



## Single-index copulas

Jean-David Fermanian<sup>a,\*</sup>, Olivier Lopez<sup>b</sup>

<sup>a</sup> CREST-ENSAE, 5, avenue Henry-le-Chatelier, 91764 Palaiseau cedex, France

<sup>b</sup> Université Pierre-et-Marie-Curie Paris 6, 4 place Jussieu, 75005 Paris, France



### ARTICLE INFO

#### Article history:

Received 16 May 2017

Available online 5 December 2017

#### AMS subject classifications:

62F12

62G05

#### Keywords:

Conditional copulas

Conditional Kendall's tau

Kernel smoothing

Single-index models

### ABSTRACT

We introduce so-called single-index copulas. They are semi-parametric conditional copulas whose parameter is an unknown link function of a univariate index only. We propose estimates of this link function and of the finite-dimensional unknown parameter. The asymptotic properties of the latter estimates are stated. Thanks to some properties of conditional Kendall's tau, we illustrate our technical conditions with several usual copula families.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Since Sklar's theorem [42], copula modeling has emerged as a very active field of theoretical and applied research. Applications in finance, insurance, biology, medicine, hydrology, etc., are legions. The origin of this success is the ability to split specification/inference/testing of a (complex) multivariate model into two separate simpler problems: the management of marginal distributions on one hand, and the modeling of the dependence structure (copula) on the other. See, e.g., the books of Joe [23] or Nelsen [29] for an introduction to the field.

In practice, multivariate models often involve explanatory variables (also called covariates), particularly in econometrics and financial risk management. To study the effect of these covariates on the underlying copulas, we immediately need the concept of conditional copula [32]. Conditional copulas are a natural way of linking conditional marginal distributions to get a multivariate conditional law and they have been applied extensively [33,34]. Recently, the rise of vine models [1] has extended the scope and the importance of conditional copulas.

Until now, most conditional copula models were parametric. They might specify, say, a functional link between the copula parameters and an index  $\beta^\top \mathbf{z}$ ,  $\mathbf{z}$  being the underlying vector of covariates; see, e.g., [5,24,32,37]. A fully nonparametric approach has also been proposed in [15,17], which relies on kernel smoothing, local polynomials or other functional estimation tools. As the number of covariates increases, however, such methods suffer from the well-known curse of dimensionality, and they become unfeasible in practice.

In this paper, we propose an intermediate solution through a single-index assumption on the underlying copula parameter. Therefore, only a finite-dimensional parameter and a univariate link function must be estimated, avoiding the curse of dimensionality. Note that Acar et al. [3,4] and Abegaz et al. [2] have proposed another alternative through local linear approximations of the link function between covariates and copula parameters. Nonetheless, the latter approach is based on a linearization (thus approximative) procedure and the number of unknown parameters rises very quickly with

\* Corresponding author.

E-mail addresses: [jean-david.fermanian@ensae.fr](mailto:jean-david.fermanian@ensae.fr) (J.-D. Fermanian), [olivier.lopez0@upmc.fr](mailto:olivier.lopez0@upmc.fr) (O. Lopez).

the dimension of  $\mathbf{z}$ . Moreover, Sabeti et al. [38] have recently introduced additive models for copula, and have adapted the Bayesian-type estimation procedure proposed by Craiu and Sabeti [6]. Inference for single-index copula models has been recently studied in [27] in a Bayesian framework.

### 1.1. The framework of single-index dependence functions

To fix ideas and notation, let us consider an iid sample of observations  $(\mathbf{X}_1, \mathbf{Z}_1), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$  in  $\mathbb{R}^d \times \mathbb{R}^p$  drawn from the law of  $(\mathbf{X}, \mathbf{Z})$ . The vector  $\mathbf{X}$  represents the endogenous vector, and  $\mathbf{Z}$  is the vector of covariates. We are interested in the evaluation of the law of  $\mathbf{X}$  conditional on  $\mathbf{Z} = \mathbf{z}$ , for arbitrary vectors  $\mathbf{z}$ . This conditional CDF is denoted  $F(\cdot|\mathbf{z})$ . For each  $k \in \{1, \dots, d\}$ , the (marginal) law of  $X_k$  given  $\mathbf{Z} = \mathbf{z}$  is denoted  $F_k(\cdot|\mathbf{z})$ ; these marginal distributions are assumed to be continuous, for convenience. We introduce the unobserved random vector  $\mathbf{U}_\mathbf{z} = (U_{1,\mathbf{z}}, \dots, U_{d,\mathbf{z}})$ , where  $U_{k,\mathbf{z}} = F_k(X_k|\mathbf{z})$  for each  $k \in \{1, \dots, d\}$ . To simplify notation, and when there is no ambiguity,  $\mathbf{U}_\mathbf{z}$  will often be denoted  $\mathbf{U}$ . By definition, the law of  $\mathbf{U}_\mathbf{z}$  knowing  $\mathbf{Z} = \mathbf{z}$  is the conditional copula of  $\mathbf{X}$  knowing  $\mathbf{Z} = \mathbf{z}$ , denoted  $C(\cdot|\mathbf{z})$ .

First, we consider a parametric framework. A natural model specification would be to assume that, for any  $\mathbf{u} \in [0, 1]^d$  and any  $\mathbf{z} \in \mathbb{R}^p$ ,  $C(\mathbf{u}|\mathbf{z}) = C_{\theta(\mathbf{z})}(\mathbf{u})$ , where  $\theta : \mathbb{R}^p \rightarrow \mathbb{R}^q$  maps the vector of covariates to the (true) parameter of the conditional copula knowing  $\mathbf{Z} = \mathbf{z}$ , and  $\mathcal{C} = \{C_\theta : \theta \in \Theta \subset \mathbb{R}^q\}$  denotes a parametric family of copulas. The copula density of  $C_\theta$  is assumed to exist and is denoted  $c_\theta$ . To simplify, this density is assumed to be continuous for every  $\theta \in \Theta$ , and  $\Theta$  is taken to be a compact subset.

Second, since the single-index assumption is related to the dependence function among the components of  $\mathbf{X}$ , given the covariates, this means there exists an unknown function  $\psi$  such that

$$\theta(\mathbf{z}) = \psi(\beta_0, \beta_0^\top \mathbf{z}), \quad (1)$$

where the true parameter  $\beta_0 \in \mathcal{B}$ , a compact subset in  $\mathbb{R}^m$ . The single-index relation (1) is assumed to be fulfilled in the whole paper. To identify the parameter  $\beta_0$ , let us assume that the first component of  $\beta_0$ , i.e.,  $\beta_{0,1}$ , is equal to 1, and the estimate of  $\beta_{0,1}$  is always 1, obviously. Note that it is necessary to indicate the dependency of  $\theta(\mathbf{z})$  on  $\beta_0$  explicitly in (1). Indeed, given  $\mathbf{Z} = \mathbf{z}$ , we need to know the parameter value to be able to compute the index, and then to evaluate the underlying conditional copula. Then, under Assumption (1),  $C(\cdot|\mathbf{z})$  depends on  $(\beta, \beta^\top \mathbf{z})$  if the underlying parameter is assumed to be  $\beta$ . This function will equivalently be denoted  $C_\beta(\cdot|\beta^\top \mathbf{z})$ .

**Example 1.** Assume that the conditional copula of  $\mathbf{X}$  given  $\mathbf{Z} = \mathbf{z}$  is a  $d$ -dimensional Gaussian copula  $C_\Sigma$ . Knowing the true parameter value is  $\beta_0$ , its associated correlation matrix is

$$\Sigma = [\mathbf{1}(i=j) + \mathbf{1}(i \neq j)h(\beta_0, \beta_0^\top \mathbf{z})]_{1 \leq i, j \leq d},$$

for some unknown function  $h : \mathbb{R}^{m+1} \rightarrow (-1/(d-1), 1)$ .

We stress that Assumption (1) does not mean that  $C(\cdot|\mathbf{z})$ , the conditional copula of  $\mathbf{X}$  knowing  $\mathbf{Z} = \mathbf{z}$ , is equal to the conditional copula of  $\mathbf{X}$  knowing  $\beta_0^\top \mathbf{Z} = \beta_0^\top \mathbf{z}$ , denoted  $\tilde{C}(\cdot|\beta_0^\top \mathbf{z})$ . Indeed, in the former case, the relevant margins are the CDFs  $F_1(\cdot|\mathbf{z}), \dots, F_d(\cdot|\mathbf{z})$  and in the latter case, we need to consider the CDFs  $\tilde{F}_k(\cdot|\beta_0^\top \mathbf{z}) : x_k \mapsto \Pr(X_k \leq x_k | \beta_0^\top \mathbf{z})$ . To avoid any confusion, let us denote  $\tilde{\mathbf{U}}_\beta = (\tilde{F}_1(X_1|\beta^\top \mathbf{Z}), \dots, \tilde{F}_d(X_d|\beta^\top \mathbf{Z}))$ , and  $\tilde{C}(\cdot|\beta^\top \mathbf{Z} = y)$  is the copula of  $\tilde{\mathbf{U}}_\beta$  knowing  $\beta^\top \mathbf{Z} = y$ . The conditional copulas  $C(\cdot|\mathbf{z})$  and  $\tilde{C}(\cdot|\beta_0^\top \mathbf{z})$  are identical only when  $\mathbf{Z}$  yields the same information as  $\beta_0^\top \mathbf{Z}$  to explain every margin  $X_k$ , i.e., when  $F_k(\cdot|\mathbf{z}) = \tilde{F}_k(\cdot|\beta_0^\top \mathbf{z})$  a.e. for every  $k$  [15].

Our single-index copula assumption (1) is relevant theoretically and empirically. Indeed, in general, covariates have a different influence on the conditional margins and on the underlying conditional copula. See [4] for a discussion, and empirical illustrations. Mainly for practical reasons, some authors assume that the conditional copula of  $\mathbf{X}$  given  $\mathbf{Z} = \mathbf{z}$  does not depend on  $\mathbf{z}$ . This particular case is the so-called “simplifying assumption” (see, e.g., [9]), commonly used in vine models.

### 1.2. The M-estimate criterion

Single-index models are well-known in the world of semiparametric statistics. The theory of M-estimators was initiated by the seminal papers of Klein and Spady [26] in the case of the so-called binary response model, and Ichimura [21] for the general single-index regression model. Sherman [41] and Delecroix and Hristache [7] extended this approach. Härdle et al. [18] and Delecroix et al. [8] discussed the choice of the bandwidth for the nonparametric estimation of the link function. As an alternative, the so-called average derivative method was developed in parallel by Stoker [43], Powell et al. [35], Härdle and Stoker [19], among others.

In this paper, we rely on M-estimators of single-index models, but related to the parameter of the underlying copula only. If we were able to observe a sample of the random vector  $\mathbf{U}$ , say  $\mathbf{U}_1, \dots, \mathbf{U}_n$ , then our “naive” estimator of  $\beta_0$  could be

$$\hat{\beta}_{naive} = \arg \max_{\beta \in \mathcal{B}} \sum_{i=1}^n \ln c_{\hat{\psi}(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i),$$

for some function  $\hat{\psi}$  that estimates  $\psi$  consistently.

Since we do not observe realizations of  $\mathbf{U}$ , we have to replace the unknown vectors  $\mathbf{U}_i$  by some estimates  $\hat{\mathbf{U}}_i$ , given  $\mathbf{Z}_i$ , providing a so-called pseudo-sample  $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_n$ . Then, a natural idea is to define our estimator by

$$\hat{\beta} = \arg \max_{\beta \in \mathcal{B}} \sum_{i=1}^n \hat{\omega}_{i,n} \ln c_{\hat{\psi}(\beta, \beta^\top \mathbf{z}_i)}(\hat{\mathbf{U}}_i), \tag{2}$$

for some sequence of trimming functions  $\hat{\omega}_{i,n}$ . Typically, they are of the type  $\hat{\omega}_{i,n} = \mathbf{1}(\hat{\mathbf{U}}_i \in \mathcal{E}_n, \mathbf{Z}_i \in \mathcal{Z})$ , for some non decreasing sequence of subsets  $\mathcal{E}_n$  in  $[0, 1]^d$ , and some  $\mathcal{Z} \subset \mathbb{R}^p$ . Such trimming functions are usually necessary to approximate  $\mathbf{U}_i$  by  $\hat{\mathbf{U}}_i$ , uniformly for all  $i \in \{1, \dots, n\}$ . For technical reasons (see the remark below), we choose strictly increasing trimmings on the  $\mathbf{U}$ -side such that  $\cup_n \mathcal{E}_n = (0, 1)^d$ . This choice makes it necessary to control explicitly the behavior of  $\mathbf{U}$  close to the boundary of  $[0, 1]^d$ . This pretty delicate task requires several regularity assumptions but the challenge has already been met in the literature; see, e.g., [44]. Moreover, we set a fixed trimming for  $\mathcal{Z}$  (i.e.,  $\mathcal{Z} \subset \mathbb{R}^p$  strictly), because we need uniformity with respect to the  $\mathbf{Z}_i$ , too. This will not create any bias, because the law of the  $\mathbf{U}$  knowing  $\mathbf{Z} \in \mathcal{Z}$  is just  $c_{\psi(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}) \mathbf{1}(\mathbf{z} \in \mathcal{Z}) / \Pr(\mathbf{Z} \in \mathcal{Z})$ . Thus, this law depends on the true parameter  $\beta_0$ . See [Assumption 1](#).

**Remark 1.** Actually, fixed trimming functions for  $\hat{\mathbf{U}}_i$  could be chosen instead, i.e.,  $\mathcal{E}_n = \mathcal{E} \subset [a, 1 - a]^d$  for some  $a > 0$  and every  $n$ . They would induce consistent estimates without having to impose regularity conditions on the copula density close to the frontier of  $[0, 1]^d$ . But the asymptotic behavior of  $\hat{\beta}$  would be more complex. Typically, it would be asymptotically normal, but after removing an annoying bias that cannot be evaluated easily. Moreover, apart from a small loss of efficiency, this would preclude modeling the tail dependence behavior, a feature that is important in many applications. That is why we have chosen  $\hat{\beta}$ , as defined by (2).

The consistency and the asymptotic normality of our estimate  $\hat{\beta}$  are established in Sections 2 and 3, respectively. To be able to check some of our technical assumptions, we provide some useful results about conditional Kendall's tau in [Appendix C](#). The paper concludes with a short simulation study reported in Section 4.

## 2. Consistency of the estimators

### 2.1. The convergence of single-index estimators

The first assumption is related to the trimming functions.

**Assumption 1.** Let us set  $\mathcal{E}_n = [v_n, 1 - v_n]^d$  for some positive sequence  $(v_n, v_n \in (0, 1/2), v_n \rightarrow 0$ . Moreover,  $\mathcal{Z}$  is a compact subset in  $\mathbb{R}^p$ . The trimming functions are  $\omega_n : [0, 1]^d \times \mathbb{R}^p \rightarrow [0, 1], (\mathbf{u}, \mathbf{z}) \mapsto \mathbf{1}(\mathbf{u} \in \mathcal{E}_n, \mathbf{z} \in \mathcal{Z})$ .

We set  $\hat{\omega}_{i,n} = \omega_n(\hat{\mathbf{U}}_i, \mathbf{Z}_i)$ . For the sake of completeness, we introduce  $\omega_{i,n} = \omega_n(\mathbf{U}_i, \mathbf{Z}_i)$ , the trimming function when  $\mathbf{U}_i$  is known, and  $\omega_i = \omega_{i,\infty} = \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z})$ . Typically,  $\mathcal{Z}$  is chosen so that the density of  $\mathbf{Z}$  with respect to the Lebesgue measure exists and is bounded away from zero, i.e.,  $\inf_{\mathbf{z} \in \mathcal{Z}} f_{\mathbf{Z}}(\mathbf{z}) \geq f_0 > 0$ . This will be assumed hereafter, even if this is not mandatory at this stage.

**Assumption 2.** The parameter  $\beta_0$  is identifiable, i.e., two different parameters induce two different laws of  $\mathbf{U}_{\mathbf{Z}}$ , knowing  $\mathbf{Z} \in \mathcal{Z}$ . The function  $M : \mathcal{B} \rightarrow \mathbb{R}, \beta \mapsto E\{\ln c_{\psi(\beta, \beta^\top \mathbf{z})}(\mathbf{U}_{\mathbf{Z}}) | \mathbf{Z} \in \mathcal{Z}\}$  is continuous and uniquely maximized at  $\beta = \beta_0$ . There exist a measurable function  $h$  and  $\alpha > 1$  such that, for every  $\mathbf{z} \in \mathcal{Z}$ ,

$$\sup_{\beta \in \mathcal{B}} |\ln c_{\psi(\beta, \beta^\top \mathbf{z})}(\mathbf{U}_{\mathbf{Z}})| \leq h(\mathbf{U}_{\mathbf{Z}}, \mathbf{z}), \quad \text{with } E\{h^\alpha(\mathbf{U}_{\mathbf{Z}}, \mathbf{Z}) \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty. \tag{3}$$

The latter assumption is usual for maximum likelihood estimation (MLE) purposes. In particular, the identifiability of  $\beta$  implies that the particular case of a constant  $\psi$  function is not possible in our framework. In other words,  $\theta(\mathbf{z})$  is varying with  $\mathbf{z}$  by assumption.

The limiting objective function (when  $n \rightarrow \infty$ ) is here

$$M(\beta) = E\{\ln c_{\psi(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i) | \mathbf{Z}_i \in \mathcal{Z}\},$$

which is maximized at  $\beta = \beta_0$ . Obviously, all expectations  $E(\cdot)$  have to be understood with respect to the true law of  $(\mathbf{X}, \mathbf{Z})$  that depends on  $\beta_0$  only. When the true parameter is  $\beta$ , we denote expectations by  $E_\beta(\cdot)$ . Note that, due to our trimming functions, we are dealing with a M-estimator of  $\beta$  instead of a usual MLE formally, at the cost of a (small) loss of efficiency. The next assumption imposes the uniform weak (resp. strong) consistency of  $\hat{\psi}$  (resp.  $\hat{\mathbf{U}}_i$ ), our chosen estimator of  $\psi$  (resp.  $\mathbf{U}_i$ ).

### Assumption 3.

$$\sup_{\mathbf{z} \in \mathcal{Z}} \sup_{\beta \in \mathcal{B}} |\hat{\psi}(\beta, \beta^\top \mathbf{z}) - \psi(\beta, \beta^\top \mathbf{z})| = o_p(1). \tag{4}$$

Moreover, the pseudo-observation  $\hat{U}_{i,k}$  belongs to  $(0, 1)$  for all  $k \in \{1, \dots, d\}$  and  $i \in \{1, \dots, n\}$ , and there exists a deterministic sequence  $(\delta_n)$ ,  $\delta_n = o(v_n)$ , such that

$$\limsup_{n \rightarrow \infty} \max_{i \in \{1, \dots, n\}} |\hat{U}_i - \mathbf{U}_i| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) / \delta_n \leq 1 \text{ a.e.} \tag{5}$$

In particular, (5) implies that  $\max_{i \in \{1, \dots, n\}} |\hat{U}_i - \mathbf{U}_i| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) = O_P(\delta_n)$ . These assumptions have to be checked for any particular single-index model and for any particular estimate of the marginal CDFs.

Now, we recall the definition of reproducing U-shaped functions, as introduced in Tsukahara [44]. This concept is usual for the semi-parametric inference of copula models. Here, it will be necessary to obtain the consistency of  $\hat{\beta}$  (Theorem 1).

**Definition 1.**

- (a) A function  $f : (0, 1) \rightarrow (0, \infty)$  is called U-shaped if it is symmetric about  $1/2$  and decreasing on  $(0, 1/2]$ .
- (b) For  $\beta \in (0, 1)$  and a U-shaped function  $r$ , define

$$r_\beta(u) = \begin{cases} r(\beta u) & \text{if } 0 < u \leq 1/2, \\ r\{1 - \beta(1 - u)\} & \text{if } 1/2 < u \leq 1. \end{cases}$$

If, for every  $\beta > 0$  in a neighborhood of 0, there exists a constant  $A_\beta$  such that  $r_\beta < A_\beta r$  on  $(0, 1)$ , then  $r$  is called a reproducing U-shaped function.

- (c) We denote by  $\mathcal{R}$  the set of univariate reproducing U-shaped functions. The set  $\mathcal{R}_d$  is the set of functions  $r : (0, 1)^d \rightarrow \mathbb{R}^+$ ,  $r(\mathbf{u}) = \prod_{k=1}^d r_k(u_k)$ , and  $r_k \in \mathcal{R}$  for every  $k$ . Moreover,  $r_\beta(\mathbf{u}) = \prod_{k=1}^d r_{k,\beta}(u_k)$ .

Typically, the usual functions in  $\mathcal{R}$  are of the type  $r(u) = C_r u^{-a}(1 - u)^{-a}$ , for some positive constants  $a$  and  $C_r$ .

**Assumption 4.** There exist functions  $r, \tilde{r}_1, \dots, \tilde{r}_d$  in  $\mathcal{R}_d$  such that, for every  $\mathbf{u} \in (0, 1)^d$ ,

$$\sup_{\theta \in \Theta} |\nabla_\theta \ln c_\theta(\mathbf{u})| \leq r(\mathbf{u}), \quad E\{r(\mathbf{U}_Z) \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty,$$

$$\forall_{k \in \{1, \dots, d\}} \sup_{\theta \in \Theta} |\partial_{u_k} \ln c_\theta(\mathbf{u})| \leq \tilde{r}_k(\mathbf{u}), \quad \max_{k \in \{1, \dots, d\}} E\{U_k(1 - U_k)\tilde{r}_k(\mathbf{U}_Z) \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty.$$

The latter conditions on moments are easily satisfied for most copula models. They are close to those of Assumption (A.1) in [44]. The following result is proved in Appendix B.1.

**Theorem 1.** Under Assumptions 1–4, the estimator  $\hat{\beta}$  given by (2) tends to  $\beta_0$  in probability when  $n \rightarrow \infty$ .

Until now, we have not specified how we estimate the link function  $\psi$  and the pseudo-observations  $\hat{U}_i$ . This will be the subject of the next two subsections.

2.2. Estimation of the link function  $\psi$

For inference purposes, we need a relationship between the link function  $\psi$  and some quantities that can be estimated empirically. Typically, there are two possibilities in practice: for any  $(\beta, \mathbf{z}) \in \mathbb{R}^m \times \mathcal{Z}$  and for some known “explicit” functional  $\Psi$ ,

$$\psi(\beta, \beta^\top \mathbf{z}) = \Psi\{C_\beta(\cdot | \beta^\top \mathbf{z})\}, \tag{6}$$

$$\psi(\beta, \beta^\top \mathbf{z}) = \Psi\{H_\beta(\cdot | \beta^\top \mathbf{z})\}, \tag{7}$$

where  $H_\beta(\cdot | y)$  denotes the CDF of  $(\mathbf{X}, \mathbf{Z})$  conditional on  $\beta^\top \mathbf{Z} = y$  and given  $\beta$ . These relations define two classes of functions  $\psi$ . In this paper, we assume that  $\psi$  belongs to one of the two latter classes of functions, called (A1) and (A2). Note that (A1) is a subset of (A2), due to Sklar’s theorem. Both situations are distinguished for pedagogical reasons (see the discussion about Kendall’s tau below).

In numerous practical situations, Assumptions (6) and (7) are simply moment-like conditions, as in the GMM methodology: there is a map  $g : \mathbb{R}^{\tilde{m}} \rightarrow \mathbb{R}^q$ , with  $\tilde{m} \geq m$ , such that

$$\theta(\mathbf{z}) = g\{m_1(\beta_0, \beta_0^\top \mathbf{z}), \dots, m_{\tilde{m}}(\beta_0, \beta_0^\top \mathbf{z})\},$$

where  $m_1(\beta, y), \dots, m_{\tilde{m}}(\beta, y) \in \mathbb{R}$  are “moment” relations based on the underlying distributions. In case (6), these moment relations are directly linked to conditional copulas, viz.

$$\begin{aligned} m_k(\beta, y) &= E_\beta\{\chi_k(\mathbf{U}_Z, \beta^\top \mathbf{Z}) | \beta^\top \mathbf{Z} = y\} = E_\beta[E_\beta\{\chi_k(\mathbf{U}_Z, y) | \mathbf{Z}\} | \beta^\top \mathbf{Z} = y] \\ &= E_\beta\left\{\int \chi_k(\mathbf{u}, y) C(d\mathbf{u} | \mathbf{Z}) | \beta^\top \mathbf{Z} = y\right\} = \int \chi_k(\mathbf{u}, y) C_\beta(d\mathbf{u} | \beta^\top \mathbf{Z} = y), \end{aligned} \tag{8}$$

for some known functions  $\chi_1, \dots, \chi_{\tilde{m}}$ .

In case (7), there exist some “moments”  $m_k(\beta, y) \in \mathbb{R}$  based on the underlying distribution of  $(\mathbf{X}, \mathbf{Z})$  given  $\beta^\top \mathbf{Z} = y$ , viz.

$$m_k(\beta, y) = E_\beta \{ \chi_k(\mathbf{X}, \mathbf{Z}) | \beta^\top \mathbf{Z} = y \} = \int \chi_k(\mathbf{x}, \mathbf{z}) H_\beta(d\mathbf{x}, d\mathbf{z} | \beta^\top \mathbf{Z} = y).$$

During the estimation procedure, the latter moments  $m_k$ , or more generally the CDFs  $C_\beta(\cdot | \beta^\top \mathbf{z})$  and  $H_\beta(\cdot | \beta^\top \mathbf{z})$  in (6) and (7), will be replaced by some empirical counterparts. The formalism of (A2) behaves nicer than (A1), because it is simpler to work with the observations  $(\mathbf{X}_i, \mathbf{Z}_i)$  directly rather than with vectors  $\mathbf{U}_i$  (i.e., some iid realizations of the random vector  $\mathbf{U}_Z$ ). Indeed, since  $\mathbf{U}_Z$  cannot be observed, the latter quantities  $\mathbf{U}_i$  have to be estimated too, adding another level of complexity.

**Example 2 (Spearman’s Rho).** A natural candidate for a moment is given by  $m_k(\beta, \beta^\top \mathbf{z}) = \rho(\beta, \beta^\top \mathbf{z})$ , a multivariate extension of the usual Spearman’s rho, defined by

$$\rho(\beta, y) = \int \left\{ C_\beta(\mathbf{u} | \beta^\top \mathbf{Z} = y) - \prod_{j=1}^d u_j \right\} d\mathbf{u}.$$

Through a  $d$ -dimensional integration by parts, this moment can be verified to be of the type (8). Therefore, we work in (A1). Other definitions of Spearman’s rho are possible with an arbitrary dimension  $d$ : see, e.g., [39]. Note that when  $d = 2$ ,  $\rho(\beta, y)$  is simply the correlation between  $F_1(X_1 | \mathbf{Z})$  and  $F_2(X_2 | \mathbf{Z})$  given  $\beta^\top \mathbf{Z}$ . Therefore, it can be estimated relatively easily, at least when the dimension of  $\mathbf{Z}$  is “reasonable”.

**Example 3 (Kendall’s Tau).** To fix ideas, assume  $d = 2$ . The Kendall’s tau of  $\mathbf{X}$  conditional on  $\mathbf{Z} = \mathbf{z}$  appears in several papers in the literature; see, e.g., [47]. Under the single-index assumption, it is written here

$$\tau(\beta, \beta^\top \mathbf{z}) = -1 + 4 \int C(\mathbf{u} | \mathbf{z}) C(d\mathbf{u} | \mathbf{z}) = -1 + 4 \int C_\beta(\mathbf{u} | \beta^\top \mathbf{z}) C_\beta(d\mathbf{u} | \beta^\top \mathbf{z}). \tag{9}$$

Then, managing Kendall’s tau, we work in (A1) usually. The parameter  $\beta$  and then  $\psi(\beta, \beta^\top \mathbf{z})$  can be estimated empirically, replacing  $C_\beta(\cdot | \beta^\top \mathbf{z})$  by an empirical counterpart in the previous integral.

If  $(\mathbf{X}, \mathbf{Z})$  and  $(\mathbf{Y}, \mathbf{Z})$  denote independent copies knowing  $\mathbf{Z}$ , note that

$$\begin{aligned} E_\beta \{ \mathbf{1}(X_1 > Y_1, X_2 > Y_2) | \beta^\top \mathbf{Z} = y \} &= E_\beta [ E_\beta \{ \mathbf{1} \{ F_1(X_1 | \mathbf{Z}) > F_1(Y_1 | \mathbf{Z}), F_2(X_2 | \mathbf{Z}) > F_2(Y_2 | \mathbf{Z}) \} | \mathbf{Z} \} | \beta^\top \mathbf{Z} = y ] \\ &= E_\beta \left\{ \int C(\mathbf{u} | \mathbf{Z}) C(d\mathbf{u} | \mathbf{Z}) | \beta^\top \mathbf{Z} = y \right\} = \int C_\beta(\mathbf{u} | y) C_\beta(d\mathbf{u} | y). \end{aligned}$$

This implies that the Kendall’s tau of  $\mathbf{X}$  given  $\beta^\top \mathbf{Z} = y$  is  $\tau(\beta, y)$ , under (1). Incidentally, we have proved that

$$\tau(\beta, \beta^\top \mathbf{z}) = -1 + 4 \int \tilde{C}_\beta(\mathbf{u} | y) \tilde{C}_\beta(d\mathbf{u} | y), \tag{10}$$

where  $\tilde{C}(\cdot | y)$  is the conditional copula of  $\mathbf{X}$  knowing  $\beta_0^\top \mathbf{Z} = y$ . Moreover, since

$$E_\beta \{ \mathbf{1}(X_1 > Y_1, X_2 > Y_2) | \beta^\top \mathbf{Z} = y \} = \int H_\beta(\mathbf{x}, \infty | \beta^\top \mathbf{Z} = y) H_\beta(d\mathbf{x}, \infty | \beta^\top \mathbf{Z} = y),$$

we identify a functional of  $H_\beta$  as in (A2), and then

$$\tau(\beta, \beta^\top \mathbf{z}) = -1 + 4 \int H_\beta(\mathbf{x}, +\infty | \beta^\top \mathbf{z}) H_\beta(d\mathbf{x}, +\infty | \beta^\top \mathbf{z}). \tag{11}$$

In other terms, Kendall’s tau belongs to the two classes (A1) and (A2) simultaneously. Both relations (10) and (11) are very useful in practice. Indeed, the estimation of  $H_\beta(\cdot | y)$  or  $\tilde{C}_\beta(\cdot | y)$  is less demanding than the nonparametric estimation of  $C_\beta(\cdot | \beta^\top \mathbf{z})$ : an empirical counterpart of  $H_\beta(\mathbf{x} | y)$  or  $\tilde{C}_\beta(\mathbf{u} | y)$  does not suffer from the curse of dimensionality because it requires only conditioning subsets in  $\mathbb{R}$ , contrary to  $C_\beta(\mathbf{u} | y)$  that involves conditioning with respect to  $\mathbf{z} \in \mathbb{R}^p$  to manage its marginal laws.

In dimension  $d$ , many Kendall’s tau can be built. Logically, these Kendall’s tau may be associated to any couple of variables  $(X_i, X_j)$  with  $i, j \in \{1, \dots, d\}$  and  $i \neq j$ . Or they can be defined formally as in (9), with  $d$ -dimension integrals, or even  $d'$ -dimension integrals,  $d' < d$  if we focus on some sub-vectors of  $\mathbf{X}$ . Globally, all such quantities are linear function of  $\int C(\mathbf{u}_I, \mathbf{1}_{\bar{I}} | \mathbf{z}) C(d\mathbf{u}_I, \mathbf{1}_{\bar{I}} | \mathbf{z})$ , where  $I$  is a subset of  $\{1, \dots, d\}$  and  $\bar{I}$  is its complement. Obviously,  $\mathbf{u}_I, \mathbf{1}_{\bar{I}}$  denotes a  $d$ -dimensional vector whose components are  $u_k$  when  $k \in I$ , and are equal to 1 otherwise. These dependence measures are candidates to yield convenient moments. Note the two usual generalizations of Kendall’s tau in dimension  $d$ : the first one has been proposed by Joe [22] as

$$\tau_d(\mathbf{z}) = \frac{1}{2^d - 1} \left\{ -1 + 2^d \int C(\mathbf{u} | \mathbf{z}) C(d\mathbf{u} | \mathbf{z}) \right\}, \tag{12}$$

and the second one has been introduced by Kendall and Babington Smith [25] as the average value of Kendall’s tau over all possible couples  $(X_k, X_\ell)$  with  $k, \ell \in \{1, \dots, d\}$  and  $k \neq \ell$ . See Genest et al. [16] for details and complementary results. In every case, the same arguments as above apply, providing straightforward  $d$ -dimensional extensions of (10) and (11). As a consequence, such generalized Kendall’s tau belong to the two classes (A1) and (A2) simultaneously and can be written as functionals of  $\hat{C}_\beta$  itself.

Now, let us specify our estimator  $\hat{\psi}$ . The simplest solution we adopt is to invoke kernel-type regression functions. In (A1), we can replace simply the conditional copula  $C_\beta(\cdot|\beta^\top \mathbf{Z} = y)$  by a consistent estimator  $\hat{C}(\cdot|\beta^\top \mathbf{Z} = y)$ . Several candidates exist in the literature. Historically, Fermanian and Wegkamp [15] studied a nearest-neighbor estimator of conditional copulas. Gijbels et al. [17] proposed other nonparametric estimates, including Nadaraya–Watson, Gaßer–Müller, etc.

In the class (A2), for every  $\beta \in \mathcal{B}$  and  $y \in \mathbb{R}$ , set  $\hat{\psi}(\beta, y) = \Psi\{\hat{H}_\beta(\cdot|y)\}$ , where

$$\begin{aligned} \hat{H}_\beta(\mathbf{x}, \mathbf{z}|y) &= \sum_{j=1}^n w_{\beta,j,n}(y) \mathbf{1}(\mathbf{X}_j \leq \mathbf{x}, \mathbf{Z}_j \leq \mathbf{z}), \\ w_{\beta,j,n}(y) &= K\left(\frac{\beta^\top \mathbf{Z}_j - y}{h_n}\right) / \sum_{\ell=1}^n K\left(\frac{\beta^\top \mathbf{Z}_\ell - y}{h_n}\right), \end{aligned} \tag{13}$$

for some kernel function  $K : \mathbb{R} \rightarrow \mathbb{R}$  and some bandwidth sequence  $(h_n)$ ,  $h_n > 0$ . Hereafter, we remove the latter sub-index  $n$ , i.e.,  $h = h_n$  simply for any bandwidth.

To satisfy Condition (4), we have to rely on the functional link between the parameter  $\psi$  and the underlying distributions, as evaluated in (6) and/or (7). This depends on the regularity of the corresponding functionals  $\Psi$  and on the uniform distance between the conditional empirical CDFs and true ones.

For instance, assume in (A2) that  $\Psi$  is Lipschitz, as in the case of Spearman’s rho and Kendall’s tau, with a Lipschitz constant  $\lambda$  (at least when  $\beta \in \mathcal{B}$  and  $\mathbf{z} \in \mathcal{Z}$ , and then  $\beta^\top \mathbf{z}$  belongs to a real compact subset). For such couples  $(\beta, \mathbf{z})$ , we have

$$|\hat{\psi}(\beta, \beta^\top \mathbf{z}) - \psi(\beta, \beta^\top \mathbf{z})| \leq \lambda \|\hat{H}_\beta(\cdot|\beta^\top \mathbf{z}) - H_\beta(\cdot|\beta^\top \mathbf{z})\|_\infty.$$

Assuming  $\hat{H}_\beta$  is given by (13) and applying Corollary 3 in Einmahl and Mason [14], we obtain

$$\max_{i \in \{1, \dots, n\}} |\hat{\psi}(\beta, \beta^\top \mathbf{Z}_i) - \psi(\beta, \beta^\top \mathbf{Z}_i)| \omega_{i,n} \leq \lambda \max_{i \in \{1, \dots, n\}} \|\hat{H}_\beta(\cdot|\beta^\top \mathbf{Z}_i) - H_\beta(\cdot|\beta^\top \mathbf{Z}_i)\|_{X_\infty} \omega_{i,n} \rightarrow 0,$$

a.e. and uniformly with respect to  $\beta \in \mathcal{B}$ . This is sufficient to satisfy (4).

More generally, in (A2) and if  $\Psi$  is Hadamard differentiable, there exist continuous linear maps  $\hat{\psi}_i$  such that

$$\begin{aligned} \hat{\psi}(\beta, \beta^\top \mathbf{Z}_i) - \psi(\beta, \beta^\top \mathbf{Z}_i) &= \Psi\{\hat{H}_\beta(\cdot|\beta^\top \mathbf{Z}_i)\} - \Psi\{H_\beta(\cdot|\beta^\top \mathbf{Z}_i)\} \\ &= \hat{\psi}_i\{(\hat{H} - H)_\beta(\cdot|\beta^\top \mathbf{Z}_i)\} + o\{\|(\hat{H} - H)_\beta(\cdot|\beta^\top \mathbf{Z}_i)\|\}. \end{aligned}$$

Under some additional conditions (particularly on the  $\hat{\psi}_i$ ), we typically get the uniformity of the latter identity with respect to  $\mathbf{Z}_i \in \mathcal{Z}$ . But, thanks to Theorem 3 in [14], there exists a sequence of positive numbers  $(a_n)$ ,  $a_n \rightarrow 0$ , such that

$$a_n \sup_{\beta \in \mathcal{B}} \sup_{\mathbf{z} \in \mathcal{Z}} \|(\hat{H} - H)_\beta(\cdot|\beta^\top \mathbf{z})\|_\infty \rightarrow 0$$

a.e. when  $n \rightarrow \infty$ . The latter result is true uniformly with respect to bandwidth sequences  $(h_n)$  such that  $nh_n/\ln n \gg 1$  and  $h_n \rightarrow 0$ . Therefore, (4) is usually satisfied when  $\Psi$  is Hadamard differentiable.

Note that the uniform consistency of the conditional copula function, simultaneously with respect to its argument and the conditioning value, is not available in the literature. Therefore, checking Condition (4) with (A1) is more difficult than with (A2).

### 2.3. The choice of the pseudo-observations $\hat{\mathbf{U}}$

By assumption,  $\beta$  is the index of the underlying dependence functions (copulas) only. Therefore,  $\hat{\mathbf{U}}_i$  does not depend on  $\beta$ . Now, let us discuss the possible choices for  $\hat{\mathbf{U}}_1, \dots, \mathbf{U}_n$ , in particular to satisfy (5). Actually, in Section 3, we consider a generic class of estimates such that, for all  $k \in \{1, \dots, d\}$ ,

$$\hat{F}_k(x|\mathbf{z}) - F_k(x|\mathbf{z}) = \frac{1}{n} \sum_{j=1}^n a_{k,n}(\mathbf{X}_j, \mathbf{Z}_j, x, \mathbf{z}) + r_n(x, \mathbf{z}), \tag{14}$$

for some sequence  $r_n(x, \mathbf{z})$  that tends to zero sufficiently quickly uniformly in probability, and for some particular functions  $a_{k,n}$ . Then, we set  $\hat{U}_{i,k} = \hat{F}_k(X_{i,k}|\mathbf{Z}_i)$  for all  $i \in \{1, \dots, n\}$  and  $k \in \{1, \dots, d\}$ .

A lot of estimators of  $F_k$  may be built and satisfy (14). A first example of such estimates is given by parametric marginal conditional distributions: for every  $x$  and  $\mathbf{z}$ ,  $F_k(x|\mathbf{z}) = G_{k,\theta_k(\mathbf{z})}(x)$ , for some family of CDFs  $\mathcal{G}_k = \{G_{k,\theta_k} : \theta_k \in \Theta_k\}$ . Since this model is parametric, the function  $\theta_k$  depends on a vector of parameters  $\eta_k \in \mathbb{R}^{m_k}$ . With a slight abuse of notations, set

$\theta_k(\mathbf{z}) = \theta_k(\mathbf{z}, \eta_k)$ , and  $\theta_k(\cdot, \eta)$  is known for every  $\eta$ . Assume we have found a consistent and asymptotically normal estimate  $\hat{\eta}_k$ , and set  $\hat{F}_k(\mathbf{x}|\mathbf{z}) = G_{k, \theta_k(\mathbf{z}, \hat{\eta}_k)}(\mathbf{x})$ . This implies  $\hat{U}_{i,k} = G_{k, \theta_k(\mathbf{z}_i, \hat{\eta}_k)}(X_{i,k})$ .

Clearly, for every  $i \in \{1, \dots, n\}$ , there exists  $\theta_{k,i}^*$  and  $\eta_k^*$  such that

$$|\hat{U}_{i,k} - U_{i,k}| \leq |\nabla_{\theta} G_{k, \theta_{k,i}^*}(X_{i,k})| \times |\partial_2 \theta_k(\mathbf{Z}_i, \eta_k^*)| \times |\hat{\eta}_k - \eta_k|,$$

where  $|\theta_k(\mathbf{Z}_i, \eta_k) - \theta_{k,i}^*| \leq |\theta_k(\mathbf{Z}_i, \hat{\eta}_k) - \theta_k(\mathbf{Z}_i, \eta_k)|$  and  $|\eta_k - \eta_k^*| \leq |\hat{\eta}_k - \eta_k|$ . Typically, if  $\sup_{\theta_k} |\nabla_{\theta} G_{k, \theta_k}(X_{i,k})|$  and  $\sup_{\eta_k} |\partial_2 \theta_k(\mathbf{Z}_i, \eta_k)|$  are bounded in probability, the condition (5) is satisfied, even without trimming. Moreover, in a lot of usual cases (M-estimates, for instance), it can be checked by a limited expansion that the functions  $\hat{F}_k(\mathbf{x}|\mathbf{z})$  satisfy (14). Typically, for asymptotically normal estimators, we have  $nr_n(\mathbf{x}, \mathbf{z}) = O_p(1)$ , and this result may be uniform under some conditions of regularity concerning  $G$  and  $\theta_k$ . Such a choice of conditional margins induces the estimator we label  $\hat{\beta}^{(1)}$ .

A second candidate is obtained by nonparametric estimates of conditional expectations. A usual kernel-based nonparametric estimator of  $F(\cdot|\mathbf{z})$  on  $\mathbb{R}^d$  is given by

$$\hat{F}(\mathbf{x}|\mathbf{z}) = \sum_{j=1}^n w_{j,n}(\mathbf{z}) \mathbf{1}(X_j \leq \mathbf{x}),$$

with the weights

$$w_{j,n}(\mathbf{z}) = \mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}) / \sum_{\ell=1}^n \mathbf{K}(\mathbf{Z}_{\ell} - \mathbf{z}, \mathbf{h}), \tag{15}$$

where  $\mathbf{K}$  is a multivariate kernel and  $\mathbf{h} = (h_1, \dots, h_p)$  is a  $p$ -vector of bandwidths  $h_k > 0$ . To simplify, we can restrict ourselves to products of  $p$  univariate kernels  $K_k$ :

$$\mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}) = \frac{1}{h_1 \dots h_p} \prod_{k=1}^p K_k \left( \frac{Z_{j,k} - z_k}{h_k} \right). \tag{16}$$

Therefore, some nonparametric estimators of every marginal conditional CDF  $F_k(\mathbf{x}|\mathbf{z})$  are obtained by setting  $\hat{F}_k(\mathbf{x}|\mathbf{z}) = \hat{F}(\mathbf{x}, \infty_{(-k)}|\mathbf{z})$ . The marginal “unfeasible” observations are  $U_{i,k} = F_k(X_{i,k}|\mathbf{Z}_i)$ , and their estimated versions are  $\hat{U}_{i,k} = \hat{F}_k(X_{i,k}|\mathbf{Z}_i)$ . In this case, it can be verified that (14) is satisfied, even if  $\hat{F}(\mathbf{x}|\mathbf{z})$  suffers from the curse of dimensionality along the  $\mathbf{z}$  dimension; see Lemma 4 in Appendix A. As a consequence of this lemma, Condition (5) is satisfied for the nonparametric versions on  $\hat{U}_i$  and for a wide range of bandwidths. Let us denote the associated estimator by  $\hat{\beta}^{(2)}$ .

Alternatively, it could be assumed that some conditional distributions  $F_k(\cdot|\mathbf{Z})$  are given by particular single-index models, but with some parameter  $\beta_k \in \mathbb{R}^{m_k}$  that are different from  $\beta$ . If the latter index  $\beta_k$  is estimated consistently by  $\hat{\beta}_k$ , then we can adapt easily the previous nonparametric kernel estimator: for any real number  $y$ ,

$$\hat{F}_{k, \hat{\beta}_k}(x|y) = \sum_{j=1}^n w_{\hat{\beta}_k, j, n}(y) \mathbf{1}(X_{j,k} \leq x),$$

where

$$w_{\hat{\beta}_k, j, n}(y) = K \left( \frac{\hat{\beta}_k^{\top} \mathbf{Z}_j - y}{h} \right) / \sum_{\ell=1}^n K \left( \frac{\hat{\beta}_k^{\top} \mathbf{Z}_{\ell} - y}{h} \right),$$

for some kernel function  $K : \mathbb{R} \rightarrow \mathbb{R}$  and some bandwidth  $h > 0$ . Obviously,  $\hat{F}_{k, \hat{\beta}_k}(x|y)$  yields a nonparametric estimator of the CDF  $F_{k, \beta_k}(x|y)$ . In this case,  $U_{k, \mathbf{z}} = F_k(X_k|\beta_k^{\top} \mathbf{z})$ . To deal with pseudo-observations, we set  $U_{i,k, \beta_k} = F_{k, \beta_k}(X_{i,k}|\beta_k^{\top} \mathbf{Z}_i)$ , and  $\hat{U}_{i,k} = \hat{F}_{k, \hat{\beta}_k}(X_{i,k}|\hat{\beta}_k^{\top} \mathbf{Z}_i)$ . For some conditions of regularity, (14) can be verified; see, e.g., Du and Akritas [12] for such a representation in the more general case of censored data. When all margins are assumed single-index, let us denote by  $\hat{\beta}^{(3)}$  the corresponding  $\beta$  estimator.

Now, let us verify the conditions of Theorem 1, particularly Assumption 3, in some particular cases.

### 2.4. Examples

Let us illustrate the previous ideas with a few standard copula models.

**Example 4 (The Gaussian Copula).** For every  $\mathbf{u}$  and  $\mathbf{z}$  and with usual notations, given  $\beta_0$ , the underlying copula is

$$C_{\beta_0}(\mathbf{u}|\mathbf{z}) = \Phi_{\Sigma(\mathbf{z})} \{ \Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d) \},$$

where the correlation matrix  $\Sigma(\mathbf{z}) = [\theta_{k, \ell}(\mathbf{z})]_{1 \leq k, \ell \leq d}$  depends on the index  $\beta_0^{\top} \mathbf{z}$  only. With our previous notations,  $\Sigma(\mathbf{z}) = \psi(\beta_0, \beta_0^{\top} \mathbf{z})$ . It is well-known that every component  $\theta_{k, \ell}(\mathbf{z})$  of  $\Sigma(\mathbf{z})$  is a function of the conditional Kendall’s tau

that is associated to  $(X_k, X_\ell)$ , knowing  $\beta_0^\top \mathbf{Z} = \beta_0^\top \mathbf{z}$ , viz.  $\theta_{k,\ell}(\mathbf{z}) = \sin\{\pi \tau_{k,\ell}(\beta_0, \beta_0^\top \mathbf{z})/2\}$ . The latter quantity can be estimated by standard nonparametric techniques, and then

$$\hat{\Sigma}(\mathbf{z}) = \hat{\psi}(\beta, \beta^\top \mathbf{z}) = \left[ \sin \left\{ \frac{\pi}{2} \hat{\tau}_{k,\ell}(\beta, \beta^\top \mathbf{z}) \right\} \right].$$

Even if the latter estimated matrix  $\hat{\Sigma}(\mathbf{z})$  may not be positive definite, this is not required for proving the consistency of  $\hat{\beta}$ . At the opposite, it will be necessary to obtain its asymptotic normality and some estimates of its limiting law (see below). To be specific, we can choose

$$\hat{\tau}_{k,\ell}(\beta, y) = -1 + 4 \int \hat{C}_{k,\ell}(u, v | \beta^\top \mathbf{Z} = y) \hat{C}_{k,\ell}(du, dv | \beta^\top \mathbf{Z} = y),$$

for some estimator  $\hat{C}_{k,\ell}(\cdot | \beta^\top \mathbf{Z} = y)$  of the conditional copula of  $(X_k, X_\ell)$  given  $\beta^\top \mathbf{Z} = y$ . Alternatively, we can invoke an asymptotically equivalent estimator

$$\hat{\tau}_{k,\ell}(\beta, \beta^\top \mathbf{z}) = -1 + 4 \sum_{i=1}^n \sum_{j=1}^n w_{i,h}(\beta^\top \mathbf{z}) w_{j,h}(\beta^\top \mathbf{z}) \mathbf{1}(X_{k,i} < X_{k,j}, X_{\ell,i} < X_{\ell,j}),$$

for some weights, for instance the standard Nadaraya–Watson kernel

$$w_{i,h}(y) = K \left( \frac{y - \beta^\top \mathbf{Z}_i}{h} \right) / \sum_{\ell=1}^n K \left( \frac{y - \beta^\top \mathbf{Z}_\ell}{h} \right).$$

See [17] for alternative weights and estimators.

Once we have stated  $\hat{\psi}$ , it remains to set the marginal CDFs  $\hat{U}_1, \dots, \hat{U}_d$  to be able to compute our estimator  $\hat{\beta}$ . To fix ideas, we rely on the standard univariate kernel-based conditional distributions, as given in (15):  $\hat{U}_{i,k} = \hat{F}(X_{i,k} | \mathbf{Z}_i)$  and our estimator is then  $\hat{\beta}^{(2)}$ .

Concerning Assumption 2, the only thing to be checked is (3). This is guaranteed when the random matrix  $\Sigma^{-1}(\mathbf{Z})$  is staying “under control”, e.g., when all eigenvalues of  $\Sigma(\mathbf{Z})$  are uniformly bounded from below almost surely. It is sufficient to assume that

$$\sup_{\mathbf{z} \in \mathcal{Z}} \sup_{\beta \in \mathcal{B}} \lambda_{\min}\{\psi(\beta, \beta^\top \mathbf{z})\} \geq \underline{\lambda} > 0, \tag{17}$$

where  $\lambda_{\min}(\Sigma)$  denotes the smallest eigenvalue of any nonnegative matrix  $\Sigma$ . In this case, it is easy to bound the log-density of  $\mathbf{X}$  (conditional on  $\mathbf{Z}$ ) from above, and to satisfy (3).

Assumption 3 is the trickiest. In Lemma 12, some sufficient conditions are given to satisfy (4). It remains to verify (5). We can apply our Lemma 4: under its conditions and if all the bandwidths we consider in  $\hat{\mathbf{U}}$  behave as the same power of  $n$ , say  $n^{-\pi}$  (the usual case), there exists a constant  $C$  such that, with probability 1,

$$\limsup_n \max_{i \leq n} |\hat{\mathbf{U}}_i - \mathbf{U}_i| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) / \delta_n \leq C, \quad \text{where } \delta_n = \sqrt{\ln(n)} n^{-(1-p\pi)/2} + n^{-\pi s}.$$

Note that, for consistency purposes, we can choose any  $\pi$  such that  $\pi < 1/p$ , and  $v_n$  can be chosen arbitrarily as long as we have  $v_n \gg \delta_n$ , and then the condition (5) is satisfied.

Assumption 4 is satisfied for the Gaussian copula, as in most usual copula families. In our case and under (17), we choose

$$r(\mathbf{u}) \propto \sum_{k=1}^d \{\Phi^{-1}(u_k)\}^2 \quad \text{and} \quad \tilde{r}_k(\mathbf{u}) \propto \Phi^{-1}(u_k) \sum_{\ell=1, \ell \neq k}^d \{\Phi^{-1}(u_\ell)\} \times \{\phi \circ \Phi^{-1}(u_k)\}^{-1}.$$

Therefore, the estimator  $\hat{\beta}^{(2)}$  is consistent under the Gaussian copula framework.

**Example 5 (The Clayton Copula).** The Clayton copula is often useful in finance, because it induces left-tail dependence, a common feature of asset returns. When the values of its parameter are strictly positive, the conditional Clayton copula is written, for all  $\mathbf{u} \in (0, 1)^d$ ,

$$C(\mathbf{u} | \mathbf{z}) = \left( \sum_{k=1}^d u_k^{-\theta(\mathbf{z})} - d + 1 \right)^{-1/\theta(\mathbf{z})},$$

with  $\theta(\mathbf{z}) = \psi(\beta, \beta^\top \mathbf{z})$  under the single-index assumption. As with the Gaussian copula model, we can evaluate  $\hat{\psi}$  with conditional Kendall’s tau, because of their one-to-one mapping. Indeed, invoking Example 1 in Genest et al. [16], the Kendall’s tau for a Clayton model is equal to

$$\tau_d = \frac{1}{2^{d-1} - 1} \left\{ -1 + 2^d \prod_{k=0}^{d-1} \left( \frac{1 + k\theta}{2 + k\theta} \right) \right\}.$$



It is easy to check that the latter mapping between  $\tau$  and  $\theta$  is one-to-one. The density of the Clayton copula with parameter  $\theta > 0$  is given by

$$\ln c_\theta(\mathbf{u}|\mathbf{z}) = \sum_{k=1}^{d-1} \ln(1 + k\theta) - (\theta + 1) \sum_{k=1}^d \ln(u_k) - \left(\frac{1}{\theta} + d\right) \ln\left(\sum_{k=1}^d u_k^{-\theta} - 1\right).$$

Assume that there exist  $\underline{\theta}$  and  $\bar{\theta}$  such that, for every  $\mathbf{z} \in \mathcal{Z}$  and every  $\beta \in \mathcal{B}$ ,  $\underline{\theta} \leq \psi(\beta, \beta^\top \mathbf{z}) \leq \bar{\theta}$ . Then Assumption 2 is satisfied. Indeed, note that

$$0 \leq \ln\left(\sum_{k=1}^d u_k^{-\theta} - d + 1\right) \leq \sum_{k=1}^d \ln(du_k^{-\theta}) \leq d \ln(d) - \bar{\theta} \sum_{k=1}^d \ln(u_k).$$

Letting  $V$  denote a  $\mathcal{U}(0, 1)$  random variable, we have

$$E[\ln\{F_k(X_k|\mathbf{Z})\}] = E_Z [E_{X_k|\mathbf{Z}}[\ln\{F_k(X_k|\mathbf{Z})\}|\mathbf{Z}]] = E_Z \{E_{X_k|\mathbf{Z}}[\ln V]\} = (-1),$$

and (3) follows.

Assumption 3 is satisfied with the same arguments as for the Gaussian copula. Assumption 4 can be verified relatively easily. Concerning  $\nabla_\theta \ln c_\theta(\mathbf{u}|\mathbf{z})$ , the relevant reproducing U-shaped function is given by the product of the functions  $r_k(u) \propto -\ln(u_k)\mathbf{1}(u_k \in (0, 1/2]) - \ln(1 - u_k)\mathbf{1}(u_k \in (1/2, 1))$  with  $k \in \{1, \dots, d\}$ . To see this, use the following inequality: for every  $\mathbf{u} \in (0, 1)^d$ ,

$$\frac{|\sum_{k=1}^d u_k^{-\theta} \ln u_k|}{\sum_{k=1}^d u_k^{-\theta} - d + 1} \leq \frac{\max_k u_k^{-\theta} \sum_{k=1}^d |\ln u_k|}{\sum_{k=1}^d u_k^{-\theta} - d + 1} \leq -\sum_{k=1}^d \ln u_k.$$

To manage  $\nabla_{u_k} \ln c_\theta(\mathbf{u}|\mathbf{z})$ , the relevant reproducing U-shaped function is obtained by replacing  $r_k$  above by  $\tilde{r}_k(u) \propto u_k^{-1}(1 - u_k)^{-1}$ . Assumption 4 follows by setting  $\tilde{r}_k(\mathbf{u}) = \tilde{r}_k(u_k) \prod_{l \neq k} r_l(u_l)$ .

**Example 6 (The Gumbel Copula).** The  $d$ -dimensional Gumbel copula is given by

$$C_\theta(\mathbf{u}) = \exp\left\{-\left(\sum_{k=1}^d |\ln u_k|^\theta\right)^{-1/\theta}\right\},$$

for some parameter  $\theta > 1$ . It exhibits right-tail dependence.

Its Kendall's tau in dimension  $d$ , as defined by (12) has been calculated by Genest et al. [16] as

$$\tau_d = \frac{1}{2^d - 1} \left[ -1 + 2^d \sum C_{\vec{m}} \frac{(m-1)!}{(d-1)!} \left(\frac{1}{2\theta}\right)^{m-1} \prod_{q=1}^d \left\{ \prod_{\ell=1}^{q-1} (k - 1/\theta) \right\}^{m_q} \right],$$

where  $\vec{m} = (m_1, \dots, m_d)$  with  $m = m_1 + \dots + m_d$ , and the summation is taken over all  $d$ -uplets of integers such that  $m_1 + 2m_2 + \dots + dm_d = d$ . For every  $\vec{m}$ ,  $C_{\vec{m}}$  denotes a positive constant. Note that

$$\left(\frac{1}{\theta}\right)^{m-1} \prod_{q=1}^d \left\{ \prod_{\ell=1}^{q-1} (k - 1/\theta) \right\}^{m_q} = \left(\frac{1}{\theta}\right)^{d-1} \prod_{q=2}^d \left\{ \prod_{\ell=1}^{q-1} (k\theta - 1) \right\}^{m_q} = \chi_{\vec{m}}(\theta),$$

and

$$(\ln \chi_{\vec{m}})'(\theta) \propto -(d-1) + \sum_{q=2}^d \sum_{k=1}^{q-1} \frac{km_q}{k - 1/\theta} > -(d-1) + \sum_{q=2}^d \sum_{k=1}^{q-1} m_q = 0.$$

Therefore, every function  $\chi_{\vec{m}}$  above is invertible, and the mapping between  $\theta$  and  $\tau$  is one-to-one, as usual. We can use the empirical (conditional) Kendall's tau to evaluate the under parameter  $\theta$  (or  $\theta(\mathbf{z})$  more generally).

The Gumbel copula density is a linear combination of the functions

$$c_j(\mathbf{u}) = C_\theta(\mathbf{u}) \left(\sum_{k=1}^d |\ln u_k|^\theta\right)^{j/\theta-d} \prod_{k=1}^d \frac{|\ln u_k|^{\theta-1}}{u_k},$$

for some  $j \in \{1, \dots, d\}$ . In the single-index model,  $\theta$  is a function of  $\mathbf{z}$ . Assume that  $\theta(\mathbf{z})$  belongs to a fixed interval  $[\underline{\theta}, \bar{\theta}] \subset (1, +\infty)$  almost everywhere. Therefore, the density  $c_{\theta(\mathbf{z})}$  of a Gumbel copula satisfies

$$c_{\theta(\mathbf{z})}(\mathbf{u}) \leq K_0 \times C(\mathbf{u}) \max_{\theta \in \{\underline{\theta}, \bar{\theta}\}} \left\{ \left(\sum_{k=1}^d |\ln u_k|^\theta\right)^{j/\theta-d} \prod_{k=1}^d \frac{|\ln u_k|^{\theta-1}}{u_k} \right\},$$

for every  $\mathbf{u} \in (0, 1)^d$  and some constant  $K_0$ . By taking the logarithm of the previous right-hand side, it is easy to check that (3), and then Assumption 2, are satisfied.

Assumption 3 is satisfied with the same arguments as above. After lengthy calculations, we can check Assumption 4 too, by noticing that

$$\sup_{\theta \in [\underline{\theta}, \bar{\theta}]} |\partial_{\mathbf{u}_k} c_{\theta}(\mathbf{u})| \leq K_1 \times h_k(\mathbf{u}) C_{\theta}(\mathbf{u}) / u_k^2 = \tilde{r}_k(\mathbf{u}),$$

for some constant  $K_1$  and some slowly varying functions  $h_k$  (deduced from the powers of the functions  $u_{\ell} \mapsto \ln u_{\ell}$  with  $\ell \in \{1, \dots, d\}$ ). The function  $\tilde{r}_k$  belongs to  $\mathcal{R}_d$  since  $C_{\theta}(\mathbf{u})$  behaves as  $u_k$  when  $u_k$  tends to zero. Therefore  $E\{U_k(1 - U_k)\tilde{r}_k(\mathbf{U})\} < \infty$ .

### 3. Asymptotic normality of the estimators

#### 3.1. Notations and assumptions

For convenience, set  $\psi_i = \psi(\beta_0, \beta_0^{\top} \mathbf{Z}_i)$  and  $\hat{\psi}_i = \hat{\psi}(\beta_0, \beta_0^{\top} \mathbf{Z}_i)$ .

**Assumption 5.** For every  $\mathbf{z} \in \mathcal{Z}$ , assume that  $\psi_{\mathbf{z}} : \mathcal{B} \rightarrow \Theta, \beta \mapsto \psi(\beta, \beta^{\top} \mathbf{z})$  is three times continuously differentiable. Moreover, set  $\ln c : (0, 1)^d \times \Theta \rightarrow \mathbb{R}, (\mathbf{u}, \theta) \mapsto \ln c_{\theta}(\mathbf{u})$ . Assume that  $\nabla_{\mathbf{u}} \nabla_{\theta}^2 \ln c_{\theta}(\mathbf{u})$  exists on  $(0, 1)^d \times \Theta$ . Assume that  $\beta \mapsto M(\beta)$  is twice continuously differentiable. Its Hessian matrix at point  $\beta_0$  is denoted  $\Sigma = \nabla_{\beta}^2 M(\beta_0)$ , and is invertible.

Such conditions of regularity are standard and can be easily checked most often. Simple calculations yield

$$\Sigma = \frac{1}{\Pr(\mathbf{Z} \in \mathcal{Z})} E \left[ \left\{ \nabla_{\theta}(\ln c_{\theta})_{|\theta=\psi_1}(\mathbf{U}_i) \nabla_{\beta}^2 \psi(\beta, \beta^{\top} \mathbf{Z}_i) + \nabla_{\theta}^2(\ln c_{\theta})_{|\theta=\psi_1}(\mathbf{U}_i) \nabla_{\beta} \psi_i \nabla_{\beta}^{\top} \psi_i \right\} \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) \right].$$

**Assumption 6.** Assume that, for  $k \in \{0, 1, 2\}$ ,

$$\begin{aligned} \sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} \left| \nabla_{\beta}^k \hat{\psi}(\beta, \beta^{\top} \mathbf{z}) - \nabla_{\beta}^k \psi(\beta, \beta^{\top} \mathbf{z}) \right| &= o_P(1), \\ \sup_{\mathbf{z} \in \mathcal{Z}} |\hat{\psi}(\beta_0, \beta_0^{\top} \mathbf{z}) - \psi(\beta_0, \beta_0^{\top} \mathbf{z})| &= O_P(\eta_{1n}), \\ \sup_{\mathbf{z} \in \mathcal{Z}} |\nabla_{\beta} \hat{\psi}(\beta, \beta^{\top} \mathbf{z})_{|\beta=\beta_0} - \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z})_{|\beta=\beta_0}| &= O_P(\eta_{2n}), \end{aligned} \tag{18}$$

with  $\delta_n^{1-\alpha} \eta_{jn} = o(n^{-1/2})$ , for  $j \in \{1, 2\}$ ,  $\eta_{1n}^2 = o(n^{-1/2})$ , and  $\eta_{1n} \eta_{2n} = o(n^{-1/2})$ .

The latter assumption requires that the rate of uniform weak consistency of  $\hat{\psi}$  (and some of its derivatives) is sufficiently quick. For the clarity of these developments, some additional technical assumptions have been postponed into Appendix B.2. They are related to the regularity of the underlying model distributions/functions and some of their estimates. Broadly speaking, the key point is to check Assumption 6 in practice.

#### 3.2. Main results

Now, let us state the main theoretical result, that will be key to obtain the asymptotic normality of  $\hat{\beta}$ .

**Theorem 2.** Under Assumptions 1 to 12,

$$\hat{\beta} - \beta_0 = -\frac{\Sigma^{-1}}{n} \sum_{i=1}^n \omega_{i,n} \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i)_{|\beta=\beta_0} + o_P(n^{-1/2}).$$

The proof of Theorem 2 is given in Appendix B.2. Invoking a few additional assumptions (see Appendix B.3), we finally obtain the asymptotic normality of  $\hat{\beta}$ .

**Corollary 3.** Under Assumptions 1 to 16, we have  $n^{1/2}\{\Sigma(\hat{\beta} - \beta_0) + b_n\} \rightsquigarrow \mathcal{N}(0, S)$  as  $n \rightarrow \infty$ , where  $S = E(\omega_1 \mathcal{M}_1 \mathcal{M}_1^{\top})$ , and

$$\mathcal{M}_1 = \frac{\nabla_{\theta} c_{\theta}}{c_{\theta}} \Big|_{\theta=\psi_1}(\mathbf{U}_1) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_1)_{|\beta=\beta_0} + \nabla_{\theta} \nabla_{\theta}(\ln c_{\theta})_{|\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{Z}_1)}(\mathbf{U}_1) W(\mathbf{Z}_1, \mathbf{X}_1) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_1)_{|\beta=\beta_0},$$

for some  $d$ -dimensional random vector  $W(\mathbf{Z}_1, \mathbf{X}_1)$  that is defined in Assumption 15 of Appendix B.3, and

$$b_n = E\{(\omega_{1,n} - \omega_i) \mathcal{M}_1\} = E\{\mathbf{1}(\mathbf{U}_1 \in [0, 1]^d) \varepsilon_n, \mathbf{Z}_1 \in \mathcal{Z}\} \mathcal{M}_1.$$

Moreover, if

$$E \left[ \nabla_{\mathbf{u}} \nabla_{\theta} (\ln c_{\theta})_{|\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{z}_1)}(\mathbf{U}_1) W(\mathbf{Z}_1, \mathbf{X}_1) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z}_1)_{|\beta=\beta_0} \right. \\ \left. \times \{ \mathbf{1}(|U_{k,1} - v_n| < \delta_n) + \mathbf{1}(|1 - U_{k,1} - v_n| < \delta_n) \} \right] = o(n^{-1/2}), \tag{19}$$

for every  $k \in \{1, \dots, d\}$ , then  $n^{1/2}b_n = o(1)$  and  $n^{1/2}(\hat{\beta} - \beta_0) \rightsquigarrow \mathcal{N}(0, \Sigma^{-1}S\Sigma^{-1})$  as  $n \rightarrow \infty$ .

This result is proved in [Appendix B.2](#). Note that the bias  $b_n$  cannot be removed in general, even if  $E(\mathbf{a}_{i,j}) = 0$ . Indeed, the trimming part  $E\{(\omega_{i,n} - \omega_i)\mathcal{M}_i\}$  is typically of order  $\delta_n$  and has no reason to be  $o(n^{-1/2})$ . To remove the asymptotic bias, we need (19). The latter condition is easily satisfied with purely parametric or nonparametric estimates, because  $W(\mathbf{Z}, \mathbf{X})$  is zero or most often negligible in such cases; see the discussion of [Assumption 15](#) in [Appendix B.3](#).

The asymptotic variance of  $\hat{\beta}$  in [Corollary 3](#) can be estimated easily in most cases. Indeed, set

$$\forall_{k \in \{0,1,2,\}} \forall_{i \in \{1,\dots,n\}} \nabla_{\beta}^k \check{\psi}_i = \nabla_{\beta}^k \hat{\psi}(\beta, \beta^{\top} \mathbf{z}_i)_{|\beta=\hat{\beta}}.$$

We can consistently estimate  $\Sigma$  by

$$\hat{\Sigma} = \frac{1}{n\text{Pr}_n(\mathbf{Z} \in \mathcal{Z})} \sum_{i=1}^n \left\{ \nabla_{\theta} (\ln c_{\theta})_{|\theta=\check{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta}^2 \check{\psi}_i + \nabla_{\theta}^2 (\ln c_{\theta})_{|\theta=\check{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta} \check{\psi}_i \nabla_{\beta^{\top}} \check{\psi}_i \right\} \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}),$$

where  $\text{Pr}_n(\mathbf{Z} \in \mathcal{Z})$  is the empirical probability of observing  $\mathbf{Z}$  in  $\mathcal{Z}$ . Note that  $\hat{\Sigma}$  is not definite positive in general. Nonetheless, several techniques exist to build definite positive approximations of any symmetric matrix; see, e.g., [20,36]. Moreover, estimate  $S$  by the definite positive matrix  $\hat{S} = n^{-1} \sum_{i=1}^n \hat{\omega}_{i,n} \hat{\mathcal{M}}_i \hat{\mathcal{M}}_i^{\top}$ , where

$$\hat{\mathcal{M}}_i = \nabla_{\theta} (\ln c_{\theta})_{|\theta=\check{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta} \check{\psi}_i + \nabla_{\mathbf{u}} \nabla_{\theta} (\ln c_{\theta})_{|\theta=\check{\psi}_i}(\hat{\mathbf{U}}_i) \cdot \hat{W}(\mathbf{Z}_i, \mathbf{X}_i) \nabla_{\beta} \check{\psi}_i,$$

for some functions  $\hat{W}$  that consistently estimate  $W$ . It is difficult to specify  $\hat{W}$  because its analytic expression depends on the selected definition of the pseudo-observations  $\hat{\mathbf{U}}_i$ . Nonetheless, note that we can set  $\hat{W} = 0$  when the  $\hat{\mathbf{U}}_i$  are obtained through kernel-based estimates; see [Appendix B.3](#).

It is possible to verify that the conditions above (1 to 16) apply to get the asymptotic normality of  $\hat{\beta}$  in the case of the copula families in [Section 2.4](#); see [Appendix B.4](#).

#### 4. A short simulation study

To investigate the finite-sample behavior of our estimate  $\hat{\beta}$ , we consider the following simulation setting. Let  $\mathbf{Z} \in \mathbb{R}^4$  be a random vector with independent margins that are  $\mathcal{U}(0, 1)$ . The conditional (marginal) distributions of  $X_1$  and  $X_2$  given  $\mathbf{Z} = \mathbf{z}$  are respectively taken as

$$X_1 | \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(\alpha^{\top} \mathbf{z}, 1), \quad X_2 | \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(\alpha^{\top} \mathbf{z}, 1),$$

with  $\alpha = (0.2, -0.3, 0.1, 0.1)^{\top}$ . Concerning the conditional copula model, the dependence structure of  $(X_1, X_2)$  given  $\mathbf{Z} = \mathbf{z}$  is a Clayton copula of parameter  $\beta_0^{\top} \mathbf{z}$ . We recall that a Clayton copula with a parameter  $\theta > 0$  is defined as  $C_{\theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ . Set  $\beta_0 = (\beta_{0,1}, \beta_{0,2}, \beta_{0,3}, \beta_{0,4}) = (1, 0.7, 1.2, 0.8)^{\top}$ . We then assume

$$\theta(\mathbf{z}) = \psi(\beta_0, \beta_0^{\top} \mathbf{z}) = \frac{\beta_0^{\top} \mathbf{z} / 3.7}{1 - \beta_0^{\top} \mathbf{z} / 7.4}.$$

Therefore, the values of the random parameter  $\theta(\mathbf{Z})$  belong to  $(0, 2)$ . The corresponding values of the conditional Kendall's tau coefficient (which depends of  $\mathbf{Z}$ ) lies between 0 and 1/2. Under our simulation scheme (independent covariates  $\mathbf{Z}$  uniformly distributed over  $(0, 1)$ ), this leads to an average value of 1/4 for  $\tau(\beta_0, \beta_0^{\top} \mathbf{Z})$ . The sample size is  $n = 1000$ . We consider 1000 replications of this simulation scheme. For each simulated sample, we compute our estimator  $\hat{\beta}$  as defined in (2). The conditional margins are estimated using a kernel estimator as in (A.1), with products of univariate Gaussian kernels  $K_1 = \dots = K_4 = \phi$ . The bandwidth used in the estimation of the margins is taken as  $h = 0.12$  as per Scott's rule in  $\mathbb{R}^4$ ; see [40], p. 164. In contrast, we consider different values of the bandwidth for computing the conditional Kendall's  $\tau$  estimators, to understand its impact on the performances. To measure the estimation errors, for each simulated sample and each value of the bandwidth, we introduce

$$e(\hat{\beta}) = \frac{1}{3} \sum_{i=2}^4 \frac{|\hat{\beta}_i - \beta_{0,i}|}{|\beta_{0,i}|}.$$

The latter number measures the mean relative absolute distance between the true parameter  $\beta_0$  and its estimate. The mean value of  $e(\hat{\beta})$  over the 1000 replications is then calculated, providing  $\bar{e}(\hat{\beta})$ . For a grid of realistic bandwidths between 0.03 and 0.2, the obtained range of  $\bar{e}(\hat{\beta})$  is [1.3%, 6.1%], with an average value of 3.5%. If we look at the results component by

component (i.e., if, for each sample, we compute  $e_i(\hat{\beta}) = |\hat{\beta}_i - \beta_{0,i}| |\beta_{0,i}|^{-1}$  for  $i \in \{2, 3, 4\}$ ), we obtain similar levels of errors:  $e_2(\hat{\beta})$  lies between [2.0%, 7.8%] with an average value of 4.7%,  $e_3(\hat{\beta})$  between [0.8%, 4.9%] with an average value of 2.8% and  $e_4(\hat{\beta})$  between [1.0%, 5.9%] with an average value of 3.8%. Regarding the standard deviations and defining  $s_i(\hat{\beta}) = \text{var}(\hat{\beta}_i)^{1/2}$  for  $i = 2, 3, 4$ , we have  $s_2/\beta_{0,2} \in [0.9\%, 4.8\%]$  with an average value of 3.8%,  $s_3/\beta_{0,3} \in [0.1\%, 3.2\%]$  with an average value of 2.5%, and  $s_4/\beta_{0,4} \in [0.7\%, 3.6\%]$  with an average value of 2.9%.

As another error measurement criterion, we considered the empirical estimation of the quadratic error  $E\{(\hat{\beta} - \beta_0)^2\}$ . This error goes from 0.0040 to 0.0049 for our range of bandwidths. A bias/variance decomposition of this error, i.e.,  $E\{(\hat{\beta} - \beta_0)^2\} = \text{var}(\hat{\beta}) + \{E(\hat{\beta}) - \beta_0\}^2$ , may give us indications on the relative influence of the bias with respect to the error. The empirical estimation of  $\{E(\hat{\beta}) - \beta_0\}^2$  goes from 24.7% of the global quadratic error to 36.3%. Such error levels are very reasonable. However, the asymptotic normality of  $\hat{\beta}$  may not be achieved for this sample size, since the null hypothesis of a Shapiro–Wilk test is strongly rejected. A larger sample size may be required to obtain this asymptotic normality. As expected, there is an impact of the bandwidth on the practical behavior. However, we focus here on estimating  $\beta_0$ , which is relatively weakly sensitive to the choice of the bandwidth. Indeed, it is based on the maximization of a criterion where the nonparametric kernel estimators are averaged.

**Acknowledgments**

The authors thank Axel Bücher, Holger Dette, Christian Genest, Johan Segers, Alexandre Tsybakov, and Stanislav Volgushev for helpful discussions, as well as numerous seminar participants, particularly at the 2015 European Meeting of Statisticians in Amsterdam, and at the SFB 2015 workshop in Bochum. Moreover, we received many relevant remarks and suggestions from the Associate Editor and three referees. This research has been supported by the Labex Ecodec (reference project 11-LABX-0047).

**Appendix A. Technical lemmas**

**Lemma 4.** For each  $k \in \{1, \dots, d\}$ , define  $\hat{F}_k$  as

$$\hat{F}_k(x|\mathbf{z}) = \sum_{j=1}^n w_{j,n}(\mathbf{z}) \mathbf{1}(X_{j,k} \leq x), \tag{A.1}$$

with the weights given by (16). Assume that, for some  $s \geq 2$ , the following conditions hold:

- (i)  $f_{\mathbf{Z}}$ , the density of  $\mathbf{Z}$ , exists and is strictly positive on  $\mathcal{Z}$ . Moreover, it is  $s$ -times continuously differentiable.
- (ii) For every real  $x$  and every  $k$ , the function  $h_k(x, \cdot) : \mathbf{z} \mapsto \Pr(X_k \leq x | \mathbf{Z} = \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z})$ , defined on  $\mathcal{Z}$ , is  $s$ -times continuously differentiable. Moreover,

$$\sup_{x \in \mathbb{R}} \sup_{\mathbf{z} \in \mathcal{Z}} |d_{\mathbf{z}}^s h_k(x, \mathbf{z})| \text{ is bounded.}$$

- (iii) The underlying kernel  $\mathbf{K}(\cdot, \mathbf{1})$  is continuous, bounded, of bounded variation,  $\int \mathbf{K}(\mathbf{z}, \mathbf{1}) d\mathbf{z} = 1$  and it is compactly supported. To be specific, this kernel has to be “regular” in the sense of Einmahl and Mason [14], i.e., it has to satisfy Assumptions K.i–K.iv. Moreover, it is a multivariate  $s$ -order kernel, i.e.,

$$\int \prod_{j=1}^p z_j^{\alpha_j} \mathbf{K}(\mathbf{z}, \mathbf{1}) d\mathbf{z} = 0,$$

for every  $p$ -uplet of integers  $(\alpha_1, \dots, \alpha_p)$  such that  $\alpha_j \in \{1, \dots, s - 1\}$  for some index  $j$ .

Then, for all  $k \in \{1, \dots, d\}$ , we have

$$\begin{aligned} \hat{F}_k(x|\mathbf{z}) - F_k(x|\mathbf{z}) &= \frac{1}{n} \sum_{j=1}^n a_{k,n}(\mathbf{X}_j, \mathbf{Z}_j, x, \mathbf{z}) + r_n(x, \mathbf{z}), \\ a_{k,n}(\mathbf{X}_j, \mathbf{Z}_j, x, \mathbf{z}) &= \frac{1}{f_{\mathbf{Z}}(\mathbf{z})} \left[ \mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}) \mathbf{1}(X_{j,k} \leq x) - E\{\mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}) \mathbf{1}(X_{j,k} \leq x)\} \right. \\ &\quad \left. - \Pr(X_k \leq x | \mathbf{Z} = \mathbf{z}) [\mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}) - E\{\mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h})\}] \right], \end{aligned}$$

where, by setting

$$u_{n,1} = \left[ \frac{n \prod_{\ell=1}^p h_{\ell}}{\max\{-\ln(\prod_{\ell=1}^p h_{\ell}), \ln \ln n\}} \right]^{1/2}, \quad \text{and } u_{n,2} = \frac{1}{\max_{\ell \in \{1, \dots, p\}} h_{\ell}^s},$$

there exist positive constants  $C_1, C_2$  such that

$$\limsup_n \min(u_{n,1}^2, u_{n,2}) \sup_{x \in \mathbb{R}, \mathbf{z} \in \mathcal{Z}} |r_n(x, \mathbf{z})| \leq C_1 \tag{A.2}$$

and

$$\limsup_n \min(u_{n,1}, u_{n,2}) \sup_{x \in \mathbb{R}, \mathbf{z} \in \mathcal{Z}} |\hat{F}_k(x|\mathbf{z}) - F_k(x|\mathbf{z})| \leq C_2 \text{ a.e.}$$

**Proof.** Eq. (A.2) is deduced directly from Theorem 2 in [14]. Moreover, by straightforward calculations, we get

$$r_{n,k}(x, \mathbf{z}) = r_{n,k}^{(1)}(x, \mathbf{z}) + r_{n,k}^{(2)}(x, \mathbf{z}),$$

$$r_{n,k}^{(1)}(x, \mathbf{z}) = \frac{E\hat{h}(x, \mathbf{z})(\hat{g} - E\hat{g})^2(\mathbf{z})}{(E\hat{g})^2\hat{g}(\mathbf{z})} - \frac{(\hat{h} - E\hat{h})(x, \mathbf{z})(\hat{g} - E\hat{g})(\mathbf{z})}{\hat{g}(\mathbf{z})E\hat{g}(\mathbf{z})},$$

$$r_{n,k}^{(2)}(x, \mathbf{z}) = \frac{E\hat{h}(x, \mathbf{z})}{E\hat{g}(\mathbf{z})} - F(x_k|\mathbf{z}),$$

$$\hat{h}(x, \mathbf{z}) = \frac{1}{n} \sum_{j=1}^n \mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}) \mathbf{1}(X_{j,k} \leq x), \quad \hat{g}(\mathbf{z}) = \frac{1}{n} \sum_{j=1}^n \mathbf{K}(\mathbf{Z}_j - \mathbf{z}, \mathbf{h}),$$

that tends typically to  $g = f_{\mathbf{Z}}$  and  $h_k(x, \mathbf{z}) = \Pr(X_k \leq x|\mathbf{Z} = \mathbf{z})g(\mathbf{z})$ . By invoking Eqs. (3.7)–(3.8) in the proof of Theorem 2 in [14], we get the uniform convergence of  $\hat{h}$  (resp.  $\hat{g}$ ) towards  $E\hat{h}$  (resp.  $E\hat{g}$ ) almost surely, at the same rate  $u_n$ . Note their Remark 8 justifies the choice of different bandwidths for every component of  $\mathbf{Z}$ .

Moreover, by usual limited expansion of  $E\hat{g} - g$  and  $E\hat{h} - h$ , we can deal with the bias term. Due to our assumptions concerning the order of the kernel  $\mathbf{K}$  and the regularity conditions on the underlying laws, we obtain easily that  $r_{n,k}^{(2)}(x, \mathbf{z}) = O(\max_{\ell \in \{1, \dots, p\}} h_\ell^s)$ , providing the result.  $\square$

**Lemma 5.** Consider an integrable function  $\chi$  on  $(0, 1)^d \times \mathcal{Z}$ . Assume that there exist two deterministic sequences  $(\xi_n)$  and  $(\delta_n)$ ,  $\xi_n \rightarrow 0, \delta_n = o(v_n)$ , such that

$$\Pr \left( \max_{i \in \{1, \dots, n\}} |\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z} \right) \rightarrow 0$$

when  $n \rightarrow \infty$ , and

$$E \left[ |\chi(\mathbf{U}_i, \mathbf{Z}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) \{ \mathbf{1}(|U_{i,k} - v_n| \leq 2\delta_n) + \mathbf{1}(|1 - v_n - U_{i,k}| \leq 2\delta_n) \} \right] \leq \xi_n,$$

for all  $k \in \{1, \dots, d\}$ . Then  $\sum_{i=1}^n |\chi(\mathbf{U}_i, \mathbf{Z}_i) \cdot (\hat{\omega}_{i,n} - \omega_{i,n})|/n = O_P(\xi_n)$ .

**Proof.** Let us fix  $\varepsilon > 0$ . For any constant  $A > 0$ , we have

$$\Pr \left\{ \frac{1}{n} \sum_{i=1}^n |\chi(\mathbf{U}_i, \mathbf{Z}_i)| \times |\hat{\omega}_{i,n} - \omega_{i,n}| > A\xi_n \right\} \leq \Pr \left\{ \frac{1}{n} \sum_{i=1}^n |\chi(\mathbf{U}_i, \mathbf{Z}_i)| \times |\hat{\omega}_{i,n} - \omega_{i,n}| \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) > A\xi_n \right\} \\ + \Pr \left\{ \max_{i \in \{1, \dots, n\}} |\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z} \right\} \equiv P_1 + P_2.$$

First, we have

$$P_1 \leq \sum_{k=1}^d \Pr \left\{ \frac{1}{n} \sum_{i=1}^n |\chi(\mathbf{U}_i, \mathbf{Z}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}, |U_{i,k} - v_n| \leq |\hat{U}_{i,k} - U_{i,k}|) > A\xi_n/(2d) \right\} \\ + \sum_{k=1}^d \Pr \left\{ \frac{1}{n} \sum_{i=1}^n |\chi(\mathbf{U}_i, \mathbf{Z}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}, |1 - v_n - U_{i,k}| \leq |\hat{U}_{i,k} - U_{i,k}|) > A\xi_n/(2d) \right\} \\ \leq \frac{2d}{A\xi_n} \sum_{k=1}^d E \left[ |\chi(\mathbf{U}_i, \mathbf{Z}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) \{ \mathbf{1}(|U_{i,k} - v_n| \leq 2\delta_n) + \mathbf{1}(|1 - v_n - U_{i,k}| \leq 2\delta_n) \} \right] \\ \leq 2d/A,$$

which is less than  $\varepsilon > 0$  for  $A$  large enough. This means  $P_1 = O_P(\xi_n)$ . Second, by assumption,  $P_2$  is less than  $\varepsilon$  when  $n$  is sufficiently large, proving the result.  $\square$

**Remark 2.** In particular, it is tempting to define, with obvious notations,

$$\xi_n = \max_k \mathbb{E} \left[ \sup_{u_k, |u_k - v_n| \leq 2\delta_n} |\chi(u_k, \mathbf{U}_{i,-k}, \mathbf{Z}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) \right] + \max_k \mathbb{E} \left[ \sup_{u_k, |u_k - 1 + v_n| \leq 2\delta_n} |\chi(u_k, \mathbf{U}_{i,-k}, \mathbf{Z}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) \right],$$

or even, when this term tends to zero,

$$\begin{aligned} \xi_n &= \max_k \sup_{u_k, |u_k - v_n| \leq 2\delta_n} \sup_{\mathbf{u}_{-k} \in [v_n - 2\delta_n, 1 - v_n + 2\delta_n]^{d-1}} \sup_{\mathbf{z} \in \mathcal{Z}} |\chi(u_k, \mathbf{u}_{-k}, \mathbf{z})| \\ &\quad + \max_k \sup_{u_k, |u_k - 1 + v_n| \leq 2\delta_n} \sup_{\mathbf{u}_{-k} \in [v_n - 2\delta_n, 1 - v_n + 2\delta_n]^{d-1}} \sup_{\mathbf{z} \in \mathcal{Z}} |\chi(u_k, \mathbf{u}_{-k}, \mathbf{z})|. \end{aligned}$$

**Lemma 6.** Under the assumptions of [Theorem 2](#),

$$\sup_{\beta \in \mathcal{B}} |\nabla_{\beta}^2 M_n(\beta) - \nabla_{\beta}^2 M(\beta)| = o_P(1).$$

**Proof.** We have

$$\begin{aligned} \frac{n_i + 1}{n} \nabla_{\beta}^2 M_n(\beta) &= \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \nabla_{\theta}^2 (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta} \hat{\psi}_i \nabla_{\beta^{\top}} \hat{\psi}_i \hat{\omega}_{i,n} \equiv B_{1,n}(\beta) + B_{2,n}(\beta). \end{aligned}$$

Furthermore,

$$B_{1,n}(\beta) - \frac{1}{n} \sum_{i=1}^n \frac{\nabla_{\theta} c_{\theta}}{c_{\theta}}_{|\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n}$$

is equal to

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \left\{ \nabla_{\theta} (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) - \nabla_{\theta} (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\mathbf{U}_i) + \nabla_{\theta} (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\mathbf{U}_i) - \nabla_{\theta} (\ln c_{\theta})_{|\theta=\psi_i}(\mathbf{U}_i) \right\} \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \nabla_{\mathbf{u},\theta}^2 (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\mathbf{U}_i^*) \cdot (\hat{\mathbf{U}}_i - \mathbf{U}_i) + \nabla_{\theta,\theta}^2 (\ln c_{\theta})_{|\theta=\psi_i^*}(\mathbf{U}_i) \cdot (\hat{\psi}_i - \psi) \right\} \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n}, \end{aligned}$$

for some  $\mathbf{U}_i^*$  and  $\psi_i^*$  such that  $\|\mathbf{U}_i^* - \mathbf{U}_i\| < \|\hat{\mathbf{U}}_i - \mathbf{U}_i\|$ ,  $|\psi_i^* - \psi_i| < |\hat{\psi}_i - \psi_i|$ . From [Assumption 11](#), and using the same arguments as in the proof of [Theorem 1](#) (see the term  $T_2(\beta)$ ), we get

$$\sup_{\beta} \left| \frac{1}{n} \sum_{i=1}^n \nabla_{\mathbf{u},\theta}^2 (\ln c_{\theta})_{|\theta=\hat{\psi}_i}(\mathbf{U}_i^*) (\hat{\mathbf{U}}_i - \mathbf{U}_i) \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n} \right| = o_P(1).$$

From [Assumptions 9](#) and the uniform consistency of  $\hat{\psi}(\beta, \beta^{\top} \mathbf{z})$  (see [\(18\)](#)), we have

$$\sup_{\beta} \left| \frac{1}{n} \sum_{i=1}^n \nabla_{\theta,\theta}^2 (\ln c_{\theta})_{|\theta=\psi_i^*}(\mathbf{U}_i) (\hat{\psi}_i - \psi) \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n} \right| = o_P(1),$$

and we deduce

$$\sup_{\beta \in \mathcal{B}} \left| B_{1,n}(\beta) - \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta}^2 \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n} \right| = o_P(1),$$

Invoking [Assumption 6](#), Eq. [\(18\)](#), we get

$$\sup_{\beta \in \mathcal{B}} \left| B_{1,n}(\beta) - \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta}^2 \psi(\beta, \beta^{\top} \mathbf{Z}_i) \hat{\omega}_{i,n} \right| = o_P(1).$$

Since the score function is uniformly integrable ([Assumption 4](#)) and applying [Lemma 5](#) (or the dominated convergence theorem simply), we can replace  $\hat{\omega}_{i,n}$  by  $\omega_i$ . Therefore,  $\sup_{\beta} |B_{1,n}(\beta) - B_1(\beta)| = o_P(1)$ , with

$$B_1(\beta) = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta}^2 \psi(\beta, \beta^{\top} \mathbf{Z}_i) \omega_i.$$

Similarly, one can deduce from Assumptions 11 and 6 that  $\sup_{\beta} |B_{2,n}(\beta) - B_2(\beta)| = o_p(1)$ , with

$$B_2(\beta) = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta, \theta}^2 (\ln c_{\theta})|_{\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) \nabla_{\beta^{\top}} \psi(\beta, \beta^{\top} \mathbf{Z}_i) \omega_i.$$

From Assumption 10 and (B.2) and (B.3) in Assumption 9, we can apply Example 19.7 and Theorem 19.4 in [45] to deduce that

$$\sup_{\beta \in \mathcal{B}} |B_1(\beta) - E\{B_1(\beta)\} + B_2(\beta) - E\{B_2(\beta)\}| = o_p(1).$$

Since  $(n_i + 1)/n$  tends to  $\Pr(\mathbf{Z} \in \mathcal{Z})$  a.e. and  $\Sigma = [E\{B_1(\beta)\} + E\{B_2(\beta)\}]/\Pr(\mathbf{Z} \in \mathcal{Z})$ , we obtain the result.  $\square$

**Lemma 7.** Let  $c_0(\mathbf{u}, v)$  denote the first order partial derivative of  $C_{\beta_0}^M(\mathbf{u}|w)$  with respect to  $w$  evaluated at point  $w = v$ , where  $C_{\beta}^M(\mathbf{u}|w)$  denotes the conditional copula function of  $\mathbf{U}$  given  $\beta^{\top} \mathbf{Z}$  and  $\|\mathbf{Z}\|_{\infty} \leq M$  (i.e.,  $\mathbf{Z} \in \mathcal{Z}$ ). We have

$$\nabla_{\beta} C_{\beta}^M(\mathbf{u}|\beta^{\top} \mathbf{Z})|_{\beta=\beta_0} = c_0(\mathbf{u}, \beta_0^{\top} \mathbf{Z})\{\mathbf{Z} - E(\mathbf{Z}|\beta_0^{\top} \mathbf{Z}, \mathbf{Z} \in \mathcal{Z})\}.$$

**Proof.** The proof is similar to the proof of Lemma 5A in [11], and of Lemma 3.4 in [28]. Observe that

$$C_{\beta}^M(\mathbf{u}|\beta^{\top} \mathbf{Z}) = E(\mathbf{1}_{\mathbf{U} \leq \mathbf{u}}|\beta^{\top} \mathbf{Z}, \mathbf{Z} \in \mathcal{Z}) = E\{E(\mathbf{1}_{\mathbf{U} \leq \mathbf{u}}|\mathbf{Z})|\beta^{\top} \mathbf{Z}, \mathbf{Z} \in \mathcal{Z}\} = E\{C_{\beta_0}^M(\mathbf{u}|\beta_0^{\top} \mathbf{Z})|\beta^{\top} \mathbf{Z}, \mathbf{Z} \in \mathcal{Z}\},$$

where we used the single-index assumption. Next, let

$$\Gamma_{\mathbf{u}, \mathbf{Z}}(\beta_1, \beta_2) = E\{C_{\beta_0}^M(\mathbf{u}|\alpha(\mathbf{Z}, \beta_1) + \beta_2^{\top} \mathbf{Z})|\beta_2^{\top} \mathbf{Z}, \mathbf{Z} \in \mathcal{Z}\},$$

where  $\alpha(\mathbf{Z}, \beta_1) = \beta_0^{\top} \mathbf{Z} - \beta_1^{\top} \mathbf{Z}$ . Hence,  $C_{\beta}^M(\mathbf{u}|\beta^{\top} \mathbf{Z}) = \Gamma_{\mathbf{u}, \mathbf{Z}}(\beta, \beta)$ . As a consequence,

$$\nabla_{\beta} C_{\beta}^M(\mathbf{u}|\beta^{\top} \mathbf{Z})|_{\beta=\beta_0} = \nabla_1 \Gamma_{\mathbf{u}, \mathbf{Z}}(\beta, \beta_0)|_{\beta=\beta_0} + \nabla_2 \Gamma_{\mathbf{u}, \mathbf{Z}}(\beta_0, \beta)|_{\beta=\beta_0},$$

where  $\nabla_j$  represents the gradient vector with respect to  $\beta_j$ . Observe that

$$\nabla_1 \Gamma_{\mathbf{u}, \mathbf{Z}}(\beta, \beta_0)|_{\beta=\beta_0} = -E\{\mathbf{Z} c_0(\mathbf{u}, \beta_0^{\top} \mathbf{Z})|\beta_0^{\top} \mathbf{Z}\}.$$

Moreover,  $\Gamma_{\mathbf{u}, \mathbf{Z}}(\beta_0, \beta) = C_{\beta_0}^M(\mathbf{u}|\beta^{\top} \mathbf{Z})$ , which leads to  $\nabla_2 \Gamma_{\mathbf{u}, \mathbf{Z}}(\beta_0, \beta)|_{\beta=\beta_0} = \mathbf{Z} c_0(\mathbf{u}, \beta_0^{\top} \mathbf{Z})$ , and the result follows.  $\square$

**Lemma 8.** Assume that the transformation  $\Psi$  is Hadamard differentiable. Then, for all  $v$ ,

$$\int \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z})|_{\beta=\beta_0} d\Pr_{\mathbf{Z}|\beta_0^{\top} \mathbf{Z}}(\mathbf{z}|v) = 0.$$

**Proof.** Let  $\dot{\Psi}\{C(\cdot)\}[D(\cdot)]$  denote the Hadamard derivative of  $\Psi$  at point  $C$ , applied to function  $D$ . Recall that  $\psi(\beta, \beta^{\top} \mathbf{z}) = \Psi\{C_{\beta}^M(\cdot|\beta^{\top} \mathbf{z})\}$ . Hence, using Lemma 7,

$$\nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z})|_{\beta=\beta_0} = \{\mathbf{z} - E(\mathbf{Z}|\beta_0^{\top} \mathbf{Z} = \beta_0^{\top} \mathbf{z})\} \dot{\Psi}\{C_{\beta_0}^M(\cdot|\beta_0^{\top} \mathbf{z})\}[c_0(\cdot|\beta_0^{\top} \mathbf{z})].$$

This shows that  $\nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z})|_{\beta=\beta_0} = \{\mathbf{z} - E(\mathbf{Z}|\beta_0^{\top} \mathbf{Z} = \beta_0^{\top} \mathbf{z})\} \Lambda(\beta_0^{\top} \mathbf{z})$ , for some transform  $\Lambda$ , and the result of Lemma 8 follows.  $\square$

Finally, Lemma 14 invokes two propositions from [14], that we recall here.

**Proposition 9.** Let  $\mathcal{G}$  denote a class of functions bounded by 1, and let  $\sigma_{\mathcal{G}}^2 = \sup_{g \in \mathcal{G}} \text{var}\{g(\mathbf{X}, \mathbf{Z})\}$ . Then, for all  $t > 0$ ,

$$\Pr \left\{ \sup_{g \in \mathcal{G}} \left| \sum_{i=1}^n g(\mathbf{X}_i, \mathbf{Z}_i) - E\{g(\mathbf{X}_i, \mathbf{Z}_i)\} \right| \geq A_1(G_{\varepsilon} + t) \right\} \leq 2 \left\{ \exp\left(-\frac{A_2 t^2}{n \sigma_{\mathcal{G}}^2}\right) + \exp(-A_2 t) \right\},$$

for some universal constants  $A_1$  and  $A_2$ , and

$$G_{\varepsilon} = E \left\{ \sup_{g \in \mathcal{G}} \left| \sum_{i=1}^n g(\mathbf{X}_i, \mathbf{Z}_i) \varepsilon_i \right| \right\},$$

where  $(\varepsilon_i)_{1 \leq i \leq n}$  are iid Rademacher variables independent from  $(\mathbf{X}_1, \mathbf{Z}_1), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$ .

**Proposition 10.** Assume that  $\mathcal{G}$  is a class of functions satisfying the assumptions of Proposition 9 and such that  $N(\varepsilon, \mathcal{G}) \leq C \varepsilon^{-\nu}$  for  $C > 0$  and  $\nu > 0$ . Moreover, assume that there exists  $\sigma^2 \leq 1$  such that  $\sup_{g \in \mathcal{G}} E\{g(\mathbf{X}, \mathbf{Z})^2\} \leq \sigma^2$ . Then,  $G_{\varepsilon} \leq An^{1/2} \sigma \ln(1/\sigma)$ .

**Proposition 11.** For each  $k \in \{1, \dots, p\}$ , let  $\hat{F}_k(x|\mathbf{z})$  denote the kernel estimator of the conditional distribution function  $F_k(x|\mathbf{z})$  as given in Eq. (A.1), i.e.,  $\hat{F}_k(x|\mathbf{z}) = \hat{N}_k(x|\mathbf{z})/\hat{f}(\mathbf{z})$  with

$$\hat{N}_k(x|\mathbf{z}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{x_k \leq x} \mathbf{K}(\mathbf{Z}_i - \mathbf{z}, \mathbf{h}), \quad \hat{f}(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^n \mathbf{K}(\mathbf{Z}_i - \mathbf{z}, \mathbf{h}), \quad \mathbf{K}(\mathbf{Z}_i - \mathbf{z}, \mathbf{h}) = \frac{1}{h_1 \cdots h_p} \prod_{k=1}^p K_k\left(\frac{Z_{i,k} - z_k}{h_k}\right).$$

Define

$$P(t) = \Pr \left\{ \sup_{x \in \mathbb{R}, \mathbf{z} \in \mathcal{Z}} \left| \hat{F}_k(x|\mathbf{z}) - F_k(x|\mathbf{z}) \right| \geq t \right\},$$

and assume that

$$\sup_{x \in \mathbb{R}, \mathbf{z} \in \mathcal{Z}} \left| E\{\hat{N}_k(x|\mathbf{z})\}/E\{\hat{f}(\mathbf{z})\} - F_k(x|\mathbf{z}) \right| = b_n,$$

for some sequence  $b_n \rightarrow 0$ . Then, for  $B$  large enough and  $t \geq \max\{2b_n, Bn^{-1/2} \ln(1/\min_k h_k)(h_1 \cdots h_p)^{-1/2}\}$ , there exist positive constants  $(\alpha, \beta, \gamma, \delta)$  such that

$$P(t) \leq 4\{\exp(-\alpha n h_1 \cdots h_p t^2) + \exp(-\beta n h_1 \cdots h_p t) + \exp(-\gamma n h_1 \cdots h_p) + \exp(-\delta n h_1 \cdots h_p)\}.$$

**Proof.** Let

$$\tilde{P}(t) = \Pr \left\{ \sup_{x, \mathbf{z}} \left| \hat{F}_k(x|\mathbf{z}) - \frac{E\{\hat{N}_k(x|\mathbf{z})\}}{E\{\hat{f}(\mathbf{z})\}} \right| \geq t \right\}, \quad P_1(t) = \Pr \left\{ \sup_{x, \mathbf{z}} |\hat{N}_k(x|\mathbf{z}) - E\{\hat{N}_k(x|\mathbf{z})\}| \geq t \right\},$$

and

$$P_2(t) = \Pr \left\{ \sup_{\mathbf{z}} |\hat{f}(\mathbf{z}) - E\{\hat{f}(\mathbf{z})\}| \geq t \right\}.$$

We have

$$P(t) \leq \tilde{P}(t/2) + \Pr \left\{ \sup_{x, \mathbf{z}} \left| \frac{E\{\hat{N}_k(x|\mathbf{z})\}}{E\{\hat{f}(\mathbf{z})\}} - F_k(x|\mathbf{z}) \right| \geq t/2 \right\},$$

where the last probability is zero for  $t/2 \geq b_n$ . Hence, an upper bound for  $P(t)$  can be deduced from an upper bound on  $\tilde{P}(t)$ . Define the classes of functions

$$\mathcal{G}_1 = \{(x, \mathbf{z}) \in \mathbb{R} \times \mathcal{Z} \mapsto \mathbf{1}_{x' \leq x} \mathbf{K}(\mathbf{z} - \mathbf{z}', \mathbf{h}) : \mathbf{z}' \in \mathcal{Z}, \mathbf{h} = (h_1, \dots, h_p) \in \mathbb{R}_+^p, x' \in \mathbb{R}\},$$

$$\text{and } \mathcal{G}_2 = \{\mathbf{z} \in \mathcal{Z} \mapsto \mathbf{K}(\mathbf{z} - \mathbf{z}', \mathbf{h}) : \mathbf{z}' \in \mathcal{Z}, \mathbf{h} = (h_1, \dots, h_p) \in \mathbb{R}_+^p, x \in \mathbb{R}\}.$$

These two classes of functions satisfy the Assumption of Proposition 10 with  $\sigma^2 \propto \prod_{k=1}^p h_k$ . Hence, from Proposition 9, we get, for each  $j \in \{1, 2\}$ ,

$$P_j(t) \leq 2[\exp\{-Cn(h_1 \cdots h_p)t^2\} + \exp\{-Cn(h_1 \cdots h_p)t\}],$$

for  $t \geq An^{-1/2} \ln(1/\min_k h_k)/(h_1 \cdots h_p)^{1/2}$ , with  $A$  large enough and  $C > 0$  some constant. Decompose

$$\tilde{P}(t) \leq \Pr \left\{ \sup_{x, \mathbf{z}} \left| \frac{\hat{N}_k(x|\mathbf{z}) - E\{\hat{N}_k(x|\mathbf{z})\}}{E\{\hat{f}(\mathbf{z})\}} \right| \geq t/2 \right\} + \Pr \left\{ \sup_{x, \mathbf{z}} \left| \frac{\hat{N}_k(x|\mathbf{z})[\hat{f}(\mathbf{z}) - E\{\hat{f}(\mathbf{z})\}]}{\hat{f}(\mathbf{z})E\{\hat{f}(\mathbf{z})\}} \right| \geq t/2 \right\} \tag{A.3}$$

$$\leq \tilde{P}_1(t/2) + \tilde{P}_2(t/2). \tag{A.4}$$

Next, recall that  $\inf_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{z}) \geq f_0 > 0$ . Moreover, the bias of the kernel estimator of the density tends to 0 uniformly on  $\mathcal{Z}$ , i.e.,  $\sup_{\mathbf{z} \in \mathcal{Z}} |E\{\hat{f}(\mathbf{z})\} - f(\mathbf{z})| \rightarrow 0$  as  $n$  grows to infinity. This, combined with the bound obtained on  $P_2(t)$ , shows that  $\sup_{\mathbf{z} \in \mathcal{Z}} |\hat{f}(\mathbf{z}) - f(\mathbf{z})| \rightarrow 0$ . Hence,  $\inf_{\mathbf{z} \in \mathcal{Z}} E\{\hat{f}(\mathbf{z})\} > f_0/2$  for  $n$  large enough, which leads to

$$\tilde{P}_1(t/2) \leq P_1(tf_0/4), \tag{A.5}$$

and

$$\tilde{P}_2(t/2) \leq \Pr \left\{ \sup_{x, \mathbf{z}} |\hat{f}(\mathbf{z}) - E\{\hat{f}(\mathbf{z})\}| \geq f_0/4 \right\} + \Pr \left\{ \sup_{x, \mathbf{z}} |\hat{N}_k(x|\mathbf{z})[\hat{f}(\mathbf{z}) - E\{\hat{f}(\mathbf{z})\}]| \geq f_0^2 t/16 \right\}.$$

Since  $\sup_{x, \mathbf{z}} |E\{\hat{N}_k(x|\mathbf{z})\}| \leq \|\mathbf{K}\|_\infty < \infty$ , we have

$$\tilde{P}_2(t/2) \leq P_2(f_0/4) + P_2\{f_0^2 t/(16\|\mathbf{K}\|_\infty)\}. \tag{A.6}$$

Gathering (A.4), (A.5) and (A.6) leads to the result of the proposition.  $\square$



**Appendix B. Proofs of the main results**

Introduce the set of indicator functions

$$\mathcal{H} = \left\{ \mathbf{g} : [0, 1]^d \times \mathbb{R}^p \rightarrow [0, 1], (\mathbf{u}, \mathbf{z}) \mapsto \mathbf{1}(\mathbf{u} \in B_{\mathbf{a},\mathbf{b}}, \mathbf{z} \in \tilde{B}_{\mathbf{c},\mathbf{d}}), \right. \\ \left. \text{for some } B_{\mathbf{a},\mathbf{b}} = \prod_{k=1}^d [a_k, b_k] \subset [0, 1]^d \text{ and } \tilde{B}_{\mathbf{c},\mathbf{d}} = \prod_{k=1}^p [c_k, d_k] \subset \mathbb{R}^p \right\}.$$

Since all the subsets we consider in  $\mathcal{H}$  are boxes, it is simple to verify that  $\mathcal{H}$  is universally Donsker (for instance, see Example 2.6.1 and apply Lemma 2.6.17 in [46]).

*B.1. Proof of the consistency of  $\hat{\beta}$  (Theorem 1)*

For inference purposes and a given sample, the sample size that we use is actually  $\hat{n}_i = \sum_{i=1}^n \hat{\omega}_{i,n}$ . This random number is close to  $n_i = \sum_{i=1}^n \omega_{i,n}$ , the sample size if the  $\mathbf{U}_i$  were observable. Let us introduce

$$M_n(\beta) = \frac{1}{n_i + 1} \sum_{i=1}^n \hat{\omega}_{i,n} \ln c_{\hat{\psi}(\beta, \beta^\top \mathbf{z}_i)}(\hat{\mathbf{U}}_i), \\ M_n^*(\beta) = \frac{1}{n_i + 1} \sum_{i=1}^n \hat{\omega}_{i,n} \ln c_{\psi(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i), \\ M_n^{**}(\beta) = \frac{1}{n_i + 1} \sum_{i=1}^n \omega_i \ln c_{\psi(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i).$$

Note that  $\hat{\beta}$  is the optimizer of  $M_n$  because neither  $n_i$  nor  $\hat{n}_i$  is a function of the underlying parameter  $\beta$ . By assumption,  $\beta_0$  maximizes  $M(\beta)$  over  $\mathcal{B}$ . To prove the consistency of  $\hat{\beta}$ , it is sufficient to show that  $\sup_{\beta \in \mathcal{B}} |M_n(\beta) - M(\beta)| = o_p(1)$ .

We first show that  $\sup_{\beta \in \mathcal{B}} |M_n(\beta) - M_n^*(\beta)| = o_p(1)$ . Simple calculations yield

$$|M_n(\beta) - M_n^*(\beta)| \leq \frac{1}{n_i + 1} \sum_{i=1}^n \hat{\omega}_{i,n} \sup_{\theta \in \Theta} |\nabla_\theta \ln c_\theta(\hat{\mathbf{U}}_i)| \times |\hat{\psi}(\beta, \beta^\top \mathbf{z}_i) - \psi(\beta, \beta^\top \mathbf{z}_i)| \\ + \frac{1}{n_i + 1} \sum_{i=1}^n \left| \hat{\omega}_{i,n} \frac{\nabla_{\mathbf{u}} c_{\psi(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i^*)}{c_{\psi(\beta, \beta^\top \mathbf{z}_i)}}(\hat{\mathbf{U}}_i - \mathbf{U}_i) \right| \equiv T_1(\beta) + T_2(\beta),$$

for some vectors  $\mathbf{U}_i^*$  such that  $|\mathbf{U}_i - \mathbf{U}_i^*| \leq |\mathbf{U}_i - \hat{\mathbf{U}}_i|$  for all  $i \in \{1, \dots, n\}$ .

Let us deal with  $T_1(\beta)$ . By (4),  $\sup_{\beta} \max_{i \in \{1, \dots, n\}} |\hat{\psi}(\beta, \beta^\top \mathbf{z}_i) - \psi(\beta, \beta^\top \mathbf{z}_i)| = o_p(1)$ . Then, it suffices to prove that

$$\frac{1}{n_i + 1} \sum_{i=1}^n \hat{\omega}_{i,n} \sup_{\theta \in \Theta} |\nabla_\theta \ln c_\theta(\hat{\mathbf{U}}_i)| = O_p(1).$$

For every  $\varepsilon > 0$  and  $A > 0$ , we have

$$\Pr \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\omega}_{i,n} \sup_{\theta \in \Theta} |\nabla_\theta \ln c_\theta(\hat{\mathbf{U}}_i)| > A \right\} \leq \Pr \left( \max_{i \in \{1, \dots, n\}} |\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z} \right) \\ + \Pr \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\omega}_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \sup_{\theta \in \Theta} |\nabla_\theta \ln c_\theta(\hat{\mathbf{U}}_i)| > A \right\}.$$

By (5), the first term on the right-hand side above is less than  $\varepsilon$  when  $n$  is large. To manage the last term on the right-hand side, consider an arbitrary index  $i$  such that  $|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n$  and  $\hat{\omega}_{in} = 1$ . Since  $\delta_n = o(\nu_n)$ , we can assume that, for every  $k \in \{1, \dots, d\}$ , we have

$$U_{i,k}/2 \leq \hat{U}_{i,k} \text{ if } \hat{U}_{i,k} \leq 1/2, \quad \text{and} \quad (1 - U_{i,k})/2 \leq (1 - \hat{U}_{i,k}) \text{ if } \hat{U}_{i,k} > 1/2.$$

For  $r_k$ , the  $k$ th of the U-shaped functions that define  $r$ , we deduce

$$r_k(\hat{U}_{i,k}) \leq r_k(U_{i,k}/2) \text{ if } \hat{U}_{i,k} \leq 1/2, \quad \text{and} \quad r_k(\hat{U}_{i,k}) \leq r_k\{1 - (1 - U_{i,k})/2\} \text{ if } \hat{U}_{i,k} > 1/2.$$

In other words,  $r_k(\hat{U}_{i,k}) \leq r_{k,1/2}(U_{i,k})$  for such  $i$  and every  $k$ . Then, [Assumption 4](#) implies

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} |\nabla_{\theta} \ln c_{\theta}(\hat{\mathbf{U}}_i)| \hat{\omega}_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) &\leq \frac{1}{n} \sum_{i=1}^n r(\hat{\mathbf{U}}_i) \hat{\omega}_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \\ &\leq \frac{1}{n} \sum_{i=1}^n r_{1/2}(\mathbf{U}_i) \omega_i \leq \frac{M_{1/2}^d}{n} \sum_{i=1}^n r(\mathbf{U}_i) \omega_i, \end{aligned}$$

which is integrable. We get

$$\Pr \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\omega}_{i,n} \sup_{\theta \in \Theta} |\nabla_{\theta} \ln c_{\theta}(\hat{\mathbf{U}}_i)| > A \right\} \leq \varepsilon + \frac{M_{1/2}^d \mathbb{E}\{r(\mathbf{U}_i)\omega_i\}}{A} < 2\varepsilon,$$

when  $A$  and  $n$  are sufficiently large. Since  $n_i/n$  tends to a positive constant a.e., we deduce  $\sup_{\beta} T_1(\beta) = o_P(1)$ .

By a slightly more subtle reasoning, we can obtain  $\sup_{\beta} T_2 = o_P(1)$ . For every  $\varepsilon > 0$ ,

$$\begin{aligned} \Pr\{T_2(\beta) > \varepsilon\} &\leq \Pr \left( \max_{i \in \{1, \dots, n\}} |\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z} \right) \\ &\quad + \Pr \left\{ \frac{1}{n_i + 1} \sum_{i=1}^n \hat{\omega}_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \sup_{\beta} \left| \frac{\nabla_{\mathbf{u}} c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}(\mathbf{U}_i^*) (\hat{\mathbf{U}}_i - \mathbf{U}_i)}{c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}} \right| > \varepsilon \right\}, \end{aligned}$$

and the first term on the right-hand side is smaller than  $\varepsilon$  when  $n$  is large. Due to [Assumption 4](#) and for every  $\varepsilon > 0$ , there exists  $\eta \in (0, 1/2)$  such that

$$\max_{k \in \{1, \dots, d\}} \mathbb{E}[\tilde{r}_{k,1/2}(\mathbf{Z}) \{U_k \mathbf{1}(U_k < \eta) + (1 - U_k) \mathbf{1}(U_k > 1 - \eta)\} \mathbf{1}(\mathbf{Z} \in \mathcal{Z})] < \varepsilon^2.$$

By applying Markov's inequality, we deduce that

$$\begin{aligned} \Pr \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\omega}_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \sup_{\beta} \left| \frac{\nabla_{\mathbf{u}} c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}(\mathbf{U}_i^*) (\hat{\mathbf{U}}_i - \mathbf{U}_i)}{c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}} \right| > \varepsilon \right\} \\ \leq \Pr \left\{ \frac{1}{n} \sum_{k=1}^d \sum_{i=1}^n \hat{\omega}_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) |\tilde{r}_k(\mathbf{U}_i^*)| \times |\hat{U}_{i,k} - U_{i,k}| > \varepsilon \right\} \\ \leq \Pr \left\{ \frac{1}{n} \sum_{k=1}^d \sum_{i=1}^n \omega_i \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \tilde{r}_{k,1/2}(\mathbf{U}_i) \right. \\ \quad \left. \times \{|\hat{U}_{i,k} - U_{i,k}| \mathbf{1}(\eta \leq U_{i,k} \leq 1 - \eta) + U_{i,k} \mathbf{1}(U_{i,k} < \eta) + (1 - U_{i,k}) \mathbf{1}(U_{i,k} > 1 - \eta)\} > \varepsilon \right\} \\ \leq \Pr \left[ \frac{2\delta_n}{n} \sum_{k=1}^d \sum_{i=1}^n \omega_i \tilde{r}_{k,1/2}(\mathbf{U}_i) \mathbf{1}\{\eta \leq U_{i,k} \leq 1 - \eta\} > \varepsilon/2 \right] + \frac{2\varepsilon^2}{\varepsilon} \\ \leq \Pr \left\{ \frac{2\delta_n}{n\eta} \sum_{k=1}^d \sum_{i=1}^n \omega_i U_{i,k} (1 - U_{i,k}) \tilde{r}_{k,1/2}(\mathbf{U}_i) > \varepsilon/2 \right\} + 2\varepsilon, \end{aligned}$$

and then

$$\Pr\{T_2(\beta) > \varepsilon\} \leq 3\varepsilon + \frac{4d\delta_n \max_k \mathbb{E}\{\omega_i U_{i,k} (1 - U_{i,k}) \tilde{r}_{k,1/2}(\mathbf{U}_i)\}}{\eta\varepsilon},$$

that is less than  $4\varepsilon$  when  $n$  is sufficiently large, because of [Assumption 4](#). Note that we have used  $\hat{U}_{i,k} \in (0, 1)$  for every  $i \in \{1, \dots, n\}$  and  $k \in \{1, \dots, d\}$  above. Since  $\varepsilon$  may be arbitrarily small and  $n_i/n$  tends to a constant a.e., we get  $\sup_{\beta} T_2(\beta) = o_P(1)$ , and we have proved that  $\sup_{\beta \in \mathcal{B}} |M_n(\beta) - M_n^*(\beta)| = o_P(1)$ .

Second, due to [Assumption 2](#) and for every  $\varepsilon > 0$ , we have

$$\begin{aligned} \Pr \left\{ \sup_{\beta} \left| \frac{1}{n} \sum_{i=1}^n (\hat{\omega}_{i,n} - \omega_{i,n}) \ln c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}(\mathbf{U}_i) \right| > \varepsilon \right\} \\ \leq \Pr \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{U}_i \in \mathcal{E}_n, \hat{\mathbf{U}}_i \notin \mathcal{E}_n, \mathbf{Z}_i \in \mathcal{Z}) \sup_{\beta} |\ln c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}(\mathbf{U}_i)| > \varepsilon/2 \right\} \\ + \Pr \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{U}_i \notin \mathcal{E}_n, \hat{\mathbf{U}}_i \in \mathcal{E}_n, \mathbf{Z}_i \in \mathcal{Z}) \sup_{\beta} |\ln c_{\psi(\beta, \beta^{\top} \mathbf{z}_i)}(\mathbf{U}_i)| > \varepsilon/2 \right\} \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{2}{\varepsilon} E[\{\mathbf{1}(\mathbf{U}_i \in \mathcal{E}_n, \hat{\mathbf{U}}_i \notin \mathcal{E}_n) + \mathbf{1}(\mathbf{U}_i \notin \mathcal{E}_n, \hat{\mathbf{U}}_i \in \mathcal{E}_n)\} \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}) h(\mathbf{U}_i, \mathbf{Z}_i)] \\
 &\leq \frac{4}{\varepsilon} E[\mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z}) h(\mathbf{U}_i, \mathbf{Z}_i)] \\
 &\quad + \frac{2}{\varepsilon} E[\{\mathbf{1}(\mathbf{U}_i \in \mathcal{E}_n, \hat{\mathbf{U}}_i \notin \mathcal{E}_n) + \mathbf{1}(\mathbf{U}_i \notin \mathcal{E}_n, \hat{\mathbf{U}}_i \in \mathcal{E}_n)\} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n, \mathbf{Z}_i \in \mathcal{Z}) h(\mathbf{U}_i, \mathbf{Z}_i)].
 \end{aligned} \tag{B.1}$$

But, we have for any  $i$ ,

$$\begin{aligned}
 &\mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \{\mathbf{1}(\mathbf{U}_i \notin \mathcal{E}_n, \hat{\mathbf{U}}_i \in \mathcal{E}_n) + \mathbf{1}(\mathbf{U}_i \in \mathcal{E}_n, \hat{\mathbf{U}}_i \notin \mathcal{E}_n)\} \\
 &\leq 2 \sum_{k=1}^d \{\mathbf{1}(U_{i,k} \in [v_n - 2\delta_n, v_n + 2\delta_n]) + \mathbf{1}(1 - U_{i,k} \in [v_n - 2\delta_n, v_n + 2\delta_n])\},
 \end{aligned}$$

that tends to zero a.e. when  $n \rightarrow \infty$ . Invoking the Dominated Convergence Theorem and (3), this proves that the second term on the right-hand side of (B.1) is less than  $\varepsilon$  when  $n$  is large enough.

Moreover, due to Assumption 2 and Hölder’s inequality,

$$E\{\mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z}) h(\mathbf{U}_i, \mathbf{Z}_i)\} \leq E\{h(\mathbf{U}, \mathbf{Z})^\alpha \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\}^{1/\alpha} \Pr\{|\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z}\}^{1-1/\alpha},$$

that is less than  $\varepsilon$  when  $n$  is large enough (Assumption 5). This yields

$$\sup_{\beta} \left| \frac{1}{n} \sum_{i=1}^n (\hat{\omega}_{i,n} - \omega_{i,n}) \ln c_{\psi(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i) \right| = o_P(1).$$

Similarly, we prove  $\sup_{\beta} |\sum_{i=1}^n (\omega_{i,n} - \omega_i) \ln c_{\psi(\beta, \beta^\top \mathbf{z}_i)}(\mathbf{U}_i)|/n = o_P(1)$ . We deduce easily that  $\sup_{\beta \in \mathcal{B}} |M_n^*(\beta) - M_n^{**}(\beta)| = o_P(1)$  because  $n_i/n$  tends to a constant a.e.

To conclude the proof, we can apply a usual Uniform Law of Large Numbers. For instance, Lemma 2.4 in [30] tells us that (3) insures that  $\sup_{\beta \in \mathcal{B}} |M_n^{**}(\beta) - M(\beta)| = o_P(1)$ . Therefore, we get that  $\beta$  tends to  $\beta_0$  in probability.  $\square$

### B.2. Proof of Theorem 2

First, we need to introduce some additional assumptions. They require that the functions we manipulate (or their estimates) are “sufficiently regular”: to be Lipschitz, upper bounded by U-shaped functions, or a member of some Donsker classes of functions, etc. Since we are managing several indices ( $\theta, \mathbf{u}, \mathbf{z}$  and  $\beta$ ), the uniformity with respect to some of them is most often imposed.

**Assumption 7.** Let the functions on  $(0, 1)^d \times \mathcal{Z}$  be defined by

$$f(\mathbf{u}, \mathbf{z}) = \nabla_{\theta} \ln c_{\theta|\beta=\psi(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}) \quad \text{and} \quad \hat{f}(\mathbf{u}, \mathbf{z}) = \nabla_{\theta} \ln c_{\theta|\beta=\hat{\psi}(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}).$$

For almost every realization, the functions  $f$  and  $\hat{f}$  belong to a Donsker class for the underlying law of  $(\mathbf{X}, \mathbf{Z})$ , that is denoted  $\mathcal{F}_1$ .

**Assumption 8.** Let the functions on  $\mathcal{Z}$  be defined by

$$p : \mathbf{z} \mapsto p(\mathbf{z}) = \nabla_{\beta} \psi(\beta, \beta^\top \mathbf{z})|_{\beta=\beta_0} \quad \text{and} \quad \hat{p} : \mathbf{z} \mapsto \hat{p}(\mathbf{z}) = \nabla_{\beta} \hat{\psi}(\beta, \beta^\top \mathbf{z})|_{\beta=\beta_0}.$$

For almost every realization, the functions  $p$  and  $\hat{p}$  belong to a Donsker class for the underlying law of  $(\mathbf{X}, \mathbf{Z})$ , that is denoted  $\mathcal{F}_2$ .

**Assumption 9.** Assume that  $E\{\sup_{\theta \in \Theta} |\nabla_{\theta}^2 \ln c_{\theta}(\mathbf{U}_{\mathbf{Z}})| \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty$ . Moreover, for every  $(\mathbf{u}, \mathbf{u}') \in (0, 1)^{2d}$ , we have

$$|\nabla_{\theta} \ln c_{\theta}(\mathbf{u}) - \nabla_{\theta} \ln c_{\theta'}(\mathbf{u})| \leq \xi(\mathbf{u}) |\theta - \theta'|, \tag{B.2}$$

$$|\nabla_{\theta}^2 \ln c_{\theta}(\mathbf{u}) - \nabla_{\theta}^2 \ln c_{\theta'}(\mathbf{u})| \leq \xi(\mathbf{u}) |\theta - \theta'|, \tag{B.3}$$

for some function  $\Phi$  such that  $E\{\xi(\mathbf{U})\} < \infty$ . Moreover, there exists a function  $r_3$  in  $\mathcal{R}_d$  such that, for every  $\mathbf{u} \in (0, 1)^d$ ,

$$\sup_{\theta \in \Theta} |\nabla_{\theta}^3 \ln c_{\theta}(\mathbf{u})| \leq r_3(\mathbf{u}), \quad E\{r_3(\mathbf{U}_{\mathbf{Z}}) \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty.$$

**Assumption 10.** Assume that, for every  $(\beta_1, \beta_2) \in \mathcal{B}^2$  and  $j \in \{1, 2\}$ , there exists a finite constant  $C$  such that

$$\sup_{\mathbf{z} \in \mathcal{Z}} |\nabla_{\beta}^j \psi(\beta_1, \beta_1^\top \mathbf{z}) - \nabla_{\beta}^j \psi(\beta_2, \beta_2^\top \mathbf{z})| \leq C |\beta_1 - \beta_2|.$$

**Assumption 11.** For every  $k \in \{1, \dots, d\}$ , there exists a function  $\Gamma_k \in \mathcal{R}_d$  and some  $\alpha \in [0, 1)$  such that

$$\sup_{\theta \in \Theta} |\partial_{u_k} \nabla_{\theta}(\ln c_{\theta})(\mathbf{u})| + \sup_{\theta \in \Theta} |\partial_{u_k} \nabla_{\theta}^2(\ln c_{\theta})(\mathbf{u})| \leq \Gamma_k(\mathbf{u}), \quad E\{U_k^{\alpha}(1 - U_k)^{\alpha} \Gamma_k(\mathbf{U}_{\mathbf{Z}}) \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty.$$

The next assumption is less intuitive and highly technical. It is required to control for the behavior of (unobservable)  $\mathbf{U}_i$  close to the boundary of  $[0, 1]^d$ , i.e., when  $n$  is large and the trimming function “fills the hypercube”.

**Assumption 12.** For any  $\mathbf{u} \in \mathbb{R}^d$ , set

$$g(\mathbf{u}, \mathbf{z}) = \sup_{\theta \in B(\theta_0(\mathbf{z}), \eta_{1n})} \sup_{\mathbf{v} \in B(\mathbf{u}, 2\delta_n)} |\nabla_{\theta} \ln c_{\theta}(\mathbf{v})|,$$

where  $B(\mathbf{u}, \delta)$  (resp.  $B(\theta, \eta)$ ) denotes the closed ball of center  $\mathbf{u}$  (resp.  $\theta$ ) and radius  $\delta$  (resp.  $\eta$ ). Assume

$$\sup_{k \in \{1, \dots, d\}} E\{g(\mathbf{U}_i, \mathbf{Z}_i) \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z}, |U_{i,k} - v_n| < 2\delta_n)\} = o(n^{-1/2}), \tag{B.4}$$

and similarly after having replaced  $v_n$  by  $1 - v_n$ .

The latter assumption is usually satisfied with a lot of usual copula models. Broadly speaking and when  $c_{\theta}$  is continuous with respect to its arguments and  $\theta$  itself, it means that

$$\delta_n \int |\nabla_{\theta} c_{\theta}(\mathbf{u}_{-k}, v_n | \mathbf{z})|_{\theta=\theta_0(\mathbf{z})} \mathbf{1}(\mathbf{z} \in \mathcal{Z}) d\mathbf{u}_{-k} d\text{Pr}_{\mathbf{Z}}(\mathbf{z}) = o(n^{-1/2}),$$

and the same replacing  $v_n$  by  $1 - v_n$ . We have denoted by  $(\mathbf{u}_{-k}, v_n)$  the  $d$ -dimensional vector whose components are  $u_j$ , when  $j \neq k$ , and whose  $k$ th component is  $v_n$ .

Second, let us prove [Theorem 2](#). By definition of  $\hat{\beta}$ ,  $\nabla_{\beta} M_n(\hat{\beta}) = 0$ . Next, a first order Taylor expansion leads to  $-\nabla_{\beta} M_n(\beta_0) = (\hat{\beta} - \beta_0) \nabla_{\beta}^2 M_n(\tilde{\beta})$ , where  $\tilde{\beta} = \beta_0 + o_p(1)$ , using the consistency of  $\hat{\beta}$ . From [Lemma 6](#), we have  $\nabla_{\beta}^2 M_n(\tilde{\beta}) = \nabla_{\beta}^2 M(\tilde{\beta}) + o_p(1)$ . Moreover, from [Assumption 5](#) and the consistency of  $\hat{\beta}$  (hence the consistency of  $\tilde{\beta}$ ), we get  $\nabla_{\beta}^2 M_n(\tilde{\beta}) = \Sigma + o_p(1)$ . Next, we have

$$\nabla_{\beta} M_n(\beta_0) = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta} \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \hat{\omega}_{i,n}.$$

**A. From the trimming functions  $\hat{\omega}_{i,n}$  to  $\omega_{i,n}$ .** Under [Assumptions 3](#) and [12](#), we can apply [Lemma 5](#) with the function

$$\chi(\mathbf{U}_i, \mathbf{Z}_i) = \sup_{\theta \in B_{i,\theta}} \sup_{\mathbf{v} \in B_{i,d}} |\nabla_{\theta} \ln c_{\theta}(\mathbf{v})| \sup_{\beta \in \mathcal{B}} \sup_{\mathbf{z} \in \mathcal{Z}} |\nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z})|$$

and  $B_{i,\theta} = B(\theta_0(\mathbf{Z}_i), \eta_{1n})$  and  $B_{i,d} = B(\mathbf{U}_i, 2\delta_n)$ . This implies

$$\nabla_{\beta} M_n(\beta_0) = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) \nabla_{\beta} \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} + o_p(n^{-1/2}).$$

Now, decompose  $\nabla_{\beta} M_n(\beta_0) \equiv A_{1n} + A_{2n} + R_{1n} + R_{2n} + R_{3n}$ , where

$$\begin{aligned} A_{1n} &= \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n}, \\ A_{2n} &= \frac{1}{n} \sum_{i=1}^n \{ \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) - \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \} \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n}, \\ R_{1n} &= \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) \{ \nabla_{\beta} \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} - \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \} \omega_{i,n}, \\ R_{2n} &= \frac{1}{n} \sum_{i=1}^n \{ \nabla_{\theta} \ln c_{\theta|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) - \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) \} \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n}, \end{aligned}$$

and

$$R_{3n} = \frac{1}{n} \sum_{i=1}^n \{ \nabla_{\theta} \ln c_{\theta|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) - \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) \} \times \{ \nabla_{\beta} \hat{\psi}(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} - \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \} \omega_{i,n}.$$

In this decomposition, we show that only the first two terms ( $A_{1n}$  and  $A_{2n}$ ) matter, and that for each  $j \in \{1, 2, 3\}$ , the term  $R_{jn}$  is  $o_p(n^{-1/2})$ .

**B. Study of  $R_{1n}$ .** First observe that

$$R_{1n} = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \{\hat{p}(\mathbf{Z}_i) - p(\mathbf{Z}_i)\} \omega_{i,n} + R'_{1n},$$

$$R'_{1n} = \frac{1}{n} \sum_{i=1}^n \{\nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) - \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i)\} \{\hat{p}(\mathbf{Z}_i) - p(\mathbf{Z}_i)\} \omega_{i,n}.$$

By a limited expansion, we have

$$R'_{1n} = \frac{1}{n} \sum_{i=1}^n \{\nabla_{\mathbf{u}} \nabla_{\theta} (\ln c_{\theta})_{\theta=\psi_i}(\mathbf{U}_i^*) (\hat{\mathbf{U}}_i - \mathbf{U}_i)\} \{\hat{p}(\mathbf{Z}_i) - p(\mathbf{Z}_i)\} \omega_{i,n},$$

for some  $\mathbf{U}_i^*$  such that  $|\mathbf{U}_i^* - \mathbf{U}_i| < |\hat{\mathbf{U}}_i - \mathbf{U}_i|$ . Reasoning as in the proof of [Theorem 1](#), we can write

$$\begin{aligned} \Pr(n^{1/2}|R'_{1n}| > \varepsilon) &\leq \Pr\left(\max_{i \in \{1, \dots, n\}} |\hat{\mathbf{U}}_i - \mathbf{U}_i| > 2\delta_n, \mathbf{Z}_i \in \mathcal{Z}\right) \\ &\quad + \Pr\left\{\frac{1}{n^{1/2}} \sum_{i=1}^n |\nabla_{\mathbf{u},\theta}^2 (\ln c_{\theta})_{\theta=\psi_i}(\mathbf{U}_i^*)| |\hat{\mathbf{U}}_i - \mathbf{U}_i| \omega_{i,n} |\hat{p}(\mathbf{Z}_i) - p(\mathbf{Z}_i)| \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) > \varepsilon\right\} \\ &\leq \varepsilon + \Pr\left\{\frac{\eta_{1n}}{n^{1/2}} \sum_{i=1}^n \sum_{k=1}^d |\Gamma_{k,1/2}(\mathbf{U}_i) t| |\hat{U}_{i,k} - U_{i,k}| \omega_{i,n} \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) > \varepsilon\right\}, \end{aligned}$$

for  $n$  large enough and invoking [Assumption 6](#). Note that

$$|\hat{U}_{i,k} - U_{i,k}| \mathbf{1}(|\hat{\mathbf{U}}_i - \mathbf{U}_i| \leq 2\delta_n) \omega_{i,n} \leq C_{\alpha} U_{i,k}^{\alpha} (1 - U_{i,k})^{\alpha} |\hat{U}_{i,k} - U_{i,k}|^{1-\alpha}, \text{ a.e.}$$

for some constant  $C_{\alpha}$ , when  $n$  is sufficiently large,  $i \in \{1, \dots, n\}$  and  $k \in \{1, \dots, d\}$ . This yields

$$\begin{aligned} \Pr(n^{1/2}|R'_{1n}| > \varepsilon) &\leq \varepsilon + \Pr\left\{\frac{C'_{\alpha} \eta_{1n} \delta_n^{1-\alpha}}{n^{1/2}} \sum_{i=1}^n \sum_{k=1}^d \Gamma_{k,1/2}(\mathbf{U}_i) U_{i,k}^{\alpha} (1 - U_{i,k})^{\alpha} \omega_i > \varepsilon\right\} \\ &\leq \varepsilon + \frac{dC'_{\alpha} n^{1/2} \eta_{1n} \delta_n^{1-\alpha}}{\varepsilon} \max_k E\{\Gamma_{k,1/2}(\mathbf{U}_i) U_{i,k}^{\alpha} (1 - U_{i,k})^{\alpha} \omega_i\} \end{aligned}$$

for some constant  $C'_{\alpha}$ . Thanks to [Assumption 11](#),  $\Pr(n^{1/2}|R'_{1n}| > \varepsilon) < 2\varepsilon$  when  $n$  is large enough, implying  $R'_{1n} = o_p(n^{-1/2})$ . Moreover, with obvious notations,  $R_{1n}$  can be rewritten as

$$R_{1n} = \frac{1}{n} \sum_{i=1}^n \{\tilde{g}_n(\mathbf{X}_i, \mathbf{Z}_i) - \tilde{g}(\mathbf{X}_i, \mathbf{Z}_i)\} \omega_{i,n} + R'_{1n},$$

where  $\tilde{g}_n$  and  $\tilde{g}$  both belong to  $\mathcal{F}_3 = \mathcal{F}_1 \mathcal{F}_2 \mathcal{H}$ , which is a Donsker class of functions. Indeed, the fact that  $\mathcal{F}_3$  is a Donsker class follows from the permanence properties of Examples 2.10.10 and 2.10.7 in [\[46\]](#). Moreover, from [Assumption 6](#),

$$\sup_{\mathbf{x} \in \mathbb{R}^d, \mathbf{z} \in \mathcal{Z}} |\tilde{g}_n(\mathbf{x}, \mathbf{z}) - \tilde{g}(\mathbf{x}, \mathbf{z})| = o_p(1).$$

Therefore, the asymptotic equicontinuity of Donsker classes (see Section 2.1.2 in [\[46\]](#)) yields,

$$R_{1n} = \int \frac{\nabla_{\theta} c_{\theta}}{c_{\theta}} \Big|_{\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{z})}(\mathbf{u}) \{\hat{p}(\mathbf{z}) - p(\mathbf{z})\} \omega_n(\mathbf{u}, \mathbf{z}) d\Pr_{(\mathbf{U}, \mathbf{Z})}(\mathbf{u}, \mathbf{z}) + o_p(n^{-1/2}).$$

We can replace  $\omega_n(\mathbf{u}, \mathbf{z})$  above by  $\mathbf{1}(\mathbf{z} \in \mathcal{Z})$  if

$$\eta_{2n} \int |\nabla_{\theta} c_{\theta}(\mathbf{u})|_{\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{z})} \times |\omega_n(\mathbf{u}, \mathbf{z}) - \omega_{\infty}(\mathbf{u}, \mathbf{z})| d\mathbf{u} d\Pr_{\mathbf{Z}}(\mathbf{z}) = o(n^{-1/2}).$$

This is guaranteed under [Assumption 12](#). Then, under our assumptions, we can apply Fubini's theorem. This yields

$$\begin{aligned} &\int \nabla_{\theta} \ln c_{\theta|\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{z})}(\mathbf{u}) \{\hat{p}(\mathbf{z}) - p(\mathbf{z})\} \mathbf{1}(\mathbf{z} \in \mathcal{Z}) d\Pr_{(\mathbf{U}, \mathbf{Z})}(\mathbf{u}, \mathbf{z}) \\ &= \int \{\hat{p}(\mathbf{z}) - p(\mathbf{z})\} d\Pr_{\mathbf{Z}}(\mathbf{z}) \left\{ \int \nabla_{\theta} \ln c_{\theta|\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{z})}(\mathbf{u}) \mathbf{1}(\mathbf{z} \in \mathcal{Z}) d\Pr_{(\mathbf{U}|\mathbf{Z}=\mathbf{z})}(\mathbf{u}) \right\} = 0, \end{aligned}$$

by definition of  $\psi(\beta_0, \beta_0^\top \mathbf{z})$ , which maximizes  $E\{\ln c_\theta(\mathbf{U}_z) | \mathbf{Z} = \mathbf{z}\}$  with respect to  $\theta$ , for any  $\mathbf{z} \in \mathcal{Z}$ . This shows that  $R_{1n} = o_P(n^{-1/2})$ , and is therefore negligible.

**C. Study of  $R_{2n}$ .** Write, from [Assumption 11](#) and with obvious notations,

$$R_{2n} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\nabla_\theta c_\theta}{c_\theta} \Big|_{\theta=\hat{\psi}_i}(\mathbf{U}_i) - \frac{\nabla_\theta c_\theta}{c_\theta} \Big|_{\theta=\psi_i}(\mathbf{U}_i) \right\} \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} + R'_{2n}, \tag{B.5}$$

where

$$\begin{aligned} R'_{2n} &= \frac{1}{n} \sum_{i=1}^n \{ \nabla_\theta \ln c_{\theta|\theta=\hat{\psi}_i}(\hat{\mathbf{U}}_i) - \nabla_\theta \ln c_{\theta|\theta=\psi_i}(\hat{\mathbf{U}}_i) \} \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \{ \nabla_\theta \ln c_{\theta|\theta=\hat{\psi}_i}(\mathbf{U}_i) - \nabla_\theta \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \} \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} \\ &= \frac{1}{n} \sum_{i=1}^n \{ \nabla_\theta^2(\ln c_\theta) |_{\theta=\psi_i}(\hat{\mathbf{U}}_i) - \nabla_\theta^2(\ln c_\theta) |_{\theta=\psi_i}(\mathbf{U}_i) \} (\hat{\psi}_i - \psi_i) \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} \\ &\quad + \frac{1}{2n} \sum_{i=1}^n \{ \nabla_\theta^3(\ln c_\theta) |_{\theta=\psi_i^*}(\hat{\mathbf{U}}_i) \nabla_\theta^3(\ln c_\theta) |_{\theta=\psi_i}(\mathbf{U}_i) \} (\hat{\psi}_i - \psi_i)^{(2)} \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} \\ &= \frac{1}{n} \sum_{i=1}^n \nabla_{\mathbf{u}} \nabla_\theta^2(\ln c_\theta) |_{\theta=\psi_i}(\mathbf{U}_i^*) (\hat{\mathbf{U}}_i - \mathbf{U}_i) (\hat{\psi}_i - \psi_i) \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} \\ &\quad + O_P \left\{ \max_{i \in \{1, \dots, n\}} |\hat{\psi}_i - \psi_i|^2 \right\}, \end{aligned}$$

for some  $\mathbf{U}_i^*, \psi_i^*$  and  $\tilde{\psi}_i$  such that  $|\mathbf{U}_i^* - \mathbf{U}_i| < |\hat{\mathbf{U}}_i - \mathbf{U}_i|$ ,  $|\psi_i^* - \psi_i| < |\hat{\psi}_i - \psi_i|$  and  $|\tilde{\psi}_i - \psi_i| < |\hat{\psi}_i - \psi_i|$ . Note that we have invoked [Assumption 9](#) to bound the last term on the right-hand side in probability. The main term on the right-hand side is  $O_P(\eta_{1n} \delta_n^{1-\alpha}) = o_P(n^{-1/2})$  from [Assumptions 11](#) and [6](#) (mimic the treatment of  $R'_{1n}$  as above). We deduce  $R'_{2n} = o_P(n^{-1/2})$ .

Next, invoking [Assumptions 6](#) and [11](#), the first term on the right-hand side of [\(B.5\)](#) can be rewritten as

$$\frac{1}{n} \sum_{i=1}^n \{ h_n(\mathbf{U}_i, \mathbf{Z}_i) - h(\mathbf{U}_i, \mathbf{Z}_i) \} \omega_{i,n},$$

where  $\sup_{\mathbf{u}, \mathbf{z}} |h_n(\mathbf{u}, \mathbf{z}) - h(\mathbf{u}, \mathbf{z})| = o_P(1)$ , and  $h_n$  and  $h$  both belong to  $\mathcal{F}_4 = p.\mathcal{H}.\mathcal{F}_1$ , as a consequence of [Assumption 7](#). This is a Donsker class from Example 2.10.10 in [\[46\]](#). The asymptotic equicontinuity of the Donsker class  $\mathcal{F}_4$  allows to write

$$R_{2n} = \int \{ \nabla_\theta \ln c_{\theta|\theta=\hat{\psi}(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}) - \nabla_\theta \ln c_{\theta|\theta=\psi(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}) \} \times \nabla_\beta \psi(\beta, \beta^\top \mathbf{z}) |_{\beta=\beta_0} \omega_n(\mathbf{u}, \mathbf{z}) d\Pr_{(\mathbf{U}, \mathbf{Z})}(\mathbf{u}, \mathbf{z}) + o_P(n^{-1/2}).$$

Decompose  $\omega_n(\mathbf{u}, \mathbf{z})$  as  $\omega_v(\mathbf{u})\omega_M(\mathbf{z})$ , where  $\omega_v(\mathbf{u}) = \mathbf{1}_{\min_k \min(1-u_k, u_k) \geq v_n}$ , and  $\omega_M(\mathbf{z}) = \mathbf{1}_{|\mathbf{z}| \leq M}$ . The function

$$\phi_n(\mathbf{z}) = \int \left\{ \frac{\nabla_\theta c_\theta}{c_\theta} \Big|_{\theta=\hat{\psi}(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}) - \frac{\nabla_\theta c_\theta}{c_\theta} \Big|_{\theta=\psi(\beta_0, \beta_0^\top \mathbf{z})}(\mathbf{u}) \right\} \omega_v(\mathbf{u}) d\Pr_{(\mathbf{U} | \mathbf{Z}=\mathbf{z})}(\mathbf{u}),$$

is a function of  $\beta_0^\top \mathbf{z}$  only. This is due to the fact that the distribution of  $\mathbf{U}$  given  $\mathbf{Z}$  only depends on  $\beta_0^\top \mathbf{Z}$ , because of the single-index assumption. With a slight abuse in notations, denote  $\phi_n(\mathbf{z}) = \phi_n(\beta_0^\top \mathbf{z})$ . This leads to

$$R_{2n} = \int \phi_n(v) \left\{ \int \nabla_\beta \psi(\beta, \beta^\top \mathbf{z}) |_{\beta=\beta_0} \omega_M(\mathbf{z}) d\Pr_{(\mathbf{Z} | \beta_0^\top \mathbf{Z}=v)}(\mathbf{z} | v) \right\} d\Pr_{\beta_0^\top \mathbf{Z}}(v) + o_P(n^{-1/2}).$$

Next, as a consequence of [Lemma 8](#), use that

$$\int \nabla_\beta \psi(\beta, \beta^\top \mathbf{z}) |_{\beta=\beta_0} \omega_M(\mathbf{z}) d\Pr_{(\mathbf{Z} | \beta_0^\top \mathbf{Z}=v)}(\mathbf{z}) = 0.$$

This implies  $R_{2n} = o_P(n^{-1/2})$ .

**D. Study of  $R_{3n}$ .** By the same reasoning as for  $R_{2n}$ , we get

$$R_{3n} = \frac{1}{n} \sum_{i=1}^n \{ \nabla_\theta \ln c_{\theta|\theta=\hat{\psi}_i}(\mathbf{U}_i) - \nabla_\theta \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \} \times \{ \nabla_\beta \hat{\psi}(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} - \nabla_\beta \psi(\beta, \beta^\top \mathbf{Z}_i) |_{\beta=\beta_0} \} \omega_{i,n} + o_P(n^{-1/2}).$$

Due to [Assumptions 6](#) and [9](#) (see Eq. [\(B.2\)](#)), we obtain  $R_{3n} = o_P(n^{-1/2})$ .  $\square$

B.3. Proof of the asymptotic normality of  $\hat{\beta}$  (Corollary 3)

First, we need to introduce the way we estimate  $\mathbf{U}_i$  by pseudo-observations  $\hat{\mathbf{U}}_i$ . Again, additional technical assumptions are required.

**Assumption 13.** For every  $k \in \{1, \dots, d\}$ , there exists a function  $\zeta_k \in \mathcal{R}_d$  (see Definition 1) such that

$$\sup_{\theta \in \Theta} \left| \partial_{u_k}^2 \nabla_{\theta} (\ln c_{\theta})(\mathbf{u}) \right| \leq \zeta_k(\mathbf{u}) \quad \text{and} \quad \mathbb{E}\{U_k^{\gamma} (1 - U_k)^{\gamma} \zeta_k(\mathbf{U}_{\mathbf{Z}}) \mathbf{1}(\mathbf{Z} \in \mathcal{Z})\} < \infty,$$

for some  $\gamma \in [0, 1]$ . Moreover,  $\delta_n^{2-\gamma} = o(n^{-1/2})$ .

The latter assumption is of the same type as Assumption 11. Now, we impose that the estimated conditional margins can be rewritten as i.i.d. expansions, a rather light requirement in general.

**Assumption 14.** For every  $k \in \{1, \dots, d\}$ ,  $x \in \mathbb{R}$  and  $\mathbf{z} \in \mathcal{Z}$ , we can write

$$\hat{F}_k(x|\mathbf{z}) - F_k(x|\mathbf{z}) = \frac{1}{n} \sum_{j=1}^n a_{k,n}(\mathbf{X}_j, \mathbf{Z}_j, x, \mathbf{z}) + r_n(x, \mathbf{z}),$$

for some particular functions  $a_{k,n}$  and for some sequence  $(r_n)$  such that

$$\sup_{x \in \mathbb{R}} \sup_{\mathbf{z} \in \mathcal{Z}} |r_n(x, \mathbf{z})| = r_{n,\infty} = o_P(n^{-1/2}).$$

The latter assumption implies that, for every  $i \in \{1, \dots, n\}$  and  $k \in \{1, \dots, d\}$ ,

$$\hat{U}_{i,k} - U_{i,k} = \frac{1}{n} \sum_{j=1}^n a_{k,n}(\mathbf{X}_j, \mathbf{Z}_j, X_{i,k}, \mathbf{Z}_i) + r_{n,i}, \quad n^{1/2} \max_{i \in \{1, \dots, n\}} |r_{n,i}| = o_P(1).$$

Denote  $\mathbf{a}_n(\mathbf{X}_j, \mathbf{Z}_j, \mathbf{X}_i, \mathbf{Z}_i)$ , or  $\mathbf{a}_{i,j}$  even shorter, the  $d$ -vector whose components are  $a_{k,n}(\mathbf{X}_j, \mathbf{Z}_j, X_{i,k}, \mathbf{Z}_i)$  for all  $k \in \{1, \dots, d\}$ .

In the case of the kernel-based estimates  $\hat{F}_k$  of Lemma 4, Assumption 14 is satisfied by using  $s$ -order kernels  $\mathbf{K}$  such that  $\max_k h_k = o(n^{-1/(2s)})$  and  $n^{1/2} \prod_{k=1}^p h_k \gg n^a$  for some  $a > 0$ . If  $h_k = n^{-\pi}$  for all  $k \in \{1, \dots, d\}$ , this necessitates  $s > p$  and  $\pi \in (1/(2s), 1/(2p))$ .

Now, we require that the expectation of the previous terms  $\mathbf{a}_n(\mathbf{X}_j, \mathbf{Z}_j, \mathbf{X}_i, \mathbf{Z}_i)$  in the expansion tends towards a deterministic function sufficiently quickly and uniformly with respect to  $i \in \{1, \dots, n\}$ .

**Assumption 15.** Define  $\Lambda_{\psi(\beta_0, \beta_0^{\top} \mathbf{z})} = \nabla_{\mathbf{u}} \nabla_{\theta} (\ln c_{\theta})|_{\theta=\psi(\beta_0, \beta_0^{\top} \mathbf{z})}$ , and assume that

$$r_{n,\infty} \mathbb{E}\{|\Lambda_{\psi(\beta_0, \beta_0^{\top} \mathbf{Z}_i)}(\mathbf{U}_i)| \omega_n(\mathbf{U}_i, \mathbf{Z}_i)\} = o(n^{-1/2}). \tag{B.6}$$

Assume that there exists a function  $W$  such that

$$\sup_{\mathbf{x} \in \mathbb{R}^d, \mathbf{z} \in \mathcal{Z}} |\mathbb{E}\{\mathbf{a}_n(\mathbf{X}_i, \mathbf{Z}_i, \mathbf{x}, \mathbf{z})\} - W(\mathbf{z}, \mathbf{x})| = W_{n,\infty} = o(n^{-1/2}),$$

and such that

$$W_{n,\infty} \mathbb{E} \left\{ \left| \Lambda_{\psi(\beta_0, \beta_0^{\top} \mathbf{Z}_i)}(\mathbf{U}_i) W(\mathbf{Z}_i, \mathbf{X}_i) \cdot \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i)|_{\beta=\beta_0} \right| \omega_{i,n} \right\} < \infty. \tag{B.7}$$

Choosing the kernel-based estimates  $\hat{F}_k$  of Lemma 4, we see that  $\mathbb{E}\{a_n(\mathbf{X}_j, \mathbf{Z}_j, \mathbf{x}, \mathbf{z})\} = W(\mathbf{z}, \mathbf{x}) = 0$  and Assumption 15 is automatically satisfied. This is most often the case with parametric marginal models, too.

Moreover, (B.6) and (B.7) are often easily satisfied when  $\mathbb{E}\{|\Lambda_{\psi(\beta_0, \beta_0^{\top} \mathbf{Z}_i)}(\mathbf{U}_i)| \mathbf{1}(\mathbf{Z}_i \in \mathcal{Z})\} < \infty$ . Note that the Gaussian copula model does not fulfill the latter condition. Nonetheless, Assumption 15 will be satisfied with a convenient choice of bandwidths, kernels and trimming sequences; see Appendix B.4.

When Assumption 15 was related to the bias of  $\hat{\mathbf{U}}_i$ , the next one seeks to control its variance.

**Assumption 16.** Assume  $v_n^2 = \mathbb{E}[|\mathbf{a}_n(\mathbf{X}_2, \mathbf{Z}_2, \mathbf{X}_1, \mathbf{Z}_1) - \mathbb{E}\{\mathbf{a}_n(\mathbf{X}_2, \mathbf{Z}_2, \mathbf{X}_1, \mathbf{Z}_1) | \mathbf{X}_1, \mathbf{Z}_1\}|^2] < \infty$  and

$$v_n^2 \mathbb{E}\{|\Lambda_{\psi(\beta_0, \beta_0^{\top} \mathbf{Z}_i)}(\mathbf{U}_i)|^2 \omega_{i,n}\} / n = o(1).$$

We are ready to prove Corollary 3. We use the same notations as in the proof of Theorem 2. Recall that

$$A_{2,n} = \frac{1}{n} \sum_{i=1}^n \{\nabla_{\theta} \ln c_{\theta}|_{\theta=\psi_i}(\hat{\mathbf{U}}_i) - \nabla_{\theta} \ln c_{\theta}|_{\theta=\psi_i}(\mathbf{U}_i)\} \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i)|_{\beta=\beta_0} \omega_{i,n},$$

which can be rewritten as

$$A_{2,n} = \frac{1}{n} \sum_{i=1}^n \Lambda_{\psi_i}(\mathbf{U}_i) (\hat{\mathbf{U}}_i - \mathbf{U}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} + O_P(\delta_n^{2-\gamma}) \equiv A'_{2,n} + o_P(n^{-1/2}),$$

thanks to a limited expansion and invoking [Assumptions 15](#) and [13](#). Next, under [\(B.6\)](#), we have

$$A'_{2,n} = \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \Lambda_{\psi_i}(\mathbf{U}_i) \mathbf{a}_{i,j} \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} + o_P(n^{-1/2}).$$

The leading term in  $A'_{2,n}$  can be decomposed into  $A'_{21} + A'_{22}$ , where

$$A'_{21} = \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \Lambda_{\psi_i}(\mathbf{U}_i) E(\mathbf{a}_{i,j} | \mathbf{Z}_i, \mathbf{X}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n}$$

and

$$A'_{22} = \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1, i \neq j}^n \Lambda_{\psi_i}(\mathbf{U}_i) \{\mathbf{a}_{i,j} - E(\mathbf{a}_{i,j} | \mathbf{Z}_i, \mathbf{X}_i)\} \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n}.$$

Due to [Assumption 15](#), Eq. [\(B.7\)](#), we have

$$A'_{21} = \frac{1}{n} \sum_{i=1}^n \Lambda_{\psi_i}(\mathbf{U}_i) W(\mathbf{Z}_i, \mathbf{X}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \omega_{i,n} + o_P(n^{-1/2}).$$

Next, observe that the main term of  $A'_{22}$  is of the form  $\sum_{i < j} \mathfrak{U}(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Z}_j, \mathbf{X}_j)$ , after symmetrization, where

$$E\{\mathfrak{U}(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Z}_j, \mathbf{X}_j) | \mathbf{Z}_j, \mathbf{X}_j\} = E\{\mathfrak{U}(\mathbf{Z}_i, \mathbf{X}_i, \mathbf{Z}_j, \mathbf{X}_j) | \mathbf{Z}_i, \mathbf{X}_i\} = 0.$$

So,  $A'_{22}$  is a degenerate  $U$ -process of order 2. It can be easily verified that its expectation is zero and

$$\text{var}(A'_{22}) = O \left\{ \frac{v_n^2}{n^2} \int |\Lambda_{\psi(\beta_0, \beta_0^{\top} \mathbf{z})}(\mathbf{u})|^2 |\nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{z}) |_{\beta=\beta_0}|^2 \omega_n(\mathbf{u}, \mathbf{z}) d\text{Pr}_{(U, Z)}(\mathbf{u}, \mathbf{z}) \right\}.$$

Under [Assumptions 16](#), we get  $A'_{22} = o_P(n^{-1/2})$ . We have obtained

$$\begin{aligned} A_{1n} + A_{2n} &= \frac{1}{n} \sum_{i=1}^n \omega_{i,n} \nabla_{\theta} \ln c_{\theta|\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \omega_{i,n} \Lambda_{\psi_i}(\mathbf{U}_i) \cdot W(\mathbf{Z}_i, \mathbf{X}_i) \nabla_{\beta} \psi(\beta, \beta^{\top} \mathbf{Z}_i) |_{\beta=\beta_0} + o_P(n^{-1/2}) \equiv \frac{1}{n} \sum_{i=1}^n \omega_i \mathcal{M}_i + B_n + o_P(n^{-1/2}), \end{aligned}$$

by introducing a bias term  $B_n = \sum_{i=1}^n (\omega_{i,n} - \omega_i) \mathcal{M}_i / n$ , due to the trimming procedure. Its expectation is denoted  $b_n = E\{(\omega_{1,n} - \omega_1) \mathcal{M}_1\}$ , and its variance is  $O(n^{-1} \delta_n)$ . The asymptotic bias is negligible under [\(19\)](#), by recalling [Assumption 12](#), and then applying [Lemma 5](#). In every case, the result follows from a standard Central Limit Theorem, recalling the expansion of [Theorem 2](#).  $\square$

#### B.4. Examples (continued)

Here, we analyze the conditions [1](#) to [16](#) to obtain the asymptotic normality of  $\hat{\beta}$  in the case of the copula families in [Section 2.4](#). The reader will note that all such assumptions can be checked in practical terms, even if such a task may be slightly boring.

**Example 4** ((Continued): The Gaussian Copula). Obviously, [Assumptions 5](#), [9](#) and [10](#) are satisfied. This is the case for [Assumption 7](#) too, because  $\Sigma \mapsto \ln(|\Sigma|)$  is Lipschitz under [\(17\)](#) and invoking [Example 19.7](#) in [\[45\]](#).

To deal with [Assumption 8](#), note that  $p$  and  $\hat{p}$  are Lipschitz transforms of conditional Kendall's tau  $\tau(\beta, \beta^{\top} \mathbf{z})$  and  $\hat{\tau}(\beta, \beta^{\top} \mathbf{z})$ , respectively. From [Example 19.20](#) in [\[45\]](#), it is sufficient to show that the functions  $\mathbf{z} \mapsto \nabla_{\beta} \tau(\beta_0, \beta_0^{\top} \mathbf{z})$  and  $\mathbf{z} \mapsto \nabla_{\beta} \hat{\tau}(\beta_0, \beta_0^{\top} \mathbf{z})$  belong to a Donsker class a.e., assuming the underlying dimension  $d$  is two. It follows from [Lemma 7](#) and from the relation  $\tau(\beta_0, \beta_0^{\top} \mathbf{z}) = -1 + 4 \int C_{\beta_0}(\mathbf{u} | \beta_0^{\top} \mathbf{z}) C_{\beta_0}(\mathbf{du} | \beta_0^{\top} \mathbf{z})$  that

$$\forall \mathbf{z} \in \mathcal{Z} \quad \nabla_{\beta} \tau(\beta_0, \beta_0^{\top} \mathbf{z}) = f_1(\beta_0^{\top} \mathbf{z}) + \mathbf{z} f_2(\beta_0^{\top} \mathbf{z}),$$

with

$$f_1(v) = -E(\mathbf{Z} | \beta_0^{\top} \mathbf{Z} = v, \mathbf{Z} \in \mathcal{Z}) \left\{ \int c_0(\mathbf{u}, v) C_{\beta_0}(\mathbf{du} | v) + \int C_{\beta_0}(\mathbf{u} | v) c_0(\mathbf{du}, v) \right\},$$



$$f_2(v) = \mathbf{z} \left\{ \int c_0(\mathbf{u}, v) C_{\beta_0}(\mathbf{d}\mathbf{u}|v) + \int C_{\beta_0}(\mathbf{u}|v) c_0(\mathbf{d}\mathbf{u}, v) \right\},$$

using the notations of Lemma 7. In a Gaussian copula family, the maps  $\mathbf{z} \mapsto f_j(\beta_0^\top \mathbf{z})$  and  $\mathbf{z} \mapsto f'_j(\beta_0^\top \mathbf{z})$  are uniformly bounded on  $\mathcal{Z}$ . Therefore,  $\nabla_\beta \tau(\beta_0, \beta_0^\top \mathbf{z})$  belongs to the class  $\mathcal{G} = \{ \mathbf{z} \in \mathcal{Z} \rightarrow f(\beta_0^\top \mathbf{z}) + \mathbf{z}g(\beta_0^\top \mathbf{z}), f, g \in \mathcal{C}(M) \}$  with  $\mathcal{C}(M) = \{ f : \|f\|_\infty + \|f'\|_\infty \leq M \}$ . The collection  $\mathcal{C}(M)$  is a Donsker class from Theorem 2.7.1 in van der Vaart and Wellner [46]. Moreover,  $\mathcal{G}$  is Donsker from Examples 2.10.7–2.10.8 in [46].

It is also the case for  $\nabla_\beta \hat{\tau}$ . Indeed, with the notations of Appendix C, we can write

$$\hat{\tau}(\beta, \beta^\top \mathbf{z}) = -1 + \frac{4}{n^2 \hat{f}_\beta^2(\beta^\top \mathbf{z})} \sum_{i,j=1}^n \mathbf{1}(X_j \leq \mathbf{X}_i) \tilde{K}_{\tilde{h}}(\beta^\top \mathbf{Z}_j - \beta^\top \mathbf{z}) \tilde{K}_{\tilde{h}}(\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}).$$

A differentiation with respect to  $\beta$  easily shows that  $\nabla_\beta \hat{\tau}(\beta_0, \beta_0^\top \mathbf{z})$  is of the form  $\nabla_\beta \hat{\tau}(\beta_0, \beta_0^\top \mathbf{z}) = \hat{f}_1(\beta_0^\top \mathbf{z}) + \mathbf{z} \hat{f}_2(\beta_0^\top \mathbf{z})$ . The results of Appendix C allow to show that, for  $j \in \{1, 2\}$ ,

$$\sup_{\mathbf{z} \in \mathcal{Z}} |\hat{f}_j(\beta_0^\top \mathbf{z}) - f_j(\beta_0^\top \mathbf{z})| = O_p\{\tilde{h}_n^2 + (\ln n)^{1/2} n^{-1/2} \tilde{h}_n^{-3/2}\}$$

and

$$\sup_{\mathbf{z} \in \mathcal{Z}} |\hat{f}'_j(\beta_0^\top \mathbf{z}) - f'_j(\beta_0^\top \mathbf{z})| = O_p\{\tilde{h}_n^2 + (\ln n)^{1/2} n^{-1/2} \tilde{h}_n^{-5/2}\}.$$

Therefore,  $\mathbf{z} \in \mathcal{Z} \mapsto \nabla_\beta \hat{\tau}(\beta_0, \beta_0^\top \mathbf{z})$  belongs to the Donsker class  $\mathcal{G}$  when  $n\tilde{h}_n^5 \rightarrow 0$ .

Eq. (18) of Assumption 6 is coming from the results of Appendix C, and simple calculations prove that Assumption 11 is satisfied for every  $\alpha > 0$ . Recalling the notations of Appendix C, we have

$$\sup_{\mathbf{z} \in \mathcal{Z}} |\hat{\tau}(\beta_0, \beta_0^\top \mathbf{z}) - \tau(\beta_0, \beta_0^\top \mathbf{z})| = O_p\{\tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-1/2}\} = O_p(\eta_{1n})$$

and

$$\sup_{\mathbf{z} \in \mathcal{Z}} |\nabla_\beta \hat{\tau}(\beta_0, \beta_0^\top \mathbf{z}) - \nabla_{\beta_0} \tau(\beta_0, \beta_0^\top \mathbf{z})| = O_p\{\tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-3/2}\} = O_p(\eta_{2n}).$$

To fix the ideas, assume  $\tilde{h} \sim n^{-\tilde{\pi}}$ , for some  $\tilde{\pi} > 0$ . Then, to satisfy  $\eta_{1n} \eta_{2n} = o(n^{-1/2})$ , it is sufficient to have  $4\tilde{s}\tilde{\pi} > 1, \tilde{s} \geq 2$  and  $4\tilde{\pi} < 1$ . Recall that we had set  $\delta_n \sim n^{-\pi s} + \ln_2 n \cdot n^{-(1-p\pi)/2}$ . To satisfy  $\delta_n^{1-\alpha} \eta_{jn} = o(n^{-1/2})$  for  $j \in \{1, 2\}$ , it is sufficient to have  $1 < (1 - \alpha) \min(2s\pi, 1 - p\pi) + \min(2\tilde{s}\tilde{\pi}, 1 - 3\tilde{\pi})$ .

Concerning Assumption 12, it can be verified that the left-hand side of (B.4) is  $O[\delta_n v_n \{\Phi^{-1}(v_n)\}^2]$ . Nonetheless,  $\Phi^{-1}(v_n) \sim -\sqrt{(-2) \ln v_n}$ , when  $v_n \rightarrow 0$ ; see [10]. A sufficient condition is then  $\delta_n v_n \ln(v_n) = o(n^{-1/2})$ .

Assumptions 14 and 15 are trivially satisfied because we have chosen nonparametric marginal CDFs and we apply Lemma 4, for which we have seen that we set  $W(\mathbf{z}, \mathbf{x}) = 0$ .

Assumption 13 is the most demanding and cannot be obtained by the same reasoning as for Assumption 12. Actually, we recall that the former one has been requested only in the proof of Corollary 3 to show that

$$\frac{1}{n} \sum_{i=1}^n \nabla_{\mathbf{u}} \nabla_{\theta}^2 (\ln c_{\theta})|_{\theta=\psi_i}(\mathbf{U}_i^*) (\hat{\mathbf{U}}_i - \mathbf{U}_i)^2 \nabla_{\beta} \psi(\beta, \beta^\top \mathbf{Z}_i)|_{\beta=\beta_0} \omega_{i,n} = o_p(n^{-1/2}),$$

for some random vectors  $\mathbf{U}_i^*, |\mathbf{U}_i^* - \mathbf{U}_i| \leq |\hat{\mathbf{U}}_i - \mathbf{U}_i|$ . Due to Assumption 3, it is sufficient to check that

$$\delta_n^2 E\{|\nabla_{\mathbf{u}} \nabla_{\theta}^2 (\ln c_{\theta})|_{\theta=\psi_i}(\mathbf{U}_i) \nabla_{\beta} \psi(\beta, \beta^\top \mathbf{Z}_i)|_{\beta=\beta_0} | \omega_{i,n}\} = o(n^{-1/2}).$$

Due to the boundedness of  $c_{\theta}$ , the latter expectation is less than a constant times

$$\int_{\Phi^{-1}(v_n)}^{\Phi^{-1}(1-v_n)} |t| \exp(t^2/2) dt.$$

The latter integral behaves as  $\exp\{\Phi^{-1}(v_n)\}^2/2$ . Since  $\Phi^{-1}(v_n) \sim -\sqrt{(-2) \ln v_n}$ , it is sufficient to satisfy  $\delta_n^2/v_n = o(n^{-1/2})$ . Usual variance calculations for kernel densities prove that Assumption 16 is true when  $nh^p = n^{1-p\pi} \rightarrow \infty$ , i.e., when  $p\pi < 1$ .

Gathering all the previous constraints, we can exhibit explicit combinations of parameters. For instance, we can set

$$s = 2p, \quad \tilde{s} = 4, \quad \pi = 1/(2s + p), \quad \tilde{\pi} = 1/9, \quad h_n \sim n^{-1/(2s+p)} = n^{-1/5p}, \quad \tilde{h}_n \sim n^{-4/9},$$

for some  $\alpha < 1/2$ , implying  $\delta_n \sim n^{-2/5}$  and we choose  $v_n = n^{-1/5}$ . Note that we need high-order kernels in general, even in the bivariate case ( $p = 2$ ).

Similar reasonings allow to exhibit explicit tuning parameters to manage Clayton and/or Gumbel copula models. They are left to the reader as an exercise.

### Appendix C. Conditional Kendall's tau

In this section, we show how to check [Assumption 6](#) in general, when the conditional margins are estimated nonparametrically. Incidentally, we prove some theoretical results related to the estimation of conditional Kendall's tau, that are valuable per se.

We consider the situation of a  $d$ -dimensional random vector  $\mathbf{X}$ , whose conditional copula is parameterized by  $\tau(\beta, \beta^\top \mathbf{z})$ , the conditional Kendall's tau coefficient of this vector as defined in (11) when  $d = 2$ , and (12) more generally. In other words, we consider the case where  $\psi(\beta, \beta^\top \mathbf{z}) = g\{\tau(\beta, \beta^\top \mathbf{z})\}$  for some "sufficiently regular" function  $g$ . Indeed, Kendall's tau are commonly used for inference purposes of parametric copulas, particularly Archimedean and elliptical copulas. Moreover, as explained in Section 2.2, (6) and (7) are satisfied in such cases. Finally, we do not suffer from the curse of dimensionality because conditional Kendall's tau is that associated to the copula of  $\mathbf{X}$  knowing  $\beta^\top \mathbf{Z}$ .

Introducing a kernel estimator  $\hat{F}_\beta$  of  $F_\beta(\mathbf{x}|y) = \Pr(\mathbf{X} \leq \mathbf{x} | \beta^\top \mathbf{Z} = y)$  as  $\hat{F}_\beta(\mathbf{x}|y) = \hat{H}_\beta(\mathbf{x}, \infty | y)$  (recall (13)), define

$$\hat{\tau}(\beta, \beta^\top \mathbf{z}) = \frac{1}{2^d - 1} \left\{ -1 + 2^d \int \hat{F}_\beta(\mathbf{x} | \beta^\top \mathbf{z}) \hat{F}_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) \right\}.$$

In [Lemma 12](#), we show that the uniform consistency of the conditional Kendall's tau coefficient is obtained, provided that we have some convenient convergence rates for  $\hat{F}_\beta$ .

**Lemma 12.** *Assume that*

$$\sup_{\mathbf{x} \in \mathbb{R}^d, \beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\hat{F}_\beta(\mathbf{x} | \beta^\top \mathbf{z}) - F_\beta(\mathbf{x} | \beta^\top \mathbf{z})| = O_P(\varepsilon_{n,0}). \quad (\text{C.1})$$

Then

$$\sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\hat{\tau}(\beta, \beta^\top \mathbf{z}) - \tau(\beta, \beta^\top \mathbf{z})| = O_P(\varepsilon_{n,0}).$$

**Proof.** Decompose

$$(2^d - 1) \{ \hat{\tau}(\beta, \beta^\top \mathbf{z}) - \tau(\beta, \beta^\top \mathbf{z}) \} = 2^d \int \{ \hat{F}_\beta(\mathbf{x} | \beta^\top \mathbf{z}) - F_\beta(\mathbf{x} | \beta^\top \mathbf{z}) \} \hat{F}_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) \\ + 2^d \int F_\beta(\mathbf{x} | \beta^\top \mathbf{z}) \{ \hat{F}_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) - F_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) \}.$$

The first term is  $O_P(\varepsilon_{n,0})$  due to (C.1). For the second, observe that

$$\int F_\beta(\mathbf{x} | \beta^\top \mathbf{z}) \{ \hat{F}_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) - F_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) \} = (-1)^{d-1} \int \{ \hat{F}_\beta(\mathbf{x} | \beta^\top \mathbf{z}) - F_\beta(\mathbf{x} | \beta^\top \mathbf{z}) \} F(d\mathbf{x} | \beta^\top \mathbf{z}),$$

which is less than  $\sup_{\mathbf{x}, \beta, \mathbf{z}} |\hat{F}_\beta(\mathbf{x} | \beta^\top \mathbf{z}) - F_\beta(\mathbf{x} | \beta^\top \mathbf{z})|$ , and we use again (C.1).  $\square$

[Lemma 12](#) yields some tools to verify the first part of [Assumption 6](#), if one assumes that the function  $g$  is regular enough (i.e., Hölder with some high enough Hölder exponent). Similarly, we can derive the uniform consistency of  $\nabla_\beta^j \hat{\tau}$  for  $j \in \{1, 2\}$ , which allows to check the remaining conditions in [Assumption 6](#).

**Lemma 13.** *Assume that*

$$\sup_{\mathbf{x} \in \mathbb{R}^d, \beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\nabla_\beta^j \hat{F}_\beta(\mathbf{x} | \beta^\top \mathbf{z}) - \nabla_\beta^j F_\beta(\mathbf{x} | \beta^\top \mathbf{z})| = O_P(\varepsilon_{n,j}), \quad (\text{C.2})$$

for  $j \in \{1, 2\}$ , and that

$$\max_{j \in \{1, 2\}} \int \left| \nabla_\beta^j F_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) \right| + \left| \nabla_\beta^j \hat{F}_\beta(d\mathbf{x} | \beta^\top \mathbf{z}) \right| \leq C_0,$$

for some positive constant  $C_0$ . Then

$$\sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\nabla_\beta \hat{\tau}(\beta, \beta^\top \mathbf{z}) - \nabla_\beta \tau(\beta, \beta^\top \mathbf{z})| = O_P\{\max(\varepsilon_{n,1}, \varepsilon_{n,0})\}$$

and

$$\sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\nabla_\beta^2 \hat{\tau}(\beta, \beta^\top \mathbf{z}) - \nabla_\beta^2 \tau(\beta, \beta^\top \mathbf{z})| = O_P\{\max(\varepsilon_{n,2}, \varepsilon_{n,1}, \varepsilon_{n,0})\}.$$

**Proof.** This is a consequence of applying the  $\nabla$ -operator to  $\hat{\tau}(\beta, \beta^\top \mathbf{z})$ , and of the compactness of  $\mathcal{Z}$ .  $\square$

The next step is to verify that, under reasonable conditions, (C.1) and (C.2) hold. To this aim, let us introduce some assumptions.

**Assumption 17.** Let  $\tilde{K}$  denote a univariate symmetric kernel function of order  $\tilde{s}, \tilde{s} \geq 2$ . It is twice continuously differentiable with bounded derivatives up to order 2. Moreover,  $\tilde{h}_n$  denotes a bandwidth sequence, where  $\tilde{h}_n = O(n^{-a})$  for some  $a > 0$  and  $n\tilde{h}_n \rightarrow \infty$ .

Note that, in general, the latter triplet  $(\tilde{K}, \tilde{h}, \tilde{s})$  is different from the similar quantities  $(K, h, s)$  that have been invoked to define the pseudo-observations  $\hat{\mathbf{U}}_i$ ; see Lemma 4.

**Assumption 18.** Let  $f_\beta(y)$  denote the density of  $\beta^\top \mathbf{Z}$  evaluated at point  $y$ . Assume that  $\inf_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} \inf_y f_\beta(y) > c$ , for some  $c > 0$ . Moreover, assume that  $f_\beta$  is  $s$ -times continuously differentiable, with uniformly bounded derivatives.

The latter assumption is satisfied most of the time, because  $\beta^\top \mathbf{z}$  belongs to a compact subset when  $\beta \in \mathcal{B}$  and  $\mathbf{z} \in \mathcal{Z}$ . For instance, assume the arguments  $y$  above belong to a fixed interval  $[a, b]$  and that  $\mathbf{Z}$  follows a Gaussian  $\mathcal{N}(0, \Sigma)$ . Then  $\beta^\top \mathbf{Z} \sim \mathcal{N}(0, \beta^\top \Sigma \beta)$  and  $f_\beta(y) = \exp\{-y^2/2(\beta^\top \Sigma \beta)\}/(\sqrt{2\pi} \beta^\top \Sigma \beta)$ . Since  $\beta^\top \Sigma \beta$  belongs to a compact  $[c, d], c > 0$ , the latter density is larger than  $\exp\{-b^2/(2d^2)\}/(\sqrt{2\pi}d) > 0$ .

In the single-index literature, some authors relaxed this assumption, by only assuming  $\inf_{\mathbf{z}} \inf_y f_{\beta_0}(y) > c$ . Nevertheless, Assumption 18 requires to introduce a trimming procedure, in order to avoid parts of the distribution for which some  $f_\beta(\beta^\top \mathbf{Z}_i)$  are too close to zero. Such trimming procedures (generally working in two-steps), that can be extended straightforwardly in our case, have been investigated in detail for example in Lopez et al. [28].

Let  $\mathcal{A}$  denote a generic set of functions with envelope  $F$ . For a probability measure  $\mathbb{Q}$ , let  $\mathcal{N}(\varepsilon, \mathcal{A}, \|\cdot\|_{2, \mathbb{Q}})$  denote the number of  $L^2(\mathbb{Q})$ -balls required to cover the set of functions  $\mathcal{A}$ , and

$$N(\varepsilon, \mathcal{A}) = \sup_{\mathbb{Q}: \|F\|_{2, \mathbb{Q}} < \infty} \mathcal{N}(\varepsilon \|F\|_{2, \mathbb{Q}}, \mathcal{A}, \|\cdot\|_{2, \mathbb{Q}}).$$

**Assumption 19.**  $\mathcal{A}$  is a class of functions bounded by 1 such that  $N(\varepsilon, \mathcal{A}) \leq C\varepsilon^{-\nu}$ . Moreover, for  $\phi \in \mathcal{A}$ , let  $m_\phi(y) = E\{\phi(\mathbf{X}, \mathbf{Z}) | \beta^\top \mathbf{Z} = y\}$ . Assume that the functions  $m_\phi$  are twice continuously differentiable, and their derivatives up to order 2 are upper bounded by some finite constant  $M$  that does not depend on  $\phi$ .

We first state Lemma 14, that yields consistency rates for kernel weighted sums.

**Lemma 14.** Let  $\mathcal{L}$  denote a class of functions satisfying Assumption 19. Under Assumption 17, we have

$$\frac{1}{n\tilde{h}} \sup_{\lambda \in \mathcal{L}} \sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} \left| \sum_{i=1}^n \lambda(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right) - E \left\{ \lambda(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right) \right\} \right| = O_P\{(\ln n)^{1/2} n^{-1/2} \tilde{h}^{-1/2}\}.$$

**Proof.** Let

$$\mathcal{B} = \sup_{\beta, \mathbf{z}, \lambda} \left| \sum_{i=1}^n \lambda(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right) - E \left\{ \lambda(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right) \right\} \right|,$$

and

$$\mathcal{B}_\varepsilon = E \left\{ \sup_{\beta, \mathbf{z}, \lambda} \left| \sum_{i=1}^n \varepsilon_i \lambda(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right) \right| \right\},$$

where  $(\varepsilon_i)_{1 \leq i \leq n}$  are iid Rademacher variables. Due to Proposition 9, we have

$$\Pr \{ \mathcal{B} \geq A_1(\mathcal{B}_\varepsilon + t) \} \leq 2[\exp\{-A_2' t^2/(n\tilde{h})\} + \exp(-A_2 t)], \tag{C.3}$$

where  $A_2'$  is a constant. Indeed, since the functions  $\lambda$  are uniformly bounded by 1,

$$\sup_{\beta, \mathbf{z}, \lambda} \text{var} \left\{ \lambda(\mathbf{X}, \mathbf{Z}) \tilde{K} \left( \frac{\beta^\top \mathbf{Z} - \beta^\top \mathbf{z}}{\tilde{h}} \right) \right\} = O(\tilde{h}).$$

Next, observe that the class of functions

$$\mathcal{L}_{\tilde{K}} = \left\{ g : \mathbb{R}^d \times \mathcal{Z} \rightarrow \mathbb{R}, (\mathbf{x}, \mathbf{z}) \mapsto \lambda(\mathbf{x}, \mathbf{z}) \tilde{K} \left( \frac{\beta^\top \mathbf{z} - \beta^\top \mathbf{u}}{\tilde{h}} \right) : \mathbf{u} \in \mathcal{Z}, \beta \in \mathcal{B}, \tilde{h} \in \mathbb{R}^+ \right\},$$

satisfies the assumptions of Proposition 10 with  $\sigma^2 = O(h)$  and that, for some  $C$  and  $\nu$ ,

$$N(\varepsilon, \mathcal{L}_{\tilde{K}}) \leq C\varepsilon^{-\nu}. \tag{C.4}$$

The property (C.4) can be obtained from the following: Lemma 22 in Nolan and Pollard [31] shows that  $N(\varepsilon, \mathcal{K}) \leq C_2 \varepsilon^{-\nu_2}$ , where

$$\mathcal{K} = \left\{ (\mathbf{x}, \mathbf{z}) \in \mathbb{R}^d \times \mathcal{Z} \mapsto \tilde{K} \left( \frac{\beta^\top \mathbf{z} - \beta^\top \mathbf{u}}{\tilde{h}} \right) : \mathbf{u} \in \mathcal{Z}, \beta \in \mathcal{B}, \tilde{h} \in \mathbb{R}^+ \right\}.$$

Using Assumption 19 and Lemma A.1 in Einmahl and Mason [13], we get that  $\mathcal{L}_{\tilde{K}} = \mathcal{L}\mathcal{K}$  satisfies (C.4).

Therefore, we can apply Proposition 10 to deduce that

$$\mathcal{B}_\varepsilon \leq A' n^{1/2} \tilde{h}^{1/2} (-\ln \tilde{h})^{1/2} = A'' n^{1/2} \tilde{h}^{1/2} (\ln n)^{1/2}. \tag{C.5}$$

It follows from (C.5) that, for  $t_1 > 2A_1 A''$ ,

$$\Pr\{\mathcal{B} \geq t_1 n^{1/2} \tilde{h}^{1/2} (\ln n)^{1/2}\} \leq \Pr\{\mathcal{B} \geq A_1 \mathcal{B}_\varepsilon + t_1 n^{1/2} \tilde{h}^{1/2} (\ln n)^{1/2} / 2\}.$$

Applying (C.3) with  $t = t_1 n^{1/2} \tilde{h}^{1/2} (\ln n)^{1/2} / (2A_1)$ , we get

$$\Pr\{\mathcal{B} \geq t_1 n^{1/2} \tilde{h}^{1/2} (\ln n)^{1/2}\} \leq 2[\exp\{-A_2' t_1^2 \ln n / (4A_1^2)\}] + \exp\{-A_2 t_1 n^{1/2} \tilde{h}^{1/2} (\ln n)^{1/2} / (2A_1)\},$$

and the result follows.  $\square$

This lemma is the cornerstone of Lemma 15, which ensures consistency rates for  $\hat{F}_\beta$  and its derivatives.

**Lemma 15.** *Let  $\mathcal{A}$  denote a class of functions satisfying Assumption 19. Then, under Assumptions 17 and 18,*

$$\sup_{\phi \in \mathcal{A}} \sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} \left| \int \phi(\mathbf{x}, \mathbf{z}) \{\hat{F}_\beta(d\mathbf{x}|\beta^\top \mathbf{z}) - F_\beta(d\mathbf{x}|\beta^\top \mathbf{z})\} \right| = O_p\{\tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-1/2}\}.$$

**Proof.** Write

$$\hat{m}_\phi(\beta^\top \mathbf{z}) = \int \phi(\mathbf{x}, \mathbf{z}) \hat{F}_\beta(d\mathbf{x}|\beta^\top \mathbf{z}) = \frac{1}{n \hat{f}_\beta(\beta^\top \mathbf{z})} \sum_{i=1}^n \phi(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right),$$

where

$$\hat{f}_\beta(\beta^\top \mathbf{z}) = \frac{1}{n \tilde{h}} \sum_{i=1}^n \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right),$$

is an estimator of the density  $f_\beta(\beta^\top \mathbf{z})$  of  $\beta^\top \mathbf{Z}$  evaluated at  $\beta^\top \mathbf{z}$ . Let

$$\hat{m}_\phi(\beta^\top \mathbf{z}) = \frac{1}{n \tilde{h}} \sum_{i=1}^n \phi(\mathbf{X}_i, \mathbf{Z}_i) \tilde{K} \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right) = \hat{m}_\phi(\beta^\top \mathbf{z}) \hat{f}_\beta(\beta^\top \mathbf{z}),$$

and  $m_\phi(\beta^\top \mathbf{z}) = m_\phi(\beta^\top \mathbf{z}) f_\beta(\beta^\top \mathbf{z})$ . It follows from Lemma 14 that

$$\sup_{\beta, \mathbf{z}, \phi} |\hat{m}_\phi(\beta^\top \mathbf{z}) - E\{\hat{m}_\phi(\beta^\top \mathbf{z})\}| + \sup_{\beta, \mathbf{z}} |\hat{f}_\beta(\beta^\top \mathbf{z}) - E\{\hat{f}_\beta(\beta^\top \mathbf{z})\}| = O_p \left\{ \frac{(\ln n)^{1/2}}{n^{1/2} \tilde{h}^{1/2}} \right\}.$$

Moreover, using classical arguments on kernel estimators (and Assumptions 19 and 17), we have

$$\sup_{\beta, \mathbf{z}, \phi} |E\{\hat{m}_\phi(\beta^\top \mathbf{z})\} - m_\phi(\beta^\top \mathbf{z})| + \sup_{\beta, \mathbf{z}} |E\{\hat{f}_\beta(\beta^\top \mathbf{z})\} - f_\beta(\beta^\top \mathbf{z})| = O(\tilde{h}^{\tilde{s}}).$$

The result of the lemma follows from the fact that the density  $f_\beta(\beta^\top \mathbf{z})$  is bounded away from zero by Assumption 18.  $\square$

Lemma 15 allows to verify Condition (C.1) by considering  $\phi(\mathbf{x}, \mathbf{z}) = \mathbf{1}(\mathbf{x} \leq \mathbf{x}_0)$ , for some constant vectors  $\mathbf{x}_0$ . This shows that, in this case,  $\varepsilon_{n,0} = \tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-1/2}$ . It also permits to obtain the uniform consistency rates for  $\nabla_\beta^j \hat{F}_\beta$  for  $j \in \{1, 2\}$ , with

$$\varepsilon_{n,1} = \tilde{h}^{\tilde{s}} + \frac{(\ln n)^{1/2}}{n \tilde{h}^{3/2}}, \quad \varepsilon_{n,2} = \tilde{h}^{\tilde{s}} + \frac{(\ln n)^{1/2}}{n \tilde{h}^{5/2}}.$$

Indeed,

$$\nabla_\beta \hat{m}_\phi(\beta^\top \mathbf{z}) = \frac{1}{n \tilde{h}^2} \sum_{i=1}^n \mathbf{1}(\mathbf{X}_i \leq \mathbf{x}) (\mathbf{Z}_i - \mathbf{z}) \tilde{K}' \left( \frac{\beta^\top \mathbf{Z}_i - \beta^\top \mathbf{z}}{\tilde{h}} \right),$$

and the convergence of this term can be studied using Lemma 15, but replacing  $\tilde{K}$  by  $\tilde{K}'$ , and setting  $\phi(\mathbf{X}, \mathbf{Z}) = \mathbf{1}(\mathbf{X} \leq \mathbf{x})(\mathbf{Z} - \mathbf{z})$ . The latter function is indexed by  $(\mathbf{x}, \mathbf{z})$  that lives into  $\mathbb{R}^d \times \mathcal{Z}$ , defining the convenient class  $\mathcal{A}$  to apply Lemma 15. The other terms obtained by differentiation can be studied in the same way.

Hence, the latter results allow to verify whether [Assumption 6](#) holds. Indeed, under some (light) conditions of regularity, we have obtained that

$$\sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\hat{\tau}(\beta, \beta^\top \mathbf{z}) - \tau(\beta, \beta^\top \mathbf{z})| = O_p\{\tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-1/2}\},$$

$$\sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\nabla_\beta \hat{\tau}(\beta, \beta^\top \mathbf{z}) - \nabla_\beta \tau(\beta, \beta^\top \mathbf{z})| = O_p\{\tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-3/2}\},$$

and

$$\sup_{\beta \in \mathcal{B}, \mathbf{z} \in \mathcal{Z}} |\nabla_\beta^2 \hat{\tau}(\beta, \beta^\top \mathbf{z}) - \nabla_\beta^2 \tau(\beta, \beta^\top \mathbf{z})| = O_p\{\tilde{h}^{\tilde{s}} + (\ln n)^{1/2} n^{-1/2} \tilde{h}^{-5/2}\}.$$

## References

- [1] K. Aas, C. Czado, A. Frigessi, H. Bakken, Pair-copula constructions of multiple dependence, *Insurance Math. Econom.* 44 (2009) 182–198.
- [2] F. Abegaz, I. Gijbels, N. Veraverbeke, Semiparametric estimation of conditional copulas, *J. Multivariate Anal.* 110 (2012) 43–73.
- [3] E.F. Acar, R.V. Craiu, F. Yao, Dependence calibration in conditional copulas: A nonparametric approach, *Biometrics* 67 (2011) 445–453.
- [4] E.F. Acar, R.V. Craiu, Y. Yao, Statistical testing of covariate effects in conditional copula models, *Electron. J. Stat.* 7 (2013) 2822–2850.
- [5] S.M. Bartram, S.J. Taylor, Y.H. Wang, The Euro and the European financial market dependence, *J. Bank. Finance* 31 (2012) 1461–1481.
- [6] R.V. Craiu, A. Sabeti, In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes, *J. Multivariate Anal.* 110 (2012) 106–120.
- [7] M. Delecroix, M. Hristache, M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique, *Bull. Belg. Math. Soc.* 6 (1999) 161–185.
- [8] M. Delecroix, M. Hristache, V. Patilea, On semiparametric M-estimation in single-index regression, *J. Statist. Plann. Inference.* 136 (2006) 730–769.
- [9] A. Derumigny, J.-D. Fermanian, About tests of the “simplifying” assumption for conditional copulas, 2016. arXiv:1612.07349.
- [10] D. Dominici, The inverse of the cumulative standard normal probability function, *Integral Transforms Spec. Funct.* 14 (2003) 281–292.
- [11] J. Dornik, R.P. Sherman, Some convergence theory for iterative estimation procedures with an application to semiparametric estimation, *Econometric Theory* 21 (2005) 838–863.
- [12] Y. Du, M.G. Akritas, IID representations of the conditional Kaplan–Meier process for arbitrary distributions, *Math. Method. Statist.* 11 (2002) 152–182.
- [13] U. Einmahl, D. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators, *J. Theoret. Probab.* 13 (2000) 1–37.
- [14] U. Einmahl, D. Mason, Uniform in bandwidth consistency of kernel-type function estimators, *Ann. Statist.* 33 (2005) 1380–1403.
- [15] J.-D. Fermanian, M. Wegkamp, Time-dependent copulas, *J. Multivariate Anal.* 110 (2012) 19–29.
- [16] C. Genest, J. Nešlehová, N. Ben Ghorbal, Estimators based on Kendall’s tau in multivariate copula models, *Aust. New Zealand J. Stat.* 53 (2011) 157–177.
- [17] I. Gijbels, N. Veraverbeke, M. Omelka, Conditional copulae, association measures and their applications, *Comput. Statist. Data Anal.* 55 (2011) 1919–1932.
- [18] W.K. Härdle, P.J. Hall, H. Ichimura, Optimal smoothing in single-index models, *Ann. Statist.* 21 (1993) 157–178.
- [19] W.K. Härdle, T.M. Stoker, Investigating smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.* 84 (1989) 986–995.
- [20] N. Higham, Computing the nearest correlation matrix – a problem from finance, *IMA J. Numer. Anal.* 22 (2002) 329–343.
- [21] H. Ichimura, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *J. Econometrics* 58 (1993) 71–120.
- [22] H. Joe, Multivariate concordance, *J. Multivariate Anal.* 35 (1990) 12–30.
- [23] H. Joe, *Multivariate Models and Dependence Concepts*, Chapman & Hall, London, 1997.
- [24] E. Jondeau, M. Rockinger, The copula-GARCH model of conditional dependencies: An international stock market application, *J. Int. Money Finance* 25 (2006) 827–853.
- [25] M.G. Kendall, B. Babington Smith, On the method of paired comparisons, *Biometrika* 31 (1940) 324–345.
- [26] R.L. Klein, R.H. Spady, An efficient semiparametric estimator for binary response models, *Econometrica* 61 (1993) 387–421.
- [27] E. Levi, R.V. Craiu, Gaussian process single-index models for conditional copulas, 2016. arXiv:1603.03028.
- [28] O. Lopez, V. Patilea, I. van Keilegom, Single-index regression models in the presence of censoring depending on the covariates, *Bernoulli* 19 (2013) 721–747.
- [29] R.B. Nelsen, *An Introduction to Copulas*, Springer, New York, 1998.
- [30] W. Newey, D. McFadden, Large sample estimation and hypothesis testing, in: *Handbook of Econometrics*, Vol. IV, Elsevier, 1994, pp. 2111–2245.
- [31] D. Nolan, D. Pollard, U-processes: Rates of convergence, *Ann. Statist.* 15 (1987) 780–799.
- [32] A.J. Patton, Modelling asymmetric exchange rate dependence, *Internat. Econom. Rev.* 47 (2006) 527–556.
- [33] A.J. Patton, Copula-based models for financial time series, in: *Handbook of Financial Time Series*, Springer, Berlin, 2009, pp. 767–785.
- [34] A.J. Patton, A review of copula models for economic time series, *J. Multivariate Anal.* 110 (2012) 4–18.
- [35] J.L. Powell, J.H. Stock, T.M. Stoker, Semiparametric estimation of index coefficients, *Econometrica* 57 (1989) 1403–1430.
- [36] B. Rémillard, *Statistical Methods for Financial Engineering*, Chapman & Hall, London, 2013.
- [37] J.C. Rodriguez, Measuring financial contagion: A copula approach, *J. Empir. Finance* 14 (2007) 401–423.
- [38] A. Sabeti, M. Wei, R. Craiu, Additive models for conditional copulas, *STAT* 3 (2014) 300–312.
- [39] F. Schmid, R. Schmidt, Multivariate extensions of Spearman’s rho and related statistics, *Statist. Probab. Lett.* 77 (2007) 407–416.
- [40] D.W. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley, Chichester, 2015.
- [41] R.P. Sherman, U-processes in the analysis of a generalized semiparametric regression estimator, *Econometric Theory* 10 (1994) 372–395.
- [42] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* 8 (1959) 229–231.
- [43] T.M. Stoker, Consistent estimation of scaled coefficients, *Econometrica* 54 (1986) 1461–1481.
- [44] H. Tsukahara, Semiparametric estimation in copula models, *Canad. J. Statist.* 33 (2005) 357–375.
- [45] A. van der Vaart, *Asymptotic Statistics*, University Press, Cambridge, 2007.
- [46] A. van der Vaart, J. Wellner, *Weak Convergence and Empirical Processes*, Springer, New York, 1996.
- [47] N. Veraverbeke, M. Omelka, I. Gijbels, Estimation of a conditional copula and association measures, *Scand. J. Stat.* 38 (2011) 766–780.