $\varphi(0) = \infty$  and  $\varphi(1) = 0$  (see [23] Section 4.6 p. 151 for more about Archimedean copulas). It is easy to see that these copulas are exchangeable. Hence, for those copulas, the smallest asymptotic variance is obtained when there are no weights (that is, when  $\widehat{\mathbf{W}}$  is the identity matrix  $\mathbf{I}_p$ ).

Another situation exists where the weights have no effect. If there are as many pairs as parameters, then  $\hat{\boldsymbol{\theta}}$  is explicit and does not depend on  $\widehat{\mathbf{W}}$ , as the next proposition shows.

**Proposition 3.** *Suppose that the assumptions of Theorem 1 hold. If  $p = q$  then, as  $n \rightarrow \infty$  and with probability tending to one,*

$$\hat{\boldsymbol{\theta}} = \mathcal{D}^{-1}(\widehat{\mathcal{D}}), \quad (9)$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \left(\dot{\mathcal{D}}^T \dot{\mathcal{D}}\right)^{-1} \dot{\mathcal{D}}^T \boldsymbol{\Sigma} \dot{\mathcal{D}} \left(\dot{\mathcal{D}}^T \dot{\mathcal{D}}\right)^{-1}\right). \quad (10)$$

This special case will be illustrated in Section 3.3.

## 2.2 Asymptotic properties for the estimators based on dependence coefficients

This section is devoted to the study of the estimator  $\hat{\boldsymbol{\theta}}$  defined in (3), but with  $\mathcal{D}_{i,j}$  being replaced by  $\rho_{i,j}$ ,  $\tau_{i,j}$  or  $\lambda_{i,j}$ , respectively denoting the Spearman's rho, Kendall's tau, and upper tail dependence coefficients between the variables  $U_i$  and  $U_j$ .

The definitions of the considered dependence coefficients are recalled below. Given a bivariate copula  $C_{i,j}$ , the Spearman's rho and the Kendall's tau are respectively defined by

$$\rho_{i,j}(\boldsymbol{\theta}) = 12 \int_{[0,1]^2} C_{i,j}(u,v) du dv - 3, \quad (11)$$

$$\tau_{i,j}(\boldsymbol{\theta}) = 4 \int_{[0,1]^2} C_{i,j}(u,v) dC_{i,j}(u,v) - 1. \quad (12)$$

They belong to  $[-1, 1]$  and measure the dependence between the variables associated to  $C_{i,j}$ . Negative (respectively positive) values indicate negative (respectively positive) dependence. A value of zero indicates independence (see e.g. [23], chapter 5).

If  $C_{i,j}$  is an extreme-value copula, it is natural (see e.g. [11] Section 6.4) to measure the dependence with the upper tail dependence coefficient, or simply tail dependence coefficient, defined by

$$\lambda_{i,j}(\boldsymbol{\theta}) = 2 + \log C_{i,j}(e^{-1}, e^{-1}). \quad (13)$$

This coefficient belongs to  $[0, 1]$  and a value closer to 1 indicates more (positive) dependence, and a value closer to 0 indicates more independence (there is no negative dependence for extreme-value copulas). Recall that a copula  $C$  is an extreme-value copula if there exists a copula  $\tilde{C}$  such that

$$C(u_1, \dots, u_d) = \lim_{n \uparrow \infty} \tilde{C}^n(u_1^{1/n}, \dots, u_d^{1/n}), \quad (14)$$



for every  $(u_1, \dots, u_d) \in [0, 1]^d$ . See for instance [11] for more about extreme-value copulas.

Empirical estimators have to be chosen to estimate nonparametrically the dependence coefficients. Common ones are considered in the propositions below (one proposition per type of dependence coefficient). In each of these, it is shown that Assumption (A4) holds.

**Proposition 4** (Spearman's rho). *Let  $\mathcal{D}_{i,j}(\boldsymbol{\theta}) = \rho_{i,j}(\boldsymbol{\theta})$ . Also, define*

$$\widehat{\mathcal{D}}_{i,j} = \widehat{\rho}_{i,j} = \frac{\sum_{k=1}^n (U_i^{(k)} - \bar{U}_i) (U_j^{(k)} - \bar{U}_j)}{\left[ \sum_{k=1}^n (U_i^{(k)} - \bar{U}_i)^2 \sum_{k=1}^n (U_j^{(k)} - \bar{U}_j)^2 \right]^{1/2}}, \quad (15)$$

where  $\bar{U}_i = \sum_{k=1}^n U_i^{(k)} / n$ ,  $\bar{U}_i^2 = \sum_{k=1}^n (U_i^{(k)})^2 / n$  and  $\overline{U_i U_j} = \sum_{k=1}^n U_i^{(k)} U_j^{(k)} / n$ . Put  $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_{1,2}, \dots, \widehat{\rho}_{d-1,d})$  and  $\boldsymbol{\rho} = (\rho_{1,2}, \dots, \rho_{d-1,d})$  where  $\rho_{i,j} = \rho_{i,j}(\boldsymbol{\theta}_0)$ . Then as  $n \rightarrow \infty$

$$\begin{aligned} \widehat{\boldsymbol{\rho}} &\xrightarrow{P} \boldsymbol{\rho} \text{ and} \\ \sqrt{n}(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}) &\xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \quad (16)$$

where  $\boldsymbol{\Sigma}$  is some symmetric and positive definite matrix.

Although the convergence to a normal distribution is established, the expression for the asymptotic covariance matrix is unknown. It is possible to replace the empirical means and variances in formula (15) by their true values to get a closed form expression for  $\boldsymbol{\Sigma}$ . Let

$$\widehat{\rho}_{i,j} = \frac{12}{n} \sum_{k=1}^n \left( U_i^{(k)} - \frac{1}{2} \right) \left( U_j^{(k)} - \frac{1}{2} \right). \quad (17)$$

By the multivariate central limit theorem, (16) holds true with

$$\begin{aligned} \Sigma_{i,j;k,l} &= 144 \mu_{i,j;k,l} - \rho_{i,j} \rho_{k,l}, \\ \mu_{i,j;k,l} &= \mathbb{E} \left[ \left( U_i - \frac{1}{2} \right) \left( U_j - \frac{1}{2} \right) \left( U_k - \frac{1}{2} \right) \left( U_l - \frac{1}{2} \right) \right]. \end{aligned} \quad (18)$$

In this particular case, the empirical estimator  $\widehat{\rho}_{i,j}$  is the empirical mean of the variables  $12(U_i^{(k)} - \frac{1}{2})(U_j^{(k)} - \frac{1}{2})$ ,  $k = 1, \dots, n$ , hence, the estimator  $\widehat{\boldsymbol{\rho}}$  can be embedded within the framework of the generalized method of moments [12].

**Proposition 5** (Kendall's tau). *Let  $\mathcal{D}_{i,j}(\boldsymbol{\theta}) = \tau_{i,j}(\boldsymbol{\theta})$ . Also, define*

$$\widehat{\mathcal{D}}_{i,j} = \widehat{\tau}_{i,j} = \binom{n}{2}^{-1} \sum_{k < l} \text{sign} \left( (U_i^{(k)} - U_i^{(l)})(U_j^{(k)} - U_j^{(l)}) \right), \quad (19)$$

where  $\text{sign}(x) = 1$  if  $x > 0$ ,  $-1$  if  $x < 0$  and  $0$  if  $x = 0$ . Put  $\widehat{\boldsymbol{\tau}} = (\widehat{\tau}_{1,2}, \dots, \widehat{\tau}_{d-1,d})$  and  $\boldsymbol{\tau} = (\tau_{1,2}, \dots, \tau_{d-1,d})$  where  $\tau_{i,j} = \tau_{i,j}(\boldsymbol{\theta}_0)$ . Then as  $n \rightarrow \infty$

$$\begin{aligned} \widehat{\boldsymbol{\tau}} &\xrightarrow{P} \boldsymbol{\tau} \text{ and} \\ \sqrt{n}(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}) &\xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned}$$

where

$$\begin{aligned} \Sigma_{i,j;k,l} = & 4\mathbb{E}[(4C_{i,j}(U_i, U_j) + 1 - \tau_{i,j} - 2U_i - 2U_j) \\ & \times (4C_{k,l}(U_k, U_l) + 1 - \tau_{k,l} - 2U_k - 2U_l)]. \end{aligned} \quad (20)$$

**Proposition 6** (tail dependence coefficient). *Assume that the copula of interest  $C$  is an extreme-value copula and let  $\mathcal{D}_{i,j}(\boldsymbol{\theta}) = \lambda_{i,j}(\boldsymbol{\theta})$ . Also, define*

$$\widehat{\mathcal{D}}_{i,j} = \widehat{\lambda}_{i,j} = 3 - \frac{1}{1 - \sum_{k=1}^n \max(U_i^{(k)}, U_j^{(k)})/n}, \quad (21)$$

put  $\widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_{1,2}, \dots, \widehat{\lambda}_{d-1,d})$  and  $\boldsymbol{\lambda} = (\lambda_{1,2}, \dots, \lambda_{d-1,d})$  where  $\lambda_{i,j} = \lambda_{i,j}(\boldsymbol{\theta}_0)$ . Then as  $n \rightarrow \infty$

$$\begin{aligned} \widehat{\boldsymbol{\lambda}} & \xrightarrow{P} \boldsymbol{\lambda} \text{ and} \\ \sqrt{n}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) & \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned}$$

where

$$\Sigma_{i,j;k,l} = (3 - \lambda_{i,j})^2(3 - \lambda_{k,l})^2 \text{Cov}(\max(U_i, U_j), \max(U_k, U_l)). \quad (22)$$

To summarize, we have shown that each of the Spearman's rho, Kendall's tau, and tail dependence coefficients can form a weighted least square estimator verifying Assumption (A4).

### 3 Applications to three statistical questions

This section illustrates how the results derived in Section 2 can serve to assess the uncertainty of risk measures in credit risk modeling (Section 3.1) and hydrology (Section 3.3). In Section 3.2 it is shown that one can improve inference in presence of missing observations by using weights. The numerical computations were carried out with the R software for statistical computing [27], especially with the help of the packages [7, 14, 16, 26].

In each of the subsequent sections, the asymptotic covariance matrix  $\boldsymbol{\Sigma}$  of Assumption (A4) will need to be calculated. The formulas to compute the  $(i, j; k, l)$  element of this matrix are given in the propositions of Section 2.2. Nevertheless, due to the complexity of these formulas, we chose to base the computation of  $\boldsymbol{\Sigma}$  on Monte-Carlo simulations. More precisely, we generated 10,000 observations of the underlying copula model in order to approximate  $\Sigma_{i,j;kl}$  by its corresponding Monte-Carlo empirical mean. Generating more simulations (as many as 100,000) did not change significantly the calculations.

#### 3.1 Measuring uncertainty for loss probabilities in credit risk modeling

In credit risk modeling, a *threshold model* (see [22] Section 8.3) is characterized by a random vector  $\mathbf{X} = (X_1, \dots, X_d)$  and a threshold vector  $\mathbf{a} = (a_1, \dots, a_d)$ . The variable  $X_i$  represents the asset value of the firm  $i$  at a certain maturity date. This firm defaults when its asset value is smaller than the threshold  $a_i$ , that is,

when  $X_i \leq a_i$ . Assume that all firms have the same probability  $\pi$  of defaulting, in other words,  $P(X_i \leq a_i) = \pi$ . In risk analysis, the distribution of the loss associated to a portfolio plays a major role. Consider an homogeneous portfolio whose loss  $L$  is given by the number of defaults. Hence,  $L = \sum_{i=1}^d Y_i$  where  $Y_i = 1_{\{X_i \leq a_i\}}$  for  $i = 1, \dots, d$ . Observe that the distribution of  $L$  depends on the joint distribution of  $\mathbf{X}$ . More precisely, it depends on the copula associated to  $\mathbf{X}$  (see [22] Section 8.4). If the copula associated to  $\mathbf{X}$  is a  $d$ -dimensional Clayton copula (that is, a copula for which (8) holds with  $\varphi_\theta(t) = t^\theta - 1$ ) with parameter  $\theta > 0$ , then the probability of experiencing  $l$  defaults,  $l = 0, 1, \dots, d$ , is given by (see [22] Section 8.4)

$$\tilde{p}_l(\theta) := \frac{\binom{d}{l}}{\Gamma(1/\theta)(\pi^{-\theta} - 1)} \int_0^1 t^{\frac{l-1+1}{(\pi^{-\theta}-1)}} (1-t)^{d-l} \left( \frac{\log(t)}{1-\pi^{-\theta}} \right)^{1/\theta-1}. \quad (23)$$

These probabilities are represented Figure 1 with  $\theta = 2$  and  $\pi = 0.05$ . If  $\pi$  is known, this probability is a functional of  $\theta$ , and an estimator of the parameter gives an estimator of the loss probability. Let  $\theta_0$  be the true parameter. We wish to estimate it with  $\hat{\theta}$  defined in (3) by using the pairwise Kendall's tau. Hence the framework of Proposition 5 applies. Since the underlying copula is exchangeable and the parameter is one-dimensional, an explicit formula for  $\hat{\theta}$  exists. First, by Corollary 2, one can set  $\widehat{\mathbf{W}}$  to the identity matrix  $\mathbf{I}_p$ . Second, since  $q = 1$  and the underlying copula is exchangeable,  $\boldsymbol{\tau}(\theta)$  is equal to  $\tau(\theta)\mathbf{1}_p$ , where  $\tau(\theta) = \theta/(\theta + 2)$  for  $\theta > 0$ . Hence one can minimize the loss function of (3) by equating its derivative to zero to get

$$\hat{\theta} = \tau^{-1} \left( \frac{1}{p} \sum_{i < j} \hat{\tau}_{i,j} \right) \quad (24)$$

where  $\hat{\tau}_{i,j}$  is given in (19). By substituting (24) into (23), one gets an estimator  $\tilde{p}_l(\hat{\theta})$  of the true loss probability  $\tilde{p}_l := \tilde{p}_l(\theta_0)$ .

How to report uncertainty measures for  $\tilde{p}_l(\hat{\theta})$ ? It is easily checked that the assumptions of Theorem 1 hold, hence, as  $n \rightarrow \infty$ ,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N \left( 0, \frac{\mathbf{1}_p^T \boldsymbol{\Sigma} \mathbf{1}_p}{(p\tau')^2} \right),$$

where  $\boldsymbol{\Sigma}$  has been given in (20) and  $\tau'$  denotes the derivative of  $\tau(\cdot)$  at  $\theta_0$ . An application of the delta-method yields, as  $n \rightarrow \infty$ ,

$$\sqrt{n}(\tilde{p}_l(\hat{\theta}) - \tilde{p}_l) \xrightarrow{d} N(0, \eta_l^2), \text{ where}$$

$$\eta_l^2 = \left( \frac{d\tilde{p}_l(\theta)}{d\theta} \Big|_{\theta=\theta_0} \right)^2 \frac{\mathbf{1}_p^T \boldsymbol{\Sigma} \mathbf{1}_p}{(p\tau')^2}.$$

Thus, one can approximate (for  $n$  large enough) the distribution of  $\tilde{p}_l(\hat{\theta})$  by a normal distribution. This finite sample approximation was verified by simulation. We generated 500 replications of a dataset of size  $n = 200$ , dimension  $d = 5$  and parameter  $\theta_0 = 2$  and checked that the intervals  $\tilde{p}_l \pm 1.96\eta_l/\sqrt{n}$  contain 95% of the estimates  $\tilde{p}_l(\hat{\theta})$ . The results were 94.8, 94.4, 95.0, 94.6, 95.2 and 94.2 percent at  $l = 0, 1, 2, 3, 4, 5$  respectively. Figure 1 illustrates these intervals and the loss probabilities.

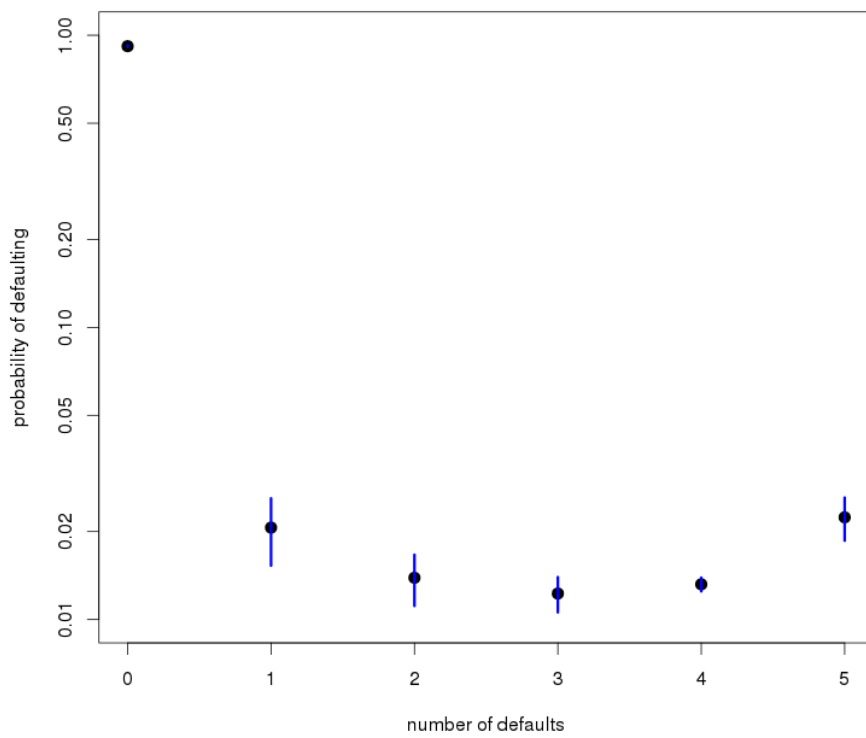


Figure 1: Loss probabilities  $\tilde{p}_l$  for  $l = 0, 1, \dots, 5$  and associated asymptotic 95% confidence intervals with  $n = 200$ .

### 3.2 Weighting to improve the inference of Archimedean copulas in presence of missing values

Consider the following situation. We wish to estimate the parameter  $\theta_0$  of a Clayton copula (as in Section 3.1) with the help of a Spearman's rho empirical estimator like (17), but our sample contains missing values.

A way to formalize the presence of missing values is to introduce independent Bernoulli random variables. Let

$$Z_i^{(k)} = \begin{cases} 1 & \text{if } U_i^{(k)} \text{ is not missing,} \\ 0 & \text{if } U_i^{(k)} \text{ is missing,} \end{cases}$$

for  $i = 1, \dots, d$  and  $k = 1, \dots, n$ . Define  $P(Z_i^{(k)} = 1) = b_i$ . If the  $b_i$  are known then an unbiased empirical estimator of the Spearman's rho for the pair  $(U_i, U_j)$  in presence of missing observations is

$$\widehat{\mathcal{D}}_{ij} = \widehat{\rho}_{ij} = \frac{12 \frac{1}{n} \sum_{k=1}^n Z_i^{(k)} Z_j^{(k)} (U_i^{(k)} - \frac{1}{2})(U_j^{(k)} - \frac{1}{2})}{b_i b_j}. \quad (25)$$

Define  $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_{1,2}, \dots, \widehat{\rho}_{d-1,d})$  and  $\rho(\theta) := 12 \int_{[0,1]^2} C_{1,2}(u, v; \theta) du dv - 3$  where  $C_{1,2}$  denotes the bivariate marginal of the underlying copula  $C$  between the variables  $U_1$  and  $U_2$ . Let  $\rho = \rho(\theta_0)$  be the true Spearman's dependence coefficient and  $\rho'$  be the derivative of  $\rho(\cdot)$  at  $\theta_0$ . Following the arguments of Section 3.1, one can obtain

$$\widehat{\theta} = \rho^{-1} \left( \frac{\widehat{\boldsymbol{\rho}}^T \widehat{\mathbf{W}} \mathbf{1}_p}{\mathbf{1}_p^T \widehat{\mathbf{W}} \mathbf{1}_p} \right). \quad (26)$$

By applying, the law of large numbers and the multivariate central limit theorem, it is easily checked that Assumption (A4) holds with

$$\begin{aligned} \Sigma_{i,j;k,l} &= 144 b_k^{-\delta_{i,k} - \delta_{j,k}} b_l^{-\delta_{i,l} - \delta_{j,l}} \mu_{i,j;k,l} - \rho_{i,j} \rho_{k,l}, \\ \mu_{i,j;k,l} &= \mathbb{E} \left[ \left( U_i - \frac{1}{2} \right) \left( U_j - \frac{1}{2} \right) \left( U_k - \frac{1}{2} \right) \left( U_l - \frac{1}{2} \right) \right], \end{aligned} \quad (27)$$

where  $\delta_{i,k}$  denotes the Kronecker symbol, that is,  $\delta_{i,k} = 1$  if  $i = k$  and 0 otherwise. If  $b_i = 1$ , (27) reduces to (18). The asymptotic properties of  $\widehat{\theta}$  can be derived by applying Theorem 1. We have, as  $n \rightarrow \infty$

$$\sqrt{n}(\theta(\widehat{\mathbf{W}}) - \theta_0) \xrightarrow{d} N \left( 0, \frac{\mathbf{1}_p^T \mathbf{W} \Sigma \mathbf{W} \mathbf{1}_p}{(\rho' \mathbf{1}_p^T \mathbf{W} \mathbf{1}_p)^2} \right).$$

By Proposition 1,  $\widehat{\mathbf{W}} = \widehat{\boldsymbol{\Sigma}}^{-1}$  would ensure an optimal asymptotic distribution. Note that the assumptions of Corollary 2 do not hold because the distribution of (25) does not depend only on that of  $(U_i, U_j)$ : it depends also on  $b_i$  and  $b_j$ . We tested the estimators presented in Algorithm 1 by performing a numerical simulation experiment. We generated 500 datasets with 200 observations of 5 variables according to a Clayton copula with parameter  $\theta_0 = 1$ . The  $b_i$  were set to  $0.9 - (i-1)(0.9-0.1)/(d-1)$  for  $i = 1, \dots, d-1$  respectively. For each dataset  $\widehat{\theta}^{(t)}$  was computed for  $t = 0, 1, 2, 3$  and the squared errors  $(\widehat{\theta}^{(t)} - \theta_0)^2$  were derived. The results reported Figure 2 show that the one-step estimator  $\widehat{\theta}^{(1)}$  allows to reduce the errors significantly. Also, considering more steps do not seem to improve further the inference.

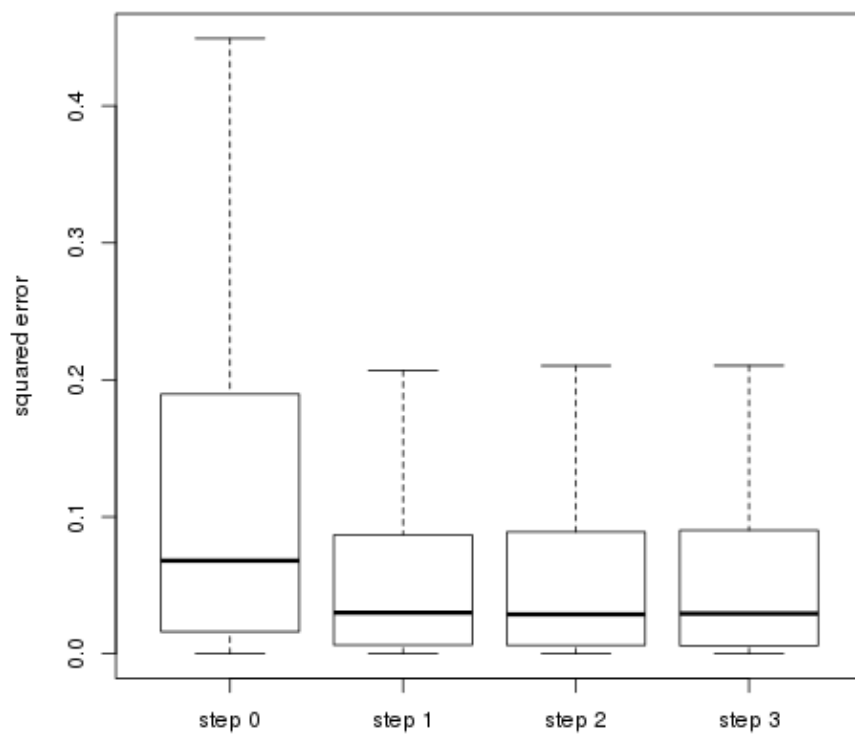


Figure 2: Box plots of the squared errors  $(\hat{\theta}^{(t)} - \theta_0)^2$  for  $t = 0, 1, 2, 3$  (see text for details).

### 3.3 Measuring uncertainty for multivariate return periods in hydrology

In hydrology, it is of interest to quantify the severity and frequency of extreme events. Such potentially dangerous events are underlain by the behavior of a vector  $(U_1, \dots, U_d)$  whose distribution is a copula  $C$ , determined by a parameter vector  $\boldsymbol{\theta}$  in  $\Theta$ . For a certain potentially dangerous event, define the *return period*  $T$  and the *critical level*  $\tilde{p}$  through the relationship

$$T = \frac{1}{1 - K_{\boldsymbol{\theta}}(\tilde{p})}, \quad (28)$$

where  $K_{\boldsymbol{\theta}}(t) = P(C(U_1, \dots, U_d) \leq t)$ ,  $t \in [0, 1]$ .  $K_{\boldsymbol{\theta}}$  is called the Kendall's distribution function associated to  $C$ , see [24]. 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 to this question, it suffices to invert (28) to get  $\tilde{p}$  as a function of  $T$

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

Let  $\boldsymbol{\theta}_0$  denote the true parameter vector and let  $\tilde{p}_T = \tilde{p}_T(\boldsymbol{\theta}_0)$ . The estimation of  $\tilde{p}_T$ , or, in other words, the estimation of  $\boldsymbol{\theta}_0$ , was performed in [6]. The parametric model proposed for  $C$  was the extreme-value copula

$$C(u_1, \dots, u_d) = \left( \prod_{i=1}^d u_i^{1-\theta_i} \right) \min_{i=1, \dots, d} (u_i^{\theta_i}), \quad \theta_i \in [0, 1], i = 1, \dots, d, \quad (29)$$

which allowed to study the dependence between  $d = 3$  sites in Italy (Airole, Merelli and Poggi). The authors chose to base the inference on the Kendall's tau, and, in our work, it corresponds to the framework of Proposition 5. For  $\boldsymbol{\theta}$  in  $[0, 1]^d$ , the Kendall's tau is given by

$$\tau_{ij}(\boldsymbol{\theta}) = \frac{\theta_i \theta_j}{\theta_i + \theta_j - \theta_i \theta_j}, \quad i < j. \quad (30)$$

By inverting (30), one obtains

$$\hat{\theta}_i = \frac{1}{2} \left( 1 + \frac{1}{\hat{\tau}_{ij}} + \frac{1}{\hat{\tau}_{ik}} - \frac{1}{\hat{\tau}_{jk}} \right), \quad (31)$$

where  $i, j, k$  denote the indexes of the three sites and  $\hat{\tau}_{ij}$  is given by (19). Observe that this is the solution of the equation (9), and, under the light of Proposition 3 (since  $p = q = d = 3$ ), we see that this estimator has the smallest asymptotic variance within the class (3). However, in [6], 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  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$  can be established by applying Theorem 1. It suffices to verify that Assumption (A3) holds, which is easily checked from (30). Hence, as  $n \rightarrow \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Xi}), \quad (32)$$

where  $\Xi$  is given by (10) and (20). Now, the derivation of the asymptotic behavior of the critical levels is straightforward. From (32), we get by the delta-method that, as  $n \rightarrow \infty$

$$\sqrt{n} \left( \tilde{p}_T(\hat{\boldsymbol{\theta}}) - \tilde{p}_T \right) \xrightarrow{d} N(0, s_T^2), \quad (33)$$

where  $s_T^2 = \dot{\tilde{p}}_T \Xi \dot{\tilde{p}}_T^T$ ,  $\dot{\tilde{p}}_T$  being the Jacobian of  $\tilde{p}_T(\cdot)$  at the true parameter value. It follows that confidence intervals can be computed from the finite-sample approximation of (33), provided that the sample size is large enough. In [6], 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<sup>1</sup> (Figure 3.3).

The test based on Corollary 3 has no power to detect a wrong model in this situation. Indeed, since  $\mathcal{D}(\hat{\boldsymbol{\theta}}) = \hat{\mathcal{D}}$ , the test statistic is always zero. Other tests can be performed to achieve such a task, see the original paper [6].

When studying extreme events, it is common to have only a limited amount of data. For instance, in [6], only  $n = 34$  (multivariate) observations were available. With such a small sample size, the approximation to a normal law for the distribution of  $\sqrt{n}(\tilde{p}_T(\hat{\boldsymbol{\theta}}) - \tilde{p}_T)$  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 (29) with  $\boldsymbol{\theta}_0 = (0.6, 0.7, 0.2)$ . For the  $m$ -th dataset ( $m = 1, \dots, N$ ), the parameter vector estimate  $\hat{\boldsymbol{\theta}}^{(m)}$  was computed. Let  $s_T(\hat{\boldsymbol{\theta}}^{(m)})$  be the asymptotic covariance matrix in (33) at  $\hat{\boldsymbol{\theta}}^{(m)}$  where  $s_T(\boldsymbol{\theta})$  is regarded as a function of  $\boldsymbol{\theta}$ . The critical levels  $\tilde{p}_T(\hat{\boldsymbol{\theta}}^{(m)})$  together with the 95% confidence bands  $\tilde{p}_T(\hat{\boldsymbol{\theta}}^{(m)}) \pm 1.96 s_T(\hat{\boldsymbol{\theta}}^{(m)})/\sqrt{n}$  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  $s_T^2(\hat{\boldsymbol{\theta}}^{(m)})$ . Therefore, these were dropped from the experiment. The results reported Table 3.3 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 (31) to reduce this vexing effect. For instance, one could consider an estimator based on both the Kendall's tau and the Spearman's rho. The asymptotic distribution of such an estimator could be deduced from Theorem 1 but would require the joint asymptotic distribution of the empirical Kendall's tau and Spearman's rho.

---

<sup>1</sup>The authors thank Fabrizio Durante and Gianfausto Salvadori for sharing their dataset.



$n$	pair $T$	(1,2)			(1,3)			(2,3)		
		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 1: Proportion for the true value  $\tilde{p}_T$  to be included within the 95% confidence intervals. The labels for the pairs (1,2), (1,3), and (2,3) are Airole-Merelli, Airole-Poggi and Merelli-Poggi respectively.

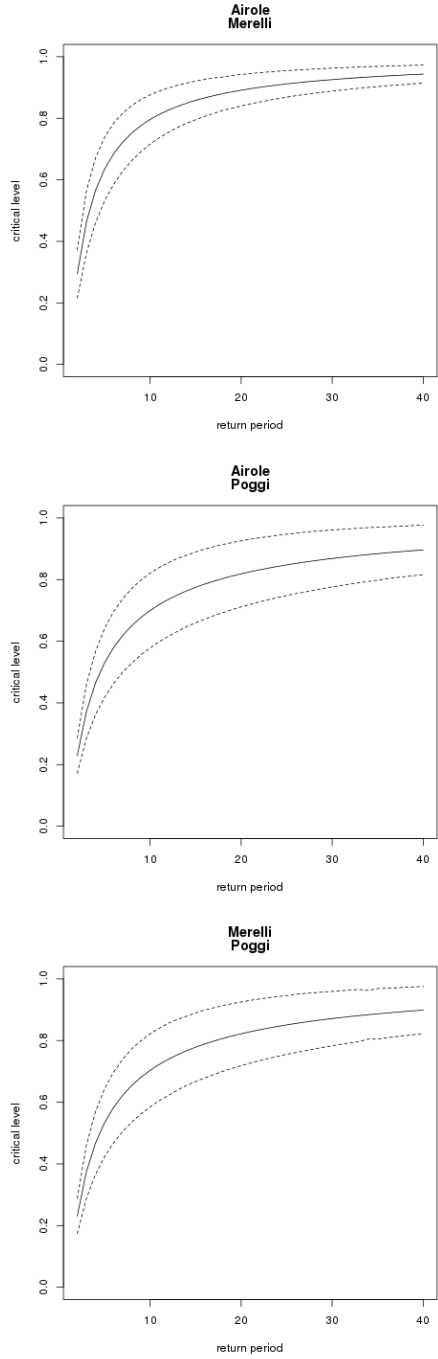


Figure 3: Critical levels  $\tilde{p}_T(\hat{\theta})$  for  $T = 2, \dots, 40$  together with 95% confidence intervals.

## Appendix

### Proof of Theorem 1

In order to prove Theorem 1, we first establish two lemmas. These lemmas, as well as their proofs, are similar to the results in [8]. It will appear that the proof of the theorem is a straightforward application of these lemmas.

Let  $\Theta$  and  $\varepsilon_0$  as in Assumption (A1). Define the vector map

$$\begin{aligned} \varphi : \Theta \subset \mathbb{R}^q &\rightarrow \varphi(\Theta) \subset \mathbb{R}^p \\ \varphi(\boldsymbol{\theta}) &\mapsto (\varphi_1(\boldsymbol{\theta}), \dots, \varphi_p(\boldsymbol{\theta}))^T, \end{aligned} \quad (34)$$

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

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

be a random vector in  $\mathbb{R}^p$  depending on an integer  $n$  and assume that  $\mathbf{Y}_n \xrightarrow{P} \varphi(\boldsymbol{\theta}_0)$  as  $n \rightarrow \infty$ . 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  $n \rightarrow \infty$ . Then  $\widehat{\mathbf{W}} = \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}}$  for some  $p \times p$  matrix  $\widetilde{\mathbf{W}}$ . Denote by  $\hat{\Theta}_n$  the set of all minimizers of the loss function

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= (\mathbf{Y}_n - \varphi(\boldsymbol{\theta}))^T \widehat{\mathbf{W}} (\mathbf{Y}_n - \varphi(\boldsymbol{\theta})) \\ &= \left\| \widetilde{\mathbf{W}} (\mathbf{Y}_n - \varphi(\boldsymbol{\theta})) \right\|^2, \quad \boldsymbol{\theta} \in \Theta, \end{aligned} \quad (35)$$

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  $\ell_n$  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

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

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

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

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  $\dot{\varphi}$  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  $\boldsymbol{\theta}_0$ .

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

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

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  $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$ ,

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

the last inequality holding because, since the second order derivatives of the  $\varphi_i$ '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  $n \rightarrow \infty$ ,

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

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

(ii) Notice that since  $\dot{\varphi}$  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  $\Theta$ ,  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$  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

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

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

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

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  $\overline{B}_\delta(\boldsymbol{\theta}_0) \cap \overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$ , the matrix  $\mathbf{H}_n(\boldsymbol{\theta})$  is positive definite. By (i),  $P(A) \rightarrow 1$  as  $n \rightarrow \infty$ , which concludes the proof.

**Lemma 2.** (i) If  $\varphi$  in (34) is an homeomorphism, then for all  $\varepsilon$  such that  $0 < \varepsilon \leq \varepsilon_0$ , as  $n \rightarrow \infty$ ,

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

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

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

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

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

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

then

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

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

$$\|\widetilde{\mathbf{W}}(\varphi(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta}_0))\| \leq \|\widetilde{\mathbf{W}}(\varphi(\boldsymbol{\theta}) - \mathbf{Y}_n)\| + \|\widetilde{\mathbf{W}}(\mathbf{Y}_n - \varphi(\boldsymbol{\theta}_0))\|$$

implies

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

Therefore

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

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  $P(A_n) \rightarrow 1$ . 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

$$\left\{ \hat{\Theta} \neq \emptyset, \hat{\Theta} \subset \overline{B}_\eta(\boldsymbol{\theta}_0) \text{ and } \mathbf{H}_n(\boldsymbol{\theta}) \text{ is positive definite for all } \boldsymbol{\theta} \text{ in } \overline{B}_\eta(\boldsymbol{\theta}_0) \right\} \quad (36)$$

happens. Let  $\boldsymbol{\theta} \in \overline{B}_\eta(\boldsymbol{\theta}_0)$  and  $\boldsymbol{\theta}^*$  be a vector in  $\hat{\Theta}$ . A Taylor expansion of  $\ell_n$  in (35) at  $\boldsymbol{\theta}^*$  gives

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

where  $\tilde{\boldsymbol{\theta}} = t\boldsymbol{\theta} + (1-t)\boldsymbol{\theta}^*$ ,  $t \in (0, 1)$  and  $\nabla \ell_n$  denotes the gradient of  $\ell_n$ . In view of Lemma 2 (i),  $\boldsymbol{\theta}^*$  is in some open neighborhood of  $\boldsymbol{\theta}_0$  and thus  $\nabla \ell_n(\boldsymbol{\theta}^*) = 0$ . 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  $\ell_n(\boldsymbol{\theta}) > \ell_n(\boldsymbol{\theta}^*)$  for all  $\boldsymbol{\theta}$  in  $\overline{B}_\eta(\boldsymbol{\theta}_0)$ . This implies that the cardinal of  $\hat{\Theta}$  is 1. By Lemma 1 (ii) and Lemma 2 (i), the event (36) has probability tending to 1. Hence  $P[\text{card } \hat{\Theta} = 1] \rightarrow 1$ . 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

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

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  $\ell_n$  in equation (35) 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,

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

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

$$\begin{aligned} P \left[ \sqrt{n} \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \right] &= P \left[ \sqrt{n} \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{card } \hat{\Theta} = 1 \right] \\ &\quad + P \left[ \sqrt{n} \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{card } \hat{\Theta} \neq 1 \right]. \end{aligned} \quad (37)$$

Since the second term in the sum in the right hand side of (37) tends to 0, we have that

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

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

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

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

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

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  $\tilde{\boldsymbol{\theta}} \in \overline{B}_{\delta}(\boldsymbol{\theta}_0)$  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.

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

## Proof of Corollary 1

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

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

By assumption,  $\dot{\mathcal{D}}(\cdot)$  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  $\Theta$ , we have  $\Xi(\boldsymbol{\theta}) = \Xi^{1/2}(\boldsymbol{\theta})\Xi^{1/2}(\boldsymbol{\theta})$  where  $\Xi^{1/2}(\boldsymbol{\theta})$  is also invertible. Therefore, by Theorem 1, as  $n \rightarrow \infty$ ,

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

leading to the desired result.

(ii) By Assumption (A4),

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

as  $n \rightarrow \infty$ . 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  $\mathcal{D}_0 := \mathcal{D}(\boldsymbol{\theta}_0)$  and write

$$\mathcal{D}(\hat{\boldsymbol{\theta}}) - \hat{\mathcal{D}} = \mathcal{D}(\hat{\boldsymbol{\theta}}) - \mathcal{D}_0 + \mathcal{D}_0 - \hat{\mathcal{D}}. \quad (38)$$

A Taylor expansion yields

$$\mathcal{D}(\hat{\boldsymbol{\theta}}) - \mathcal{D}_0 = \tilde{\mathcal{D}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad (39)$$

where  $\tilde{\mathcal{D}} := \dot{\mathcal{D}}(\tilde{\boldsymbol{\theta}})$  with  $\tilde{\boldsymbol{\theta}}$  being a vector between  $\hat{\boldsymbol{\theta}}$  and  $\boldsymbol{\theta}_0$ . Substitute (39) into (38) to get

$$\mathcal{D}(\hat{\boldsymbol{\theta}}) - \hat{\mathcal{D}} = \tilde{\mathcal{D}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{D}_0 - \hat{\mathcal{D}}. \quad (40)$$

From (39), we have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\tilde{\mathcal{D}}^T \tilde{\mathcal{D}})^{-1} \tilde{\mathcal{D}}^T (\mathcal{D}(\hat{\boldsymbol{\theta}}) - \mathcal{D}_0). \quad (41)$$

Substitute (41) into (40) to obtain

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

Since

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

we have

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

Define  $\mathbf{A}(\boldsymbol{\theta}) := \mathbf{I}_p - \dot{\mathcal{D}}(\boldsymbol{\theta}) \left( \dot{\mathcal{D}}(\boldsymbol{\theta})^T \dot{\mathcal{D}}(\boldsymbol{\theta}) \right)^{-1} \dot{\mathcal{D}}(\boldsymbol{\theta})^T$ ,  $\tilde{\mathbf{A}} := \mathbf{A}(\tilde{\boldsymbol{\theta}})$  and  $\hat{\mathbf{A}} := \mathbf{A}(\hat{\boldsymbol{\theta}})$ .

By Assumption (A4) and because  $\mathcal{D}$  is continuously differentiable, as  $n \rightarrow \infty$ ,

$$(\mathbf{I}_p - \tilde{\mathbf{A}}) \sqrt{n} (\mathcal{D}(\hat{\boldsymbol{\theta}}) - \hat{\mathcal{D}}) \xrightarrow{d} N(\mathbf{0}, (\mathbf{I}_p - \mathbf{A}) \Sigma (\mathbf{I}_p - \mathbf{A}))$$



where  $\mathbf{A} := \dot{\mathcal{D}}_0 \left( \dot{\mathcal{D}}_0^T \dot{\mathcal{D}}_0 \right)^{-1} \dot{\mathcal{D}}_0$  and  $\dot{\mathcal{D}}_0 := \mathcal{D}(\theta_0)$ . Since  $\mathbf{I}_p - \mathbf{A}(\theta)$  is idempotent, that is  $(\mathbf{I}_p - \mathbf{A}(\theta))^2 = \mathbf{I}_p - \mathbf{A}(\theta)$ , we have for all  $\theta$  in  $\Theta$

$$[(\mathbf{I}_p - \mathbf{A}(\theta))\Sigma(\theta)(\mathbf{I}_p - \mathbf{A}(\theta))]^{-1/2} = (\mathbf{I}_p - \mathbf{A}(\theta))^{-1}\Sigma^{-1}(\theta)(\mathbf{I}_p - \mathbf{A}(\theta))^{-1}.$$

By continuity  $\Sigma(\hat{\theta}) \xrightarrow{P} \Sigma$ . Since  $\hat{\mathbf{A}} \xrightarrow{P} \mathbf{A}$  as well,

$$(\mathbf{I}_p - \hat{\mathbf{A}})^{-1}\Sigma^{-1}(\hat{\theta})\sqrt{n} \left( \mathcal{D}(\hat{\theta}) - \hat{\mathcal{D}} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p)$$

as  $n \rightarrow \infty$ , yielding the desired result.

### Proof of Proposition 1

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

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

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

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

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

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

which is easily seen to be nonnegative definite.

## Proof of Proposition 2

It suffices to show that  $\mathbf{W}^* = \mathbf{I}_p$  verifies (7). Let us first assume that

$$\Sigma \mathbf{1} \propto \mathbf{1} \quad (42)$$

(this will be shown afterwards for exchangeable copulas). Then (42) is also true with  $\Sigma^{-1}$  in place of  $\Sigma$ . Hence, by substituting (42) into (7), we get  $\mathbf{W}^* \propto \mathbf{I}_p$ . To show (42), recall that  $\Sigma$  is the asymptotic variance matrix of  $\sqrt{n}(\widehat{\mathcal{D}} - \mathcal{D})$ . Keeping in mind the representation (5), we have that  $\Sigma_{i,j;k,l}$  is the limit of  $\text{Cov}(\sqrt{n}(\widehat{\mathcal{D}}_{i,j} - \mathcal{D}_{i,j}), \sqrt{n}(\widehat{\mathcal{D}}_{k,l} - \mathcal{D}_{k,l}))$ . Thus  $\Sigma_{i,j;k,l}$  only depends on the distribution of  $(U_i, U_j, U_k, U_l)$ . But since the underlying copula is exchangeable, at most three distinct distributions can arise: the distribution of  $(U_1^2, U_2, U_3)$  (the pairs  $(U_i, U_j)$  and  $(U_k, U_l)$  share one variable), that of  $(U_1^2, U_2^2)$  (the pairs  $(U_i, U_j)$  and  $(U_k, U_l)$  share two variables), and that of  $(U_1, U_2, U_3, U_4)$  (the pairs  $(U_i, U_j)$  and  $(U_k, U_l)$  share no variables). It follows that all lines of  $\Sigma$  are equal up to permutations of  $\{1, \dots, d\}$ , which implies (42).

## Proof of Proposition 3

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

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

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

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

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

## Proof of Proposition 4

Note that formula (15) is nothing else than the standard Pearson correlation coefficient. We recall how to obtain its asymptotic normality. Observe that

$$\hat{\rho}_{i,j} = \frac{\overline{U_i U_j} - \overline{U_i} \overline{U_j}}{\left( (\overline{U_i^2} - \overline{U_i}^2) (\overline{U_j^2} - \overline{U_j}^2) \right)^{1/2}},$$

where  $\overline{U_i} = \sum_{k=1}^n U_i^{(k)} / n$ ,  $\overline{U_i^2} = \sum_{k=1}^n (U_i^{(k)})^2 / n$  and  $\overline{U_i U_j} = \sum_{k=1}^n U_i^{(k)} U_j^{(k)} / n$ . By the multivariate central limit theorem, the vector

$$\sqrt{n} \left( (\overline{U_i}, \overline{U_j}, \overline{U_i^2}, \overline{U_j^2}, \overline{U_i U_j}) - (E(U_i), E(U_j), E(U_i^2), E(U_j^2), E(U_i U_j)) \right)$$

converges to a normal distribution. The convergence of (15) follows by the delta-method. To obtain the convergence of  $\hat{\rho}_{1,2}, \dots, \hat{\rho}_{d-1,d}$ , one can adapt the above arguments.

## Proof of Proposition 5

The proof of Proposition 5 uses the results of [13] on U-statistics. Recall that  $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(n)}$  is a sample of  $d$ -dimensional vectors. A U-statistics of order 2 is defined by

$$\mathcal{U}(\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(n)}) = \binom{n}{2}^{-1} \sum_{1 \leq \alpha < \beta \leq 2} \Phi(\mathbf{U}^{(\alpha)}, \mathbf{U}^{(\beta)})$$

where the symmetric map  $\Phi : (\mathbb{R}^d)^2 \rightarrow \mathbb{R}$  is called the kernel of  $\mathcal{U}$ . It has been shown in [13] that a multivariate vector of properly centered and scaled U-statistics converge in law to a normal distribution. Before using these results, it is convenient to introduce the following. Define  $\Psi_{ij}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) = \Phi_{ij}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) - \tau_{ij}$  and  $\Psi_{ij,1}(\mathbf{u}^{(1)}) = \mathbb{E}[\Psi_{ij}(\mathbf{u}^{(1)}, \mathbf{U}^{(2)})]$ . To make use of the results in [13], notice that  $\hat{\tau}_{ij}$  is a U-statistic with symmetric kernel

$$\begin{aligned} \Phi_{ij}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) &= \Phi_{ij}\left(\left(u_i^{(1)}, u_j^{(1)}\right), \left(u_i^{(2)}, u_j^{(2)}\right)\right) \\ &= \text{sign}\left(\left(u_i^{(1)} - u_i^{(2)}\right)\left(u_j^{(1)} - u_j^{(2)}\right)\right) \end{aligned}$$

for  $i < j$  and  $\mathbf{u}^{(k)} = \left(u_1^{(k)}, \dots, u_d^{(k)}\right)$ ,  $k = 1, \dots, n$ . By Theorem 7.1 of [13],

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

where  $\boldsymbol{\Sigma}$  writes as in (5) and

$$\boldsymbol{\Sigma}_{i,j;k,l} = 4 \mathbb{E}\left[\Psi_{ij,1}(\mathbf{U}^{(1)}) \Psi_{kl,1}(\mathbf{U}^{(1)})\right]. \quad (43)$$

To simplify the formula (43), observe that, almost surely,

$$\begin{aligned} &\Psi_{ij}(\mathbf{u}^{(1)}, \mathbf{U}^{(2)}) \\ &= \Phi_{ij}(\mathbf{u}^{(1)}, \mathbf{U}^{(2)}) - \tau_{ij} \\ &= \text{sign}\left(\left(u_i^{(1)} - U_i^{(2)}\right)\left(u_j^{(1)} - U_j^{(2)}\right)\right) - \tau_{ij} \\ &= \begin{cases} 1 - \tau_{ij} & \text{if } U_i^{(2)} < u_i^{(1)} \text{ and } U_j^{(2)} < u_j^{(1)} \text{ or } U_i^{(2)} > u_i^{(1)} \text{ and } U_j^{(2)} > u_j^{(1)} \\ -1 - \tau_{ij} & \text{otherwise,} \end{cases} \end{aligned}$$

and

$$\begin{aligned} &P\left[U_i^{(2)} < u_i^{(1)} \text{ and } U_j^{(2)} < u_j^{(1)} \text{ or } U_i^{(2)} > u_i^{(1)} \text{ and } U_j^{(2)} > u_j^{(1)}\right] \\ &= 2C_{ij}(u_i^{(1)}, u_j^{(1)}) + 1 - u_i^{(1)} - u_j^{(1)}. \end{aligned}$$

Therefore,

$$\Psi_{ij,1}(\mathbf{u}^{(1)}) = \mathbb{E}\left[\Psi_{ij}(\mathbf{u}^{(1)}, \mathbf{U}^{(2)})\right] = 4C_{ij}(u_i^{(1)}, u_j^{(1)}) + 1 - \tau_{ij} - 2u_i^{(1)} - 2u_j^{(1)}. \quad (44)$$

By substituting (44) into (43), one obtains (20). An application of Lemma 2 with  $\mathbf{Y}_n = \hat{\boldsymbol{\tau}}$  and  $\boldsymbol{\varphi}(\boldsymbol{\theta}) = \boldsymbol{\tau}(\boldsymbol{\theta})$  concludes the proof of (ii).

## Proof of Proposition 6

It has been shown in [9] that (recall that  $C$  here is an extreme-value copula)

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n \max(U_i^{(k)}, U_j^{(k)}) &\xrightarrow{P} \frac{2 - \lambda_{ij}}{3 - \lambda_{ij}}, \text{ and,} \\ \sqrt{n} \left( \frac{1}{n} \sum_{k=1}^n \max(U_i^{(k)}, U_j^{(k)}) - \frac{2 - \lambda_{ij}}{3 - \lambda_{ij}} \right) &\xrightarrow{d} N(0, \text{Var}(\max(U_i, U_j))). \end{aligned} \quad (45)$$

By the multivariate central limit theorem and the multivariate delta method with the continuous map  $\mathbf{g} : [0, 1]^p \rightarrow (-\infty, 2]^p$  such that  $\mathbf{g}(x_1, \dots, x_p) = (g(x_1), \dots, g(x_p))^T$  with  $g(x) = 3 - 1/(1 - x)$ , we have

$$\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

where  $\boldsymbol{\Sigma} = \mathbf{G}\boldsymbol{\Omega}\mathbf{G}^T$ ,  $\mathbf{G}$  is the Jacobian matrix of  $\mathbf{g}$  at  $\left(\frac{2-\lambda_{12}}{3-\lambda_{12}}, \dots, \frac{2-\lambda_{d-1,d}}{3-\lambda_{d-1,d}}\right)^T$ , that is,  $\mathbf{G} = \text{diag}(-(3 - \lambda_{12})^2, \dots, -(3 - \lambda_{d-1,d})^2)$  and  $\boldsymbol{\Omega}$  is the variance-covariance matrix for the limit distribution of the  $p$ -dimensional vector whose components are the left members in (45). Here  $\text{diag}(x_1, \dots, x_p)$  denotes the matrix with  $(x_1, \dots, x_p)$  on the diagonal and with zero's elsewhere. It is easily seen that  $\boldsymbol{\Sigma}$  writes as in (5) and (22).

## References

- [1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.
- [2] D. Berg. Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701, 2009.
- [3] A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186, May 2012.
- [4] S. Demarta and A. J. McNeil. The t copula and related copulas. *International statistical review*, 73(1):111–129, 2005.
- [5] F. Durante. A new class of symmetric bivariate copulas. *Nonparametric Statistics*, 18(7-8):499–510, 2006.
- [6] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.
- [7] D. Eddelbuettel and R. François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.
- [8] 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.
- [9] M. Ferreira. Nonparametric estimation of the tail-dependence coefficient. *REVSTAT-Statistical Journal*, 11(1):1–16, 2013.

- [10] 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.
- [11] G. Gudendorf and J. Segers. Extreme-value copulas. In *Copula Theory and Its Applications*, page 127–145. Springer, 2010.
- [12] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.
- [13] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.
- [14] M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula: Multivariate Dependence with Copulas*, 2014. R package version 0.999-8.
- [15] M. Hofert, M. Mächler, and A. J. McNeil. Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1):25–63, 2012.
- [16] M. Hofert and M. Maechler. *simsalapar: Tools for Simulation Studies in Parallel with R*, 2014. R package version 1.0-2.
- [17] M. Hofert, M. Mächler, and A. J. McNeil. Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133–150, 2012.
- [18] M. Hofert and F. Vriens. Sibuya copulas. *Journal of Multivariate Analysis*, 114:318–337, 2013.
- [19] H. Joe. *Multivariate models and dependence concepts*. Chapman & Hall/CRC, Boca Raton, FL, 2001.
- [20] P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101, 2013.
- [21] J.-F. Mai and M. Scherer. Lévy-frailty copulas. *Journal of Multivariate Analysis*, 100(7):1567–1585, August 2009.
- [22] A. J. Mc Neil. *Quantitative risk management: concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton, N.J, 2005.
- [23] R. B. Nelsen. *An introduction to copulas*. Springer, New York, 2006.
- [24] 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.
- [25] 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.

- [26] Gilbert P. and Varadhan R. *numDeriv: Accurate Numerical Derivatives*, 2014. R package version 2012.9-1.
- [27] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [28] G. Salvadori and C. De Michele. Estimating strategies for multiparameter multivariate extreme value copulas. *Hydrology and Earth System Sciences*, 15(1):141–150, 2011.
- [29] C. Savu and M. Tiede. Hierarchies of archimedean copulas. *Quantitative Finance*, 10(3):295–304, 2010.
- [30] R.L. Smith. Max-stable processes and spatial extremes. *Unpublished manuscript*, 1990.