UNIVERSITY of York

This is a repository copy of *Inference for semiparametric Gaussian copula model adjusted for linear regression using residual ranks*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/157509/

Version: Accepted Version

Article:
Zhao, Yue, Gijbels, Irène and van Keilegom, Ingrid (2020) Inference for semiparametric Gaussian copula model adjusted for linear regression using residual ranks. Bernoulli. pp. 2815-2846. ISSN 1350-7265

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Inference for semiparametric Gaussian copula model adjusted for linear regression using residual ranks

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. E-mail:* [*] yue.zhao@kuleuven.be; [**] 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. E-mail:* irene.gijbels@kuleuven.be

We investigate the inference of the copula parameter in the semiparametric Gaussian copula model when the copula component, subject to the influence of a covariate, is only indirectly observed as the response in a linear regression model. We consider estimators based on residual ranks instead of the usual but unobservable oracle ranks. We first study two such estimators for the copula correlation matrix, one via inversion of Spearman's rho and the other via normal scores rank correlation estimator. We show that these estimators are asymptotically equivalent to their counterparts based on the oracle ranks. Then, for the copula correlation matrix under constrained parametrizations, we show that the classical one-step estimator in conjunction with the residual ranks remains semiparametrically efficient for estimating the copula parameter. The accuracy of the estimators based on residual ranks is confirmed by simulation studies.

MSC 2010 subject classifications: Primary 62F12, 62G20; secondary 62G30, 62H20.
Keywords: Gaussian copula, normal scores rank correlation estimator, residual rank, semiparametric efficiency, Spearman's rho, U-process.

1. Introduction

1.1. Background

Let $\mathbf{E} = (E_1, \ldots, E_p)^\top \in \mathbb{R}^p$ be a random vector; we assume throughout that E_k, $k \in \{1, \ldots, p\}$ has absolutely continuous marginal distribution function F_k, and $\mathbf{E}$ has joint distribution function F. Sklar's theorem (e.g., Theorem 2.10.9 in [28]) states that the dependence structure of $\mathbf{E}$ can be uniquely described by its associated copula C, via

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_p(x_p)), \quad \mathbf{x} = (x_1, \ldots, x_p)^\top \in \overline{\mathbb{R}}^p$$

where $\overline{\mathbb{R}}$ denotes the extended real line. By the "inverse" Sklar's theorem (e.g., Corollary 2.10.10 in [28]),

$$C(\mathbf{u}) = F(F_1^{\leftarrow}(u_1), \ldots, F_p^{\leftarrow}(u_p)), \quad \mathbf{u} = (u_1, \ldots, u_p)^\top \in [0,1]^p, \tag{1.1}$$

where for $k \in \{1, \ldots, p\}$, we denote the left-continuous inverse of F_k by

$$F_k^{\leftarrow}(t) = \inf\{x : F_k(x) \geq t\}, \quad t \in [0,1]. \tag{1.2}$$

The copula C is equivalently the joint distribution function of the transformed random vector $(F_1(E_1), \ldots, F_p(E_p))^\top$, and it remains unchanged if (univariate) strictly increasing transformations are applied to the individual components of $\mathbf{E}$ (see, e.g., Theorem 2.4.3 in [28]). Copulas provide a modular approach to multivariate modeling, in that the dependence structure of a multivariate distribution can be summarized by a copula, irrespective of the behaviors of its marginals. A collection of multivariate distributions in $\mathbb{R}^p$ is called a semiparametric copula model if they share a copula smoothly parametrized by a finite-dimensional Euclidean parameter $\boldsymbol{\theta}$ while their marginals range over all p-tuples of absolutely continuous univariate distribution functions.

This paper focuses specifically on the semiparametric Gaussian copula model (or simply Gaussian copula model), where the parametric copula is restricted to be a Gaussian copula. We say that the random vector $\mathbf{E} \in \mathbb{R}^p$ has a Gaussian copula if for a *copula correlation matrix* $\mathbf{R} \in \mathbb{R}^{p \times p}$ that uniquely characterizes the copula of $\mathbf{E}$,

$$(\Phi^{-1}(F_1(E_1)), \ldots, \Phi^{-1}(F_p(E_p)))^\top \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R}). \tag{1.3}$$

Throughout the paper the symbol "$\sim$" denotes equality in distribution and Φ^{-1} denotes the standard normal quantile function. By (1.3), the Gaussian copula model obviously contains all multivariate normal distributions, and hence further encompasses all distributions that can be obtained from multivariate normal distributions through strictly increasing transformations of the marginals. Combining (1.1) and (1.3), simple algebra yields that the copula $C_{\mathbf{R}}$ associated with $\mathbf{E}$ is

$$C_{\mathbf{R}}(\mathbf{u}) = \Phi_{\mathbf{R}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)), \quad \mathbf{u} \in [0,1]^p. \tag{1.4}$$

Here the function $\Phi_{\mathbf{R}}$ is the distribution function of the $\mathcal{N}_p(\mathbf{0}, \mathbf{R})$ distribution.

To discuss semiparametric efficiency we will further treat the copula correlation matrix as being parametrized through $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$ for the *copula parameter* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\boldsymbol{\Theta} \subset \mathbb{R}^d$ is some parameter space, and d is regarded as the intrinsic dimension of $\mathbf{R}$. For brevity, we will often suppress the dependence of $\mathbf{R}$ on $\boldsymbol{\theta}$. An important special case is when each element of the upper-triangular portion of $\mathbf{R}$ is a free parameter, so $d = p(p-1)/2$; $\mathbf{R}$ is then said to obey an unrestricted model. In contrast to the unrestricted model, by a constrained parametrization we mean $\mathbf{R}(\cdot)$ is (usually) a continuously differentiable $\mathbb{R}^{p \times p}$-valued function and $\boldsymbol{\Theta}$ is (usually) within a lower-dimensional Euclidean space (so $d < p(p-1)/2$). Even under a constrained parametrization $\mathbf{R}(\boldsymbol{\theta})$, in practice one often first constructs a preliminary estimator $\widetilde{\mathbf{R}}$ of $\mathbf{R}$ without assuming any particular structure

of $\mathbf{R}$ (i.e., as if one were working with an unrestricted model for $\mathbf{R}$; see Eq. (3.22) for an example where $\widetilde{\mathbf{R}} = \widehat{\mathbf{R}}_n$ for $\widehat{\mathbf{R}}_n$ to be introduced shortly); then, one can estimate the copula parameter $\boldsymbol{\theta}$ through, e.g., the pseudo-likelihood estimation or the one-step method to be discussed later, based on the parametrization $\mathbf{R}(\boldsymbol{\theta})$ with the preliminary estimator $\widetilde{\mathbf{R}}$ as input.

The study of the Gaussian copula model has enjoyed continued interest in the last couple of decades. In the classical fixed-dimensional setting, a major research focus has been the (asymptotically) semiparametrically efficient estimation of $\mathbf{R}$ or of the copula parameter $\boldsymbol{\theta}$ when the marginal distribution functions F_k are left unspecified as infinite dimensional nuisance parameters, that is, when the problem is fully semiparametric. Progressively along the work of, e.g., [6, 21, 18, 34], we have now a fairly complete understanding of the tight semiparametric lower bound for $\mathbf{R}$ or $\boldsymbol{\theta}$. Typically, the semiparametric lower bound is larger than its counterpart within a corresponding parametric problem with known margins, due to the loss of information through the unknown margins. [34] also studied an estimator that achieves the semiparametric lower bound by adapting the *one-step method* (see, e.g., Section 25.8 in [38]).

More recently, the Gaussian copula model has also received much attention in a graphical model setting. Recall that the locations of the zeros of the precision matrix $\mathbf{S}$ (which is the inverse of the covariance matrix) of a Gaussian random vector $\mathbf{G}$ encode the conditional independence structure of the Gaussian graphical model associated with $\mathbf{G}$ [24]. Through (1.3), it is readily seen that such an encoding of the Gaussian graphical model via the precision matrix naturally extends to the Gaussian copula model [26, 39]. More precisely, for a random vector $\mathbf{E}$ that has a Gaussian copula with (copula) precision matrix $\mathbf{S} = \mathbf{R}^{-1}$, its two components E_k and E_ℓ are conditionally independent given all the other coordinates if and only if the (k, ℓ)th element of $\mathbf{S}$ is zero. Thus, inferring the graphical model associated with $\mathbf{E}$ is intrinsically tied to the problem of finding an accurate estimator of $\mathbf{S}$. In high dimensions, it is typically assumed that (the parametrization of) $\mathbf{R}$ corresponds to a sparse inverse but is otherwise unrestricted. Finding an accurate estimator of $\mathbf{R}$ in this context is important because in order to estimate $\mathbf{S}$, an estimator of $\mathbf{R}$ almost always serves as the input to sparsity-inducing optimization programs.

1.2. Research objective: semiparametric Gaussian copula model adjusted for linear regression

Suppose that a $p \times 1$ random vector $\mathbf{E} = (E_1, \ldots, E_p)^\top$ has a Gaussian copula with copula correlation matrix $\mathbf{R}$. We will refer to the case when the sample of $\mathbf{E}$ is directly observable as the *ordinary* (semiparametric) Gaussian copula model. Now, let a $p \times 1$ response vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ and a $q \times 1$ covariate vector $\mathbf{X} = (X_1, \ldots, X_q)^\top$ be linked to $\mathbf{E}$ through the linear regression model

$$\mathbf{Y} = \mathbf{B}^\top \mathbf{X} + \mathbf{E}, \tag{1.5}$$

where $\mathbf{B}$ is a $q \times p$ unknown coefficient matrix. We assume throughout that $\mathbf{X}$ is independent from $\mathbf{E}$. As before, we will still focus on the statistical inference problem for the

dependence structure summarized in $\mathbf{R}$. However, in contrast to the ordinary Gaussian copula model, assume that we do not directly observe the sample of $\mathbf{E}$. What is at our disposal instead is a sample of the perturbed response $\mathbf{Y}$ and the covariate $\mathbf{X}$. Such a scenario arises frequently in practice, because we cannot realistically expect to always be able to directly observe the signal of interest, namely $\mathbf{E}$. In this case, to correctly infer about $\mathbf{R}$, it is essential to adjust for the covariate. To see this more precisely, observe that the covariances of $\mathbf{Y}$ and $\mathbf{E}$ are related as follows:

$$\operatorname{cov}(\mathbf{Y}) = \mathbf{B}^\top \operatorname{cov}(\mathbf{X})\mathbf{B} + \operatorname{cov}(\mathbf{E}). \tag{1.6}$$

Although our interest is $\mathbf{R}$, not $\operatorname{cov}(\mathbf{E})$, this simple relationship nevertheless demonstrates that the dependence structure reflected by $\mathbf{Y}$ is not the same as that reflected by $\mathbf{E}$. Naively taking the copula correlation matrix calculated from the sample of $\mathbf{Y}$ as the estimator of the true $\mathbf{R}$ can cause significant error, as our simulation study will demonstrate later.

Recall that the p components of $\mathbf{E}$ have univariate distribution functions $F_1, \ldots, F_p$ respectively. Moreover, let the covariate $\mathbf{X}$ have distribution function $F_{\mathbf{X}}$. Then, the finite-dimensional parameter $\mathbf{B}$, the infinite-dimensional distribution functions $F_1, \ldots, F_p$ and $F_{\mathbf{X}}$ are all nuisance parameters in Model (1.5). At times we will refer to Model (1.5) colloquially as the "regression setting". To estimate $\mathbf{R}$ in this context, we will rely on the ranks of the residuals in a preliminary linear regression step, or simply *residual ranks*, from which we construct rank-based estimators of $\mathbf{R}$. Moreover, we will show that to a large extent the complication introduced by the additional regression structure in fact does not affect the (semiparametrically efficient) estimation of $\mathbf{R}$ or of the copula parameter $\boldsymbol{\theta}$, even for estimator with a divergent score function; this theme will be repeated throughout this paper.

Finally, as one referee pointed out, Gaussian copulas are often employed to model heavy-tailed or highly skewed distributions. In these cases, $\operatorname{cov}(\mathbf{E})$ may no longer be an appropriate dependence measure or may not even exist. In contrast, $\mathbf{R}$ is a copula property and is a dependence measure that is entirely free from the specification of the marginal distributions, and could thrive in these situations. However, in the presence of covariate, whether an additive noise model such as our Model (1.5) is reasonable for a particular application involving heavy-tailed or skewed $\mathbf{E}$ should be carefully examined. Moreover, alternative method to introduce a copula structure into linear regression, such as the *Gaussian copula regression model* to be described shortly, also exists. Nevertheless, when Model (1.5) holds, both our theoretical result and simulation study, in particular the last example in Section 5.1, demonstrate that our residual-based procedure can work very well even for extremely heavy-tailed or skewed $\mathbf{E}$.

1.3. Relation to existing studies

1.3.1. Copula models adjusted for covariate

Gaussian copulas have been extended in a number of ways to incorporate a covariate, and we briefly review some representative work in this area. In what follows, again

$\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ denotes a $p \times 1$ response vector and $\mathbf{X} = (X_1, \ldots, X_q)^\top$ denotes a $q \times 1$ covariate vector, although not necessarily in a linear regression setting. We will not consider parametrically specified marginals (e.g., [27, 35]) or the case where it is $\mathbf{R}$ that depends on the covariate (e.g., [3]).

Recently, [7, 11] treated the linear regression model (1.5) but they restricted $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$ to be jointly multivariate normal. It is further assumed that none of the sample of $\mathbf{Y}, \mathbf{X}, \mathbf{E}$ is available. Instead, [7, 11] considered a sample of $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ linked to the (unavailable) sample of $(\mathbf{Y}, \mathbf{X})$ through strictly increasing transformations of their components. To be precise, for arbitrary integer $r \geq 1$, let $\mathcal{G}_r$ be the collection of functions $g(x_1, \ldots, x_r) = (g_1(x_1), \ldots, g_r(x_r))^\top : \mathbb{R}^r \to \mathbb{R}^r$ such that $g_1, \ldots, g_r$ are all strictly increasing univariate functions. Then [7, 11] studied the estimation of $\mathbf{B}$ in (1.5) from a sample of $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$, where $\widetilde{\mathbf{Y}} = g(\mathbf{Y})$ and $\widetilde{\mathbf{X}} = h(\mathbf{X})$ for fixed but *unknown* marginal transformations $g \in \mathcal{G}_p$ and $h \in \mathcal{G}_q$. [7] called this model the Gaussian copula regression model, because the observable $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ has a Gaussian copula.

Note that our Model (1.5) and that of [7, 11] extend the linear regression model (1.5) in quite distinct ways: we impose a copula structure on $\mathbf{E}$ while the latter imposes a copula structure on $(\mathbf{Y}, \mathbf{X})$. This distinction reflects the different motivations of the two approaches: in our model, the copula component $\mathbf{E}$ is the object of interest and it is the object of interest that has been perturbed. In contrast, in [7, 11] the (Gaussian) $\mathbf{E}$ is a nuisance for estimating $\mathbf{B}$. Technically, our model is arguably more challenging because we never observe a sample of the Gaussian copula component $\mathbf{E}$ directly (whereas [7, 11] do observe a sample of the Gaussian copula component $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$), but only a sample that has been perturbed by the covariate. Showing that we can still obtain accurate estimators of $\mathbf{R}$ despite such perturbations is at the core of our analysis.

Beyond the Gaussian copula model, there are a number of papers dealing with copula inference, not necessarily in a linear regression setting, based on some form of ranks of residuals adjusted for a covariate. Here we briefly address the representative papers [13, 14] in the *conditional copula* framework. They assumed that a scalar covariate X affects only the marginal distributions but not the dependence structure of a bivariate response $\mathbf{Y}$, and the copula of $\mathbf{Y}$ when conditioning on X is a totally unspecified (i.e., nonparametric) copula. [13, 14] studied the resulting *empirical copula process* based on the estimated conditional distribution of $\mathbf{Y}$ given X. It was shown that, under suitable conditions, the empirical copula process based on such estimated quantities has the same asymptotics as the case when the conditional distribution is known. The conditional copula *model* certainly encompasses our linear regression model (1.5) as a special case. On the other hand, while estimating, an estimator of (an element of) $\mathbf{R}$ is often obtained through a multivariate rank order statistics which, after centering at $\mathbf{R}$ and scaling by $\sqrt{n}$, is then equivalent to the integral of a particular score function (specified in Section 2.2.1) against a bivariate empirical copula process. However, when the score function is unbounded (as in the case for the normal scores rank correlation estimator), the weak convergence of the empirical copula process alone is not even sufficient to establish the asymptotic normality of the resulting integral (see, e.g., Section 3.1 in [4]), and hence existing literature on conditional copula does not address the objectives of our

paper. Moreover, we will provide explicit rates of convergence.

A recent manuscript [30] considered pseudo-likelihood estimation for the copula parameter (of a given parametric copula) adjusted for a parametric location-scale model, and they allowed the score function to have non-trivial divergence. However, our new result on the rates of the remainder terms in a decomposition of the residual empirical process in Section 3.1, and our U-process approach in Section 3.2, enabled us to obtain much finer convergence rates.

Having discussed related literature beyond the Gaussian copula model (not necessarily in a linear regression setting), we would like to point out that our technical analysis can easily accommodate different choices of score functions, and so only partially relies on the Gaussian copula assumption. For instance, we expect that our result in Section 3.1 can be adapted to the framework of [30] without too much difficulty. However, we believe that the prevalence of Gaussian copulas (because of the possibilities for Spearman's rho-based estimation and to achieve semiparametric efficiency, and the connection to graphical models) warrants a dedicated treatment, and the specific linear regression form also allows for potential generalization to high-dimensional settings, as we will briefly discuss in Section F in the supplement.

1.3.2. Efficient estimation in linear regression

In the regression setting, semiparametric lower bound for the copula parameter was briefly addressed in Section 6.2 in [34], but no detailed calculations (regarding differentiability in quadratic mean, projections, etc.) were given; moreover the paper did not study semiparametrically efficient estimator in this setting. We will address these problems in full details in Section 4. Semiparametrically efficient estimation of the regression coefficient vector in a univariate response linear regression model has been covered at least as early as in [5]. In all these studies, one crucial ingredient that makes semiparametrically efficient estimation possible is the realization that the tangent spaces generated by the regression coefficient and by the error term are orthogonal. However, in contrast to our setting, in [5] the regression coefficient, rather than the error term, was the object of interest; moreover, [5] did not consider (residual) *rank*-based procedures for estimation. Thus, [5] does not address the objectives of our paper either.

1.4. Organization of the paper

Section 2 formally introduces our model and estimation procedures. Section 3 presents the asymptotics of two estimators of the copula correlation matrix $\mathbf{R}$ based on the residual ranks, one via inversion of Spearman's rho in Section 3.2, and another via normal scores rank correlation estimator in Section 3.3. The latter study is based on new result on residual empirical process presented in Section 3.1. As has already been mentioned in Section 1.1, and as we will elaborate in Section 3.4, these estimators are stepping stones to obtain estimators of the copula parameter $\boldsymbol{\theta}$.

Next, Section 4 derives the semiparametric lower bound for estimating the copula parameter $\boldsymbol{\theta}$, and then shows that the one-step method in conjunction with the residual

ranks yields an estimator that achieves this lower bound. Section 5 presents a simulation study comparing the estimators based on the residual ranks to the oracle estimators (to be introduced in Section 2.2.1). Section 6 concludes and Section F in the supplement provides a short discussion on the high-dimensional extension of the current project. Due to space constraint most technical analysis and some supporting materials are deferred to the supplement and a technical report [15], as will be explained throughout the main text. Sections in the supplement are labeled by Roman alphabet.

1.5. Notations

We will always use $\lesssim$ to denote inequality that holds with an absolute constant (that is, independent of sample sizes, dimensions, and any parameter we consider) as the factor. For any positive integer a, we use $[a]$ to denote the set $\{1, \ldots, a\}$.

For a matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, we use $(\mathbf{A})_{kk'}$ to denote its (k, k')th element. For a two-dimensional array of numbers $a_{kk'}$, $k, k' \in [p]$, we use $[a_{kk'}]_{k,k'\in[p]}$ to denote a matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ with $(\mathbf{A})_{kk'} = a_{kk'}$. When acting on vectors and matrices, $\|\cdot\|$ denotes the Euclidean norm, and when acting on functions, $\|\cdot\|_{L_\infty}$ denotes the supremum norm.

2. Model setup and (residual) rank-based estimation

2.1. Formal model setup

Recall that the copula correlation matrix is parametrized as $\mathbf{\Theta} \to \mathbb{R}^{p\times p} : \boldsymbol{\theta} \to \mathbf{R}(\boldsymbol{\theta})$ for some parameter space $\mathbf{\Theta} \in \mathbb{R}^d$. We say that $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$ has a joint law $\mathrm{P}_{\boldsymbol{\theta},\mathbf{B},F_1,\ldots,F_p,F_{\mathbf{X}}}$ if the following conditions hold:

(i) $\mathbf{E}$ has a Gaussian copula $C_{\boldsymbol{\theta}} \equiv C_{\mathbf{R}(\boldsymbol{\theta})}$, where $C_{\mathbf{R}(\boldsymbol{\theta})}$ is given in (1.4) with copula correlation matrix $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$; the inverse $\mathbf{S}(\boldsymbol{\theta})$ of $\mathbf{R}(\boldsymbol{\theta})$ exists. The mapping $\boldsymbol{\theta} \to \mathbf{R}(\boldsymbol{\theta})$ is one-to-one.
(ii) For each $k \in [p]$, the kth component E_k of $\mathbf{E}$ has absolutely continuous marginal distribution function F_k that corresponds to a marginal density function f_k.
(iii) $\mathbf{X}$ and $\mathbf{E}$ are independent.
(iv) The covariate $\mathbf{X}$ has an absolutely continuous joint distribution function $F_{\mathbf{X}}$ in $\mathbb{R}^q$, corresponding to a density $f_{\mathbf{X}}$. In addition, $\mathbb{E}[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^\top]$ is non-singular.
(v) Equation (1.5) holds (with $\mathbf{B} \in \mathbb{R}^{q\times p}$).

Throughout the paper we assume that both p and q are fixed; the only exception occurs in Section F in the supplement where we discuss high-dimensional generalization. The non-singularity requirement on $\mathbb{E}[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^\top]$ is the usual identification condition for linear regression up to the intercept term, see, e.g., Example 3 in [5]. With this condition (each column of) $\mathbf{B}$ is identified. Then, because both $\mathbf{Y}$ and $\mathbf{X}$ are observed (and hence identified), the copula component $\mathbf{E}$, and consequently the law P, are identified as well. Moreover, with condition (iv) above, the covariate $\mathbf{X}$ cannot contain an intercept term;

consequently, no location constraint is placed on the marginals $F_1, \ldots, F_p$. Note that a nonzero intercept term is not identifiable because a change in the intercept could equally well be a change in the marginals.

For brevity henceforth we abbreviate $\mathrm{P}_{\boldsymbol{\theta},\mathbf{B},F_1,\ldots,F_p,F_{\mathbf{X}}}$ simply as P.

2.2. Rank-based estimation procedures

Because copulas are invariant to strictly increasing marginal transformations, it is desirable for an estimator of $\mathbf{R}$ to maintain such invariance. Therefore we concentrate on rank-based methods. (For non-rank-based approaches, see, e.g., [9] for a method based on parametric sieves that also achieves semiparametric efficiency.) In Section 2.2.1 we review some rank-based estimators of $\mathbf{R}$ in the ordinary Gaussian copula model, and in Section 2.2.2 we develop their counterparts in Model (1.5). 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 $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,q})^\top$.

2.2.1. Oracle estimators for the ordinary Gaussian copula model

If the sample $\mathbf{E}_i$, $i \in [n]$ of the copula component $\mathbf{E}$ were directly observable, a $\sqrt{n}$-consistent and asymptotically normal estimator of $\mathbf{R}$ can be derived in a number of classical ways. For each $k \in [p]$, we define the empirical marginal distribution function for the kth coordinate of $\mathbf{E}$, and its rescaled version, as

$$F_{n,k}(t) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{E_{i,k} \leq t\}, \quad F^{\mathrm{r}}_{n,k}(t) = \frac{1}{n+1} \sum_{i \in [n]} \mathbb{1}\{E_{i,k} \leq t\}, \quad t \in \mathbb{R}$$

respectively. We need the rescaled version so that applying Φ^{-1} to it later will always result in finite values. We will refer to $F_{n,k}(E_{i,k})$ and $F^{\mathrm{r}}_{n,k}(E_{i,k})$ as the (normalized) *oracle ranks*. Henceforth the qualifier "oracle" denotes quantities that could be computed in the ordinary Gaussian copula model, and the qualifier "normalized," which we will omit almost throughout, refers to the multiplication by the factor $1/n$ or $1/(n+1)$.

The first estimator of $\mathbf{R} \equiv [r_{kk'}]_{k,k' \in [p]}$ we consider is the (oracle) normal scores rank correlation estimator $\mathbf{R}_n = [r_{n,kk'}]_{k,k' \in [p]}$, defined in, e.g., Eq. (7) on p. 123 in [17] as

$$r_{n,kk'} = \frac{\phi_n}{n} \sum_{i \in [n]} \Phi^{-1}(F^{\mathrm{r}}_{n,k}(E_{i,k})) \Phi^{-1}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})), \quad \forall k, k' \in [p]. \tag{2.1}$$

Here $\phi_n = [n^{-1} \sum_{i \in [n]} \{\Phi^{-1}(\frac{i}{n+1})\}^2]^{-1} = 1 + \mathcal{O}(n^{-1} \log(n))$ is a deterministic correction factor. The correction ϕ_n is asymptotically insignificant, but with it the matrices $\mathbf{R}_n$, and analogously $\widehat{\mathbf{R}}_n$ given later in (2.5), have unit diagonal elements, and they become genuine correlation matrices. We can write $\mathbf{R}_n$ explicitly in matrix form as

$$\mathbf{R}_n = \frac{\phi_n}{n} \sum_{i \in [n]} \mathbf{Z}^{(n)}_i \mathbf{Z}^{(n)\top}_i, \quad \mathbf{Z}^{(n)}_i \equiv (\Phi^{-1}(F^{\mathrm{r}}_{n,1}(E_{i,1})), \ldots, \Phi^{-1}(F^{\mathrm{r}}_{n,p}(E_{i,p})))^\top. \tag{2.2}$$

The elements of $\mathbf{R}_n$ belong to multivariate rank order statistics (in this case, with score function Φ^{-1}) that are common in the literature; see [16, 33] for some early references. The asymptotic distribution of $\sqrt{n}(\mathbf{R}_n - \mathbf{R})$ is a zero mean (matrix) Gaussian; see the discussion following Theorem 3.4. In the unrestricted model for $\mathbf{R}$, the estimator $\mathbf{R}_n$ coincides with the pseudo-likelihood estimator, and it is semiparametrically efficient in the ordinary Gaussian copula model. We refer to Section 3.4 for more details, and point out that, beyond the unrestricted model, the semiparametric efficiency of the pseudo-likelihood estimator may or may not hold for a given constrained parametrization $\mathbf{R}(\boldsymbol{\theta})$.

Alternatively, we can estimate $\mathbf{R}$ by inversion of Spearman's rho. As described, e.g., in [20, 23], the population version of Spearman's rho between the kth and k'th coordinates of $\mathbf{E}$, for any $k, k' \in [p]$, is defined as

$$\rho_{kk'} = 3\left\{\mathbb{P}((E_{1,k} - E_{2,k})(E_{1,k'} - E_{3,k'}) > 0) - \mathbb{P}((E_{1,k} - E_{2,k})(E_{1,k'} - E_{3,k'}) < 0)\right\}.$$

Then, let the (oracle) estimator of $\rho_{kk'}$ be

$$\begin{aligned}\rho_{n,kk'} &= \frac{12n}{n^2-1}\sum_{i\in[n]}\left\{F_{n,k}(E_{i,k}) - \frac{n+1}{2n}\right\}\left\{F_{n,k'}(E_{i,k'}) - \frac{n+1}{2n}\right\} \\ &= 1 - 6\frac{n}{n^2-1}\sum_{i\in[n]}\left\{F_{n,k}(E_{i,k}) - F_{n,k'}(E_{i,k'})\right\}^2; \end{aligned} \tag{2.3}$$

see, e.g., Eqs. (11) and (12) on p. 124 in [17]. The elements $\rho_{n,kk'}$, $k, k' \in [p]$ again belong to multivariate rank order statistics (in this case, with score function $\sqrt{12}(u-1/2)$). The asymptotic distribution of $\sqrt{n}[\rho_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$ is a zero mean (matrix) Gaussian, and will be hinted at following Theorem 3.3.

Furthermore, the following well-known equality connect the elements of the copula correlation matrix $\mathbf{R} = [r_{kk'}]_{k,k'\in[p]}$ to the elements of Spearman's rho matrix $[\rho_{kk'}]_{k,k'\in[p]}$:

$$r_{kk'} = 2\sin\left(\frac{\pi}{6}\rho_{kk'}\right), \quad \forall k, k' \in [p];$$

see, e.g., Corollary 4.1 in [19]. A plug-in estimator of $\mathbf{R}$ via inversion of Spearman's rho is then given by $\mathbf{R}_n^\rho = [r_{n,kk'}^\rho]_{k,k'\in[p]}$ with

$$r_{n,kk'}^\rho = 2\sin\left(\frac{\pi}{6}\rho_{n,kk'}\right), \quad \forall k, k' \in [p].$$

By the Delta method, the asymptotic distribution of $\sqrt{n}(\mathbf{R}_n^\rho - \mathbf{R})$ is also a zero mean (matrix) Gaussian.

2.2.2. Residual-based estimators for the Gaussian copula model adjusted for linear regression

In Model (1.5), the sample of the copula component $\mathbf{E}$, and therefore the oracle ranks, are not directly observable. Instead our sample consists of $(\mathbf{Y}_i, \mathbf{X}_i)$, $i \in [n]$. Therefore, we rely on this sample to estimate the sample of $\mathbf{E}$ and the oracle ranks. We let $\widehat{\mathbf{B}} = \widehat{\mathbf{B}}^{(n)}$

be an estimator (sequence) of $\mathbf{B}$. We denote the kth column of $\mathbf{B}$ and $\widehat{\mathbf{B}}$ by $\mathbf{B}_k$ and $\widehat{\mathbf{B}}_k$ respectively. For sample size $n \in \{1, 2, \dots\}$ and $i \in [n]$, let $\widehat{\mathbf{E}}_i = \widehat{\mathbf{E}}_i^{(n)} = (\widehat{E}_{i,1}, \dots, \widehat{E}_{i,q})^\top$ be the residual of the ith sample defined as

$$\widehat{E}_{i,k} = Y_{i,k} - \widehat{\mathbf{B}}_k^\top \mathbf{X}_i = E_{i,k} - (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i, \quad \forall k \in [p]. \tag{2.4}$$

We regard $\widehat{\mathbf{E}}_i$ as an estimator of $\mathbf{E}_i$. For brevity we will suppress the dependence of $\widehat{\mathbf{E}}_i$ on n (which it inherits from $\widehat{\mathbf{B}}$). Then, for each $k \in [p]$, we define the (empirical marginal) residual distribution function for the kth coordinate of $\mathbf{E}$, and its rescaled version, as

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

respectively. We regard $\widehat{F}_{n,k}$ and $\widehat{F}^{\mathrm{r}}_{n,k}$ as the estimator of $F_{n,k}$ and $F^{\mathrm{r}}_{n,k}$ respectively. We will refer to $\widehat{F}_{n,k}(\widehat{E}_{i,k})$ and $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$ as the (normalized) residual ranks.

Now, we let $\widehat{\mathbf{R}}_n = [\widehat{r}_{n,kk'}]_{k,k' \in [p]}$ be the normal scores rank correlation estimator of $\mathbf{R}$ based on the residual ranks given by

$$\widehat{r}_{n,kk'} = \frac{\phi_n}{n} \sum_{i \in [n]} \Phi^{-1}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})) \Phi^{-1}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'})), \quad \forall k, k' \in [p], \tag{2.5}$$

where ϕ_n is the same as in (2.1). Similar to $\mathbf{R}_n$, we can write $\widehat{\mathbf{R}}_n$ in matrix form as

$$\widehat{\mathbf{R}}_n = \frac{\phi_n}{n} \sum_{i \in [n]} \widehat{\mathbf{Z}}_i \widehat{\mathbf{Z}}_i^\top, \quad \widehat{\mathbf{Z}}_i = \widehat{\mathbf{Z}}_i^{(n)} \equiv (\Phi^{-1}(\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1})), \dots, \Phi^{-1}(\widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p})))^\top. \tag{2.6}$$

Note that $\widehat{\mathbf{R}}_n$ is obtained from (2.1), or in matrix form from (2.2), through substitution of the oracle ranks by the residual ranks.

Next, let the estimator of Spearman's rho $\rho_{kk'}$ based on the residual ranks be

$$\widehat{\rho}_{n,kk'} = 1 - 6 \frac{n}{n^2 - 1} \sum_{i \in [n]} \left\{ \widehat{F}_{n,k}(\widehat{E}_{i,k}) - \widehat{F}_{n,k'}(\widehat{E}_{i,k'}) \right\}^2. \tag{2.7}$$

Similar to (2.5), $\widehat{\rho}_{n,kk'}$ is obtained from (2.3) through substitution by the residual ranks. Then, a plug-in estimator of $\mathbf{R}$ via inversion of Spearman's rho, now based on the residual ranks, is given by $\widehat{\mathbf{R}}^\rho_n = [\widehat{r}^\rho_{n,kk'}]_{k,k' \in [p]}$ with

$$\widehat{r}^\rho_{n,kk'} = 2 \sin\left(\frac{\pi}{6} \widehat{\rho}_{n,kk'} \right), \quad \forall k, k' \in [p]. \tag{2.8}$$

The remainder of this paper will focus on showing the closeness of the residual-based estimators to their oracle counterparts (e.g., $\widehat{\mathbf{R}}_n$ to $\mathbf{R}_n$, and $\widehat{\mathbf{R}}^\rho_n$ to $\mathbf{R}^\rho_n$), and the implication for efficient estimation under the regression setting.

3. Asymptotics of estimators of copula correlation matrix based on residual ranks

In Section 2 we have introduced two estimators of the copula correlation matrix $\mathbf{R}$ based on the residual ranks. In this section we present asymptotic results for these estimators. Some preliminary results regarding (i) residual-based estimation of F_k and (ii) the residual ranks are presented in Section 3.1. These results concern the individual coordinates of $\mathbf{E}$, in fact do not rely on the Gaussian copula dependence structure (i.e., condition (i) under the law P described at the beginning of Section 2), and could be of independent interest. The most important results of Section 3 are presented in Sections 3.2 and 3.3. First, Section 3.2 treats the estimator $\widehat{\mathbf{R}}_n^{\rho}$ in (2.8) via inversion of Spearman's rho. Then, Section 3.3 treats the normal scores rank correlation estimator $\widehat{\mathbf{R}}_n$ in (2.5). Finally, Section 3.4 provides a short discussion, including on how, from the estimators of $\mathbf{R}$, we could obtain estimators of the copula parameter $\boldsymbol{\theta}$ under constrained parametrizations $\mathbf{R}(\boldsymbol{\theta})$.

Because the actual proofs for Sections 3.2 and 3.3 are long, we very briefly summarize our strategies here. Our proof for the estimator $\widehat{\mathbf{R}}_n^{\rho}$ in Section 3.2 proceeds through an order-3 U-process that arises from the expansion of the residual-based estimator around the oracle one. This U-process approach is novel as compared to the existing methods for conditional copula (e.g., [14, 30]), though parallels could be drawn with the analysis of R-estimators via an order-2 U-process; see, e.g., [29]. For analyzing $\widehat{\mathbf{R}}_n^{\rho}$, the U-process approach is advantageous because the Hoeffding decomposition provides clear, hierarchical rates for the different terms in the decomposition. However, the estimator $\widehat{\mathbf{R}}_n$ considered in Section 3.3 is more difficult to analyze with the same approach due to the unbounded score function Φ^{-1}. Here we fall back on our results established in Section 3.1 for the distance between the residual ranks and the oracle ones in order to bound the distance between $\widehat{\mathbf{R}}_n$ and $\mathbf{R}_n$. This approach is somewhat analogous to (but is much more involved than) how, when no regression structure is present, [33] derived the distribution for the distance between an oracle multivariate rank order statistic and its expectation.

We do not require a specific form of the estimator $\widehat{\mathbf{B}}$, but in order for our estimators of $\mathbf{R}$ to be consistent, naturally we require $\widehat{\mathbf{B}}$ to be consistent for $\mathbf{B}$. Even though in linear regression a $n^{-1/2}$ convergence rate for $\widehat{\mathbf{B}}$ is typical, a different convergence rate is possible. For instance, Rousseeuw's robust *least median of squares regression* introduced in [32], at least initially, yields a slower, $n^{-1/3}$ convergence rate for $\widehat{\mathbf{B}}$. As another example, if we consider a case where the complexity of the law P may increase with the sample size n, then $\delta_{\mathbf{B},n}$ could reflect the "effective" convergence rate of $\widehat{\mathbf{B}}$ in terms of n; in Section F in the supplement, this topic is briefly considered when specifically the ambient dimensions p and q vary with n. In any case, we thought it would be interesting to pinpoint the effect of the quality of $\widehat{\mathbf{B}}$ on the convergence of our estimators. To this end, and to demonstrate the flexibility of our estimators, we simply set the convergence rate of $\widehat{\mathbf{B}}$ as in Assumption 3.1. This assumption and some other general assumptions for this section are collected together below. We let $\mathcal{X}_n$ denote the σ-field generated by the collection of random vectors $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i\in[n]}$. All probabilities

are stated under the (arbitrary but fixed) law P unless stated otherwise.

Assumption 3.1. *$\widehat{\mathbf{B}}$ is $\mathcal{X}_n$-measurable, and under the law* P *is a $\delta_{\mathbf{B},n}^{-1}$-consistent estimator of $\mathbf{B}$, that is, $\|\widehat{\mathbf{B}}-\mathbf{B}\| = \mathcal{O}_{\mathrm{p}}(\delta_{\mathbf{B},n})$. Here, $\delta_{\mathbf{B},n}$, $n \geq 1$ is a sequence of deterministic constants that is required to satisfy $\log(n)/n \leq \delta_{\mathbf{B},n} = o(1)$.*

Here the sole purpose for lower bounding $\delta_{\mathbf{B},n}$ is to simplify certain expressions later.

Assumption 3.2. *For each $k \in [p]$, the marginal density function f_k satisfies $\|f_k\|_{L_\infty} < \infty$, and is Lipschitz continuous with Lipschitz constant L_k on $\mathbb{R}$.*

Assumption 3.3. *Under the law* P*,*

(i) the covariate $\mathbf{X}$ satisfies $\mathbb{E}[\|\mathbf{X}\|^2] < \infty$, i.e., the second moment of $\|\mathbf{X}\|$ is finite.
(ii) If $n^{-1/2} = o(\delta_{\mathbf{B},n})$ in Assumption 3.1, i.e., if $\widehat{\mathbf{B}}$ is not $\sqrt{n}$-consistent, then additionally $\max_{i\in[n]} \|\mathbf{X}_i\|\delta_{\mathbf{B},n} = o_{\mathrm{p}}(1)$.

We remark that Assumption 3.2 holds for densities with polynomial decay of the form

$$f_k(t) = \frac{a-1}{2}\frac{1}{(1+|t|)^a} \tag{3.1}$$

when $a > 1$, and the (skewed) Gamma density with shape parameter at least 2.

3.1. Results on the residual ranks

Proposition 3.1 presents a uniform bound for the difference between the empirical process based on the residual $\widehat{E}_{i,k}$'s, or simply the *residual empirical process*, and the empirical process based on the unobserved $E_{i,k}$'s. Based on Proposition 3.1, Proposition 3.2 further establishes a uniform bound for the difference between the residual ranks $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$ and the corresponding unobserved oracle ranks $F^{\mathrm{r}}_{n,k}(E_{i,k})$.

We first introduce some quantities that will appear in these propositions. For $k \in [p]$, let $\widehat{E}_k = E_k - (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}$, and write $\mathbb{E}\left[\mathbb{1}\left\{E_k \leq \cdot + (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}\right\} | \mathcal{X}_n\right] = \mathbb{E}\left[\mathbb{1}\left\{\widehat{E}_k \leq \cdot\right\} | \mathcal{X}_n\right]$ which equals the conditional probability $\mathbb{P}(\widehat{E}_k \leq \cdot | \mathcal{X}_n)$. Then, let the "oscillation-like" remainder term (e.g., p. 243 in [25] or Lemma 1 in [1]) common in the analysis of residual empirical process be, for $t \in \mathbb{R}$,

$$\begin{aligned} r_{1n,k}(t) &= \widehat{F}_{n,k}(t) - F_{n,k}(t) - \mathbb{P}(\widehat{E}_k \leq t | \mathcal{X}_n) + F_k(t) \\ &= \frac{1}{n}\sum_{i\in[n]} \left\{ \mathbb{1}\{\widehat{E}_{i,k} \leq t\} - \mathbb{1}\{E_{i,k} \leq t\} - \mathbb{P}(\widehat{E}_k \leq t | \mathcal{X}_n) + \mathbb{P}(E_k \leq t) \right\}. \end{aligned} \tag{3.2}$$

Analogous to (3.2), and for analyzing the estimator $\widehat{\mathbf{R}}_n$, let $r^{\mathrm{r}}_{1n,k}$ be obtained by replacing $\widehat{F}_{n,k}$ in $r_{1n,k}$ with $\widehat{F}^{\mathrm{r}}_{n,k}$. Next, let the additional remainder terms be

$$r_{2n,k}(t) = \mathbb{P}(\widehat{E}_k \leq t | \mathcal{X}_n) - F_k(t) - f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}], \quad t \in \mathbb{R}, \tag{3.3}$$

$$r_{3n,k,i} = F_{n,k}(\widehat{E}_{i,k}) - F_k(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k}) + F_k(E_{i,k}), \tag{3.4}$$

$$\begin{aligned} r_{4n,k,i} = {} & F_k(\widehat{E}_{i,k}) - F_k(E_{i,k}) + f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top (\mathbf{X}_i - \mathbb{E}[\mathbf{X}]) \\ & + f_k(\widehat{E}_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]. \end{aligned} \tag{3.5}$$

Proposition 3.1. *Under the law* P*, for all* $n \geq 1$*,* $k \in [p]$ *and* $t \in \mathbb{R}$ *the equalities*

$$\widehat{F}_{n,k}(t) - F_k(t) = F_{n,k}(t) - F_k(t) + f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}] + r_{1n,k}(t) + r_{2n,k}(t), \tag{3.6}$$

$$\widehat{F}^{\mathrm{r}}_{n,k}(t) - F_k(t) = F_{n,k}(t) - F_k(t) + f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}] + r^{\mathrm{r}}_{1n,k}(t) + r_{2n,k}(t) \tag{3.7}$$

hold. If in addition Assumptions 3.1, 3.2 and 3.3(i) hold, then for all $k \in [p]$*,*

$$\sup_{t \in \mathbb{R}} \frac{\max\{|r_{1n,k}(t)|, |r^{\mathrm{r}}_{1n,k}(t)|\}}{\log^{1/2}\left(\delta^{-1}_{\mathbf{B},n}\right) n^{-1/2} \left\{ f_k^{1/2}(t) \delta^{1/2}_{\mathbf{B},n} + \delta_{\mathbf{B},n} \right\} + \log(n)\, n^{-1}} = \mathcal{O}_{\mathrm{p}}(1), \tag{3.8}$$

$$\sup_{t \in \mathbb{R}} |r_{2n,k}(t)| = \mathcal{O}_{\mathrm{p}}\left(\delta^2_{\mathbf{B},n}\right). \tag{3.9}$$

Proof. The proof is deferred to Section A.1 in the supplement. □

We provide a few technical remarks about Proposition 3.1; readers more interested in our development of residual rank-based techniques are encouraged to jump to the paragraph just above Proposition 3.2.

- First, the decomposition (3.6) consists of the leading terms that are the first three terms on the right hand side, and the remainder terms $r_{1n,k}$ and $r_{2n,k}$. Decompositions similar to this (but not the rates for the reminder terms) are common in the literature on residual empirical process; see, e.g., Theorem 1 in [1], Theorem 1 in [8], and Lemma 5 in [30], which consider different extensions of the classical result of, e.g., Theorem 2.3.1 in [22]. (Even in a linear regression setting, as is the case in [8], sometimes the term $n^{1/2} f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]$ is replaced by the asymptotically equivalent quantity $n^{-1/2} f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \sum_{i \in [n]} \mathbf{X}_i$. We point out that [8] in fact studied fixed design, with fixed $\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top$ being the rows of a $n \times q$ design matrix, so additional care is necessary for a truly precise comparison.)
- Next, by replacing $f_k(t)$ with $\|f_k\|_{L_\infty} < \infty$, we can simplify (3.8) to a weaker form as

$$\sup_{t \in \mathbb{R}} |r_{1n,k}(t)| = \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(\delta^{-1}_{\mathbf{B},n}) n^{-1/2} \delta^{1/2}_{\mathbf{B},n} \right). \tag{3.10}$$

  In particular, when $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$, (3.10) implies $\sup_{t \in \mathbb{R}} |r_{1n,k}(t)| = \mathcal{O}_{\mathrm{p}}(\log^{1/2}(n) n^{-3/4})$. In this case, $\sup_{t \in \mathbb{R}} |r_{2n,k}(t)| = \mathcal{O}_{\mathrm{p}}\left(n^{-1}\right)$. Such rates on the remainder terms are

strictly faster than $o_{\mathrm{p}}(n^{-1/2})$, and improve upon existing results in the literature. For instance, Theorem 1 in [1], Theorem 1 in [8] and Lemma 5 in [30], when taken at face value, simply state the remainder terms as being $o_{\mathrm{p}}(n^{-1/2})$. We stress that such faster rates on the remainder terms are important for establishing the rate (and hence equivalence) between the residual-based estimator $\widehat{\mathbf{R}}_n$ and the oracle $\mathbf{R}_n$ in Theorem 3.4.

- Finally, for t such that the term $\left\{\log(\delta_{\mathbf{B},n}^{-1})n^{-1}f_k(t)\delta_{\mathbf{B},n}\right\}^{1/2}$ in the denominator on the left hand side of (3.8) dominates (i.e., when $f_k(t) \geq \delta_{\mathbf{B},n} \vee (\log^2(n)\log^{-1}(\delta_{\mathbf{B},n}^{-1})n^{-1}\delta_{\mathbf{B},n}^{-1})$), the bound in (3.8) on $r_{1n,k}(t)$ and $r_{1n,k}^{\mathrm{r}}(t)$ is additionally weighted by an approximate standard deviation factor $f_k^{1/2}(t)$. (For t such that the aforementioned term no longer dominates, the bound on $r_{1n,k}(t)$ and $r_{1n,k}^{\mathrm{r}}(t)$ is $\mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(\delta_{\mathbf{B},n}^{-1})n^{-1/2}\delta_{\mathbf{B},n} + \log(n)\,n^{-1}\right)$, which when $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$ is very fast at $\mathcal{O}_{\mathrm{p}}(\log(n)\,n^{-1})$.) This idea of weighing further sharpens the bound (3.8) compared to the simplified version (3.10), and is a prominent feature that further distinguishes our result from the literature (e.g., [1, 8]). It is also quite similar to how the convergence of the standard empirical process $\sqrt{n}(F_{n,k} - F_k)$ is often measured under a stronger, weighted metric $\|\cdot/w\|_{L_\infty}$ where the weight function w can almost be as small as the standard deviation factor $(F_k \wedge (1-F_k))^{1/2}$ of $\sqrt{n}(F_{n,k} - F_k)$; see, e.g., Lemma C.4 in the supplement. When integrating over a score function that becomes unbounded toward the boundary (such as the function Φ^{-1} in the normal scores rank correlation estimator $\widehat{\mathbf{R}}_n$), the weighted version (3.8) tames the unboundedness of the score function and so allows for establishing a faster convergence rate for the resulting integral, as can be seen from the derivation of (A.65) and (A.73) in the supplement. Again, this is similar to how the convergence of the standard empirical process $\sqrt{n}(F_{n,k} - F_k)$ under a weighted metric can be helpful in analyzing the classical multivariate rank order statistics with unbounded score functions (see, e.g., how [33] relies on its Lemma 4.1).

The leading term in (3.6) and (3.7) that reflects the uncertainty $\widehat{\mathbf{B}} - \mathbf{B}$ in estimating $\mathbf{B}$, namely the term $f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]$, is proportional to $\mathbb{E}[\mathbf{X}]$. Interestingly, in the corresponding term in (3.11) in Proposition 3.2 below, $\mathbb{E}[\mathbf{X}]$ is replaced by $\mathbf{X}_i - \mathbb{E}[\mathbf{X}]$. Thus for the residual ranks this term will behave as if the covariate $\mathbf{X}$ were centered. (Intuitively, a common average relating to all the residuals should not affect the *ranks* of the residuals to first order.) This centering effect will negate the leading contribution of $\widehat{\mathbf{B}} - \mathbf{B}$ to $\widehat{\mathbf{R}}_n - \mathbf{R}_n$, allowing us to conclude the asymptotic equivalence between $\widehat{\mathbf{R}}_n$ and $\mathbf{R}_n$ in Theorem 3.4, as long as the rate of $\widehat{\mathbf{B}} - \mathbf{B}$ is not much slower than $n^{-1/2}$.

Proposition 3.2. *Under the law* P*, for each* $n \geq 1$*,* $k \in [p]$ *and* $i \in [n]$ *the equality*

$$\begin{aligned}\widehat{F}_{n,k}^{\mathrm{r}}(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k}) = &-f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top(\mathbf{X}_i - \mathbb{E}[\mathbf{X}]) \\ &+ r_{1n,k}^{\mathrm{r}}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i} \qquad (3.11)\end{aligned}$$

holds. (Eq. (3.11) *also holds with the substitutions of* $\widehat{F}_{n,k}^{\mathrm{r}}$ *and* $r_{1n,k}^{\mathrm{r}}$ *by* $\widehat{F}_{n,k}$ *and* $r_{1n,k}$

respectively.) If in addition Assumptions 3.1, 3.2 and 3.3 hold, then for all $k \in [p]$,

$$\max_{i\in[n]} \left| \frac{r_{3n,k,i}}{\log^{1/2}(n)\, n^{-1/2} \left\{ f_k^{1/2}(E_{i,k}) \|\mathbf{X}_i\|^{1/2} \delta_{\mathbf{B},n}^{1/2} + \|\mathbf{X}_i\| \delta_{\mathbf{B},n} \right\} + \log(n)\, n^{-1}} \right| = \mathcal{O}_{\mathrm{p}}(1), \tag{3.12}$$

$$\max_{i\in[n]} \left| \frac{r_{4n,k,i}}{\|\mathbf{X}_i\|(\|\mathbf{X}_i\| + \|\mathbb{E}[\mathbf{X}]\|)\delta_{\mathbf{B},n}^2} \right| = \mathcal{O}_{\mathrm{p}}(1). \tag{3.13}$$

Proof. The proof is deferred to Section A.1 in the supplement. □

Two additional remainder terms, $r_{3n,k,i}$ and $r_{4n,k,i}$, appear in the decomposition (3.11). Similar to the remark on the term $r_{1n,k,i}$ below Proposition 3.1, the term $r_{3n,k,i}$ (i) has a rate strictly faster than $o_{\mathrm{p}}(n^{-1/2})$ (so long as $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-\epsilon})$ for some $\epsilon > 0$) and (ii) is again weighted by an approximate standard deviation factor $f_k^{1/2}$. We have already commented on the utilities of the same features on $r_{1n,k,i}$ below Proposition 3.1, and these comments carry over to $r_{3n,k,i}$.

3.2. Inversion of Spearman's rho

Theorem 3.3. *Assume the law* P*, and Assumptions 3.1, 3.2 and 3.3. Then,*

$$\begin{aligned} &\sqrt{n}\,(\widehat{\rho}_{n,kk'} - \rho_{n,kk'}) \\ &= \mathcal{O}_{\mathrm{p}}\left(n^{1/2}\delta_{\mathbf{B},n}^2 + \delta_{\mathbf{B},n} + n^{-1/2}\delta_{\mathbf{B},n}^{1/2}\log(n) + n^{-1/2} \right), \quad \forall k, k' \in [p]. \end{aligned} \tag{3.14}$$

Thus, if furthermore $\delta_{\mathbf{B},n} = o(n^{-1/4})$, then $\sqrt{n}\,(\widehat{\rho}_{n,kk'} - \rho_{n,kk'}) = o_{\mathrm{p}}(1), \forall k, k' \in [p]$.

Proof. The proof is deferred to Section A.2 in the supplement. □

Now, assume that the conditions of Theorem 3.3 hold, and the uncertainty $\widehat{\mathbf{B}} - \mathbf{B}$ is not too large (precisely, when $\delta_{\mathbf{B},n} = o(n^{-1/4})$, which in particular allows a rate slower than $n^{-1/2}$). Then the theorem immediately yields that the asymptotic distributions of the matrices $\sqrt{n}[\widehat{\rho}_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$ relating to Spearman's rho and its oracle counterpart $\sqrt{n}[\rho_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$ coincide. As is well-known, the asymptotic distribution of the latter matrix is a zero mean (matrix) Gaussian with a correlation structure given by, e.g., Theorem 2.2 in [10]. From these observations and the Delta method, the asymptotically normal distribution of $\sqrt{n}(\widehat{\mathbf{R}}_n^{\rho} - \mathbf{R})$, where $\widehat{\mathbf{R}}_n^{\rho}$ is the plug-in estimator given by (2.8) via inversion of Spearman's rho, will follow and will coincide with that of $\sqrt{n}(\mathbf{R}_n^{\rho} - \mathbf{R})$, where $\mathbf{R}_n^{\rho}$ is the oracle counterpart of $\widehat{\mathbf{R}}_n^{\rho}$. This is our first concrete instance where the complication introduced by the additional regression structure does not affect the estimation of $\mathbf{R}$, a theme already mentioned in Section 1.2. Moreover, when $\delta_{\mathbf{B},n} = n^{-1/2}$, the usual parametric rate, our U-process approach in fact yields via (3.14) that $\widehat{\rho}_{n,kk'} - \rho_{n,kk'}$ converges much faster at $\mathcal{O}_{\mathrm{p}}\left(n^{-1}\right)$.

3.3. Normal scores rank correlation estimator

The analysis of the normal scores rank correlation estimator $\widehat{\mathbf{R}}_n$ in (2.5) will involve two additional constants τ and γ. Their particular values are irrelevant for the construction of $\widehat{\mathbf{R}}_n$, but they relate to Assumptions 3.4 and 3.5, the quantity δ_n in (3.17), and the rate in Theorem 3.4 below.

Assumption 3.4. *Under the law* P*, the covariate* $\mathbf{X}$ *satisfies* $\mathbb{E}[\|\mathbf{X}\|^2] < \infty$*. Moreover,* $\max_{i\in[n]} \|\mathbf{X}_i\| \delta_{\mathbf{B},n} = \mathcal{O}_{\mathrm{p}}(n^{-\tau})$ *for some constant* $\tau > 1/4$ *.*

Assumption 3.4 is stronger than its counterpart Assumption 3.3 for analyzing the estimator $\widehat{\mathbf{R}}_n^{\rho}$ via inversion of Spearman's rho, and it necessitates

$$\delta_{\mathbf{B},n} = \mathcal{O}(n^{-\tau}). \tag{3.15}$$

Typically, $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$; then, by reasoning similar to the proof of Lemma A.3 in the supplement, for any $\tau \leq 1/2$, Assumption 3.4 is implied by the condition $\mathbb{E}\left[\|\mathbf{X}\|^{\frac{1}{1/2-\tau}}\right] < \infty$. Thus if we would further like $\tau = 1/4 + \epsilon$ for some small $\epsilon > 0$, then Assumption 3.4 is implied by a condition slightly stronger than the finite fourth moment of $\|\mathbf{X}\|$.

Assumption 3.5. *There exists a constant* γ *satisfying* $1/2 < \gamma < \min\{2\tau, 1\}$ *(note that such a* γ *exists when* $\tau > 1/4$*) such that, for each* $k \in [p]$*, and as* $\delta \downarrow 0$*,*

$$\sup_{u\in(\delta,1-\delta)} \frac{f_k \circ F_k^{\leftarrow}(u)}{u \wedge (1-u)} = o\left(\delta^{-\frac{1}{\gamma}\tau}\right). \tag{3.16}$$

If Assumption 3.5 is satisfied for some γ, define a partition of the interval $(0,1)$ into $A_1 = A_1^{(n)} = (0, n^{-\gamma}] \cup [1-n^{-\gamma}, 1)$ and $A_2 = A_2^{(n)} = (n^{-\gamma}, 1-n^{-\gamma})$. Then, we introduce the non-decreasing sequence δ_n, $n \geq 1$ as

$$\delta_n = \max_{k\in[p]} \int_{A_2} \left\{ \frac{f_k \circ F_k^{\leftarrow}(u)}{u \wedge (1-u)} \right\}^2 \left\{ \left|\Phi^{-1}(u)\right| \vee 1 \right\}^2 \mathrm{d}u. \tag{3.17}$$

We elaborate on Assumption 3.5 and the quantity δ_n in (3.17). Assumption 3.5 requires that at any quantile $u \in (0,1)$, the marginal density f_k cannot be too large compared to the value $u \wedge (1-u)$ which measures how close the quantile level u is to the boundary of the distribution of E_k. We first provide an example where this assumption fails. If we can take $\tau = 1/2$ (as can be done if $\widehat{\mathbf{B}}$ is $\sqrt{n}$-consistent and the support of $\mathbf{X}$ is bounded), then for E_k following the uniform distribution, the left hand side of (3.16) is of the order $1/\delta$, while if we take γ according to Assumption 3.5, then for the right hand side $1/\delta^{\tau/\gamma} = 1/\delta^{1/(2\gamma)}$ which is of a smaller order because $\gamma > 1/2$, violating Assumption 3.5. Why is the uniform marginal distribution problematic? Recall the decomposition (3.11) of the distance between the residual ranks and the oracle ranks; the first term on the right hand side of (3.11) tells us that a part of this distance is weighted by the marginal

density f_k. The density of a uniform distribution does not decay toward the boundary; this, when coupled with an unbounded score function Φ^{-1}, leads to too large a distance between the residual ranks and the oracle ranks for our current techniques to handle.

On the other hand, Assumption 3.5 should hold for any distribution whose density decays reasonably fast toward the boundary. Indeed it holds for the normal density, densities with polynomial decay of the form already introduced in (3.1) when $a > 1$, and again the Gamma density with shape parameter at least 2. Moreover, in the first two of these cases, and in the last case with the shape parameter larger than 2, the integral on the right hand side of (3.17) over the entire interval $(0, 1)$ evaluates to a finite constant that then upper bounds all of δ_n, $n \geq 1$. In fact, it is not too strong to make $\delta_n = \mathcal{O}(1)$ a requirement, because as we will also discuss later in Remark 2, δ_n is quite similar to the left hand side of (4.6), and the latter quantity should be finite for us to carry out the lower bound analysis in Section 4.

We introduce

$$\Delta_n = \log(n)\, n^{1/2-\gamma} + n^{1/4}\delta_{\mathbf{B},n}\delta_n^{1/2} + \log^{3/2}(n)\, \delta_{\mathbf{B},n}^{1/2} + \log^{1/2}(n)\, n^{1/2}\delta_{\mathbf{B},n}^2\delta_n. \tag{3.18}$$

Theorem 3.4. *Assume the law* P*, and Assumptions 3.1, 3.2, 3.4 and 3.5. Then,*

$$\sqrt{n}(\widehat{r}_{n,kk'} - r_{n,kk'}) = \mathcal{O}_{\mathrm{p}}(\Delta_n), \quad \forall k, k' \in [p]. \tag{3.19}$$

Thus, if furthermore $\delta_n = o(n^{2\tau-1/2}\log^{-1/2}(n))$*, then* $\Delta_n = o(1)$ *and (component-wise)*

$$\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R}_n) = o_{\mathrm{p}}(1). \tag{3.20}$$

Proof. The proof is deferred to Section A.3 in the supplement. □

Similar to the discussion following Theorem 3.3 on the estimator $\widehat{\mathbf{R}}_n^{\rho}$, Theorem 3.4 immediately yields that, when the conditions of the theorem (including the one on δ_n) hold — with Assumption 3.4 necessitating that the uncertainty $\widehat{\mathbf{B}} - \mathbf{B}$ is not too large, but again allowing a rate slower than $n^{-1/2}$ — the asymptotic distributions of the matrices $\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R}) = \sqrt{n}[\widehat{r}_{n,kk'} - r_{kk'}]_{k,k'\in[p]}$ and its oracle counterpart $\sqrt{n}(\mathbf{R}_n - \mathbf{R}) = \sqrt{n}[r_{n,kk'} - r_{kk'}]_{k,k'\in[p]}$ coincide. This is our second concrete instance where the additional regression structure does not affect the estimation of $\mathbf{R}$, this time for estimator with a divergent score function. Theorem 3.1 in [21], relying on an earlier result in [33], established the asymptotic normality and covariance of $\sqrt{n}(\mathbf{R}_n - \mathbf{R})$, from which the same properties of $\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R})$ could follow.

3.4. Discussion

From Section 3 alone, it appears that by our current technique, the residual-based estimator via inversion of Spearman's rho converges faster to its oracle counterpart at the rate (3.14) than the residual-based normal scores estimator to its oracle counterpart at the rate (3.19). For instance, when $\delta_{\mathbf{B},n} = n^{-1/2}$, we have already mentioned earlier that

$\widehat{\rho}_{n,kk'} - \rho_{n,kk'} = \mathcal{O}_{\mathrm{p}}(n^{-1})$, while at best $\widehat{r}_{n,kk'} - r_{n,kk'} = \mathcal{O}_{\mathrm{p}}(\log^{3/2}(n)n^{-3/4})$. The former estimator also requires weaker conditions than the latter, cf. Assumption 3.3 for the former and Assumptions 3.4 and 3.5 for the latter. However, we wish to point out that the *asymptotic covariance* of $\mathbf{R}_n^{\rho}$ is larger than that of $\mathbf{R}_n$ (indeed the latter is semiparametrically efficient under the unrestricted model), and hence the asymptotic covariance of $\widehat{\mathbf{R}}_n^{\rho}$ is larger than that of $\widehat{\mathbf{R}}_n$. Therefore, asymptotically, $\widehat{\mathbf{R}}_n$ should still be preferred over $\widehat{\mathbf{R}}_n^{\rho}$. Moreover, the pseudo-likelihood method and the one-step estimator will more closely involve the estimator $\widehat{\mathbf{R}}_n$, as we will see in this section and in Section 4.2.

Both $\widehat{\mathbf{R}}_n$ and $\widehat{\mathbf{R}}_n^{\rho}$ are natural estimators of the parametrization $\mathbf{R}(\boldsymbol{\theta})$ in the unrestricted model (defined in Section 1.1) where the copula parameter $\boldsymbol{\theta}$ simply corresponds to the elements in the upper-triangular portion of $\mathbf{R}$: we can just estimate $\boldsymbol{\theta}$ by the corresponding elements in $\widehat{\mathbf{R}}_n$ or $\widehat{\mathbf{R}}_n^{\rho}$. (In the case of $\widehat{\mathbf{R}}_n$, this intuitive conclusion is more formally justified by the pseudo-likelihood method described below.) Both $\widehat{\mathbf{R}}_n$ and $\widehat{\mathbf{R}}_n^{\rho}$ can also serve as the starting point to estimate $\boldsymbol{\theta}$ under constrained parametrizations $\mathbf{R}(\boldsymbol{\theta})$; we refer to p. 2 in [34] for a brief summary of existing methods.

Of these, the pseudo-likelihood estimation (PLE) method (see [12] for an early reference) is particularly interesting. Here, as in a parametric case, we estimate $\boldsymbol{\theta}$ by the maximizer of the likelihood function corresponding to the density $c_{\boldsymbol{\theta}}$ of the copula distribution $C_{\mathbf{R}} = C_{\mathbf{R}(\boldsymbol{\theta})}$ in (1.4). However, because the sample $(F_1(E_{i,1}), \ldots, F_p(E_{i,p}))^{\top}$, $i \in [n]$ from the distribution $C_{\mathbf{R}(\boldsymbol{\theta})}$ is unobservable, we replace it by the residual ranks $(\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1}), \ldots, \widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p}))^{\top}$. Formally, given "sample covariance matrix" $\widehat{\mathbf{R}} \in \mathbb{R}^{p\times p}$, let the function $\mathbb{M}(\cdot; \widehat{\mathbf{R}}) : \mathbb{R}^{p\times p} \to \mathbb{R}$ be

$$\mathbb{M}(\mathbf{R}'; \widehat{\mathbf{R}}) = -\frac{p\log(2\pi)}{2} - \frac{1}{2}\log(\det \mathbf{R}') - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}'^{-1}\widehat{\mathbf{R}}\right). \tag{3.21}$$

Using $\log c_{\boldsymbol{\theta}}$ given in (B.1) in the supplement, and the aforementioned substitutions by the residual ranks, the likelihood function at parameter value $\boldsymbol{\theta}'$ becomes, in terms of $\mathbb{M}$,

$$\frac{1}{n}\sum_{i\in[n]} \log c_{\boldsymbol{\theta}'}(\widehat{F}_{n,1}(\widehat{E}_{i,1}), \ldots, \widehat{F}_{n,p}(\widehat{E}_{i,p})) = \mathbb{M}(\mathbf{R}(\boldsymbol{\theta}'); \phi_n^{-1}\widehat{\mathbf{R}}_n).$$

Then, the pseudo-likelihood estimator is defined as the M-estimator

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} = \operatorname*{argmax}_{\boldsymbol{\theta}'\in\boldsymbol{\Theta}} \mathbb{M}(\mathbf{R}(\boldsymbol{\theta}'); \widehat{\mathbf{R}}_n), \tag{3.22}$$

where we note the appearance of the natural estimator $\widehat{\mathbf{R}}_n$ under the unrestricted model. (In (3.22), we have intentionally left out one factor of ϕ_n so only $\widehat{\mathbf{R}}_n$ remains.) Consider a p-variate normal distribution with fixed zero mean and with unknown covariance as the parameter. Then the quantity $\mathbb{M}(\mathbf{R}'; \widehat{\mathbf{R}})$ in (3.21) is precisely the value of the likelihood function of this distribution evaluated at the parameter value $\mathbf{R}'$ against the "sample covariance matrix" $\widehat{\mathbf{R}}$. As a remark, the maximizer of $\mathbb{M}(\mathbf{R}'; \widehat{\mathbf{R}}_n)$ over all positive definite $\mathbf{R}' \in \mathbb{R}^{p\times p}$ (instead of over $\mathbf{R}(\boldsymbol{\theta}')$ for $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$ as in (3.22)) is $\widehat{\mathbf{R}}_n$ itself (with ones on the

diagonal). Thus, in the unrestricted model, the maximizer $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ of (3.22) is given by the corresponding off-diagonal elements in $\widehat{\mathbf{R}}_n$, and $\mathbf{R}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}})$ equals $\widehat{\mathbf{R}}_n$.

To discuss the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$, we first introduce the oracle pseudo-likelihood estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ in the ordinary Gaussian copula model. The oracle estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ is obtained analogously to $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ by replacing $\widehat{\mathbf{R}}_n$ with its oracle counterpart $\mathbf{R}_n$ in (3.22). In the unrestricted model, by an argument identical to the above, $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ is given by the corresponding off-diagonal elements in $\mathbf{R}_n$; then, obviously, when (3.20) holds,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} - \boldsymbol{\theta}_n^{\mathrm{PLE}}) = o_{\mathrm{p}}(1). \tag{3.23}$$

What happens when the parametrization $\mathbf{R}(\boldsymbol{\theta})$ is constrained? Even though $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ and $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ are solutions to two different M-estimation problems, we expect that they are close if the "inputs" $\mathbf{R}_n$ and $\widehat{\mathbf{R}}_n$ to (3.22) are close. Indeed, under (3.20), classical M-estimation theory (e.g., Theorem 3.3 in [31]) again yields (3.23). The asymptotic normality of the (oracle) pseudo-likelihood estimator (in general copula models) was established in [12]; this, under (3.23), then implies the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} - \boldsymbol{\theta})$. Furthermore, in the ordinary Gaussian copula model, the conditions for the semiparametric efficiency of $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ were established in [21, 18, 34] and this will also have consequences for the semiparametric efficiency of $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$; we will come back to this point below Proposition 4.1.

4. Semiparametrically efficient estimation

In this section we study the (asymptotically) semiparametrically efficient estimation of the copula parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^\top$ in $\mathbf{R}(\boldsymbol{\theta})$. Our treatment is rooted in the classical Hàjek-Le Cam theory of asymptotics of statistical experiments as adapted to the semiparametric setting; see, e.g., Chapter 3 in [6] or Chapter 25 in [38]. Section 4.1 is concerned with the semiparametric lower bound for estimating $\boldsymbol{\theta}$ in the presence of the nuisance parameters $\mathbf{B}$, $F_1, \ldots, F_p$ and $F_{\mathbf{X}}$. The picture on the semiparametric lower bound is not complete without a matching estimator (see the discussion following Definition 2.8 in [37]). Thus in Section 4.2, relying on Theorem 3.4 developed earlier, we set out to show that the one-step estimator in conjunction with the residual ranks is a matching estimator and so remains semiparametrically efficient for estimating $\boldsymbol{\theta}$ in the regression setting. Both by space constraint and by technical reasons outlined below Theorem 4.2, and because the semiparametric lower bound in the regression setting was briefly addressed in Section 6.2 in [34], we limit our presentation on this subject in Section 4.1 to some key results and we defer a full treatment to Section B in the supplement. Again all probabilities are stated under the (arbitrary but fixed) law P unless stated otherwise.

4.1. Semiparametric lower bound

We call an estimator (sequence) $\widehat{\boldsymbol{\theta}}_n$ *regular* if it has the same asymptotic distribution under any sequence of local alternatives (e.g., p. 365 in [38]). Following Lemma 25.25 in [38],

we denote the *efficient score (function)* for $\boldsymbol{\theta}$ evaluated at P by $\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P}) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$; then the *efficient influence function* is given by $\mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P})$, where $\mathbf{I}^*(\boldsymbol{\theta})$ is the *efficient information matrix* $\mathbf{I}^*(\boldsymbol{\theta}) = \mathbb{E}[(\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}\dot{\mathbf{l}}^{*\top}_{\boldsymbol{\theta}})(\mathbf{Y}, \mathbf{X}; \mathrm{P})]$. Following Definition 2.8 and Lemma 2.9 in [37], under suitable regularity conditions (differentiability in quadratic mean of suitable local parametric submodels passing through P suffices), an estimator $\widehat{\boldsymbol{\theta}}_n$ is *(asymptotically) semiparametrically efficient* at the law P (which has the built-in requirement that $\widehat{\boldsymbol{\theta}}_n$ is regular at this law) for estimating $\boldsymbol{\theta}$ if and only if it is asymptotically linear in the efficient influence function, that is,

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) = \frac{1}{\sqrt{n}} \sum_{i \in [n]} \mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\mathbf{Y}_i, \mathbf{X}_i; \mathrm{P}) + o_{\mathrm{p}}(1). \tag{4.1}$$

By the Hàjek-Le Cam convolution theorem (e.g., Theorem 25.20 in [38]), the asymptotic distribution of every regular estimator of $\boldsymbol{\theta}$ (after centering at $\boldsymbol{\theta}$ and scaling by $\sqrt{n}$) is the convolution of $\mathcal{N}_d(\mathbf{0}, \mathbf{I}^{*-1}(\boldsymbol{\theta}))$ and another estimator-specific probability distribution M. If an estimator $\widehat{\boldsymbol{\theta}}_n$ satisfies (4.1), then with $\rightsquigarrow$ denoting weak convergence,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \rightsquigarrow \mathcal{N}_d(\mathbf{0}, \mathbf{I}^{*-1}(\boldsymbol{\theta})); \tag{4.2}$$

thus, M is degenerate at $\mathbf{0}$. Therefore $\widehat{\boldsymbol{\theta}}_n$ is optimal among regular estimators.

Denote by $\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(\cdot; \boldsymbol{\theta}) = (\dot{l}^{*\mathrm{o}}_{\boldsymbol{\theta},m}(\cdot; \boldsymbol{\theta}))_{m=1}^d : [0,1]^p \to \mathbb{R}^d$ the efficient score in the ordinary Gaussian copula model when all margins are Unif$(0,1)$ distributions (but when this information is not known). This function, which determines the semiparametric lower bound in the ordinary Gaussian copula model, is derived in Section 2.4 in [34]. Define the matrices of partial derivatives $\dot{\mathbf{R}}_1(\boldsymbol{\theta}), \ldots, \dot{\mathbf{R}}_d(\boldsymbol{\theta})$ of $\mathbf{R}(\boldsymbol{\theta})$, and the matrices of partial derivatives $\dot{\mathbf{S}}_1(\boldsymbol{\theta}), \ldots, \dot{\mathbf{S}}_d(\boldsymbol{\theta})$ of $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{R}(\boldsymbol{\theta})^{-1}$ by

$$(\dot{\mathbf{R}}_m(\boldsymbol{\theta}))_{kk'} = \frac{\partial}{\partial \theta_m}(\mathbf{R}(\boldsymbol{\theta}))_{kk'}, \quad (\dot{\mathbf{S}}_m(\boldsymbol{\theta}))_{kk'} = \frac{\partial}{\partial \theta_m}(\mathbf{S}(\boldsymbol{\theta}))_{kk'}, \quad k, k' \in [p],\, m \in [d]$$

when they exist. Further define $\Phi_\bullet^{-1} : [0,1]^p \to \mathbb{R}^p$ as $\Phi_\bullet^{-1}(\mathbf{u}) = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p))^\top$ for $\mathbf{u} = (u_1, \ldots, u_p)^\top \in [0,1]^p$. Then specifically

$$\dot{l}^{*\mathrm{o}}_{\boldsymbol{\theta},m}(\mathbf{u}; \boldsymbol{\theta}) = \frac{1}{2}\Phi_\bullet^{-1}(\mathbf{u})^\top \Big\{\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{g}_m(\boldsymbol{\theta})) - \dot{\mathbf{S}}_m(\boldsymbol{\theta})\Big\}\Phi_\bullet^{-1}(\mathbf{u}), \quad \mathbf{u} \in [0,1]^p, m \in [d]. \tag{4.3}$$

In (4.3), the vector $\mathbf{g}_m(\boldsymbol{\theta}) = (g_{1,m}(\boldsymbol{\theta}), \ldots, g_{p,m}(\boldsymbol{\theta}))^\top$ and the matrix $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{b}) \in \mathbb{R}^{p \times p}$ are

$$\mathbf{g}_m(\boldsymbol{\theta}) = -\left\{\mathbf{I}_p + \mathbf{R}(\boldsymbol{\theta}) \circ_{\mathrm{H}} \mathbf{S}(\boldsymbol{\theta})\right\}^{-1} \Big\{\dot{\mathbf{R}}_m(\boldsymbol{\theta}) \circ_{\mathrm{H}} \mathbf{S}(\boldsymbol{\theta})\Big\}\boldsymbol{\iota}_p, \tag{4.4}$$

$$\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{b}) = \mathbf{S}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{b}) + \mathrm{diag}(\mathbf{b})\mathbf{S}(\boldsymbol{\theta}). \tag{4.5}$$

In (4.4) and (4.5), $\boldsymbol{\iota}_p \in \mathbb{R}^p$ is a vector of all ones, $\circ_{\mathrm{H}}$ denotes the Hadamard product, and $\mathrm{diag}(\mathbf{b})$ denotes the diagonal matrix with the elements of $\mathbf{b}$ arranged on the diagonal.

As Proposition 4.1 below shows, under regularity conditions stated in Assumption 4.1, the efficient scores in our regression setting and in the ordinary Gaussian copula model

are related in a simple way, and the efficient information matrices in the two cases are identical. In Assumption 4.1, conditions (i) to (iv) are identical to those in Assumption 2.1 in [34], and under which the parametric Gaussian copula model for $\mathbf{E}$ with known, uniform margins is regular (Lemma 2.2 in [34]). On the other hand, we need the additional conditions (v) and (vi) to ensure differentiability in quadratic mean of suitable local parametric submodels passing through P in our regression setting.

Assumption 4.1. *For the mapping $\boldsymbol{\theta} \to \mathbf{R}(\boldsymbol{\theta}) : \boldsymbol{\Theta} \subset \mathbb{R}^d \to \mathbb{R}^{p\times p}$, suppose that*

(i) $\boldsymbol{\Theta}$ is open, and $\boldsymbol{\theta} \to \mathbf{R}(\boldsymbol{\theta})$ is one-to-one.
(ii) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the inverse $\mathbf{S}(\boldsymbol{\theta})$ of $\mathbf{R}(\boldsymbol{\theta})$ exists.
(iii) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the matrices $\dot{\mathbf{R}}_1(\boldsymbol{\theta}), \ldots, \dot{\mathbf{R}}_d(\boldsymbol{\theta})$ exist and are continuous in $\boldsymbol{\theta}$.
(iv) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the matrices $\dot{\mathbf{R}}_1(\boldsymbol{\theta}), \ldots, \dot{\mathbf{R}}_d(\boldsymbol{\theta})$ are linearly independent.

Furthermore, the law P *holds, and*

(v) The covariate $\mathbf{X}$ satisfies $\mathbb{E}\left[\|\mathbf{X}\|^2\right] < \infty$.
(vi) For each $k \in [p]$, f_k is continuous, is supported on an interval (a_k, b_k) where $-\infty \le a_k < b_k \le \infty$, and on this interval f_k is strictly positive and differentiable with derivative $\dot{f}_k$. In addition, f_k has finite information for location, that is, $\int_{[a_k,b_k]}(\dot{f}_k^2/f_k)(t)\mathrm{d}t < \infty$. Moreover,

$$\limsup_{\epsilon\to 0}\left\{\int_{(0,1/2]}\frac{1}{\phi^2(\Phi^{-1}(u))}\sup_{\delta:|\delta|\le\epsilon}\left\{f_k\circ F_k^{\leftarrow}((1+\delta)u)\right\}^2\mathrm{d}u\right.$$
$$\left.+\int_{(1/2,1)}\frac{1}{\phi^2(\Phi^{-1}(u))}\sup_{\delta:|\delta|\le\epsilon}\left\{f_k\circ F_k^{\leftarrow}(1-(1+\delta)(1-u))\right\}^2\mathrm{d}u\right\} < \infty. \quad (4.6)$$

Remark 1. *Under Assumption 4.1(ii) and (iii), the matrices $\dot{\mathbf{S}}_1(\boldsymbol{\theta}), \ldots, \dot{\mathbf{S}}_d(\boldsymbol{\theta})$ also exist and are continuous in $\boldsymbol{\theta}$ (e.g., the remark below Assumption 2.1 in [34]).*

Remark 2. *Using Inequality (C.2) in the supplement to bound the factor $1/\phi^2(\Phi^{-1}(u))$ in (4.6), it can be seen that the left hand side of (4.6) is quite similar to δ_n, introduced in (3.17), in our analysis of the estimator $\widehat{\mathbf{R}}_n$. Likewise condition (4.6) is satisfied by the normal density, densities with polynomial decay of the form (3.1) when $a > 1$, and the Gamma density with shape parameter larger than 2. The perturbation by δ in condition (4.6) is the price we pay by jointly perturbing the coefficient matrix $\mathbf{B}$ and the marginals F_k, $k \in [p]$ when constructing local parametric submodels passing through* P*; see Section B.2 in the supplement.*

Let $F_\bullet : \mathbb{R}^p \to \mathbb{R}^p$ be $F_\bullet(\mathbf{z}) = (F_1(z_1), \ldots, F_p(z_p))^\top$ for $\mathbf{z} = (z_1, \ldots, z_p)^\top \in \mathbb{R}^p$.

Proposition 4.1. *Under Assumption 4.1, the efficient score $\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P})$ is given by*

$$\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{x}; \mathrm{P}) = \dot{\mathbf{l}}^{*\circ}_{\boldsymbol{\theta}}(F_\bullet(\mathbf{y} - \mathbf{B}^\top\mathbf{x}); \boldsymbol{\theta}), \quad \forall(\mathbf{y}, \mathbf{x}) \in \mathbb{R}^p \times \mathbb{R}^q, \quad (4.7)$$

for $\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(\cdot;\boldsymbol{\theta}) = (\dot{l}^{*\mathrm{o}}_{\boldsymbol{\theta},m}(\cdot;\boldsymbol{\theta}))_{m=1}^{d}$ *given in* (4.3)*. Hence the efficient information matrix* $\mathbf{I}^{*}(\boldsymbol{\theta})$ *coincides with the one in the ordinary Gaussian copula model (given by (2.20) in [34]).*

Proof. The proof is deferred to Section B.4 in the supplement. □

The simple relationship (4.7) has practical consequence for characterizing the semiparametric efficiency of an estimator of $\boldsymbol{\theta}$ in our regression setting. We will take the pseudo-likelihood estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{PLE}}_n$ from Section 3.4 as an example. Suppose that under the ordinary Gaussian copula model the oracle pseudo-likelihood estimator $\boldsymbol{\theta}^{\mathrm{PLE}}_n$ is semiparametrically efficient. Then $\boldsymbol{\theta}^{\mathrm{PLE}}_n$ is asymptotically linear in its efficient influence function, which by isometry (e.g., Eq. (55) in Section 4.7 in [6]) is given by $\mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\cdot);\boldsymbol{\theta})$. Thus (4.8) given later holds with $\widehat{\boldsymbol{\theta}}_n$ replaced by $\boldsymbol{\theta}^{\mathrm{PLE}}_n$. Now we impose the regression structure, and moreover assume that (3.20) holds, which also implies (3.23) (as discussed in Section 3.4). Then by (1.5) and (4.7), Equation (4.1) with $\widehat{\boldsymbol{\theta}}_n$ replaced by $\widehat{\boldsymbol{\theta}}^{\mathrm{PLE}}_n$ holds as well, implying the semiparametric efficiency of $\widehat{\boldsymbol{\theta}}^{\mathrm{PLE}}_n$ in the regression setting. Therefore, we have essentially reduced the characterization of the semiparametric efficiency of the estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{PLE}}_n$ to that of $\boldsymbol{\theta}^{\mathrm{PLE}}_n$ (under the ordinary Gaussian copula model), and the latter has been extensively studied in, e.g., [21, 18, 34]. In particular, we conclude right away that, in the unrestricted model under the regression setting, $\mathbf{R}(\widehat{\boldsymbol{\theta}}^{\mathrm{PLE}}_n) = \widehat{\mathbf{R}}_n$ is semiparametrically efficient for $\mathbf{R}$, so long as (3.20) holds.

4.2. Semiparametrically efficient estimator

By Proposition 4.1 and Eq. (4.1), an estimator $\widehat{\boldsymbol{\theta}}_n$ is semiparametrically efficient at the law P for estimating $\boldsymbol{\theta}$ if and only if

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) = \frac{1}{\sqrt{n}} \sum_{i\in[n]} \mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i);\boldsymbol{\theta}) + o_{\mathrm{p}}(1). \tag{4.8}$$

The one-step method that updates an initial $\sqrt{n}$-consistent estimator $\boldsymbol{\theta}^{*}_n$ to produce an efficient estimator has a long history; see, e.g., Section 25.8 in [38] for a textbook treatment. In the ordinary Gaussian copula model, [34] constructed and established the semiparametric efficiency of an one-step estimator (see [36] for a different update technique). In the regression setting, because we would like the one-step estimator to be constructed on the residual ranks, we require the initial estimator $\boldsymbol{\theta}^{*}_n$ to be so constructed as well, as stated in Assumption 4.2. Section 3.4 discusses some natural candidates for the initial estimator $\boldsymbol{\theta}^{*}_n$; the pseudo-likelihood estimator is one such example.

Assumption 4.2. *The initial estimator* $\boldsymbol{\theta}^{*}_n$ *is constructed from the (rescaled) residual ranks* $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$, $i \in [n]$, $k \in [p]$*. Moreover,* $\sqrt{n}(\boldsymbol{\theta}^{*}_n - \boldsymbol{\theta}) = \mathcal{O}_{\mathrm{p}}(1)$ *under* P*.*

For an estimator $\boldsymbol{\theta}_n^*$ of $\boldsymbol{\theta}$, let $\widetilde{\boldsymbol{\theta}}_n$ be a discretized version of $\boldsymbol{\theta}_n^*$ obtained by rounding $\boldsymbol{\theta}_n^*$ to the nearest $n^{-1/2}\mathbb{Z}^d$ grid. The one-step estimator is defined as

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{OSE}} = \widetilde{\boldsymbol{\theta}}_n + \mathbf{I}^{*-1}(\widetilde{\boldsymbol{\theta}}_n)\left\{\frac{\phi_n}{n}\sum_{i\in[n]} \dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(\widehat{F}_{n,1}^{\mathrm{r}}(\widehat{E}_{i,1}),\ldots,\widehat{F}_{n,p}^{\mathrm{r}}(\widehat{E}_{i,p});\widetilde{\boldsymbol{\theta}}_n)\right\}, \tag{4.9}$$

where the second term on the right is the update term. The one-step estimator above is essentially obtained from (3.3) in [34] via substitution of the (for us, unobservable) oracle ranks $F_{n,k}^{\mathrm{r}}(E_{i,k})$ by the (rescaled) residual ranks $\widehat{F}_{n,k}^{\mathrm{r}}(\widehat{E}_{i,k})$. Just as in the ordinary Gaussian copula model, by Proposition 4.2 below the one-step estimator above is semiparametrically efficient for $\boldsymbol{\theta}$ now in the regression setting.

Remark 3. *The update term in* (4.9) *conforms a usual representation of the one-step method, but can also be rewritten explicitly in terms of the normal scores rank correlation estimator* $\widehat{\mathbf{R}}_n$*. To see this, the* m*th component,* $m \in [d]$*, of the term in the curly bracket in* (4.9) *can be written alternatively as* $\frac{1}{2}\operatorname{tr}(\{\mathbf{D}_{\widetilde{\boldsymbol{\theta}}_n}(\mathbf{g}_m(\widetilde{\boldsymbol{\theta}}_n)) - \dot{\mathbf{S}}_m(\widetilde{\boldsymbol{\theta}}_n)\}\widehat{\mathbf{R}}_n)$*.*

Proposition 4.2. *Suppose that Assumptions 4.1 and 4.2 hold. Moreover assume the conditions in Theorem 3.4 (including the one on* δ_n*), so* (3.20) *holds. Then (under* P*) the one-step estimator* $\widehat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$ *satisfies* (4.1) *(or equivalently* (4.8)*) and* (4.2) *with* $\widehat{\boldsymbol{\theta}}_n$ *replaced by* $\widehat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$*. In particular,* $\widehat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$ *is a semiparametrically efficient estimator of* $\boldsymbol{\theta}$ *at* P*.*

Proof. The proof is deferred to Section B.5 in the supplement. □

Interestingly, by Proposition 4.2 and under the extra conditions in Theorem 3.4, the one-step estimator $\widehat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$ satisfies (4.8) and hence is asymptotically linear in the efficient influence function $\mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(F_\bullet(\cdot);\boldsymbol{\theta})$ in the ordinary Gaussian copula model. Thus, it appears as if (with the extra conditions in Theorem 3.4) the task of estimating $\boldsymbol{\theta}$ is no more difficult under the additional regression structure. Even without a dedicated lower bound analysis (as in Section 4.1), these observations already suggest that the semiparametric lower bound in our regression setting should largely coincide with that in the ordinary Gaussian copula model. This is partially why we decide to defer the formal but somewhat tedious lower bound analysis that supplements Section 4.1 to Section B in the supplement. In contrast to the discussion above, the analysis in Sections 4.1 and B does not require a matching, efficient estimator such as the $\widehat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$, and so in particular does not require the associated conditions in Theorem 3.4 (such as that placed by Assumption 3.4 which potentially requires bounded moment of $\|\mathbf{X}\|$ higher than the second order).

5. Numerical performance

We carried out a small simulation study to demonstrate the finite-sample accuracy of our estimation procedures based on the residual ranks. We consider first an unrestricted model in Section 5.1 and then, for constrained parametrization $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$, a Toeplitz model in Section 5.2 and a factor model in Section 5.3.

5.1. Unrestricted model

In this scenario, we first consider an unrestricted model with $p = 3$ where each of the $d = p(p-1)/2$ elements of the upper-triangular portion of $\mathbf{R}$ is a free parameter. We defer a brief discussion of a higher dimensional case $p = 10$ to the very end of this section. Specifically, for $p = 3$,

- We generate $\mathbf{R}$ so that the elements in the upper-triangular portion of $\mathbf{R}$ are drawn independently from a normal distribution with standard deviation 0.5; we repeat this procedure if necessary until we obtain a positive definite matrix. The particular $\mathbf{R}$ generated is

$$\mathbf{R} = \begin{pmatrix} 1.0000 & 0.2674 & 0.1791 \\ 0.2674 & 1.0000 & 0.1709 \\ 0.1791 & 0.1709 & 1.0000 \end{pmatrix}.$$

We next specify the nuisance parameters that will be taken from the following set of possible combinations (not every single combination will be studied):

- For the $q \times p$ coefficient matrix $\mathbf{B}$ (recall that $p = 3$), we consider $q = 2$ or $q = 10$. For each case, the elements of $\mathbf{B}$ are drawn independently from a standard normal. The particular $\mathbf{B}$ generated are recorded in Section E.1 in the supplement.
- The distribution function $F_{\mathbf{X}}$ of the covariate $\mathbf{X} \in \mathbb{R}^q$ is either a multivariate normal distribution or a multivariate t-distribution (with 3 degrees of freedom) whose covariance or shape matrix has unit diagonal elements and off-diagonal elements equal to $\rho = 0.1, 0.5$ or 0.9.
- The marginal distribution functions F_k, $k \in [p]$ of $\mathbf{E}$ are chosen to have either the same standard normal or the Cauchy distribution (i.e., t-distribution with 1 degree of freedom), or these distributions scaled by a constant factor $1/5$.

For each combination studied, $N = 1000$ Monte-Carlo repetitions are performed, with sample sizes $n = 50$ or $n = 250$. For each repetition, an independent sample of $\mathcal{N}(\mathbf{0}, \mathbf{R})$ distribution is drawn, and the marginals of this sample are subsequently adjusted according to the specification of F_k, $k \in [p]$ above to produce the sample of $\mathbf{E}$. Next a sample of $\mathbf{X}$ is drawn independently from $\mathbf{E}$. Finally, the sample of $\mathbf{Y}$ is determined via (1.5).

Note that when estimating $\mathbf{B}$, the copula component $\mathbf{E}$ is considered as the noise, and intuitively a smaller copula component should yield better estimation of $\mathbf{B}$. In contrast, in the estimation of $\mathbf{R}$ the copula component is instead considered as the signal. Our current theory does not reveal which effect will dominate (as we simply take the rate $\|\widehat{\mathbf{B}} - \mathbf{B}\|$ as given and do not consider the special, possibly $\mathbf{E}$-dependent properties that $\widehat{\mathbf{B}}$ may have). Therefore, for the marginal distributions, the scaling by $1/5$ is intended to clarify the effect of the scale of the copula component on the estimation of $\mathbf{R}$.

We focus on the comparison between the normal scores rank correlation estimator $\widehat{\mathbf{R}}_n$, which coincides with the pseudo-likelihood estimator in this scenario as discussed in Section 3.4, and its oracle counterpart $\mathbf{R}_n$ based on the (unobservable) oracle ranks. Compared to $\widehat{\mathbf{R}}_n$, the estimator $\widehat{\mathbf{R}}_n^{\rho}$ via inversion of Spearman's rho performs slightly

worse. We omit presenting the results related to $\widehat{\mathbf{R}}_n^{\rho}$, except at the very end of this section when we consider $p = 10$.

Our first simulation considers the case $q = 2$ and $\rho = 0.1$, and also considers the effect of different regression methods on the estimation of $\mathbf{B}$ which in turn affects the estimation of $\mathbf{R}$. The results are summarized in Figure 1. In particular, in each subfigure, the results for $\widehat{\mathbf{B}}$ produced using the ordinary least squares (OLS) are plotted in the first six boxplots, while those produced using the least absolute deviation (LAD, or equivalently quantile regression at the quantile level 0.5) are plotted in the next six boxplots. These two six-boxplot sets share a common x-label pattern. Each x-label is a two-tuple that indicates the distribution of the covariate $\mathbf{X}$ and the marginal distribution functions F_k, $k \in [p]$ of $\mathbf{E}$. The first letter in the two-tuple is either "n" or "3," which indicates that $\mathbf{X}$ is either drawn from the multivariate normal or from the multivariate t-distribution (with 3 degrees of freedom) respectively as described earlier. The second letter in the two-tuple is either "n", "1", "n/5" or "1/5," which indicates that each of F_k, $k \in [p]$ is either the standard normal, the Cauchy distribution, the standard normal scaled by 1/5, or the Cauchy distribution scaled by 1/5 respectively, again as described earlier.

The performance of the estimators is measured by various Frobenius norms. The first row in Figure 1 plots $\|\widehat{\mathbf{R}}_n - \mathbf{R}\|_F$. Here we additionally consider a naive estimator $\mathbf{R}_n^{\mathbf{Y}}$, which is the normal scores rank correlation estimator produced directly from a sample of $\mathbf{Y}$ without taking into account the covariate $\mathbf{X}$. The particular sample of $\mathbf{Y}$ is taken from the "(n,1)" specification described above. The performances of the oracle estimator $\mathbf{R}_n$ and the naive estimator $\mathbf{R}_n^{\mathbf{Y}}$ are denoted by "Ora." and "Y" on the x-label respectively. For the second row we plot the ratio $\|\widehat{\mathbf{R}}_n - \mathbf{R}\|_F / \|\mathbf{R}_n - \mathbf{R}\|_F$, and for the third row we consider the estimation of $\mathbf{B}$ and plot $\|\hat{\mathbf{B}} - \mathbf{B}\|_F$. In Figure 1, all three rows consist of two columns corresponding to sample sizes $n = 50$ and $n = 250$ respectively, and all subfigures are produced with the y-axis in logarithm scale.

From the first row of Figure 1 we immediately observe that the naive estimator $\mathbf{R}_n^{\mathbf{Y}}$ performs substantially worse than any of the other estimators, even though the sample of $\mathbf{Y}$ is produced with a light-tailed $\mathbf{X}$ and a heavy-tailed $\mathbf{E}$ to intentionally minimize the perturbation by the sample of $\mathbf{X}$ on the sample of $\mathbf{E}$. For this reason we exclude the naive estimator in subsequent simulation studies.

Next, from the third row of Figure 1 and also as expected, when $\mathbf{E}$ is heavy-tailed the OLS produces a rather inaccurate estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}$, which further leads to inaccurate (relatively speaking especially at $n = 250$) estimator $\widehat{\mathbf{R}}_n$ of $\mathbf{R}$. On the other hand, when $\mathbf{E}$ is light-tailed, the OLS produces a somewhat more accurate estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}$ than the LAD, but this leads to no appreciable improvement in the estimator $\widehat{\mathbf{R}}_n$. The latter phenomenon is also expected because our studies have shown that the rate of $\widehat{\mathbf{B}} - \mathbf{B}$ does not affect the estimation of $\mathbf{R}$ (at the first order). We thus also exclude $\widehat{\mathbf{B}}$ produced by the non-robust OLS in future studies.

Having considered the case $q = 2$ and $\rho = 0.1$, and excluded the naive estimator and the estimators involving the OLS estimation of $\mathbf{B}$, we now consider more correlated covariates with higher ρ, and also possibly larger covariate dimension. Specifically, Figure 1 in Section E in the supplement presents the results for the common value of $\rho = 0.5$ and

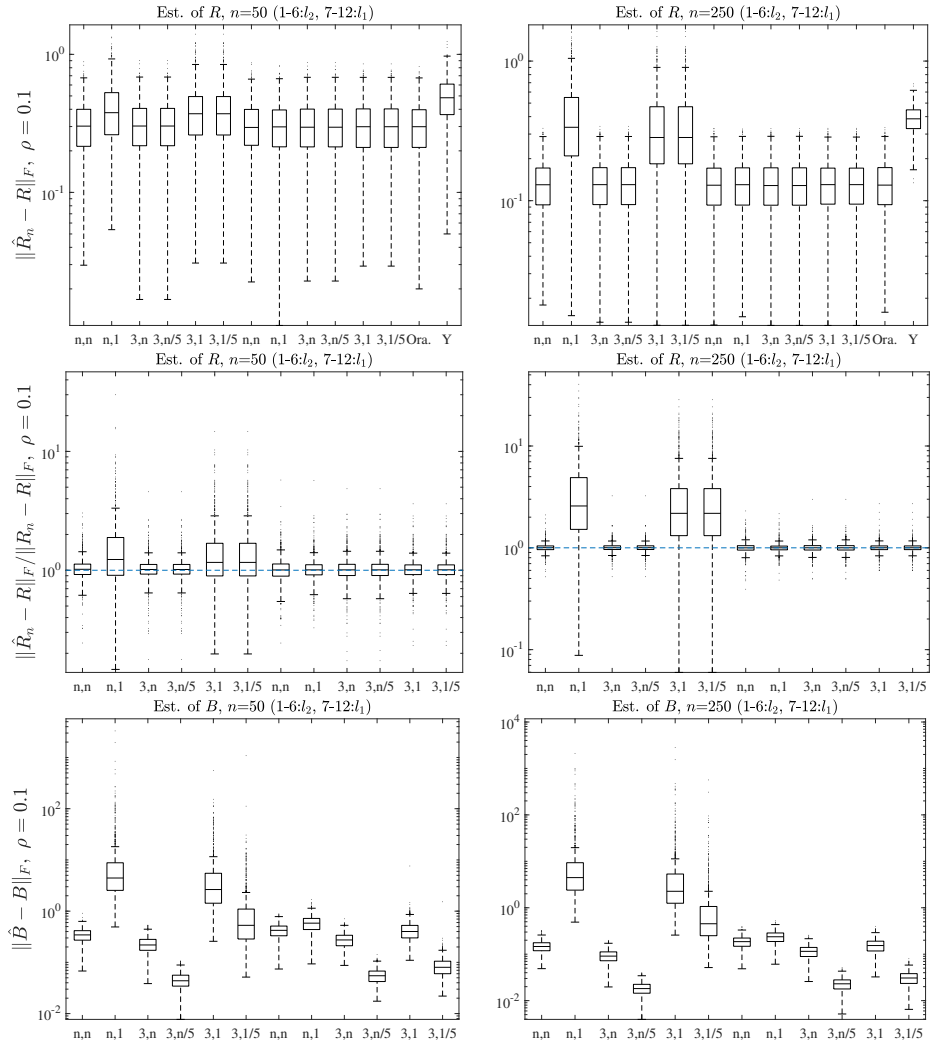

Figure 1. Simulation results for the unrestricted model for $(p, q) = (3, 2)$ and $\rho = 0.1$. The meanings of the labels and of ρ are explained in the main text. All subfigures are produced with the y-axis in logarithm scale.

for $(p, q) = (3, 2)$ or $(p, q) = (3, 10)$, while Figure 2 in the same section presents the results for the same (p, q) specifications but for $\rho = 0.9$. Collectively, these latter results and the ones from Figure 1 show that, as long as a reasonable $\widehat{\mathbf{B}}$ (such as that from the LAD) is used in computing the residual ranks, the estimator $\widehat{\mathbf{R}}_n$ is nearly as good as the oracle estimator $\mathbf{R}_n$, even with high covariate dimension $q = 10$ (which results in a

total number of $q \times p = 30$ free parameters in $\mathbf{B}$) and high correlation $\rho = 0.9$ among the covariates at a relatively small sample size $n = 50$. Moreover, within each figure the estimators $\widehat{\mathbf{R}}_n$ under various specifications of the distribution of $\mathbf{X}$ and the marginals of $\mathbf{E}$ perform quite similarly.

Next, we consider the case $p = 10$ and $q = 2$ or $q = 10$, and bring the estimators via inversion of Spearman's rho into comparison. Here we only consider the case when the distribution function $F_{\mathbf{X}}$ is a multivariate t-distribution (with 3 degrees of freedom) whose shape matrix has unit diagonal elements and off-diagonal elements equal to $\rho = 0.5$, and when the marginal distribution functions F_k, $k \in [p]$ are the Cauchy distribution. We again consider $N = 1000$ Monte-Carlo repetitions and sample sizes $n = 50$ or $n = 250$. We randomly generate each element in the upper-triangular portion of $\mathbf{R}$ to be either $1/4$ or $-1/4$, each with probability $1/2$, until we obtain a positive-definite matrix. We randomly generate the elements in $\mathbf{B}$ to be either 1 or -1, each with probability $1/2$. The same $\mathbf{R}$ and $\mathbf{B}$ are kept for all the $N = 1000$ Monte-Carlo repetitions.

We consider four estimators in total: the residual-based normal scores (rank correlation) estimator $\widehat{\mathbf{R}}_n$ and the residual-based estimator via inversion of Spearman's rho $\widehat{\mathbf{R}}_n^{\rho}$, and their two oracle counterparts $\mathbf{R}_n$ and $\mathbf{R}_n^{\rho}$. We compare their performances, summarized in Table 1, by the Frobenius norm of the deviations of these estimators from the true $\mathbf{R}$. In this larger p setting, again the performances of the residual-based estimators closely follow their oracle counterparts, even at $q = 10$. Moreover, for both the oracle version and the residual-based version, the estimator via inversion of Spearman's rho lags slightly behind the normal scores estimator. This follows from theoretical prediction simply because the normal scores estimators (whether the oracle version or the residual-based version) are semiparametrically efficient, while the estimators via inversion of Spearman's rho are not.

$q = 2$

n	$\mathbf{R}_n$	$\mathbf{R}_n^{\rho}$	$\widehat{\mathbf{R}}_n$	$\widehat{\mathbf{R}}_n^{\rho}$
50	1.24 (1.14, 1.39)	1.31 (1.19, 1.45)	1.27 (1.15, 1.40)	1.33 (1.20, 1.46)
250	0.55 (0.49, 0.61)	0.58 (0.53, 0.64)	0.56 (0.50, 0.61)	0.58 (0.53, 0.64)

$q = 10$

n	$\mathbf{R}_n$	$\mathbf{R}_n^{\rho}$	$\widehat{\mathbf{R}}_n$	$\widehat{\mathbf{R}}_n^{\rho}$
50			1.40 (1.26, 1.53)	1.48 (1.33, 1.61)
250			0.56 (0.50, 0.62)	0.59 (0.53, 0.65)

Table 1. Medians, the 25% and the 75% quantiles of the Frobenius norm of the deviations of the four estimators described in the maintext from the true $\mathbf{R}$. The values associated with the oracle estimators are obviously the same for $q = 2$ and $q = 10$, and hence are omitted for the latter case.

Finally, we consider a case with highly skewed $\mathbf{E}$. Our setup here is entirely identical to the previous case involving Spearman's rho, except that we take any positive entry in the sample of $\mathbf{E}$ to the third power. This results in an *extremely* heavy-tailed and skewed distribution. The performance of the same four estimators considered in the previous case is summarized in Table 2. As we can see from the table, the performance of our residual-based estimators is hardly affected, a testament to their remarkable robustness.

$q = 2$

n	$\mathbf{R}_n$	$\mathbf{R}_n^\rho$	$\widehat{\mathbf{R}}_n$	$\widehat{\mathbf{R}}_n^\rho$
50	1.24 (1.14, 1.39)	1.31 (1.19, 1.45)	1.28 (1.15, 1.41)	1.34 (1.21, 1.46)
250	0.55 (0.49, 0.61)	0.58 (0.53, 0.64)	0.55 (0.49, 0.61)	0.58 (0.53, 0.64)

$q = 10$

n	$\mathbf{R}_n$	$\mathbf{R}_n^\rho$	$\widehat{\mathbf{R}}_n$	$\widehat{\mathbf{R}}_n^\rho$
50			1.44 (1.29, 1.56)	1.52 (1.37, 1.65)
250			0.56 (0.51, 0.62)	0.59 (0.53, 0.65)

Table 2. Medians, the 25% and the 75% quantiles of the Frobenius norm of the deviations of the four estimators described in the maintext from the true $\mathbf{R}$, now with highly skewed $\mathbf{E}$ but otherwise with conditions identical to those corresponding to Table 1. Again, the values associated with the oracle estimators are obviously the same for $q = 2$ and $q = 10$, and hence are omitted for the latter case.

5.2. Toeplitz model

Our first scenario for constrained parametrization considers a Toeplitz model, which is a $(p-1)$-parameter model with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p-1})^\top$ such that $(\mathbf{R})_{kk'} = \theta_{|k-k'|}$ for $k \neq k'$. We consider the case $p = 4$, which is particularly interesting because as stated in [34], here the oracle pseudo-likelihood estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ of $\boldsymbol{\theta}$ in the ordinary Gaussian copula model can be quite inefficient. Specifically, [34] verified that at the particular value $\boldsymbol{\theta} = \boldsymbol{\theta}^* = (0.4945, 0.4593, -0.8462)^\top$, the asymptotic relative efficiencies of $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ with respect to the information bound are equal to $18.3\%, 19.8\%, 96.9\%$ for θ_1, θ_2 and θ_3 respectively. Recall that the oracle pseudo-likelihood estimator and the pseudo-likelihood estimator $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ have the same asymptotic distribution (see Section 3.4), the efficient information matrices in the ordinary Gaussian copula model and in our regression setting coincide (see Proposition 4.1), and the one-step estimator (discussed in Section 4.2) is semiparametrically efficient. Thus, the one-step estimator should substantially outperform $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$.

For our specific simulation study, we let $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ be as discussed in the last paragraph, and specify the nuisance parameters similar to Section 5.1:

- The $q \times p$ coefficient matrix $\mathbf{B}$ is as generated in Section 5.1, now with $p = 4$, and again with $q = 2$ and $q = 10$. The particular $\mathbf{B}$ generated are recorded in Section E.2 in the supplement.
- The distribution function $F_{\mathbf{X}}$ is a multivariate t-distribution (with 3 degrees of freedom) whose shape matrix has unit diagonal elements and off-diagonal elements equal to $\rho = 0.5$ or 0.9.
- The marginal distribution functions F_k, $k \in [p]$ are the Cauchy distribution.

We again consider $N = 1000$ Monte-Carlo repetitions and sample sizes $n = 50$ or $n = 250$.

Figure 2 presents the results when $\rho = 0.5$ and $(p, q) = (4, 2)$. Figure 3 in Section E in the supplement presents the results for the same ρ but for $(p, q) = (4, 10)$. Figures 4 and 5 in the same section present the results for the same (p, q) specifications but for $\rho = 0.9$. The ℓth row of each figure, where $\ell \in \{1, 2, 3\}$, considers the estimation of θ_ℓ, and compares the deviations of the oracle pseudo-likelihood estimator, the oracle one-step estimator, the pseudo-likelihood estimator, and the one-step estimator (as indicated

by the x-labels) from θ_ℓ.

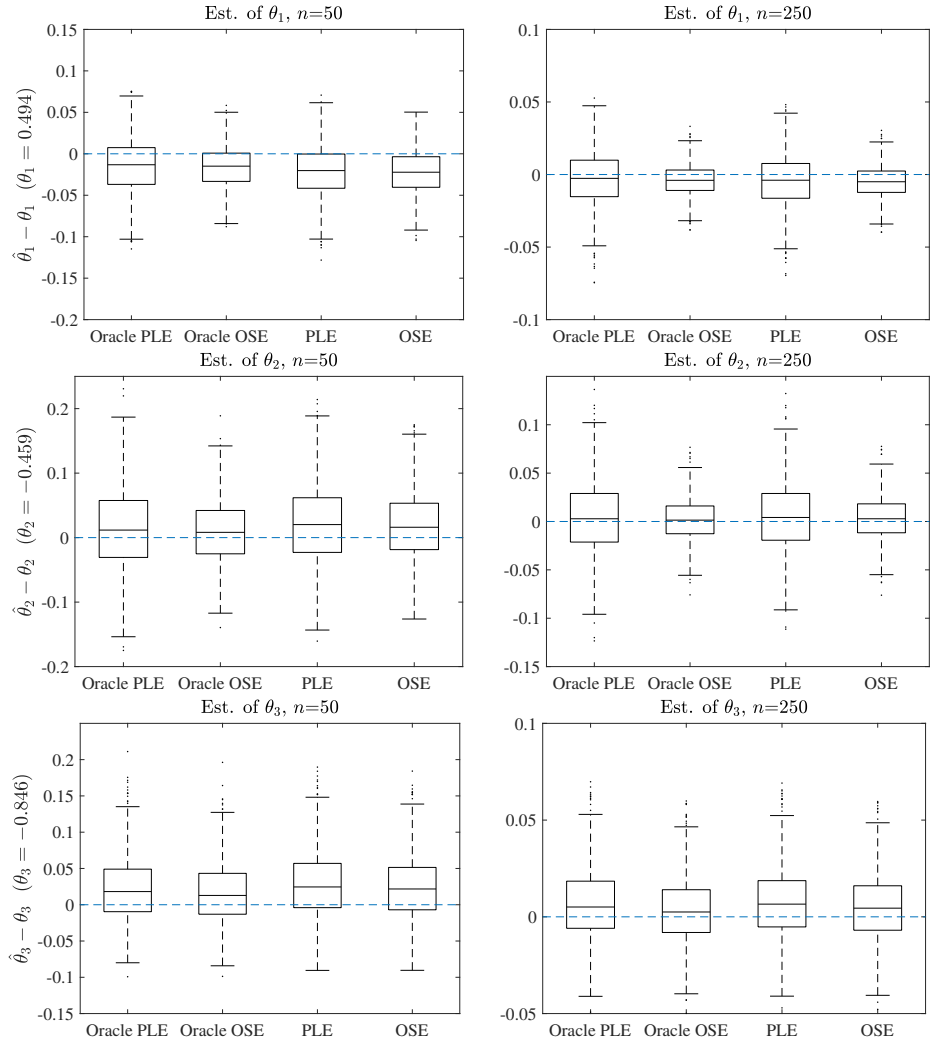

Figure 2. Simulation results for the Toeplitz model under the parameters specified in the main text, for $(p, q) = (4, 2)$ and $\rho = 0.5$.

Similar to [34], but in our regression setting, we observe that the one-step estimators outperform the respective pseudo-likelihood estimators in particular for θ_ℓ, $\ell = 1$ or 2, just as expected. More importantly, at least when the covariate dimension is low with $q = 2$, the one-step estimators perform almost as well as the corresponding oracle one-step estimators. When the covariate dimension is high with $q = 10$, both the one-step

estimators and the respective pseudo-likelihood estimators start out with a relatively large bias when the sample size is small, but the bias improves at a larger sample size.

5.3. Factor model

Our second scenario for constrained parametrization is a factor model which states that $\mathbf{R}$ admits a low-rank plus diagonal decomposition as $\mathbf{R} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \mathbf{D}$ where $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times r}$ with $r < p$ and $\mathbf{D}$ is a diagonal matrix with non-negative diagonal elements. We consider a simple case with $p = 5$ and $r = 1$. Then $\boldsymbol{\Lambda} = \boldsymbol{\theta} = (\theta_1, \ldots, \theta_5)^\top$ is identified (when at least three elements of $\boldsymbol{\theta}$ are nonzero to the precise; see Theorem 5.5 in [2]), so long as we fix the sign of θ_1. (For identification conditions in a general factor model, see., e.g., [2].) Example 5.5 in [34] shows that (under the factor model) the oracle pseudo-likelihood estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ is semiparametrically efficient for $\boldsymbol{\theta}$. By the earlier discussion following Proposition 4.1, we then conclude that the residual-based pseudo-likelihood estimator $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ is also semiparametrically efficient for $\boldsymbol{\theta}$ in the regression setting.

For our specific simulation study, we first construct $\mathbf{R}$ and the copula parameter $\boldsymbol{\theta}$ as follows. We set $\boldsymbol{\theta}^* = (1, -1, 1, 1, 1)^\top$, $\mathbf{D}^*$ a diagonal matrix with elements $1, \ldots, 5$ on the diagonal, and the covariance matrix $\mathbf{R}^* = \boldsymbol{\theta}^*\boldsymbol{\theta}^{*\top} + \mathbf{D}^*$. Then, we let $\mathbf{R}$ be the correlation matrix corresponding to $\mathbf{R}^*$. Specifically, we let $\widetilde{\mathbf{D}}$ be a scaling diagonal matrix with diagonal elements equal to the reciprocal of the square root of the diagonal elements of $\mathbf{R}^*$, that is, we let the diagonal elements of $\widetilde{\mathbf{D}}$ be $1/\sqrt{2}, \ldots, 1/\sqrt{6}$. Then, $\mathbf{R} = \widetilde{\mathbf{D}}\mathbf{R}^*\widetilde{\mathbf{D}}$. Finally we set the copula parameter

$$\boldsymbol{\theta} = \widetilde{\mathbf{D}}\boldsymbol{\theta}^* = (0.707, -0.577, 0.500, 0.447, 0.408)^\top.$$

We again consider $N = 1000$ Monte-Carlo repetitions but only sample size $n = 250$, and $q = 2$ or $q = 10$. The nuisance parameters are specified as follows:

- We randomly generate the elements in $\mathbf{B}$ to be either 1 or -1, each with probability $1/2$. The same $\mathbf{B}$ is kept for all the $N = 1000$ Monte-Carlo repetitions.
- The distribution function $F_{\mathbf{X}}$ is a multivariate t-distribution (with 3 degrees of freedom) whose shape matrix has unit diagonal elements and off-diagonal elements equal to $\rho = 0.5$.
- The marginal distribution functions F_k, $k \in [p]$ are the Cauchy distribution.

The performance of the estimators $\boldsymbol{\theta}_n^{\mathrm{PLE}} = (\theta_{n,1}^{\mathrm{PLE}}, \ldots, \theta_{n,5}^{\mathrm{PLE}})^\top$ and $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} = (\widehat{\theta}_{n,1}^{\mathrm{PLE}}, \ldots, \widehat{\theta}_{n,5}^{\mathrm{PLE}})^\top$ of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_5)^\top$ is compared in the two panels, corresponding to $q = 2$ and $q = 10$ respectively, in Figure 3. In each panel we plot the deviations $\theta_{n,m}^{\mathrm{PLE}} - \theta_m$ and $\widehat{\theta}_{n,m}^{\mathrm{PLE}} - \theta_m$ next to each other, and sequentially for $m = 1, \ldots, 5$. Even under $q = 10$, we can hardly discern any significant deterioration of the residual-based estimator as compared to its oracle counterpart. (Some minor differences do exist, such as the slightly larger distance between the whiskers for $\widehat{\theta}_{n,5}^{\mathrm{PLE}}$ at $q = 10$ as compared to the same estimator at $q = 2$ and its oracle counterpart.)

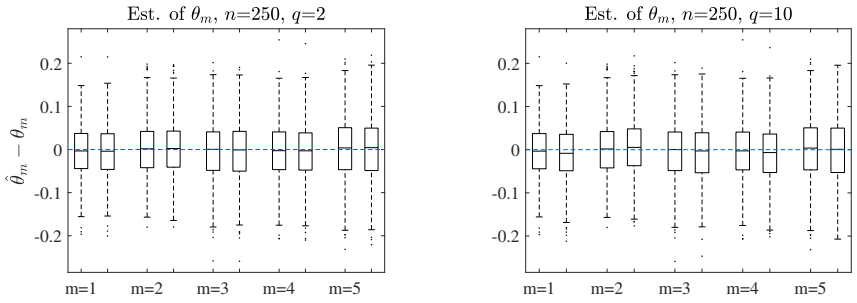

Figure 3. Simulation results for the factor model under the parameters specified in the main text, for $(p, q) = (5, 2)$ or $(p, q) = (5, 10)$, and $\rho = 0.5$.

6. Conclusion

In this paper we have studied the (semiparametrically efficient) estimation of the copula parameter in the Gaussian copula model when adjusted for linear regression. We have provided explicit rates for the distances between the residual-based estimators $\widehat{\mathbf{R}}_n^{\rho}$ and $\widehat{\mathbf{R}}_n$ on the one hand, and their oracle counterparts on the other. Under mild condition on $\widehat{\mathbf{B}}$, the residual-based estimators and their oracle counterparts are asymptotically equivalent. New result on residual empirical process, which could be of independent interest, is also derived. Finally, we have shown that the one-step estimator in conjunction with the residual ranks remains semiparametrically efficient.

Supplement

Supplement to the paper: "Inference for semiparametric Gaussian copula model adjusted for linear regression using residual ranks". The supplement contains more detailed derivations for Section 4 on semiparametric efficiency, most proofs for the paper, some additional results for Section 5, and potential generalization to high dimensions.

Acknowledgements

Research supported by the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409), by the IAP Research Network P7/06 of the Belgian State, and by the GOA-project 12/014 from the Research Council KU Leuven.

References

[1] Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.*, 28(3):549–567.

[2] Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, pages 111–150, Berkeley, Calif. University of California Press.

[3] Basrak, B. and Klaassen, C. A. J. (2013). *Efficient estimation in the semiparametric normal regression-copula model with a focus on QTL mapping*, volume Volume 9 of *Collections*, pages 20–32. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

[4] Berghaus, B., Bücher, A., and Volgushev, S. (2017). Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli*, 23(1):743–772.

[5] Bickel, P. (1982). On adaptive estimation. *Ann. Statist.*, 10(3):647–671.

[6] 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.

[7] Cai, T. T. and Zhang, L. (2018). High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Stat. Sinica*, 28:963–993.

[8] Chen, G. and Lockhart, R. A. (2001). Weak convergence of the empirical process of residuals in linear models with many parameters. *Ann. Statist.*, 29(3):748–762.

[9] Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *J. Amer. Statist. Assoc.*, 101:1228–1240.

[10] El Maache, H. and Lepage, Y. (2003). Spearman's rho and Kendall's tau for multivariate data sets. In *Mathematical statistics and applications: Festschrift for Constance van Eeden*, volume 42 of *IMS Lecture Notes Monogr. Ser.*, pages 113–130. Inst. Math. Statist., Beachwood, OH.

[11] Fan, J., Xue, L., and Zou, H. (2016). Multitask quantile regression under the transnormal model. *J. Amer. Statist. Assoc.*, 111:1726–1735.

[12] 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.

[13] Gijbels, I., Omelka, M., and Veraverbeke, N. (2014). Preadjusted non-parametric estimation of a conditional distribution function. *J. R. Stat. Soc. Ser. B*, 76(2):399–438.

[14] Gijbels, I., Omelka, M., and Veraverbeke, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics*, 42(4):1109–1126.

[15] Gijbels, I., Van Keilegom, I., and Zhao, Y. (2018). Inference for covariate-adjusted semiparametric gaussian copula model using residual ranks. FEB Research Report KBI_1829 accessible at https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS2344364&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US.

[16] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Prague: Academia.

[17] Hájek, J., Šidák, Z., and Sen, P. K. (1999). *Theory of Rank Tests*. Academic Press,

3 edition.
[18] Hoff, P. D., Niu, X., and Wellner, J. A. (2014). Information bounds for Gaussian copulas. *Bernoulli*, 20(2):604–622.
[19] Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. Appl. Probab.*, 34:587–608.
[20] Kendall, M. G. and Gibbons, J. D. (1990). *Rank Correlation Methods*. Oxford University Press, U.S.A., 5th edition.
[21] 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.
[22] Koul, H. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*.
[23] Kruskal, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.*, 53:814–861.
[24] Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, New York.
[25] Lee, S. and Wei, C.-Z. (1999). On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist.*, 27(1):237–261.
[26] Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.*, 40:2293–2326.
[27] Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression. *Electron. J. Statist.*, 6:1517–1549.
[28] Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, 2nd edition.
[29] Omelka, M. (2006). *Second order properties of some M-estimators and R-estimators*. PhD thesis, Charles University in Prague.
[30] Omelka, M., Šárka Hudecová, and Neumeyer, N. (2019). Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models. arXiv:1903.04221.
[31] Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–57.
[32] Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.*, 79:871–880.
[33] Ruymgaart, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *Ann. Statist.*, 2:892 – 910.
[34] 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(5):1911–1940.
[35] Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics*, 65(1):60–68.
[36] Susyanto, N. and Klaassen, C. A. J. (2017). Semiparametrically efficient estimation of constrained Euclidean parameters. *Electron. J. Statist.*, 11(2):3120–3140.
[37] van der Vaart, A. (2002). *Semiparametric Statistics*, pages 331–457. Number 1781 in Lecture Notes in Math. Springer. MR1915446.
[38] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
[39] Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann. Statist.*, 40:2541–2571.