

MAXIMUM PSEUDO-LIKELIHOOD ESTIMATION BASED ON ESTIMATED RESIDUALS IN COPULA SEMIPARAMETRIC MODELS

MAREK OMELKA¹, ŠÁRKA HUDECOVÁ¹, NATALIE NEUMEYER²

¹ Department of Probability and Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic

² Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

March 12, 2019

ABSTRACT. This paper deals with a situation when one is interested in the dependence structure of a multidimensional response variable in the presence of a multivariate covariate. It is assumed that the covariate affects only the marginal distributions through regression models while the dependence structure, which is described by a copula, is unaffected. A parametric estimation of the copula function is considered with focus on the maximum pseudo-likelihood method. It is proved that under some appropriate regularity assumptions the estimator calculated from the residuals is asymptotically equivalent to the estimator based on the unobserved errors. In such case one can ignore the fact that the response is first adjusted for the effect of the covariate. A Monte Carlo simulation study explores (among others) situations where the regularity assumptions are not satisfied and the claimed result does not hold. It shows that in such situations the maximum pseudo-likelihood estimator may behave poorly and the moment estimation of the copula parameter is of interest. Our results complement the results available for nonparametric estimation of the copula function.

Keywords and phrases: asymptotic normality, copula, moment estimation, pseudo-likelihood, residuals.

1. INTRODUCTION

Consider a d -dimensional vector $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$ of responses and an associated q -dimensional vector of the covariates $\mathbf{X} = (X_1, \dots, X_q)^\top$. For instance in insurance applications one can consider that the response represents various type of payments related to a given car accident (medical benefits, income replacement benefits, and allocated expenses for a claimant) and the covariates present some additional information (claimants age, gravity of accident, number of people injured in the accident, ...).

Often we are interested in the conditional distribution of \mathbf{Y} given the value of the covariate. To simplify the situation it is often assumed that \mathbf{X} affects only the marginal distributions of Y_j ($j = 1, \dots, d$), but does not affect the dependence structure of \mathbf{Y} . More formally, it is assumed that there exists a copula C such that the joint conditional distribution of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$ can be for all $\mathbf{x} \in S_{\mathbf{X}}$ (the support of \mathbf{X}) written as

$$H_{\mathbf{x}}(y_1, \dots, y_d) = P(Y_1 \leq y_1, \dots, Y_d \leq y_d \mid \mathbf{X} = \mathbf{x}) = C(F_{1\mathbf{x}}(y_1), \dots, F_{d\mathbf{x}}(y_d))$$

where $F_{j\mathbf{x}}(y_j) = P(Y_j \leq y_j \mid \mathbf{X} = \mathbf{x})$, $j = 1, \dots, d$. Using this assumption one can proceed in two steps. In the first step one models the effect of the covariate on each of the marginal distributions separately (i.e. estimating $F_{j\mathbf{x}}$ for each $j \in \{1, \dots, d\}$ separately). Having $\hat{F}_{j\mathbf{x}}$ one estimates the copula function C in the second step.

Nonparametric estimation of the copula function C (for $d = 2$ and $q = 1$) was in detail considered in Gijbels et al. (2015). The most interesting result is as follows. Suppose that the marginal distributions follow the parametric or even non-parametric location scale models, i.e.

$$(1) \quad Y_j = m_j(X) + s_j(X)\varepsilon_j, \quad \text{where } \varepsilon_j \text{ is independent with } X.$$

Note that then C is the copula function corresponding to the random vector $(\varepsilon_1, \varepsilon_2)^T$. Then Gijbels et al. (2015) proved that (under some regularity assumptions) the empirical copula \hat{C}_n based on the estimated residuals from model (1) is asymptotically equivalent to the empirical copula \tilde{C}_n calculated from the unobserved errors ε_{ji} . More precisely it was proved that

$$(2) \quad \sup_{(u_1, u_2) \in [0, 1]^2} \sqrt{n} |\hat{C}_n(u_1, u_2) - \tilde{C}_n(u_1, u_2)| = o_P(1).$$

This result was generalized to time-series setting by Neumeyer et al. (2019). In Portier and Segers (2018) the authors were even able to drop the location-scale assumption (1) but at the cost of deriving only a slightly weaker result (the supremum in (2) is replaced with $\sup_{[\gamma, 1-\gamma]^2}$ where γ can be taken arbitrarily small but positive). On the other hand Côté et al. (2019) concentrated on the parametric form of the location scale model (1) and generalized the results to $d > 2$, $q > 1$ and at the same time relaxed assumptions on $f_{j\varepsilon}$ (the density of ε_{ji}).

To complement the results on nonparametric estimation of C one is naturally interested if analogous results hold also for parametric estimation of C . More precisely suppose that the copula function C belongs to the family $\mathcal{C} = \{C(\cdot; \mathbf{a}) : \mathbf{a} \in \Theta\}$ and we are interested in estimating the unknown parameter. Denote $\boldsymbol{\alpha}$ the true value of the parameter, $\hat{\boldsymbol{\alpha}}_n$

the estimator based on the residuals $(\widehat{\varepsilon}_{ji})$ and $\widetilde{\alpha}_n$ its counterpart based on the true (but unobserved) errors (ε_{ji}) from the location-scale model (1). Then in analogy to (2) one would expect that $\widehat{\alpha}_n$ is (the first-order) asymptotically equivalent to $\widetilde{\alpha}_n$, i.e.

$$(3) \quad \sqrt{n}(\widehat{\alpha}_n - \widetilde{\alpha}_n) = o_P(1).$$

Although the conjecture (3) seems to be natural, to the best of our knowledge there are only limited results specifying the regularity assumptions that are needed so that (3) holds. Some results for the moment-like estimators that can be deduced from the convergence of the empirical copula \widehat{C}_n can be found in Neumeyer et al. (2019) and Côté et al. (2019).

In this paper (similarly as in Côté et al., 2019) we assume the parametric form of the location-scale model (1) and concentrate on **maximum pseudo-likelihood estimation**. This method of estimation was in the context of copula models popularised by Genest et al. (1995) and in more detail investigated in Tsukahara (2005). This method is often preferred to moment-like estimation because the resulting estimator has usually a lower asymptotic variance.

In the econometric (time-series) literature the inference based on the residuals is also known as univariate (marginal) filtering (see e.g., Bücher et al., 2015) and the result (3) is supported by many simulation studies. The result is formulated already in Chen and Fan (2006a) but there it is presented more on an intuitive level and the precise assumptions (as well as reasoning) are missing. This lack of of rigorousness were to some extent redeemed in the subsequent paper Chan et al. (2009) where the authors concentrated on the multivariate GARCH-models and presented a lot of interesting ideas how to deal with the technical difficulties. But a careful reading of the paper reveals that (probably due to the broad scope of the presented results) some of the crucial steps in the proofs are missing.

In our paper we will explore in detail the assumptions that are needed so that (3) holds in the standard i.i.d. setting. Even in this relatively simply setting one has to handle many technical difficulties. The thing is that it is not clear how to make use to of the recent deep results in empirical copula estimation (see e.g., Berghaus et al., 2017; Radulović et al., 2017) as the densities of many standard copulas are unbounded. The only remarkable exception in this aspect is Theorem 3.3 of (Berghaus et al., 2017), but the authors considered only two dimensional copulas and no covariates.

We show that although the assumptions that guarantess (3) are mild, they are not satisfied for some combinations of commonly used copula functions and marginal densities. Roughly speaking we illustrate that an unbounded copula density has to be compensated

with marginal densities that are well behaved not only in the supports of the corresponding distributions, but also at the border points of the supports. We are convinced that exploring this problem in this settings is not only of independence interest, but it provides also insights to understand what might go wrong when switching to more complicated econometric or time-series models (see also the discussion in Section 4).

The paper is organised as follows. The main result and the needed assumptions are formulated in Section 2. The theoretical results are illustrated in a simulation study in Section 3. All the proofs are given in the Appendices.

2. MAIN RESULT

In what follows we assume that for each $j \in \{1, \dots, d\}$ there exists a **known** transformation T_j increasing on the support of Y_j and **known** functions $m_j(\mathbf{x}; \boldsymbol{\theta}_j)$ and $s_j(\mathbf{x}; \boldsymbol{\theta}_j)$ depending only on an unknown (finite-dimensional) parameter $\boldsymbol{\theta}_j$ such that the random variable

$$\varepsilon_j = \frac{T_j(Y_j) - m_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)},$$

is independent of \mathbf{X} with cumulative distribution function $F_{j\varepsilon}$. The distribution of the random vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)^\top$ has continuous margins and the copula corresponding to $\boldsymbol{\varepsilon}$ belongs to the families of copulas $\mathcal{C} = \{C(\cdot; \mathbf{a}) : \mathbf{a} \in \Theta\}$ and $\Theta \subset \mathbb{R}^p$.

Our task is to estimate the true value of the copula parameter (say $\boldsymbol{\alpha}$) based on the observations $(\mathbf{Y}_1), \dots, (\mathbf{Y}_n)$ that are assumed to be mutually independent copies of the vector (\mathbf{Y}) .

Let $\mathbf{Y}_i = (Y_{1i}, \dots, Y_{di})^\top$. As the parameters $\boldsymbol{\theta}_j$ ($j \in \{1, \dots, d\}$) are in practice unknown, we work with the residuals

$$\widehat{\varepsilon}_{ji} = \frac{T_j(Y_{ji}) - m_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j)}{s_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j)}, \quad i = 1, \dots, n; \quad j = 1, \dots, d,$$

where $\widehat{\boldsymbol{\theta}}_j$ is a suitable estimate of $\boldsymbol{\theta}_j$. For $j \in \{1, \dots, d\}$ let $\widehat{F}_{j\widehat{\varepsilon}}$ be the marginal empirical distribution function of the estimated residuals, i.e.

$$\widehat{F}_{j\widehat{\varepsilon}}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{\widehat{\varepsilon}_{ji} \leq y\}.$$

Then the **maximum pseudo-likelihood estimator** based on the residuals is defined as

$$\widehat{\boldsymbol{\alpha}}_n = \arg \max_{\mathbf{a} \in \Theta} \sum_{i=1}^n \log \{c(\widehat{\mathbf{U}}_i; \mathbf{a})\},$$

where

$$(4) \quad \widehat{\mathbf{U}}_i = (\widehat{U}_{1i}, \dots, \widehat{U}_{di})^\top = \frac{n}{n+1} (\widehat{F}_{1\widehat{\varepsilon}}(\widehat{\varepsilon}_{1i}), \dots, \widehat{F}_{d\widehat{\varepsilon}}(\widehat{\varepsilon}_{di}))^\top$$

are the estimated pseudo-observations and $c(\mathbf{u}; \mathbf{a})$ is the density of the assumed copula family. As it is common in the maximum likelihood theory we will consider the estimator $\widehat{\boldsymbol{\alpha}}_n$ to be an appropriately chosen root of the estimating equations

$$(5) \quad \sum_{i=1}^n \boldsymbol{\psi}(\widehat{\mathbf{U}}_i; \widehat{\boldsymbol{\alpha}}_n) = \mathbf{0}_p, \quad \text{where} \quad \boldsymbol{\psi}(\mathbf{u}; \mathbf{a}) = \frac{\partial \log\{c(\mathbf{u}; \mathbf{a})\}}{\partial \mathbf{a}}.$$

Analogously let $\widetilde{\boldsymbol{\alpha}}_n$ be the corresponding estimator based on the true (but unobserved) errors ε_{ji} . I.e. $\widetilde{\boldsymbol{\alpha}}_n$ is defined as (an appropriately chosen) root of the estimating equations

$$(6) \quad \sum_{i=1}^n \boldsymbol{\psi}(\widetilde{\mathbf{U}}_i; \widetilde{\boldsymbol{\alpha}}_n) = \mathbf{0}_p,$$

where

$$(7) \quad \widetilde{\mathbf{U}}_i = (\widetilde{U}_{1i}, \dots, \widetilde{U}_{di})^\top = \frac{n}{n+1} (\widehat{F}_{1\varepsilon}(\varepsilon_{1i}), \dots, \widehat{F}_{d\varepsilon}(\varepsilon_{di}))^\top$$

and $\widehat{F}_{j\varepsilon}$ is the marginal empirical distribution function of the (unobserved) errors, i.e.

$$\widehat{F}_{j\varepsilon}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{\varepsilon_{ji} \leq y\}, \quad j = 1, \dots, d.$$

2.1. Regularity assumptions on the marginal distributions. In general we need to assume that the density of the error term ε_j should be ‘well-behaved’ on the border of its support. The following assumption is close to assumption F(iii) in Appendix A of Einmahl and Van Keilegom (2008). But our assumption is weaker as it allows for distributions with supports different from a real line.

Assumption (F_{jε}): For each $j \in \{1, \dots, d\}$ the density function $f_{j\varepsilon}$ of ε_j is continuous on the support of ε_j and there exists $\beta \in [0, \frac{1}{2})$ such that

$$(8) \quad \sup_{u \in (0,1)} \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) (1 + |F_{j\varepsilon}^{-1}(u)|)}{u^\beta (1-u)^\beta} < \infty$$

and

$$\sup_{u \in (0,1/2)} \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(2u))}{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))} < \infty \quad \text{and} \quad \sup_{u \in (1/2,1)} \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(1-2u))}{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(1-u))} < \infty.$$

Further for some u_1, u_2 in $(0, 1)$ the function $f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))$ is non-decreasing on $(0, u_1)$ and non-increasing on $(u_2, 1)$.

Note that assumption $(\mathbf{F}_{j\varepsilon})$ with $\beta = 0$ allows also for distributions with non-continuous but bounded densities (e.g. exponential and uniform). But as we show later, for copula families with unbounded densities one needs to assume that $\beta > 0$.

Remark 1. The assumption $(\mathbf{F}_{j\varepsilon})$ is formulated so that it covers the general case when both the conditional mean as well as the conditional variance of $T_j(Y_{ji})$ depends on \mathbf{X}_i . From the proofs given in the appendix it follows that if one rightly assumes that the conditional variance does not depend on \mathbf{X}_i , then one does only location adjustment (i.e. $\widehat{\varepsilon}_{ji} = T_j(Y_{ji}) - m_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j)$) and assumption (8) simplifies to

$$\sup_{u \in (0,1)} \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))}{u^\beta(1-u)^\beta} < \infty.$$

On the other hand if one rightly assumes that the conditional mean is zero then one does only scale adjustment (i.e. $\widehat{\varepsilon}_{ji} = \frac{T_j(Y_{ji})}{s_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j)}$) and it is sufficient to assume

$$\sup_{u \in (0,1)} \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) |F_{j\varepsilon}^{-1}(u)|}{u^\beta(1-u)^\beta} < \infty.$$

This last assumption is close to the assumption 2. formulated just before Theorem 2.1 of Chan et al. (2009). But similarly as when comparing with assumption F(iii) in Appendix A of Einmahl and Van Keilegom (2008), our assumption does not require that the support of the distribution is a real line.

Remark 2. As in assumption $(\mathbf{F}_{j\varepsilon})$ the function $f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))$ is supposed to be monotone when u is close to zero or close to one, then the integrability of $f_{j\varepsilon}$ (see Lemma 12) implies that

$$\lim_{|x| \rightarrow \infty} |x| f_{j\varepsilon}(x) = 0.$$

Thus if

$$(9) \quad \lim_{u \rightarrow 0_+} F_{j\varepsilon}^{-1}(u) = -\infty \quad \left(\lim_{u \rightarrow 1_-} F_{j\varepsilon}^{-1}(u) = \infty \right),$$

then one gets

$$(10) \quad \lim_{u \rightarrow 0_+} f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) (1 + |F_{j\varepsilon}^{-1}(u)|) = 0 \quad \left(\lim_{u \rightarrow 1_-} f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) (1 + |F_{j\varepsilon}^{-1}(u)|) = 0 \right).$$

Note that the above equations are also automatically satisfied if $\beta > 0$ even if (9) does not hold. Thus one can conclude that if (10) does not hold, then $\beta = 0$ and the corresponding border of the support is finite, i.e.,

$$\lim_{u \rightarrow 0_+} F_{j\varepsilon}^{-1}(u) > -\infty \quad \left(\lim_{u \rightarrow 1_-} F_{j\varepsilon}^{-1}(u) < \infty \right).$$

2.2. Regularity assumptions on m_j and s_j . The next assumption states that the parametric models can be estimated at the standard \sqrt{n} -rate and that the location and scale functions are sufficiently smooth and integrable.

Assumption (ms): For each $j \in \{1, \dots, d\}$ $\hat{\boldsymbol{\theta}}_j$ is a \sqrt{n} -consistent estimate of the parameter $\boldsymbol{\theta}_j \in \mathbb{R}^{p_j}$. The functions $m_j(\mathbf{x}; \mathbf{t})$ and $s_j(\mathbf{x}; \mathbf{t})$ are (once) differentiable with respect to \mathbf{t} and the derivatives are denoted as $m'_j(\mathbf{x}; \mathbf{t})$ and $s'_j(\mathbf{x}; \mathbf{t})$. Further there exists a neighborhood $U(\boldsymbol{\theta}_j)$ of the true value of the parameter $\boldsymbol{\theta}_j$ such that $\inf_{\mathbf{x} \in S_{\mathbf{X}}, \mathbf{t} \in U(\boldsymbol{\theta}_j)} s_j(\mathbf{x}; \mathbf{t}) > 0$ and there exists a function $M_j : S_{\mathbf{X}} \rightarrow \mathbb{R}$ such that for each $\mathbf{x} \in S_{\mathbf{X}}$:

$$\sup_{\mathbf{t} \in U(\boldsymbol{\theta}_j)} \left\| \frac{m'_j(\mathbf{x}; \mathbf{t})}{s_j(\mathbf{x}; \mathbf{t})} \right\| \leq M_j(\mathbf{x}), \quad \sup_{\mathbf{t} \in U(\boldsymbol{\theta}_j)} \left\| \frac{s'_j(\mathbf{x}; \mathbf{t})}{s_j(\mathbf{x}; \mathbf{t})} \right\| \leq M_j(\mathbf{x}),$$

and $\mathbb{E}[M_j(\mathbf{X})]^r < \infty$ for some $r \geq 2$. Finally, for each $K > 0$ the derivatives $m'_j(\mathbf{x}; \mathbf{t})$ and $s'_j(\mathbf{x}; \mathbf{t})$ viewed as functions of \mathbf{t} are continuous at $\boldsymbol{\theta}_j$ uniformly in $\mathbf{x} \in \{\tilde{\mathbf{x}} \in S_{\mathbf{X}} : \|\tilde{\mathbf{x}}\| \leq K\}$.

2.3. Regularity assumptions about the copula family \mathcal{C} . To formulate the main regularity assumptions about the copula family it is useful to introduce the following set of functions.

Definition (Class of \mathcal{J} - and $\tilde{\mathcal{J}}^{\beta_1, \beta_2}$ -functions). A function $\varphi : (0, 1)^d \rightarrow \mathbb{R}$ is called a \mathcal{J} -function if φ is continuous on $(0, 1)^d$ and there exist $\eta \in [0, 1)$ and a finite constant M_1 such that for all $\mathbf{u} \in (0, 1)^d$

$$|\varphi(u_1, \dots, u_d)| \leq \sum_{j=1}^d \frac{M_1}{[\min\{u_j, 1 - u_j\}]^\eta}.$$

Let $\beta_1 \in [0, 1/2)$ and $\beta_2 \geq 0$ be fixed. We say that a function $\varphi : (0, 1)^d \rightarrow \mathbb{R}$ is a $\tilde{\mathcal{J}}^{\beta_1, \beta_2}$ -function if it is continuous on $(0, 1)^d$ and there exists a finite constant M_2 such that for all $\mathbf{u} \in (0, 1)^d$

$$|\varphi(u_1, \dots, u_d)| \leq \sum_{j=1}^d \frac{M_2}{[\min\{u_j, 1 - u_j\}]^{\beta_1}}.$$

Further $|\varphi^{(j)}(u_1, \dots, u_d)| u_j^{\beta_2} (1 - u_j)^{\beta_2}$ is a \mathcal{J} -function for all $j \in \{1, \dots, d\}$, where

$$\varphi^{(j)}(u_1, \dots, u_d) = \frac{\partial \varphi(u_1, \dots, u_d)}{\partial u_j}.$$

Now we are ready to formulate the needed regularity assumptions about the copula family. Recall that $\Theta \subset \mathbb{R}^p$, $\boldsymbol{\alpha}$ is the true value of the parameter, and $c(\mathbf{u}; \mathbf{a})$ is a density corresponding to the copula function $C(\mathbf{u}; \mathbf{a})$.

Assumptions C:

C1. $c(\mathbf{u}; \mathbf{a}_1) = c(\mathbf{u}; \mathbf{a}_2)$ for almost all $\mathbf{u} \in (0, 1)^d$ only if $\mathbf{a}_1 = \mathbf{a}_2$.

C2. The function $\log\{c(\mathbf{u}; \mathbf{a})\}$ is continuously differentiable with respect to \mathbf{a} for all $\mathbf{u} \in (0, 1)^d$.

Denote the k th element of the vector function $\boldsymbol{\psi}(\mathbf{u}; \mathbf{a}) = \partial \log\{c(\mathbf{u}; \mathbf{a})\} / \partial \mathbf{a}$ by $\psi_k(\mathbf{u}; \mathbf{a})$.

C3. For each $k \in \{1, \dots, p\}$, the function $\psi_k(\cdot; \boldsymbol{\alpha}) \in \tilde{\mathcal{J}}^{\beta_1, \beta_2}$, where $\beta > \max\{\beta_1 + \frac{1}{r-1}, \beta_2\}$, for β introduced in assumption $(\mathbf{F}_{j\varepsilon})$ and r in assumption (\mathbf{ms}) .

C4. The function $\boldsymbol{\psi}(\mathbf{u}; \mathbf{a})$ is assumed to be continuously differentiable with respect to \mathbf{a} for all $\mathbf{u} \in (0, 1)^d$. Further there exist an open neighborhood $\mathcal{U} \subset \Theta$ of $\boldsymbol{\alpha}$ and a dominating function $h(\mathbf{u}) \in \mathcal{J}$ such that $\partial \boldsymbol{\psi}(\mathbf{u}; \mathbf{a}) / \partial \mathbf{a}^\top$ is continuous in $(0, 1)^d \times \mathcal{U}$ and

$$\max_{k, \ell \in \{1, \dots, p\}} \sup_{\mathbf{a} \in \mathcal{U}} \left| \frac{\partial \psi_k(\mathbf{u}; \mathbf{a})}{\partial a_\ell} \right| \leq h(\mathbf{u}).$$

C5. The $p \times p$ (Fisher information) matrix $I(\boldsymbol{\alpha}) = -\mathbf{E} \left\{ \partial \boldsymbol{\psi}(\mathbf{U}; \mathbf{a}) / \partial \mathbf{a}^\top \Big|_{\mathbf{a}=\boldsymbol{\alpha}} \right\}$, where

$$\mathbf{U} = (U_1, \dots, U_d)^\top = (F_{1\varepsilon}(\varepsilon_1), \dots, F_{d\varepsilon}(\varepsilon_d))^\top,$$

is finite and nonsingular.

Remark 3. Note that the score functions of the commonly used one-parameter bivariate copula families with unbounded densities (e.g. Clayton, Gumbel, Normal, Student, ...) can be bounded by

$$|\psi(u_1, u_2; a)| \leq M_3 \sum_{j=1}^2 |\log(u_j) + \log(1 - u_j)|$$

and its derivative as

$$|\psi^{(j)}(u_1, u_2; a)| \leq \frac{M_3}{[\min\{u_j, 1 - u_j\}]} + M_3 \sum_{j'=1}^2 |\log(u_{j'}) + \log(1 - u_{j'})|, \quad j = 1, 2$$

for a sufficiently large but finite constant M_3 (see also Chen and Fan, 2006b). Thus in Assumption **C3** one can consider β_1 and β_2 arbitrarily close to zero but positive.

Assumption **C3** is inspired by Chan et al. (2009). Note that generally speaking this assumption is more strict than the corresponding assumptions of Tsukahara (2005) that are based on U -shaped functions. The advantage of assumption **C3** is that it enables to derive bounds that depend only on the marginal distributions. The price that we pay for

this advantage does not seem to be big because we are not aware of a standard copula family that does not meet **C3** with β_1 and β_2 arbitrarily small positive constants.

Note that assumption **C3** implies that $\beta > 0$, which does not allow for marginal densities $f_{j\varepsilon}$ that are bounded but possibly discontinuous at a border point (e.g. exponential or uniform distributions). As shown in simulations in Section 3 the aimed result (3) indeed does not hold in general when the marginal densities $f_{j\varepsilon}$ are not continuous.

Nevertheless a closer inspection of the proof shows that $\beta > 0$ is needed to get a control over a possibly unbounded score function $\psi(\mathbf{u}; \mathbf{a})$. But there are commonly used copula families (e.g. Frank, Ali-Mikhail-Haq, Plackett) for which the score function $\psi(\mathbf{u}; \mathbf{a})$ and its derivatives are bounded. It is of interest to formulate an alternative to assumptions **C3** and **C4** separately as it allows for $\beta = 0$ in assumption $(\mathbf{F}_{j\varepsilon})$,

C6. The function $\psi(\mathbf{u}; \mathbf{a})$ is bounded and continuously differentiable with respect to \mathbf{a} for all $\mathbf{u} \in (0, 1)^d$. Further there exists an open neighborhood \mathcal{U} of $\boldsymbol{\alpha}$ such that $\partial\psi(\mathbf{u}; \mathbf{a})/\partial\mathbf{a}^\top$ is continuous in $(0, 1)^d \times \mathcal{U}$ and

$$\max_{k, \ell \in \{1, \dots, p\}} \sup_{\mathbf{a} \in \mathcal{U}} \sup_{\mathbf{u} \in (0, 1)^d} \left| \frac{\partial \psi_k(\mathbf{u}; \mathbf{a})}{\partial a_\ell} \right| < \infty \quad \text{and} \quad \max_{j \in \{1, \dots, d\}} \max_{k \in \{1, \dots, p\}} \sup_{\mathbf{u} \in (0, 1)^d} \left| \frac{\partial \psi_k(\mathbf{u}; \boldsymbol{\alpha})}{\partial u_j} \right| < \infty.$$

2.4. Main results. Now we are ready to formulate the main results of the paper.

Theorem 1. *Suppose that assumptions (\mathbf{ms}) , **C1-C5** and $(\mathbf{F}_{j\varepsilon})$ with $\beta > 0$ are satisfied. Then with probability going to one there exist consistent roots (say $\hat{\boldsymbol{\alpha}}_n$ and $\tilde{\boldsymbol{\alpha}}_n$) of the estimating equations (5) and (6). Further $\hat{\boldsymbol{\alpha}}_n$ and $\tilde{\boldsymbol{\alpha}}_n$ satisfy (3).*

The next theorem says that if assumption **C6** is satisfied then one can also include the case $\beta = 0$ in assumption $(\mathbf{F}_{j\varepsilon})$. Thus for instance if one (rightly) assumes that C is a Frank copula then the marginal distributions of the errors are allowed to be also uniform or exponential.

Theorem 2. *Suppose that assumptions (\mathbf{ms}) , **C1**, **C2**, **C5**, **C6** and $(\mathbf{F}_{j\varepsilon})$ are satisfied. Then the statement of Theorem 1 holds.*

The above theorems imply that when fitting the copula C one can (under the stated assumptions) ignore the fact that he/she is working with estimated residuals $(\hat{\varepsilon}_{ij})$ instead of unobserved errors (ε_{ij}) . As it is known (and it also follows from the proof of Theorem 1) the asymptotic distribution of $\tilde{\boldsymbol{\alpha}}_n$ is normal. Thus thanks to (3) one can conclude that also $\hat{\boldsymbol{\alpha}}_n$ is asymptotically normal.

Corollary 1. *Suppose that the assumptions either of Theorem 1 or 2 hold. Then with probability going to one there exists a consistent root $\hat{\alpha}_n$ of (5). This root satisfies*

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow[n \rightarrow \infty]{d} \mathbf{N}_p(\mathbf{0}, \Sigma), \quad \Sigma = I^{-1}(\alpha) \text{var}(\tilde{\psi}(\mathbf{U})) I^{-1}(\alpha),$$

where $\tilde{\psi}(\mathbf{u}) = (\tilde{\psi}_1(\mathbf{u}), \dots, \tilde{\psi}_p(\mathbf{u}))^\top$ with

$$(11) \quad \tilde{\psi}_k(\mathbf{u}) = \psi_k(\mathbf{u}; \alpha) + \sum_{j=1}^d \int_{[0,1]^d} [\mathbf{1}\{u_j \leq v_j\} - v_j] \psi_k^{(j)}(\mathbf{v}; \alpha) dC(\mathbf{v}), \quad k = 1, \dots, p.$$

3. SIMULATION STUDY

A Monte Carlo study was conducted in order to illustrate the theoretical conclusions and to show how the finite sample performance of the maximum pseudo-likelihood estimator depends on the level of violation of the regularity assumptions.

3.1. Settings. To keep the presentation as clear as possible we concentrate on a bivariate response variable (some results for a three-dimensional case can be found in the Supplementary material) following the model

$$(12) \quad Y_{1i} = \theta_{10} + \theta_{11}X_i + \varepsilon_{1i}, \quad Y_{2i} = \theta_{20} + \theta_{21}X_i + \varepsilon_{2i}, \quad i = 1, \dots, n.$$

The joint cumulative distribution function $H(y_1, y_2)$ of the random vector $(\varepsilon_{1i}, \varepsilon_{2i})^\top$ is $C(F_{1\varepsilon}(y_1), F_{2\varepsilon}(y_2))$, where C is a copula and $F_{1\varepsilon}, F_{2\varepsilon}$ are marginal distribution functions. The following five copula families were considered for C : Clayton, Frank, Gumbel, Gaussian, and Student with 5 degrees of freedom. The copula parameter α is chosen such that the corresponding Kendall's tau is $\tau = 0.5$ or $\tau = 0.75$. The marginal distributions were chosen one of the following:

- $F_{1\varepsilon}$ is standard normal and $F_{2\varepsilon}$ exponential with mean 1 (denoted as N+E),
- $F_{1\varepsilon}$ is standard normal and $F_{2\varepsilon}$ uniform on $[-1, 1]$ (denoted as N+U),
- $F_{1\varepsilon}$ and $F_{2\varepsilon}$ are both Student t with 5 degrees of freedom (denoted as t).

The first two situations satisfy the assumption $(\mathbf{F}_{j\varepsilon})$ only with $\beta = 0$. Hence, the result of Theorem 2 applies only if **(C6)** holds. From the five considered copula families, this is the case only for the Frank copula. On the other hand, the t marginals satisfy $(\mathbf{F}_{j\varepsilon})$ with $\beta > 0$ and the assumptions of Theorem 1 hold. Hence, these marginals provide a useful regular benchmark for a comparison with the first two situations.

The covariate X_i is generated from the standard normal distribution (Poisson distribution with mean 5 was considered as well, but the results are almost identical and are not reported). The presented results correspond to the particular choice $\theta_{10} = 1, \theta_{20} = -1$,

$\theta_{11} = 1$, and $\theta_{21} = 2$. The unobserved errors ε_{ji} are estimated as the residuals after fitting the regression lines (marginally) where the parameters are estimated with the help of the least squares method assuming $s_j \equiv 1$, $j = 1, 2$, cf. Remark 1.

The following estimators of the parameter α are compared:

- (i) (oracle) inversion of Kendall's tau based on the unobserved errors $\tilde{\alpha}^{(ik)}$;
- (ii) inversion of Kendall's tau based on the residuals $\hat{\alpha}^{(ik)}$;
- (iii) (oracle) maximum pseudo-likelihood estimator based on the unobserved errors $\tilde{\alpha}^{(pl)}$;
- (iv) maximum pseudo-likelihood method estimator on the residuals $\hat{\alpha}^{(pl)}$;
- (v) modified maximum pseudo-likelihood estimator based on the residuals $\hat{\alpha}^{(pl*)}$.

The latter estimator $\hat{\alpha}^{(pl*)}$ is inspired by the estimator introduced in the context of single index conditional copulas by Fermanian and Lopez (2018). In our situation this estimator coincides with the maximum pseudo-likelihood estimator computed only from $\hat{\mathbf{U}}_i$ which lie in $[\delta_n, 1 - \delta_n]^2$, where $\delta_n = Dn^{-1/\lambda}$. Note that this choice corresponds to the choice δ_n in the proof of Theorem 1. In the presented simulations we choose $D = 1/4$ and $\lambda = 1.9$, thus in view of Remark 3 the statement of Theorem 1 (or 2) holds also for $\hat{\alpha}^{(pl*)}$ provided that the corresponding regularity assumptions hold.

In order to have more comparable results for the various copula families, the estimates of the parameters are presented on the Kendall's tau scale. The performance of the estimators is measured by the bias, the standard error (SD), and the root mean square error (RMSE), which are estimated from 1 000 random samples of sample sizes $n = 100, 1\,000, 10\,000$ and whose 100 multiplies are reported, because the obtained quantities are typically of order 10^{-2} . The obtained results for Clayton, Frank and Gaussian copulas are listed in Tables 1, 2, and 3, while tables for Gumbel and Student copula can be found in the Supplementary material. The Monte Carlo simulations were run in R statistical computing environment (R Core Team, 2018). The same starting seed was always used so that the estimates based on the true (but unobserved) errors ε_{ij} are the same regardless the choice of the marginals $F_{1\varepsilon}$ and $F_{2\varepsilon}$. These 'oracle' estimates are denoted as "inov" in the tables and provide benchmarks for the estimators calculated from the estimated residuals.

3.2. Findings. As it is well known (Genest et al., 1995; Tsukahara, 2005) in case of no covariates the maximum pseudo-likelihood is usually more efficient than the moment like estimators. This is illustrated by the performance of the estimators $\tilde{\alpha}^{(ik)}$ and $\tilde{\alpha}^{(pl)}$ that are calculated from the errors ε_{ij} . The question of interest is if this property continues to hold also for estimators that are calculated from the residuals (i.e., in the presence of covariates).

τ	margins	estim	$n = 100$			$n = 1\,000$			$n = 10\,000$		
			bias	SD	RMSE	bias	SD	RMSE	bias	SD	RMSE
0.50	inov	$\tilde{\alpha}^{(ik)}$	-0.03	5.54	5.54	0.00	1.64	1.64	-0.01	0.53	0.53
		$\tilde{\alpha}^{(pl)}$	0.33	4.90	4.91	0.01	1.49	1.49	0.00	0.48	0.48
	N+E	$\hat{\alpha}^{(ik)}$	-1.25	5.62	5.76	-0.27	1.64	1.67	-0.05	0.53	0.53
		$\hat{\alpha}^{(pl)}$	-3.91	5.54	6.78	-2.26	2.08	3.08	-0.80	0.75	1.10
		$\hat{\alpha}^{(pl*)}$	-1.94	5.30	5.65	-1.23	1.81	2.19	-0.44	0.63	0.77
	N+U	$\hat{\alpha}^{(ik)}$	-0.21	5.55	5.55	-0.03	1.63	1.63	-0.02	0.53	0.53
		$\hat{\alpha}^{(pl)}$	-0.84	4.86	4.93	-0.61	1.53	1.65	-0.22	0.51	0.55
		$\hat{\alpha}^{(pl*)}$	0.02	5.00	5.00	-0.13	1.50	1.51	-0.05	0.49	0.49
	t	$\hat{\alpha}^{(ik)}$	-0.15	5.58	5.58	-0.01	1.64	1.64	-0.02	0.53	0.53
		$\hat{\alpha}^{(pl)}$	0.10	4.96	4.96	-0.02	1.50	1.50	-0.01	0.48	0.48
		$\hat{\alpha}^{(pl*)}$	0.38	5.05	5.06	0.06	1.51	1.51	0.02	0.48	0.48
0.75	inov	$\tilde{\alpha}^{(ik)}$	0.02	3.40	3.40	-0.01	1.01	1.01	0.01	0.31	0.31
		$\tilde{\alpha}^{(pl)}$	-0.77	3.12	3.21	-0.16	0.93	0.94	-0.01	0.28	0.28
	N+E	$\hat{\alpha}^{(ik)}$	-2.14	3.70	4.27	-0.48	1.08	1.18	-0.07	0.32	0.33
		$\hat{\alpha}^{(pl)}$	-9.19	5.85	10.89	-4.19	2.88	5.09	-1.57	1.14	1.94
		$\hat{\alpha}^{(pl*)}$	-6.26	4.95	7.98	-2.86	2.36	3.71	-1.07	0.94	1.43
	N+U	$\hat{\alpha}^{(ik)}$	-0.24	3.39	3.40	-0.06	1.01	1.01	0.00	0.31	0.31
		$\hat{\alpha}^{(pl)}$	-2.99	3.27	4.43	-1.22	1.18	1.70	-0.44	0.41	0.60
		$\hat{\alpha}^{(pl*)}$	-1.63	3.15	3.55	-0.60	1.01	1.17	-0.20	0.33	0.39
	t	$\hat{\alpha}^{(ik)}$	-0.22	3.45	3.45	-0.05	1.01	1.01	0.01	0.31	0.31
		$\hat{\alpha}^{(pl)}$	-1.21	3.21	3.43	-0.22	0.93	0.95	-0.02	0.28	0.28
		$\hat{\alpha}^{(pl*)}$	-1.04	3.24	3.40	-0.17	0.93	0.95	-0.01	0.28	0.28

TABLE 1. Model (12) with Clayton copula, quantities multiplied by 100.

Generally speaking one can conclude that in agreement with our theoretical results the maximum pseudo-likelihood estimator $\hat{\alpha}^{(pl)}$ outperforms $\hat{\alpha}^{(ik)}$ in situations for which our regularity assumptions are satisfied (see Table 2 and the rows corresponding to t -marginals in Tables 1 and 3). For these situations the modified maximum pseudo-likelihood estimator $\hat{\alpha}^{(pl*)}$ is of no interest.

On the other hand the performance of $\hat{\alpha}^{(pl)}$ may deteriorate significantly if the regularity assumptions are not met. The problems are generally worse for larger values of Kendall's tau (a stronger dependence). It is also interesting that exponential margins (rows denoted as N+E) are much more problematic than uniform margins (rows denoted as N+U).

As illustrated in Table 1 one should be in particular careful when fitting the Clayton copula (and also the Gumbel copula as illustrated in the Supplementary material). Then

τ	margins	estim	$n = 100$			$n = 1\,000$			$n = 10\,000$		
			bias	SD	RMSE	bias	SD	RMSE	bias	SD	RMSE
0.50	inov	$\tilde{\alpha}^{(ik)}$	-0.03	4.62	4.62	0.01	1.44	1.43	0.01	0.45	0.45
		$\tilde{\alpha}^{(pl)}$	-0.03	4.51	4.50	0.01	1.42	1.42	0.01	0.45	0.45
	N+E	$\hat{\alpha}^{(ik)}$	-0.45	4.68	4.70	-0.05	1.44	1.44	0.00	0.45	0.45
		$\hat{\alpha}^{(pl)}$	-0.45	4.55	4.57	-0.05	1.43	1.43	0.00	0.45	0.45
		$\hat{\alpha}^{(pl*)}$	-0.21	4.84	4.84	-0.04	1.46	1.46	0.00	0.45	0.45
	N+U	$\hat{\alpha}^{(ik)}$	-0.08	4.65	4.65	0.00	1.44	1.43	0.00	0.45	0.45
		$\hat{\alpha}^{(pl)}$	-0.08	4.53	4.53	0.00	1.42	1.42	0.01	0.45	0.45
		$\hat{\alpha}^{(pl*)}$	0.09	4.85	4.85	0.01	1.45	1.45	0.01	0.45	0.45
0.75	inov	$\tilde{\alpha}^{(ik)}$	-0.11	2.50	2.50	0.00	0.74	0.74	0.00	0.23	0.23
		$\tilde{\alpha}^{(pl)}$	-0.53	2.45	2.50	-0.06	0.74	0.74	0.00	0.23	0.22
	N+E	$\hat{\alpha}^{(ik)}$	-1.17	2.79	3.02	-0.14	0.76	0.77	-0.01	0.23	0.23
		$\hat{\alpha}^{(pl)}$	-1.59	2.77	3.19	-0.19	0.76	0.78	-0.02	0.23	0.23
		$\hat{\alpha}^{(pl*)}$	-1.42	2.90	3.23	-0.17	0.77	0.79	-0.01	0.23	0.23
	N+U	$\hat{\alpha}^{(ik)}$	-0.25	2.53	2.54	-0.01	0.74	0.74	0.00	0.23	0.23
		$\hat{\alpha}^{(pl)}$	-0.69	2.50	2.59	-0.07	0.74	0.74	0.00	0.23	0.23
		$\hat{\alpha}^{(pl*)}$	-0.57	2.62	2.68	-0.05	0.76	0.76	0.00	0.23	0.23

TABLE 2. Model (12) with Frank copula, quantities multiplied by 100.

$\hat{\alpha}^{(pl)}$ performs significantly worse than $\hat{\alpha}^{(ik)}$ in cases of non-regular margins combined with a strong dependence ($\tau = 0.75$). The problems can be to some extent prevented by considering the modified estimator $\hat{\alpha}^{(pl*)}$ in particular in case of uniform margins (N+U). Thus while for Frank copula the modified estimator $\hat{\alpha}^{(pl*)}$ is of no interest, for the Clayton (and the Gumbel) copula it presents an interesting alternative to the ‘standard’ pseudo maximum-likelihood estimator.

The results for the Gaussian copula (see Table 3) are of independence interest. Note that although the density of the copula function is unbounded, the estimator $\hat{\alpha}^{(pl)}$ performs better than $\hat{\alpha}^{(ik)}$ for $\tau = 0.5$ even in case of exponential margins (N+E). And this holds true for uniform margins (N+U) even for $\tau = 0.75$. This raises a question whether a milder assumptions than $(\mathbf{F}_{j\varepsilon})$ would be sufficient for the Gaussian copula.

An analogous simulation study was conducted also for a system of three linear regressions, where the vector of innovations was sampled from $C(F_{1\varepsilon}(y_1), F_{2\varepsilon}(y_2), F_{3\varepsilon}(y_3))$ with the marginals $F_{1\varepsilon}$ and $F_{2\varepsilon}$ being standard normal and $F_{3\varepsilon}$ either exponential (with mean 1) or uniform on $[-1, 1]$. As the obtained results are very similar to the results for model (12), they are not presented here, but can be found in the Supplementary material. The common

τ	margins	estim	$n = 100$			$n = 1\,000$			$n = 10\,000$		
			bias	SD	RMSE	bias	SD	RMSE	bias	SD	RMSE
0.50	inov	$\tilde{\alpha}^{(ik)}$	0.03	4.94	4.94	-0.07	1.53	1.53	0.00	0.48	0.48
		$\tilde{\alpha}^{(pl)}$	1.07	4.51	4.63	0.10	1.39	1.40	0.03	0.44	0.44
	N+E	$\hat{\alpha}^{(ik)}$	-0.43	4.97	4.99	-0.17	1.53	1.54	-0.02	0.48	0.48
		$\hat{\alpha}^{(pl)}$	0.32	4.54	4.55	-0.21	1.41	1.43	-0.06	0.45	0.45
		$\hat{\alpha}^{(pl*)}$	0.99	4.90	5.00	0.08	1.46	1.46	0.03	0.45	0.45
	N+U	$\hat{\alpha}^{(ik)}$	-0.06	4.97	4.97	-0.08	1.53	1.53	0.00	0.48	0.48
		$\hat{\alpha}^{(pl)}$	0.87	4.53	4.62	-0.01	1.40	1.39	-0.01	0.44	0.44
		$\hat{\alpha}^{(pl*)}$	1.36	4.86	5.05	0.22	1.46	1.47	0.07	0.45	0.45
	t	$\hat{\alpha}^{(ik)}$	0.04	4.99	4.98	-0.07	1.53	1.53	0.00	0.48	0.48
		$\hat{\alpha}^{(pl)}$	1.08	4.55	4.67	0.09	1.40	1.40	0.02	0.44	0.44
		$\hat{\alpha}^{(pl*)}$	1.42	4.90	5.10	0.21	1.46	1.48	0.06	0.45	0.45
0.75	inov	$\tilde{\alpha}^{(ik)}$	0.16	2.79	2.80	-0.02	0.89	0.89	0.00	0.27	0.27
		$\tilde{\alpha}^{(pl)}$	0.06	2.53	2.53	-0.02	0.80	0.80	0.00	0.25	0.25
	N+E	$\hat{\alpha}^{(ik)}$	-1.02	2.93	3.10	-0.24	0.90	0.93	-0.04	0.27	0.27
		$\hat{\alpha}^{(pl)}$	-1.81	2.81	3.34	-0.73	0.95	1.20	-0.21	0.30	0.37
		$\hat{\alpha}^{(pl*)}$	-1.01	2.77	2.95	-0.40	0.88	0.97	-0.10	0.27	0.29
	N+U	$\hat{\alpha}^{(ik)}$	-0.08	2.80	2.80	-0.05	0.89	0.89	-0.01	0.27	0.27
		$\hat{\alpha}^{(pl)}$	-0.48	2.52	2.56	-0.27	0.82	0.86	-0.09	0.25	0.27
		$\hat{\alpha}^{(pl*)}$	0.00	2.61	2.60	-0.05	0.81	0.81	-0.01	0.25	0.25
	t	$\hat{\alpha}^{(ik)}$	0.14	2.82	2.82	-0.02	0.89	0.89	-0.01	0.27	0.27
		$\hat{\alpha}^{(pl)}$	0.03	2.56	2.56	-0.02	0.79	0.79	0.00	0.25	0.25
		$\hat{\alpha}^{(pl*)}$	0.20	2.62	2.62	0.02	0.80	0.80	0.02	0.25	0.25

TABLE 3. Model (12) with Gaussian copula, quantities multiplied by 100.

important finding is that the pseudo-likelihood estimator $\hat{\alpha}^{(pl)}$ may perform poorly (and noticeably worse compared to $\hat{\alpha}^{(ik)}$) for copula families with unbounded densities even in cases when only one of the marginals does not satisfy the regularity assumption while the remaining ones are regular.

4. CONCLUSIONS AND FURTHER DISCUSSIONS

As illustrated in the previous section one should be careful when a copula with an unbounded density is fitted with the help of the maximum pseudo-likelihood method. Although the assumptions of Theorem 1 are not strict one should keep in mind that they are not satisfied for distributions with a non-continuous error density function $f_{j\varepsilon}$ (e.g., uniform distribution, exponential distribution, ...). Although such situations are probably

rare in practice, there are applications in which for instance uniform errors can naturally appear (see e.g., Schechtman and Schechtman, 1986).

One of the possible next steps would be to generalize the results into the time-series context and to find the assumptions so that the results claimed in Chen and Fan (2006a) hold. Based on our results for i.i.d. setting and our simulation study we conjecture that the method of the pseudo-likelihood estimation can be problematic when the marginal models have exponential innovations (or more generally positive or bounded innovations with discontinuous density) (see e.g. Lawrance and Lewis, 1985; Davis and McCormick, 1989; Anděl, 1989, 1992; Nielsen and Shephard, 2003) and one uses \sqrt{n} -consistent estimators of the model parameters.

Note that in models where (based on our findings) the use of maximum pseudo-likelihood estimation is questionable, one can consider the method of moments (see e.g., Section 5.5.1 of McNeil et al., 2005; Brahimi and Necir, 2012). As proved in Côté et al. (2019) many moment estimators based on residuals satisfy (3) under less restrictive assumptions on the marginal error density $f_{j\varepsilon}$. In particular for standard two-dimensional copulas the method of the inversion of Kendall's tau can present a 'robust' alternative. It is usually only slightly less efficient if no covariates are present, but in the presence of covariates it can perform significantly better than the maximum pseudo-likelihood estimator.

For the sake of brevity we concentrated only on estimation of the copula parameter. We conjecture that also other procedures (e.g., procedures for goodness-of-fit testing) that make use of the maximum pseudo-likelihood estimator $\hat{\alpha}_n$ calculated from the residuals will be valid provided that next to our assumptions also some standard regularity assumptions for these procedures are satisfied.

ACKNOWLEDGMENTS

M. Omelka gratefully acknowledges support from the grant GACR 19-00015S. The research of Š. Hudecová was supported by the grant GACR 18-01781Y. N. Neumeyer gratefully acknowledges support from the DFG (Research Unit FOR 1735 Structural Inference in Statistics: Adaptation and Efficiency).

APPENDIX A. PROOFS OF THE MAIN RESULTS

Note that the estimated pseudoobservations $\hat{\mathbf{U}}_i$ given by (4) can be viewed as estimates of 'unobserved' pseudoobservations $\tilde{\mathbf{U}}_i$ (given in (7)) which can be further viewed as estimates

of \mathbf{U}_i , given by

$$\mathbf{U}_i = (U_{1i}, \dots, U_{di})^\top = (F_{1\varepsilon}(\varepsilon_{1i}), \dots, F_{d\varepsilon}(\varepsilon_{di}))^\top.$$

To prove Theorem 1 we need some technical results about the ‘closeness’ of \widehat{U}_{ji} (the j -th element of $\widehat{\mathbf{U}}_i$) to \widetilde{U}_{ji} and U_{ji} .

As we will show later one does not need to handle \widehat{U}_{ji} if either U_{ji} is close to zero or one or if $M_j(\mathbf{X}_i)$ is too large. This is formalised as follows. Introduce the set of indices

$$(A1) \quad J_{jn}^X = \{i \in \{1, \dots, n\} : U_{ji} \in [\delta_n, 1 - \delta_n], M_j(\mathbf{X}_i) \leq a_n\},$$

where

$$(A2) \quad \delta_n = \frac{1}{n^{1/\lambda}}, \quad \text{and} \quad a_n = n^{1/(\lambda_x r)}, \quad \text{for some } \lambda \geq 1 \text{ and } 0 < \lambda_x \leq \lambda.$$

The following lemma gives an upper bound on the number of indices i for which it holds that $U_{ji} \notin [\delta_n, 1 - \delta_n]$ or $M_j(\mathbf{X}_i) > a_n$.

Lemma 1. *Let δ_n and a_n satisfy (A2) and assumption **(ms)** holds. Then*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{U_{ji} \notin [\delta_n, 1 - \delta_n] \text{ or } M_j(\mathbf{X}_i) > a_n\} = O_P\left(\frac{1}{n^{1/\lambda}}\right),$$

which further implies that

$$\mathbb{P}\left(\sum_{i=1}^n \mathbf{1}\{U_{ji} \notin [\delta_n, 1 - \delta_n] \text{ or } M_j(\mathbf{X}_i) > a_n\} \leq n^{1-1/\lambda} \log n\right) \xrightarrow{n \rightarrow \infty} 1.$$

Proof. Denote

$$p_n = \mathbb{P}(U_{ji} \notin [\delta_n, 1 - \delta_n] \text{ or } M_j(\mathbf{X}_i) > a_n)$$

and note that thanks to (A2) and Markov’s inequality (applied to $M_j^r(\mathbf{X}_i)$)

$$\begin{aligned} p_n &\leq \mathbb{P}(U_{ji} \notin [\delta_n, 1 - \delta_n]) + \mathbb{P}(M_j(\mathbf{X}_i) > a_n) \\ &\leq 2\delta_n + \mathbb{E} \frac{M_j^r(\mathbf{X}_i)}{a_n^r} = O\left(\frac{1}{n^{1/\lambda}}\right) + O\left(\frac{1}{n^{1/\lambda_x}}\right) = O\left(\frac{1}{n^{1/\lambda}}\right). \end{aligned}$$

Now as the random variable $\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{U_{ji} \notin [\delta_n, 1 - \delta_n] \text{ or } M_j(\mathbf{X}_i) > a_n\}$ is non-negative one can use once more Markov’s inequality to conclude that

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{U_{ji} \notin [\delta_n, 1 - \delta_n] \text{ or } M_j(\mathbf{X}_i) > a_n\} = O_P(p_n).$$

□

A.1. Some results on statistics with ranks calculated from residuals.

Lemma 2. *Suppose that assumptions $(\mathbf{F}_{j\varepsilon})$ and (\mathbf{ms}) hold and that φ is a \mathcal{J} -function. Then*

$$\frac{1}{n} \sum_{i=1}^n \varphi(\widehat{\mathbf{U}}_i) \xrightarrow[n \rightarrow \infty]{P} \mathbb{E} \varphi(\mathbf{U}).$$

Proof. As φ is a \mathcal{J} -function, it is easy to show that the expectation $\mathbb{E} \varphi(\mathbf{U})$ exists and is finite. Thus thanks to the law of large numbers it is sufficient to show

$$(A3) \quad D_n = \left| \frac{1}{n} \sum_{i=1}^n \varphi(\widehat{\mathbf{U}}_i) - \frac{1}{n} \sum_{i=1}^n \varphi(\mathbf{U}_i) \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Let J_{jn}^X and δ_n be as in (A1) and (A2), where λ and λ_x are chosen so that they satisfy the assumptions of Lemma 6. Then this lemma together with the standard Glivenko-Cantelli theorem for the empirical distribution function $\widehat{F}_{j\varepsilon}$ implies that

$$(A4) \quad \begin{aligned} & \max_{j \in \{1, \dots, d\}} \max_{i \in J_{jn}^X} |\widehat{U}_{ji} - U_{ji}| \\ & \leq \max_{j \in \{1, \dots, d\}} \max_{i \in J_{jn}^X} |\widehat{U}_{ji} - \widetilde{U}_{ji}| + \max_{j \in \{1, \dots, d\}} \max_{i \in J_{jn}^X} |\widetilde{U}_{ji} - U_{ji}| = o_P(1). \end{aligned}$$

Now introduce

$$(A5) \quad J_n^X = \cap_{j=1}^d J_{jn}^X, \quad \text{and} \quad K_n^X = \{1, \dots, n\} \setminus J_n^X$$

and note that with the help of (A4)

$$(A6) \quad \max_{i \in J_n^X} \|\widehat{\mathbf{U}}_i - \mathbf{U}_i\| = o_P(1).$$

As the above equation is not guaranteed for $i \in K_n^X$, we need to take care about the sets of indices J_n^X and K_n^X separately. That is why we bound D_n given by (A3) as

$$(A7) \quad D_n \leq \frac{1}{n} \sum_{i \in K_n^X} |\varphi(\widehat{\mathbf{U}}_i)| + \frac{1}{n} \sum_{i \in K_n^X} |\varphi(\mathbf{U}_i)| + \left| \frac{1}{n} \sum_{i \in J_n^X} \varphi(\widehat{\mathbf{U}}_i) - \frac{1}{n} \sum_{i \in J_n^X} \varphi(\mathbf{U}_i) \right|.$$

In what follows we show that each term on the right-hand side of (A7) is asymptotically negligible.

Dealing with the first term in (A7)

As φ is a \mathcal{J} -function one can bound

$$\frac{1}{n} \sum_{i \in K_n^X} |\varphi(\widehat{\mathbf{U}}_i)| \leq \sum_{j=1}^d \frac{M_1}{n} \sum_{i \in K_n^X} \frac{1}{[\min\{\widehat{U}_{ji}, 1 - \widehat{U}_{ji}\}]^\eta}.$$

Now by Lemma 1 (with probability going to one) there are at most $dn^{1-1/\lambda} \log n$ indices i for which there exists $j \in \{1, \dots, d\}$ such that $U_{ji} \notin [\delta_n, 1 - \delta_n]$ or $M_j(\mathbf{X}_i) > a_n$. Thus one can choose the indices i for which $\frac{1}{[\min\{\widehat{U}_{ji}, 1 - \widehat{U}_{ji}\}]^\eta}$ takes the biggest values and gets that (with probability going to one)

$$(A8) \quad \begin{aligned} \frac{1}{n} \sum_{i \in K_n^X} |\varphi(\widehat{\mathbf{U}}_i)| &\leq \sum_{j=1}^d \frac{M_1}{n} \left[\sum_{i=1}^{\lceil \frac{d}{2} n^{1-1/\lambda} \log n \rceil} \frac{1}{(\frac{i}{n+1})^\eta} + \sum_{i=\lfloor n - \frac{d}{2} n^{1-1/\lambda} \log n \rfloor}^n \frac{1}{(1 - \frac{i}{n+1})^\eta} \right] \\ &\leq 2d^2 M_1 n^{-(1-\eta)/\lambda} (\log n)^{1-\eta} (1 + o(1)) = o(1). \end{aligned}$$

Dealing with the second term in (A7)

Note that $\mathbf{E}|\varphi(\mathbf{U}_i)| < \infty$ implies that

$$\begin{aligned} \mathbf{E} \left[\frac{1}{n} \sum_{i \in K_n^X} |\varphi(\mathbf{U}_i)| \right] &= \mathbf{E} \left[|\varphi(\mathbf{U}_i)| \mathbf{1}\{\mathbf{U}_i \notin [\delta_n, 1 - \delta_n]^d \text{ or } \max_{1 \leq j \leq d} M_j(\mathbf{X}_i) > a_n\} \right] \\ &\leq \mathbf{E} \left[|\varphi(\mathbf{U}_i)| \mathbf{1}\{\mathbf{U}_i \notin [\delta_n, 1 - \delta_n]^d\} \right] + \mathbf{E} |\varphi(\mathbf{U}_i)| \mathbf{P} \left(\max_{1 \leq j \leq d} M_j(\mathbf{X}_i) > a_n \right) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Thus $\frac{1}{n} \sum_{i \in K_n^X} |\varphi(\mathbf{U}_i)| = o_P(1)$ follows from Markov's inequality.

Dealing with the third term in (A7)

We use the continuity of the function φ . To be able to do that we need to stay in the interior of $[0, 1]^d$. Thus for a given $\delta \in (0, 1/2)$ (that will be specified later on), consider the set

$$(A9) \quad \mathbf{I}_\delta = \{\mathbf{u} : \mathbf{u} \in [\delta, 1 - \delta]^d\}.$$

and introduce the corresponding sets of indices

$$(A10) \quad J_\delta = \{i \in \{1, \dots, n\} : \mathbf{U}_i \in \mathbf{I}_\delta\}, \quad K_\delta = \{1, \dots, n\} \setminus J_\delta,$$

where for simplicity of notation we do not stress that both J_δ and K_δ depends on n . Now one can bound

$$(A11) \quad \begin{aligned} &\left| \frac{1}{n} \sum_{i \in J_n^X} \varphi(\widehat{\mathbf{U}}_i) - \frac{1}{n} \sum_{i \in J_n^X} \varphi(\mathbf{U}_i) \right| \\ &\leq \frac{1}{n} \sum_{i \in J_n^X \cap J_\delta} |\varphi(\widehat{\mathbf{U}}_i) - \varphi(\mathbf{U}_i)| + \frac{1}{n} \sum_{i \in J_n^X \cap K_\delta} |\varphi(\widehat{\mathbf{U}}_i)| + \frac{1}{n} \sum_{i \in J_n^X \cap K_\delta} |\varphi(\mathbf{U}_i)|. \end{aligned}$$

Note that by the uniform continuity of the function $\varphi(\cdot)$ on $[\delta/2, 1 - \delta/2]^d$ and (A6) one gets that the first term on the right-hand side of (A11) converges to zero in probability.

To deal with the second term on the right-hand side of (A11) note that thanks to (A6) with probability going to one

$$J_n^X \cap K_\delta \subseteq \{i \in \{1, \dots, n\} : \widehat{\mathbf{U}}_i \notin [2\delta, 1 - 2\delta]^d\}$$

Thus one can bound

$$\begin{aligned} \frac{1}{n} \sum_{i \in J_n^X \cap K_\delta} |\varphi(\widehat{\mathbf{U}}_i)| &\leq \sum_{j=1}^d \frac{M_1}{n} \left[\sum_{i=1}^{\lceil (n+1)2\delta \rceil} \frac{1}{\left(\frac{i}{n+1}\right)^\eta} + \sum_{i=\lfloor n-(n+1)2\delta \rfloor}^n \frac{1}{\left(1 - \frac{i}{n+1}\right)^\eta} \right] \\ &\leq 2d M_1 \frac{(2\delta)^{1-\eta}}{1-\eta} (1 + o(1)), \end{aligned}$$

which can be made arbitrarily small by taking δ small enough.

Finally with the help of law of large numbers the third term on the right-hand side of (A11) can be bounded by

$$\begin{aligned} \frac{1}{n} \sum_{i \in J_n^X \cap K_\delta} |\varphi(\mathbf{U}_i)| &\leq \frac{1}{n} \sum_{i \in K_\delta} |\varphi(\mathbf{U}_i)| \\ &\leq \sum_{j=1}^d \frac{M_1}{n} \left[\sum_{i=1}^n \frac{1}{U_{ji}^\eta} \mathbf{1}\{U_{ji} \leq \delta\} + \sum_{i=1}^n \frac{1}{(1 - U_{ji})^\eta} \mathbf{1}\{U_{ji} \geq 1 - \delta\} \right] \\ &= 2d M_1 \left(\frac{\delta^{1-\eta}}{1-\eta} + o_P(1) \right), \end{aligned}$$

which can be also made arbitrarily small by taking δ sufficiently small and n sufficiently large. \square

Lemma 3. *Suppose that assumptions $(\mathbf{F}_{j_\varepsilon})$ and (\mathbf{ms}) hold. Let φ be a $\widetilde{\mathcal{J}}^{\beta_1, \beta_2}$ -function such that $\mathbf{E}\{\varphi(\mathbf{U})\} = 0$ and $\beta > \max\{\beta_1 + \frac{1}{r-1}, \beta_2\}$. Then*

$$(A12) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\widehat{\mathbf{U}}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\widetilde{\mathbf{U}}_i) + o_P(1).$$

Proof. Let J_n^X and K_n^X be defined as in (A5). Then similarly as in (A8) of the proof of Lemma 2 one can bound

$$(A13) \quad \frac{1}{\sqrt{n}} \sum_{i \in K_n^X} \varphi(\widehat{\mathbf{U}}_i) = O_P(n^{\frac{1}{2} - \frac{1-\beta_1}{\lambda}} \log n), \quad \frac{1}{\sqrt{n}} \sum_{i \in K_n^X} \varphi(\widetilde{\mathbf{U}}_i) = O_P(n^{\frac{1}{2} - \frac{1-\beta_1}{\lambda}} \log n),$$

where the role of η is now taken by β_1 .

In what follows we take λ so that

$$2(1 - \beta + \frac{1}{r-1}) < \lambda < 2(1 - \beta_1)$$

and λ_x satisfies (B30). Such choices of λ and λ_x guarantee that the right-hand sides of (A13) are of order $o_P(1)$ and at the same time the assumptions of Lemma 5 are satisfied and one can make use of Lemmas 6 and 7.

It is sufficient to show that

$$\frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi(\widehat{\mathbf{U}}_i) = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi(\widetilde{\mathbf{U}}_i) + o_P(1).$$

Note that

$$J_n^X = \{i \in \{1, \dots, n\} : \mathbf{U}_i \in [\delta_n, 1 - \delta_n]^d, \max_{1 \leq j \leq d} M_j(\mathbf{X}_i) \leq a_n\},$$

where δ_n and a_n are given in (A2).

Now by the mean value theorem

$$(A14) \quad \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi(\widehat{\mathbf{U}}_i) = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi(\widetilde{\mathbf{U}}_i) + \sum_{j=1}^d \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) (\widehat{U}_{ji} - \widetilde{U}_{ji}),$$

where U_{ji}^* lies between \widehat{U}_{ji} and \widetilde{U}_{ji} . Thus to prove the lemma it is sufficient to show that the second term on the right-hand side of (A14) diminishes in probability.

With the help of Lemma 6 for a fixed $j \in \{1, \dots, d\}$ one gets

$$\frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) (\widehat{U}_{ji} - \widetilde{U}_{ji}) = A_n + B_n + C_n,$$

where

$$(A15) \quad A_n = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(\varepsilon_{ji}) \left\{ \mathbf{E}_X \left[\frac{m'_j(\mathbf{X}, \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right] + \varepsilon_{ji} \mathbf{E}_X \left[\frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right] \right\}^\top (\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j),$$

$$(A16) \quad B_n = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(\varepsilon_{ji}) (\widehat{\varepsilon}_{ji} - \varepsilon_{ji}),$$

$$(A17) \quad C_n = \frac{o_P(1)}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) U_{ji}^{\beta-\gamma} (1 - U_{ji})^{\beta-\gamma} (1 + M_j(\mathbf{X}_i)),$$

and $\gamma > 0$ is taken sufficiently small so that $\beta - \gamma > \beta_2$. In what follows we show that C_n and $A_n + B_n$ are asymptotically negligible.

Dealing with C_n . With the help of Lemma A3 of Shorack (1972) and Lemma 7 for each $\varepsilon > 0$ there exists a positive constant L such that the quantity C_n given by (A17) can be

with probability at least $1 - \varepsilon$ bounded by

$$\begin{aligned}
 |C_n| &\leq \frac{o_P(1)}{n} \sum_{i \in J_n^X} |\varphi^{(j)}(\mathbf{U}_i^*)(U_{ji}^*)^{\beta_2}(1 - U_{ji}^*)^{\beta_2}| \frac{U_{ji}^{\beta-\gamma}(1 - U_{ji})^{\beta-\gamma}}{(U_{ji}^*)^{\beta_2}(1 - U_{ji}^*)^{\beta_2}} (1 + M_j(\mathbf{X}_i)) \\
 &\leq \frac{o_P(1)}{n} \sum_{i \in J_n^X} \frac{M_1}{[\min_{j=1,\dots,d} \min\{U_{ji}^*, 1 - U_{ji}^*\}]^\eta} \frac{1}{L^{\beta_2}} (1 + M_j(\mathbf{X}_i)) \\
 &= \frac{o_P(1)}{n} \sum_{i \in J_n^X} \frac{M_1}{[\min_{j=1,\dots,d} \min\{U_{ji}, 1 - U_{ji}\}]^\eta} \frac{1}{L^{\beta_2+\eta}} (1 + M_j(\mathbf{X}_i)) \\
 &= o_P(1) O_P(1) = o_P(1),
 \end{aligned}$$

where the law of large numbers is used on the last line.

Thus one can concentrate on the quantities A_n and B_n .

Dealing with A_n . Note that A_n given by (A15) can be rewritten as

$$\begin{aligned}
 (A18) \quad A_n &= \sqrt{n} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)^\top \mathbf{E}_{\mathbf{X}} \left[\frac{m'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right] \frac{1}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(\varepsilon_{ji}) \\
 &\quad + \sqrt{n} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)^\top \mathbf{E}_{\mathbf{X}} \left[\frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right] \frac{1}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(\varepsilon_{ji}) \varepsilon_{ji}.
 \end{aligned}$$

Now analogously as in the proof of Lemma 2 one can show that

$$\begin{aligned}
 (A19) \quad \frac{1}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(\varepsilon_{ji}) &= \frac{1}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_{ji})) \\
 &= \mathbf{E} [\varphi^{(j)}(\mathbf{U}) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_j))] + o_P(1)
 \end{aligned}$$

and also

$$(A20) \quad \frac{1}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) f_{j\varepsilon}(\varepsilon_{ji}) \varepsilon_{ji} = \mathbf{E} [\varphi^{(j)}(\mathbf{U}) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_j)) F_{j\varepsilon}^{-1}(U_j)] + o_P(1).$$

Combining (A18), (A19), (A20) and the fact that the estimator $\hat{\boldsymbol{\theta}}_j$ is \sqrt{n} -consistent yields

$$\begin{aligned}
 (A21) \quad A_n &= \sqrt{n} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)^\top \mathbf{E} [\varphi^{(j)}(\mathbf{U}) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_j))] \mathbf{E}_{\mathbf{X}} \left[\frac{m'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right] \\
 &\quad + \sqrt{n} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)^\top \mathbf{E} [\varphi^{(j)}(\mathbf{U}) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_j)) F_{j\varepsilon}^{-1}(U_j)] \mathbf{E}_{\mathbf{X}} \left[\frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right] + o_P(1).
 \end{aligned}$$

Dealing with B_n . Now have a look at the term B_n defined in (A16). One can proceed analogously as above and show that

$$(A22) \quad B_n = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i) f_{j\varepsilon}(\varepsilon_{ji}) (\hat{\varepsilon}_{ji} - \varepsilon_{ji}) + o_P(1) = B_{n1} + B_{n2} + o_P(1),$$

where

$$B_{n1} = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_{ji})) \left[\frac{m_j(\mathbf{X}_i; \boldsymbol{\theta}_j) - m_j(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_j)}{s_j(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_j)} \right],$$

$$B_{n2} = \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_{ji})) F_{j\varepsilon}^{-1}(U_{ji}) \left[\frac{s_j(\mathbf{X}_i; \boldsymbol{\theta}_j) - s_j(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_j)}{s_j(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_j)} \right].$$

Now similarly as in the proof of Lemma 5 one can show that

$$\begin{aligned} B_{n1} &= \sqrt{n}(\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j)^\top \frac{1}{n} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_{ji})) \frac{m'_j(\mathbf{X}_i; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_j)} + o_P(1) \\ (A23) \quad &= \sqrt{n}(\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j)^\top \mathbb{E}[\varphi^{(j)}(\mathbf{U}) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_j))] \mathbb{E}\left[\frac{m'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)}\right] + o_P(1) \end{aligned}$$

and analogously also

$$(A24) \quad B_{n2} = \sqrt{n}(\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j)^\top \mathbb{E}[\varphi^{(j)}(\mathbf{U}) f_{j\varepsilon}(F_{j\varepsilon}^{-1}(U_j)) F_{j\varepsilon}^{-1}(U_j)] \mathbb{E}\left[\frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)}\right] + o_P(1).$$

Now (A21), (A22), (A23) and (A24) yields that $B_n = -A_n + o_P(1)$, which was to be proved. □

The following lemma will be useful for copula families with ‘nicely bounded’ score functions.

Lemma 4. *Suppose that assumptions $(\mathbf{F}_{j\varepsilon})$ and (\mathbf{ms}) hold. Let φ be a $\tilde{\mathcal{J}}^{0,0}$ -function such that $\mathbb{E}\{\varphi(\mathbf{U})\} = 0$ and $\varphi^{(j)}$ is bounded for each $j \in \{1, \dots, p\}$. Then the statement of Lemma 3 holds.*

Proof. By the mean value theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\hat{\mathbf{U}}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\tilde{\mathbf{U}}_i) + \sum_{j=1}^d \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi^{(j)}(\mathbf{U}_i^*) (\hat{U}_{ji} - \tilde{U}_{ji}).$$

Now take $\lambda > 2(1 + \frac{1}{r})$ and recall the sets of indices J_n^X of K_n^X introduced in (A5). Then

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi^{(j)}(\mathbf{U}_i^*) (\hat{U}_{ji} - \tilde{U}_{ji}) \\ (A25) \quad &= \frac{1}{\sqrt{n}} \sum_{i \in J_n^X} \varphi^{(j)}(\mathbf{U}_i^*) (\hat{U}_{ji} - \tilde{U}_{ji}) + \frac{1}{\sqrt{n}} \sum_{i \in K_n^X} \varphi^{(j)}(\mathbf{U}_i^*) (\hat{U}_{ji} - \tilde{U}_{ji}). \end{aligned}$$

Now with the help of Lemma 9 one can show that the second term on the right-hand side of (A25) can be bounded as the preceding equation is $o_P(1)$

$$\frac{1}{\sqrt{n}} \sum_{i \in K_n^X} |\varphi^{(j)}(\mathbf{U}_i^*)(\widehat{U}_{ji} - \widetilde{U}_{ji})| \leq \frac{O_P(1)}{n} \sum_{i \in K_n^X} (1 + M_j(\mathbf{X}_i)) = o_P(1),$$

where the last equation follows from Markov's inequality and

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \sum_{i \in K_n^X} (1 + M_j(\mathbf{X}_i)) \right] &= \mathbb{E} \left[(1 + M_j(\mathbf{X})) \mathbf{1} \{ \mathbf{U} \notin [\delta_n, 1 - \delta_n]^d \text{ or } \max_{1 \leq j \leq d} M_j(\mathbf{X}) > a_n \} \right] \\ &= o(1). \end{aligned}$$

Finally the first term on the right-hand side of (A25) can be handled analogously as in the proof of Lemma 3. \square

Corollary 2. *Suppose that assumptions of Lemma 3 or Lemma 4 are satisfied. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\widehat{\mathbf{U}}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{\varphi}(\mathbf{U}_i) + o_P(1),$$

where

$$\widetilde{\varphi}(\mathbf{u}) = \varphi(\mathbf{u}) + \sum_{j=1}^d \int_{[0,1]^d} [\mathbf{1}\{u_j \leq v_j\} - v_j] \varphi^{(j)}(\mathbf{v}) dC(\mathbf{v}).$$

Proof. With the help of (A12) it is sufficient to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\widetilde{\mathbf{U}}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{\varphi}(\mathbf{U}_i) + o_P(1).$$

But this can be proved component-wise by mimicking the proof of Lemma 2 of Gijbels et al. (2017), where the situation with $d = 2$ but a more general φ depending possibly also on \mathbf{X}_i is considered. \square

A.2. Proofs of Theorems 1 and 2.

Proof of Theorem 1. With the help of Lemmas 2 and 3 the proof can closely follow the proof of Lemma 3 in Gijbels et al. (2017). In order to do that define

$$(A26) \quad \mathbf{W}_n(\mathbf{a}) = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\psi}(\widehat{\mathbf{U}}_i; \mathbf{a}) \quad \text{and} \quad \mathbf{W}(\mathbf{a}) = \mathbb{E} \boldsymbol{\psi}(\mathbf{U}; \mathbf{a}).$$

In what follows we show that assumptions of Theorem A.10.2 of Bickel et al. (1993) are satisfied for \mathbf{W}_n and \mathbf{W} given by (A26).

It follows from the standard maximum likelihood theory that Assumption (GM0) is satisfied thanks to Assumption **C1**. Moreover, Assumptions **C4** and **C5** imply Assumption (GM3). Assumption (GM2) is also satisfied as thanks to assumption **C3** one can for each $k \in \{1, \dots, p\}$ apply Corollary 2 to $\varphi(\mathbf{u}) = \psi_k(\mathbf{u}; \boldsymbol{\alpha})$ and get

$$\frac{1}{n} \sum_{i=1}^n \psi(\widehat{\mathbf{U}}_i; \boldsymbol{\alpha}) = \frac{1}{n} \sum_{i=1}^n \tilde{\psi}(\mathbf{U}_i) + o_P(n^{-1/2}),$$

where $\tilde{\psi}(\mathbf{u})$ was introduced in Corollary 1.

Thus, it remains to check Assumption (U) from Theorem A.10.2. Therefore for each $\varepsilon > 0$ and for each $k, \ell \in \{1, \dots, p\}$, it is sufficient to find a neighborhood $\mathcal{U}_\varepsilon = \{\mathbf{a} \in \mathcal{U} : \|\mathbf{a} - \boldsymbol{\alpha}\| < \varepsilon\}$ such that

$$\sup_{\mathbf{a} \in \mathcal{U}_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^n \frac{\partial \psi_k(\widehat{\mathbf{U}}_i; \mathbf{a})}{\partial a_\ell} - I^{(j, \ell)}(\mathbf{a}) \right| \leq \varepsilon + o_P(1),$$

where $I^{(j, \ell)}(\mathbf{a})$ stands for the (j, ℓ) element of $I(\mathbf{a})$.

For simplicity of notation, let us put $g_{k, \ell}(\mathbf{u}; \mathbf{a}) = \partial \psi_k(\mathbf{u}; \mathbf{a}) / \partial a_\ell$. Assumption **C4** allows to adapt Lemma 2, which gives

$$\frac{1}{n} \sum_{i=1}^n g_{k, \ell}(\widehat{\mathbf{U}}_i; \boldsymbol{\alpha}) - I^{(k, \ell)}(\boldsymbol{\alpha}) = o_P(1).$$

Hence, it remains to show

$$(A27) \quad D_n = \sup_{\mathbf{a} \in \mathcal{U}_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^n g_{k, \ell}(\widehat{\mathbf{U}}_i; \mathbf{a}) - \frac{1}{n} \sum_{i=1}^n g_{k, \ell}(\widehat{\mathbf{U}}_i; \boldsymbol{\alpha}) \right| \leq \varepsilon + o_P(1).$$

For a given $\delta \in (0, 1/4)$ (that will be specified later on), let us introduce the sets \mathbf{I}_δ and \mathbf{J}_δ as in (A9) and (A10). Then the left-hand side of (A27) can be bounded by

$$(A28) \quad D_n \leq \sup_{\mathbf{a} \in \mathcal{U}_\varepsilon} \left| \frac{1}{n} \sum_{i \in \mathbf{J}_\delta \cap \mathbf{J}_n^X} g_{k, \ell}(\widehat{\mathbf{U}}_i; \mathbf{a}) - \frac{1}{n} \sum_{i \in \mathbf{J}_\delta \cap \mathbf{J}_n^X} g_{k, \ell}(\widehat{\mathbf{U}}_i; \boldsymbol{\alpha}) \right| + \frac{2}{n} \sum_{i \notin \mathbf{J}_\delta \cap \mathbf{J}_n^X} h(\widehat{\mathbf{U}}_i),$$

where \mathbf{J}_n^X was introduced in (A5) and h in Assumption **C4**. Now with probability going to one for each sufficiently large n , if $\mathbf{U}_i \in \mathbf{I}_\delta$, then $\widehat{\mathbf{U}}_i \in \mathbf{I}_{\delta/2}$. Thus for each $\delta \in (0, 1/4)$ the term on the right-hand side of (A28) can be made arbitrarily small (Assumption **C4**) up to $o_P(1)$ term by considering a sufficiently small neighbourhood \mathcal{U}_ε .

Finally, analogously as in the proof of Lemma 2, one can show that

$$\frac{1}{n} \sum_{i \notin \mathbf{J}_\delta \cap \mathbf{J}_n^X} h(\widehat{\mathbf{U}}_i) \leq r(\delta),$$

where $r(\delta) \rightarrow 0$ as $\delta \rightarrow 0_+$.

Thus we have verified the assumptions of Theorem A.10.2 of Bickel et al. (1993) which yields that there exists a consistent root (say $\hat{\alpha}_n$) of the estimating equation (5) which has the following asymptotic representation

$$\sqrt{n}(\hat{\alpha}_n - \alpha) = \{I(\alpha)\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\psi}(U_i) + o_P(1),$$

where the elements of the vector function $\tilde{\psi}$ are given in (11). Note that completely analogously one can show that there exists a consistent root (say $\tilde{\alpha}_n$) of the estimating equation (6) which has the same asymptotic representation. This finally implies the statement of the theorem. \square

Proof of Theorem 2. The proof is completely analogous to the proof of Theorem 2. The only difference is that one uses Lemma 4 instead of Lemma 3. In fact the proof is even simpler as thanks to assumption **C6** one can take a finite constant instead of the function h . \square

APPENDIX B. SOME RESULTS ON $\hat{F}_{j\hat{\varepsilon}}$ AND \hat{U}_{ji}

In what follows let $x_+ = \max\{x, 0\}$.

Lemma 5. *Suppose that assumptions $(\mathbf{F}_{j\varepsilon})$ and (\mathbf{ms}) hold. Then for $\delta_n = n^{-1/\lambda}$ where $\lambda > 2(1 - \beta + \frac{1}{r-1})$ it holds uniformly in $u \in [\delta_n/2, 1 - \delta_n/2]$*

$$\begin{aligned} \hat{F}_{j\hat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) &= \hat{F}_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) + f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) \mathbf{E}_{\mathbf{X}} \left[\frac{m'_j(\mathbf{X}; \theta_j)}{s_j(\mathbf{X}; \theta_j)} + F_{j\varepsilon}^{-1}(u) \frac{s'_j(\mathbf{X}; \theta_j)}{s_j(\mathbf{X}; \theta_j)} \right]^\top (\hat{\theta}_j - \theta_j) \\ (B1) \quad &+ u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+} o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

for each $\gamma > 0$ and $j \in \{1, \dots, d\}$.

Proof. We will show the statement for $u \in [\frac{\delta_n}{2}, \frac{1}{2}]$. The proof would be completely analogous for $u \in [\frac{1}{2}, 1 - \frac{\delta_n}{2}]$.

Note that

$$\hat{F}_{j\hat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) = \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left\{ \varepsilon_{ji} \leq \frac{m_j(\mathbf{X}_i; \hat{\theta}_j) - m_j(\mathbf{X}_i; \theta_j)}{s_j(\mathbf{X}_i; \theta_j)} + \frac{F_{j\varepsilon}^{-1}(u) s_j(\mathbf{X}_i; \hat{\theta}_j)}{s_j(\mathbf{X}_i; \theta_j)} \right\}.$$

In what follows we need to take care of the fact that the majorant $M_j(\mathbf{x})$ from assumption (\mathbf{ms}) can be unbounded. Let $a_n = n^{1/(\lambda_x r)}$, where λ_x will be specified later. Then

similarly as in the proof of Lemma 1 one can use Markov's inequality to bound

$$\begin{aligned} & \left| \widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) - \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left\{ \varepsilon_{ji} \leq \frac{m_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j) - m_j(\mathbf{X}_i; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}_i; \boldsymbol{\theta}_j)} + \frac{F_{j\varepsilon}^{-1}(u) s_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j)}{s_j(\mathbf{X}_i; \boldsymbol{\theta}_j)}, M_j(\mathbf{X}_i) \leq a_n \right\} \right| \\ & \leq \frac{1}{n} \sum_{i=1}^n \mathbf{1} \{M_j(\mathbf{X}_i) > a_n\} \leq \frac{1}{n} \sum_{i=1}^n \frac{M_j^r(\mathbf{X}_i)}{a_n^r} \mathbf{1} \{M_j(\mathbf{X}_i) > a_n\} = o_P\left(\frac{1}{n^{1/\lambda_x}}\right). \end{aligned}$$

Note that thanks to the assumption $\lambda > 2(1 - \beta + \frac{1}{r-1})$ it is straightforward to verify that $\frac{1}{2} + \frac{\beta}{\lambda} < r(\frac{1}{2} - \frac{1-\beta}{\lambda})$. In the following we will take λ_x such that

$$(B2) \quad \frac{1}{2} + \frac{\beta}{\lambda} < \frac{1}{\lambda_x} < r\left(\frac{1}{2} - \frac{1-\beta}{\lambda}\right).$$

Now with the help of (B2) one can conclude that

$$\begin{aligned} \widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left\{ U_{ji} \leq F_{j\varepsilon} \left(\frac{m_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j) - m_j(\mathbf{X}_i; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}_i; \boldsymbol{\theta}_j)} + \frac{F_{j\varepsilon}^{-1}(u) s_j(\mathbf{X}_i; \widehat{\boldsymbol{\theta}}_j)}{s_j(\mathbf{X}_i; \boldsymbol{\theta}_j)} \right), M_j(\mathbf{X}_i) \leq a_n \right\} \\ (B3) \quad &+ u^{(\beta-\gamma)+} (1-u)^{(\beta-\gamma)+} o_P(n^{-1/2}), \end{aligned}$$

for $u \in [\delta_n/2, 1/2]$.

Now for simplicity of notation introduce

$$(B4) \quad y_{j\mathbf{x}}(\mathbf{t}, u) = \frac{m_j(\mathbf{x}; \mathbf{t}) - m_j(\mathbf{x}; \boldsymbol{\theta}_j)}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} + F_{j\varepsilon}^{-1}(u) \frac{s_j(\mathbf{x}; \mathbf{t})}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)}.$$

Further for $u \in (0, 1]$ and $\mathbf{t} \in \mathbb{R}^{p_j}$ put

$$w(u) = \min\{u, 1-u\}^{(\beta-\gamma)+}, \quad \text{and} \quad \mathbf{t}^{(n)} = \boldsymbol{\theta}_j + \mathbf{t}/n^{1/2-\eta},$$

where $\eta > 0$ is sufficiently small. Note that the function w is increasing on $(0, \frac{1}{2})$ and decreasing on $(\frac{1}{2}, 1)$ for $\beta - \gamma > 0$. Finally let

$$u^{(n)} = \max\{u, \delta_n/2\}$$

and for $i \in \{1, \dots, n\}$ introduce the processes

$$Z_{ni}(\mathbf{t}, u) = \frac{1}{w(u^{(n)})\sqrt{n}} \mathbf{1} \{U_{ji} \leq F_{j\varepsilon}(y_{j\mathbf{x}_i}(\mathbf{t}^{(n)}, u^{(n)})), |M_j(\mathbf{X}_i)| \leq a_n\}$$

that are indexed by the set $\mathcal{F} = T_1 \times (0, 1/2]$, where $T_1 = \{\mathbf{t} \in \mathbb{R}^{p_j} : \|\mathbf{t}\| \leq 1\}$.

Note that assumption **(ms)** guarantees that $n^{1/2-\eta}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) \xrightarrow[n \rightarrow \infty]{P} 0$ for each $\eta \in (0, \frac{1}{2})$, which further implies that $P(\|n^{1/2-\eta}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)\| \leq 1) \xrightarrow[n \rightarrow \infty]{} 1$. Put

$$\widehat{\boldsymbol{\theta}}_n = n^{1/2-\eta}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j).$$

Then with the help of (B3) one can (with probability going to one) write that for $u \in [\delta_n/2, 1/2]$

$$(B5) \quad \widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) = \frac{w(u)}{\sqrt{n}} \sum_{i=1}^n Z_{ni}(\widehat{\boldsymbol{\vartheta}}_n, u) + w(u) o_P(n^{-1/2}).$$

Now equip the space \mathcal{F} with the semimetric ρ given by

$$(B6) \quad \rho((\mathbf{t}_1, u_1), (\mathbf{t}_2, u_2)) = K \sqrt{\|\mathbf{t}_1 - \mathbf{t}_2\| + \frac{u_2 - u_1}{w^2(u_2)} + \left(\frac{1}{w(u_1)} - \frac{1}{w(u_2)}\right)^2 u_1}, \quad \text{for } u_1 \leq u_2,$$

where K is a finite constant that will be specified afterwards.

Later we show that the assumptions of Theorem 2.11.11 of van der Vaart and Wellner (1996) are satisfied for the empirical process indexed by \mathcal{F} , which implies that the process is asymptotically tight. Further as $\sup_{u \in (0, \frac{1}{2}]} \rho((\widehat{\boldsymbol{\vartheta}}_n, u), (\mathbf{0}, u)) = o_P(1)$, one gets that uniformly in $u \in (0, 1/2]$

$$(B7) \quad \sum_{i=1}^n Z_{ni}(\widehat{\boldsymbol{\vartheta}}_n, u) - \sum_{i=1}^n Z_{ni}(\mathbf{0}, u) - \sum_{i=1}^n \mathbb{E}_{U, \mathbf{X}} [Z_{ni}(\widehat{\boldsymbol{\vartheta}}_n, u) - Z_{ni}(\mathbf{0}, u)] = o_P(1),$$

where $\mathbb{E}_{U, \mathbf{X}}$ stands for the expectation with respect to U_{ji} 's and \mathbf{X}_i 's (while considering $\widehat{\boldsymbol{\vartheta}}_n$ being fixed).

In what follows we concentrate on $u \in [\delta_n/2, 1/2]$. If not stated otherwise all the following results hold uniformly for u from this interval.

Note that similarly as in (B5) one can argue that

$$\widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) = \frac{w(u)}{\sqrt{n}} \sum_{i=1}^n Z_{ni}(\mathbf{0}, u) + w(u) o_P\left(\frac{1}{\sqrt{n}}\right).$$

This together with (B3) and (B7) implies

$$(B8) \quad \widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) = \widehat{F}_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) + w(u) \sqrt{n} \mathbb{E}_{U, \mathbf{X}} [Z_{n1}(\widehat{\boldsymbol{\vartheta}}_n, u) - Z_{n1}(\mathbf{0}, u)] + w(u) o_P\left(\frac{1}{\sqrt{n}}\right).$$

Thus to finish the proof it remains to deal with the second term on the right-hand side of (B8). As $\sqrt{n}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) = O_P(1)$ one can use the mean value theorem which guarantees that (with probability going to one) there exists $\mathbf{t}_* \in T_1$ such that

$$\begin{aligned} & w(u) \sqrt{n} \mathbb{E}_{U, \mathbf{X}} [Z_{n1}(\widehat{\boldsymbol{\vartheta}}_n, u) - Z_{n1}(\mathbf{0}, u)] \\ &= \mathbb{E}_{\mathbf{X}} \left[[F_{j\varepsilon}(y_{j\mathbf{X}}(\widehat{\boldsymbol{\theta}}_j, u)) - u] \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \\ (B9) \quad &= \mathbb{E}_{\mathbf{X}} \left[f_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_*^{(n)}, u)) \left(\frac{m'_j(\mathbf{X}; \mathbf{t}_*^{(n)})}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} + F_{j\varepsilon}^{-1}(u) \frac{s'_j(\mathbf{X}; \mathbf{t}_*^{(n)})}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right)^\top \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] (\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j). \end{aligned}$$

Note that for \mathbf{x} such that $M_j(\mathbf{x}) \leq a_n$ one has

$$(B10) \quad \left| \frac{m_j(\mathbf{x}; \mathbf{t}^{(n)}) - m_j(\mathbf{x}; \boldsymbol{\theta}_j)}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} \right| \leq M_j(\mathbf{x}) \|\mathbf{t}^{(n)} - \boldsymbol{\theta}_j\| \leq a_n n^{-1/2+\eta} = n^{1/(\lambda_x r) - 1/2 + \eta}$$

and also

$$(B11) \quad \left| \frac{s_j(\mathbf{x}; \mathbf{t}^{(n)})}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} - 1 \right| \leq M_j(\mathbf{x}) \|\mathbf{t}^{(n)} - \boldsymbol{\theta}_j\| \leq a_n n^{-1/2+\eta} = n^{1/(\lambda_x r) - 1/2 + \eta},$$

where both inequalities hold uniformly in $\mathbf{t} \in T_1$ and $\mathbf{x} \in \{\tilde{\mathbf{x}} : M_j(\tilde{\mathbf{x}}) \leq a_n\}$. Thus with the help of Lemma 11

$$(B12) \quad \sup_{\mathbf{t}_* \in T_1} \sup_{\mathbf{x} \in \{\tilde{\mathbf{x}} : M_j(\tilde{\mathbf{x}}) \leq a_n\}} \sup_{u \in [\delta_n/2, 1/2]} \frac{|f_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}_*, u)) - f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))|}{u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+}} = o_P(1)$$

and also

$$(B13) \quad \sup_{\mathbf{t}_* \in T_1} \sup_{\mathbf{x} \in \{\tilde{\mathbf{x}} : M_j(\tilde{\mathbf{x}}) \leq a_n\}} \sup_{u \in [\delta_n/2, 1/2]} \frac{|f_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}_*, u)) F_{j\varepsilon}^{-1}(u) - f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) F_{j\varepsilon}^{-1}(u)|}{u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+}} = o_P(1).$$

Now combining the above findings with assumption **(ms)** yields that (B9) can be simplified to

$$\begin{aligned} & w(u) \sqrt{n} \mathbf{E}_{U, \mathbf{X}} [Z_{n1}(\hat{\boldsymbol{\theta}}_n, u) - Z_{n1}(\mathbf{0}, u)] \\ &= f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) \mathbf{E}_{\mathbf{X}} \left[\frac{m'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} + F_{j\varepsilon}^{-1}(u) \frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right]^\top (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) + w(u) o_P(n^{-1/2}), \end{aligned}$$

which together with (B8) implies (B1).

Verifying assumptions of Theorem 2.11.11 of van der Vaart and Wellner (1996)

First of all we need to show that the semimetric ρ defined in (B6) is *Gaussian-dominated*. To prove that it is sufficient to show that (see p. 212 of van der Vaart and Wellner, 1996)

$$(B14) \quad \int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}, \rho)} d\epsilon < \infty,$$

where $N(\epsilon, \mathcal{F}, \rho)$ is the covering number of \mathcal{F} .

It is known (see Example 2.11.15 of van der Vaart and Wellner, 1996) that (B14) holds true if \mathcal{F} is replaced with $(0, 1/2]$ and ρ with

$$(B15) \quad \rho_0(u_1, u_2) = \sqrt{\frac{u_2 - u_1}{w^2(u_2)} + \left(\frac{1}{w(u_1)} - \frac{1}{w(u_2)} \right)^2} u_1, \quad \text{for } u_1 \leq u_2,$$

as ρ_0 is Gaussian. But from the definition of ρ in (B6) it follows that one can bound

$$\begin{aligned} N(\epsilon, \mathcal{F}, \rho) &\leq N(\epsilon^2/(4K^2), T_1, \|\cdot\|) N(\epsilon/(2K), (0, 1/2], \rho_0) \\ &= O(\epsilon^{-2p_j}) N(\epsilon/(2K), (0, 1/2], \rho_0), \end{aligned}$$

thus also (\mathcal{F}, ρ) satisfies (B14).

Next we need to check the three assumptions of Theorem 2.11.11 of van der Vaart and Wellner (1996). As in our situations the processes Z_{n1}, \dots, Z_{nn} are identically distributed, the assumptions can be rewritten as follows.

(I) For each $\zeta > 0$

$$(B16) \quad n \mathbb{E} \left[\|Z_{n1}\|_{\mathcal{F}} \mathbf{1}\{\|Z_{n1}\|_{\mathcal{F}} > \zeta\} \right] \xrightarrow{n \rightarrow \infty} 0.$$

(II) For each $(\mathbf{t}_1, u_1), (\mathbf{t}_2, u_2) \in \mathcal{F}$

$$(B17) \quad n \mathbb{E} (Z_{n1}(\mathbf{t}_2, u_2) - Z_{n1}(\mathbf{t}_1, u_1))^2 \leq \rho^2((\mathbf{t}_2, u_2), (\mathbf{t}_1, u_1)).$$

(III) For every ρ -ball $B(\epsilon) \subset \mathcal{F}$ of radius less than ϵ

$$(B18) \quad n \sup_{v>0} v^2 \mathbb{P} \left(\sup_{(\mathbf{t}_1, u_1), (\mathbf{t}_2, u_2) \in B(\epsilon)} |Z_{n1}(\mathbf{t}_2, u_2) - Z_{n1}(\mathbf{t}_1, u_1)| > v \right) \leq \epsilon^2.$$

Note that **the first assumption** (B16) is easy to check as

$$\|Z_{n1}\|_{\mathcal{F}} \leq \sup_{(\mathbf{t}, u) \in \mathcal{F}} |Z_{n1}(\mathbf{t}, u)| \leq \sup_{u \in (0, \frac{1}{2}]} \frac{1}{\sqrt{n} w(u^{(n)})} \leq \frac{1}{\sqrt{n} w(\delta_n/2)} \xrightarrow{n \rightarrow \infty} 0.$$

To verify **the second assumption** (B17) fix $\mathbf{t}_1, \mathbf{t}_2$ and u_1, u_2 (so that $u_1 \leq u_2$) and calculate

$$\begin{aligned} &n \mathbb{E} (Z_{n1}(\mathbf{t}_2, u_2) - Z_{n1}(\mathbf{t}_1, u_1))^2 \\ &= \mathbb{E} \left[\left(\frac{\mathbf{1}\{U_j \leq F_{j\epsilon}(y_j \mathbf{X}(\mathbf{t}_2^{(n)}, u_2^{(n)}))\}}{w(u_2^{(n)})} - \frac{\mathbf{1}\{U_j \leq F_{j\epsilon}(y_j \mathbf{X}(\mathbf{t}_1^{(n)}, u_1^{(n)}))\}}{w(u_1^{(n)})} \right)^2 \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \\ &\leq 2 \mathbb{E} \left[\left(\frac{\mathbf{1}\{U_j \leq F_{j\epsilon}(y_j \mathbf{X}(\mathbf{t}_2^{(n)}, u_2^{(n)}))\}}{w(u_2^{(n)})} - \frac{\mathbf{1}\{U_j \leq F_{j\epsilon}(y_j \mathbf{X}(\mathbf{t}_1^{(n)}, u_1^{(n)}))\}}{w(u_2^{(n)})} \right)^2 \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \\ &\quad + 2 \mathbb{E} \left[\left(\frac{\mathbf{1}\{U_j \leq F_{j\epsilon}(y_j \mathbf{X}(\mathbf{t}_1^{(n)}, u_1^{(n)}))\}}{w(u_2^{(n)})} - \frac{\mathbf{1}\{U_j \leq F_{j\epsilon}(y_j \mathbf{X}(\mathbf{t}_1^{(n)}, u_1^{(n)}))\}}{w(u_1^{(n)})} \right)^2 \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \end{aligned}$$

$$\begin{aligned}
\text{(B19)} \quad &= \frac{2}{w^2(u_2^{(n)})} \mathbb{E} \left[|F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_2^{(n)}, u_2^{(n)})) - F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u_1^{(n)}))| \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \\
&+ 2 \left(\frac{1}{w(u_2^{(n)})} - \frac{1}{w(u_1^{(n)})} \right)^2 \mathbb{E} [F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u_1^{(n)})) \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\}].
\end{aligned}$$

Now we will have a look at *the first term* on the right-hand side of (B19). For a given $u \in [\delta_n/2, 1/2]$ by the mean value theorem there exists \mathbf{t}_* between \mathbf{t}_1 and \mathbf{t}_2 such that

$$\begin{aligned}
\text{(B20)} \quad &\mathbb{E} \left[|F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_2^{(n)}, u)) - F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u))| \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \\
&\leq \mathbb{E} \left[f_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_*^{(n)}, u)) \left(M_j(\mathbf{X}) + |F_{j\varepsilon}^{-1}(u)| M_j(\mathbf{X}) \right) \right] \|\mathbf{t}_1^{(n)} - \mathbf{t}_2^{(n)}\|
\end{aligned}$$

Now with the help of (B4), (B10), (B11) and Lemma 10 one can conclude that with probability going to one

$$\text{(B21)} \quad \sup_{\mathbf{t} \in T_1} \sup_{\mathbf{x} \in \{\tilde{\mathbf{x}}: M_j(\tilde{\mathbf{x}}) \leq a_n\}} y_{j\mathbf{x}}(\mathbf{t}^{(n)}, u) \leq F_{j\varepsilon}^{-1}(2u),$$

which together with (B20) implies that

$$\begin{aligned}
\text{(B22)} \quad &\mathbb{E} \left[|F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_2^{(n)}, u^{(n)})) - F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u^{(n)}))| \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right] \\
&\leq O(n^{-1/2+\eta}) \mathbb{E} M_j(\mathbf{X}) \|\mathbf{t}_1 - \mathbf{t}_2\| O((u^{(n)})^\beta) \leq O(n^{-1/2+\eta}) \|\mathbf{t}_1 - \mathbf{t}_2\| w(u^{(n)})
\end{aligned}$$

uniformly in u .

Now fix \mathbf{t} and \mathbf{x} . Then by the mean value theorem there exists \tilde{u} between $u_1^{(n)}$ and $u_2^{(n)}$ such that

$$\begin{aligned}
\text{(B23)} \quad &|F_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}^{(n)}, u_1^{(n)})) - F_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}^{(n)}, u_2^{(n)}))| \\
&\leq f_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}^{(n)}, \tilde{u})) \frac{s_j(\mathbf{x}; \mathbf{t}^{(n)})}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} \frac{1}{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(\tilde{u}))} |u_1^{(n)} - u_2^{(n)}|,
\end{aligned}$$

which together with

$$\left| \frac{s_j(\mathbf{x}; \mathbf{t}^{(n)})}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} \right| = \left| 1 + \frac{s_j(\mathbf{x}; \mathbf{t}^{(n)}) - s_j(\mathbf{x}; \boldsymbol{\theta}_j)}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} \right| \leq 1 + M_j(\mathbf{x}) \|\mathbf{t}^{(n)} - \boldsymbol{\theta}_j\| \leq 1 + a_n n^{-1/2+\eta},$$

assumption $(\mathbf{F}_{j\varepsilon})$ and (B23) implies that

$$\text{(B24)} \quad |F_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}^{(n)}, u_1^{(n)})) - F_{j\varepsilon}(y_{j\mathbf{x}}(\mathbf{t}^{(n)}, u_2^{(n)}))| \leq O(1) |u_1^{(n)} - u_2^{(n)}|$$

uniformly in \mathbf{t} and \mathbf{x} .

Now combining the inequalities (B21), (B22) and (B24) implies that

$$\text{(B25)} \quad \frac{1}{w^2(u_2^{(n)})} \mathbb{E} \left[|F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_2^{(n)}, u_2^{(n)})) - F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u_1^{(n)}))| \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\} \right]$$

$$\leq \frac{O(n^{-1/2+\eta})\|\mathbf{t}_1 - \mathbf{t}_2\| w(u_2^{(n)})}{w^2(u_2^{(n)})} + \frac{O(1)(u_2^{(n)} - u_1^{(n)})}{w^2(u_2^{(n)})} = O(1) \left(\|\mathbf{t}_1 - \mathbf{t}_2\| + \frac{u_2^{(n)} - u_1^{(n)}}{w^2(u_2^{(n)})} \right).$$

Now turn our attention to *the second term* on the right-hand side of (B19). Analogously as above one can bound

$$\begin{aligned} & \mathbb{E} [F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u_1^{(n)})) \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\}] \\ & \leq \mathbb{E} [|F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_1^{(n)}, u_1^{(n)})) - F_{j\varepsilon}(y_{j\mathbf{X}}(\boldsymbol{\theta}_j, u_1^{(n)}))| \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\}] \\ & \quad + \mathbb{E} [F_{j\varepsilon}(y_{j\mathbf{X}}(\boldsymbol{\theta}_j, u_1^{(n)})) \mathbf{1}\{M_j(\mathbf{X}) \leq a_n\}] \\ & \leq O(1)\|\mathbf{t}_1^{(n)} - \boldsymbol{\theta}_j\| + u_1^{(n)} = O(n^{-1/2+\eta}) + u_1^{(n)} \leq 2u_1^{(n)}. \end{aligned}$$

Combining this with (B19) and (B25) one gets

$$\begin{aligned} n \mathbb{E} (Z_{n1}(\mathbf{t}_2, u_2) - Z_{n1}(\mathbf{t}_1, u_1))^2 & \leq O(1) \left[\|\mathbf{t}_1 - \mathbf{t}_2\| + \frac{u_2^{(n)} - u_1^{(n)}}{w^2(u_2^{(n)})} + \left(\frac{1}{w(u_2^{(n)})} - \frac{1}{w(u_1^{(n)})} \right)^2 u_1^{(n)} \right] \\ & \leq O(1) \left[\|\mathbf{t}_1 - \mathbf{t}_2\| + \rho_0^2(u_1^{(n)}, u_2^{(n)}) \right] \leq O(1) \left[\|\mathbf{t}_1 - \mathbf{t}_2\| + 2\rho_0^2(u_1, u_2) \right], \end{aligned}$$

where the last inequality follows by Lemma 13(iii) in Appendix D.

Finally we show that also **the third assumption** (B18) is satisfied. Let $B(\epsilon)$ be a fixed ϵ -ball. Then from the properties of the Euclidean norm and the function ρ_0 (see Lemma 13(iv) in Appendix D), there exist $\mathbf{t}_0 \in T_1$ and $u_L, u_U \in (0, \frac{1}{2}]$ such that

$$\mathcal{B}(\epsilon) \subset T_\epsilon \times [u_L, u_U], \quad \text{where } T_\epsilon = \{\mathbf{t} : \sqrt{\|\mathbf{t} - \mathbf{t}_0\|} \leq \frac{\epsilon}{K}\} \quad \text{and} \quad \rho_0(u_L, u_U) < \frac{2\epsilon}{K}.$$

Then one can bound

$$\begin{aligned} & n \sup_{v>0} v^2 \mathbb{P} \left(\sup_{(\mathbf{t}_1, u_1), (\mathbf{t}_2, u_2) \in B(\epsilon)} |Z_{ni}(\mathbf{t}_2, u_2) - Z_{ni}(\mathbf{t}_1, u_1)| > v \right) \\ (B26) \quad & \leq 2 \sup_{v>0} v^2 \mathbb{P} \left(\sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \sqrt{n} |Z_{ni}(\mathbf{t}, u) - Z_{ni}(\mathbf{t}_0, u_U)| > v/2 \right) \end{aligned}$$

To deal with the last probability introduce

$$G_{j\mathbf{X}}^{(L)} = \inf_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}^{(n)}, u^{(n)})), \quad G_{j\mathbf{X}}^{(U)} = \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} F_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}^{(n)}, u^{(n)})).$$

Then one can bound

$$\begin{aligned} & \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \sqrt{n} |Z_{ni}(\mathbf{t}, u) - Z_{ni}(\mathbf{t}_0, u_U)| \\ & = \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \left| \frac{\mathbf{1}\{U_{ji} \leq F_{j\varepsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))\}}{w(u^{(n)})} - \frac{\mathbf{1}\{U_{ji} \leq F_{j\varepsilon}(y_{j\mathbf{X}_i}(\mathbf{t}_0^{(n)}, u_U^{(n)}))\}}{w(u_U^{(n)})} \right| \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \end{aligned}$$

$$\begin{aligned}
&\leq \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \left| \frac{\mathbf{1}\{U_{ji} \leq F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))\} - \mathbf{1}\{U_{ji} \leq F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}_0^{(n)}, u_U^{(n)}))\}}{w(u_U^{(n)})} \right| \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \\
&\quad + \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \left| \mathbf{1}\{U_{ji} \leq F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))\} \left(\frac{1}{w(u^{(n)})} - \frac{1}{w(u_U^{(n)})} \right) \right| \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \\
&\quad (B27) \\
&\leq \left[\frac{\mathbf{1}\{G_{j\mathbf{X}_i}^{(L)} \leq U_{ji} \leq G_{j\mathbf{X}_i}^{(U)}\}}{w(u_U^{(n)})} + \mathbf{1}\{U_{ji} \leq u_L\} \left(\frac{1}{w(u_L^{(n)})} - \frac{1}{w(u_U^{(n)})} \right) \right] \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \\
&\quad + \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \mathbf{1}\{u_L \leq U_{ji} \leq F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))\} \left(\frac{1}{w(u^{(n)})} - \frac{1}{w(u_U^{(n)})} \right) \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \\
&= V_{n1} + V_{n2},
\end{aligned}$$

where V_{n1} , V_{n2} stand for the first and second term on the right-hand side of (B27) respectively.

Now similarly as in (B25) one can bound the second moment of V_{n1} as

$$\begin{aligned}
\mathbb{E} V_{n1}^2 &\leq \mathbb{E} \left[\frac{2(G_{j\mathbf{X}_i}^{(U)} - G_{j\mathbf{X}_i}^{(L)})}{w^2(u_U^{(n)})} \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \right] + 2u_L \left(\frac{1}{w(u_L^{(n)})} - \frac{1}{w(u_U^{(n)})} \right)^2 \\
&\leq \frac{O(1)}{w^2(u_U^{(n)})} \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} [\|\mathbf{t} - \mathbf{t}_0\| w(u_U^{(n)}) O(n^{-1/2+\eta}) + |u_U - u_L|] \\
&\quad + 2u_L \left(\frac{1}{w(u_L^{(n)})} - \frac{1}{w(u_U^{(n)})} \right)^2 \\
&= O(1) \left[\frac{\epsilon^2}{K} + \frac{u_U - u_L}{w^2(u_U)} \right] + 2u_L \left(\frac{1}{w(u_L)} - \frac{1}{w(u_U)} \right)^2 = O\left(\frac{\epsilon^2}{K}\right) + O\left(\frac{\rho_0^2(u_L, u_U)}{K}\right) \leq \frac{\epsilon^2}{64},
\end{aligned}$$

provided that K in the definition of the semimetric (B6) is taken sufficiently large.

Thus also by Markov's inequality

$$(B28) \quad \sup_{v>0} v^2 \mathbb{P}(V_{n1} > \frac{v}{4}) \leq \frac{\epsilon^2}{4}.$$

Now we can concentrate on the second term in (B27). To do so note that from the definition of the semimetric ρ_0 in (B15) it follows that for each $u \in [u_L, u_U]$

$$\left(\frac{1}{w(u)} - \frac{1}{w(u_U)} \right)^2 \leq \frac{\rho_0^2(u, u_U)}{u} \leq \frac{4\epsilon^2}{K^2 u},$$

which further implies that

$$\left(\frac{1}{w(u^{(n)})} - \frac{1}{w(u_U^{(n)})} \right) \leq \frac{2\epsilon}{K\sqrt{u}}.$$

Using the above inequality one can bound (with probability going to one)

$$\begin{aligned}
V_{n2} &\leq \frac{2\epsilon \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \mathbf{1}\{u_L \leq U_{ji} \leq F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))\}}{K\sqrt{u^{(n)}}} \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\} \\
&\leq \frac{2\epsilon \sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \mathbf{1}\{u_L \leq U_{ji} \leq F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))\}}{K\sqrt{F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))}} \sqrt{\frac{F_{j\epsilon}(y_{j\mathbf{X}_i}(\mathbf{t}^{(n)}, u^{(n)}))}{u^{(n)}}} \mathbf{1}\{M_j(\mathbf{X}_i) \leq a_n\}
\end{aligned}$$

$$\leq \frac{2\epsilon}{K\sqrt{U_{ji}}} \sqrt{2},$$

where we have used that thanks to (B21)

$$\sqrt{\frac{F_{j\epsilon}(y_j \mathbf{x}_i(\mathbf{t}^{(n)}, u^{(n)}))}{u^{(n)}}} \leq \sqrt{\frac{2u^{(n)}}{u^{(n)}}} \leq \sqrt{2}$$

and for each \mathbf{t}, u

$$\frac{\mathbf{1}\{u_L \leq U_{ji} \leq F_{j\epsilon}(y_j \mathbf{x}_i(\mathbf{t}^{(n)}, u^{(n)}))\}}{\sqrt{F_{j\epsilon}(y_j \mathbf{x}_i(\mathbf{t}^{(n)}, u^{(n)}))}} \leq \frac{\mathbf{1}\{u_L \leq U_{ji} \leq F_{j\epsilon}(y_j \mathbf{x}_i(\mathbf{t}^{(n)}, u^{(n)}))\}}{\sqrt{U_{ji}}} \leq \frac{1}{\sqrt{U_{ji}}}.$$

Thus we can bound

$$(B29) \quad \sup_{v>0} v^2 \mathbf{P}\left(V_{n2} > \frac{v}{4}\right) \leq \sup_{v>0} v^2 \mathbf{P}\left(\frac{2\epsilon\sqrt{2}}{K\sqrt{U_{ji}}} > \frac{v}{4}\right) = \sup_{v>0} v^2 \mathbf{P}\left(U_{ji} < \frac{128\epsilon^2}{K^2 v^2}\right) \leq \frac{\epsilon^2}{4}$$

for a sufficiently large K . Now combining (B28) and (B29) yields that

$$2 \sup_{v>0} v^2 \mathbf{P}\left(\sup_{(\mathbf{t}, u) \in T_\epsilon \times [u_L, u_U]} \sqrt{n} |Z_{ni}(\mathbf{t}, u) - Z_{ni}(\mathbf{t}_0, u_U)| > v/2\right) \leq \epsilon^2,$$

which together with (B26) implies that also (B18) is satisfied. \square

Note that while λ_x is only a cleverly chosen constant in Lemma 5 that is not involved in the statement, in the following lemmas we will speak about J_{jn}^X and thus we need to be more specific about λ_x . Thus in what follows we often assume that

$$(B30) \quad \frac{1}{\lambda_x r} < \frac{1}{2} - \frac{1-\beta}{\lambda}.$$

Lemma 6. *Suppose that the assumptions of Lemma 5 are satisfied and λ_x satisfies (B30). Then it holds uniformly in $k \in J_{jn}^X$*

$$\begin{aligned} \widehat{U}_{jk} - \widetilde{U}_{jk} &= f_{j\epsilon}(\varepsilon_{jk}) \mathbf{E}_{\mathbf{X}} \left[\frac{m'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} + \varepsilon_{jk} \frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right]^\top (\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) + f_{j\epsilon}(\varepsilon_{jk}) (\widehat{\varepsilon}_{jk} - \varepsilon_{jk}) \\ &\quad + U_{jk}^{(\beta-\gamma)+} (1 - U_{jk})^{(\beta-\gamma)+} [M_j(\mathbf{X}_k) + 1] o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

for each $\gamma > 0$ and $j \in \{1, \dots, d\}$.

Proof. The lemma will be shown by substitution of $u = F_{j\epsilon}(\widehat{\varepsilon}_{jk})$ into the approximation (B1) stated in Lemma 5. Note that all the following statements holds uniformly in $k \in J_{jn}^X$.

The proof will be divided into **four steps**. First we show that with probability going to one

$$(B31) \quad F_{j\epsilon}(\widehat{\varepsilon}_{jk}) \in [\delta_n/2, 1 - \delta_n/2]$$

to justify the substitution into (B1). Second

$$(B32) \quad F_{j\varepsilon}(\widehat{\varepsilon}_{jk})^{(\beta-\gamma)+} (1 - F_{j\varepsilon}(\widehat{\varepsilon}_{jk}))^{(\beta-\gamma)+} o_P(1) = U_{jk}^{(\beta-\gamma)+} (1 - U_{jk})^{(\beta-\gamma)+} o_P(1).$$

Next we show that

$$(B33) \quad f_{j\varepsilon}(\widehat{\varepsilon}_{jk}) = f_{j\varepsilon}(\varepsilon_{jk}) + U_{jk}^{(\beta-\gamma)+} (1 - U_{jk})^{(\beta-\gamma)+} o_P(1),$$

$$(B34) \quad f_{j\varepsilon}(\widehat{\varepsilon}_{jk}) \widehat{\varepsilon}_{jk} = f_{j\varepsilon}(\varepsilon_{jk}) F_{j\varepsilon}^{-1}(U_{jk}) + U_{jk}^{(\beta-\gamma)+} (1 - U_{jk})^{(\beta-\gamma)+} o_P(1),$$

and finally we derive

$$(B35) \quad \begin{aligned} \widehat{F}_{j\varepsilon}(\widehat{\varepsilon}_{jk}) &= \widehat{F}_{j\varepsilon}(\varepsilon_{jk}) + f_{j\varepsilon}(\varepsilon_{jk})(\widehat{\varepsilon}_{jk} - \varepsilon_{jk}) \\ &\quad + U_{jk}^{(\beta-\gamma)+} (1 - U_{jk})^{(\beta-\gamma)+} (M_j(\mathbf{X}_k) + 1) o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

and realise that $\widehat{U}_{jk} = \widehat{F}_{j\varepsilon}(\widehat{\varepsilon}_{jk})$ and $\widetilde{U}_{jk} = \widehat{F}_{j\varepsilon}(\varepsilon_{jk})$.

Showing (B31).

Analogously as in (B22) for $k \in J_{jn}^X$

$$\begin{aligned} |F_{j\varepsilon}(\widehat{\varepsilon}_{jk}) - U_{jk}| &= |F_{j\varepsilon}(y_j \mathbf{X}_k(\widehat{\boldsymbol{\theta}}_j, F_{j\varepsilon}^{-1}(U_{jk}))) - F_{j\varepsilon}(y_j \mathbf{X}_k(\boldsymbol{\theta}_j, F_{j\varepsilon}^{-1}(U_{jk})))| \\ &\leq O_P(1) M_j(\mathbf{X}_k) \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\| U_{jk}^\beta (1 - U_{jk})^\beta \\ &\leq O_P(n^{1/(\lambda_x r) - 1/2}) U_{jk}^\beta (1 - U_{jk})^\beta, \end{aligned}$$

This further implies that

$$(B36) \quad \frac{|F_{j\varepsilon}(\widehat{\varepsilon}_{jk}) - U_{jk}|}{U_{jk}(1 - U_{jk})} \leq O_P(n^{1/(\lambda_x r) - 1/2}) \delta_n^{\beta-1} = O_P(n^{1/(\lambda_x r) - 1/2 + (1-\beta)/\lambda}) = o_P(1),$$

where we have used that λ_x satisfies (B30). Thus for a sufficiently large n one gets that

$$F_{j\varepsilon}(\widehat{\varepsilon}_{jk}) \geq \frac{U_{jk}}{2} \geq \frac{\delta_n}{2}$$

and analogously also

$$F_{j\varepsilon}(\widehat{\varepsilon}_{jk}) \leq U_{jk} + \frac{1}{2}(1 - U_{jk}) \leq 1 - \delta_n + \frac{\delta_n}{2} = 1 - \frac{\delta_n}{2}.$$

Showing (B32).

Note that with the help of (B36) one can conclude that

$$F_{j\varepsilon}(\widehat{\varepsilon}_{jk}) \in \left(\frac{1}{2} U_{jk}, \frac{3}{2} U_{jk}\right), \quad \text{and} \quad 1 - F_{j\varepsilon}(\widehat{\varepsilon}_{jk}) \in \left(\frac{1}{2}(1 - U_{jk}), \frac{3}{2}(1 - U_{jk})\right),$$

which implies (B32).

Showing (B33) and (B34). This follows from (B31), (B12) and (B13).

Showing (B35).

Without loss of generality consider only those $k \in J_{jn}^X$ for which $U_{jk} \leq \frac{1}{2}$. Now for $\eta \in (0, \frac{1}{2} - \frac{\beta}{\lambda} - \frac{1}{\lambda_x r})$ introduce

$$r_n = n^{1/2-1/(\lambda_x r)-\eta}.$$

Similarly as in the proof of Lemma 5 define for $i \in \{1, \dots, n\}$ the processes

$$Z_{ni}(\mathbf{t}, u) = \frac{1}{w(u^{(n)})\sqrt{n}} \mathbf{1}\left\{U_{ji} \leq F_{j\varepsilon}\left(F_{j\varepsilon}^{-1}(u^{(n)})\left(1 + \frac{t_1}{r_n}\right) + \frac{t_2}{r_n}\right)\right\}$$

that are indexed by the set $\mathcal{F} = [-1, 1]^2 \times (0, \frac{1}{2}]$. Now one can write $\hat{F}_{j\varepsilon}(\hat{\varepsilon}_{jk})$ as

$$\hat{F}_{j\varepsilon}(\hat{\varepsilon}_{jk}) = \frac{w(\hat{u}^{(n)})}{\sqrt{n}} \sum_{i=1}^n Z_{ni}(\hat{\mathbf{t}}_n, \hat{u}),$$

where

$$\hat{\mathbf{t}}_n = \left(r_n \left[\frac{s_j(\mathbf{X}_k; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}_k; \boldsymbol{\theta}_j)} - 1 \right], \frac{r_n [m_j(\mathbf{X}_k; \boldsymbol{\theta}_j) - m_j(\mathbf{X}_k; \hat{\boldsymbol{\theta}}_j)]}{s_j(\mathbf{X}_k; \boldsymbol{\theta}_j)} \right), \quad \hat{u} = F_{j\varepsilon}(\varepsilon_{jk}).$$

Note that for $k \in J_{jn}^X$

$$(B37) \quad \hat{F}_{j\varepsilon}(\hat{\varepsilon}_{jk}) - \hat{F}_{j\varepsilon}(\varepsilon_{jk}) = \frac{w(\hat{u}_n)}{\sqrt{n}} \sum_{i=1}^n [Z_{ni}(\hat{\mathbf{t}}_n, \hat{u}_n) - Z_{ni}(\mathbf{0}, \hat{u}_n)].$$

Now equip the space \mathcal{F} with the semimetric ρ given by

$$\rho((\mathbf{t}_1, u_1), (\mathbf{t}_2, u_2)) = K \sqrt{\|\mathbf{t}_1 - \mathbf{t}_2\| + \frac{u_2 - u_1}{w^2(u_2)} + \left(\frac{1}{w(u_1)} - \frac{1}{w(u_2)}\right)^2 u_1}, \quad \text{for } u_1 \leq u_2,$$

where K is a sufficiently large but finite constant. Then completely analogously as in the proof of Lemma 5 one can verify the assumptions of Theorem 2.11.11 of van der Vaart and Wellner (1996). Thus $\sup_{u \in (0, \frac{1}{2}]} \rho((\hat{\mathbf{t}}_n, u), (\mathbf{0}, u)) = o_P(1)$, implies that

$$\begin{aligned} \sum_{i=1}^n Z_{ni}(\hat{\mathbf{t}}_n, \hat{u}_n) - Z_{ni}(\mathbf{0}, \hat{u}_n) &= \sum_{i=1}^n \mathbb{E}_{U, \mathbf{X}} [Z_{ni}(\hat{\mathbf{t}}_n, \hat{u}_n) - Z_{ni}(\mathbf{0}, \hat{u}_n)] + o_P(1) \\ &= \frac{\sqrt{n}}{w(U_{jk})} [F_{j\varepsilon}(\hat{\varepsilon}_{jk}) - F_{j\varepsilon}(\varepsilon_{jk})] + o_P(1), \end{aligned}$$

which together with (B37) implies that

$$(B38) \quad \hat{F}_{j\varepsilon}(\hat{\varepsilon}_{jk}) - \hat{F}_{j\varepsilon}(\varepsilon_{jk}) = F_{j\varepsilon}(\hat{\varepsilon}_{jk}) - F_{j\varepsilon}(\varepsilon_{jk}) + w(U_{jk}) o_P\left(\frac{1}{\sqrt{n}}\right).$$

Now the right-hand side of the above equations can be with the help of (B12) and (B13) rewritten as

$$F_{j\varepsilon}(\hat{\varepsilon}_{jk}) - F_{j\varepsilon}(\varepsilon_{jk}) = f_{j\varepsilon}(\varepsilon_{jk})(\hat{\varepsilon}_{jk} - \varepsilon_{jk}) + w(U_{jk}) M_j(\mathbf{X}_k) o_P\left(\frac{1}{\sqrt{n}}\right),$$

which combined with (B38) implies (B35). \square

Lemma 7. *Suppose that the assumptions of Lemma 6 are satisfied and $\beta > 0$. Then for each $\epsilon > 0$ there exists $L_\epsilon > 0$ such that for each $j \in \{1, \dots, d\}$ for all sufficiently large n*

$$\mathbb{P}(\forall_{k \in J_{jn}^X} L_\epsilon U_{jk} \leq \widehat{U}_{jk} \leq 1 - L_\epsilon (1 - U_{jk})) \geq 1 - \epsilon.$$

Proof. We concentrate on the inequality $L_\epsilon U_{jk} \leq \widehat{U}_{jk}$. Showing the upper inequality for \widehat{U}_{jk} would be analogous.

By Lemma 6 one gets $\widehat{U}_{jk} \geq \widetilde{U}_{jk} - |R_{jk}|$, where

$$(B39) \quad \begin{aligned} R_{jk} = & f_{j\epsilon}(\varepsilon_{jk}) \mathbf{E}_{\mathbf{X}} \left[\frac{m'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} + \varepsilon_{jk} \frac{s'_j(\mathbf{X}; \boldsymbol{\theta}_j)}{s_j(\mathbf{X}; \boldsymbol{\theta}_j)} \right]^\top (\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) + f_{j\epsilon}(\varepsilon_{jk}) (\widehat{\varepsilon}_{jk} - \varepsilon_{jk}) \\ & + U_{jk}^{\beta-\gamma} (1 - U_{jk})^{\beta-\gamma} [M_j(\mathbf{X}_k) + 1] o_P\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

and $\gamma > 0$ can be taken arbitrarily small.

Now by Lemma A3 of Shorack (1972) for each $\epsilon > 0$ there exists $\widetilde{L} \in (0, 1)$ such that

$$\mathbb{P}(\forall_{k \in \{1, \dots, n\}} : \widetilde{U}_{jk} \geq \widetilde{L} U_{jk}) \geq 1 - \epsilon/2.$$

Thus one can take $L_\epsilon = \widetilde{L}/2$ provided we show that

$$\mathbb{P}(\forall_{k \in J_{jn}^X} : |R_{jk}| \leq \frac{\widetilde{L} U_{jk}}{2}) \geq 1 - \epsilon/2.$$

To do that one can consider each of the summands on the right-hand side of (B39) separately. Thus for instance one has that uniformly in $k \in J_{jn}^X$

$$\begin{aligned} \left| \frac{f_{j\epsilon}(\varepsilon_{jk})(\widehat{\varepsilon}_{jk} - \varepsilon_{jk})}{U_{jk}} \right| &\leq U_{jk}^{\beta-1} M_j(\mathbf{X}_k) O_P\left(\frac{1}{\sqrt{n}}\right) \leq n^{(1-\beta)/\lambda} a_n O_P\left(\frac{1}{\sqrt{n}}\right) \\ &= O_P\left(n^{(1-\beta)/\lambda+1/(\lambda_x r)-1/2}\right) = o_P(1), \end{aligned}$$

as λ_x satisfies (B30). The other summands on the right-hand side of (B39) can be handled analogously. \square

Some results useful when $(\mathbf{F}_{j\epsilon})$ holds with $\beta = 0$.

Lemma 8. *Suppose that assumptions $(\mathbf{F}_{j\epsilon})$ and (\mathbf{ms}) hold. Then for each $j \in \{1, \dots, d\}$*

$$(B40) \quad \sup_{u \in (0,1)} \sqrt{n} \left| \widehat{F}_{j\epsilon}(F_{j\epsilon}^{-1}(u)) - \widehat{F}_{j\epsilon}(F_{j\epsilon}^{-1}(u)) \right| = O_P(1).$$

Proof. Let $U(\boldsymbol{\theta}_j)$ be the neighborhood of $\boldsymbol{\theta}_j$ introduced in (\mathbf{ms}) . Now consider the set of functions

$$\mathcal{F} = \left\{ (\mathbf{x}, e) \mapsto \mathbf{1}\left\{ e \leq \frac{m_j(\mathbf{x}; \mathbf{t}) - m_j(\mathbf{x}; \boldsymbol{\theta}_j)}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} + F_{j\epsilon}^{-1}(u) \frac{s_j(\mathbf{x}; \mathbf{t})}{s_j(\mathbf{x}; \boldsymbol{\theta}_j)} \right\}; u \in (0, 1), \mathbf{t} \in U(\boldsymbol{\theta}_j) \right\}$$

and denote its elements as $f_{\mathbf{t},u}$. Then one can write

$$\widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) = \frac{1}{n} \sum_{i=1}^n f_{\widehat{\boldsymbol{\theta}}_j,u}(\mathbf{X}_i, \varepsilon_{ji}) \stackrel{\text{Say}}{=} P_n(f_{\widehat{\boldsymbol{\theta}}_j,u}).$$

Similarly as in the proof of Theorem 4 of Gijbels et al. (2015) one can argue that the set \mathcal{F} is P -Donsker. Further similarly as in the proof of Lemma 5 one can show that

$$\sup_{u \in (0,1)} \text{var}_{U,\mathbf{X}}(f_{\widehat{\boldsymbol{\theta}}_j,u} - f_{\boldsymbol{\theta}_j,u}) \leq \mathbf{E}_{\mathbf{X}} |F_{j\varepsilon}(y_{j\mathbf{X}}(\widehat{\boldsymbol{\theta}}_j, u)) - F_{j\varepsilon}(y_{j\mathbf{X}}(\boldsymbol{\theta}_j, u))| = o_P(1),$$

which further implies that uniformly in $u \in (0,1)$

$$(B41) \quad \sqrt{n}[\widehat{F}_{j\widehat{\varepsilon}}(F_{j\varepsilon}^{-1}(u)) - \widehat{F}_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))] = \sqrt{n}[P(f_{\widehat{\boldsymbol{\theta}}_j,u}) - P(f_{\boldsymbol{\theta}_j,u})] + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Now by the mean value theorem there exists \mathbf{t}_* between $\widehat{\boldsymbol{\theta}}_j$ and $\boldsymbol{\theta}_j$ such that

$$\begin{aligned} & \sup_{u \in (0,1)} \left| \sqrt{n}[P(f_{\widehat{\boldsymbol{\theta}}_j,u}) - P(f_{\boldsymbol{\theta}_j,u})] \right| \\ &= \sqrt{n} \sup_{u \in (0,1)} \mathbf{E}_{\mathbf{X}} [|F_{j\varepsilon}(y_{j\mathbf{X}}(\widehat{\boldsymbol{\theta}}_j, u)) - F_{j\varepsilon}(y_{j\mathbf{X}}(\boldsymbol{\theta}_j, u))|] \\ &\leq \sqrt{n} \sup_{u \in (0,1)} \mathbf{E}_{\mathbf{X}} \left[f_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}_*, u)) \left(M_j(\mathbf{X}) + |F_{j\varepsilon}^{-1}(u)| M_j(\mathbf{X}) \right) \right] \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\| \\ &\leq \sqrt{n} \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\| \sup_{u \in (0,1)} \mathbf{E}_{\mathbf{X}} \left[f_{j\varepsilon}(y_{j\mathbf{X}}(\mathbf{t}^*, u)) (1 + |y_{j\mathbf{X}}(\mathbf{t}^*, u)|)^{\frac{M_j(\mathbf{X})(1+|F_{j\varepsilon}^{-1}(u)|)}{1+|y_{j\mathbf{X}}(\mathbf{t}^*, u)|}} \right] \\ &\leq O_P(1) \mathbf{E} M_j(\mathbf{X}) = O_P(1) O(1) = O_P(1), \end{aligned}$$

which together with (B41) implies (B40). \square

Lemma 9. *Suppose that the assumptions of Lemma 8 are satisfied. Then for each $j \in \{1, \dots, d\}$*

$$\max_{k \in \{1, \dots, n\}} \left| \frac{\widehat{U}_{jk} - \widetilde{U}_{jk}}{1 + M_j(\mathbf{X}_k)} \right| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Proof. The proof follows by substitution of $u = F_{j\varepsilon}(\widehat{\varepsilon}_{jk})$ into (B40) and following the proof of Lemma 6. \square

APPENDIX C. FURTHER AUXILIARY RESULTS

Lemma 10. *Suppose that assumption $(\mathbf{F}_{j\varepsilon})$ holds. Let λ satisfy $\lambda > 2(1 - \beta + \frac{1}{r-1})$ and λ_x satisfies (B30). Further for $\eta > 0$ introduce $b_n = n^{\frac{1}{\lambda_x r} - \frac{1}{2} - \eta}$. Then there exists $\eta > 0$ such that for all sufficiently large n for all $u \in [\frac{\delta_n}{2}, \frac{1}{2}]$ for each $j \in \{1, \dots, d\}$*

$$F_{j\varepsilon}^{-1}\left(\frac{u}{2}\right) \leq b_n + [1 + b_n \text{sign}(F_{j\varepsilon}^{-1}(u))] F_{j\varepsilon}^{-1}(u) \leq F_{j\varepsilon}^{-1}(2u),$$

and

$$F_{j\varepsilon}^{-1}(1 - \frac{u}{2}) \geq -b_n + [1 - b_n \operatorname{sign}(F_{j\varepsilon}^{-1}(1 - u))] F_{j\varepsilon}^{-1}(1 - u) \geq F_{j\varepsilon}^{-1}(1 - 2u).$$

Proof. We show only that

$$b_n + [1 + b_n \operatorname{sign}(F_{j\varepsilon}^{-1}(u))] F_{j\varepsilon}^{-1}(u) \leq F_{j\varepsilon}^{-1}(2u),$$

as the remaining inequalities could be proved analogously. Thus we need to show that

$$(C1) \quad b_n + b_n |F_{j\varepsilon}^{-1}(u)| \leq F_{j\varepsilon}^{-1}(2u) - F_{j\varepsilon}^{-1}(u).$$

Now by the mean value theorem

$$F_{j\varepsilon}^{-1}(2u) - F_{j\varepsilon}^{-1}(u) = \frac{u}{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(\tilde{u}))},$$

where \tilde{u} is between u and $2u$. Thus with the help of (C1) it remains to show that

$$\frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(\tilde{u}))(1 + |F_{j\varepsilon}^{-1}(u)|)}{u} \leq \frac{1}{b_n} = n^{\frac{1}{2} - \frac{1}{\lambda_x r} - \eta}.$$

Now by assumption $(\mathbf{F}_{j\varepsilon})$ and using the fact that $u \geq \delta_n$

$$\frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(\tilde{u}))(1 + |F_{j\varepsilon}^{-1}(u)|)}{u} = O(u^{\beta-1}) \leq O(n^{\frac{1-\beta}{\lambda}}) = o(n^{\frac{1}{2} - \frac{1}{\lambda_x r} - \eta}),$$

where we have used that λ_x satisfies (B30), which guarantees that one can find $\eta > 0$ sufficiently small so that $\frac{1}{2} - \frac{1}{\lambda_x r} - \eta > \frac{1-\beta}{\lambda}$ holds. □

Lemma 11. *Suppose that the assumptions of Lemma 10 are satisfied. Then there exists $\eta > 0$ such that for all $\gamma > 0$ for each $j \in \{1, \dots, d\}$*

$$\sup_{s_1, s_2 \in \{-1, 1\}} \sup_{u \in [\frac{\delta_n}{2}, 1 - \frac{\delta_n}{2}]} \left| \frac{f_{j\varepsilon}(s_1 b_n + (1 + s_2 b_n) F_{j\varepsilon}^{-1}(u)) - f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))}{u^{(\beta-\gamma)+} (1-u)^{(\beta-\gamma)+}} \right| = o(1)$$

and also

$$\sup_{s_1, s_2 \in \{-1, 1\}} \sup_{u \in [\frac{\delta_n}{2}, 1 - \frac{\delta_n}{2}]} \left| \frac{[f_{j\varepsilon}(s_1 b_n + (1 + s_2 b_n) F_{j\varepsilon}^{-1}(u)) - f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))] F_{j\varepsilon}^{-1}(u)}{u^{(\beta-\gamma)+} (1-u)^{(\beta-\gamma)+}} \right| = o(1)$$

as $n \rightarrow \infty$.

Proof. We will prove only that

$$\sup_{s_1, s_2 \in \{-1, 1\}} \sup_{u \in [\frac{\delta_n}{2}, \frac{1}{2}]} \left| \frac{[f_{j\varepsilon}(s_1 b_n + (1 + s_2 b_n) F_{j\varepsilon}^{-1}(u)) - f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))] F_{j\varepsilon}^{-1}(u)}{u^{(\beta-\gamma)+} (1-u)^{(\beta-\gamma)+}} \right| = o(1)$$

as the remaining cases can be shown analogously.

First suppose that $\lim_{u \rightarrow 0+} f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) > 0$. Then from Remark 2 one can conclude that $\lim_{u \rightarrow 0+} F_{j\varepsilon}^{-1}(u) > -\infty$ and $\beta = 0$. Thus also $(\beta - \gamma)_+ = 0$ and the statement follows from the continuity of $f_{j\varepsilon}$.

Now suppose that $\lim_{u \rightarrow 0+} f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) = 0$. Note that for a given $u_U \in (0, \frac{1}{2})$

$$\sup_{s_1, s_2 \in \{-1, 1\}} \sup_{u \in [\frac{u_U}{2}, \frac{1}{2}]} \left| \frac{[f_{j\varepsilon}(s_1 b_n + (1 + s_2 b_n) F_{j\varepsilon}^{-1}(u)) - f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))] F_{j\varepsilon}^{-1}(u)}{u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+}} \right| = o(1),$$

which follows from the continuity of the function $f_{j\varepsilon}$.

Now let $\epsilon > 0$ be given and $\gamma > 0$ fixed. Thanks to assumption $(\mathbf{F}_{j\varepsilon})$ one can choose u_U so that

$$\sup_{u \in (0, 2u_U]} \left| \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) F_{j\varepsilon}^{-1}(u)}{u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+}} \right| < \frac{\epsilon}{M},$$

where $M = \sup_{u \in (0, 1/2)} \frac{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(2u))}{f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u))}$. Now thanks to Lemma 10 one can conclude that also

$$\begin{aligned} \sup_{s_1, s_2 \in \{-1, 1\}} \sup_{u \in [\frac{\delta n}{2}, u_U]} \left| \frac{f_{j\varepsilon}(s_1 b_n + (1 + s_2 b_n) F_{j\varepsilon}^{-1}(u)) F_{j\varepsilon}^{-1}(u)}{u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+}} \right| \\ \leq \sup_{u \in (0, 2u_U]} \left| \frac{M f_{j\varepsilon}(F_{j\varepsilon}^{-1}(u)) F_{j\varepsilon}^{-1}(u)}{u^{(\beta-\gamma)_+} (1-u)^{(\beta-\gamma)_+}} \right| < \epsilon, \end{aligned}$$

which finishes the proof of the lemma. \square

Lemma 12. *Suppose that the density $f_{j\varepsilon}$ satisfies assumption $(\mathbf{F}_{j\varepsilon})$. Then*

$$\lim_{|x| \rightarrow \infty} |x| f_{j\varepsilon}(x) = 0.$$

Proof. We will consider only $x \rightarrow \infty$. The remaining case would be handled analogously.

First, note that one can assume that $\lim_{u \rightarrow 1-} F_{j\varepsilon}^{-1}(u) = \infty$, otherwise the proof is trivial. Now suppose that

$$\lim_{x \rightarrow \infty} x f_{j\varepsilon}(x) \neq 0.$$

Then one can find a positive constant a and a sequence $\{z_n\}_{n=1}^{\infty}$ monotonically going to infinity such that

$$z_n f(z_n) \geq a, \quad \forall n \in \mathbb{N}.$$

Note that by assumption $(\mathbf{F}_{j\varepsilon})$ the function $f_{j\varepsilon}(x)$ is non-increasing for $x > F_{j\varepsilon}^{-1}(u_2)$. In what follows we will assume that $z_1 > F_{j\varepsilon}^{-1}(u_2)$ and that $z_{n+1} \geq 2z_n$ (otherwise one can

take an appropriate subsequence of $\{z_n\}$). Now one can bound

$$\begin{aligned} \int_{z_1}^{\infty} f_{j\varepsilon}(x) dx &= \sum_{n=1}^{\infty} \int_{z_n}^{z_{n+1}} \frac{x f_{j\varepsilon}(x)}{x} dx \geq \sum_{n=1}^{\infty} \int_{z_n}^{z_{n+1}} \frac{a}{z_{n+1}} dx \\ &= a \sum_{n=1}^{\infty} \frac{z_{n+1} - z_n}{z_{n+1}} = a \sum_{n=1}^{\infty} \left(1 - \frac{z_n}{z_{n+1}}\right) \geq a \sum_{n=1}^{\infty} \left(1 - \frac{1}{2}\right) = \infty, \end{aligned}$$

which is in contradiction with the fact, that $f_{j\varepsilon}$ is a density. \square

APPENDIX D. SOME PROPERTIES OF ρ_0 FUNCTION

Recall the definition of ρ_0 in (B15) and for simplicity of notation put $b = (\beta - \gamma)_+$. Then for each u_1, u_2 satisfying $0 < u_1 \leq u_2 \leq \frac{1}{2}$ one has $\rho_0(u_1, u_2) = r_0(u_1, u_2)$, where

$$r_0(u_1, u_2) = \sqrt{\frac{u_2 - u_1}{u_2^{2b}} + \left(\frac{1}{u_1^b} - \frac{1}{u_2^b}\right)^2 u_1}.$$

Lemma 13. *Let $u_0 \in (0, \frac{1}{2})$ and $b \in [0, \frac{1}{2})$ be fixed. Then the following statements hold.*

- (i). *The function $g_R(u) = r_0^2(u_0, u)$ is increasing for $u \in (u_0, \frac{1}{2})$.*
- (ii). *For $b > 0$ the function $g_L(u) = r_0^2(u, u_0)$ is increasing on $(0, u_*)$ and decreasing on (u_*, u_0) , where $u_* = u_0 \left(\frac{1-2b}{2(1-b)}\right)^{1/b}$.*
- (iii). *For each $0 \leq u_1 < u_2 < u_0 \leq \frac{1}{2}$ it holds that $r_0^2(u_2, u_0) \leq 2 r_0^2(u_1, u_0)$.*
- (iv). *For each $\epsilon > 0$ the set $U(u_0, \epsilon) = \{u \in [0, \frac{1}{2}] : \rho_0(u, u_0) \leq \epsilon\}$ is contained in a set $[u_L, u_U]$ such that $r_0(u_L, u_U) \leq 2\epsilon$.*

Proof. The proof of (i) follows directly from the definition of the function g , as

$$g_R(u) = r_0^2(u_0, u) = \frac{u - u_0}{u^{2b}} + \left(\frac{1}{u_0^b} - \frac{1}{u^b}\right)^2 u_0 = u^{1-2b} + u_0^{1-2b} - \frac{2u_0^{1-b}}{u^b},$$

which is evidently an increasing function on $(u_0, \frac{1}{2}]$.

For the proof of (ii) rewrite

$$g_L(u) = r_0^2(u, u_0) = \frac{u_0 - u}{u_0^{2b}} + \left(\frac{1}{u^b} - \frac{1}{u_0^b}\right)^2 u = u_0^{1-2b} + u^{1-2b} - \frac{2u^{1-b}}{u_0^b}.$$

Now it is straightforward to find that the function g_L has exactly one local maximum in the point u_* and meets the claimed properties.

Now we show (iii). Note that thanks to (ii) the function $g_L(u)$ is decreasing on (u_*, u_0) , thus the inequality trivially holds if $u_* \leq u_1 < u_2$.

Thus suppose that $u_1 < u_*$. From (ii) we further know that $u_* = u_0 a$, where $a < 1$. Thus we can bound

$$\begin{aligned} r_0^2(u_2, u_0) &\leq r_0^2(u_*, u_0) = u_0^{1-2b} [1 - a + (1 - a^b)^2 a^{1-2b}] \\ &\leq 2 u_0^{1-2b} = 2 r_0^2(0, u_0) \leq 2 r_0^2(u_1, u_0), \end{aligned}$$

which was to be proved.

To prove (iv) first note that from (i) there exists u_U such that

$$\{u \in [u_0, \tfrac{1}{2}] : \rho_0(u, u_0) \leq \epsilon\} = [u_0, u_U] \quad \text{and} \quad \rho_0(u, u_U) \leq \epsilon.$$

When searching for u_L one has to be more careful as the function g_L is not decreasing on $(0, u_0)$. We need to distinguish two cases. First, let $\epsilon < r_0(0, u_0)$. Then one can find u_L in a similar way as u_U was found. Second, suppose that $\epsilon \geq r_0(0, u_0)$. Then we take simply $u_L = 0$.

Now it remains to check that $r_0(u_L, u_U) \leq 2\epsilon$. To do that bound

$$\begin{aligned} r_0^2(u_L, u_U) &= \frac{u_U - u_L}{u_U^{2b}} + \left(\frac{1}{u_L^b} - \frac{1}{u_U^b}\right)^2 u_L \\ &\leq \frac{u_U - u_0}{u_U^{2b}} + \frac{u_0 - u_L}{u_0^{2b}} + 2\left(\frac{1}{u_L^b} - \frac{1}{u_0^b}\right)^2 u_L + 2\left(\frac{1}{u_0^b} - \frac{1}{u_U^b}\right)^2 u_0 \\ &\leq 2 r_0^2(u_L, u_0) + 2 r_0^2(u_0, u_U) \leq 4\epsilon^2. \end{aligned}$$

□

REFERENCES

- Anděl, J. (1989). Non-negative autoregressive processes. *J. Time Series Anal.*, 10(1):1–11.
- Anděl, J. (1992). Nonnegative multivariate AR(1) processes. *Kybernetika*, 28(3):213–226.
- Berghaus, B., Bücher, A., and Volgushev, S. (2017). Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli*, 23(1):743–772.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.
- Brahimi, B. and Necir, A. (2012). A semiparametric estimation of copula models based on the method of moments. *Stat. Methodol.*, 9(4):467–477.
- Bücher, A., Jäschke, S., and Wied, D. (2015). Nonparametric tests for constant tail dependence with an application to energy and finance. *J. Econometrics*, 187(1):154–168.
- Chan, N.-H., Chen, J., Chen, X., Fan, Y., and Peng, L. (2009). Statistical inference for multivariate residual copula of GARCH models. *Statist. Sinica*, 19:53–70.

- Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *J. Econometrics*, 135:125–154.
- Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models. *J. Econometrics*, 130(2):307–335.
- Côté, M.-P., Genest, C., and Omelka, M. (2019). Rank-based inference tools for copula regression, with property and casualty insurance applications. *Insurance Math. Econom.* Under revision.
- Davis, R. A. and McCormick, W. P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Process. Appl.*, 31(2):237–250.
- Einmahl, J. H. and Van Keilegom, I. (2008). Specification tests in nonparametric regression. *J. Econometrics*, 143(1):88–102.
- Fermanian, J.-D. and Lopez, O. (2018). Single-index copulas. *J. Multivariate Anal.*, 165:27–55.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.
- Gijbels, I., Omelka, M., Pešta, M., and Veraverbeke, N. (2017). Score tests for covariate effects in conditional copulas. *J. Multivariate Anal.*, 159:111–133.
- Gijbels, I., Omelka, M., and Veraverbeke, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scand. J. Statist.*, 42:1109–1126.
- Lawrance, A. and Lewis, P. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 47(2):165–202.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance, Princeton.
- Neumeyer, N., Omelka, M., and Hudecová, Š. (2019). A copula approach for dependence modeling in multivariate nonparametric time series. *J. Multivariate Anal.*, 171:139–162.
- Nielsen, B. and Shephard, N. (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations. *J. Time Series Anal.*, 24(3):337–344.
- Portier, F. and Segers, J. (2018). On the weak convergence of the empirical conditional copula under a simplifying assumption. *J. Multivariate Anal.*, 166:160–181.

- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Radulović, D., Wegkamp, M., and Zhao, Y. (2017). Weak convergence of empirical copula processes indexed by functions. *Bernoulli*, 23(4B):3346–3384.
- Schechtman, E. and Schechtman, G. (1986). Estimating the parameters in regression with uniformly distributed errors. *J. Comput. Graph. Stat.*, 26(3-4):269–281.
- Shorack, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.*, 43:412–427.
- Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, 33:357–375.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.