Série des Documents de Travail

n° 2015-12
Single-index copulae

J-D.Fermanian[1]

O.Lopez[2]

[1] CREST, Laboratoire de Finance et d'Assurance, France, E-mail: jean-david.fermanian@ensae.fr.

[2] CREST, Laboratoire de Finance et d'Assurance, France, E-mail: olivier.lopez@ensae.fr.

Single-index copulae

Jean-David Fermanian[1], Olivier Lopez[2]

December 22, 2015

Abstract

We introduce so-called "single-index copulae". They are semi-parametric conditional copulae whose parameter is an unknown "link" function of a univariate index only. We provide 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.

Key words: Conditional copulae, single-index models, kernel smoothing.

Short title: Single-index copulae.

[1] CREST, Laboratoire de Finance et d'Assurance, France, E-mail: jean-david.fermanian@ensae.fr.

[2] CREST, Laboratoire de Finance et d'Assurance, France, E-mail: olivier.lopez@ensae.fr.

1 Introduction

1.1 The framework of single-index dependence functions

Since Sklar's theorem (1959), copula modeling has emerged as a very active field in theoretical and applied research. Applications in finance, insurance, biology, medicine, hydrology, etc., are now countless. The origin of this success is the ability of splitting specification/inference/testing of a (complex) multivariate model into two separate (simpler) problems: the management of marginal distributions on one side, and the modelling of the dependence structure (copula) on the other side. See the books of Joe (1997) or Nelsen (1998) for a rigorous presentation of this field.

In practice, it is usual to introduce explanatory variables (also called "covariates") in a multivariate model, particularly in econometrics or financial risk management. When we focus on the effect of these covariates on the underlying copulae, we need the concept of conditional copulae immediately (Patton, 2006). Conditional copulae are a natural way of linking conditional marginal distributions to get a multivariate conditional law and they have been applied extensively (see the surveys of Patton 2009, 2012). Recently, the rise of vine models (Aas et alii, 2009) has extended the scope and the importance of conditional copulae.

Until now, most conditional copula models were parametric. For instance, they specify a given functional link between the copula parameters and an index $\beta'\boldsymbol{z}$, $\boldsymbol{z}$ being the underlying vector of covariates: see Rockinger and Jondeau (2006), Patton (2006), Rodriguez (2007), Batram (2007), among others. Alternatively, a fully nonparametric point of view has been proposed by Fermanian and Wegkamp (2012) or Gijbels et alii (2011). Such techniques rely on kernel smoothing, local polynomials or other tools in functional estimation. As a consequence, when the dimension of the vector of covariates of larger than three, such methods suffer from the well-know curse of dimension, 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 have to be estimated, avoiding the curse of dimension. Note that Acar et alii (2011, 2013) 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 is increasing quickly with the dimension of $\boldsymbol{z}$.

To fix the ideas and the notations, let us consider an i.i.d. sample of observations $(\boldsymbol{X}_i, \boldsymbol{Z}_i)$ in $\mathbb{R}^d \times \mathbb{R}^p$, that are drawn from the law of $(\boldsymbol{X}, \boldsymbol{Z})$. The vector $\boldsymbol{X}$ represents the endogenous vector, and $\boldsymbol{Z}$ is the vector of covariates. We will be interested in the evaluation of the law of $\boldsymbol{X}$ conditional on $\boldsymbol{Z} = \boldsymbol{z}$, for arbitrary vectors $\boldsymbol{z}$. This conditional cdf is denoted by $F(\cdot|\boldsymbol{z})$. The (marginal) law of X_k, $k = 1, \ldots, d$, given $\boldsymbol{Z} = \boldsymbol{z}$, will be denoted by $F_k(\cdot|\boldsymbol{z})$. We introduce the unobserved random vector $\boldsymbol{U}_{\boldsymbol{z}} = (U_{1,\boldsymbol{z}}, \ldots, U_{d,\boldsymbol{z}})$, where $U_{k,\boldsymbol{z}} = F_k(X_k|\boldsymbol{z})$, $k = 1, \ldots, d$. To simplify notations and when there is no ambiguity, $\boldsymbol{U}_{\boldsymbol{z}}$ will be often denoted by $\boldsymbol{U}$. By definition, the law of $\boldsymbol{U}_{\boldsymbol{z}}$ conditional on $\boldsymbol{Z} = \boldsymbol{z}$ is the conditional copula of $\boldsymbol{X}$ knowing $\boldsymbol{Z} = \boldsymbol{z}$, denoted by $C(\cdot|\boldsymbol{z})$.

First, we live in a parametric framework. A natural model specification would be to assume that, for any $\boldsymbol{u} \in [0,1]^d$ and any $\boldsymbol{z} \in \mathbb{R}^p$,

$$C(\boldsymbol{u}|\boldsymbol{z}) = C_{\theta(\boldsymbol{z})}(\boldsymbol{u}),$$

where $\theta : \mathbb{R}^p \to \mathbb{R}^q$ maps the vector of covariates to the (true) parameter of the conditional copula knowing $\boldsymbol{Z} = \boldsymbol{z}$, and $\mathcal{C} = \{C_\theta : \theta \in \Theta \subset \mathbb{R}^q\}$ denotes a parametric family of copulae. The copula density of C_θ is supposed to exist and is denoted by c_θ. To simplify, this density is assumed to be continuous for every $\theta \in \Theta$, and Θ will be a compact subset.

Second, since the single-index assumption will be related to the dependence function among the components of $\boldsymbol{X}$, given the covariates, this means there exists an unknown function ψ s.t.

$$\theta(\boldsymbol{z}) = \psi(\beta_0, \beta_0'\boldsymbol{z}), \tag{1.1}$$

where the true parameter $\beta_0 \in \mathcal{B}$, a compact subset in $\mathbb{R}^m$. To identify the parameter β_0, let us assume that the first component of β_0, that is $\beta_{0,1}$, is equal to one. Under the single-index assumption (1.1), $C(\cdot|\boldsymbol{z})$ does depend on $(\beta, \beta'\boldsymbol{z})$ if the underlying parameter is β. Therefore, this function will be denoted equivalently $C_\beta(\cdot|\beta'\boldsymbol{z})$ too.

We stress that Assumption (1.1) does not mean that $C(\cdot|\boldsymbol{z})$, the conditional copula of $\boldsymbol{X}$ knowing $\boldsymbol{Z} = \boldsymbol{z}$, is equal to the conditional copula of $\boldsymbol{X}$ knowing $\beta_0'\boldsymbol{Z} = \beta_0'\boldsymbol{z}$ (denoted by $\tilde{C}(\cdot|\beta_0'\boldsymbol{z})$). Indeed, in the former case, the relevant margins are the cdfs' $F_k(\cdot|\boldsymbol{z})$, $k = 1, \ldots, d$, and in the latter case, we need to consider the cdfs' $\tilde{F}_k(\cdot|\beta_0'\boldsymbol{z}) : x_k \mapsto P(X_k \leq x_k|\beta_0'\boldsymbol{z})$. To avoid any confusion, let us denote $\tilde{\boldsymbol{U}}_\beta = (\tilde{F}_1(X_1|\beta'\boldsymbol{Z}), \ldots, \tilde{F}_d(X_d|\beta'\boldsymbol{Z}))$, and $\tilde{C}(\cdot|\beta'\boldsymbol{Z} = y)$ will be the copula of $\tilde{\boldsymbol{U}}_\beta$ knowing $\beta'\boldsymbol{Z} = y$. The conditional copulae $C(\cdot|\boldsymbol{z})$ and $\tilde{C}(\cdot|\beta_0'\boldsymbol{z})$ are identical only when $\boldsymbol{Z}$ provides the same information as $\beta_0'\boldsymbol{Z}$ to explain every margin X_k, i.e. when $F_k(\cdot|\boldsymbol{z}) = \tilde{F}_k(\cdot|\beta_0'\boldsymbol{z})$ a.e. for every k: see Fermanian and Wegkamp (2012) for a discussion.

1.2 The M-estimate criterion

Single-index models are well-known in the world of semiparametric statistics. The theory of M-estimators has started with the seminal papers of Klein and Spady (1993) in the case of the so-called binary response model, and Ichimura (1993) for the general single-index regression model. Sherman (1994), Delecroix and Hristache (1999) extended this approach. Härdle et alii (1993) and Delecroix et alii (1999) discussed the choice of the bandwidth for the nonparametric estimation of the link function. Alternatively, the so-called average derivative method has been developped in parallel by Stoker (1986), Powell, Stock and Stoker (1989), Härdle and Stoker (1989), etc.

In this paper, we will 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 $\boldsymbol{U}$, i.e. $\boldsymbol{U}_i$, $i = 1, \ldots, n$, then our "naive" estimator of β_0 could be

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

for some function $\hat{\psi}$ that estimates $\psi(\cdot,\cdot)$ consistently.

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

$$\hat{\beta} = \arg\max_{\beta\in\mathcal{B}} \sum_{i=1}^n \hat{\omega}_{i,n} \ln c_{\hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)}(\hat{\boldsymbol{U}}_i), \tag{1.2}$$

for some sequence of trimming functions $\hat{\omega}_{i,n}$. Typically, they are of the type $\hat{\omega}_{i,n} = \mathbf{1}(\hat{\boldsymbol{U}}_i \in \mathcal{E}_n, \boldsymbol{Z}_i \in \mathcal{Z})$, 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 allow to control some boundary effects and the uniform convergence of our kernel estimates. For technical reasons, we will choose strictly increasing trimmings on the $\boldsymbol{U}$-side, i.e. $\cup_n \mathcal{E}_n = (0,1)^d$. This choice makes it necessary to control explicitly the behavior of $\boldsymbol{U}$ close to the boundary of $[0,1]^d$. This pretty delicate task will require several regularity assumptions but the problem has already been met in the literature (see Tsukahara 2005, for instance). Moreover, we will set a fixed trimming for $\mathcal{Z}$ (i.e. $\mathcal{Z} \subset \mathbb{R}^p$ strictly). This will not create any bias, because the law of the $\boldsymbol{U}$

knowing $\boldsymbol{Z} \in \mathcal{Z}$, is just $c_{\psi(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u})\mathbf{1}(\boldsymbol{z} \in \mathcal{Z})/\mathbb{P}(\boldsymbol{Z} \in \mathcal{Z})$. Thus, it depends on the true parameter β_0. See Assumption 1.

Remark 1.1 *Actually, fixed trimming functions for $\hat{\boldsymbol{U}}_i$ 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, beside a small loss of efficiency, this would forbid to model the tail dependence behaviors, a feature that is important in a lot of fields. That is why we have chosen $\hat{\beta}$, as defined by (1.2).*

2 Consistency

2.1 The convergence of single-index estimators

Assumption 1 *Let us set $\mathcal{Z} := [-M, M]^p$ and $\mathcal{E}_n = [\nu_n, 1-\nu_n]^d$ for some positive sequence (ν_n), $\nu_n \in (0, 1/2)$, $\nu_n \to 0$. The trimming functions are $\omega_n : [0,1]^d \times \mathbb{R}^p \to [0,1]$, $(\boldsymbol{u}, \boldsymbol{z}) \mapsto \mathbf{1}(\boldsymbol{u} \in \mathcal{E}_n, \boldsymbol{z} \in \mathcal{Z})$.*

We set $\hat{\omega}_{i,n} = \omega_n(\hat{\boldsymbol{U}}_i, \boldsymbol{Z}_i)$ simply. For the sake of completeness, we introduce $\omega_{i,n} := \omega_n(\boldsymbol{U}_i, \boldsymbol{Z}_i)$, the trimming function when $\boldsymbol{U}_i$ is known, and $\omega_i = \omega_{i,\infty} = \mathbf{1}(\boldsymbol{Z}_i \in \mathcal{Z})$.

Assumption 2 *The parameter β_0 is identifiable, i.e. two different parameters induce two different laws of $\boldsymbol{U}_{\boldsymbol{Z}}$, knowing $\boldsymbol{Z} \in \mathcal{Z}$. The function $M : \mathcal{B} \to \mathbb{R}$, $\beta \mapsto E[\ln c_{\psi(\beta,\beta'\boldsymbol{Z})}(\boldsymbol{U}_{\boldsymbol{Z}}) \,|\, \boldsymbol{Z} \in \mathcal{Z}]$ is continuous and uniquely maximized at $\beta = \beta_0$. There exists a measurable function h s.t., for every $\boldsymbol{z} \in \mathcal{Z}$,*

$$\sup_{\beta \in \mathcal{B}} |\ln c_{\psi(\beta,\beta'\boldsymbol{z})}(\boldsymbol{U}_{\boldsymbol{z}})| \leq h(\boldsymbol{U}_{\boldsymbol{z}}, \boldsymbol{z}), \text{ with } E[h(\boldsymbol{U}_{\boldsymbol{Z}}, \boldsymbol{Z}).\mathbf{1}(\boldsymbol{Z} \in \mathcal{Z})] < \infty. \tag{2.1}$$

The latter assumption is usual for maximum likelihood estimation purpose. The limiting objective function is here

$$M(\beta) := E\left[\ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i) \,|\, \boldsymbol{Z}_i \in \mathcal{Z}\right].$$

Note that, due to our trimming functions, we are dealing with a M-estimator of β instead of a usual MLE formally, at the cost of a (small) loss of efficiency.

Assumption 3

$$\sup_{\boldsymbol{z}\in\mathcal{Z}}\sup_{\beta\in\mathcal{B}}\left|\hat\psi(\beta,\beta'\boldsymbol{z})-\psi(\beta,\beta'\boldsymbol{z})\right| = o_P(1). \tag{2.2}$$

Moreover, the pseudo-observations $\hat U_{i,k}$ *belong to* $(0,1)$*,* $k=1,\ldots,d$*,* $i=1,\ldots,n$*, and there exists a deterministic sequence* (δ_n)*,* $\delta_n = o(\nu_n)$*, s.t.*

$$\sup_i |\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i|.\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z}) \leq \delta_n \;\; a.e. \tag{2.3}$$

These assumptions have to be checked for any particular single-index model (see the assumptions (A1) and (A2) in Subsection 2.2) and for any particular estimate of the marginal cdfs'.

Now, we recall the definition of reproducing u-shaped functions, as introduced in Tsukahara (2005).

Definition 2.1

- *A function* $f:(0,1)\to(0,\infty)$ *is called u-shaped if it is symmetric about* $1/2$ *and decreasing on* $(0,1/2]$.
- *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* M_β *such that* $r_\beta < M_\beta r$ *on* $(0,1)$*, then* r *is called a reproducing u-shaped function.*

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

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 some functions* $r,\tilde r_1,\ldots,\tilde r_d$ *in* $\mathcal{R}_d$ *s.t., for every* $\boldsymbol{u}\in(0,1)^d$*,*

$$\sup_{\theta\in\Theta}|\nabla_\theta \ln c_\theta(\boldsymbol{u})| \leq r(\boldsymbol{u}),\; E\left[r(\boldsymbol{U}_{\boldsymbol{Z}})\mathbf{1}(\boldsymbol{Z}\in\mathcal{Z})\right]<\infty,$$

$$\sup_{\theta\in\Theta}|\partial_{u_k}\ln c_\theta(\boldsymbol{u})| \leq \tilde r_k(\boldsymbol{u}),\; \textit{for every } k=1,\ldots,d, \; \textit{and}$$

$$\sup_{k=1,\ldots,d} E\left[U_k(1-U_k)\tilde r_k(\boldsymbol{U}_{\boldsymbol{Z}})\mathbf{1}(\boldsymbol{Z}\in\mathcal{Z})\right]<\infty.$$

The latter conditions of moments are easily satisfied for most copula models. They are close to those of Assumption (A.1) in Tsukahara (2005).

Theorem 2.2 *Under the assumptions 1-4, the estimator $\hat\beta$ given by (1.2) tends to β_0 in probability, when n tends to the infinity.*

Proof. For inference purpose 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 $\boldsymbol{U}_i$ were observable. Let us introduce

$$\begin{aligned} M_n(\beta) &:= \frac{1}{n_i+1}\sum_{i=1}^n \hat\omega_{i,n} \ln c_{\hat\psi(\beta,\beta'\boldsymbol{z}_i)}(\hat{\boldsymbol{U}}_i), \\ M_n^*(\beta) &:= \frac{1}{n_i+1}\sum_{i=1}^n \hat\omega_{i,n} \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i), \\ M_n^{**}(\beta) &:= \frac{1}{n_i+1}\sum_{i=1}^n \omega_i \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i). \end{aligned}$$

Note that $\hat\beta$ is the optimizer of $M_n(\cdot)$ because neither n_i or $\hat n_i$ is a function of the underlying parameter β. By assumption, β_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 provide

$$\begin{aligned} |M_n(\beta) - M_n^*(\beta)| &\le \frac{1}{n_i+1}\sum_{i=1}^n \hat\omega_{i,n} \sup_{\theta\in\Theta}\left|\frac{\nabla_\theta c_\theta(\hat{\boldsymbol{U}}_i)}{c_\theta(\hat{\boldsymbol{U}}_i)}\right| \cdot \left|\hat\psi(\beta,\beta'\boldsymbol{Z}_i) - \psi(\beta,\beta'\boldsymbol{Z}_i)\right| \\ &+ \frac{1}{n_i+1}\sum_{i=1}^n \left|\hat\omega_{i,n}\frac{\nabla_{\boldsymbol{u}} c_{\psi(\beta,\beta'\boldsymbol{z}_i)}}{c_{\psi(\beta,\beta'\boldsymbol{z}_i)}}(\boldsymbol{U}_i^*)\cdot(\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i)\right| := T_1(\beta) + T_2(\beta), \end{aligned}$$

for some vectors $\boldsymbol{U}_i^*$ s.t. $|\boldsymbol{U}_i - \boldsymbol{U}_i^*| \le |\boldsymbol{U}_i - \hat{\boldsymbol{U}}_i|$ for all i.

Due to Assumption 3, the vectors $\hat{\boldsymbol{U}}_i$ and $\boldsymbol{U}_i^*$ we consider in the summations above belong to a neighborhood of $\boldsymbol{U}_i$ whose size may be chosen uniformly wrt i. To be specific, since $\delta_n = o(\nu_n)$, we can assume that, for every i s.t. $\hat\omega_{in} = 1$ and every $k = 1,\ldots,d$, we have

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

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

$$r_k(\hat U_{i,k}) \le r_k(U_{i,k}/2) \quad \text{if } \hat U_{i,k} \le 1/2, \ \text{ and}$$

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

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

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

which is integrable. Since n_i/n tends to a positive constant a.e., and due to Assumption (2.2), we deduce $\sup_\beta T_1(\beta) = o_P(1)$.

By a slightly more subtle reasoning, we can obtain $\sup_\beta T_2 = o_P(1)$. Indeed, due to Assumption 4 and for every $\varepsilon > 0$, there exists $\eta \in (0,1/2)$ s.t.

$$\sup_{k=1,\ldots,d} E[\tilde{r}_k(\boldsymbol{U}_{\boldsymbol{Z}})\{U_k\mathbf{1}(U_k<\eta)+(1-U_k)\mathbf{1}(U_k>1-\eta)\}.\mathbf{1}(\boldsymbol{Z}\in\mathcal{Z})] < \varepsilon.$$

Invoking Assumptions 3 and the Law of Large Numbers, we have

$$\begin{aligned}\sup_{\beta\in\mathcal{B}} T_2(\beta) &\leq \frac{1}{n_i+1}\sum_{k=1}^d\sum_{i=1}^n \hat{\omega}_{i,n}\,|\tilde{r}_k(\boldsymbol{U}_i^*)|\cdot\left|\hat{U}_{i,k}-U_{i,k}\right| \\ &\leq \; \frac{1}{n_i+1}\sum_{k=1}^d\sum_{i=1}^n \omega_i\,|\tilde{r}_k(\boldsymbol{U}_i)|\cdot\Big(\left|\hat{U}_{i,k}-U_{i,k}\right|\mathbf{1}\{\eta\leq U_{i,k}\leq 1-\eta\} \\ &+ \; U_{i,k}\mathbf{1}(U_{i,k}<\eta)+(1-U_{i,k})\mathbf{1}(U_{i,k}>1-\eta)\Big) \\ &\leq \; \frac{\delta_n}{n_i+1}\sum_{k=1}^d\sum_{i=1}^n \omega_i\,|\tilde{r}_k(\boldsymbol{U}_i)|\cdot\mathbf{1}\{\eta\leq U_{i,k}\leq 1-\eta\}+2\varepsilon,\end{aligned}$$

for n sufficiently large and a.e. Note the r.h.s. of the latter inequality does not depend on β, and that we have used $\hat{U}_{i,k}\in(0,1)$ for every $i=1,\ldots,n$ and $k=1,\ldots,d$. Moreover, we have

$$\frac{1}{n_i+1}\sum_{k=1}^d\sum_{i=1}^n \omega_i\,|\tilde{r}_k(\boldsymbol{U}_i)|\cdot\mathbf{1}\{\eta\leq U_{i,k}\leq 1-\eta\} \leq \frac{\eta}{n_i+1}\sum_{k=1}^d\sum_{i=1}^n \omega_i U_{i,k}(1-U_{i,k})\,|\tilde{r}_k(\boldsymbol{U}_i)|\,,$$

that is $O_P(1)$ due to the LLN and Assumption 4. Since ε may be arbitrarily small, we get $\sup_\beta T_2(\beta) = o_P(1)$, and we have proved $\sup_{\beta\in\mathcal{B}}|M_n(\beta)-M_n^*(\beta)| = o_P(1)$.

Second, due to Assumption 2 and for every $\varepsilon > 0$, we have

$$
\begin{aligned}
&\mathbb{P}\left(|\frac{1}{n}\sum_{i=1}^n (\hat\omega_{i,n} - \omega_{i,n}) \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i)| > \varepsilon \right) \\
&\leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}(\boldsymbol{U}_i \in \mathcal{E}_n, \hat{\boldsymbol{U}}_i \notin \mathcal{E}_n, \boldsymbol{Z}_i \in \mathcal{Z}) | \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i)| > \varepsilon/2 \right) \\
&+ \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}(\boldsymbol{U}_i \notin \mathcal{E}_n, \hat{\boldsymbol{U}}_i \in \mathcal{E}_n, \boldsymbol{Z}_i \in \mathcal{Z}) | \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i)| > \varepsilon/2 \right) \\
&\leq \frac{2}{\varepsilon} E\left[\left\{ \mathbf{1}(\boldsymbol{U}_i \in \mathcal{E}_n, \hat{\boldsymbol{U}}_i \notin \mathcal{E}_n) + \mathbf{1}(\boldsymbol{U}_i \notin \mathcal{E}_n, \hat{\boldsymbol{U}}_i \in \mathcal{E}_n) \right\} \cdot \mathbf{1}(\boldsymbol{Z}_i \in \mathcal{Z}) | \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i)| \right].
\end{aligned}
$$

But, due to (2.3), we have for any i

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

that tends to zero a.e. when n tends to the infinity. Invoking the dominated convergence Theorem and (2.1), we get

$$
\frac{1}{n}\sum_{i=1}^n (\hat\omega_{i,n} - \omega_{i,n}) \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i) = o_P(1).
$$

Similarly, we prove $n^{-1}\sum_{i=1}^n (\omega_{i,n} - \omega_i) \ln c_{\psi(\beta,\beta'\boldsymbol{z}_i)}(\boldsymbol{U}_i) = o_P(1)$. We deduce easily $\sup_{\beta\in\mathcal{B}} |M_n^*(\beta) - M_n^{**}(\beta)| = o_P(1)$ 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 Newey and McFadden (1994) tells us that (2.1) insures that $\sup_{\beta\in\mathcal{B}} |M_n^{**}(\beta) - M(\beta)| = o_P(1)$. Therefore, we get that $\hat\beta$ tends to β_0 in probability. ∎

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

2.2 Estimation of the link function ψ

For inference purpose, we need a relationship between the previous link function ψ and some quantities that can be estimated empirically. Typically, there are two possibilities.

(A1) There exists a known functional Ψ s.t., for any $\beta \in \mathbb{R}^m$,

$$
\psi(\beta, \beta'\boldsymbol{z}) = \Psi\left(C_\beta(\cdot|\beta'\boldsymbol{z}) \right). \tag{2.4}
$$

(A2) There exists a known functional Ψ s.t., for any $\beta \in \mathbb{R}^m$,

$$\psi(\beta, \beta' \boldsymbol{z}) = \Psi\left(H_\beta(\cdot|\beta' \boldsymbol{z})\right), \tag{2.5}$$

where $H_\beta(\cdot|y)$ denotes the cdf of $(\boldsymbol{X}, \boldsymbol{Z})$ conditional on $\beta' \boldsymbol{Z} = y$.

In numerous practical situations, Assumptions (2.4) and (2.5) are simply moment-like conditions, as in the GMM methodology: there is a map $g : \mathbb{R}^{\bar{m}} \to \mathbb{R}^q$, $\bar{m} \geq m$, such that

$$\theta(\boldsymbol{z}) = g(m_1(\beta_0, \beta_0' \boldsymbol{z}), \dots, m_{\bar{m}}(\beta_0, \beta_0' \boldsymbol{z})),$$

where $m_k(\beta, y) \in \mathbb{R}$, $k = 1, 2, \dots$, are "moment" relations based on the underlying distributions. In the case of (2.4), these moment relations are directly linked to conditional copulae by

$$\begin{aligned} m_k(\beta, y) &= E[\chi_k(\boldsymbol{U}_{\boldsymbol{Z}}, \beta' \boldsymbol{Z})|\beta' \boldsymbol{Z} = y] = E[E[\chi_k(\boldsymbol{U}_{\boldsymbol{Z}}, \beta' \boldsymbol{Z})|\boldsymbol{Z}]|\beta' \boldsymbol{Z} = y] \\ &= E[\int \chi_k(\boldsymbol{u}, \beta' \boldsymbol{Z})\, C(d\boldsymbol{u}|\boldsymbol{Z})|\beta' \boldsymbol{Z} = y] = \int \chi_k(\boldsymbol{u}, y)\, C_\beta(d\boldsymbol{u}|\beta' \boldsymbol{Z} = y), \end{aligned} \tag{2.6}$$

for some known functions χ_k, $k = 1, \dots, \bar{m}$.

In the case of (2.5), there exist some "moments" $m_k(\beta, y) \in \mathbb{R}$, $k = 1, 2, \dots$, based on the underlying distribution of $(\boldsymbol{X}, \boldsymbol{Z})$ given $\beta' \boldsymbol{Z} = y$:

$$m_k(\beta, y) = E[\chi_k(\boldsymbol{X}, \boldsymbol{Z})|\beta' \boldsymbol{Z} = y] = \int \chi_k(\boldsymbol{x}, \boldsymbol{z})\, H_\beta(d\boldsymbol{x}, d\boldsymbol{z}|\beta' \boldsymbol{Z} = y). \tag{2.7}$$

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

Example: Spearman's rho.

A natural candidate is given by $m_k(\beta, \beta' \boldsymbol{z}) = \rho(\beta, \beta' \boldsymbol{z})$, a multivariate extension of the usual Spearman's rho, defined by

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

Through a d-dimensional integration by parts, check that this moment is of the type (2.6). Therefore, we work under (A1). Other definitions of Spearman's rho are possible with an arbitrary dimension d: see Schmidt and Schmid (2007), for instance. Note that, when $d = 2$, $\rho(\beta, y)$ is simply the correlation between $F_1(X_1|\boldsymbol{Z})$ and $F_2(X_2|\boldsymbol{Z})$ given $\beta'\boldsymbol{Z}$. Therefore, it can be estimated relatively easily, at least when the dimension of $\boldsymbol{Z}$ is "reasonable".

Example: Kendall's tau. To fix the ideas, let us assume $d = 2$. The Kendall's tau of $\boldsymbol{X}$ conditional on $\boldsymbol{Z} = \boldsymbol{z}$ is

$$\tau_{\boldsymbol{z}} = 4\int C(\boldsymbol{u}|\boldsymbol{z})C(d\boldsymbol{u}|\boldsymbol{z}) - 1 = 4\int C_\beta(\boldsymbol{u}|\beta'\boldsymbol{z})C_\beta(d\boldsymbol{u}|\,\beta'\boldsymbol{z}) - 1. \tag{2.8}$$

Since it depends only on $\beta'\boldsymbol{z}$, it is denoted by $\tau(\beta, \beta'\boldsymbol{z})$. Then, managing Kendall's tau, we work under Assumption (A1) usually. The parameter β and then $\psi(\beta, \beta'\boldsymbol{z})$ can be estimated empirically, replacing $C_\beta(\cdot|\beta'\boldsymbol{z})$ by an empirical counterpart in the previous integral.

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

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

This implies that the Kendall's tau of $\boldsymbol{X}$ given $\beta'\boldsymbol{Z} = y$ is $\tau(\beta, y)$, under (1.1). Incidentally, we have proved that

$$\int C_\beta(\boldsymbol{u}|y)\,C_\beta(d\boldsymbol{u}|y) = \int \tilde{C}_\beta(\boldsymbol{u}|y)\,\tilde{C}_\beta(d\boldsymbol{u}|y), \text{ and}$$

$$\tau(\beta, \beta'\boldsymbol{z}) = 4\int \tilde{C}_\beta(\boldsymbol{u}|y)\,\tilde{C}_\beta(d\boldsymbol{u}|y) - 1. \tag{2.9}$$

Moreover, since

$$E[\mathbf{1}(X_1 > Y_1, X_2 > Y_2)|\beta'\boldsymbol{Z} = y] = \int H_\beta(\boldsymbol{x}, +\infty|\beta'\boldsymbol{Z} = y)\,H_\beta(d\boldsymbol{x}, +\infty|\beta'\boldsymbol{Z} = y),$$

we recognize Assumption (A2), and

$$\tau(\beta, \beta'\boldsymbol{z}) = 4\int H_\beta(\boldsymbol{x}, +\infty|\beta'\boldsymbol{z})\,H_\beta(d\boldsymbol{x}, +\infty|\beta'\boldsymbol{z}) - 1. \tag{2.10}$$

In other terms, Kendall's tau are of the two types (A1) and (A2) simultaneously. And the relations (2.9) and (2.10) will be very useful in practice. Indeed, the estimation of $H_\beta(\cdot|y)$ or $\tilde{C}_\beta(\cdot|y)$ is less demanding than the non parametric estimation of $C_\beta(\cdot|\beta'\boldsymbol{z})$: an empirical counterpart of $H_\beta(\boldsymbol{x}|y)$ or $\tilde{C}_\beta(\boldsymbol{u}|y)$ does not suffer from the curse of dimension because it necessitates only conditioning subsets in $\mathbb{R}$, contrary to $C_\beta(\boldsymbol{u}|y)$ that involves conditioning wrt $\boldsymbol{z} \in \mathbb{R}^p$ to manage its marginal laws.

In dimension d, many Kendall's tau can be built, but the same reasonings and conclusions apply. These Kendall's tau may be associated to any couple of variables (X_i, X_j), $i, j = 1, \ldots, d$, $i \neq j$. Or they can be defined formally as in (2.8), with d-dimension integrals, or even d'-dimension integrals, $d' < d$ if we focus on some sub-vectors of $\boldsymbol{X}$. Globally, all such quantities are linear function of $\int C(\boldsymbol{u}_I, \mathbf{1}_{\bar{I}}|\boldsymbol{z})\, C(d\boldsymbol{u}_I, \mathbf{1}_{\bar{I}}|\boldsymbol{z})$, where I is a subset of $\{1, \ldots, d\}$ and $\bar{I}$ is its complement [1]. These dependence measures are candidates to provide convenient moments. Note the two usual generalizations of Kendall's tau in dimension d: the first one has been proposed by Joe (1990) as

$$\tau_d(\boldsymbol{z}) := \frac{1}{2^d - 1} \left\{ 2^d \int C(\boldsymbol{u}|\boldsymbol{z}) C(d\boldsymbol{u}|\boldsymbol{z}) - 1 \right\}, \tag{2.11}$$

and the second one has been introduced by Kendall and Babington Smith (1940) as the average value of Kendall's tau over all possible couples (X_k, X_l), $k, l = 1, \ldots, d$, $k \neq l$. See Genest et al. (2011) for details and complementary results.

In practice, the underlying copulae often depend on a few parameters only, say one or two (Archimedean copulae, typically). In the latter case, their Kendall's tau and/or Spearman's rho are sufficient to identify the underlying copula parameters. And there often exists an explicit one-to-one relationship between θ and the latter dependence measures. But, obviously, other moments may be considered, particularly some functionals of the conditional copula functions only.

Now, let us specify our estimator $\hat{\psi}$. The simplest solution we adopt is to invoke kernel-type regression functions. Under (A1), we can replace simply the conditional copula $C_\beta(\cdot|\beta'\boldsymbol{Z} = y)$ by a consistent estimator $\hat{C}(\cdot|\beta'\boldsymbol{Z} = y)$. Several candidates exist in the literature. Historically, Fermanian and Wegkamp (2006, published in 2012) were the first ones to propose a nearest neigbour estimator of conditional copulae. Gijbels et alii (2011) introduced other non-parametric estimates, including Nadaraya-Watson, Gasser-Müller, etc.

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

$$\hat{H}_\beta(\boldsymbol{x}, \boldsymbol{z}|y) = \sum_{j=1}^{n} w_{\beta,j,n}(y) \mathbf{1}(\boldsymbol{X}_j \leq \boldsymbol{x}, \boldsymbol{Z}_j \leq \boldsymbol{z}), \tag{2.12}$$

[1] Obviously, $\boldsymbol{u}_I, \mathbf{1}_{\bar{I}}$ denotes a d-dimensional vector whose components are u_k when $k \in I$, and are equal to one otherwise.

$$w_{\beta,j,n}(y) = K\left(\frac{\beta'\boldsymbol{Z}_j - y}{h_n}\right) / \sum_{l=1}^{n} K\left(\frac{\beta'\boldsymbol{Z}_l - y}{h_n}\right),$$

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

To check Condition (2.2), we have to rely on the functional link between the parameter ψ and the underlying distributions, as evaluated under (A1) and/or (A2). This depends on the regularity of the corresponding functionals Ψ and on the uniform distance between the conditional empirical cdfs' and true ones.

For instance, under (A2), assume Ψ is Lipschitz, with a Lipschitz constant λ (at least when $\beta \in \boldsymbol{\mathcal{B}}$ and $\boldsymbol{z} \in \mathcal{Z}$, and then $\boldsymbol{z}'\beta$ belongs to a real compact subset). For such couples $(\beta, \boldsymbol{z})$, we have

$$|\hat{\psi}(\beta, \beta'\boldsymbol{z}) - \psi(\beta, \beta'\boldsymbol{z})| \leq \lambda \|\hat{H}_\beta(\cdot|\beta'\boldsymbol{z}) - H_\beta(\cdot|\beta'\boldsymbol{z})\|_\infty$$

Assuming $\hat{H}_\beta$ is given by (2.12) and applying Corollary 3 in Einmahl and Mason (2005), we obtain

$$\sup_i |\hat{\psi}(\beta, \beta'\boldsymbol{Z}_i) - \psi(\beta, \beta'\boldsymbol{Z}_i)|\omega_{i,n} \leq \lambda \sup_i \|\hat{H}_\beta(\cdot|\beta'\boldsymbol{Z}_i) - H_\beta(\cdot|\beta'\boldsymbol{Z}_i)\|_\infty \omega_{i,n} \longrightarrow 0,$$

a.e. and uniformly wrt $\beta \in \boldsymbol{\mathcal{B}}$. This will be sufficient to satisfy (2.2).

Note that Ψ is Lipschitz in the case of Kendall's tau. Indeed, through an integration by parts and for two cdfs' H and H' (for which H or H' is continuous), we observe that

$$\begin{aligned} & |\int H(\cdot|\boldsymbol{z})\, dH(\cdot, |\boldsymbol{z}) - \int H'(\cdot|\boldsymbol{z})\, dH'(\cdot, |\boldsymbol{z})| \\ \leq\ & |\int (H - H')(\cdot|\boldsymbol{z})\, dH(\cdot, |\boldsymbol{z})| + |\int (H - H')(\cdot|\boldsymbol{z})\, dH'(\cdot, |\boldsymbol{z})| \\ \leq\ & \|(H - H')(\cdot|\boldsymbol{z})\|_\infty \cdot \left(\int |dH|(\cdot, |\boldsymbol{z}) + \int |dH'|(\cdot, |\boldsymbol{z})\right) \leq 2\|(H - H')(\cdot|\boldsymbol{z})\|_\infty. \end{aligned}$$

More generally, under (A2) and if Ψ is Hadamard differentiable, there exist continuous linear maps $\dot{\Psi}_i$ s.t.

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

Under some additional conditions (particularly on the $\dot{\Psi}_i$), we get typically the uniformity of the latter identity wrt $\boldsymbol{Z}_i \in \mathcal{Z}$. But, thanks to Theorem 3 in Einmahl and Mason (2005), there exists a sequence of positive numbers (a_n), $a_n \to 0$, s.t.

$$a_n \sup_{\beta\in\mathcal{B}} \sup_{\boldsymbol{z}\in\mathcal{Z}} \|(\hat{H}-H)_\beta(\cdot|\beta'\boldsymbol{z})\|_\infty \longrightarrow 0$$

a.e. when $n \to 0$. The latter result is true uniformly wrt bandwith sequences (h_n) s.t. $nh_n/\ln n >> 1$ and $h_n \to 0$. Therefore, (2.2) is usually satisfied when Ψ is Hadamard differentiable.

Note that the uniform consistency of the conditional copula function, simultaneously wrt to its argument and the conditioning value, is not available in the literature. Therefore, checking Condition (2.2) under (A1) is more difficult than under (A2).

2.3 The choice of the pseudo-estimations $\hat{\boldsymbol{U}}$

By assumption, β will be the index of the underlying dependence functions (copulae) only. Therefore, $\hat{\boldsymbol{U}}_i$ will not depend on β. Now, let us discuss the possible choices for $\hat{\boldsymbol{U}}_i$, $i=1,\dots,n$. Actually, in Section 3, we will consider a generic class of estimates s.t., for all $k=1,\dots,d$,

$$\hat{F}_k(x|\boldsymbol{z}) - F_k(x|\boldsymbol{z}) = \frac{1}{n}\sum_{j=1}^n a_{k,n}(\boldsymbol{X}_j,\boldsymbol{Z}_j,x,\boldsymbol{z}) + r_n(x,\boldsymbol{z}), \tag{2.13}$$

for some sequence $(r_n(x,\boldsymbol{z}))$ that tend to zero sufficiently quickly [2] uniformly in probability, and some particular functions $a_{k,n}$. Then, we will set $\hat{U}_{i,k} := \hat{F}_k(X_{i,k}|\boldsymbol{Z}_i)$, $i=1,\dots,n$, $k=1,\dots,d$. A lot of estimators of F_k, and then of $\boldsymbol{U}_i$, may be built and satisfy (2.13).

A first example of such estimates is given by parametric marginal conditional distributions: for every x and $\boldsymbol{z}$, $F_k(x|\boldsymbol{z}) = G_{k,\theta_k(\boldsymbol{z})}(x)$, for some family of cdfs' $\mathcal{G}_k := \{G_{k,\theta_k},\ \theta_k \in \Theta_k\}$. Since this model is parametric, the function θ_k depends on a vector of parameters $\eta_k \in \mathbb{R}^{m_k}$. With a light abuse of notations, set $\theta_k(\boldsymbol{z}) = \theta_k(\boldsymbol{z},\eta_k)$, and $\theta_k(\cdot,\eta)$ is known for every η. Assume we have found a consistent and asymptotically normal estimate $\hat{\eta}_k$, and set $\hat{F}_k(x|\boldsymbol{z}) = G_{k,\theta_k(\boldsymbol{z},\hat{\eta}_k)}(x)$. This implies $\hat{U}_{i,k} = G_{k,\theta_k(\boldsymbol{z}_i,\hat{\eta}_k)}(X_{i,k})$.

Clearly, for every i, there exists $\theta^*_{k,i}$ and η^*_k s.t.

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

where $|\theta_k(\boldsymbol{Z}_i,\eta_k) - \theta^*_{k,i}| \leq |\theta_k(\boldsymbol{Z}_i,\hat{\eta}_k) - \theta_k(\boldsymbol{Z}_i,\eta_k)|$ and $|\eta_k - \eta^*_k| \leq |\hat{\eta}_k - \eta_k|$.

[2] in particular to satisfy (2.3)

Typically, if $\sup_{\theta_k} |\nabla_\theta G_{k,\theta_k}(X_{i,k})|$ and $\sup_{\eta_k} |\partial_2 \theta_k(\boldsymbol{Z}_i, \eta_k)|$ are bounded in probability, the condition (2.3) is satisfied, even without trimming.

Moreover, in a lot of usual cases (M-estimates, e.g.), it can be checked by a limited expansion that the functions $\hat{F}_k(x|\boldsymbol{z})$ satisfy (2.13). Typically, for asymptotically normal estimators, we observe $nr_n(x, \boldsymbol{z}) = O_P(1)$, and this result may be uniform under some conditions of regularity concerning G and $\theta_k(\cdot)$. Such a choice of conditional margins induces the so-called estimator $\hat{\beta}^{(1)}$.

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

$$\hat{F}(\boldsymbol{x}|\boldsymbol{z}) = \sum_{j=1}^{n} w_{j,n}(\boldsymbol{z}) \mathbf{1}(\boldsymbol{X}_j \leq \boldsymbol{x}), \tag{2.14}$$

with weights

$$w_{j,n}(\boldsymbol{z}) = \boldsymbol{K}\left(\boldsymbol{Z}_j - \boldsymbol{z}, \boldsymbol{h}\right) / \sum_{l=1}^{n} \boldsymbol{K}\left(\boldsymbol{Z}_l - \boldsymbol{z}, \boldsymbol{h}\right), \tag{2.15}$$

where $\boldsymbol{K}$ is a multivariate kernel and $\boldsymbol{h} := (h_1, \ldots, h_p)$ is a p-vector of bandwidths $h_k > 0$. To simplify and w.l.o.g., we can restrict ourselves on products of p univariate kernels K_k i.e.

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

Therefore, nonparametric estimators of every marginal cdf $F_k(\boldsymbol{x}|\boldsymbol{z})$ are obtained by setting $\hat{F}_k(x|\boldsymbol{z}) = \hat{F}(x, +\infty_{(-k)}|\boldsymbol{z})$. The marginal "unfeasible" observations will be $U_{i,k} = F_k(X_{i,k}|\boldsymbol{Z}_i)$, and their estimated versions will be $\hat{U}_{i,k} = \hat{F}_k(X_{i,k}|\boldsymbol{Z}_i)$. In this case, it can be checked that (2.13) is satisfied.

Lemma 2.3 *For $k = 1, \ldots, d$, define $\hat{F}_k$ as*

$$\hat{F}_k(x|\boldsymbol{z}) = \sum_{j=1}^{n} w_{j,n}(\boldsymbol{z}) \mathbf{1}(X_{j,k} \leq x), \tag{2.16}$$

with the weights given by (2.15). If

- *$f_{\boldsymbol{Z}}$, the density of $\boldsymbol{Z}$, exists and is strictly positive on $\mathcal{Z}$. Moreover, it is s-times continuously differentiable, $s \geq 2$.*

- *For every real x and every k, the function $h(x, \cdot) : \boldsymbol{z} \mapsto P(X_k \leq x | \boldsymbol{Z} = \boldsymbol{z}) f_{\boldsymbol{Z}}(\boldsymbol{z})$, defined on $\mathcal{Z}$, is s-times continuously differentiable. Moreover,*

$$\sup_{x \in \mathbb{R}} \sup_{\boldsymbol{z} \in \mathcal{Z}} |d^s_{\boldsymbol{z}} h(x, \boldsymbol{z})| \text{ is bounded.}$$

- *the underlying kernel $K(\cdot,\mathbf{1})$ is continuous, bounded, $\int K(\boldsymbol{z},\mathbf{1})\,d\boldsymbol{z}=1$, of bounded variation and compactly supported* [3]. *Moreover, it is a multivariate s-order kernel, i.e.*

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

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

Then, for any $k=1,\ldots,d$, we have

$$\hat{F}_k(x|\boldsymbol{z})-F_k(x|\boldsymbol{z})=\frac{1}{n}\sum_{j=1}^{n} a_{k,n}(\boldsymbol{X}_j,\boldsymbol{Z}_j,x,\boldsymbol{z})+r_n(x,\boldsymbol{z}), \tag{2.17}$$

$$a_{k,n}(\boldsymbol{X}_j,\boldsymbol{Z}_j,\boldsymbol{x},\boldsymbol{z})=\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\left(\boldsymbol{K}\left(\boldsymbol{Z}_j-\boldsymbol{z},\boldsymbol{h}\right)\mathbf{1}(X_{j,k}\leq x)-E[\boldsymbol{K}\left(\boldsymbol{Z}_j-\boldsymbol{z},\boldsymbol{h}\right)\mathbf{1}(X_{j,k}\leq x)]\right.$$
$$\left.-\quad P(X_k\leq x|\boldsymbol{Z}=\boldsymbol{z})\{\boldsymbol{K}\left(\boldsymbol{Z}_j-\boldsymbol{z},\boldsymbol{h}\right)-E[\boldsymbol{K}\left(\boldsymbol{Z}_j-\boldsymbol{z},\boldsymbol{h}\right)]\}\right),$$

$$\sup_{x\in\mathbb{R},\boldsymbol{z}\in\mathcal{Z}}|r_n(x,\boldsymbol{z})|\leq C_1|\frac{\max(-\ln(\prod_{l=1}^p h_l),\ln\ln n)}{n\prod_{l=1}^p h_l}+\max_{l=1,\ldots,p}h_l^s|\ \text{a.e., and}$$

$$\sup_{x\in\mathbb{R},\boldsymbol{z}\in\mathcal{Z}}|\hat{F}_k(x|\boldsymbol{z})-F_k(x|\boldsymbol{z})|\leq C_2\left(\frac{\max(-\ln(\prod_{l=1}^p h_l),\ln\ln n)}{n\prod_{l=1}^p h_l}\right)^{1/2}+C_3\max_{l=1,\ldots,p}h_l^s,\ \text{a.e.} \tag{2.18}$$

for some positive constants C_1, C_2 and C_3.

Proof. By straightforward calculations, we get

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

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

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

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

that tends typically to $g=f_{\boldsymbol{Z}}$ and $h(x,\boldsymbol{z})=P(X_k\leq x|\boldsymbol{Z}=\boldsymbol{z})g(\boldsymbol{z})$. By invoking the equations (3.7) and (3.8) in the proof of Theorem 2 in Einmahl and Mason (2005), we get

[3] To be specific, this kernel has to be "regular" in the sense of Einmahl and Mason (2005), i.e. it has to satisfy their assumptions $K.i-K.iv$.

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, where

$$u_n^2 := \frac{n\prod_{l=1}^p h_l}{\max(-\ln(\prod_{l=1}^p h_l), \ln\ln n)}.$$

Note their remark 8 justifies the choice of different bandwidths for every component of $\boldsymbol{Z}$.

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 K and the regularity conditions on the underlying laws, we obtain easily that $r_{n,k}^{(2)}(x, \boldsymbol{z}) = O(\max_{l=1,\dots,p} h_l^s)$, providing the result. ■

As a consequence, the condition (2.3) is satisfied for the nonparametric versions on $\hat{\boldsymbol{U}}_i$ and for a wide range of bandwiths. Let us denote the associated estimator by $\hat{\beta}^{(2)}$.

Between the two previous polar cases, there exist a lot of candidates. For instance, to avoid the curse of dimension, it may be assumed that some marginal conditional distribution, say the k-th, will be given by a particular single-index model, but with a parameter $\beta_k \in \mathbb{R}^{m_k}$ that is different of β. Assume the latter index β_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

$$w_{\hat{\beta}_k,j,n}(y) = K\left(\frac{\hat{\beta}_k'\boldsymbol{Z}_j - y}{h}\right) / \sum_{l=1}^n K\left(\frac{\hat{\beta}_k'\boldsymbol{Z}_l - y}{h}\right),$$

for some kernel function $K : \mathbb{R} \to \mathbb{R}$ and some bandwidth $h > 0$. Obviously, $\hat{F}_{k,\hat{\beta}_k}(x|y)$ provides a nonparametric estimator of the cdf $F_{k,\beta_k}(x|y)$. In this case, $U_{k,\boldsymbol{z}} = F_k(X_k|\beta_k'\boldsymbol{z})$. To deal with pseudo-observations, we set $U_{i,k,\beta_k} = F_{k,\beta_k}(X_{i,k}|\beta_k'\boldsymbol{Z}_i)$, and $\hat{U}_{i,k} = \hat{F}_{k,\hat{\beta}_k}(X_{i,k}|\hat{\beta}_k'\boldsymbol{Z}_i)$. For some conditions of regularity, (2.13) can be verified, see for example Du and Akritas (2002) for such a representation in the more general case where censored data is present. When all margins are assumed single-index, let us denote by $\hat{\beta}^{(3)}$ the corresponding β estimator.

Now, let us check the conditions of Theorem 2.2, particularly Assumption 3, in some particular cases.

2.4 Examples

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

Example 1: the Gaussian copula

Let us consider a d-dimensional conditional copula model: for every $\boldsymbol{u}$ and $\boldsymbol{z}$ and with usual notations, the true underlying copula is

$$C_{\beta_0}(\boldsymbol{u}|\boldsymbol{Z}=\boldsymbol{z}) = C^G_{\Sigma(\boldsymbol{z})}(\boldsymbol{u}) = \Phi_{\Sigma(\boldsymbol{z})}\left(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)\right),$$

where the correlation matrix $\Sigma(\boldsymbol{z}) = [\theta_{k,l}(\boldsymbol{z})]_{1\leq k,l\leq d}$ depends on the index $\beta_0'\boldsymbol{z}$ only. With our previous notations, $\Sigma(\boldsymbol{z}) = \psi(\beta_0,\beta_0'\boldsymbol{z})$. It is well-known that every component $\theta_{k,l}(\boldsymbol{z})$ of $\Sigma(\boldsymbol{z})$ is a function of a Kendall's tau: $\theta_{k,l}(\boldsymbol{z}) = \sin(\pi\tau_{k,l}(\beta_0,\beta_0'\boldsymbol{z})/2)$, the conditional Kendall's tau that is associated to (X_k,X_l), knowing $\beta_0'\boldsymbol{Z} = \beta_0'\boldsymbol{z}$. The latter quantity can be estimated by standard nonparametric techniques, and then

$$\hat\psi(\beta,\beta'\boldsymbol{z}) = \left[\sin(\frac{\pi}{2}\hat\tau_{k,l}(\beta,\beta'\boldsymbol{z}))\right].$$

To be specific, we can choose

$$\hat\tau_{k,l}(\beta,y) := 4\int \hat C_{k,l}(u,v|\beta'\boldsymbol{Z}=y)\,\hat C_{k,l}(du,dv|\beta'\boldsymbol{Z}=y) - 1,$$

for some estimator $\hat C_{k,l}(\cdot|\beta'\boldsymbol{Z}=y)$ of the conditional copula of (X_k,X_l) given $\beta'\boldsymbol{Z}=y$. Alternatively, we can invoke an asymptotically equivalent estimator

$$\hat\tau_{k,l}(\beta,\beta'\boldsymbol{z}) := 4\sum_{i=1}^n\sum_{j=1}^n w_{i,h}(\beta'\boldsymbol{z})w_{j,h}(\beta'\boldsymbol{z})\mathbf{1}(X_{k,i}<X_{k,j},X_{l,i}<X_{l,j}) - 1,$$

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

$$w_{i,h}(y) := K\left(\frac{y-\beta'\boldsymbol{Z}_i}{h}\right)/\sum_{l=1}^n K\left(\frac{y-\beta'\boldsymbol{Z}_l}{h}\right).$$

See Gijbels et al. (2011) for alternative weights and estimators.

Once we have stated $\hat\psi$, it remains to set the marginal cdfs' $\hat U_k$, $k=1,\dots,d$, to be able to calculate our estimator $\hat\beta$. To fix the ideas, we rely on the standard univariate kernel-based conditional distributions, as given in (2.15): $\hat U_{i,k} := \hat F(X_{i,k}|\boldsymbol{Z}_i)$ and our estimator is then $\hat\beta^{(2)}$.

Concerning Assumption 2, the only thing to check is (2.1). This is guaranteed when the random matrix $\Sigma^{-1}(\boldsymbol{Z})$ is staying "under control", for instance when all eigenvalues of $\Sigma(\boldsymbol{Z})$ are uniformly bounded from below almost surely. It is sufficient to assume that

$$\sup_{\boldsymbol{z}\in\mathcal{Z}}\sup_{\beta\in\mathcal{B}}\lambda_{\min}(\psi(\beta,\beta'\boldsymbol{z})) \geq \underline{\lambda} > 0, \tag{2.19}$$

where $\lambda_{\min}(\Sigma)$ denotes the smallest eigenvalue of any nonnegative matrix Σ. In this case, it is easy to bound the log-density of $\boldsymbol{X}$ (conditional on $\boldsymbol{Z}$) from above, and to satisfy (2.1).

Assumption 3 is the most tricky one. In Section 4.1, some sufficient conditions are given to satisfy (2.2). It remains to check (2.3). We can apply our Lemma 2.3: under its conditions and if all the bandwidths we consider in $\hat{\boldsymbol{U}}_i$ behave as the same power of n, say $n^{-\pi}$ (the usual case), there exists a constant C s.t.

$$\sup_{i=1,\dots,n} |\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i|.\mathbf{1}(\boldsymbol{Z}_i \in \mathcal{Z}) \leq C\left(\sqrt{\ln(n)}n^{-(1-p\pi)/2} + n^{-\pi s}\right) := \delta_n \text{ a.s.}$$

Note that, for consistency purpose, we can choose any π s.t. $\pi < 1/p$. And ν_n can be chosen arbitrarily as long as we have $\nu_n >> \delta_n$, and then the condition (2.3) is satisfied.

Assumption 4 is satisfied for the Gaussian copula, as in most usual copula families. In our case and under (2.19), we choose $r(\boldsymbol{u}) \propto \prod_{k=1}^d \left(\Phi^{-1}(u_k)\right)^2$, and $\tilde{r}_k(\boldsymbol{u}) \propto \Phi^{-1}(u_k)\prod_{l=1,l\neq k}^d \left(\Phi^{-1}(u_l)\right)^2 / \left(\phi \circ \Phi^{-1}(u_k)\right)$.

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

Example 2: 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

$$C(\boldsymbol{u}|\boldsymbol{z}) = \left(\sum_{k=1}^d u_k^{-\theta(\boldsymbol{z})} - d + 1\right)^{-1/\theta(\boldsymbol{z})}, \ \boldsymbol{u} \in (0,1)^d,$$

with $\theta(\boldsymbol{z}) = \psi(\beta, \beta'\boldsymbol{z})$ 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. (2011), the Kendall tau of 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 to check that the latter mapping between τ and θ is one-to-one. The density of the Clayton copula with parameter $\theta > 0$ is given by

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

Assume that there exists $\underline{\theta}$ and $\overline{\theta}$ s.t., for every $\boldsymbol{z} \in \mathcal{Z}$ and every $\beta \in \mathcal{B}$, $\underline{\theta} \leq \psi(\beta, \beta'\boldsymbol{z}) \leq \overline{\theta}$. Then Assumption 2 is satisfied. Indeed, note that

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

Denoting V a r.v. that is uniform on $(0,1)$, we have

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

and (2.1) follows.

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

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

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

Example 3: the Gumbel copula

The d-dimensional Gumbel copula is given by

$$C_\theta(\boldsymbol{u}) := \exp\left(-\left[\sum_{k=1}^d (-\ln u_k)^\theta\right]^{-1/\theta}\right),$$

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

Its Kendall's tau in dimension d, as defined by (2.11) has been calculated in Genest et al. (2011):

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

where $\vec{m} := (m_1, \ldots, m_d)$, $m = m_1 + \ldots + m_d$, and the summation is taken over all d-uplets of integers s.t. $m_1 + 2m_2 + \ldots, +dm_d = d$. For every $\vec{m}$, $\mathcal{C}_{\vec{m}}$ denotes a positive constant. But note that

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

and

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

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

The Gumbel copula density is a linear combination of the functions

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

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

$$
c_{\theta(\boldsymbol{z})}(\boldsymbol{u}) \leq Cst.C(\boldsymbol{u}) \max_{\theta \in \{\underline{\theta}, \overline{\theta}\}} \left\{ \left[\sum_{k=1}^{d} (-\ln u_k)^\theta \right]^{j/\theta - d} \prod_{k=1}^{d} \frac{(-\ln u_k)^{\theta-1}}{u_k} \right\},
$$

for every $\boldsymbol{u} \in (0,1)^d$ and some constant Cst. By taking the logarithm of the previous r.h.s., it is easy to check that (2.1), and then Assumption 2, are satisfied.

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

$$
\sup_{\theta \in [\underline{\theta}, \overline{\theta}]} |\partial u_k c_\theta(\boldsymbol{u})| \leq Cst.h_k(\boldsymbol{u}) C_\theta(\boldsymbol{u}) / u_k^2 := \tilde{r}_k(\boldsymbol{u}),
$$

for some slowly varying functions h_k (deduced from the powers of the functions $u_l \mapsto \ln u_l$, $l = 1, \ldots, d$). The function $\tilde{r}_k$ belongs to $\mathcal{R}_d$ since $C_\theta(\boldsymbol{u})$ behaves as u_k when u_k tends to zero. Therefore $E[U_k(1-U_k)\tilde{r}_k(\boldsymbol{U})] < \infty$.

3 Asymptotic normality

3.1 Notations and assumptions

For convenience, we will denote $\psi_i = \psi(\beta_0, \beta_0' \boldsymbol{Z}_i)$ and $\hat{\psi}_i = \hat{\psi}(\beta_0, \beta_0' \boldsymbol{Z}_i)$.

Introduce the set of indicator functions

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

Since all the subsets we consider in $\mathcal{H}$ are boxes, it is simple to check that $\mathcal{H}$ is universally Donsker (for instance, see Example 2.6.1 and apply Lemma 2.6.17 in van der Vaart and Wellner (1996)).

Assumption 5 *For every $\boldsymbol{z} \in \mathcal{Z}$, assume that $\psi_{\boldsymbol{z}} : \mathcal{B} \to \Theta, \beta \mapsto \psi(\beta, \beta'\boldsymbol{z})$ is three times continuously differentiable. Moreover, set $\ln c : (0,1)^d \times \Theta \to \mathbb{R}, (\boldsymbol{u}, \theta) \mapsto \ln c_\theta(\boldsymbol{u})$. Assume that $\nabla_{\boldsymbol{u}} \nabla^2_\theta \ln c_\theta(\boldsymbol{u})$ exists on $(0,1)^d \times \Theta$.*

Assumption 6 *Let the functions on $(0,1)^d \times \mathcal{Z}$ defined by*

$$f(\boldsymbol{u}, \boldsymbol{z}) = \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta = \psi(\beta_0, \beta_0'\boldsymbol{z})}(\boldsymbol{u}), \text{ and } \hat{f}(\boldsymbol{u}, \boldsymbol{z}) = \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta = \hat{\psi}(\beta_0, \beta_0'\boldsymbol{z})}(\boldsymbol{u}).$$

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

Assumption 7 *Let the functions on $\mathcal{Z}$ defined by*

$$p : \boldsymbol{z} \to p(\boldsymbol{z}) = \nabla_\beta \psi(\beta, \beta'\boldsymbol{z})_{|\beta = \beta_0}, \quad \text{and } \hat{p} : \boldsymbol{z} \to \hat{p}(\boldsymbol{z}) = \nabla_\beta \hat{\psi}(\beta, \beta'\boldsymbol{z})_{|\beta = \beta_0}.$$

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

Assumption 8 *Assume that*

$$E\left[\sup_{\theta \in \Theta} |\nabla^j_\theta \ln c_\theta(\boldsymbol{U}_{\boldsymbol{Z}})| . \mathbf{1}(\boldsymbol{Z} \in \mathcal{Z})\right] < +\infty, \quad j = 2, 3.$$

Moreover, for every $(\boldsymbol{u}, \boldsymbol{u}') \in (0,1)^{2d}$, we have

$$|\nabla_\theta \ln c_\theta(\boldsymbol{u}) - \nabla_\theta \ln c_{\theta'}(\boldsymbol{u})| \leq \Phi(\boldsymbol{u}).|\theta - \theta'|, \tag{3.1}$$

$$\left|\nabla^2_\theta \ln c_\theta(\boldsymbol{u}) - \nabla^2_\theta \ln c_{\theta'}(\boldsymbol{u})\right| \leq \Phi(\boldsymbol{u}).|\theta - \theta'|, \tag{3.2}$$

for some function Φ s.t. $E[\Phi(\boldsymbol{U})] < \infty$.

Assumption 9 *Assume that, for every* $(\beta_1, \beta_2) \in \mathcal{B}^2$ *and* $j = 1, 2$,

$$\sup_{\boldsymbol{z} \in \mathcal{Z}} |\nabla^j_\beta \psi(\beta_1, \beta_1' \boldsymbol{z}) - \nabla^j_\beta \psi(\beta_2, \beta_2' \boldsymbol{z})| \leq C.|\beta_1 - \beta_2|,$$

where C *is a finite constant.*

Assumption 10 *Assume that*

$$\sup_{\beta \in \mathcal{B}, \boldsymbol{z} \in \mathcal{Z}} \left| \psi(\beta, \beta' \boldsymbol{z}) - \hat{\psi}(\beta, \beta' \boldsymbol{z}) \right| = o_P(1), \tag{3.3}$$

$$\sup_{\beta \in \mathcal{B}, \boldsymbol{z} \in \mathcal{Z}} \left| \nabla_\beta \psi(\beta, \beta' \boldsymbol{z}) - \nabla_\beta \hat{\psi}(\beta, \beta' \boldsymbol{z}) \right| = o_P(1), \tag{3.4}$$

$$\sup_{\beta \in \mathcal{B}, \boldsymbol{z} \in \mathcal{Z}} \left| \nabla^2_\beta \psi(\beta, \beta' \boldsymbol{z}) - \nabla^2_\beta \hat{\psi}(\beta, \beta' \boldsymbol{z}) \right| = o_P(1). \tag{3.5}$$

Assumption 11 *For every* $k = 1, \ldots, d$, *there exists a function* $\Gamma_k \in \mathcal{R}_d$ *s.t.*

$$\sup_{\theta \in \Theta} |\partial_{u_k} \nabla_\theta (\ln c_\theta)(\boldsymbol{u})| + \sup_{\theta \in \Theta} \left| \partial_{u_k} \nabla^2_\theta (\ln c_\theta)(\boldsymbol{u}) \right| \leq \Gamma_k(\boldsymbol{u}),$$

$$E\left[U_k^\alpha (1 - U_k)^\alpha \Gamma_k(\boldsymbol{U}_{\boldsymbol{Z}}).\mathbf{1}(\boldsymbol{Z} \in \mathcal{Z}) \right] < \infty,$$

for some $\alpha \in [0, 1[$.

Assumption 12 *Assume that*

$$\sup_{\boldsymbol{z} \in \mathcal{Z}} |\hat{\psi}(\beta_0, \beta_0' \boldsymbol{z}) - \psi(\beta_0, \beta_0' \boldsymbol{z})| = O_P(\eta_{1n}),$$

$$\sup_{\boldsymbol{z} \in \mathcal{Z}} |\hat{p}(\boldsymbol{z}) - p(\boldsymbol{z})| = O_P(\eta_{2n}),$$

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

Assumption 13 *Assume that* $\beta \mapsto M(\beta)$ *is twice continuously differentiable. Its Hessian matrix at point* β_0 *is denoted by* $\Sigma = \nabla^2_\beta M(\beta_0)$, *and is invertible.*

Assumption 14 *For any* $\boldsymbol{u} \in \mathbb{R}^d$, *set*

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

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

$$\sup_{k=1,\ldots,d} E[g(\boldsymbol{U}_i, \boldsymbol{Z}_i) \cdot \mathbf{1}(\boldsymbol{Z}_i \in \mathcal{Z}, |U_{i,k} - \nu_n| < \delta_n)] = o(n^{-1/2}), \tag{3.6}$$

and similarly after having replaced ν_n *by* $1 - \nu_n$.

The latter assumption is usually satisfied with a lot of usual copula models. Broadly speaking and when c_θ is continuous wrt its arguments and θ itself, it means that

$$\delta_n \int |\nabla_\theta c_\theta(\boldsymbol{u}_{-k}, \nu_n | \boldsymbol{z})_{|\theta=\theta_0(\boldsymbol{z})}|.\mathbf{1}(\boldsymbol{z} \in \mathcal{Z})\, d\boldsymbol{u}_{-k}\, d\mathbb{P}_{\boldsymbol{Z}}(\boldsymbol{z}) = o(n^{-1/2}),$$

and the same replacing ν_n by $1-\nu_n$. Obviously, we denote by $(\boldsymbol{u}_{-k}, \nu_n)$ the d-dimensional vector whose components are u_j, when $j \neq k$, and whose k-th component is ν_n.

3.2 Main results

Theorem 3.1 *Under Assumptions 1 to 14,*

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

Proof. By definition of $\hat\beta$, $\nabla_\beta M_n(\hat\beta) = 0$. Next, a first order Taylor expansion leads to

$$-\nabla_\beta M_n(\beta_0) = (\hat\beta - \beta_0)\nabla^2_\beta M_n(\tilde\beta),$$

where $\tilde\beta = \beta_0 + o_P(1)$, using the consistency of $\hat\beta$.

From Lemma A.3, we have $\nabla^2_\beta M_n(\tilde\beta) = \nabla^2_\beta M(\tilde\beta) + o_P(1)$. Moreover, from Assumption 13 and the consistency of $\hat\beta$ (hence the consistency of $\tilde\beta$), we get $\nabla^2_\beta M_n(\tilde\beta) = \Sigma + o_P(1)$.

Next, we have

$$\nabla_\beta M_n(\beta_0) \;=\; \frac{1}{n}\sum_{i=1}^n \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat\psi_i}(\hat{\boldsymbol{U}}_i)\nabla_\beta \hat\psi(\beta, \beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\hat\omega_{i,n}.$$

a. Switch from the trimming functions $\hat\omega_{i,n}$ to $\omega_{i,n}$.

Under Assumption 14, we can apply Lemma A.1 with the function

$$\chi(\boldsymbol{U}_i, \boldsymbol{Z}_i) := \sup_{\theta\in B_{i,\theta}} \sup_{\boldsymbol{v}\in B_{i,d}} \nabla_\theta \ln c_\theta(\boldsymbol{v}) \cdot \sup_{\beta\in\mathcal{B}} \sup_{\boldsymbol{z}\in\mathcal{Z}} |\nabla_\beta \psi(\beta, \beta'\boldsymbol{z})|, \;\text{ with}$$

$$B_{i,\theta} := B(\theta_0(\boldsymbol{Z}_i), \eta_{1n}),\;\; B_{i,d} := B(\boldsymbol{U}_i, \delta_n).$$

This implies

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

Now, decompose

$$\nabla_\beta M_n(\beta_0) = A_{1n} + A_{2n} + R_{1n} + R_{2n} + R_{3n},$$

where

$$
\begin{aligned}
A_{1n} &:= \frac{1}{n}\sum_{i=1}^n \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\boldsymbol{U}_i)\nabla_\beta \psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n},\\
A_{2n} &:= \frac{1}{n}\sum_{i=1}^n \left\{\frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\hat{\boldsymbol{U}}_i) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\boldsymbol{U}_i)\right\}\nabla_\beta \psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n},\\
R_{1n} &:= \frac{1}{n}\sum_{i=1}^n \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\hat{\boldsymbol{U}}_i)\left\{\nabla_\beta \hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0} - \nabla_\beta \psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\right\}\omega_{i,n},\\
R_{2n} &:= \frac{1}{n}\sum_{i=1}^n \left\{\frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat{\psi}_i}(\hat{\boldsymbol{U}}_i) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\hat{\boldsymbol{U}}_i)\right\}\nabla_\beta \psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n},\\
R_{3n} &:= \frac{1}{n}\sum_{i=1}^n \left\{\frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat{\psi}_i}(\hat{\boldsymbol{U}}_i) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\hat{\boldsymbol{U}}_i)\right\}\left\{\nabla_\beta \hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0} - \nabla_\beta \psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\right\}\omega_{i,n}.
\end{aligned}
$$

In this decomposition, we will show that only the first two terms (A_{1n} and A_{2n}) matter, and that the R_{jn}, $j=1,2,3$ are $o_P(n^{-1/2})$.

b. Study of R_{1n}.

First observe that

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

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

By a limited expansion, we have

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

for some $\boldsymbol{U}_i^*$ s.t. $|\boldsymbol{U}_i^* - \boldsymbol{U}_i| < |\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i|$. In addition, invoking Assumptions 11, note that

$$
|\hat{U}_{i,k} - U_{i,k}| \leq U_{i,k}^\alpha (1-U_{i,k})^\alpha |\hat{U}_{i,k} - U_{i,k}|^{1-\alpha},
$$

for n sufficiently large, uniformly wrt $i=1,\dots,n$ and $k=1,\dots,d$. Thanks to Assumption 12, we deduce

$$
R'_{1n} = O_P\left(\sup_i |\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i|^{1-\alpha}\cdot\|\hat{p}-p\|_\infty\right) = O_P(\delta_n^{1-\alpha}\eta_{2n}) = o_P(n^{-1/2}).
$$

Moreover, with obvious notations, R_{1n} can be rewritten as

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

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 van der Vaart and Wellner (1996). Moreover, from Assumption 12,

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

Therefore, the asymptotic equicontinuity of Donsker classes (see section 2.1.2 in van der Vaart and Wellner (1996) yields,

$$R_{1n} = \int \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u})\{\hat{p}(\boldsymbol{z}) - p(\boldsymbol{z})\}\omega_n(\boldsymbol{u},\boldsymbol{z})d\mathbb{P}_{(\boldsymbol{U},\boldsymbol{Z})}(\boldsymbol{u},\boldsymbol{z}) + o_P(n^{-1/2}).$$

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

$$\eta_{2n}\int |\nabla_\theta c_\theta(\boldsymbol{u})_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}|\cdot|\omega_n(\boldsymbol{u},\boldsymbol{z}) - \omega_\infty(\boldsymbol{u},\boldsymbol{z})|\,d\boldsymbol{u}\,d\mathbb{P}_{\boldsymbol{Z}}(\boldsymbol{z}) = o(n^{-1/2}).$$

This is guaranteed under our assumption 14.

Then, under our assumptions, we can apply Fubini's theorem. This provides

$$\int \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u})\{\hat{p}(\boldsymbol{z}) - p(\boldsymbol{z})\}\mathbf{1}(\boldsymbol{z}\in\mathcal{Z})d\mathbb{P}_{(\boldsymbol{U},\boldsymbol{Z})}(\boldsymbol{u},\boldsymbol{z})$$
$$= \int\{\hat{p}(\boldsymbol{z}) - p(\boldsymbol{z})\}d\mathbb{P}_{\boldsymbol{Z}}(\boldsymbol{z})\left(\int \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u})\,\mathbf{1}(\boldsymbol{z}\in\mathcal{Z})d\mathbb{P}_{(\boldsymbol{U}|\boldsymbol{Z}=\boldsymbol{z})}(\boldsymbol{u})\right) = 0,$$

by definition of $\psi(\beta_0,\beta_0'\boldsymbol{z})$, which maximizes $E[\ln c_\theta(\boldsymbol{U}_{\boldsymbol{z}})|\boldsymbol{Z}=\boldsymbol{z}]$ with respect to θ, for any $\boldsymbol{z}\in\mathcal{Z}$. 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,

$$R_{2n} = \frac{1}{n}\sum_{i=1}^n\left\{\frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat{\psi}_i}(\boldsymbol{U}_i) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\boldsymbol{U}_i)\right\}\nabla_\beta\psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n} + R_{2n}', \tag{3.7}$$

where

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

for some $\boldsymbol{U}_i^*$, ψ_i^* and $\tilde{\psi}_i$ s.t. $|\boldsymbol{U}_i^*-\boldsymbol{U}_i|<|\hat{\boldsymbol{U}}_i-\boldsymbol{U}_i|$, $|\psi_i^*-\psi_i|<|\hat{\psi}_i-\psi_i|$ and $|\tilde{\psi}_i-\psi_i|<|\hat{\psi}_i-\psi_i|$. Note that we have invoked Assumption 8 to bound the last term on the r.h.s. The main term on the r.h.s. is $O_P(\eta_{1n}\delta_n^{1-\alpha}) = o_P(n^{-1/2})$ from Assumptions, 12 and 11 We deduce $R'_{2n} = o_P(n^{-1/2})$.

Next, invoking assumptions 12 and 11, the first term on the right-hand side of (3.7) can be rewritten as

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

where $\sup_{\boldsymbol{u},\boldsymbol{z}}|h_n(\boldsymbol{u},\boldsymbol{z}) - h(\boldsymbol{u},\boldsymbol{z})| = o_P(1)$, and h_n and h both belong to $\mathcal{F}_4 = p.\mathcal{H}.\mathcal{F}_1$, as a consequence of Assumption 6. This is a Donsker class from Example 2.10.10 in van der Vaart and Wellner (1996). The asymptotic equicontinuity of the Donsker class $\mathcal{F}_4$ allows to write

$$\begin{aligned}
R_{2n} = \int&\left\{\frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat{\psi}(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u}) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u})\right\}\nabla_\beta\psi(\beta,\beta'\boldsymbol{z})_{|\beta=\beta_0} \\
&\cdot\ \omega_n(\boldsymbol{u},\boldsymbol{z})d\mathbb{P}_{(\boldsymbol{U},\boldsymbol{Z})}(\boldsymbol{u},\boldsymbol{z}) + o_P(n^{-1/2}).
\end{aligned}$$

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

$$\phi_n(\boldsymbol{z}) = \int\left\{\frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat{\psi}(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u}) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}(\boldsymbol{u})\right\}\omega_\nu(\boldsymbol{u})d\mathbb{P}_{(\boldsymbol{U}|\boldsymbol{Z}=\boldsymbol{z})}(\boldsymbol{u}),$$

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

$$R_{2n} = \int \phi_n(v)\left[\int \nabla_\beta \psi(\beta,\beta'\boldsymbol{z})_{|\beta=\beta_0}\omega_M(\boldsymbol{z})d\mathbb{P}_{(\boldsymbol{Z}|\beta_0'\boldsymbol{Z})}(\boldsymbol{z}|v)\right] d\mathbb{P}_{\beta_0'\boldsymbol{Z}}(v) + o_P(n^{-1/2}).$$

Next, as a consequence of Lemma A.5, use that

$$\int \nabla_\beta \psi(\beta,\beta'\boldsymbol{z})_{|\beta=\beta_0}\omega_M(\boldsymbol{z})d\mathbb{P}_{(\boldsymbol{Z}|\beta_0'\boldsymbol{Z}=v)}(\boldsymbol{z}) = 0.$$

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

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

$$R_{3n} = \frac{1}{n}\sum_{i=1}^n \left\{ \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\hat\psi_i}(\boldsymbol{U}_i) - \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\boldsymbol{U}_i)\right\}$$
$$\cdot \quad \left\{\nabla_\beta \hat\psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0} - \nabla_\beta \psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\right\}\omega_{i,n} + o_P(n^{-1/2}).$$

Due to Assumption 12 and Assumption 8 (see equation (3.1)), we obtain $R_{3n} = o_P(n^{-1/2})$. ■

Now, we need to introduce the way we estimate $\boldsymbol{U}_i$ by pseudo-observations $\hat{\boldsymbol{U}}_i$. Therefore, additional assumptions are required to achieve asymptotic normality.

Assumption 15 *For every $k = 1,\dots,d$, $x \in \mathbb{R}$ and $\boldsymbol{z} \in \mathcal{Z}$, we can write*

$$\hat F_k(x|\boldsymbol{z}) - F_k(x|\boldsymbol{z}) = \frac{1}{n}\sum_{j=1}^n a_{k,n}(\boldsymbol{X}_j,\boldsymbol{Z}_j,x,\boldsymbol{z}) + r_n(x,\boldsymbol{z}), \tag{3.8}$$

for some particular functions $a_{k,n}$ and for some sequence (r_n) s.t.

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

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

$$\hat U_{i,k} - U_{i,k} = \frac{1}{n}\sum_{j=1}^n a_{k,n}(\boldsymbol{X}_j,\boldsymbol{Z}_j,X_{i,k},\boldsymbol{Z}_i) + r_{n,i},\ n^{1/2}\sup_i |r_{n,i}| = O_P(1).$$

We will denote $\boldsymbol{a}_n(\boldsymbol{X}_j,\boldsymbol{Z}_j,\boldsymbol{X}_i,\boldsymbol{Z}_i)$ (or even shorter $\boldsymbol{a}_{i,j}$) the d-vector whose components are $a_{k,n}(\boldsymbol{X}_j,\boldsymbol{Z}_j,X_{i,k},\boldsymbol{Z}_i)$, $k = 1,\dots,d$.

In the case of the kernel-based estimates $\hat F_k$ of Lemma 2.3, Assumption 15 is satisfied by using s-order kernels K s.t. $\sup_k h_k = O(n^{-1/(2s)})$ and $n^{1/2}\prod_{k=1}^p h_k >> n^a$ for some $a > 0$. If $h_k = n^{-\pi}$ for all k, this necessitates $s \geq p$ and $\pi \in [1/(2s); 1/(2p)[$.

Assumption 16 *Assume that there exists a function W such that*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d,\boldsymbol{z}\in\mathcal{Z}} |E\left[a_n(\boldsymbol{X}_j,\boldsymbol{Z}_j,\boldsymbol{x},\boldsymbol{z})\right] - W(\boldsymbol{z},\boldsymbol{x})| = o(n^{-1/2}),$$

and such that

$$E\left[\left\|\Lambda_{\psi(\beta_0,\beta_0'\boldsymbol{Z}_i)}(\boldsymbol{U}_i).W(\boldsymbol{Z},\boldsymbol{X})\nabla_\beta\psi(\beta,\beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\right\|^2 .\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z})\right] < \infty,$$

with

$$\Lambda_{\psi(\beta_0,\beta_0'\boldsymbol{z})} := \nabla_{\boldsymbol{u}}\nabla_\theta(\ln c_\theta)_{|\theta=\psi(\beta_0,\beta_0'\boldsymbol{z})}.$$

In the case of the kernel-based estimates $\hat{F}_k$ of Lemma 2.3, we see that

$$E\left[a_n(\boldsymbol{X}_j,\boldsymbol{Z}_j,\boldsymbol{x},\boldsymbol{z})\right] = W(\boldsymbol{z},\boldsymbol{x}) = 0,$$

and Assumption 16 is automatically satisfied. This is most often the case with parametric marginal models too.

Assumption 17 *For every $k=1,\dots,d$, there exists a function $\zeta_k\in\mathcal{R}_d$ s.t.*

$$\sup_{\theta\in\Theta}\left|\partial^2_{u_k}\nabla_\theta(\ln c_\theta)(\boldsymbol{u})\right| \le \zeta_k(\boldsymbol{u}),$$

$$E\left[U_k^\gamma(1-U_k)^\gamma\zeta_k(\boldsymbol{U}_{\boldsymbol{Z}}).\mathbf{1}(\boldsymbol{Z}\in\mathcal{Z})\right] < \infty,$$

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

Assumption 18 *Assume that*

$$v_n^2 := E\left[\|\boldsymbol{a}_n(\boldsymbol{X}_j,\boldsymbol{Z}_j,\boldsymbol{X}_i,\boldsymbol{Z}_i) - E[\boldsymbol{a}_n(\boldsymbol{X}_j,\boldsymbol{Z}_j,\boldsymbol{X}_i,\boldsymbol{Z}_i)\,|\,\boldsymbol{X}_i,\boldsymbol{Z}_i]\|^2\right] < \infty.$$

and $v_n^2/n = o(1)$.

Corollary 3.2 *Under Assumptions 1 to 18, we have*

$$n^{1/2}\left\{\Sigma.(\hat\beta-\beta_0)+b_n\right\} \Longrightarrow \mathcal{N}(0,S),$$

where $S = E[\omega_1\mathcal{M}_1\mathcal{M}_1']$, where

$$\mathcal{M}_1 = \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_1}(\boldsymbol{U}_1)\nabla_\beta\psi(\beta,\beta'\boldsymbol{Z}_1)_{|\beta=\beta_0} + \Lambda_{\psi(\beta_0,\beta_0'\boldsymbol{Z}_1)}(\boldsymbol{U}_1).W(\boldsymbol{Z}_1,\boldsymbol{X}_1)\nabla_\beta\psi(\beta,\beta'\boldsymbol{Z}_1)_{|\beta=\beta_0},$$

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

Moreover, if

$$E\Big[\Lambda_{\psi(\beta_0,\beta_0'\boldsymbol{Z}_1)}(\boldsymbol{U}_1).W(\boldsymbol{Z}_1,\boldsymbol{X}_1)\nabla_\beta\psi(\beta,\beta'\boldsymbol{Z}_1)_{|\beta=\beta_0}$$
$$\cdot\quad\{\mathbf{1}(|U_{k,1}-\nu_n|<\delta_n)+\mathbf{1}(|1-U_{k,1}-\nu_n|<\delta_n)\}] = o(n^{-1/2}), \tag{3.9}$$

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

Note that the bias b_n cannot be removed in general, even if $E[\boldsymbol{a}_{i,j}] = 0$. Indeed, the trimming part $E[(\omega_{i,n} - \omega_i)\mathcal{M}_i]$ is of order δ_n typically, that has no reasons to be $o(n^{-1/2})$. To remove the asymptotic bias, we need (3.9). The latter condition is easily satisfied with purely parametric or nonparametric estimates, because $W(\boldsymbol{Z}, \boldsymbol{X})$ is zero or most often negligible in such cases.

Proof. We use the same notations as in the proof of Theorem 3.1. Recall that

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

which can be rewritten as

$$\begin{aligned} A_{2,n} &= \frac{1}{n}\sum_{i=1}^{n} \Lambda_{\psi_i}(\boldsymbol{U}_i).[\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i]\nabla_\beta \psi(\beta, \beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n} + O_P(\delta_n^{2-\gamma}) \\ &=: A'_{2,n} + o_P(n^{-1/2}), \end{aligned}$$

invoking Assumption 17. Next, under (3.8), we have

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

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

$$\begin{aligned} A'_{21} &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} \Lambda_{\psi_i}(\boldsymbol{U}_i).E\left[\boldsymbol{a}_{i,j}|\boldsymbol{Z}_i, \boldsymbol{X}_i\right]\nabla_\beta \psi(\beta, \beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n}, \text{ and} \\ A'_{22} &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1, i\neq j}^{n} \Lambda_{\psi_i}(\boldsymbol{U}_i).\left\{\boldsymbol{a}_{i,j} - E\left[\boldsymbol{a}_{i,j}|\boldsymbol{Z}_i, \boldsymbol{X}_i\right]\right\}.\nabla_\beta \psi(\beta, \beta'\boldsymbol{Z}_i)_{|\beta=\beta_0}\omega_{i,n} + o_P(n^{-1/2}). \end{aligned}$$

Due to Assumption 16,

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

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

$$E\left[\mathfrak{U}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{Z}_j, \boldsymbol{X}_j)|\boldsymbol{Z}_j, \boldsymbol{X}_j\right] = E\left[\mathfrak{U}(\boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{Z}_j, \boldsymbol{X}_j)|\boldsymbol{Z}_i, \boldsymbol{X}_i\right] = 0.$$

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

$$Var(A'_{22}) = O\left(\frac{v_n^2}{n^2}\cdot \int |\Lambda_{\psi(\beta_0, \beta_0'\boldsymbol{z})}(\boldsymbol{u})|^2.|\nabla_\beta \psi(\beta, \beta'\boldsymbol{z})_{|\beta=\beta_0}|^2 \omega_n(\boldsymbol{u}, \boldsymbol{z})\, d\mathbb{P}_{(\boldsymbol{U}, \boldsymbol{Z})}(\boldsymbol{u}, \boldsymbol{z})\right).$$

Under Assumptions 16 and 18, we get $A'_{22} = o_P(n^{-1/2})$.

We have obtained

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

by introducing a bias term $B_n := n^{-1}\sum_{i=1}^n \{\omega_{i,n} - \omega_i\}\mathcal{M}_i$, due to the trimming procedure. Its expectation is denoted by $b_n = E[(\omega_{1,n} - \omega_i)\mathcal{M}_1]$, and its variance is $O(n^{-1}\delta_n)$. The asymptotic bias is negligible under (3.9), by recalling assumption 14, and then applying Lemma A.1.

In every case, the result follows from a standard CLT, recalling the expansion of Theorem 3.1. ■

3.3 Examples cont'd

Let us check whether the conditions above apply to get the asymptotic normality of $\hat\beta$ in the case of the copula families in Subsection 2.4.

Example 1 cont'd: the Gaussian copula.

Obviously, Assumptions 5, 8 and 9 are satisfied. This is the case for Assumption 6 too, because $\Sigma \mapsto \ln(|\Sigma|)$ is Lipschitz under (2.19) and invoking Example 19.7 in van der Vaart (2007).

To deal with Assumption 7, note that p and $\hat p$ are Lipschitz transforms of conditional Kendall's tau $\tau(\beta, \beta'\boldsymbol{z})$ and $\hat\tau(\beta,\beta'\boldsymbol{z})$ respectively. Due to Example 19.20 in van der Vaart (2007), it is sufficient to show that the functions $\boldsymbol{z} \mapsto \nabla_\beta \tau(\beta_0, \beta_0'\boldsymbol{z})$ and $\boldsymbol{z} \mapsto \nabla_\beta \hat\tau(\beta_0,\beta_0'\boldsymbol{z})$ belong to a Donsker class a.e., assuming the underlying dimension d is two. It follows from Lemma A.4 and from the relation $\tau(\beta_0, \beta_0'\boldsymbol{z}) = 4\int C_{\beta_0}(\mathbf{u}|\beta_0'\boldsymbol{z})C_{\beta_0}(d\mathbf{u}|\beta_0'\boldsymbol{z}) - 1$ that

$$\nabla_\beta \tau(\beta_0, \beta_0'\boldsymbol{z}) = f_1(\beta_0'\boldsymbol{z}) + \boldsymbol{z} f_2(\beta_0'\boldsymbol{z}),\; \boldsymbol{z} \in \mathcal{Z},$$

with

$$f_1(v) = -E[\boldsymbol{Z}|\beta_0'\boldsymbol{Z}=v, \boldsymbol{Z}\in\mathcal{Z}]\left\{\int c_0(\mathbf{u},v)C_{\beta_0}(d\mathbf{u}|v) + \int C_{\beta_0}(\mathbf{u}|v)c_0(d\mathbf{u},v)\right\},$$

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

using the notations of Lemma A.4. In a Gaussian copula family, $\boldsymbol{z}\mapsto f_j(\beta_0'\boldsymbol{z})$ and $\boldsymbol{z}\mapsto f_j'(\beta_0'\boldsymbol{z})$, are uniformly bounded on $\mathcal{Z}$. Therefore, $\nabla_\beta\tau(\beta_0,\beta_0'\boldsymbol{z})$ belongs to the class $\mathcal{G}=\{\boldsymbol{z}\in\mathcal{Z}\rightarrow f(\beta_0'\boldsymbol{z})+\boldsymbol{z}g(\beta_0'\boldsymbol{z}), f,g\in\mathcal{C}(M)\}$, with $\mathcal{C}(M)=\{f:\|f\|_\infty+\|f'\|_\infty\leq M\}$. $\mathcal{C}(M)$ is a Donsker class from Theorem 2.7.1 in Van der Vaart and Wellner (1996). Moreover, $\mathcal{G}$ is Donsker from Examples 2.10.7 and 2.10.8 in Van der Vaart and Wellner (1996).

It is also the case for $\nabla_\beta\hat\tau$. Indeed, with the notations of Section 4, we can write

$$\hat\tau(\beta,\beta'\boldsymbol{z}) = \frac{4}{n^2\hat f_\beta^2(\beta'\boldsymbol{z})}\sum_{i,j=1}^n \mathbf{1}(\boldsymbol{X}_j\leq\boldsymbol{X}_i)\tilde K_{\tilde h}(\beta'\boldsymbol{Z}_j-\beta'\boldsymbol{z})\,\tilde K_{\tilde h}(\beta'\boldsymbol{Z}_i-\beta'\boldsymbol{z})-1.$$

A differentiation with respect to β easily shows that $\nabla_\beta\hat\tau(\beta_0,\beta_0'\boldsymbol{z})$ is of the form

$$\nabla_\beta\hat\tau(\beta_0,\beta_0'\boldsymbol{z}) = \hat f_1(\beta_0'\boldsymbol{z})+\boldsymbol{z}\hat f_2(\beta_0'\boldsymbol{z}).$$

The results of Section 4 allow to show that $\sup_{\boldsymbol{z}\in\mathcal{Z}}|\hat f_j(\beta_0'\boldsymbol{z})-f_j(\beta_0'\boldsymbol{z})| = O_P(\tilde h_n^2+[\log n]^{1/2}n^{-1/2}\tilde h_n^{-3/2})$, and that $\sup_{\boldsymbol{z}\in\mathcal{Z}}|\hat f_j'(\beta_0'\boldsymbol{z})-f_j'(\beta_0'\boldsymbol{z})| = O_P(\tilde h_n^2+[\log n]^{1/2}n^{-1/2}\tilde h_n^{-5/2})$, for $j=1,2$. Therefore, $\boldsymbol{z}\in\mathcal{Z}\mapsto\nabla_\beta\hat\tau(\beta_0,\beta_0'\boldsymbol{z})$ belongs to the Donsker class $\mathcal{G}$ when $n\tilde h_n^5\rightarrow 0$.

Assumption 10 is coming from the results of Section 4, and simple calculations prove that Assumption 11 is satisfied for every $\alpha>0$.

Recalling the notations of Section 4, we have

$$\sup_{\boldsymbol{z}\in\mathcal{Z}}|\hat\tau(\beta_0,\beta_0'\boldsymbol{z})-\tau(\beta_0,\beta_0'\boldsymbol{z})| = O_P(\tilde h^{\tilde s}+[\log n]^{1/2}n^{-1/2}\tilde h^{-1/2}) := O_P(\eta_{1n}),\text{ and}$$

$$\sup_{\boldsymbol{z}\in\mathcal{Z}}|\nabla_\beta\hat\tau(\beta,\beta_0'\boldsymbol{z})-\nabla_{\beta_0}\tau(\beta_0,\beta_0'\boldsymbol{z})| = O_P(\tilde h^{\tilde s}+[\log n]^{1/2}n^{-1/2}\tilde h^{-3/2}) := O_P(\eta_{2n}).$$

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 we have $4\tilde s\tilde\pi>1$, $\tilde s\geq 2$ and $4\tilde\pi<1$. Recall that we had set $\delta_n\sim n^{-\pi s}+\ln_2 n.n^{-(1-p\pi)/2}$. To satisfy $\delta_n^{1-\alpha}\eta_{jn}=o(n^{-1/2})$, $j=1,2$, 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 14, it can be checked that the l.h.s. of (3.6) is $O\left(\delta_n \nu_n [\Phi^{-1}(\nu_n)]^2\right)$. Nonetheless, $\Phi^{-1}(\nu_n) \sim -\sqrt{(-2)\ln \nu_n}$, when $\nu_n \to 0$ (see Dominici, 2003). A sufficient condition is then $\delta_n \nu_n |\ln(\nu_n)| = o(n^{-1/2})$.

Assumptions 15 and 16 are trivially satisfied because we have chosen nonparametric marginal cdfs' and we apply Lemma 2.3, for which we have seen that we set $W(\boldsymbol{z}, \boldsymbol{x}) = 0$.

Assumption 17 is the most demanding and cannot be obtained by the same reasoning as for Assumption 14. Actually, we recall that the former one has been requested only in the proof of Corollary 3.2 to show that

$$\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{u}} \nabla^2_\theta (\ln c_\theta)_{|\theta=\psi_i}(\boldsymbol{U}^*_i).[\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i]^2 \nabla_\beta \psi(\beta, \beta' \boldsymbol{Z}_i)_{|\beta=\beta_0} \omega_{i,n} = o_P(n^{-1/2}),$$

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

$$\delta_n^2 E\left[|\nabla_{\boldsymbol{u}} \nabla^2_\theta (\ln c_\theta)_{|\theta=\psi_i}(\boldsymbol{U}_i) \nabla_\beta \psi(\beta, \beta' \boldsymbol{Z}_i)_{|\beta=\beta_0}|\omega_{i,n}\right] = o(n^{-1/2}).$$

Due to the bounded-ness of c_θ, the latter expectation is less than a constant times

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

The latter integral behaves as $\exp\left([\Phi^{-1}(\nu_n)]^2/2\right)$. Since $\Phi^{-1}(\nu_n) \sim -\sqrt{(-2)\ln \nu_n}$, it is sufficient to satisfy

$$\delta_n^2/\nu_n = o(n^{-1/2}).$$

Usual variance calculations for kernel densities prove that Assumption 18 is true when $nh^p = n^{1-p\pi} \to \infty$, i.e. when $p\pi < 1$.

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

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

implying $\delta_n \sim n^{-2/5}$ and we choose $\nu = n^{-1/5}$. 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.

4 Conditional Kendall's Tau

In this section, we show how to check Assumptions 10 and 12 in general, when the conditional margins are estimated non-parametrically. 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 $\boldsymbol{X}$, whose conditional copula will be parameterized by $\tau(\beta, \beta'\boldsymbol{z})$, the conditional Kendall's tau coefficient of this vector as defined in (2.10) when $d = 2$, and (2.11) more generally. In other words, we consider the case where $\psi(\beta, \beta'\boldsymbol{z}) = \Phi(\tau(\beta, \beta'\boldsymbol{z}))$ for some "sufficiently regular" function Φ. Indeed, Kendall's tau are commonly used for inference purpose of parametric copulae, particularly Archimedean and elliptical copulae. Moreover, as explained in Subsection 2.2, (A1) and (A2) are satisfied in such cases. Finally, we do not suffer from the curse of dimension because conditional Kendall's tau are those associated to the copula of $\boldsymbol{X}$ knowing $\beta'\boldsymbol{Z}$.

Introducing a kernel estimator $\hat{F}_\beta$ of $F_\beta(\boldsymbol{x}|y) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}|\beta'\boldsymbol{Z} = y)$ as $\hat{F}_\beta(\boldsymbol{x}|y) = \hat{H}_\beta(\boldsymbol{x}, \infty|y)$ (recall (2.12)), define

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

In Lemma 4.1 below, 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 4.1 *Assume that*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d, \beta\in\mathcal{B}, \boldsymbol{z}\in\mathcal{Z}} |\hat{F}_\beta(\boldsymbol{x}|\beta'\boldsymbol{z}) - F_\beta(\boldsymbol{x}|\beta'\boldsymbol{z})| = O_P(\varepsilon_{n,0}). \tag{4.1}$$

Then,

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

Proof. Decompose

$$\begin{aligned}(2^d - 1)\left\{\hat{\tau}(\beta, \beta'\boldsymbol{z}) - \tau(\beta, \beta'\boldsymbol{z})\right\} &= 2^d \int \{\hat{F}_\beta(\boldsymbol{x}|\beta'\boldsymbol{z}) - F_\beta(\boldsymbol{x}|\beta'\boldsymbol{z})\}\hat{F}_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z}) \\ &+ 2^d \int F_\beta(\boldsymbol{x}|\beta'\boldsymbol{z})\{\hat{F}_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z}) - F_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z})\}.\end{aligned}$$

The first term is $O_P(\varepsilon_n)$ due to (4.1). For the second, observe that

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

which is less than $\sup_{\boldsymbol{x},\beta,\boldsymbol{z}} |\hat{F}_\beta(\boldsymbol{x}|\beta'\boldsymbol{z}) - F_\beta(\boldsymbol{x}|\beta'\boldsymbol{z})|$, and we use again (4.1). ■

Lemma 4.1 provides some tools to check the first part of Assumptions 10 and 12, if one assumes that the function Φ is regular enough (that is Hölder with some high enough Hölder exponent). Similarly, we can derive the uniform consistency of $\nabla^j_\beta \hat{\tau}$ for $j = 1, 2$, which allows to check the remaining conditions in Assumptions 10 and 12.

Lemma 4.2 *Assume that*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d,\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}} |\nabla^j_\beta \hat{F}_\beta(\boldsymbol{x}|\beta'\boldsymbol{z}) - \nabla^j_\beta F_\beta(\boldsymbol{x}|\beta'\boldsymbol{z})| \;=\; O_P(\varepsilon_{n,j}), \tag{4.2}$$

for $j = 1, 2$, and that

$$\sup_{j=1,2} \int \left|\nabla^j_\beta F_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z})\right| + \left|\nabla^j_\beta \hat{F}_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z})\right| \leq M,$$

for some $M > 0$. Then,

$$\sup_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}} |\nabla_\beta \hat{\tau}(\beta, \beta'\boldsymbol{z}) - \nabla_\beta \tau(\beta, \beta'\boldsymbol{z})| = O_P(\max(\varepsilon_{n,1}, \varepsilon_{n,0})), \text{ and}$$

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

Proof. This is a consequence of applying the $\nabla-$operator to $\hat{\tau}(\beta, \beta'\boldsymbol{z})$, and of the compactness of $\mathcal{Z}$. ■

The next step is to check that, under reasonable conditions, (4.1) and (4.2) hold. To this aim, let us introduce some assumptions.

Assumption 19 *Let $\tilde{K}$ denote a univariate symmetric kernel function of order $\tilde{s}$, $s \geq 2$. 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 $n\tilde{h}_n \to \infty$.*

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 $\hat{\boldsymbol{U}}_i$ (see Lemma 2.3).

Assumption 20 *Let $f_\beta(y)$ denote the density of $\beta'\boldsymbol{Z}$ evaluated at point y. Assume that $\inf_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}} f_\beta(y) > c$, for some $c > 0$. Moreover, assume that $f_\beta(u)$ is s-times continuously differentiable, with uniformly bounded derivatives.*

The latter assumption is satisfied most of the time, because $\beta'\boldsymbol{z}$ belongs to a compact subset when $\beta \in \mathcal{B}$ and $\boldsymbol{z} \in \mathcal{Z}$ [4] In the single-index literature, some authors relaxed this assumption, by only assuming $\inf_{\boldsymbol{z}} f_{\beta_0}(u) > c$. Nevertheless, Assumption 20 requires to introduce a trimming procedure, in order to avoid parts of the distribution for which some $f_\beta(\beta'\boldsymbol{Z}_i)$ 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, Patilea, Van Keilegom (2013).

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 21 *$\mathcal{A}$ 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 $m_\phi(y) = E[\phi(\boldsymbol{X}, \boldsymbol{Z})|\beta'\boldsymbol{Z} = y]$. Assume that the functions m_ϕ are twice continuously differentiable, and their derivatives up to order 2 are bounded by some finite constant M that does not depend on ϕ.*

We first state Lemma 4.3 that provides consistency rates for kernel weighted sums.

Lemma 4.3 *Let $\mathcal{L}$ denote a class of functions satisfying Assumption 21. Under Assumption 19, we have*

$$\begin{aligned}
&\frac{1}{n\tilde{h}} \sup_{\lambda\in\mathcal{L}} \sup_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}} \left| \sum_{i=1}^n \lambda(\boldsymbol{X}_i, \boldsymbol{Z}_i)\tilde{K}\left(\frac{\beta'\boldsymbol{Z}_i - \beta'\boldsymbol{z}}{\tilde{h}}\right) - E\left[\lambda(\boldsymbol{X}_i, \boldsymbol{Z}_i)\tilde{K}\left(\frac{\beta'\boldsymbol{Z}_i - \beta'\boldsymbol{z}}{\tilde{h}}\right)\right]\right| \\
&\quad = \; O_P([\log n]^{1/2} n^{-1/2}\tilde{h}^{-1/2}).
\end{aligned}$$

Proof. Let

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

and

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

where $(\varepsilon_i)_{1\leq i\leq n}$ are i.i.d. Rademacher variables. Due to Proposition A.6, we have

$$\mathbb{P}\left(\mathcal{B} \geq A_1(\mathcal{B}_\varepsilon + t)\right) \leq 2\left\{\exp(-A_2't^2/(n\tilde{h})) + \exp(-A_2 t)\right\}, \tag{4.3}$$

[4]For instance, assume the arguments y above belong to a fixed interval $[a,b]$ and that $\boldsymbol{Z}$ follows a Gaussian $\mathcal{N}(0,\Sigma)$. Then $\beta'\boldsymbol{Z} \sim \mathcal{N}(0, \beta'\Sigma\beta)$ and $f_\beta(y) = \exp(-y^2/2(\beta'\Sigma\beta)/[\sqrt{2\pi}\beta'\Sigma\beta]$. Since $\beta'\Sigma\beta$ belongs to a compact $[c,d]$, $c > 0$, the latter density is larger than $\exp(-b^2/(2d^2))/[\sqrt{2\pi}d > 0$.

where A_2' is a constant. Indeed, since the function λ are uniformly and bounded by one,

$$\sup_{\beta,\boldsymbol{z},\lambda} Var\left(\lambda(\boldsymbol{X},\boldsymbol{Z})\tilde{K}\left(\frac{\beta'\boldsymbol{Z}-\beta'\boldsymbol{z}}{\tilde{h}}\right)\right) = O(\tilde{h}).$$

Next, observe that the class of functions

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

satisfies the assumptions of Proposition A.7 with $\sigma^2 = O(h)$ and

$$N(\varepsilon, \mathcal{L}_{\tilde{K}}) \leq C\varepsilon^{-\nu}, \tag{4.4}$$

for some C and ν. The property (4.4) can be obtained from the following: Lemma 22 in Nolan and Pollard (1987) shows that $N(\varepsilon, \mathcal{K}) \leq C_2\varepsilon^{-\nu_2}$, where

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

Using Assumption 21 and Lemma A.1 in Einmahl and Mason (2000), we get that $\mathcal{L}_{\tilde{K}} = \mathcal{L} \cdot \mathcal{K}$ satisfies (4.4).

Therefore, we can apply Proposition A.7 to deduce that

$$\mathcal{B}_\varepsilon \leq A'n^{1/2}\tilde{h}^{1/2}[\log(\tilde{h}^{-1})]^{1/2} = A''n^{1/2}\tilde{h}^{1/2}[\log n]^{1/2}. \tag{4.5}$$

It follows from (4.5) that, for $t_1 > 2A_1A''$,

$$\mathbb{P}(\mathcal{B} \geq t_1n^{1/2}\tilde{h}^{1/2}[\log n]^{1/2}) \leq \mathbb{P}\left(\mathcal{B} \geq A_1\mathcal{B}_\varepsilon + t_1n^{1/2}\tilde{h}^{1/2}[\log n]^{1/2}/2\right).$$

Applying (4.3) with $t = t_1n^{1/2}\tilde{h}^{1/2}[\log n]^{1/2}/(2A_1)$, we get

$$\mathbb{P}(\mathcal{B} \geq t_1n^{1/2}\tilde{h}^{1/2}[\log n]^{1/2}) \leq 2\left\{\exp(-A_2't_1^2[\log n]/(4A_1^2)) + \exp(-A_2t_1n^{1/2}\tilde{h}^{1/2}[\log n]^{1/2}/(2A_1))\right\},$$

and the result follows. ■

This Lemma is the cornerstone of Lemma 4.4 belows, which ensures consistency rates for $\hat{F}_\beta$ and its derivatives.

Lemma 4.4 *Let $\mathcal{A}$ denote a class of functions satisfying Assumption 21. Then, under Assumptions 19 and 20,*

$$\begin{aligned}&\sup_{\phi\in\mathcal{A}}\sup_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}}\left|\int\phi(\boldsymbol{x},\boldsymbol{z})\{\hat{F}_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z}) - F_\beta(d\boldsymbol{x}|\beta'\boldsymbol{z})\}\right| \\ &= O_P\left(\tilde{h}^{\tilde{s}} + [\log n]^{1/2}n^{-1/2}\tilde{h}^{-1/2}\right).\end{aligned}$$

Proof. Write

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

where

$$\hat{f}_\beta(\beta'\boldsymbol{z}) = \frac{1}{n\tilde{h}}\sum_{i=1}^n \tilde{K}\left(\frac{\beta'\boldsymbol{Z}_i-\beta'\boldsymbol{z}}{\tilde{h}}\right), \tag{4.6}$$

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

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

and $\mathfrak{m}_\phi(\beta'\boldsymbol{z}) = m_\phi(\beta'\boldsymbol{z})f_\beta(\beta'\boldsymbol{z})$. It follows from Lemma 4.3 that

$$\sup_{\beta,\boldsymbol{z},\phi}|\hat{\mathfrak{m}}_\phi(\beta'\boldsymbol{z}) - E[\hat{\mathfrak{m}}_\phi(\beta'\boldsymbol{z})]| + \sup_{\beta,\boldsymbol{z}}|\hat{f}_\beta(\beta'\boldsymbol{z}) - E[\hat{f}_\beta(\beta'\boldsymbol{z})]| = O_P(\frac{[\log n]^{1/2}}{n^{1/2}\tilde{h}^{1/2}}).$$

Moreover, using classical arguments on kernel estimators (and Assumptions 21 and 19), we have

$$\sup_{\beta,\boldsymbol{z},\phi}|E[\hat{\mathfrak{m}}_\phi(\beta'\boldsymbol{z})] - \mathfrak{m}_\phi(\beta'\boldsymbol{z})| + \sup_{\beta,\boldsymbol{z}}|E[\hat{f}_\beta(\beta'\boldsymbol{z})] - f_\beta(\beta'\boldsymbol{z})| = O(\tilde{h}^{\tilde{s}}).$$

The result of the Lemma follows from the fact that the density $f_\beta(\beta'\boldsymbol{z})$ is bounded away from zero by Assumption 20. ■

Lemma 4.4 allows to check condition (4.1) by considering $\phi(\boldsymbol{x},\boldsymbol{z}) = 1$, showing that, in this case, $\varepsilon_{n,0} = \tilde{h}^{\tilde{s}} + [\log n]^{1/2}n^{-1/2}\tilde{h}^{-1/2}$. It also permits to obtain the uniform consistency rates for $\nabla^j_\beta \hat{F}_\beta$ for $j = 1, 2$, with

$$\varepsilon_{n,1} = \tilde{h}^{\tilde{s}} + \frac{[\log n]^{1/2}}{n\tilde{h}^{3/2}},\ \varepsilon_{n,2} = \tilde{h}^{\tilde{s}} + \frac{[\log n]^{1/2}}{n\tilde{h}^{5/2}}.$$

Indeed,

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

and the convergence of this term can be studied using Lemma 4.4, but replacing $\tilde{K}$ by $\tilde{K}'$, and setting $\phi(\boldsymbol{X},\boldsymbol{Z}) = \mathbf{1}(\boldsymbol{X} \le \boldsymbol{x}).(\boldsymbol{Z} - \boldsymbol{z})$. The latter function is indexed by $(\boldsymbol{x},\boldsymbol{z})$ that lives into $\mathbb{R}^d \times \mathcal{Z}$, defining the convenient class $\mathcal{A}$ to apply Lemma 4.4. The other terms obtained by differentiation can be studied in the same way.

Hence, the latter results allow to check whether Assumptions 10 and 12 hold. Indeed, under some (light) conditions of regularity, we have obtained that

$$\sup_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}}|\hat{\tau}(\beta,\beta'\boldsymbol{z}) - \tau(\beta,\beta'\boldsymbol{z})| = O_P(\tilde{h}^{\tilde{s}} + [\log n]^{1/2}n^{-1/2}\tilde{h}^{-1/2}),$$

$$\sup_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}} |\nabla_\beta\hat{\tau}(\beta,\beta'\boldsymbol{z}) - \nabla_\beta\tau(\beta,\beta'\boldsymbol{z})| = O_P(\tilde{h}^{\tilde{s}} + [\log n]^{1/2}n^{-1/2}\tilde{h}^{-3/2}), \text{ and}$$

$$\sup_{\beta\in\mathcal{B},\boldsymbol{z}\in\mathcal{Z}} |\nabla^2_\beta\hat{\tau}(\beta,\beta'\boldsymbol{z}) - \nabla^2_\beta\tau(\beta,\beta'\boldsymbol{z})| = O_P(\tilde{h}^{\tilde{s}} + [\log n]^{1/2}n^{-1/2}\tilde{h}^{-5/2}).$$

References

[1] K. Aas, C. Czado, A. Frigessi and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44**, 182-198.

[2] E.F. Acar, R.V. Craiu and Yao, F. (2011). Dependence Calibration in Conditional Copulas: A Nonparametric Approach. *Biometrics* **67**, 445-453.

[3] E.F. Acar, R.V. Craiu and Yao, F. (2013). Statistical testing of covariate effects in conditional copula models. *Electr. J. Statistics* **7**, 2822-2850.

[4] Delecroix, M., Hristache, M. (1999). M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique. *Bull. Belg. Math. Soc.* **6**, 161-185.

[5] Dominici, D. (2003). The inverse of the cumulative standard normal probability function. *Integral Transforms & Special Functions* **14**, 281-292.

[6] Du, Y., Akritas, M.G. (2002). I.i.d. representations of the conditional Kaplan-Meier process for arbitrary distributions. *Math. Method. Statist.* **11**, 152-182.

[7] Einmahl, U., Mason, D. (2000). An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators. *Journal of Theoretical Probability* **13**, 1-37;

[8] Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380-1403.

[9] J-D. Fermanian and Wegkamp, M. (2012). Time-dependent copulae, *J. Multivariate Anal.* **110**, 19-29.

[10] Genest, C., Nešlehová, J. and Ben Ghorbal, N. (2011). Estimators based on Kendall's tau in multivariate copula models. *Aust. N.Z. J. Stat.* **53**, 157-177.

[11] Gijbels, I., Veraverbeke, N. and Omelka, M. (2011). Conditional copulae, association measures and their applications. *Computational Stat. Data Anal.* **55**, 1919-1932.

[12] Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157-178.

[13] Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995

[14] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single- index models. *J. Econometrics* **58**, 71-120.

[15] Joe, H. (1990). Multivariate concordance. *J. Multivariate Anal.* **35**, 12-30.

[16] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Volume **73**. London: Chapman & Hall.

[17] Jondeau, E. and Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. *J. of Internat. Money and Finance* **25**, 827-853.

[18] Kendall, M.G. and Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika* **31**, 324-345.

[19] Klein, R. L. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61**, 387-421.

[20] Lopez, O., Patilea, V. and van Keilegom, I. (2013). Single-index regression models in the presence of censoring depending on the covariates. *Bernoulli* **19**, 721-747.

[21] Nelsen, R. (1998). *An introduction to copulas.* Lecture Notes in Statistics **139**. Springer.

[22] Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing, in *Handbook of Econometrics*, vol. IV, 2111-2245. Elsevier.

[23] Nolan, D. and Pollard, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15**, 780-799.

[24] Patton, A.J. (2006). Modelling asymmetric exchange rate dependence. *Int. Econ. Rev.* **47**, 527-556.

[25] Patton, A.J. (2009). Copula-based models for financial time series. In: *Handbook of Financial Time Series*, Springer, Berlin, 767-785.

[26] Patton, A.J. (2012). A review of copula models for economic time series. *J. Multivariate Anal.*, **110**, 4-18.

[27] Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57**, 1403-1430.

[28] Rodriguez, J.C. (2007). Measuring financial contagion: A copula approach. *J. of Empirical Finance*, **14(3)** 401-423.

[29] Schmid, F. and Schmidt, R. (2007). Multivariate extensions of Spearman rho and related statistics, *Statistics & Probability Letters*, **77**(4), 407-416.

[30] Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica* **54**, 1461-1481

[31] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian J. Statist.* **33**, 357-375.

[32] van der Vaart, A. (2007). *Asymptotic Statistics.* Cambrige UP.

[33] van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes.* Springer.

A Technical lemmas

Lemma A.1 *Assume that there exists a deterministic sequence* (δ_n), $\delta_n = o(\nu_n)$, *s.t.*

$$\sup_i |\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i|.\mathbf{1}(\boldsymbol{Z}_i \in \mathcal{Z}) \leq \delta_n \ \ a.e. \tag{A.1}$$

Consider an integrable function χ *on* $(0,1)^d \times \mathcal{Z}$. *Assume that there exists a sequence* (ξ_n), $\xi_n \to 0$, *s.t.*

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

for all $k = 1, \ldots, d$. *Then*

$$\frac{1}{n}\sum_{i=1}^n |\chi(\boldsymbol{U}_i, \boldsymbol{Z}_i).(\hat{\omega}_{i,n} - \omega_{i,n})| = O_P(\xi_n).$$

Proof. Note that

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n |\chi(\boldsymbol{U}_i,\boldsymbol{Z}_i)|\cdot|\hat\omega_{i,n}-\omega_{i,n}|>\varepsilon\right)\\
\leq\ & \sum_{k=1}^d \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n |\chi(\boldsymbol{U}_i,\boldsymbol{Z}_i)|\cdot\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z},|U_{i,k}-\nu_n|\leq|\hat U_{i,k}-U_{i,k}|)>\varepsilon/(2d)\right)\\
+\ & \sum_{k=1}^d \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n |\chi(\boldsymbol{U}_i,\boldsymbol{Z}_i)|\cdot\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z},|1-\nu_n-U_{i,k}|\leq|\hat U_{i,k}-U_{i,k}|)>\varepsilon/(2d)\right)\\
\leq\ & \frac{2d}{\varepsilon}\sum_{k=1}^d E\left[|\chi(\boldsymbol{U}_i,\boldsymbol{Z}_i)|\cdot\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z}).\{\mathbf{1}(|U_{i,k}-\nu_n|\leq\delta_n)+\mathbf{1}(|1-\nu_n-U_{i,k}|\leq\delta_n)\}\right]\\
\leq\ & 2d\xi_n/\varepsilon,
\end{aligned}$$

proving the result. ■

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

$$\begin{aligned}
\xi_n :=\ & \sup_k E[\sup_{u_k,|u_k-\nu_n|\leq\delta_n}|\chi(u_k,\boldsymbol{U}_{i,-k},\boldsymbol{Z}_i)|\cdot\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z})]\\
+\ & \sup_k E[\sup_{u_k,|u_k-1+\nu_n|\leq\delta_n}|\chi(u_k,\boldsymbol{U}_{i,-k},\boldsymbol{Z}_i)|\cdot\mathbf{1}(\boldsymbol{Z}_i\in\mathcal{Z})],
\end{aligned}$$

or even, when it tends to zero,

$$\begin{aligned}
\xi_n :=\ & \sup_k \sup_{u_k,|u_k-\nu_n|\leq\delta_n}\ \sup_{\boldsymbol{u}_{-k}\in[\nu_n-\delta_n,1-\nu_n+\delta_n]^{d-1}}\ \sup_{\boldsymbol{z}\in\mathcal{Z}}|\chi(u_k,\boldsymbol{u}_{-k},\boldsymbol{z})|\\
+\ & \sup_k \sup_{u_k,|u_k-1+\nu_n|\leq\delta_n}\ \sup_{\boldsymbol{u}_{-k}\in[\nu_n-\delta_n,1-\nu_n+\delta_n]^{d-1}}\ \sup_{\boldsymbol{z}\in\mathcal{Z}}|\chi(u_k,\boldsymbol{u}_{-k},\boldsymbol{z})|.
\end{aligned}$$

Lemma A.3 *Under the assumptions of Theorem 3.1,*

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

Proof. We have

$$\begin{aligned}
\nabla^2_\beta M_n(\beta) &= \frac{1}{n}\sum_{i=1}^n \nabla_\theta(\ln c_\theta)_{|\theta=\hat\psi_i}(\hat{\boldsymbol{U}}_i)\nabla^2_\beta\hat\psi(\beta,\beta'\boldsymbol{Z}_i)\,\hat\omega_{i,n}\\
&\quad+\frac{1}{n}\sum_{i=1}^n \nabla^2_\theta(\ln c_\theta)_{|\theta=\hat\psi_i}(\hat{\boldsymbol{U}}_i)\nabla_\beta\hat\psi_i\nabla_{\beta'}\hat\psi_i\,\hat\omega_{i,n}\\
&=: B_{1,n}(\beta)+B_{2,n}(\beta).
\end{aligned}$$

$$B_{1,n}(\beta) - \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla_\theta c_\theta}{c_\theta}_{|\theta=\psi_i}(\boldsymbol{U}_i)\nabla^2_\beta \hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)\hat{\omega}_{i,n} = \frac{1}{n}\sum_{i=1}^{n}\Big[\nabla_\theta(\ln c_\theta)_{|\theta=\hat{\psi}_i}(\hat{\boldsymbol{U}}_i)$$

$$- \quad \nabla_\theta(\ln c_\theta)_{|\theta=\hat{\psi}_i}(\boldsymbol{U}_i) + \nabla_\theta(\ln c_\theta)_{|\theta=\hat{\psi}_i}(\boldsymbol{U}_i) - \nabla_\theta(\ln c_\theta)_{|\theta=\psi_i}(\boldsymbol{U}_i)\Big]\,\nabla^2_\beta\hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)\,\hat{\omega}_{i,n}$$

$$= \quad \frac{1}{n}\sum_{i=1}^{n}\Big[\nabla^2_{\boldsymbol{u},\theta}(\ln c_\theta)_{|\theta=\hat{\psi}_i}(\boldsymbol{U}^*_i).(\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i) + \nabla^2_{\theta,\theta}(\ln c_\theta)_{|\theta=\psi^*_i}(\boldsymbol{U}_i).(\hat{\psi}_i - \psi)\Big]\,\nabla^2_\beta\hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)\,\hat{\omega}_{i,n},$$

for some $\boldsymbol{U}^*_i$ and ψ^*_i s.t.

$$|\boldsymbol{U}^*_i - \boldsymbol{U}_i| < |\hat{\boldsymbol{U}}_i - \boldsymbol{U}_i|, \quad |\psi^*_i - \psi_i| < |\hat{\psi} - \psi_i|.$$

From Assumption 11, we get that

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

From Assumptions 8 and the uniform consistency of $\hat{\psi}(\beta,\beta'\boldsymbol{z})$ (see (3.5)), we have

$$\sup_\beta |\frac{1}{n}\sum_{i=1}^{n}\nabla^2_{\theta,\theta}(\ln c_\theta)_{|\theta=\psi^*_i}(\boldsymbol{U}_i).(\hat{\psi}_i - \psi)\nabla^2_\beta\hat{\psi}(\beta,\beta'\boldsymbol{Z}_i)\,\hat{\omega}_{i,n}| = o_P(1),$$

and we deduce

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

Invoking Assumption 10, equation (3.5), we get

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

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

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

Similarly, one can deduce from Assumptions 11 and 10 that $\sup_\beta |B_{2,n}(\beta) - B_2(\beta)| = o_P(1)$, with

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

From Assumption 9 and (3.1) and (3.2) in Assumption 8, we can apply Example 19.7 and Theorem 19.4 in van der Vaart (2007) to deduce that

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

■

Lemma A.4 *Let $c_0(\boldsymbol{u}, v)$ denote the first order partial derivative of $C^M_{\beta_0}(\boldsymbol{u}|w)$ with respect to w evaluated at point $w = v$, where $C^M_\beta(\boldsymbol{u}|w)$ denotes the conditional copula function of $\boldsymbol{U}$ conditionally to $\beta'\boldsymbol{Z}$ and $\|\boldsymbol{Z}\|_\infty \leq M$ (that is $\boldsymbol{Z} \in \mathcal{Z}$). We have*

$$\nabla_\beta C_\beta(\boldsymbol{u}|\beta'\boldsymbol{Z})_{|\beta=\beta_0} = c_0(\boldsymbol{u}, \beta_0'\boldsymbol{Z})\left(\boldsymbol{Z} - E\left[\boldsymbol{Z}|\beta_0'\boldsymbol{Z}, \boldsymbol{Z} \in \mathcal{Z}\right]\right).$$

Proof. The proof is similar to the proof of Lemma 5A in Dominitz and Sherman (2005), and of Lemma 3.4 in Lopez, Patilea and Van Keilegom (2013). Observe that

$$\begin{aligned} C^M_\beta(\boldsymbol{u}|\beta'\boldsymbol{Z}) &= E\left[\mathbf{1}_{\boldsymbol{U}\leq\boldsymbol{u}}|\beta'\boldsymbol{Z}, \boldsymbol{Z} \in \mathcal{Z}\right] \\ &= E\left[E\left[\mathbf{1}_{\boldsymbol{U}\leq\boldsymbol{u}}|\boldsymbol{Z}\right]|\beta'\boldsymbol{Z}, \boldsymbol{Z} \in \mathcal{Z}\right] \\ &= E\left[C^M_{\beta_0}(\boldsymbol{u}|\beta_0'\boldsymbol{Z})|\beta'\boldsymbol{Z}, \boldsymbol{Z} \in \mathcal{Z}\right], \end{aligned}$$

where we used the single-index assumption for going from line 2 to line 3. Next, let

$$\Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta_1, \beta_2) = E\left[C^M_{\beta_0}(\boldsymbol{u}|\alpha(\boldsymbol{Z}, \beta_1) + \beta_2'\boldsymbol{Z})|\beta_2'\boldsymbol{Z}, \boldsymbol{Z} \in \mathcal{Z}\right],$$

where $\alpha(\boldsymbol{Z}, \beta_1) = \beta_0'\boldsymbol{Z} - \beta_1'\boldsymbol{Z}$. Hence, $C^M_\beta(\boldsymbol{u}|\beta'\boldsymbol{Z}) = \Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta, \beta)$. As a consequence,

$$\nabla_\beta C^M_\beta(\boldsymbol{u}|\beta'\boldsymbol{Z})_{|\beta=\beta_0} = \nabla_1\Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta, \beta_0)_{|\beta=\beta_0} + \nabla_2\Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta_0, \beta)_{|\beta=\beta_0},$$

where ∇_j represents the gradient vector with respect to β_j. Observe that

$$\nabla_1\Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta, \beta_0)_{|\beta=\beta_0} = -E\left[\boldsymbol{Z}c_0(\boldsymbol{u}, \beta_0'\boldsymbol{Z})|\beta_0'\boldsymbol{Z}\right].$$

Moreover, $\Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta_0, \beta) = C^M_{\beta_0}(\boldsymbol{u}|\beta'\boldsymbol{Z})$, which leads to

$$\nabla_2\Gamma_{\boldsymbol{u},\boldsymbol{Z}}(\beta_0, \beta)_{|\beta=\beta_0} = \boldsymbol{Z}c_0(\boldsymbol{u}, \beta_0'\boldsymbol{Z}),$$

and the result follows. ■

Lemma A.5 *Assume that the transformation Ψ is Hadamard differentiable. Then, for all v,*

$$\int \nabla_\beta\psi(\beta, \beta'\boldsymbol{z})_{\beta=\beta_0} d\mathbb{P}_{(\boldsymbol{Z}|\beta_0'\boldsymbol{Z})}(\boldsymbol{z}|v) = 0.$$

Proof. Let $\dot{\Psi}(C(\cdot))[D(\cdot)]$ denote the Hadamard derivative of Ψ at point C, applied to function D. Recall that

$$\psi(\beta, \beta'\boldsymbol{z}) = \Psi(C_\beta^M(\cdot|\beta'\boldsymbol{z})).$$

Hence, using Lemma A.4,

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

This shows that

$$\nabla_\beta \psi(\beta, \beta'\boldsymbol{z})_{|\beta=\beta_0} = [\boldsymbol{z} - E[\boldsymbol{Z}|\beta_0'\boldsymbol{Z} = \beta_0'\boldsymbol{z}]]\, \Lambda(\beta_0'\boldsymbol{z}),$$

and the result of Lemma A.5 follows. ■

Finally, Lemma 4.3 invokes two propositions from Einmahl and Mason (2005), that we recall here.

Proposition A.6 *Let $\mathcal{G}$ denote a class of functions bounded by 1, and let $\sigma_{\mathcal{G}}^2 = \sup_{g\in\mathcal{G}} Var(g(\boldsymbol{X}, \boldsymbol{Z}))$. Then, for all $t > 0$,*

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

for some universal constants A_1 and A_2, and

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

where $(\varepsilon_i)_{1\leq i\leq n}$ are i.i.d. Rademacher variables independent from $(\boldsymbol{X}_i, \boldsymbol{Z}_i)_{1\leq i\leq n}$.

Proposition A.7 *Assume that $\mathcal{G}$ is a class of functions satisfying the assumptions of Proposition A.6 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 $\sup_{g\in\mathcal{G}} E[g(\boldsymbol{X}, \boldsymbol{Z})^2] \leq \sigma^2$. Then,*

$$G_\varepsilon \leq A n^{1/2} \sigma \log(1/\sigma).$$