PARAMETRIC COPULA ADJUSTED FOR NON- AND SEMI-PARAMETRIC REGRESSION

By Yue Zhao[1], Irène Gijbels[2,†] and Ingrid Van Keilegom[1,*]

[1]Research Centre for Operations Research and Statistics (ORSTAT)
KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
yue.zhao@york.ac.uk; [*]ingrid.vankeilegom@kuleuven.be

[2]Department of Mathematics and Leuven Statistics Research Center (LStat)
KU Leuven, Celestijnenlaan 200B, B-3001 Leuven (Heverlee), Belgium
[†]irene.gijbels@kuleuven.be

We consider a multivariate response regression model where each coordinate is described by a location-scale non- or semi-parametric regression, and where the dependence structure of the "noise term" is described by a parametric copula. Our goal is to estimate the associated Euclidean copula parameter given a sample of the response and the covariate. In the absence of the copula sample, the usual oracle ranks are no longer computable. Instead, we study the normal scores estimator for the Gaussian copula, and generalized pseudo-likelihood estimation for general parametric copulas, both based on residual ranks calculated from preliminary non- or semi-parametric estimators of the location and scale functions. We show that the residual-based estimators are asymptotically equivalent to their oracle counterparts, and provide explicit rate of convergence. Partially to serve this objective, we also study weighted convergence of the residual empirical process under the non- or semi-parametric regression model.

1. Introduction.

Let $\mathbf{E} = (E_1, \ldots, E_p)^\top \in \mathbb{R}^p$ be a random vector; we assume throughout that E_k, $k \in [p] \equiv \{1, \ldots, p\}$ has absolutely continuous marginal distribution function F_k, and $\mathbf{E}$ has joint distribution function F. We consider a multiple-response regression model where a $p \times 1$ response vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ and a $q \times 1$ covariate vector $\mathbf{X} = (X_1, \ldots, X_q)^\top$ are linked to $\mathbf{E}$ through a coordinate-by-coordinate (location-scale, as will always be assumed) regression model

$$Y_k = m_k(\mathbf{X}) + \sigma_k(\mathbf{X}) E_k, \quad \forall k \in [p]. \tag{1}$$

We assume throughout that $\mathbf{X}$ is independent of $\mathbf{E}$. In its raw form, model (1) is a purely nonparametric regression model; by specifying particular forms of m_k (and at times simply setting $\sigma_k = 1$), model (1) also accommodates a wide range of popular non- and semi-parametric regression variants such as the partly linear regression model and the additive model. For identification purpose, we assume $\mathbb{E}\mathbf{E} = \mathbf{0}$; if the function σ_k, $k \in [p]$ is not assumed to be a known constant function, then we further assume $\mathrm{Var}(E_k) = 1$. Under model (1), we observe a sample of $(\mathbf{X}, \mathbf{Y})$, but not the sample of $\mathbf{E}$.

Given model (1) for a single $k \in [p]$, there are in general two venues of research. In the first venue, the interest is in the estimation of m_k and σ_k, utilizing as much assumed structure of m_k and σ_k as possible but treating E_k as a noise term. In the second venue, the distribution of E_k is the object of interest instead while m_k and σ_k are treated as nuisance parameters; this approach belongs to the literature on residual empirical processes.

MSC2020 subject classifications: Primary 62H20; secondary 62G08, 62G20, 62G30.

Keywords and phrases: Non- and semi-parametric regression, normal scores rank correlation estimator, parametric copula, pseudo-likelihood, residual empirical processes.

In this paper we follow the second approach in which (the distribution of) $\mathbf{E}$ is our interest, and face the natural challenge that a sample of $\mathbf{E}$ is not observed. However, instead of the individual marginals E_k, $k \in [p]$, here we are more interested in the multivariate dependence structure of $\mathbf{E}$. At first, we will assume a (semi-parametric) *Gaussian copula* model on $\mathbf{E}$ and estimate the intrinsic dependence structure of $\mathbf{E}$ described by a *copula correlation matrix* $\mathbf{R}_0$ via the normal scores estimator. Later, we broaden our study to the case when the dependence structure of $\mathbf{E}$ is described by a general parametric copula at the Euclidean *copula parameter* $\boldsymbol{\theta}_0$. We will estimate $\boldsymbol{\theta}_0$, our object of interest, by generalized pseudo-likelihood estimation (PLE). In this paper we follow the convention that copula parameters with subscript zero, for example $\mathbf{R}_0$ and $\boldsymbol{\theta}_0$, denote fixed, population quantities, while the un-subscripted versions denote their variable counterparts in a parametrization. In fact PLE for the Gaussian copula yields the normal scores estimator as the closed-form solution. However such a clean-cut solution, which results in more straightforward analysis, are usually not available for general parametric copulas. Thus, separate treatments for the Gaussian copula and the general case are natural. To estimate $\mathbf{R}_0$ or the more general copula parameter $\boldsymbol{\theta}_0$ under model (1), we will rely on the ranks of estimators $\widehat{E}_{i,k}$ defined in Section 3.1 of the sample of $\mathbf{E}$ based on some preliminary estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ of m_k and σ_k, or simply *residual ranks*, from which we construct residual (rank)-based estimators of the copula parameter.

To distinguish from model (1), we henceforth refer to the situation when the sample of $\mathbf{E}$ is directly observable as the *oracle* model, and use the qualifier "oracle" to denote quantities that could be computed in this setting. Recently, Zhao, Gijbels and Van Keilegom (2020) carried out the task of estimating $\mathbf{R}_0$ under Gaussian copula while restricting model (1) to a homoscedastic linear regression model. There the function $\sigma_k = 1$ identically and m_k admits the simple form $m_k(\mathbf{X}) = \mathbf{B}_k^\top \mathbf{X}$ with $\mathbf{B}_k$ being the kth column of a $q \times p$ unknown coefficient matrix $\mathbf{B}$. These authors showed that the residual-based normal scores estimator (see Eq. (17)) and Spearman's rho achieve the same asymptotic distribution as their oracle counterparts, even when the convergence rate of the estimator $\widehat{\mathbf{B}}$ to $\mathbf{B}$ is almost as slow as $n^{-1/4}$, and provided explicit rates of convergence. Omelka, Hudecová and Neumeyer (2020) studied model (1), but with parametrically specified m_k and σ_k. See also Veraverbeke, Omelka and Gijbels (2011); Veraverbeke, Gijbels and Omelka (2014); Gijbels, Omelka and Veraverbeke (2015) for related studies on *empirical copula process* involving a covariate.

The main contribution of our present paper is two-fold. First, we will study the estimation of copula parameters in the Gaussian and more general parametric copulas specifically under regression model (1) in its general, non- or semi-parametric form with arbitrary functions m_k and σ_k, and not just under linear regression or parametrically specified m_k and σ_k as previously done. In particular we study the residual-based normal scores estimator (for Gaussian copula) and PLE (for general parametric copulas), and handle the complications introduced by the non- or semi-parametric estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ under unbounded *score functions* (as an example, for Gaussian copula, the score function involves $\Phi^{\leftarrow}$, the standard normal quantile function). Our conclusion is that the residual-based normal scores estimator and the PLE under model (1) reach the same asymptotic distribution as their oracle counterpart under mild conditions. We provide explicit rates of convergence, and allow the dimension of the parametric component in a semi-parametric regression model to vary with n. We apply our general result to a number of popular non- and semi-parametric regression models.

Second, and partially to serve the first point above, we study residual empirical processes under model (1), in particular providing explicit and weighted rates for the remainder terms in the "asymptotic" expansion. Both the rates and the weighing feature are important for taming the unboundedness of the score functions when analyzing the residual-based estimators for copula parameters. On the other hand, residual empirical processes concern the individual marginals of $\mathbf{E}$ and hence do not rely on the copula dependence structure. Thus this aspect of our study can be more directly seen as the continuation of and improvement over related works in residual empirical processes under (1) along the lines of Akritas and Van Keilegom (2001); Neumeyer and Van Keilegom (2010) and Müller, Schick and Wefelmeyer (2009, 2012), among others. In particular we allow the dimension of the parametric component in a semi-parametric regression model to vary with n, as was done under purely parametric, linear regression in, e.g., Mammen (1996); Chen and Lockhart (2001) but to the best of our knowledge not under semi-parametric regression.

Having commented on our contribution to residual empirical processes above, we mention that Chen and Fan (2006); Neumeyer, Omelka and Šárka Hudecová (2019); Chen, Huang and Yi (2021) studied models similar to (1) but in the time series framework, with i.i.d. innovations but time-dependent covariate. Their results for the residual-based estimators of the copula parameter are weaker than ours, for instance, Neumeyer, Omelka and Šárka Hudecová (2019) and the two-stage procedure in (Chen, Huang and Yi, 2021, Section 4) cannot handle unbounded score functions (see the paragraph following Eq. (8) in Neumeyer, Omelka and Šárka Hudecová (2019)). This is partially due to their weaker results on residual empirical processes, for instance see Theorem 4 in Chen, Huang and Yi (2021). However, in their time series framework many i.i.d. tools we employ are potentially not available, and thus we refrain from a finer comparison of our manuscript with these papers.

Before getting into further details, as suggested by one referee we provide a short roadmap of our analysis. We note that under model (1) the residual estimation of the individual marginal distributions is often merely $\sqrt{n}$-consistent. The slow rate is due to the leading term (namely, the term proportional to f_k in Proposition 3.1) in a decomposition of the residual empirical processes associated with *non*- or *semi*-parametric estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$. The slow rate of the leading term becomes even worse after passing from residual distributions to the deviations between residual ranks and oracle ranks (again see the term proportional to f_k now in Eq. (8) in Proposition 3.2). However, for the latter situation the aforementioned leading term is now centered (due to subtractions by $\bar{\delta}_{n,k,\sigma}$ and $\bar{\delta}_{n,k,m}$). Thus in our eventual analysis of residual-based estimators of copula parameters, because the leading term will be summed over the sample, the resulting average (albeit still indexed by random estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$) benefits from an additional $n^{-1/2}$-scaling and so converges faster than $n^{-1/2}$. This coupled with faster rates of the remaining remainder terms in the decomposition in the residual empirical processes/residual ranks eventually lead to the equivalence between the residual-based estimators of copula parameters and their oracle counterparts.

The outline of our paper is as follows. In Section 2 we review necessary background on copula and formally introduce our model associated with (1). In Section 3 we study residual empirical processes under model (1); this section concerns the individual marginals of $\mathbf{E}$ and does not rely on the copula dependence structure. In Section 4, under the Gaussian copula, we prove the asymptotic equivalence of the residual-based normal scores estimator to its oracle counterpart. In Section 5 we present the analogous result for the generalized pseudo-likelihood estimator under general parametric copulas. In Section 6 we apply our results to several popular non- and semi-parametric regression models. We also carry out a small numerical study, with simulation studies presented in Section 7 and with a real data example deferred to Section F in the supplement. We defer all proofs to Sections A to E in the supplement.

2. Background on copula, formal model setup, and notations.

Recall that $\mathbf{E} \in \mathbb{R}^p$ has joint distribution function F and absolutely continuous marginal distribution functions F_k, $k \in [p]$. Sklar's theorem (e.g., Sklar (1959), or Corollary 2.10.10 in Nelsen (2006)) states that the dependence structure of $\mathbf{E}$ can be uniquely described by its associated copula $C : [0,1]^p \to [0,1]$, which satisfies $C(\mathbf{u}) = F(F_1^{\leftarrow}(u_1), \ldots, F_p^{\leftarrow}(u_p))$, for

$\mathbf{u} = (u_1, \dots, u_p)^\top \in [0,1]^p$. Here, for $k \in [p]$, $F_k^{\leftarrow}(t) = \inf\{x : F_k(x) \geq t\}$ denotes the left-continuous inverse of F_k for $t \in [0,1]$. The copula C is equivalently the joint distribution function of the transformed random vector $(F_1(E_1), \dots, F_p(E_p))^\top$, which clearly has uniform marginals on the unit interval. Moreover, the copula C remains unchanged if (univariate) strictly increasing transformations are applied to the individual marginals of $\mathbf{E}$. As such, copulas decouple the dependence structure of a multivariate distribution from the behaviors of its marginals, and thus present a modular approach to multivariate modeling.

Now we consider a collection of random vectors $\mathbf{E}$ on $\mathbb{R}^p$. When the copulas of $\mathbf{E}$ within the collection are not parametrically specified, and the marginals $F_1, \dots, F_p$ of $\mathbf{E}$ can range over all p-tuples of absolutely continuous univariate distribution functions, we say that the collection constitutes a non-parametric copula model. If, in addition, we restrict the copulas of $\mathbf{E}$ within the collection to be smoothly parametrized by an Euclidean copula parameter (and the copula parameter can typically vary over some set), we call the collection a semi-parametric copula model. We will exclusively focus on such a semi-parametric copula model of (that is, a collection of) $\mathbf{E}$ in which the copulas of $\mathbf{E}$ are always parametric while the marginals of $\mathbf{E}$ are modelled non-parametrically. For brevity we will not often distinguish between the *semi-parametric* copula model and the underlying *parametric* copula, and will oftentimes ignore displaying the qualifier "semi-parametric".

Arguably the most popular oracle estimators for copulas are *rank-based* because they, just as the copula at the population level, are invariant to strictly increasing marginal transformations. The estimators based on residual ranks are not strictly invariant under such transformations, due to the perturbation by the covariate, but as stated are asymptotically indistinguishable from their oracle counterparts. This paper focuses exclusively on rank-based estimators.

Now we formally set up our model associated with (1). We say that $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$ follows a joint distribution $\mathrm{P} = \mathrm{P}_{C, F_1, \dots, F_p, m_1, \dots, m_p, \sigma_1, \dots, \sigma_p, F_{\mathbf{X}}}$ if the following conditions hold:

(i) Model (1) holds. For simplicity we assume that all coordinates $k \in [p]$ follow the same general non- or semi-parametric regression model, but the functions m_k, σ_k within the model could be different across $k \in [p]$. The covariate $\mathbf{X}$ has distribution function $F_{\mathbf{X}}$ and support $\mathcal{X} \subset \mathbb{R}^q$.

(ii) $\mathbf{E} = (E_1, \dots, E_p)^\top$ has copula C.

(iii) For each $k \in [p]$, E_k has absolutely continuous marginal distribution function F_k with corresponding marginal density function f_k.

(iv) $\mathbf{X}$ and $\mathbf{E}$ are independent; for identification, $\mathbb{E}\mathbf{E} = \mathbf{0}$ and if σ_k, $k \in [p]$ is not assumed to be a known constant function, then moreover $\mathrm{Var}(E_k) = 1$.

Throughout the remainder of the paper we will assume that the law P holds, but depending on the context the copula C will admit different parametrizations (Gaussian copula in Section 4, general parametric copula in Section 5). The elements of P involving the covariate $\mathbf{X}$ (namely $F_{\mathbf{X}}$, the m_k's and σ_k's) may vary implicitly with sample size, which will allow the dimension of the parametric component in a semi-parametric regression model to vary as well; however C, $F_1, \dots, F_p$ will remain fixed.

For notations, let M, sometimes with superscript, denote an absolute constant that may change for each occurrence; such a constant may depend on various parameters we consider (e.g., fixed Lipschitz constant) but never on n or any parameter that may depend on n. Let $\lesssim$ denote an inequality that holds with such a constant M as the multiplicative factor. Let $\|\cdot\|$ denote the Euclidean norm and $\|\cdot\|_\infty$ the supremum norm of the argument. All convergences are taken along the limit $n \to \infty$. Let $\wedge/\vee$ with two vectors of the same dimension on the two sides return the minimal/maximum values of the two sides component-wise.

3. Results for residual empirical processes and residual ranks.

3.1. *Preliminaries and assumptions.* Recall that results in this section do not depend on the copula dependence structure, that is (ii) in Section 2, and moreover q is allowed to vary with n. We let $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{E}_i)$, $i \geq 1$ be independent copies of $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$, with $\mathbf{E}_i = (E_{i,1}, \ldots, E_{i,p})^\top$, $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$ and $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,q})^\top$.

Under the oracle model, our (observed) sample of size $n \geq 1$ consists of $\mathbf{E}_i$, $i \in [n]$. Then, for each $k \in [p]$ we define the empirical marginal distribution function for the kth coordinate of $\mathbf{E}$, and its rescaled version, as $F_{n,k}(t) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{E_{i,k} \leq t\}$ and $F^{\mathrm{r}}_{n,k}(t) = \frac{n}{n+1} F_{n,k}(t)$, $t \in \mathbb{R}$ respectively. Note that the rescaled $F^{\mathrm{r}}_{n,k}$, when supplied with the $E_{i,k}$'s, $i \in [n]$ as arguments as we will do shortly, takes values in the interval $[1/(n+1), n/(n+1)]$ and so stays away from the boundary points 0 and 1. Thus supplying such values further as arguments to an unbounded score function that diverges toward the boundary points, for instance a function involving $\Phi^{\leftarrow}$, always results in finite values.

Now we turn to model (1). Here we no longer have access to the sample of the copula component $\mathbf{E}$, and hence the *oracle ranks* $F^{\mathrm{r}}_{n,k}(E_{i,k})$. Instead, for sample size $n \geq 1$, our sample consists of $(\mathbf{Y}_i, \mathbf{X}_i)$, $i \in [n]$. Therefore, we rely on this sample to construct estimators of the oracle ranks.

Recall that $\widehat{m}_k$ and $\widehat{\sigma}_k$, which we assume throughout are constructed from $(\mathbf{Y}_i, \mathbf{X}_i)$, $i \in [n]$, are estimators of m_k and σ_k respectively. For $i \in [n]$, let $\widehat{\mathbf{E}}_i = (\widehat{E}_{i,1}, \ldots, \widehat{E}_{i,p})^\top$ with $\widehat{E}_{i,k} = \{Y_{i,k} - \widehat{m}_k(\mathbf{X}_i)\} / \widehat{\sigma}_k(\mathbf{X}_i)$ be the residual of the ith sample point; $\widehat{\mathbf{E}}_i$ serves to estimate $\mathbf{E}_i$. Then, for each $k \in [p]$, from $\{\widehat{E}_{1,k}, \ldots, \widehat{E}_{n,k}\}$ we construct the residual (empirical marginal) distribution function for the kth coordinate of $\mathbf{E}$, and its rescaled version, as

$$\widehat{F}_{n,k}(t) = \tfrac{1}{n} \textstyle\sum_{i \in [n]} \mathbb{1}\{\widehat{E}_{i,k} \leq t\}, \quad \widehat{F}^{\mathrm{r}}_{n,k}(t) = \tfrac{n}{n+1} \widehat{F}_{n,k}(t), \quad t \in \mathbb{R}$$

respectively. The functions $\widehat{F}_{n,k}$ and $\widehat{F}^{\mathrm{r}}_{n,k}$ serve as the estimators of $F_{n,k}$ and $F^{\mathrm{r}}_{n,k}$ respectively. The estimators $\widehat{\mathbf{E}}_i$, $\widehat{F}_{n,k}$ and $\widehat{F}^{\mathrm{r}}_{n,k}$ in turn give rise to $\widehat{F}_{n,k}(\widehat{E}_{i,k})$ and $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$ which we call the residual ranks and which serve to approximate the oracle ones.

Let $\mathcal{T}_n$ denote the σ-field generated by the collection of random vectors $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i \in [n]}$, and $\mathbb{E}[f(\mathbf{E}, \mathbf{X}) | \mathcal{T}_n]$ for a random function $f = f(\mathbf{E}, \mathbf{X})$ be the conditional expectation given $\mathcal{T}_n$, which is an expectation over $\mathbf{E}$ and $\mathbf{X}$ only but not $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i \in [n]}$. Under model (1), introduce the following quantities:

$$\begin{aligned} \delta_{n,k,m}(\mathbf{x}) &= \{(\widehat{m}_k - m_k)/\sigma_k\}(\mathbf{x}), \quad \delta_{n,k,\sigma}(\mathbf{x}) = \{(\widehat{\sigma}_k - \sigma_k)/\sigma_k\}(\mathbf{x}), \\ \delta_{n,k,m,i} &= \delta_{n,k,m}(\mathbf{X}_i), \quad \delta_{n,k,\sigma,i} = \delta_{n,k,\sigma}(\mathbf{X}_i), \\ \bar{\delta}_{n,k,m} &= \mathbb{E}\left[\delta_{n,k,m}(\mathbf{X}) | \mathcal{T}_n\right], \quad \bar{\delta}_{n,k,\sigma} = \mathbb{E}\left[\delta_{n,k,\sigma}(\mathbf{X}) | \mathcal{T}_n\right]. \end{aligned} \tag{2}$$

Here $\bar{\delta}_{n,k,m}$ and $\bar{\delta}_{n,k,\sigma}$ are expectations over $\mathbf{X}$ only while holding $\widehat{m}_k$ and $\widehat{\sigma}_k$ fixed.

When analyzing residual empirical processes, a crucial "oscillation-like" remainder (function) term (e.g., Lemma A.3 in Neumeyer and Van Keilegom (2010)) is defined as, for $t \in \mathbb{R}$,

$$\begin{aligned} r_{1n,k}(t) &= \widehat{F}_{n,k}(t) - F_{n,k}(t) - \mathbb{P}(\widehat{E}_k \leq t | \mathcal{T}_n) + F_k(t) \\ &= \tfrac{1}{n} \textstyle\sum_{i \in [n]} \left\{ \mathbb{1}\{\widehat{E}_{i,k} \leq t\} - \mathbb{1}\left\{E_{i,k} \leq t\right\} - \mathbb{P}(\widehat{E}_k \leq t | \mathcal{T}_n) + \mathbb{P}(E_k \leq t) \right\}. \end{aligned} \tag{3}$$

Here, for $k \in [p]$, in the conditional probability $\mathbb{P}(\widehat{E}_k \leq \cdot | \mathcal{T}_n) \equiv \mathbb{E}[\mathbb{1}\{\widehat{E}_k \leq \cdot\} | \mathcal{T}_n]$ the quantity $\widehat{E}_k$ is defined as $\widehat{E}_k = \{Y_k - \widehat{m}_k(\mathbf{X})\} / \widehat{\sigma}_k(\mathbf{X})$. We let $r^{\mathrm{r}}_{1n,k}$ be defined analogous to (3) but with $\widehat{F}_{n,k}$ replaced by $\widehat{F}^{\mathrm{r}}_{n,k}$; $r^{\mathrm{r}}_{1n,k}$ is useful in the context of unbounded score functions.

Next, in Section 3.2 we first present some general results on residual empirical processes. A specific result on the oscillation terms $r_{1n,k}$ and $r^{\mathrm{r}}_{1n,k}$ is deferred to the dedicated Section 3.3 that also contains a discussion of the results. Necessary assumptions for our analysis are collected below.

ASSUMPTION 3.1. *There exist sets $\mathcal{X}_n \subset \mathcal{X}$, $n \geq 1$ such that $\mathbb{P}(\cap_{i\in[n]}\{\mathbf{X}_i \in \mathcal{X}_n\}) \to 1$ and, for some $a_{n,1} = o(1)$ and $a_{n,2} = o(1)$, the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ satisfy $\|\delta_{n,k,m}\|_{\mathcal{X}_n} = \mathcal{O}_{\mathrm{p}}(a_{n,1})$ and $\|\delta_{n,k,\sigma}\|_{\mathcal{X}_n} = \mathcal{O}_{\mathrm{p}}(a_{n,2})$ where $\|\cdot\|_{\mathcal{X}_n}$ denotes the supremum norm when restricted to $\mathcal{X}_n$.*

ASSUMPTION 3.2. *The density f_k satisfies $\sup_{t\in\mathbb{R}} f_k(t)(|t|\vee 1) < \infty$ and, for some common absolute constant L, and $\forall t_1, t_2 \in \mathbb{R}$, $|f_k(t_1) - f_k(t_2)| \leq L|t_1 - t_2|/\{(1\vee|t_1|)(1\vee|t_2|)\}$.*

Assumption 3.1 will later be superseded by the stronger Assumption 4.1 which further requires $a_{n,1}, a_{n,2} = O(n^{-\tau})$ for some $1/4 < \tau < 1/2$. If bounds on $\delta_{n,k,m}$ and/or $\delta_{n,k,\sigma}$ uniform over $\mathcal{X}$ are difficult to find, possibly as a result of m_k and/or σ_k being unbounded over the full support $\mathcal{X}$, then Assumptions 3.1 and 4.1 allow for bounds on $\delta_{n,k,m}$ and $\delta_{n,k,\sigma}$ uniform only over a restricted support $\mathcal{X}_n$, so long as $\mathcal{X}_n$ grows fast enough to enclose all $\mathbf{X}_i$, $i \in [n]$ with high probability. In general, $\mathcal{X}_n$ can be allowed to grow more slowly under stronger moment condition on $\mathbf{X}$. For example, under homoscedastic linear regression discussed in Section 1 with q fixed (admittedly a simple case that is not our primary interest), if $\widehat{m}_k(\mathbf{x}) = \widehat{\mathbf{B}}_k^\top \mathbf{x}$ (and $\widehat{\sigma}_k(\mathbf{x}) = 1$ identically) with $\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| = \mathcal{O}_{\mathrm{p}}(n^{-\delta_1})$ for $\delta_1 > \frac{1}{4}$, and $\mathbb{E}[\|\mathbf{X}\|^{\delta_2}] < \infty$ for $\delta_2 > (\delta_1 - \frac{1}{4})^{-1}$, then Assumption 4.1 can be satisfied by taking $\mathcal{X}_n = [-n^{1/\delta_2}, n^{1/\delta_2}]^q$ which results in $\tau = \delta_1 - 1/\delta_2 > \frac{1}{4}$. From now on we make the blanket assumption that for each $n \geq 1$, the random variable $\mathbf{X} = \mathbf{X}^{(n)}$ actually comes from a new, truncated sequence $\mathbf{X}\mathbb{1}\{\mathbf{X} \in \mathcal{X}_n\}$, $n \geq 1$, and the i.i.d. copies $\mathbf{X}_i = \mathbf{X}_i^{(n)}$, $i \in [n]$ of $\mathbf{X}$ come from the truncated triangular array $\mathbf{X}_i\mathbb{1}\{\mathbf{X}_i \in \mathcal{X}_n\}$, $i \in [n]$, $n \geq 1$; this comment applies to our definitions of $\bar{\delta}_{n,k,m}$ and $\bar{\delta}_{n,k,\sigma}$ in (2). Since as n increases the probability of the truncation actually having taken place on some (newly defined) $\mathbf{X}_i$ over $i \in [n]$ approaches zero under Assumption 3.1, and all our analysis are equally valid under the aforementioned truncated random variable and triangular array setup, the asymptotics we will derive hold under the original, un-truncated setup. Of course, if we can establish bounds on $\delta_{n,k,m}$ and $\delta_{n,k,\sigma}$ uniform over $\mathcal{X}$, truncation is no longer necessary and we can simply take $\mathcal{X}_n \equiv \mathcal{X}$. This actually will be the case for all examples in Section 6.

The first half of Assumption 3.2 is weaker than Assumption 4.5 that we will address later. The second half is a reinforced Lipschitz condition on f_k. The reinforcement is needed to estimate $\sigma_k \neq 1$ under the heteroscedastic case, which causes not only a location but also a scale shift in the estimator $\widehat{\mathbf{E}}_i$; however, if we assume $\sigma_k = 1$ identically, then the reinforcement is no longer necessary. Intuitively, if $|t_1|$ and/or $|t_2|$ are large, then $f_k(t_1)$ and/or $f_k(t_2)$ should be small to begin with, which restricts how much $|f_k(t_1) - f_k(t_2)|$ can be. Lemma A.3 will verify this assumption with a reasonable L for (univariate) Student's t-distributions with degrees of freedom (d.o.f.) $\nu_{\mathrm{df}} \geq 1$ (and densities similarly decaying polynomially as $f_k(t) \sim (1+|t|)^{-(\nu_{\mathrm{df}}+1)}$).

3.2. *General results on residual empirical processes.* The proofs of the propositions below appear in Section B in the supplement. Introduce a remainder (function) term $r_{2n,k}$ as

$$r_{2n,k}(t) = \mathbb{P}(\widehat{E}_k \leq t|\mathcal{T}_n) - F_k(t) - f_k(t)\left\{t\bar{\delta}_{n,k,\sigma} + \bar{\delta}_{n,k,m}\right\}, \quad t \in \mathbb{R}. \tag{4}$$

PROPOSITION 3.1 (Residual empirical process). *For all $n \geq 1$, $k \in [p]$ and $t \in \mathbb{R}$ the equality $\widehat{F}^{\mathrm{r}}_{n,k}(t) = F_{n,k}(t) + f_k(t)\left\{t\bar{\delta}_{n,k,\sigma} + \bar{\delta}_{n,k,m}\right\} + r^{\mathrm{r}}_{1n,k}(t) + r_{2n,k}(t)$ holds. The same equality also holds with the simultaneous replacements of $\widehat{F}^{\mathrm{r}}_{n,k}$ by $\widehat{F}_{n,k}$, and $r^{\mathrm{r}}_{1n,k}$ by $r_{1n,k}$. Moreover, under Assumptions 3.1 and 3.2, the remainder term $r_{2n,k}$ satisfies*

$$\sup_{t\in\mathbb{R}} |r_{2n,k}(t)| = \mathcal{O}_{\mathrm{p}}(a_{n,1}^2 + a_{n,2}^2). \tag{5}$$

For residual ranks we need two other remainder terms $r_{3n,k,i}$ and $r_{4n,k,i}$:

$$r_{3n,k,i} = F_{n,k}(\widehat{E}_{i,k}) - F_k(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k}) + F_k(E_{i,k}), \tag{6}$$

$$\begin{aligned} r_{4n,k,i} = &\Big[F_k(\widehat{E}_{i,k}) - F_k(E_{i,k}) - f_k(E_{i,k})\left\{-E_{i,k}\delta_{n,k,\sigma,i} - \delta_{n,k,m,i}\right\}\Big] \\ &+ \Big[f_k(\widehat{E}_{i,k})\Big\{\widehat{E}_{i,k}\bar{\delta}_{n,k,\sigma} + \bar{\delta}_{n,k,m}\Big\} - f_k(E_{i,k})\left\{E_{i,k}\bar{\delta}_{n,k,\sigma} + \bar{\delta}_{n,k,m}\right\}\Big]. \end{aligned} \tag{7}$$

PROPOSITION 3.2 (Residual rank). *For all $n \geq 1$, $k \in [p]$ and $i \in [n]$,*

$$\begin{aligned} \widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k}) = &-f_k(E_{i,k})\left\{E_{i,k}\left(\delta_{n,k,\sigma,i} - \bar{\delta}_{n,k,\sigma}\right) + \left(\delta_{n,k,m,i} - \bar{\delta}_{n,k,m}\right)\right\} \\ &+ r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i}. \end{aligned} \tag{8}$$

Moreover, under Assumptions 3.1 and 3.2, the remainder terms $r_{3n,k,i}$ and $r_{4n,k,i}$ satisfy

$$\max_{i\in[n]} \frac{|r_{3n,k,i}|}{\log^{\frac{1}{2}}(n)n^{-\frac{1}{2}}\left[f_k(E_{i,k})\left\{|E_{i,k}|a_{n,2} + a_{n,1}\right\} + a_{n,1}^2 + a_{n,2}^2\right]^{1/2} + \frac{\log(n)}{n}} = \mathcal{O}_{\mathrm{p}}(1), \tag{9}$$

$$\max\nolimits_{i\in[n]} |r_{4n,k,i}| = \mathcal{O}_{\mathrm{p}}(a_{n,1}^2 + a_{n,2}^2). \tag{10}$$

3.3. *Bounds on the oscillation term.* We first briefly review the concept of a bracketing number. Let $(\mathcal{F}, \|\cdot\|_{\mathrm{g}})$ be a subset of a normed space of real-valued functions. The bracketing number $N_{[]}(\mu, \mathcal{F}, \|\cdot\|_{\mathrm{g}})$ is defined as the minimal number of μ-brackets as measured by the norm $\|\cdot\|_{\mathrm{g}}$ needed to cover the set $\mathcal{F}$, where a μ-bracket $[l, u]$ is the set of all functions $f \in \mathcal{F}$ with $l \leq f \leq u$ where l, u satisfy $\|u - l\|_{\mathrm{g}} \leq \mu$ (see, e.g., Definition 2.1.6 in van der Vaart and Wellner (1996)).

Denote by $L_2(F_{\mathbf{X}})$ the L_2 norm on functions from $\mathbb{R}^q$ to $\mathbb{R}$ with respect to the distribution $F_{\mathbf{X}}$. Let $\{\mathcal{D}_{1,n}\}_{n\geq 1}$, $\{\mathcal{D}_{2,n}\}_{n\geq 1}$ be two sequences of collections of functions from $\mathcal{X}_n$ to $\mathbb{R}$ in which we will embed the estimators $(\widehat{m}_k - m_k)/\sigma_k$ and $\widehat{\sigma}_k/\sigma_k$ respectively as in Assumption 3.3. Recall from (i) in Section 2 that we assume the same non- or semi-parametric regression model for all coordinates $k \in [p]$ (though the functions m_k, σ_k could be different across $k \in [p]$). Thus it is reasonable that we could embed $\widehat{m}_k$, $\widehat{\sigma}_k$ across $k \in [p]$ into the same collections, though generalization to coordinate-specific collections is straightforward.

ASSUMPTION 3.3. *The classes $\mathcal{D}_{1,n}$, $\mathcal{D}_{2,n}$ satisfy $\sup_{n\geq 1, d\in\mathcal{D}_{1,n}, \mathbf{x}\in\mathcal{X}_n} |d|(\mathbf{x}) < \infty$, $\inf_{n\geq 1, \widetilde{d}\in\mathcal{D}_{2,n}, \mathbf{x}\in\mathcal{X}_n} \widetilde{d}(\mathbf{x}) \geq 1/2$. For each $k \in [p]$ there exists a sequence of events $\{D_{n,k}\}_{n\geq 1}$ satisfying $\mathbb{P}(D_{n,k}) \to 1$ and $\frac{\widehat{m}_k - m_k}{\sigma_k} \in \mathcal{D}_{1,n}$, $\frac{\widehat{\sigma}_k}{\sigma_k} \in \mathcal{D}_{2,n}$ on $D_{n,k}$. In addition there exist an absolute constant $\varepsilon > 0$, and constants $K_1 = K_{1,n} > \varepsilon$, $K_2 = K_{2,n} \geq 0$, $\beta = \beta_n > \varepsilon$ and $\nu = \nu_n > 1$ (that may depend on n) such that*

$$\begin{aligned} &\max\{N_{[]}(\mu, \mathcal{D}_{1,n}, L_2(F_{\mathbf{X}})), N_{[]}(\mu, \mathcal{D}_{2,n}, L_2(F_{\mathbf{X}}))\} \\ &\quad \leq K_1(1/\mu)^{\beta}\exp(K_2(1/\mu)^{1/\nu}), \quad \forall \mu \in (0, 1). \end{aligned} \tag{11}$$

Given Assumption 3.3, define the function $w : (0, 1] \to \mathbb{R}^+$ as $w(u) = \beta^{1/2}\log^{1/2}(1/u)u + K_2^{1/2}(1 - \frac{1}{\nu})^{-1}u^{1-1/\nu}$, and a remainder term $\widetilde{R}_n$ as

$$\begin{aligned} \widetilde{R}_n = &\beta\log(n)n^{-1} + \beta^{1/2}K_2^{\frac{1}{2}\frac{1}{1+1/\nu}}\log^{1/2}(n)n^{-\frac{1}{2}\left(1+\frac{1}{1+1/\nu}\right)} \\ &+ K_2^{1/2}(1 - \tfrac{1}{\nu})^{-1}n^{-1/2}\left\{\beta^{1/2}\log^{1/2}(n)n^{-1/2} + (K_2 n^{-1})^{\frac{1}{2}\frac{1}{1+1/\nu}}\right\}^{1-1/\nu}. \end{aligned}$$

PROPOSITION 3.3. *Suppose that Assumptions 3.1 to 3.3 hold, and $K_1 = K_{1,n}$, $K_2 = K_{2,n}$ and $\beta = \beta_n$ satisfy*

$$\log(K_1) = \mathcal{O}(\beta \log(a_{n,1}^{-1} \wedge a_{n,2}^{-1} \wedge n)), \quad \max\{\beta \log(n), K_2\} n^{-1} = o(1). \tag{12}$$

Then for all n large enough,

$$\sup_{t \in \mathbb{R}} \frac{\max\{|r_{1n,k}(t)|, |r^{\mathrm{r}}_{1n,k}(t)|\}}{n^{-1/2} w \left[\{f_k(t)(|t| a_{n,2} + a_{n,1}) + a_{n,1}^2 + a_{n,2}^2\}^{1/2} \right] + \widetilde{R}_n} = \mathcal{O}_{\mathrm{p}}(1). \tag{13}$$

If furthermore $K_2 = K_{2,n}$, $\beta = \beta_n$ and $\nu = \nu_n$ satisfy

$$\begin{gathered} \nu > 1 + \varepsilon, \quad \beta \log(a_{n,1}^{-1} \wedge a_{n,2}^{-1}) = \mathcal{O}(K_2 (a_{n,1}^{-2/\nu} \wedge a_{n,2}^{-2/\nu})), \\ \beta \log(n) n^{-1} = \mathcal{O}((K_2 n^{-1})^{\frac{1}{1+1/\nu}}), \end{gathered} \tag{14}$$

then with $\Delta = \frac{1}{2}(1 - 1/\nu)$ and $R_n = n^{-\frac{1}{1+1/\nu}} (= o(n^{-1/2}))$,

$$\sup_{t \in \mathbb{R}} \frac{\max\{|r_{1n,k}(t)|, |r^{\mathrm{r}}_{1n,k}(t)|\}}{n^{-1/2} K_2^{1/2} \left\{ f_k(t)(|t| a_{n,2} + a_{n,1}) + a_{n,1}^2 + a_{n,2}^2 \right\}^{\Delta} + K_2^{\frac{1}{1+1/\nu}} R_n} = \mathcal{O}_{\mathrm{p}}(1). \tag{15}$$

The proof of Proposition 3.3 appears in Section B.3. We provide some remarks about Assumption 3.3 and the proposition. First, Assumption 3.3 will later be superseded by the stronger Assumption 4.2 when we analyze residual-based estimators for copula parameters. Both assumptions place conditions on the randomness of $\widehat{m}_k$ and $\widehat{\sigma}_k$ by restricting the complexity, measured by the bracketing numbers in (11) in which K_1, K_2, β and ν can all potentially vary with n, of the function classes in which we embed $\widehat{m}_k$ and $\widehat{\sigma}_k$. The constituent exponential in $1/\mu$ involving K_2 and ν in (11) aims toward bounding the complexity of the non-parametric component of $\widehat{m}_k$ and $\widehat{\sigma}_k$ in either a non- or a semi-parametric model. The required bound usually translates into a smoothness condition on (the non-parametric component of) $\widehat{m}_k$ and $\widehat{\sigma}_k$, and as such can in principle always be made to satisfy: if the targets m_k and σ_k are indeed smooth enough but the original estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ are not, we can always produce appropriately smoothed versions of $\widehat{m}_k$ and $\widehat{\sigma}_k$ (e.g., by convolution with a smooth function, as how a smoothed empirical distribution function can be obtained from the raw empirical distribution function that is a step function) that simultaneously satisfy the required bound and retain the original convergence rates. Nevertheless later in Section 6 we will verify Assumption 4.2 by directly checking the smoothness of the original $\widehat{m}_k$ and $\widehat{\sigma}_k$. In contrast, the constituent polynomial in $1/\mu$ in (11) involving K_1 and β aims toward bounding the complexity of any "leftover", parametric component in either a non- or a semi-parametric model.

As a simple illustration of Proposition 3.3, we consider a homoscedastic linear regression in $\mathbb{R}^q$. Then we can set $K_2 = 0$ and $a_{n,2} = 0$ (which is allowed by Eq. (13)) to eliminate the complication caused by the non-parametric component. Simple calculation also shows that we can set $\beta = q$ and $K_1 = M^q q^q$. Then from (13) we can recover as a special case a slight variant of (3.8) in Proposition 3.1 in Zhao, Gijbels and Van Keilegom (2020) under linear regression, but now we are also allowing β (and thus q) to depend on n; we omit the details for this recovery. At the other extreme, if (the last two parts of) (14) holds — we interpret this as the effect of the non-parametric component dominating over the parametric component — we obtain the simplification (15).

We also note that when model (1) is fixed across $n \geq 1$, we typically can embed $\widehat{m}_k$ and $\widehat{\sigma}_k$ into respectively fixed collections $\mathcal{D}_{1,n} = \mathcal{D}_1$ and $\mathcal{D}_{2,n} = \mathcal{D}_2$, resulting in fixed constants K_1,

K_2, β, ν in (11). Then (12) and (14) hold trivially. Thus intuitively (12) and (14) will also hold when model (1) does not vary too rapidly, and as a consequence the constants K_1, K_2, β, ν do not vary too greatly, with n. *In practice*, however, determining the precise dependencies of K_2 and ν on the dimension of the non-parametric component is often unwieldy. Thus in practice for simplicity for the remainder of the paper we will assume that K_2 and ν are fixed but will allow K_1 and β responsible for the parametric component to vary with n under (12) and (14). (Even with K_2 and ν fixed and with (12), (14) and also (18) later enforced, K_1 and β can still increase with n, in particular because under fixed β the dependence on n on the left-hand sides of (14) and (18) is strictly faster than that on the right-hand sides. Discarding (12), (14) and (18), thus allowing the parametric component to dominate the non-parametric component, introduces no extra technical challenge, but for brevity of presentation we do not discuss this admittedly more general case in the present paper.) The bound (11) in Assumptions 3.3 and 4.2 will be verified in Propositions 6.1, 6.2 and 6.3 for non-parametric regression, a partly linear regression variant with the dimension of the parametric component possibly increasing, and an additive model variant respectively.

The technical details $K_1 > \varepsilon$ and (12) prevent trivial cases and simplify our analysis. In addition, instead of in terms of bracketing number, Proposition 3.3 can be presented in terms of covering number (e.g., Definition 2.1.5 in van der Vaart and Wellner (1996)) as well; this requires only minor modification of the current proof.

Last but not least, we note that the first-order remainder terms $r_{1n,k}$, $r^{\mathrm{r}}_{1n,k}$ and $r_{3n,k,i}$ in Propositions 3.2 and 3.3 reveal two simultaneous features.

- They can all converge as $\mathcal{O}_{\mathrm{p}}(n^{-c})$ with $c > 1/2$, so strictly faster than $o_{\mathrm{p}}(n^{-1/2})$, given fast enough $a_{n,1}$ and $a_{n,2}$. For example, under the condition necessary for (9) and (15), if further we can take $a_{n,1}, a_{n,2} = n^{-\tau}$ for $0 < \tau < 1$, and $K_2 > 0$ and $\nu > 1$ to be fixed constants, then $\sup_{t\in\mathbb{R}} \max\{|r_{1n,k}(t)|, |r^{\mathrm{r}}_{1n,k}(t)|\} = \mathcal{O}_{\mathrm{p}}(n^{-c})$ for $c = \min\{\frac{1}{2} + \frac{1}{2}(1 - \frac{1}{\nu})\tau, \frac{1}{1+1/\nu}\} > 1/2$ and $\max_{i\in[n]} |r_{3n,k,i}| = \mathcal{O}_{\mathrm{p}}(\log^{\frac{1}{2}}(n) n^{-(\frac{1}{2}+\frac{\tau}{2})})$.
- Their rates are weighted down multiplicatively by the density f_k, which further sharpens the rates for a distribution whose density decays in the tails.

Both these features will help to tame unbounded score functions. In contrast, traditional residual empirical process theory under model (1) (e.g., Akritas and Van Keilegom (2001)) or under linear regression (e.g., Mammen (1996); Chen and Lockhart (2001)) typically only yields $o_{\mathrm{p}}(n^{-1/2})$ for these terms, and lacks the weighing feature.

4. Residual-based normal scores estimator.

In this section we first introduce in Section 4.1 the Gaussian copula and the associated oracle and residual-based normal scores estimators. In Section 4.3 we establish the asymptotic equivalence between these estimators. Necessary assumptions are collected in Section 4.2.

4.1. *Gaussian copula and the (residual-based) normal scores estimator.*

We call a copula $C = C(\cdot; \mathbf{R})$ a *Gaussian copula* uniquely characterized by a copula correlation matrix $\mathbf{R}$ if it is the copula of a p-variate normal distribution $\mathcal{N}_p(\mathbf{0}, \mathbf{R})$ with positive-definite correlation matrix $\mathbf{R}$. If $\mathbf{E}$ has Gaussian copula $C(\cdot; \mathbf{R})$, then $\Phi^{\leftarrow}(F_1(E_1))$, …, $\Phi^{\leftarrow}(F_p(E_p))$ jointly follows a $\mathcal{N}_p(\mathbf{0}, \mathbf{R})$ distribution. Then it is easily derived that $C(\cdot; \mathbf{R})$ admits the form $C(\mathbf{u}; \mathbf{R}) = \Phi_{\mathbf{R}}(\Phi^{\leftarrow}(u_1), \dots, \Phi^{\leftarrow}(u_p))$ for $\mathbf{u} \in [0, 1]^p$. Here $\Phi_{\mathbf{R}}$ is the cumulative distribution function of the $\mathcal{N}_p(\mathbf{0}, \mathbf{R})$ distribution. The (oracle, semi-parametric) Gaussian copula model, a straight extension of multivariate normal distributions, is the one where each $\mathbf{E}$ in the model has a Gaussian copula and where the associated $\mathbf{R}$ can vary over all positive-definite correlation matrices.

A primary theme on the studies of the Gaussian copula is the estimation of the population copula correlation matrix $\mathbf{R}_0$. Among the rank-based estimators for $\mathbf{R}_0$, the (oracle) normal scores estimator holds special importance because it is semi-parametrically efficient in an *unrestricted* model for $\mathbf{R}$ (where $\mathbf{R}$ is a positive-definite correlation matrix but otherwise arbitrary), meaning that it has the smallest (asymptotic) covariance matrix among all estimators of $\mathbf{R}_0$ (Klaassen and Wellner (1997); Segers, van den Akker and Werker (2014); Zhao and Genest (2019)). The version of this estimator based on residual ranks is the central object of study in this section, and we first record the form of the original, oracle estimator.

In this paper, for a matrix $\mathbf{A}$, let $(\mathbf{A})_{kk'}$ denote its (k,k')th element; for a two-dimensional array of numbers $a_{kk'}$, $k,k' \in [p]$, let $[a_{kk'}]_{k,k'\in[p]}$ denote a matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ with $(\mathbf{A})_{kk'} = a_{kk'}$. With such notations, the (oracle) normal scores estimator $\mathbf{R}_n = [r_{n,kk'}]_{k,k'\in[p]}$ of $\mathbf{R}_0 \equiv [r_{0,kk'}]_{k,k'\in[p]}$ is defined in, e.g., Eq. (7) on p. 113 in Hájek and Šidák (1967) as

$$(16)\qquad r_{n,kk'} = \tfrac{\phi_n}{n}\textstyle\sum_{i\in[n]} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})), \quad \forall k,k' \in [p].$$

Here $\phi_n = [n^{-1}\sum_{i\in[n]}\{\Phi^{\leftarrow}(\frac{i}{n+1})\}^2]^{-1} = 1 + \mathcal{O}(n^{-1}\log(n))$ is a deterministic, asymptotically insignificant correction factor to ensure that the matrix $\mathbf{R}_n$, and analogously $\widehat{\mathbf{R}}_n$ given later in (17), have unit diagonal elements, and hence are proper correlation matrices. The elements $r_{n,kk'}$, $k,k' \in [p]$ belong to multivariate rank order statistics (abbreviated as MvROS in this paper) (Hájek and Šidák (1967); Ruymgaart (1974)). In this context the product $\Phi^{\leftarrow}(\cdot)\Phi^{\leftarrow}(\cdot)$ in (16) is called a score function (this terminology is related to though not entirely identical to the score function in Section 5). Note also that we can write $\mathbf{R}_n$ explicitly in the form of a positive semi-definite sample correlation matrix as $\mathbf{R}_n = \frac{\phi_n}{n}\sum_{i\in[n]}\mathbf{Z}_{n,i}\mathbf{Z}_{n,i}^\top$, using the (non-independent) oracle *Gaussianized observations* $\mathbf{Z}_{n,i} = (\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,1}(E_{i,1})),\ldots,\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,p}(E_{i,p})))^\top$.

Already under the oracle model, the analysis of MvROSs is complicated by the facts that the estimators are not simple i.i.d. sums and the score functions are often unbounded (as is the case here). Model (1) introduces yet another layer of complication due to the necessary reliance on residual ranks. Overcoming such complication is the major focus of our paper.

Now, define $\widehat{\mathbf{R}}_n = [\widehat{r}_{n,kk'}]_{k,k'\in[p]}$, the residual-based normal scores estimator of $\mathbf{R}_0$, as

$$(17)\qquad \widehat{r}_{n,kk'} = \tfrac{\phi_n}{n}\textstyle\sum_{i\in[n]} \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'})), \quad \forall k,k' \in [p],$$

with ϕ_n as in (16). Just as $\mathbf{R}_n$ earlier, but substituting the oracle $\mathbf{Z}_{n,i}$ by the residual Gaussianized observations $\widehat{\mathbf{Z}}_{n,i} \equiv (\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1})),\ldots,\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p})))^\top$, $\widehat{\mathbf{R}}_n$ can also be written as a positive semi-definite sample correlation matrix.

4.2. *Assumptions.* We remark that some assumptions are strengthened to shorten our analysis, and thus not all are in the weakest possible form. We provide some insights on the assumptions at the end of Section 4.2. Also recall that from now on the dimension of the nonparametric component of model (1), and hence the constants K_2 and ν in Assumption 3.3, are fixed.

ASSUMPTION 4.1. *Assumption 3.1 holds, and moreover there exists an absolute constant $1/4 < \tau < 1/2$ such that $a_{n,1}, a_{n,2} = O(n^{-\tau})$.*

ASSUMPTION 4.2. *Assumption 3.3, (12) and (14) hold, and furthermore $\beta = \beta_n$ satisfies*

$$(18)\qquad \beta\log(a_{n,1}^{-1}) = \mathcal{O}(a_{n,1}^{-1/\nu}) \text{ and } \beta\log(a_{n,2}^{-1}) = \mathcal{O}(a_{n,2}^{-1/\nu}).$$

Recall the oscillation terms $r_{1n,k}$ and $r^{\mathrm{r}}_{1n,k}$ introduced in (3) and below.

ASSUMPTION 4.3. *For some absolute constants* $0 < \Delta < 1/2$, $1/2 < \xi < 1$ *and a remainder term* $R_n \leq n^{-\xi}$, *the following equality holds:*

$$\sup_{t\in\mathbb{R}} \frac{\max\{|r_{1n,k}(t)|, |r^{\mathrm{r}}_{1n,k}(t)|\}}{n^{-1/2}\{f_k(t)(|t|a_{n,2}+a_{n,1})+a^2_{n,1}+a^2_{n,2}\}^{\Delta}+R_n} = \mathcal{O}_{\mathrm{p}}(1).$$

Next, given $\gamma < 1$ in Assumption 4.4, define $\mathcal{U}_{n,\gamma} = (n^{-\gamma}, 1-n^{-\gamma})$. For $k \in [p]$, define $G_k(\cdot; a_1, a_2) : [0,1] \to \mathbb{R}^+$ indexed by $a_1, a_2 \in \mathbb{R}^+$ as $G_k(u; a_1, a_2) = f_k(F_k^{\leftarrow}(u))a_1 + f_k(F_k^{\leftarrow}(u))|F_k^{\leftarrow}(u)|a_2$.

ASSUMPTION 4.4. *Assumptions 4.1 and 4.3 hold, and moreover there exists an absolute constant* γ *satisfying* $1/2 < \gamma < \min\{2\tau, 1/2+2\Delta\tau, \xi\}$ *(note that such a* γ *exists when* $\tau > 1/4$, $\Delta > 0$ *and* $\xi > 1/2$*) such that*

$$\sup_{u\in\mathcal{U}_{n,\gamma}} G_k(u; a_{n,1}, a_{n,2})/\{u \wedge (1-u)\} = o(1), \tag{19}$$

$$n^{-\frac{1}{2}} \sup_{u\in\mathcal{U}_{n,\gamma}} G_k^{\Delta}(u; a_{n,1}, a_{n,2})/\{u \wedge (1-u)\} = o(1), \tag{20}$$

$$n^{-\frac{1}{2}} \log^{\frac{1}{2}}(n) \sup_{u\in\mathcal{U}_{n,\gamma}} G_k^{\frac{1}{2}}(u; a_{n,1}, a_{n,2})/\{u \wedge (1-u)\} = o(1). \tag{21}$$

ASSUMPTION 4.5. *For some absolute constant* $0 < \delta \leq 1/2$, *for each* $k \in [p]$ *the kth marginal satisfies* $\sup_{u\in[0,1]} G_k(u; 1, 1)/\{u \wedge (1-u)\}^{\frac{1}{2}+\delta} < \infty$.

Of the assumptions above, Assumptions 4.1 and 4.2 place probabilistic conditions on the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$, and could be verified for particular non- or semi-parametric model on a case-by-case basis. These assumptions are strengthened versions of Assumptions 3.1 and 3.3 respectively, and together with the latter have been addressed before. We only mention here in addition that Assumptions 4.1 allows convergence rates of $\widehat{m}_k$ and $\widehat{\sigma}_k$ to be slower than $n^{-1/2}$, as is typical in non- and semi-parametric models, and that (18) in Assumption 4.2 is another condition under which the effect of the non-parametric component dominates over the parametric component.

Assumption 4.3 places a condition on the convergence rate of the oscillation terms $r_{1n,k}$ and $r^{\mathrm{r}}_{1n,k}$. In fact this assumption has already been verified in Proposition 3.3 under the smoothness condition in Assumption 3.3. Nevertheless because the parameters Δ and ξ in Assumption 4.3 will appear in Assumption 4.4, we state Assumption 4.3 as a standalone assumption. As a specific example, in Section 6.1 we will verify Assumption 4.3 for $\Delta = \frac{1}{2}(1-1/\nu)$ where $\nu = 1+\alpha/q$, and $R_n = n^{-\frac{1}{1+1/\nu}}$, *when* we can choose (q, α) as the Hölder continuity indices of the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$. In addition, of course larger Δ yields faster convergence rate, but by our current technique $\Delta < 1/2$ appears to be a barrier when dealing with the non-parametric component of $\widehat{m}_k$ and $\widehat{\sigma}_k$. Thus we explicitly set $1/2$ as a strict upper bound, which will also simplify certain expressions later on.

We will verify Assumptions 4.1 to 4.3 in Section 6 for several popular non- and semi-parametric regression models under (1). Next, Assumptions 4.4 and 4.5, when coupled with the various convergence rates, are two mild conditions mostly concerning the underlying marginal distributions F_k. For illustration consider again Student's t-distributions with d.o.f. $\nu_{\mathrm{df}} \geq 1$. Here Assumption 4.5 holds quite generously for $\delta = 1/2$ (a larger δ implies a tighter bound), and (19), (21) in Assumption 4.4 are automatically satisfied, while (20) is guaranteed by the mild condition $\gamma < (1/2+\tau\Delta)/(1-\Delta)$. We will not address the mild Assumptions 4.4 and 4.5 further.

4.3. *Asymptotics of residual-based normal scores estimator.* Introduce

$$\Delta_n = n^{-\min\{\tau(1-\frac{1}{2\nu}),\Delta\tau,\frac{1}{2}+2\Delta\tau-\gamma(1-(1+2\delta)\Delta)\}} + \log(n)\, n^{-\min\{2\tau-\frac{1}{2},\gamma-\frac{1}{2}\}}.$$

THEOREM 4.1. *Suppose that Assumptions 3.2 and 4.1 to 4.5 hold. Then $\Delta_n = o(1)$ and $\sqrt{n}(\widehat{r}_{n,kk'} - r_{n,kk'}) = \mathcal{O}_{\mathrm{p}}(\Delta_n)$, $\forall k, k' \in [p]$.*

The proof of Theorem 4.1 appears in Section D.1 in the supplement and essentially proceeds as a special case of the later Proposition 5.1. The theorem shows that the asymptotic distribution of the residual-based normal scores estimator $\widehat{\mathbf{R}}_n$ adjusted for model (1) is the same as its oracle counterpart, which further implies that $\widehat{\mathbf{R}}_n$ is a semi-parametrically efficient estimator of the copula correlation matrix $\mathbf{R}_0$ (in the unrestricted model) under model (1).

Although the normal scores estimator is *semi-parametrically* efficient, not knowing the marginals still causes efficiency loss in general. For example, for a bivariate Gaussian distribution with known marginal variances, the information lower bound for estimating the correlation ρ_0 (the only unknown parameter) is $(1-\rho_0^2)^2/(1+\rho_0^2)$, but deteriorates to $(1-\rho_0^2)^2$ if the marginal variances are unknown (p. 38 in Bickel et al. (1993)). The latter coincides with the semi-parametric lower bound (achieved by the normal scores estimator) in a bivariate Gaussian copula model for estimating the off-diagonal element ρ_0 of $\mathbf{R}_0$, the only unknown Euclidean parameter. Thus, here, no efficiency is lost only at independence when $\rho_0 = 0$.

Beyond the unrestricted model, we can also estimate $\boldsymbol{\theta}_0$ in a structured model where $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$ is parametrized by a lower-dimensional copula parameter $\boldsymbol{\theta}$. This and the study of general parametric copulas naturally lead to the generalized pseudo-likelihood estimation which we now describe in Section 5.

5. Residual-based generalized pseudo-likelihood estimation. In this section we extend beyond the Gaussian copula, and let $\mathbf{E}$ in the law P admit a copula $C = C(\cdot;\boldsymbol{\theta}_0)$ where C is smoothly parametrized by a copula parameter $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_d)^\top \in \boldsymbol{\Theta}$. Here $\boldsymbol{\Theta} \subset \mathbb{R}^d$ is some parameter space, and the dimension d is considered fixed throughout. Our goal is to estimate $\boldsymbol{\theta}_0$ under model (1). To this end, we adapt the PLE method (Oakes (1994); Genest, Ghoudi and Rivest (1995)). Let $\mathbf{g}(\cdot;\boldsymbol{\theta}) = (g_1(\cdot;\boldsymbol{\theta}),\ldots,g_{\bar{d}}(\cdot;\boldsymbol{\theta}))^\top$ be an appropriate *score function*. The precise requirements on $\mathbf{g}$ will be specified later; a crucial one is that the true $\boldsymbol{\theta}_0$ is the unique solution to the population level equation $\mathbb{E}g_m(\mathbf{U};\boldsymbol{\theta}) = 0$ for each $m \in [\bar{d}]$ where $\mathbf{U} = (F_1(E_1),\ldots,F_p(E_p))^\top$. In addition, necessarily $\bar{d} \geq d$ by Theorem 5.2 later. We will call the particular $\mathbf{g}$ in the form of the traditional score function in maximum likelihood estimation, that is $\bar{d} = d$ and $g_k(\cdot;\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_k}\log\{c(\cdot;\boldsymbol{\theta})\}$ for $k \in [d]$ where $c(\cdot;\boldsymbol{\theta})$ is the density of $C(\cdot;\boldsymbol{\theta})$, the *parametric* score function. Because we can accommodate score functions $\mathbf{g}$ that are different from the parametric score function (and sometimes this will be more convenient, for instance in Sections D.6.2 and D.6.3), our method should be more appropriately referred to as the *generalized* PLE, but for brevity from now on we will omit the qualifier "generalized" and will still refer to this more general version as the PLE.

Define the residual rank vectors $\widehat{\mathbf{U}}_{n,i} \equiv (\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1}),\ldots,\widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p}))^\top$, $i \in [n]$. Then we let the residual-based PLE estimator $\widehat{\boldsymbol{\theta}}_n$ be the solution to

$$\left\|\tfrac{1}{n}\textstyle\sum_{i\in[n]}\mathbf{g}(\widehat{\mathbf{U}}_{n,i};\widehat{\boldsymbol{\theta}}_n)\right\| = \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\tfrac{1}{n}\textstyle\sum_{i\in[n]}\mathbf{g}(\widehat{\mathbf{U}}_{n,i};\boldsymbol{\theta})\right\| + o_{\mathrm{p}}(n^{-1/2}). \tag{22}$$

(All probabilistic statements in this section are taken with respect to the law P with copula $C = C(\cdot;\boldsymbol{\theta}_0)$.) Note that (22) is simply a Z-estimation problem, and thus $\widehat{\boldsymbol{\theta}}_n$ is just a Z-estimator, popular in the literature, cf. (i) in Theorem 3.3 in Pakes and Pollard (1989). Similarly, define the oracle rank vectors $\mathbf{U}_{n,i} \equiv (F^{\mathrm{r}}_{n,1}(E_{i,1}),\ldots,F^{\mathrm{r}}_{n,p}(E_{i,p}))^\top$, $i \in [n]$; then, we let the oracle PLE estimator $\boldsymbol{\theta}_n$ be obtained from (22) with the substitution of $\widehat{\mathbf{U}}_{n,i}$ by $\mathbf{U}_{n,i}$ but otherwise be identical to $\widehat{\boldsymbol{\theta}}_n$.

Now we can relate the normal scores estimator $\widehat{\mathbf{R}}_n$ under model (1) and its oracle counterpart $\mathbf{R}_n$ to the PLE. The unrestricted model for $\mathbf{R}$ in the Gaussian copula is equivalent to parametrizing the copula correlation matrix as $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \mathbb{R}^{p(p-1)/2}$ is the vectorized upper-triangular portion of $\mathbf{R}$. Under this parametrization, the PLE with the parametric score function under the oracle model and under model (1) turn out to yield closed-form solutions in the form of respectively and precisely the oracle estimator $\mathbf{R}_n$ and the residual-based estimator $\widehat{\mathbf{R}}_n$; see Section 3.4 in Zhao, Gijbels and Van Keilegom (2020) for this result.

Unfortunately, for PLE of a copula parameter under general parametric copulas, closed-form solution is the exception rather than the rule even for popular copulas such as the t-copula. For analysis of the PLE in general, then, the Z-estimation machinery as we tackle in this section becomes necessary. Genest, Ghoudi and Rivest (1995) and Tsukahara (2005) have studied the asymptotics of the oracle $\boldsymbol{\theta}_n$ in detail. Our main goal in this section is to show that under mild conditions, asymptotically the residual-based PLE estimator $\widehat{\boldsymbol{\theta}}_n$ is indistinguishable from the oracle $\boldsymbol{\theta}_n$.

Before presenting our result on $\widehat{\boldsymbol{\theta}}_n$, we first introduce Proposition 5.1 which acts as a stepping stone. The proposition also extends the asymptotic normality of the classical, oracle MvROS based on a sample of $\mathbf{E}$, namely $n^{-1}\sum_{i\in[n]} g(\mathbf{U}_{n,i})$ in the proposition, to the one based on residual ranks, namely $n^{-1}\sum_{i\in[n]} g(\widehat{\mathbf{U}}_{n,i})$ in the proposition. The latter then applies under model (1). As a concrete example, when $g = g(u_k, u_{k'}) = \Phi^{\leftarrow}(u_k)\Phi^{\leftarrow}(u_{k'})$, the scaled summation of g over the residual ranks for the (k, k')th marginal pair becomes a bivariate rank order statistic that is just the element $\widehat{r}_{n,kk'}$ in the normal scores estimator $\widehat{\mathbf{R}}_n$ (see Eq. (17)); in fact this estimator is analyzed precisely with Proposition 5.1. We refer to a remark following Theorem 5.2 for the asymptotic normality of the oracle MvROS itself.

We first collect the necessary Definitions 5.1 to 5.3 and Assumption 5.1. The conditions on the function g (intended as spatial derivatives or their bounds of a score function) stated therein concern the interaction of the function g and the underlying copula C, and are trivial if g is bounded, and hence these conditions constrain the divergence of g. The conditions associated with the definitions and the assumption, as well as some others (for instance the earlier Assumptions 4.1 to 4.3 that are on the *marginal* estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ and that do not concern the copula structure), will culminate in Proposition 5.1 and Theorem 5.2. There, all statements involving γ should be interpreted as "there exists a common $1/2 < \gamma < 1$ such that these conditions hold". Section D.4 will provide some general remarks on the verification of these new conditions. As a concrete illustration, as mentioned earlier, Theorem 4.1 is a special case of Proposition 5.1, so the new conditions involved in this proposition have already been verified for a Gaussian copula. As additional concrete illustrations, we will verify the new conditions relevant to Proposition 5.1 and Theorem 5.2 in details for estimating the correlation parameter in a bivariate t-copula (with any known d.o.f. $\nu_{\text{df}} > 0$) in Section D.5, and for three particular families of the bivariate Archimedean copulas, namely the Clayton, Frank, Gumbel copulas, in Section D.6. Section D.5 also explains how the aforementioned task generalizes to estimating the correlation matrix for a p-variate t-copula.

Recall the quantities $a_{n,1}$, $a_{n,2}$, τ, Δ, ξ from Assumptions 4.1 and 4.3, and the set $\mathcal{U}_{n,\gamma}$ and the function G_k defined above Assumptions 4.4.

DEFINITION 5.1. *We say that a function $g : [0,1]^p \to \mathbb{R}$ satisfies condition (G1) with rate $b_{n,1} = o(1)$ for the tuple $(\mathbf{E}, k)$, with $k \in [p]$ and $\mathbf{E}$ having copula C and marginal distribution functions $F_1, \ldots, F_p$ corresponding to densities $f_1, \ldots, f_p$, if*

$$\int_{[0,1]^p} \{|g|(\mathbf{u}) G_k(u_k; 1, 1)\}^2 \, \mathrm{d}C(\mathbf{u}) < \infty, \tag{23}$$

$$\int_{\mathcal{U}^p_{n,\gamma}} |g|(\mathbf{u}) G^{\Delta}_k(u_k; a_{n,1}, a_{n,2}) \mathrm{d}C(\mathbf{u}) = O(b_{n,1}), \tag{24}$$

$$\log^{1/2}(n) \int_{\mathcal{U}^p_{n,\gamma}} |g|(\mathbf{u}) G^{1/2}_k(u_k; a_{n,1}, a_{n,2}) \mathrm{d}C(\mathbf{u}) = O(b_{n,1}), \tag{25}$$

$$n^{-\min\{2\Delta\tau, 2\tau-\frac{1}{2}, \xi-\frac{1}{2}\}} \int_{\mathcal{U}^p_{n,\gamma}} |g|(\mathbf{u}) \mathrm{d}C(\mathbf{u}) = O(b_{n,1}). \tag{26}$$

For $k \in [p]$, define the function $\Delta_{n,k} : [0,1] \to \mathbb{R}^+$ as

$$\begin{aligned}\Delta_{n,k}(u) = \sqrt{\log\log(n)/n}\sqrt{u \wedge (1-u)} + G_k(u; a_{n,1}, a_{n,2}) + n^{-\frac{1}{2}} G^{\Delta}_k(u; a_{n,1}, a_{n,2}) \\ + \log^{\frac{1}{2}}(n)\, n^{-\frac{1}{2}} G^{1/2}_k(u; a_{n,1}, a_{n,2}) + n^{-\min\{\frac{1}{2}+2\Delta\tau, 2\tau, \xi\}}.\end{aligned}$$

DEFINITION 5.2. *Under the same notations as Definition 5.1, we say that a function $g : [0,1]^p \to \mathbb{R}^+$ satisfies condition (G2) with rate $b_{n,2} = o(1)$ for the triplet $(\mathbf{E}, k, k')$, if $n^{1/2} \int_{\mathcal{U}^p_{n,\gamma}} g(\mathbf{u}) \Delta_{n,k}(u_k) \Delta_{n,k'}(u_{k'}) \mathrm{d}C(\mathbf{u}) = \mathcal{O}(b_{n,2})$.*

Let $\circ$ denote the Hadamard product, and $\mathbf{1}_p \in \mathbb{R}^p$ a vector of all ones.

DEFINITION 5.3. *We call a function $\bar{g} : [0,1]^p \to \mathbb{R}^+$ reproducing if there exist common $0 < \epsilon < 1$ and $1 \leq M < \infty$ such that for all $\mathbf{u} \in [0,1]^p$, $\sup_{\widetilde{\mathbf{u}} \in [\mathbf{u} - \epsilon\{\mathbf{u} \wedge (\mathbf{1}_p - \mathbf{u})\}, \mathbf{u} + \epsilon\{\mathbf{u} \wedge (\mathbf{1}_p - \mathbf{u})\}]} \bar{g}(\widetilde{\mathbf{u}}) \leq M \bar{g}(\mathbf{u})$.*

Our definition above is essentially a generalization of the *univariate* reproducing u-shape function and related functions in the literature (e.g., Tsukahara (2005)). A function is reproducing if its value is robust against local perturbation to its argument. A typical reproducing function is $\bar{g}(\mathbf{u}) = g^{\vee}(u_1) \times \cdots \times g^{\vee}(u_p)$ where $g^{\vee}(u) = (u \wedge (1-u))^{-\eta}$ for some $\eta \geq 0$.

Define the ith oracle sample point, where $i \in [n]$, as $\mathbf{U}_i = (F_1(E_{i1}), \ldots, F_p(E_{ip}))^{\top}$. Then, define the index set $\mathcal{I}_{n,\gamma} \equiv \{i \in [n] : \mathbf{U}_i \in \mathcal{U}^p_{n,\gamma}\}$.

ASSUMPTION 5.1. *$n^{-1/2} \sum_{i \in [n] \setminus \mathcal{I}_{n,\gamma}} \{g(\widehat{\mathbf{U}}_{n,i}) - g(\mathbf{U}_{n,i})\} = \mathcal{O}_{\mathrm{p}}(b_{n,3})$ for some $b_{n,3} = o(1)$, that is the "boundary effect" is negligible.*

Often the score function g diverges so severely inside the boundary region $\mathbf{u} \in [0,1]^p \setminus \mathcal{U}^p_{n,\gamma}$ that a careful analysis based on Taylor expansion becomes unmanageable there. Instead, a much cruder analysis is performed there, taking advantage of the fact that not many sample points would fall into the boundary region to begin with. Assumption 5.1 then simply assumes that the latter analysis is still reasonable enough that the boundary effect is indeed negligible. This assumption is certainly satisfied if g is bounded (admittedly not a very interesting case) and the boundary region is small ($\gamma > 1/2$ suffices), and hence intuitively is also satisfied with reasonable divergence of g. Section D.4 will explain how Assumption 5.1 can be met.

For a function g dependent on $\mathbf{u}$, let superscript(s) k enclosed in square bracket denote the "spatial" partial derivative with respect to u_k.

PROPOSITION 5.1 (Asymptotic equivalence between the residual-based and the oracle MvROS). *Let $g : [0,1]^p \to \mathbb{R}$ be a function such that for each $k \in [p]$, its kth partial derivative $g^{[k]}$ satisfies condition (G1) with rate $b_{n,1}$ for the tuple $(\mathbf{E}, k)$, and for each tuple $(k, k') \in [p] \times [p]$, its k, k'th mixed partial derivative $g^{[k,k']}$ satisfies $|g^{[k,k']}| \leq \bar{g}^{[k,k']}$ for some $\bar{g}^{[k,k']} : [0,1]^p \to \mathbb{R}^+$ such that $\bar{g}^{[k,k']}$ is reproducing and moreover satisfies condition (G2) with rate $b_{n,2}$ for the triplet $(\mathbf{E}, k, k')$. Also suppose that Assumptions 3.2, 4.1 to 4.4, and 5.1 hold. Then $n^{-1/2} \sum_{i \in [n]} \{g(\widehat{\mathbf{U}}_{n,i}) - g(\mathbf{U}_{n,i})\} = \mathcal{O}_{\mathrm{p}}(n^{-\tau(1-\frac{1}{2\nu})} + b_{n,1} + b_{n,2} + b_{n,3}) = o_{\mathrm{p}}(1)$.*

The proof of Proposition 5.1 appears in Section D.1 in the supplement. Now we are ready to present our result on $\widehat{\boldsymbol{\theta}}_n$. For $k \in [p]$, $m \in [\bar{d}]$ and $m' \in [d]$, let the functions $g_m^{[k]}(\cdot;\boldsymbol{\theta}), g_{m,m'}(\cdot;\boldsymbol{\theta}), g_{m,m'}^{[k]}(\cdot;\boldsymbol{\theta}) : [0,1]^p \to \mathbb{R}$ be $g_m^{[k]}(\mathbf{u};\boldsymbol{\theta}) = \frac{\partial}{\partial u_k} g_m(\mathbf{u};\boldsymbol{\theta})$, $g_{m,m'}(\cdot;\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_{m'}} g_m(\cdot;\boldsymbol{\theta})$, $g_{m,m'}^{[k]}(\mathbf{u};\boldsymbol{\theta}) = \frac{\partial}{\partial u_k} g_{m,m'}(\mathbf{u};\boldsymbol{\theta})$, and the matrices $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = [\mathbb{E} g_{m,m'}(\mathbf{U};\boldsymbol{\theta})]_{m\in[\bar{d}],m'\in[d]}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\boldsymbol{\theta}_0)$. Also define $\mathbf{M}_n(\boldsymbol{\theta}_0) \equiv n^{-1}\sum_{i\in[n]} \mathbf{g}(\mathbf{U}_{n,i};\boldsymbol{\theta}_0)$, an oracle MvROS.

THEOREM 5.2 (Asymptotic equivalence between the residual-based and the oracle PLE). *Suppose that the following sets of conditions hold:*

(i) *"Z-estimation conditions": $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$, $\mathbb{E}\mathbf{g}(\mathbf{U};\boldsymbol{\theta}_0) = \mathbf{0}$, and $\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}$ $\|\mathbb{E}\mathbf{g}(\mathbf{U};\boldsymbol{\theta})\| > 0$ for all $\delta > 0$; $\boldsymbol{\Gamma}$ has full column rank, $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ is continuous in a neighborhood of $\boldsymbol{\theta}_0$; for each $m \in [\bar{d}]$, the class of functions $\{g_m(\cdot;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is $C(\cdot;\boldsymbol{\theta}_0)$-Donsker.*

(ii) *"Oracle MvROS conditions": $\mathbf{M}_n(\boldsymbol{\theta}_0) = \mathbf{N}_n + o_{\mathrm{p}}(n^{-1/2})$ where the random vector $\sqrt{n}\mathbf{N}_n \rightsquigarrow \mathcal{N}_d(\mathbf{0},\mathbf{V})$.*

(iii) *"Approximation by residual rank conditions": for each $m \in [\bar{d}]$, $g_m(\cdot;\boldsymbol{\theta}_0)$ satisfies the assumptions in Proposition 5.1 with g replaced by $g_m(\cdot;\boldsymbol{\theta}_0)$ and simply with $b_{n,1}, b_{n,2} = o(1)$, and in addition we also assume*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| n^{-1/2} \sum_{i\in[n]\setminus\mathcal{I}_{n,\gamma}} \left\{ g_m(\widehat{\mathbf{U}}_{n,i};\boldsymbol{\theta}) - g_m(\mathbf{U}_i;\boldsymbol{\theta}) \right\} \right| = o_{\mathrm{p}}(1); \tag{27}$$

for each $k \in [p]$, uniformly over $m \in [\bar{d}]$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the functions $g_m^{[k]}(\cdot;\boldsymbol{\theta})$ are bounded in magnitude by $\bar{g}^{[k]}$, and uniformly over $m \in [\bar{d}]$, $m' \in [d]$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ for $\boldsymbol{\Theta}_0$ a neighborhood of $\boldsymbol{\theta}_0$, the functions $g_{m,m'}^{[k]}(\cdot;\boldsymbol{\theta})$ are also bounded in magnitude by $\bar{g}^{[k]}$, where $\bar{g}^{[k]}$ is reproducing and satisfies $\int_{[0,1]^p} \bar{g}^{[k]}(\mathbf{u}) \{u_k \wedge (1-u_k)\} C(\mathrm{d}\mathbf{u};\boldsymbol{\theta}_0) < \infty$. Moreover, Assumptions 3.2 and 4.1 to 4.4 hold (as in Proposition 5.1).

Then, both $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ admit the asymptotic representation $-(\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma})^{-1}$ $\times \boldsymbol{\Gamma}^\top(\sqrt{n}\mathbf{N}_n) + o_{\mathrm{p}}(1)$ where only the $o_{\mathrm{p}}(1)$ term could differ, and thus $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) = o_{\mathrm{p}}(1)$.

The proof of Theorem 5.2 appears in Section D.2 in the supplement. We provide a few remarks on the theorem. First, Eq. (27) is in essence stronger than Assumption 5.1 due to the required uniformity over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Section D.4 will explain how this condition can be met.

Next, the asymptotic i.i.d. representation of the oracle MvROS $\mathbf{M}_n(\boldsymbol{\theta}_0)$ had been established in Ruymgaart (1974) for the bivariate case and stated in Proposition 3 in Tsukahara (2005) for the multivariate case, among others. The required conditions in these works are not directly comparable to those in our Theorem 5.2 due to different goals (they focus on the oracle case instead of model (1)) and proof techniques, but are in general weaker. In our context, under the conditions of Theorem 5.2, if also $\int_{\mathcal{U}_{n,\gamma}^p} g^{[k]}(\mathbf{u};\boldsymbol{\theta}_0)^2\{u_k \wedge (1-u_k)\}C(\mathrm{d}\mathbf{u};\boldsymbol{\theta}_0) = o(n)$ for all $k \in [p]$, then we recover the i.i.d. representation in Tsukahara (2005):

$$\begin{aligned} \mathbf{M}_n(\boldsymbol{\theta}_0) = \sum_{i\in[n]} \tfrac{1}{n} \Big\{ & \mathbf{g}(\mathbf{U}_i;\boldsymbol{\theta}_0) + \sum_{k\in[p]} \int_{[0,1]^p} \mathbf{g}^{[k]}(\mathbf{u};\boldsymbol{\theta}_0) \\ & \times (\mathbb{1}\{U_{i,k} \le u_k\} - u_k) C(\mathrm{d}\mathbf{u};\boldsymbol{\theta}_0) \Big\} + o_{\mathrm{p}}(n^{-1/2}). \end{aligned} \tag{28}$$

The summands on the right-hand side above are i.i.d., have expectation zero, and their second parts involving the $\mathbf{g}^{[k]}$'s represent the effect of replacing the true marginals by the empirical ones. Thus we can set the random vector $\mathbf{N}_n$ in the "oracle MvROS conditions" as the i.i.d. sum in Eq. (28); obviously $\sqrt{n}\mathbf{N}_n$ is centered and convergences weakly to a multivariate normal distribution.

Analogous to Theorem 4.1 for the normal scores estimator, Theorem 5.2 shows that the residual-based PLE is asymptotically indistinguishable from its oracle counterpart. However, not knowing the marginals again causes efficiency loss in general; moreover the PLE is typically not semi-parametrically efficient (even for the Gaussian copula beyond the unrestricted model). We will offer more detailed remarks on the optimality of our residual-based PLE in Section D.3. Finally, as stated earlier, the new conditions in Proposition 5.1 and Theorem 5.2 will be verified in Section D.5 for the bivariate t-copula, and in Section D.6 for the bivariate Clayton, Frank and Gumbel copulas. We now turn to addressing the earlier Assumptions 4.1 to 4.3.

6. Applications to non- and semi-parametric regression models. In this section we apply our result in Sections 4 and 5 to several popular non- and semi-parametric regression models: Section 6.1 deals with the non-parametric regression model, Section 6.2 deals with a partly linear regression variant, and Section 6.3 deals with an additive model variant. For the latter two models we simply assume the function $\sigma_k = 1$ identically. Our effort amounts to verifying that Assumptions 4.1 to 4.3 are satisfied under these models. (We remind the readers that these assumptions concern the regression function estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ for the individual marginals, but do not concern the copula structure that is imposed separately.) As a consequence Theorem 4.1, Proposition 5.1 and Theorem 5.2 can be applied under these models as well. We do not attempt to exhaust all possible estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ in the literature, but will only consider some specific but popular candidates; we do take these candidates in their raw forms in the literature, and impose as little extra conditions as possible.

The assumptions we will need to verify are all stated coordinate-wise for each $k \in [p]$, and thus we will verify the assumptions in a coordinate-by-coordinate fashion. In all models, we set $\mathcal{X}_n = \mathcal{X}$ in Assumption 4.1.

6.1. *Non-parametric regression with multivariate covariate.* We assume that q is fixed. There have been several studies on the estimation of the distribution of E_k, for a single $k \in [p]$, by way of some preliminary estimators of m_k and σ_k under the non-parametric regression model (1). In particular the case of univariate covariate (i.e., $q = 1$) was considered in Akritas and Van Keilegom (2001), while the general $q \geq 1$ case was considered in Müller, Schick and Wefelmeyer (2009); Neumeyer and Van Keilegom (2010) via the local polynomial estimator (Fan and Gijbels (1996)).

Following the convention in Müller, Schick and Wefelmeyer (2009), a *multi-index* $\mathbf{m} \in \mathbb{R}^q$ denotes a q-dimensional vector $\mathbf{m} = (m_1, \ldots, m_q)^\top$ with non-negative integer components. For a multi-index $\mathbf{m}$ define the function $\psi_{\mathbf{m}} : \mathbb{R}^q \to \mathbb{R}$ as $\psi_{\mathbf{m}}(\mathbf{x}) = (x_1^{m_1}/m_1!) \cdots (x_q^{m_q}/m_q!)$ for $\mathbf{x} = (x_1, \ldots, x_q)^\top$. Set $\mathbf{m}_\bullet = m_1 + \cdots + m_q$. Next, for a function $h : \mathbb{R}^q \to \mathbb{R}$ and a multi-index $\mathbf{m}$, define $D^{\mathbf{m}} h(\mathbf{x}) = (\partial^{\mathbf{m}_\bullet}/\partial x_1^{m_1} \ldots \partial x_q^{m_q}) h(\mathbf{x})$ for $\mathbf{x} = (x_1, \ldots, x_q)^\top$.

For a non-negative integer d (that is eventually specified as in Proposition 6.1), let I_d and J_d denote the set of multi-indices $\mathbf{m}$ with $0 \leq \mathbf{m}_\bullet \leq d$ and with $\mathbf{m}_\bullet = d$, respectively. Let $K_1, \ldots, K_q$ be univariate kernels, K the resulting product kernel on $\mathbb{R}^q$, and $\{c_n\}_{n \geq 1}$ a bandwidth sequence. Let the local polynomial estimator for the kth coordinate be (as in Müller, Schick and Wefelmeyer (2009); Neumeyer and Van Keilegom (2010))

$$(29)\quad \widehat{\boldsymbol{\beta}}^{(k)}(\mathbf{x}) = \operatorname{argmin}_{\boldsymbol{\beta} = (\beta_{\mathbf{m}})_{\mathbf{m} \in I_d}} \sum_{i \in [n]} \left\{ Y_{i,k} - \sum_{\mathbf{m} \in I_d} \beta_{\mathbf{m}} \psi_{\mathbf{m}} \left(\frac{\mathbf{X}_i - \mathbf{x}}{c_n} \right) \right\}^2 K \left(\frac{\mathbf{X}_i - \mathbf{x}}{c_n} \right).$$

Given $\widehat{\boldsymbol{\beta}}^{(k)} = \widehat{\boldsymbol{\beta}}^{(k)}(\mathbf{x})$ (as a function of $\mathbf{x}$), we let the estimator $\widehat{m}_k$ of the function m_k be the component $\widehat{\beta}_{\mathbf{0}}^{(k)}$ of $\widehat{\boldsymbol{\beta}}^{(k)}$ corresponding to the multi-index $\mathbf{0} = (0, \ldots, 0)$. Next, let $\widehat{\boldsymbol{\gamma}}^{(k)}$ be defined similarly as $\widehat{\boldsymbol{\beta}}^{(k)}$, but with $Y_{i,k}$ replaced by $(Y_{i,k} - \widehat{m}_k(\mathbf{X}_i))^2$; given $\widehat{\boldsymbol{\gamma}}^{(k)} = \widehat{\boldsymbol{\gamma}}^{(k)}(\mathbf{x})$, we finally let the estimator $\widehat{\sigma}_k^2$ of the function σ_k^2 be the component $\widehat{\gamma}_{\mathbf{0}}^{(k)}$. This estimator of

σ_k^2 is asymptotically equivalent to the one in Neumeyer and Van Keilegom (2010), but in our experience more stable at finite sample sizes.

The above defines the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$. Developing their theoretical properties invites some extra technicalities. For a non-negative integer d and a real number $\kappa \in (0,1]$, let $\mathcal{H}(d,\kappa)$ be the collection of functions $h : \mathcal{X} \to \mathbb{R}$ such that h has continuous partial derivatives up to order d and the dth order partial derivatives are Hölder with exponent κ. As in Section 2.7 in van der Vaart and Wellner (1996), for $h \in \mathcal{H}(d,\kappa)$ define the norm

$$\|h\|_{d,\kappa} = \max_{\mathbf{m}\in I_d} \sup_{\mathbf{x}\in\mathcal{X}} |D^{\mathbf{m}} h(\mathbf{x})| + \max_{\mathbf{m}\in J_d} \sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}:\mathbf{x}\neq\mathbf{y}} \tfrac{|D^{\mathbf{m}}h(\mathbf{x})-D^{\mathbf{m}}h(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|^{\kappa}}. \tag{30}$$

Let $\mathcal{H}_{M'}(d,\kappa)$ be the set of functions $h \in \mathcal{H}(d,\kappa)$ that further satisfy $\|h\|_{d,\kappa} \le M'$. Let $\widetilde{\mathcal{H}}_{M'}(d,\kappa)$ be similarly defined with the additional constraint that if $h \in \widetilde{\mathcal{H}}_{M'}(d,\kappa)$, then $\inf_{\mathbf{x}\in\mathcal{X}} h(\mathbf{x}) \ge 1/2$. Moreover, let $f_{\mathbf{X}}$ be the density of $\mathbf{X}$, and for $\mathbf{m},\mathbf{m}' \in I_d$, let $Q_{n,\mathbf{m},\mathbf{m}'}(\mathbf{x}) = \int \psi_{\mathbf{m}}(\mathbf{u})\psi_{\mathbf{m}'}(\mathbf{u})K(\mathbf{u})f_{\mathbf{X}}(\mathbf{x}+c_n\mathbf{u})\mathrm{d}\mathbf{u}$. Let $\mathbf{Q}_n = \mathbf{Q}_n(\mathbf{x})$ be the matrix with entries $Q_{n,\mathbf{m},\mathbf{m}'}(\mathbf{x})$, $\mathbf{m},\mathbf{m}' \in I_d$ at $\mathbf{x}$.

Assumptions 4.1 to 4.3 are verified in the following Proposition 6.1, whose proof appears in Section E.1 in the supplement. The required conditions are mostly borrowed from but are also slightly stronger than, e.g., Lemma 1 in Müller, Schick and Wefelmeyer (2009); in particular our condition on the moment of E_k is precisely two times theirs due to our estimation of σ_k (which was not considered in their context). For succinctness and better connection with Section 6.2 later, we collect a subset of the conditions below in Assumption 6.1, under which $k_n \equiv \log(n)/(nc_n^q)$ satisfies $k_n \lesssim \log^{4/3}(n)n^{-2/3}$.

ASSUMPTION 6.1. *The density $f_{\mathbf{X}}$ has uniformly bounded partial derivatives up to order $q+1$ almost everywhere and its support $\mathcal{X}$ is bounded and convex, the kernels $K_1,\dots,K_q$ are $(q+2)$-times continuously differentiable and have compact support $[-1,1]$, the matrix $\mathbf{Q}_n$ satisfies $\inf_{n\ge1}\inf_{\mathbf{x}\in\mathcal{X}}\lambda_{\min}(\mathbf{Q}_n(\mathbf{x})) > 0$ where $\lambda_{\min}$ returns the minimal eigenvalue, and finally $c_n = M(n\log(n))^{-1/(2s)}$ for $M > 0$ and s to be specified in Proposition 6.1 (in Section 6.1) or Proposition 6.2 (in Section 6.2).*

PROPOSITION 6.1. *Suppose that Assumption 6.1 holds, and (i) the functions $m_k, \sigma_k \in \mathcal{H}(d,\kappa)$ with $d \ge q+1$, $s \equiv d+\kappa > 3q/2$, $\inf_{\mathbf{x}\in\mathcal{X}}\sigma_k(\mathbf{x}) > 0$; (ii) E_k satisfies $\mathbb{E}[|E_k|^{\zeta}] < \infty$ for some $\zeta > 8s/(2s-q)$. Then:*

(a) *Assumption 4.1 is satisfied with $a_{n,1} = a_{n,2} = k_n^{1/2}$. By the bound on k_n, certainly $a_{n,1}, a_{n,2} = \mathcal{O}(n^{-\tau})$ with $\tau > 1/4$.*

(b) *For Assumption 4.2, (b1) there exists $0 < \alpha < 1$ such that by setting $\mathcal{D}_{1,n} = \mathcal{H}_{M'}(q,\alpha)$ and $\mathcal{D}_{2,n} = \widetilde{\mathcal{H}}_{M'}(q,\alpha)$ for M' large enough, the required sequence $\{D_{n,k}\}_{n\ge1}$ exists; (b2) (11), (12), (14) and (18) hold if we set K_1, K_2, β as large enough positive absolute constants, and $\nu = 1+\alpha/q > 1$.*

(c) *Assumption 4.3 is satisfied with $\Delta = \frac{1}{2}(1-1/\nu)$ for ν given above and $R_n = n^{-\frac{1}{1+1/\nu}}$.*

6.2. *Partly linear regression.* The semi-parametric variant of (1) in the form of a partly linear regression model (PLM) states that the regression function $m_k(\mathbf{w},\mathbf{x}) = \boldsymbol{\theta}_k^{\top}\mathbf{w} + \widetilde{m}_k(\mathbf{x})$; here $(\mathbf{w}^{\top},\mathbf{x}^{\top})^{\top} \in \mathcal{W}\times\mathcal{X} \subset \mathbb{R}^{q_{\mathrm{L}}+q}$ is the covariate, and the (linear) regression coefficient $\boldsymbol{\theta}_k \in \mathbb{R}^{q_{\mathrm{L}}}$ and the non-parametric component $\widetilde{m}_k : \mathbb{R}^q \to \mathbb{R}$ are the unknown parameters. We also set the scale function $\sigma_k = 1$ identically. Hence $Y_k = m_k(\mathbf{W},\mathbf{X}) + E_k = \boldsymbol{\theta}_k^{\top}\mathbf{W} + \widetilde{m}_k(\mathbf{X}) + E_k$. For identification we assume $\mathbb{E}E_k = 0$ as usual, but because $\sigma_k = 1$ we impose no condition on $\mathrm{Var}(E_k)$. We assume that q is fixed. However, we allow the dimension $q_{\mathrm{L}} = q_{\mathrm{L},n}$ of the linear component to vary with n. Let $(\mathbf{Y}_i,\mathbf{W}_i,\mathbf{X}_i,\mathbf{E}_i)$, $i \ge 1$ be independent copies of $(\mathbf{Y},\mathbf{W},\mathbf{X},\mathbf{E})$, with $\mathbf{W} = (W_1,\dots,W_{q_{\mathrm{L}}})^{\top}$; our sample consists of $(\mathbf{Y}_i,\mathbf{W}_i,\mathbf{X}_i)$,

$i \in [n]$. We let $\widehat{\boldsymbol{\theta}}_k$ be an estimator of $\boldsymbol{\theta}_k$. The particular form of $\widehat{\boldsymbol{\theta}}_k$ is quite irrelevant, as long as $\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k\|$ satisfies a mild convergence condition in Proposition 6.2; see the remark following the proposition for a concrete example.

Whenever possible we follow the same notations as in Section 6.1. Then, let the local polynomial estimator $\widehat{\boldsymbol{\beta}}^{(k)} = \widehat{\boldsymbol{\beta}}^{(k)}(\mathbf{x})$ relevant to $\widetilde{m}_k$ be obtained from (29) but with $Y_{i,k}$ replaced by $Y_{i,k} - \widehat{\boldsymbol{\theta}}_k^\top \mathbf{W}_i$ (as in the last equation on p. 554 in Müller, Schick and Wefelmeyer (2012)). Then we let the estimator $\widehat{\widetilde{m}}_k$ for the function $\widetilde{m}_k$ be the component $\beta_{\mathbf{0}}^{(k)}$ of $\widehat{\boldsymbol{\beta}}^{(k)}$ corresponding to the multi-index $\mathbf{0} = (0, \ldots, 0)$. Finally we let the estimator of $m_k(\mathbf{w}, \mathbf{x})$ be $\widehat{m}_k(\mathbf{w}, \mathbf{x}) = \widehat{\boldsymbol{\theta}}_k^\top \mathbf{w} + \widehat{\widetilde{m}}_k(\mathbf{x})$; we set the estimator $\widehat{\sigma}_k = 1$ identically.

Proposition 6.2 below is the analogy to Proposition 6.1 now in the context of the PLM variant. The required conditions are mostly borrowed from but are also slightly stronger than those in Theorem 2.1 in Müller, Schick and Wefelmeyer (2012). In particular the component-wise uniform boundedness of $\mathbf{W}$ facilitates application of deviation inequalities for supremum of empirical processes when $q_{\mathrm{L}} = q_{\mathrm{L},n}$ varies with n. Let $\boldsymbol{\mu}(\mathbf{x}) = \mathbb{E}(\mathbf{W}|\mathbf{X} = \mathbf{x})$, with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{q_{\mathrm{L}}})^\top$. Let $\mathcal{D}_{q_{\mathrm{L}}}$ be the collection $\mathcal{D}_{q_{\mathrm{L}}} = \{f(\cdot; \boldsymbol{\delta}) : \mathbb{R}^{q_{\mathrm{L}}} \to \mathbb{R}, f(\mathbf{w}; \boldsymbol{\delta}) = \mathbf{w}^\top \boldsymbol{\delta}, \boldsymbol{\delta} \in \mathbb{R}^{q_{\mathrm{L}}}, \|\boldsymbol{\delta}\| \leq 1\}$ of functions indexed by $\boldsymbol{\delta}$. We refer to, e.g, the beginning of Section 2.1 in Giné and Mason (2007) for the definition of Vapnik and Červonenkis (VC)-type of collection of functions (essentially, the collection with polynomial covering number).

PROPOSITION 6.2. *Suppose that Assumption 6.1 holds, and (i) the function* $\widetilde{m}_k \in \mathcal{H}(d, \kappa)$ *with* $d \geq q+1$, $s \equiv d + \kappa > 3q/2$*; (ii)* E_k *satisfies* $\mathbb{E}[|E_k|^\zeta] < \infty$ *for some* $\zeta > 4s/(2s-q)$*; (iii)* $\mathbf{W}$ *is component-wise uniformly bounded, and each component of* $\boldsymbol{\mu}$ *has partial derivatives up to order* $q+1$ *that are uniformly bounded, both by an absolute constant* M*;* $q_{\mathrm{L}} = q_{\mathrm{L},n}$ *satisfies* $\log(q_{\mathrm{L}}) \lesssim \log(n)$*, and for all* $k \in [p]$*, for some* $\widetilde{\tau} > 1/4$ *and a sequence* $\{\zeta_n\}_{n \geq 1}$ *satisfying* $q_{\mathrm{L}}\zeta_n = \mathcal{O}(n^{-\widetilde{\tau}})$*,* $\widehat{\boldsymbol{\theta}}_k$ *converges as* $\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k\| = \mathcal{O}_{\mathrm{p}}(\sqrt{q_{\mathrm{L}}}\zeta_n)$*; (iv) the collections of functions* $\mathcal{K}_\ell^{[m]} \equiv \left\{K_\ell^{[m]}\left(\frac{t-\cdot}{c}\right) : t \in \mathbb{R}, c \in \mathbb{R}^+\right\}$ *where* $K_\ell^{[m]}$ *is the* m*th derivative of* K_ℓ*,* $\ell \in [q]$*,* $m \in [d+1]$ *are all VC-type. Then:*

(a) *Assumption 4.1 is satisfied with* $a_{n,1} = k_n^{1/2} + q_{\mathrm{L}}\zeta_n = \mathcal{O}(k_n^{1/2} + n^{-\widetilde{\tau}})$ *for* k_n *as in Section 6.1 (so* $a_{n,1} = \mathcal{O}(n^{-\tau})$ *for some* $\tau > 1/4$*) and* $a_{n,2} = 0$*.*

(b) *For Assumption 4.2, (b1) there exists* $0 < \alpha < 1$ *such that by setting* $\mathcal{D}_{1,n}$ *as the product of the class* $\mathcal{H}_{M'}(q, \alpha)$ *for* M' *large enough (for the covariate* $\mathbf{x}$*) and the class* $\mathcal{D}_{q_{\mathrm{L}}}$ *(for the covariate* $\mathbf{w}$*), and* $\mathcal{D}_{2,n}$ *as the singleton set of the constant function one, the required sequence* $\{D_{n,k}\}_{n \geq 1}$ *exists; (b2) (11), (12), (14) and (18) hold under if we set* $K_1 = M^{q_{\mathrm{L}}} {q_{\mathrm{L}}}^{q_{\mathrm{L}}}$*,* $K_2 = M'$ *with* M*,* M' *large enough,* $\beta = q_{\mathrm{L}}$*,* $\nu = 1 + \alpha/q$*, and if* $q_{\mathrm{L}} \log(n) = \mathcal{O}(a_{n,1}^{-1/\nu} \wedge n^{\frac{1}{1+\nu}})$*. (To be notationally correct in the PLM, for Assumptions 3.3 and 4.2 we should replace* $L_2(F_{\mathbf{X}})$ *by* $L_2(F_{\mathbf{W},\mathbf{X}})$*, where* $F_{\mathbf{W},\mathbf{X}}$ *is the joint distribution function of* $(\mathbf{W}, \mathbf{X})$*, to accommodate the linear covariate* $\mathbf{W}$*.)*

(c) *Assumption 4.3 is satisfied with* $\Delta = \frac{1}{2}(1 - 1/\nu)$ *for* ν *given above and* $R_n = n^{-\frac{1}{1+1/\nu}}$ *(exactly as in Proposition 6.1).*

The proof of Proposition 6.2 appears in Section E.2 in the supplement. For the initial estimator $\widehat{\boldsymbol{\theta}}_k \in \mathbb{R}^{q_{\mathrm{L}}}$ of the regression coefficient suitable for Proposition 6.2 and specifically for the condition on the rate of $\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k\|$, Xie and Huang (2009) (among others) has studied PLM with the dimension $q_{\mathrm{L}} = q_{\mathrm{L},n}$ varying with n. (To be exact, Xie and Huang (2009) focused on the case $q = 1$ for $\widetilde{m}_k$, but extension to higher q was also discussed.) In fact the case considered there is more general: the ambient dimension of $\boldsymbol{\theta}_k$ is ${q_{\mathrm{L}}}^+ = {q_{\mathrm{L},n}}^+$ but only $q_{\mathrm{L}} = q_{\mathrm{L},n}$ components of $\boldsymbol{\theta}_k$ are non-zero. Then Theorem 2 in Xie and Huang (2009) shows that, under regularity conditions that in particular restrict ${q_{\mathrm{L}}}^+ = o(n^{1/2})$, on an event with probability tending to one, all the components of $\widehat{\boldsymbol{\theta}}_k$ corresponding to the zero components of $\boldsymbol{\theta}_k$ are correctly identified as being zero. Thus by focusing on this event, when applying Proposition 6.2 we can simply pretend that the actual dimension of $\boldsymbol{\theta}_k$ is q_{L} (rather than ${q_{\mathrm{L}}}^+$). Moreover, on this event, by the last equation display on p. 691 and the proof of Theorem 1 in Xie and Huang (2009) it is straightforward to show that $\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k\| = \mathcal{O}_{\mathrm{p}}(\sqrt{q_{\mathrm{L}}/n} + M_n^{-S_g})$. Here $M_n = o(n)$ is the number of knots in the polynomial spline to model the non-parametric component $\widetilde{m}_k$, and S_g is related to the degree of smoothness, but in principle with enough smoothness the second term on the right-hand side, $M_n^{-S_g}$, can be made comparable to the first. Thus in this case $\zeta_n = n^{-1/2}$ for our condition on $\|\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k\|$. Consequently by the condition on q_{L}, the number of non-zero components of $\boldsymbol{\theta}_k$ must satisfy $q_{\mathrm{L}} = \mathcal{O}(n^{-\widetilde{\tau}+1/2}) = o(n^{1/4})$ for some $\widetilde{\tau} > 1/4$ (though a stronger condition could apply because of the bound on $q_{\mathrm{L}} \log(n)$ required for Assumption 4.2). On the other hand as we saw earlier the ambient dimension must satisfy (the more relaxed) ${q_{\mathrm{L}}}^+ = o(n^{1/2})$, suggesting that the theory of Xie and Huang (2009) fits nicely into our framework by allowing the ambient dimension ${q_{\mathrm{L}}}^+$ to diverge faster than q_{L}, the intrinsic dimension of interest.

6.3. *Additive model.* The additive model is another variant of (1). Here, the components of the covariate $\mathbf{x} = (x_1, \ldots, x_q)^\top$ separately influences the regression outcome; specifically the location function admits the decomposition $m_k(\mathbf{x}) = m_{k,0} + m_{k,1}(x_1) + \cdots + m_{k,q}(x_q)$ with intercept $m_{k,0}$ and univariate functions $m_{k,\ell}$, $k \in [p]$, $\ell \in [q]$ satisfying $\mathbb{E}m_{k,\ell}(X_\ell) = 0$. For simplicity we again assume $\sigma_k = 1$ identically, so $Y_k = m_{k,0} + m_{k,1}(X_1) + \cdots + m_{k,q}(X_q) + E_k$. We follow the same identification condition as for the PLM, assume that q is fixed and, for simplicity, $\mathcal{X} = [0,1]^q$.

We let $\widehat{m}_{k,\ell}$, $k \in [p]$ and $\ell \in [q]$ be the smooth backfitting Nadaraya-Watson estimator of $m_{k,\ell}$ given by (12) in Mammen, Linton and Nielsen (1999). Specifically, let $\widetilde{m}_{k,\ell}$, $\widehat{p}_\ell$, and $\widehat{p}_{\ell,\ell'}$ be respectively the kernel estimator of $\mathbb{E}[Y_k|X_\ell = x_\ell]$, the density p_ℓ of X_ℓ, and the bivariate density $p_{\ell,\ell'}$ of $(X_\ell, X_{\ell'})$ given by (55), (56) and (57) in Mammen, Linton and Nielsen (1999) (we should modify the aforementioned (55) in a straightforward way to accommodate multiple coordinates $k \in [p]$ to arrive at our $\widetilde{m}_{k,\ell}$), and let $\widetilde{m}_{k,0} = n^{-1}\sum_{i\in[n]} Y_{i,k}$ be the population mean estimator of the kth coordinate. Then, the estimator $\widehat{m}_{k,\ell}$ is

$$\widehat{m}_{k,\ell}(x_\ell) = \widetilde{m}_{k,\ell}(x_\ell) - \textstyle\sum_{\ell' \in [q]\setminus\{\ell\}} \int_{[0,1]} \widehat{m}_{k,\ell'}(x_{\ell'}) \frac{\widehat{p}_{\ell,\ell'}(x_\ell, x_{\ell'})}{\widehat{p}_\ell(x_\ell)} \mathrm{d}x_{\ell'} - \widetilde{m}_{k,0}. \tag{31}$$

The final estimator of m_k is $\widehat{m}_k(\mathbf{x}) = \widetilde{m}_{k,0} + \widehat{m}_{k,1}(x_1) + \cdots + \widehat{m}_{k,q}(x_q)$.

The simple form of the estimator $\widehat{m}_{k,\ell}$ allows for a straightforward analysis. Also for simplicity we will not treat the boundary bias issue and will only focus on the estimation of m_k away from the boundary. Specifically let $\varepsilon > 0$ be an arbitrarily small but fixed absolute constant and let $\mathcal{U}_- = [\varepsilon, 1-\varepsilon]$; then we only consider the estimation of $m_{k,\ell}$ on $\mathcal{U}_-$, and hence the estimation of m_k only on $\mathcal{U}_-^q$. Then the results from the previous sections of this paper should be understood to be based on the subset of the sample $(\mathbf{Y}_i, \mathbf{X}_i)$, $i \in [n]$ satisfying $\mathbf{X}_i \in \mathcal{U}_-^q$.

For Proposition 6.3 below, the required conditions are mostly borrowed from but are also slightly stronger than (B1), (B2'), (B3') and (B4') in Theorem 4 in Mammen, Linton and Nielsen (1999). Let K be the kernel and h_n be the bandwidth used in the kernel estimators $\widetilde{m}_{k,\ell}$, $\widehat{p}_\ell$, $\widehat{p}_{\ell,\ell'}$ (the last using a bivariate product kernel built from K). Analogous to (30) earlier but restricting to univariate function h and $d = 1$, for a real number $\kappa \in (0,1]$ define the norm $\|\cdot\|_{\mathcal{U}_-,1,\kappa}$ as $\|h\|_{\mathcal{U}_-,1,\kappa} = \sup_{x\in\mathcal{U}_-}|h(x)| + \sup_{x\in\mathcal{U}_-}|\dot{h}(x)| + \sup_{x,y\in\mathcal{U}_-:x\neq y}|\dot{h}(x) - \dot{h}(y)|/\|x-y\|^\kappa$ where $\dot{h}$ denotes the derivative of h. Next, analogous to $\mathcal{H}_{M'}(d,\kappa)$ earlier in Sections 6.1 and 6.2, define $\mathcal{H}_{M',\mathcal{U}_-}(1,\kappa)$ as the collection of univariate functions h satisfying $\|h\|_{\mathcal{U}_-,1,\kappa} \le M'$.

PROPOSITION 6.3. *Suppose that (i) the second derivatives of the functions m_k, $k \in [p]$ exist and are continuous; (ii) for some $\theta > 5/2$, $\mathbb{E}(|E_k|^\theta) < \infty$, $\forall k \in [p]$; (iii) the density $f_{\mathbf{X}}$ is bounded away from zero and infinity on the support $\mathcal{X}$, and possesses continuous partial derivatives up to the second order; (iv) the kernel K is symmetric about zero, two-times differentiable with uniformly bounded mth derivative $K^{[m]}$ for $m \in \{0,1,2\}$, has compact support $[-1,1]$, satisfies $\inf_{u>0} \int_{[0,u]} K(v)\mathrm{d}v \geq 0$, and finally the collections of functions $\mathcal{K}^{[m]} \equiv \left\{K^{[m]}\left(\frac{t-\cdot}{c}\right) : t \in \mathbb{R}, c \in \mathbb{R}^+\right\}$, $m \in \{0,1,2\}$ are all VC-type; (v) the bandwidth h_n satisfies $n^{1/5}h_n \to c_h$ for a constant c_h. Then:*

(a) *Assumption 4.1 is satisfied with $a_{n,1} = \log^{1/2}(n)n^{-2/5}$ (where the supremum is understood to be taken over $\mathcal{U}_-^q$) and $a_{n,2} = 0$. Certainly $a_{n,1}, a_{n,2} = \mathcal{O}(n^{-\tau})$ with $\tau > 1/4$.*

(b) *For Assumption 4.2, (b1) by setting $\mathcal{D}_{1,n}$ as the q-time product of $\mathcal{H}_{M',\mathcal{U}_-}(1,\kappa)$ for M' large enough (with the kth factor for x_k) and any $\kappa \in (0, 1/10)$, and $\mathcal{D}_{2,n}$ as the singleton set of the constant function one, the required sequence $\{D_{n,k}\}_{n\geq 1}$ exists; (b2) (11), (12), (14) and (18) hold if we set K_1, K_2, β as large enough positive absolute constants and $\nu = 1 + \kappa$.*

(c) *Assumption 4.3 is satisfied with $\Delta = \frac{1}{2}(1 - 1/\nu)$ for ν given above and $R_n = n^{-\frac{1}{1+1/\nu}}$.*

The proof of Proposition 6.3 appears in Section E.3 in the supplement. In the additive model the required smoothness on the function m_k, namely twice continuous differentiability, is independent of q and is thus potentially much weaker than the corresponding requirement in the non-parametric regression model considered in Section 6.1. This is the appealing "free from the curse of dimensionality" consequence of the additive model.

7. Numerical study. In our simulation study we show that the performance of the residual-based normal scores estimator approaches that of its oracle counterpart as the sample size increases. We consider two cases, first, in Section 7.1, a non-parametric regression model with $q = 2$ and next, in Section 7.2, a partly linear regression variant with $q = 1$ but with the dimension of the linear component as high as $q_{\mathrm{L}} = 10$. In contrast, the naive estimator that does not adjust for the covariate performs significantly worse.

Next we study in Section F (deferred to the supplement) a real data example that examines the students' performance in various disciplines across different countries in the Programme for International Student Assessment (PISA) (OECD (2015)). We first estimate the correlation among the (performance in) different disciplines after adjusting for GDP per capita for the countries through a pure non-parametric regression in Model (1). Then, in the event that only one discipline is observed for a country, we develop procedures to predict the remaining disciplines. Such procedures naturally depend on the assumed dependence structure among the disciplines, and we will compare prediction results based on the assumed Gaussian, Gaussian copula or t-copula structure on the signal component $\mathbf{E}$.

7.1. *Non-parametric regression.* We consider a simple case with $p = 2$ and $q = 2$. We first generate the (unobservable) Gaussian copula component $\mathbf{E}$. We denote the single off-diagonal element in the copula correlation matrix $\mathbf{R}_0$ by ρ_0, and we first let $\rho_0 = 0.5$. We let $\mathbf{E}^{\mathrm{r}}$ follow a bivariate normal distribution with standard Gaussian marginals and correlation ρ_0. Then, we adjust the marginals of $\mathbf{E}^{\mathrm{r}}$ to be the t-distribution with degrees of freedom $\nu_{\mathrm{df}} = 5$ or (for even heavier tail) $\nu_{\mathrm{df}} = 3$, and further normalize the marginals to have unit variance; we let the copula component $\mathbf{E}$ follow the resulting distribution (which clearly has Gaussian copula with copula correlation matrix $\mathbf{R}_0$). We remind the readers that when estimating the scale function σ_k it is the square of Y_k (among others) that serves as the input, and hence our specification of $\mathbf{E}$ is already quite adversarial for our residual-based estimators.

For the location and scale functions, we let

$$\begin{aligned} m_1(x_1,x_2) &= 2(x_1+x_2-1)^2, & \sigma_1(x_1,x_2) &= (5+x_1^2+x_2^2)/5, \\ m_2(x_1,x_2) &= (x_1-x_2)^3, & \sigma_2(x_1,x_2) &= (7+x_1+x_2)/7. \end{aligned}$$

We let the covariate $\mathbf{X}$ follow the uniform distribution on the unit square. Finally, we build the response $\mathbf{Y}$ as in (1) based on independently drawn $\mathbf{E}$ and $\mathbf{X}$.

We compare the performance of the following three estimators of ρ_0:

- The oracle normal scores estimator ρ_n.
- The residual-based normal scores estimator $\widehat{\rho}_n$.
- The naive normal scores estimator ρ_n^{N}.

The naive estimator is built in the same way as the oracle normal scores estimator ρ_n, but using the response sample $\mathbf{Y}_i$, $i \in [n]$ in place of the oracle copula sample $\mathbf{E}_i$, $i \in [n]$, without taking into account the covariate.

We consider sample sizes $n = 50$, 100, 200, 400 or 800, and for each sample size we perform $N = 1\,000$ Monte Carlo simulations. To obtain the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ (and hence the residual-based $\widehat{\rho}_n$), we rely on the **np** package in R. Specifically, we employ a dummy approach, and for each Monte Carlo sample we simply generate the bandwidth automatically with the npregbw routine. We then feed the bandwidth to the npreg routine to obtain the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ without providing any extra guidance.

The performance of the three estimators is summarized in Table 1, where we display the biases and the RMSEs of the deviations between the upper-triangular portions (as will be assumed throughout in this section and Section 7.2) of the estimators and of the true ρ_0 gathered from the $N = 1\,000$ Monte Carlo samples. For clarity of presentation we have multiplied (the biases and the RMSEs of) the deviations by a factor of 100.

		$\nu_{\text{df}} = 5$		$\nu_{\text{df}} = 3$	
n	ρ_n	$\widehat{\rho}_n$	ρ_n^{N}	$\widehat{\rho}_n$	ρ_n^{N}
50	-1.3 (10.6)	-8.5 (15.5)	-5.3 (12.5)	-10.3 (17.4)	-6.7 (13.3)
100	-0.6 (7.7)	-4.4 (9.5)	-4.7 (9.4)	-5.7 (10.2)	-6.2 (10.4)
200	-0.5 (5.4)	-2.4 (6.4)	-4.4 (7.1)	-3.2 (6.6)	-5.8 (8.1)
400	-0.1 (3.7)	-1.4 (4.1)	-3.9 (5.6)	-1.9 (4.4)	-5.3 (6.7)
800	-0.0 (2.6)	-0.7 (2.8)	-3.8 (4.7)	-1.2 (3.0)	-5.2 (5.9)

TABLE 1

The biases and the RMSEs (the latters in parentheses) of the deviations from the true $\rho_0 = 0.5$ associated with the three estimators considered in Section 7.1, based on N=1 000 Monte Carlo simulations. For clarity numbers have been multiplied by a factor of 100.

From the table, we can make a few observations. First, as the sample size n increases, the ratio of the (RMSE of the) deviation between the residual-based estimator $\widehat{\rho}_n$ and the truth to the deviation between the oracle estimator ρ_n and the truth converges to one. This agrees with our theoretically established asymptotic equivalence between the residual-based estimator and its oracle counterpart. Second, the aforementioned convergence is slower under the heavier tail case when $\nu_{\text{df}} = 3$. This is expected because the quality of the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ obviously deteriorates under a larger "error" term E_k. Third, at the smallest sample sizes $n = 50$ and 100, the naive estimator ρ_n^{N} could actually outperform the residual-based estimator. Again, this is expected because for small sample sizes the estimators $\widehat{m}_k$ and $\widehat{\sigma}_k$ are less accurate. However, as n increases, the performance of the naive estimator starts to lag increasingly behind the residual-based estimator.

7.2. *Partly linear variant.* We consider $p = 2$, $q_{\mathrm{L}} = 3$ or $q_{\mathrm{L}} = 10$, and $q = 1$. We first let the Gaussian copula component $\mathbf{E}$ be identical to that in Section 7.1 (so $\rho_0 = 0.5$), except that because here $\sigma_k = 1$ identically we do not adjust the marginals of $\mathbf{E}$ to have unit variances. Then, for the linear component, let the regression coefficient matrix be $\boldsymbol{\Theta} \in \mathbb{R}^{q_{\mathrm{L}} \times 2}$ with

$$\boldsymbol{\Theta} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \end{pmatrix}^{\top} \quad \text{for } q_{\mathrm{L}} = 3,$$

$$\boldsymbol{\Theta} = \begin{pmatrix} 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix}^{\top} \quad \text{for } q_{\mathrm{L}} = 10.$$

Then we let the regression coefficients $\boldsymbol{\theta}_k$, $k \in \{1, 2\}$ be the kth column of $\boldsymbol{\Theta}$. We let the covariate $\mathbf{W}$ for the linear component follow a uniform distribution on the unit hypercube $[0, 1]^{q_{\mathrm{L}}}$. Next, for the non-parametric component $\widetilde{m}_k$ with argument $\mathbf{X} = X$, we let

$$\widetilde{m}_1(x) = 2(x - 0.5)^2, \quad \widetilde{m}_2(x) = 4(x - 0.5)^3.$$

Then, we simply let X follow the uniform distribution on the unit interval. Finally, we build $\mathbf{Y}$ based on independently drawn $\mathbf{E}$, $\mathbf{W}$ and X.

We consider the same three estimators of $\rho_0 = 0.5$ as in Section 7.1 now adapted to the partly linear regression variant. We consider sample sizes $n = 50$, 100, 200, 400 or 800, and for each sample size we perform $N = 1\,000$ Monte Carlo simulations. To obtain the estimators $\widehat{\boldsymbol{\theta}}_k$ and $\widehat{\widetilde{m}}_k$, we employ the **np** package as before. This time, for each Monte Carlo sample we generate the bandwidth automatically with the npplregbw routine and then feed the bandwidth to the npplreg routine, again without providing any extra guidance.

The performance of the three estimators is summarized in Table 2. Our comments on the estimators here also in general agree with the earlier ones in Section 7.1, with the following adaptations. First, the convergence of the residual-based estimator $\widehat{\rho}_n$ to the oracle counterpart under the partly linear regression variant is faster than that under the non-parametric regression model, even with the presence of the additional linear component and without the additional normalization of the marginals of $\mathbf{E}$ to have unit variances (as was done under the non-parametric regression model). This is due to a combination of the simpler, univariate covariate for the non-parametric component considered here and the assumption that $\sigma_k = 1$ identically which eliminates the scale shift in the estimator $\widehat{\mathbf{E}}_i$ defined in Section 3.1. Next, the naive estimator ρ_n^{N} simply fails. It performs slightly better under the heavier tail case when $\nu_{\mathrm{df}} = 3$, but only because here the copula component $\mathbf{E}$ becomes more dominant compared to the perturbation by the covariate. In fact, the naive estimator performs not much better than randomly guessing a positive number (which would on average deviate 0.25 from the target $\rho_0 = 0.5$).

Next we consider estimations at higher correlation ρ_0. We first make a single modification of our models introduced earlier by raising ρ_0 from 0.5 to 0.9, and repeat exactly the earlier estimation procedures. In place of Table 2, the performance of the three estimators ρ_n, $\widehat{\rho}_n$, ρ_n^{N} under the higher ρ_0 setup is summarized in Table 3. As can be seen there, for the residual-based estimator $\widehat{\rho}_n$, its performance converges to that of the oracle estimator ρ_n a bit slower than in the previous, $\rho_0 = 0.5$ case, but it outperforms the naive estimator ρ_n^{N} even more substantially than before.

Finally we consider unbounded covariate $\mathbf{W}$. Note that for the non-parametric components $\widetilde{m}_k$ in Section 6.2 and m_k, σ_k in Section 6.1, the required smoothness conditions dictate that these functions should be bounded. Thus the effect of unbounded covariate is better explored in the covariate $\mathbf{W}$ in the partly linear model. We focus on the higher-dimensional $\mathbf{W}$ case where $q_{\mathrm{L}} = 10$, and consider both $\rho_0 = 0.5$ and $\rho_0 = 0.9$. We simply generate $\mathbf{W}$ as a centered multivariate normal distribution with a 10×10 covariance matrix having unit diagonal

$q_{\rm L} = 3$

		$\nu_{\rm df} = 5$		$\nu_{\rm df} = 3$	
n	ρ_n	$\widehat{\rho}_n$	$\rho_n^{\rm N}$	$\widehat{\rho}_n$	$\rho_n^{\rm N}$
50	-1.3 (10.6)	-3.1 (11.9)	-24.3 (27.8)	-3.4 (11.9)	-20.5 (24.4)
100	-0.6 (7.7)	-1.9 (8.0)	-23.2 (25.1)	-2.1 (8.0)	-19.3 (21.5)
200	-0.5 (5.4)	-1.1 (5.6)	-23.1 (24.0)	-1.4 (5.6)	-19.0 (20.1)
400	-0.1 (3.7)	-0.5 (3.7)	-22.4 (22.9)	-0.7 (3.8)	-18.4 (19.0)
800	-0.0 (2.6)	-0.3 (2.6)	-22.3 (22.5)	-0.4 (2.6)	-18.2 (18.5)

$q_{\rm L} = 10$

		$\nu_{\rm df} = 5$		$\nu_{\rm df} = 3$	
n	ρ_n	$\widehat{\rho}_n$	$\rho_n^{\rm N}$	$\widehat{\rho}_n$	$\rho_n^{\rm N}$
50	-1.3 (10.6)	-4.0 (12.7)	-20.0 (23.8)	-4.7 (13.2)	-17.3 (21.6)
100	-0.6 (7.7)	-2.3 (8.8)	-19.0 (21.1)	-3.0 (9.0)	-16.4 (18.7)
200	-0.5 (5.4)	-1.3 (5.7)	-18.9 (19.9)	-1.8 (5.9)	-16.2 (17.3)
400	-0.1 (3.7)	-0.6 (3.8)	-18.4 (19.0)	-0.9 (3.9)	-15.7 (16.3)
800	-0.0 (2.6)	-0.3 (2.7)	-18.3 (18.6)	-0.5 (2.7)	-15.6 (15.9)

TABLE 2

The biases and the RMSEs (the latters in parentheses) of the deviations from the true $\rho_0 = 0.5$ associated with the three estimators considered in Section 7.1 now adapted to the partly linear regression variant, based on $N = 1\,000$ Monte Carlo simulations. For clarity numbers have been multiplied by a factor of 100. The performance of the oracle estimator ρ_n is obviously the same under $q_{\rm L} = 3$ and $q_{\rm L} = 10$.

$q_{\rm L} = 3$

		$\nu_{\rm df} = 5$		$\nu_{\rm df} = 3$	
n	ρ_n	$\widehat{\rho}_n$	$\rho_n^{\rm N}$	$\widehat{\rho}_n$	$\rho_n^{\rm N}$
50	-1.5 (3.5)	-4.5 (7.1)	-31.9 (33.4)	-4.6 (8.4)	-27.1 (28.6)
100	-0.8 (2.3)	-2.6 (3.8)	-30.5 (31.3)	-2.7 (4.6)	-25.6 (26.4)
200	-0.5 (1.5)	-1.5 (2.4)	-29.8 (30.2)	-1.6 (2.4)	-24.8 (25.2)
400	-0.2 (1.0)	-0.8 (1.4)	-29.2 (29.4)	-0.9 (1.5)	-24.2 (24.4)
800	-0.1 (0.7)	-0.5 (0.9)	-28.9 (29.0)	-0.5 (0.9)	-23.9 (24.0)

$q_{\rm L} = 10$

		$\nu_{\rm df} = 5$		$\nu_{\rm df} = 3$	
n	ρ_n	$\widehat{\rho}_n$	$\rho_n^{\rm N}$	$\widehat{\rho}_n$	$\rho_n^{\rm N}$
50	-1.5 (3.5)	-5.6 (8.4)	-35.4 (36.9)	-5.6 (8.8)	-30.6 (32.2)
100	-0.8 (2.3)	-3.1 (4.5)	-34.1 (34.8)	-3.2 (4.7)	-29.1 (29.9)
200	-0.5 (1.5)	-1.8 (2.6)	-33.6 (34.0)	-1.9 (2.8)	-28.5 (28.9)
400	-0.2 (1.0)	-1.0 (1.6)	-33.1 (33.3)	-1.2 (1.7)	-28.0 (28.2)
800	-0.1 (0.7)	-0.6 (0.9)	-32.8 (32.9)	-0.7 (1.0)	-27.6 (27.7)

TABLE 3

The biases and the RMSEs (the latters in parentheses) of the deviations from the true ρ_0 associated with the same three estimators considered in Table 2, under the same models for Table 2 except that now $\rho_0 = 0.9$, based on $N = 1\,000$ Monte Carlo simulations. For clarity numbers have been multiplied by a factor of 100. The performance of the oracle estimator ρ_n is obviously the same under $q_{\rm L} = 3$ and $q_{\rm L} = 10$.

elements and off-diagonal elements taking a common value 0.5, and repeat the earlier estimation procedures. The performance of the three estimators ρ_n, $\widehat{\rho}_n$, $\rho_n^{\rm N}$ in the current cases is summarized in Table 4. From the table, at least when $\rho_0 = 0.5$, the performance of the residual-based estimator $\widehat{\rho}_n$ is again quite reputable compared to the oracle estimator ρ_n, though (again) a bit worse than that presented in Table 2 for the same $\widehat{\rho}_n$ but under bounded covariate $\mathbf{W}$. At $\rho_0 = 0.9$, the performance of $\widehat{\rho}_n$ is comparable to itself at $\rho_0 = 0.5$. However here ρ_n performs much better than itself at $\rho_0 = 0.5$. The end result is that $\widehat{\rho}_n$ lags further behind ρ_n, though decreasing so at higher sample size. The naive estimator $\rho_n^{\rm N}$ performs abysmally in all cases.

$\rho_0 = 0.5, q_{\mathrm{L}} = 10$

		$\nu_{\mathrm{df}} = 5$		$\nu_{\mathrm{df}} = 3$	
n	ρ_n	$\widehat{\rho}_n$	ρ_n^{N}	$\widehat{\rho}_n$	ρ_n^{N}
50	-1.3 (10.6)	-8.3 (17.8)	-39.2 (41.6)	-8.1 (17.1)	-36.5 (39.0)
100	-0.6 (7.7)	-5.3 (11.3)	-39.9 (41.2)	-5.2 (10.6)	-37.0 (38.5)
200	-0.5 (5.4)	-3.1 (7.3)	-39.3 (40.0)	-3.2 (7.0)	-36.3 (37.0)
400	-0.1 (3.7)	-1.5 (4.3)	-39.4 (39.7)	-1.7 (4.5)	-36.3 (36.6)
800	-0.0 (2.6)	-0.9 (2.9)	-39.4 (39.5)	-1.0 (2.9)	-36.2 (36.4)

$\rho_0 = 0.9, q_{\mathrm{L}} = 10$

		$\nu_{\mathrm{df}} = 5$		$\nu_{\mathrm{df}} = 3$	
n	ρ_n	$\widehat{\rho}_n$	ρ_n^{N}	$\widehat{\rho}_n$	ρ_n^{N}
50	-1.5 (3.5)	-14.2 (20.3)	-71.2 (72.5)	-12.4 (17.9)	-66.0 (67.4)
100	-0.8 (2.3)	-8.4 (11.9)	-71.5 (72.2)	-7.6 (10.9)	-65.9 (66.7)
200	-0.5 (1.5)	-5.0 (7.1)	-70.9 (71.2)	-4.6 (6.3)	-65.0 (65.4)
400	-0.2 (1.0)	-2.8 (3.8)	-70.7 (70.9)	-2.5 (3.5)	-64.7 (64.9)
800	-0.1 (0.7)	-1.6 (2.2)	-70.7 (70.8)	-1.5 (2.0)	-64.5 (64.6)

TABLE 4

The biases and the RMSEs (the latters in parentheses) of the deviations from the true ρ_0 associated with the same three estimators considered in Tables 2 and 3, now for unbounded covariate $\mathbf{W}$, based on $N = 1\,000$ Monte Carlo simulations. For clarity numbers have been multiplied by a factor of 100. The performance of the oracle estimator ρ_n in here and in the corresponding entries in Tables 2 and 3 is obviously the same.

Acknowledgements. We thank a co-editor, an associate editor and three anonymous referees for their insightful comments and for encouraging us to carry out the study in Section 5.

Research was supported by the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409) and by C1-project C16/20/002 from the Research Council KU Leuven.

supplement. Supplement to the paper: "Parametric copula adjusted for non- and semi-parametric regression". The supplement contains most proofs for the paper (Sections A to E) as well as the real data example (Section F).

REFERENCES

AKRITAS, M. G. and VAN KEILEGOM, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.* **28** 549–567.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag New York, Inc.

CHEN, X. and FAN, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* **135** 125–154.

CHEN, X., HUANG, Z. and YI, Y. (2021). Efficient estimation of multivariate semi-nonparametric GARCH filtered copula models. *Journal of Econometrics* **222** 484-501.

CHEN, G. and LOCKHART, R. A. (2001). Weak convergence of the empirical process of residuals in linear models with many parameters. *Ann. Statist.* **29** 748-762.

FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications*. *Monographs on statistics and applied probability series* **66**. Chapman & Hall.

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. and VERAVERBEKE, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics* **42** 1109–1126.

GINÉ, E. and MASON, D. M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.* **35** 1105–1145.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Prague: Academia.

KLAASSEN, C. A. J. and WELLNER, J. A. (1997). Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli* **3** 55–77.

MAMMEN, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24** 307–335.

MAMMEN, E., LINTON, O. and NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443–1490.

MÜLLER, U. U., SCHICK, A. and WEFELMEYER, W. (2009). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statistics & Probability Letters* **79** 957–964.

MÜLLER, U. U., SCHICK, A. and WEFELMEYER, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *Journal of Statistical Planning and Inference* **142** 552–566.

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York.

NEUMEYER, N. and VAN KEILEGOM, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.* **101** 1067–1078.

NEUMEYER, N., OMELKA, M. and ŠÁRKA HUDECOVÁ (2019). A copula approach for dependence modeling in multivariate nonparametric time series. *Journal of Multivariate Analysis* **171** 139 – 162.

OAKES, D. (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics* **3** 343–354.

OECD (2015). Programme for International Student Assessment. https://databank.worldbank.org/source/education-statistics:-learning-outcomes.

OMELKA, M., HUDECOVÁ, ŠÁRKA. and NEUMEYER, N. (2020). Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models.

PAKES, A. and POLLARD, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57** 1027-57.

RUYMGAART, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *Ann. Statist.* **2** 892 – 910.

SEGERS, J., VAN DEN AKKER, R. and WERKER, B. J. M. (2014). Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *Ann. Statist.* **42** 1911–1940.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8** 229–231.

TSUKAHARA, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics* **33** 357-375.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

VERAVERBEKE, N., GIJBELS, I. and OMELKA, M. (2014). Preadjusted non-parametric estimation of a conditional distribution function. *J. R. Stat. Soc. Ser. B* **76** 399–438.

VERAVERBEKE, N., OMELKA, M. and GIJBELS, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics* **38** 766–780.

XIE, H. and HUANG, J. (2009). SCAD-penalized regression in high-dimensional partially linear models. *Ann. Statist.* **37** 673–696.

ZHAO, Y. and GENEST, C. (2019). Inference for elliptical copula multivariate response regression models. *Electron. J. Statist.* **13** 911–984.

ZHAO, Y., GIJBELS, I. and VAN KEILEGOM, I. (2020). Inference for covariate-adjusted semiparametric Gaussian copula model using residual ranks. *Bernoulli* **26** 2815–2846.