Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

INFERENCE FOR COVARIATE-ADJUSTED SEMIPARAMETRIC GAUSSIAN COPULA MODEL USING RESIDUAL RANKS*

By Irène Gijbels†, Ingrid Van Keilegom† and Yue Zhao†

KU Leuven†

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 coefficient. 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.

1. Introduction.

1.1. *Background.* Let $\mathbf{E} = (E_1, \dots, E_p)^\top \in \mathbb{R}^p$ be a random vector; we assume throughout that E_k, $k \in \{1, \dots, p\}$ 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 [33]) states that the dependence structure of $\mathbf{E}$ can be uniquely described by its associated copula C, via

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

where $\overline{\mathbb{R}}$ denotes the extended real line. By the "inverse" Sklar's theorem (e.g., Corollary 2.10.10 in [33]),

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

where for $k \in \{1, \dots, p\}$,

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

denotes the left-continuous inverse of F_k. The copula C is equivalently the joint distribution function of the transformed random vector $(F_1(E_1), \dots, F_p(E_p))^\top$, 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 [33]). Copulas provide a modular approach

*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.

MSC 2010 subject classifications: Primary 62F12, 62G20; secondary 62G30, 62H20

Keywords and phrases: Gaussian copula, normal scores rank correlation coefficient, residual rank, Spearman's rho, semiparametric efficiency

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 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 copula in the semiparametric copula model 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 $\mathbf{E}$,

$$(\Phi^{\leftarrow}(F_1(E_1)), \dots, \Phi^{\leftarrow}(F_p(E_p)))^\top \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R}); \tag{1.3}$$

throughout the paper the symbol "$\sim$" denotes equality in distribution and $\Phi^{\leftarrow}$ 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

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

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 assumes a working unrestricted model for $\mathbf{R}$, and constructs a preliminary estimator $\widetilde{\mathbf{R}}$ of $\mathbf{R}$; 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., [3, 27, 23, 37], 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. [37] also studied an estimator that achieves the semiparametric lower bound by adapting the *one-step method* (see, e.g., Section 25.8 in [41]).

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 $\mathbf{G}$ [29]. 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 [31, 43]. 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: Covariate-adjusted semiparametric Gaussian copula model.* Suppose that a $p \times 1$ random vector $\mathbf{E} = (E_1, \dots, E_p)^\top$ 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 $\mathbf{Y} = (Y_1, \dots, Y_p)^\top$ and a $q \times 1$ covariate vector $\mathbf{X} = (X_1, \dots, X_q)^\top$ 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. 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 $(\mathbf{Y}, \mathbf{X})$. As before, we will focus on the statistical inference problem for $\mathbf{R}$. Recall that the p components of $\mathbf{E}$ have univariate distribution functions $F_1, \dots, F_p$ 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 $F_1, \dots, F_p$ and $F_{\mathbf{X}}$ are all nuisance parameters. We call this model the *covariate-adjusted (semiparametric) Gaussian copula model*; later for brevity we will often refer to this model simply as the "regression setting."

To motivate this study, let us assume for the moment a simpler model where under (1.5), $\mathbf{E}$ is just Gaussian (instead of having a Gaussian copula), and we again observe a sample of $(\mathbf{Y}, \mathbf{X})$. (Then, of course, $\mathbf{E}$ still has a Gaussian copula; however, its marginals are all fixed to be univariate Gaussian.) This model is often called a covariate-adjusted/conditional Gaussian graphical model in the literature. The graphical interpretation here comes from the Gaussian vector $\mathbf{E}$. As the sample of $\mathbf{E}$ is never observed, however, one can only infer the associated Gaussian graphical model through the response $\mathbf{Y}$. The qualifier "covariate-adjusted" reflects the fact that the Gaussian graphical model underlying $\mathbf{Y}$ is obtained after removing the influence of the covariate $\mathbf{X}$. The covariate-adjusted Gaussian graphical model has recently been studied by [5, 10] among others. These authors have shown that removing the external effects associated with $\mathbf{X}$ can significantly improve the detection of the intrinsic connections associated with $\mathbf{E}$.

Given the extension of Gaussian distributions in one direction to the Gaussian copula model, and in the other direction to accommodate a covariate in the covariate-adjusted Gaussian graphical model, it is clear that the covariate-adjusted Gaussian copula model is a natural but difficult next step that combines both of the previous extensions. Moreover, our study 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 $\boldsymbol{\theta}$; this theme will be repeated throughout this paper. To achieve this goal, 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 the copula correlation matrix.

1.3. *Relation to existing studies.* Gaussian copulas have been extended in a number of ways to incorporate a covariate, sometimes in a regression structure, 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.

First, a parametrically specified Gaussian copula can be used to create a dependence structure for the different components of $\mathbf{Y}$, and the marginal distributions of $\mathbf{Y}$ can also be specified to have a parametric form. For each marginal distribution, its associated parameters may depend on the covariate $\mathbf{X}$, and may also be of interest. This is the approach taken in, e.g., [32, 38]. Because this approach requires parametrically specified marginals, it differs significantly from our approach where the marginals are unspecified.

More recently, [6, 14] also treated the linear regression model (1.5) but they restricted $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$ to be jointly multivariate normal. (To be precise, [6] considered the case $p = 1$, while [14] considered a reparametrization of (1.5), presented in their Eq. (1), in terms of the covariance matrix of $(\mathbf{Y}, \mathbf{X})$.) It is further assumed that none of the sample of $\mathbf{Y}, \mathbf{X}, \mathbf{E}$ is available. Instead, [6, 14] 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 $g : \mathbb{R}^r \to \mathbb{R}^r$ such that, for each $g \in \mathcal{G}_r$, there exist strictly increasing univariate functions $g_1, \ldots, g_r$ so that for all $x_1, \ldots, x_r \in \mathbb{R}$, one has $g(x_1, \ldots, x_r) = (g_1(x_1), \ldots, g_r(x_r))^\top$. Then [6, 14] 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 transformation functions $g \in \mathcal{G}_p$ and $h \in \mathcal{G}_q$. In turn, the estimator of $\mathbf{B}$ is used to predict $\widetilde{\mathbf{Y}}$ from a realization of $\widetilde{\mathbf{X}}$, and this is done without assuming a linear relationship between $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}$. [6] called this model the Gaussian copula regression model, because the joint distribution of the observable $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ has a Gaussian copula.

Note that our approach and that of [6, 14] 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 covariate-adjusted Gaussian copula model, the copula component $\mathbf{E}$ is the object of interest and it is the object of interest that has been perturbed, just as in the covariate-adjusted Gaussian graphical model. In contrast, in the approach of [6, 14], the (Gaussian) $\mathbf{E}$ is largely a nuisance parameter in the prediction task. Technically, our approach is arguably more challenging because we never observe a sample of the Gaussian copula component $\mathbf{E}$ directly (whereas [6, 14] 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 meaningful 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. We briefly address the representative papers [7, 19, 20] here. [7] studied asymptotic distribution of copula parameters via pseudo-likelihood estimators in their so-called semiparametric copula-based multivariate GARCH models, Here, univariate GARCH models describe marginal risk series and a parametric copula creates a dependence structure for the different marginal risks; the residual marginal risks are obtained from the observed responses after adjusting for the estimated marginal standard deviations. Note that to the best of our knowledge detailed treatment of semiparametric efficiency (including the semiparametric lower bound and the efficient one-step method) in copula models appears to be rather restricted to Gaussian copulas, and the pseudo-likelihood estimators typically do not achieve semiparametric efficiency (even within Gaussian copulas); see, e.g., [18, 23]. Thus, apart from a model setup that is different from ours, naturally [7] does not address semiparametric efficiency, which is among our research objectives. [19, 20] are set 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. [19, 20] studied the asymptotics of the resulting *empirical copula process* based on the estimated conditional distribution of $\mathbf{Y}$ given X. Conditional copulas certainly encompass our linear regression model (1.5) as a special case. Note that 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 coefficient), 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 [2]).

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. However, we believe that the prevalence of Gaussian copulas (because of the possibility to achieve semiparametric efficiency and its connection to graphical models) warrants a dedicated treatment, and the specific linear regression form also allows for a more refined analysis, as well as potential generalization to high-dimensional settings, as we briefly discuss in Section 6.

1.4. *Organization of the paper.* Section 2 formally introduces our model and estimation procedures. Section 3 presents the asymptotic normality 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 coefficient in Section 3.3. 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 the results of a simulation study comparing the estimators based on the residual ranks and the oracle estimators (to be introduced in Section 2.2.1). Section 6 concludes and also 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 a supplement, 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 $\{1, \dots, a\}$.

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 of the argument.

We let $L_2[0, 1]$ be the Hilbert space of functions on the interval $[0, 1]$ that are square-integrable with respect to the Lebesgue measure. We let $L_2^0[0, 1]$ be the closed (linear) subspace of $L_2[0, 1]$ resulting from the restriction that if $h \in L_2^0[0, 1]$, then $\int_{[0,1]} h(u)\mathrm{d}u = 0$, and let $L_2^{\mathrm{d}}[0, 1]$ be the dense subspace of $L_2[0, 1]$ consisting of continuously differentiable functions (e.g., Proposition 8.17 in [15]). It is easy to show that $L_2^{0,\mathrm{d}} \equiv L_2^0[0, 1]\cap L_2^{\mathrm{d}}[0, 1]$ is dense in $L_2^0[0, 1]$. Analogously, we let $L_2(F_{\mathbf{X}})$ be the Hilbert space of functions $h : \mathbb{R}^q \to \mathbb{R}$ that are square-integrable with respect to the measure $F_{\mathbf{X}}$ (i.e., $\int_{\mathbb{R}^q} h^2\mathrm{d}F_{\mathbf{X}} < \infty$ iff $h \in L_2(F_{\mathbf{X}})$). We let $L_2^0(F_{\mathbf{X}})$ be the closed subspace of $L_2(F_{\mathbf{X}})$ resulting from the restriction that if $h \in L_2^0(F_{\mathbf{X}})$, then $\int_{\mathbb{R}^q} h\mathrm{d}F_{\mathbf{X}} = 0$.

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 $\mathrm{P}_{\boldsymbol{\theta},\mathbf{B},F_1,\dots,F_p,F_{\mathbf{X}}}$ 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.
- (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 $f_{\mathbf{X}}$.
- (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 6 where we discuss high-dimensional generalization. Note also that with condition (iv) above, the covariate $\mathbf{X}$ cannot contain an intercept term. Consequently, no location constraint is placed on $F_1, \dots, F_p$.

For brevity henceforth we abbreviate $\mathrm{P}_{\boldsymbol{\theta},\mathbf{B},F_1,\dots,F_p,F_{\mathbf{X}}}$ 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., [11] 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 the covariate-adjusted Gaussian copula model. 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. *Procedures 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 $\mathbf{E}$ as

$$F_{n,k}(t) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{E_{i,k} \leq t\},$$

and its rescaled version as

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

We need the above rescaled empirical marginal distribution function so that applying $\Phi^{\leftarrow}$ to it later will 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 coefficient estimator $\mathbf{R}_n = [r_{n,kk'}]_{k,k' \in [p]}$, defined as

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

see, e.g., Eq. (7) on p. 113 in [22]. Here ϕ_n is a deterministic correction factor given by

$$\phi_n = \left[\frac{1}{n} \sum_{i \in [n]} \left\{\Phi^{\leftarrow}\left(\frac{i}{n+1}\right)\right\}^2\right]^{-1} = 1 + \mathcal{O}(n^{-1}\log(n)). \tag{2.2}$$

Using the Gaussianized observations $\mathbf{Z}_i^{(n)} \equiv (\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,1}(E_{i,1})), \ldots, \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,p}(E_{i,p})))^\top$, we can write $\mathbf{R}_n$ explicitly in matrix form as

$$\mathbf{R}_n = \frac{\phi_n}{n} \sum_{i \in [n]} \mathbf{Z}_i^{(n)} \mathbf{Z}_i^{(n)\top}. \tag{2.3}$$

The correction ϕ_n is asymptotically insignificant, but with it the diagonal elements of $\mathbf{R}_n$, and analogously of $\widehat{\mathbf{R}}_n$ given later in (2.6), all equal to one, and so $\mathbf{R}_n$ and $\widehat{\mathbf{R}}_n$ become genuine correlation matrices. The elements of $\mathbf{R}_n$ belong to multivariate rank order statistics (in this case, with score function $\Phi^{\leftarrow}$) that are common in the literature; see [22, 36] 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); see Section 3.4 for more details.

Alternatively, we can estimate $\mathbf{R}$ by inversion of Spearman's rho. As described, e.g., in [25, 28], 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

$$\begin{aligned}\rho_{n,kk'} &= \frac{12n}{n^2-1}\sum_{i\in[n]}\left\{F_{n,k}(E_{i,k}) - \frac{n+1}{2n}\right\}\left\{F_{n,k'}(E_{i,k'}) - \frac{n+1}{2n}\right\} \\ &= 1 - 6\frac{n}{n^2-1}\sum_{i\in[n]}\left\{F_{n,k}(E_{i,k}) - F_{n,k'}(E_{i,k'})\right\}^2; \end{aligned} \tag{2.4}$$

see, e.g., Eqs. (11) and (12) on p. 113 to 114 in [22]. 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 holds between the elements of the copula correlation matrix $\mathbf{R} = [r_{kk'}]_{k,k'\in[p]}$ and 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 [24]. A plug-in estimator of $\mathbf{R}$ via inversion of Spearman's rho is then given by $\mathbf{R}_n^\rho = [r^\rho_{n,kk'}]_{k,k'\in[p]}$ with

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

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. *Procedures for the covariate-adjusted Gaussian copula model.* In the covariate-adjusted Gaussian copula model, 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 denote the kth column of $\mathbf{B}$ by $\mathbf{B}_k$. Then the kth component of (1.5) reads

$$Y_k = \mathbf{B}_k^\top \mathbf{X} + E_k.$$

We let $\widehat{\mathbf{B}} = \widehat{\mathbf{B}}^{(n)}$ be an estimator (sequence) of $\mathbf{B}$, and denote the kth column of $\widehat{\mathbf{B}}$ by $\widehat{\mathbf{B}}_k$.

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, which we also regard as an estimator of $\mathbf{E}_i$, defined as

$$\widehat{E}_{i,k} = Y_{i,k} - \widehat{\mathbf{B}}_k^\top \mathbf{X}_i = E_{i,k} - (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i, \quad \forall k \in [p]. \tag{2.5}$$

For brevity we will suppress the dependence of $\widehat{\mathbf{E}}_i$ on n (which it inherits from $\widehat{\mathbf{B}}$). Then, for each $k \in [p]$, let $\widehat{F}_{n,k}$ be the (empirical marginal) residual distribution function for the kth coordinate of $\mathbf{E}$, which we regard as an estimator of $F_{n,k}$, defined as

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

and let $\widehat{F}^{\mathrm{r}}_{n,k}$ be the estimator of $F^{\mathrm{r}}_{n,k}$ defined as

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

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 coefficient estimator of $\mathbf{R}$ based on the residual ranks given by

$$\widehat{r}_{n,kk'} = \frac{\phi_n}{n} \sum_{i \in [n]} \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'})), \quad \forall k, k' \in [p], \tag{2.6}$$

where we recall the correction factor ϕ_n from (2.2). Using the Gaussianized observations

$$\widehat{\mathbf{Z}}^{(n)}_i = \widehat{\mathbf{Z}}_i \equiv (\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1})), \ldots, \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p})))^{\top}, \tag{2.7}$$

we can write $\widehat{\mathbf{R}}_n$ explicitly in matrix form as

$$\widehat{\mathbf{R}}_n = \frac{\phi_n}{n} \sum_{i \in [n]} \widehat{\mathbf{Z}}_i \widehat{\mathbf{Z}}_i^{\top}. \tag{2.8}$$

Note that $\widehat{\mathbf{R}}_n$ is obtained from (2.1), or in matrix form from (2.3), 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

$$\widehat{\rho}_{n,kk'} = 1 - 6\frac{n}{n^2-1} \sum_{i \in [n]} \left\{ \widehat{F}_{n,k}(\widehat{E}_{i,k}) - \widehat{F}_{n,k'}(\widehat{E}_{i,k'}) \right\}^2. \tag{2.9}$$

Similar to (2.6), $\widehat{\rho}_{n,kk'}$ is obtained from (2.4) 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

$$\widehat{r}^{\rho}_{n,kk'} = 2 \sin\left(\frac{\pi}{6} \widehat{\rho}_{n,kk'}\right), \quad \forall k, k' \in [p]. \tag{2.10}$$

The perturbation term $(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^{\top} \mathbf{X}_i$ in (2.5) will likely cause the residual ranks to deviate from the oracle ranks. However, as we will demonstrate in Section 3.1, the difference between the residual ranks and the oracle ranks is uniformly small. Consequently, the differences between the estimators based on the residual ranks and their counterparts based on the oracle ranks should be small as well, allowing us to conclude that these estimators have the same asymptotic distribution. Rigorously demonstrating this fact will be a major focus for the remainder of this paper.

3. Asymptotic normality 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 normality results for these estimators. Some preliminary results regarding the residual ranks are presented in Section 3.1. These results concern the individual coordinates of $\mathbf{E}$ and 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). 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 (2.10) via inversion of Spearman's rho. Then, Section 3.3 treats the normal scores rank correlation coefficient estimator $\widehat{\mathbf{R}}_n$ in (2.6). 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 $\mathbf{R}(\boldsymbol{\theta})$. As we will see, the estimator $\widehat{\mathbf{R}}_n^{\rho}$ is easier to analyze, but the estimator $\widehat{\mathbf{R}}_n$ is more closely related to the pseudo-likelihood method in Section 3.4 and to the one-step method in Section 4.2.

We do not require a specific form of the estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}$, 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 a linear regression model a $n^{-1/2}$ convergence rate for $\widehat{\mathbf{B}}$ is the most common one, a different convergence rate is possible. For instance, Rousseeuw's robust *least median of squares regression* introduced in [35] yields a slower, $n^{-1/3}$ convergence rate for $\widehat{\mathbf{B}}$ (see, e.g., Example 6.3 in [26]). As another example, if we consider a case where the law P may change 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 6, this topic is treated in more details when specifically the ambient dimensions p and q vary with n. 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 $\log(n)/n \leq \delta_{\mathbf{B},n} = o(1)$.*

The sole purpose for placing a lower bound on $\delta_{\mathbf{B},n}$ in Assumption 3.1 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 $\mathbb{E}[\|\mathbf{X}\|^2] \leq \infty$, 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 $\max_{i\in[n]} \|\mathbf{X}_i\|\delta_{\mathbf{B},n} = o_{\mathrm{p}}(1)$.

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 $\widehat{F}_{n,k}(\widehat{E}_{i,k})$ and the corresponding unobserved oracle ranks $F_{n,k}(E_{i,k})$.

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 $\mathbb{E}\left[\mathbb{1}\left\{E_k \le \cdot + (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}\right\} | \mathcal{X}_n\right] = \mathbb{E}\left[\mathbb{1}\left\{\widehat{E}_k \le \cdot\right\} | \mathcal{X}_n\right]$ which equals the conditional probability $\mathbb{P}(\widehat{E}_k \le \cdot | \mathcal{X}_n)$. Then, let the "oscillation-like" remainder term (e.g., p. 243 in [30] or Lemma 1 in [1]) common in the analysis of residual empirical processes be, for $t \in \mathbb{R}$,

$$\begin{aligned} r_{1n,k}(t) &= \widehat{F}_{n,k}(t) - F_{n,k}(t) - \mathbb{P}(\widehat{E}_k \le t | \mathcal{X}_n) + F_k(t) \\ &= \frac{1}{n} \sum_{i \in [n]} \left\{ \mathbb{1}\{\widehat{E}_{i,k} \le t\} - \mathbb{1}\{E_{i,k} \le t\} - \mathbb{P}(\widehat{E}_k \le t | \mathcal{X}_n) + \mathbb{P}(E_k \le t) \right\}. \end{aligned} \tag{3.1}$$

Analogous to above, for analyzing normal scores rank correlation coefficients, let

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

Next, let the additional remainder terms be

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

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

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

PROPOSITION 3.1. *Under the law* P, *for all* $n \ge 1$, $k \in [p]$ *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 and 3.2 and 3.3(i) hold, then for all $k \in [p]$,

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

furthermore,

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

PROOF. First, (3.6) follows because

$$\widehat{F}_{n,k}(t) - F_k(t) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{E_{i,k} \le t\} - F_k(t) + \frac{1}{n} \sum_{i \in [n]} \left(\mathbb{1}\{\widehat{E}_{i,k} \le t\} - \mathbb{1}\{E_{i,k} \le t\} \right)$$

$$
\begin{aligned}
&= F_{n,k}(t) - F_k(t) + \frac{1}{n}\sum_{i\in[n]}\Big(\mathbb{1}\{\widehat{E}_{i,k}\le t\} - \mathbb{1}\{E_{i,k}\le t\} - \mathbb{P}(\widehat{E}_k \le t|\mathcal{X}_n) + F_k(t)\Big)\\
&\quad + \mathbb{P}(\widehat{E}_k\le t|\mathcal{X}_n) - F_k(t)\\
&= F_{n,k}(t) - F_k(t) + f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}] + r_{1n,k}(t) + r_{2n,k}(t).
\end{aligned}
$$

The derivation of (3.7) is completely analogous. The proofs for the bounds in (3.8) and (3.9) are deferred to Proposition A.1 and Lemma A.4 in the supplement respectively. □

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 $r_{1n,k} + r_{2n,k}$. Decompositions similar to this are common in the literature on residual empirical processes; see, e.g., Theorem 1 in [1] or Theorem 1 in [9]. (Even in a linear regression setting, as is the case in [9], 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 [9] in fact studied fixed design, with fixed $\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top$ being the rows of a $n\times q$ design matrix, so additional care is necessary for a truly precise comparison.)
- Next, we can obtain the following weakened, but simpler form of (3.8), simply by replacing $f_k(t)$ with $\|f_k\|_{L_\infty} < \infty$:

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

  In particular, when $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$, (3.10) implies $\sup_{t\in\mathbb{R}} |r_{1n,k}(t)| = \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n) n^{-3/4}\right)$. 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] and Theorem 1 in [9], when taken at face value, simply state the remainder terms as being $o_{\mathrm{p}}(n^{-1/2})$.
- Finally, for t such that the term $\left\{\log(\delta_{\mathbf{B},n}^{-1}) n f_k(t) \delta_{\mathbf{B},n}\right\}^{1/2}$ in the denominator on the left hand side of (3.8) dominates (i.e., when $f_k(t) \ge \delta_{\mathbf{B},n} \vee (\log^2(n)\log^{-1}(\delta_{\mathbf{B},n}^{-1}) n^{-1} \delta_{\mathbf{B},n}^{-1})$), 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 of our result that differs from the literature. 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 $\Phi^{\leftarrow}$ in the normal scores rank

correlation coefficient 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. 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 [36] relies on its Lemma 4.2).

The leading term in (3.6) 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 terms in (3.11) and (3.12) in Proposition 3.2 below, $\mathbb{E}[\mathbf{X}]$ is replaced by $\mathbf{X}_i - \mathbb{E}[\mathbf{X}]$. Thus for the residual ranks these terms 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 the asymptotics of our estimators of $\mathbf{R}$ based on the residual ranks, and will allow us to conclude that the asymptotics of our estimators of $\mathbf{R}$ do not depend on the estimation of $\mathbf{B}$, 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 equalities*

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

$$\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k}) = -f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top(\mathbf{X}_i - \mathbb{E}[\mathbf{X}]) + r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i} \tag{3.12}$$

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

$$\max_{i\in[n]} \left| \frac{r_{3n,k,i}}{\log^{1/2}(n)\, n^{-1/2} \left\{ f_k^{1/2}(E_{i,k}) \|\mathbf{X}_i\|^{1/2} \delta^{1/2}_{\mathbf{B},n} + \|\mathbf{X}_i\| \delta_{\mathbf{B},n} \right\} + \log(n)\, n^{-1}} \right| = \mathcal{O}_{\mathrm{p}}(1), \tag{3.13}$$

$$\max_{i\in[n]} \left| \frac{r_{4n,k,i}}{\|\mathbf{X}_i\|(\|\mathbf{X}_i\| + \|\mathbb{E}[\mathbf{X}]\|)\delta^2_{\mathbf{B},n}} \right| = \mathcal{O}_{\mathrm{p}}(1)\,. \tag{3.14}$$

PROOF. We fix arbitrary $k \in [p]$, and first prove (3.11). First, (3.6) with $t = \widehat{E}_{i,k}$ gives

$$\widehat{F}_{n,k}(\widehat{E}_{i,k}) = F_{n,k}(\widehat{E}_{i,k}) + r_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + f_k(\widehat{E}_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]. \tag{3.15}$$

For the first term $F_{n,k}(\widehat{E}_{i,k})$ on the right hand side of (3.15), sequentially substituting in $r_{3n,k,i}$ (see (3.4)) and $r_{4n,k,i}$ (see (3.5)) gives

$$\begin{aligned} F_{n,k}(\widehat{E}_{i,k}) &= F_k(\widehat{E}_{i,k}) - F_k(E_{i,k}) + F_{n,k}(E_{i,k}) + r_{3n,k,i} \\ &= F_{n,k}(E_{i,k}) + r_{3n,k,i} + r_{4n,k,i} - f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top(\mathbf{X}_i - \mathbb{E}[\mathbf{X}]) \\ &\quad - f_k(\widehat{E}_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]. \end{aligned} \tag{3.16}$$

Combining (3.15) and (3.16) yields (3.11). The proof for (3.12) is completely analogous to the one for (3.11), except that we now start from (3.7) with $t = \widehat{E}_{i,k}$.

The proofs for (3.13) and (3.14) are deferred to Lemma A.2 and Lemma A.4 in the supplement respectively. □

3.2. *Inversion of Spearman's rho.*

THEOREM 3.3. *Assume the law* P, *and that Assumptions 3.1, 3.2 and 3.3 hold. Then,*

$$\sqrt{n}\left(\widehat{\rho}_{n,kk'} - \rho_{n,kk'}\right) = \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\delta_{\mathbf{B},n}^{1/2} + n^{1/2}\delta_{\mathbf{B},n}^{2}\right), \quad \forall k,k' \in [p]. \tag{3.17}$$

Thus, if furthermore

$$\delta_{\mathbf{B},n} = o(n^{-1/4}), \tag{3.18}$$

then $\sqrt{n}\left(\widehat{\rho}_{n,kk'} - \rho_{n,kk'}\right) = o_{\mathrm{p}}(1)$, $\forall k,k' \in [p]$.

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

Theorem 3.3 immediately yields that when the conditions of the theorem hold and when the uncertainty $\widehat{\mathbf{B}} - \mathbf{B}$ is not too large (precisely, when (3.18) holds, which in particular allows a rate slower than $n^{-1/2}$), the matrix $\sqrt{n}[\widehat{\rho}_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$ relating to Spearman's rho is at most $o_{\mathrm{p}}(1)$ away from its oracle counterpart $\sqrt{n}[\rho_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$. Thus in particular the asymptotic distributions of the two matrices 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 [12]. (To be precise, [12] gives the asymptotic distribution of (in their notation) $\mathcal{S}_n$ which is a (matrix) U-statistic of degree three. The difference between $\mathcal{S}_n$ and our matrix $[\rho_{n,kk'}]_{k,k'\in[p]}$ (which is D_n in [12]) is only $\mathcal{O}_p(1/n)$ though; see, e.g., the second equation display on p. 118 in [12].)

Finally, 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 (2.10) via inversion of Spearman's rho, will follow by the Delta method. This asymptotic distribution will coincide with that of $\sqrt{n}(\mathbf{R}_n^{\rho} - \mathbf{R})$, where $\mathbf{R}_n^{\rho}$ is the oracle counterpart of $\widehat{\mathbf{R}}_n^{\rho}$ in the ordinary Gaussian copula model, due to the asymptotic equivalence of the matrices $\sqrt{n}[\widehat{\rho}_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$ and $\sqrt{n}[\rho_{n,kk'} - \rho_{kk'}]_{k,k'\in[p]}$. This is our first concrete instance where the complication introduced by the addition regression structure does not affect the estimation of $\mathbf{R}$, a theme already mentioned as early as the end of Section 1.2.

3.3. *Normal scores rank correlation coefficient.* The analysis of the normal scores rank correlation coefficient estimator $\widehat{\mathbf{R}}_n$ in (2.6) 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 (3.24), and the rate in Theorem 3.4 below. We recall $\delta_{\mathbf{B},n}$ from Assumption 3.1.

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* τ *satisfying*

$$\tau > 1/4. \tag{3.19}$$

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

$$\delta_{\mathbf{B},n} = \mathcal{O}(n^{-\tau}). \tag{3.20}$$

In the canonical case, $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$; then, by reasoning similar to the proof of Lemma A.3 in the supplement, for any $\tau \le 1/2$, 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*

$$\frac{1}{2} < \gamma < \min\{2\tau, 1\} \tag{3.21}$$

(note that for τ satisfying (3.19), the range in (3.21) is not empty) such that, for each $k \in [p]$,

$$\sup_{u\in(\delta,1-\delta)} \frac{f_k \circ F_k^{\leftarrow}(u)}{u \wedge (1-u)} = o\left(\delta^{-\frac{1}{\gamma}\tau}\right) \tag{3.22}$$

as $\delta \downarrow 0$.

If Assumption 3.5 is satisfied for some γ, we define a partition of the interval $(0,1)$ into

$$A_1 = A_1^{(n)} = (0, n^{-\gamma}] \cup [1-n^{-\gamma}, 1), \quad A_2 = A_2^{(n)} = (n^{-\gamma}, 1-n^{-\gamma}). \tag{3.23}$$

Then, we introduce the non-decreasing sequence δ_n, $n \ge 1$ as

$$\delta_n = \max_{k\in[p]} \int_{A_2} \left\{\frac{f_k \circ F_k^{\leftarrow}(u)}{u \wedge (1-u)}\right\}^2 \left\{|\Phi^{\leftarrow}(u)| \vee 1\right\}^2 \mathrm{d}u. \tag{3.24}$$

We elaborate on Assumption 3.5 and the quantity δ_n in (3.24). 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. This assumption excludes certain distributions. For instance, 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 (3.22) is of the order $1/\delta$, while if we take γ according to (3.21), 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 (3.12) of the distance between the residual ranks and the oracle ranks; the first term on the right hand side of (3.12) 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 $\Phi^{\leftarrow}$, 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. For instance, it is easy to check that Assumption 3.5 holds for the normal density and for densities with polynomial decay of the form

$$f_k(t) = \frac{a-1}{2} \frac{1}{(1+|t|)^a} \tag{3.25}$$

with $a > 1$. Moreover, in these two cases, the integral on the right hand side of (3.24) 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 (4.9), and the latter quantity should be finite for us to carry out the lower bound analysis in Section 4.

We introduce

$$\Delta_n = \log(n)\, n^{1/2-\gamma} + n^{1/4}\delta_{\mathbf{B},n}\delta_n^{1/2} + \log^{3/2}(n)\, \delta_{\mathbf{B},n}^{1/2} + \log^{1/2}(n)\, n^{1/2}\delta_{\mathbf{B},n}^2\delta_n. \tag{3.26}$$

THEOREM 3.4. *Assume the law* P, *and that Assumptions 3.1, 3.2, 3.4 and 3.5 hold (with Assumption 3.4 implying* (3.20) *with* τ *satisfying* (3.19)*). Then,*

$$\sqrt{n}(\widehat{r}_{n,kk'} - r_{n,kk'}) = \mathcal{O}_{\mathrm{p}}(\Delta_n), \quad \forall k, k' \in [p]. \tag{3.27}$$

Thus, if furthermore

$$\delta_n = o(n^{2\tau-1/2}\log^{-1/2}(n)), \tag{3.28}$$

then $\Delta_n = o(1)$ *and (component-wise)*

$$\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R}_n) = o_{\mathrm{p}}(1). \tag{3.29}$$

PROOF. At a high level, the proof is similar to that of Theorem 3.3. However, the present proof is much longer due to the technicalities encountered when treating the unboundedness of $\Phi^{\leftarrow}$ in (2.6). The detailed 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 and (3.28) 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 matrix $\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R}) = \sqrt{n}[\widehat{r}_{n,kk'} - r_{kk'}]_{k,k'\in[p]}$ is at most $o_{\mathrm{p}}(1)$ away from its oracle counterpart $\sqrt{n}(\mathbf{R}_n - \mathbf{R}) = \sqrt{n}[r_{n,kk'} - r_{kk'}]_{k,k'\in[p]}$. This is our second concrete instance where the additional regression structure does not affect the estimation of $\mathbf{R}$.

It is straightforward to derive the asymptotic distributions of $\sqrt{n}(\mathbf{R}_n - \mathbf{R})$, and hence of $\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R})$. Theorem 3.1 in [27], relying on an earlier result in [36], establishes that

$$\begin{aligned}\sqrt{n}(r_{n,kk'} - r_{kk'}) = \frac{1}{\sqrt{n}}\sum_{i\in[n]} \Big[&\Phi^{\leftarrow}(F_k(E_{i,k}))\Phi^{\leftarrow}(F_{k'}(E_{i,k'})) \\ &- \frac{r_{kk'}}{2}\left\{\Phi^{\leftarrow}(F_k(E_{i,k}))^2 + \Phi^{\leftarrow}(F_{k'}(E_{i,k'}))^2\right\}\Big] + o_{\mathrm{p}}(1).\end{aligned}$$

From this and (1.3), the asymptotic distributions of $\sqrt{n}(\mathbf{R}_n - \mathbf{R})$, and hence of $\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R})$, are a zero mean (matrix) Gaussian with a correlation structure given in, e.g., [17].

3.4. *Discussion.* 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 [37] for a brief summary of existing methods.

Of these, the pseudo-likelihood estimation (PLE) method (see [16] 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

$$\mathbb{M}(\mathbf{R}'; \widehat{\mathbf{R}}) = -\frac{p\log(2\pi)}{2} - \frac{1}{2}\log(\det \mathbf{R}') - \frac{1}{2}\operatorname{tr}\left(\mathbf{R}'^{-1}\widehat{\mathbf{R}}\right). \tag{3.30}$$

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

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

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

$$\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} = \operatorname*{argmax}_{\boldsymbol{\theta}'\in\boldsymbol{\Theta}} \mathbb{M}(\mathbf{R}(\boldsymbol{\theta}'); \widehat{\mathbf{R}}_n). \tag{3.31}$$

(In (3.31), 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 (3.30) 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 (3.31)) 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 (3.31) 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 the oracle rank-based $\mathbf{R}_n$ in (3.31). 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 (3.29) holds (i.e., when $\sqrt{n}(\widehat{\mathbf{R}}_n - \mathbf{R}_n) = o_{\mathrm{p}}(1)$),

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} - \boldsymbol{\theta}_n^{\mathrm{PLE}}) = o_{\mathrm{p}}(1). \tag{3.32}$$

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 (3.31) are close. Indeed, under (3.29), classical M-estimation theory (e.g., Theorem 3.2.16 in [42]) again yields (3.32). The asymptotic normality of the (oracle) pseudo-likelihood estimator (in general copula models) was established in [16]; this, under (3.32), 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 [27, 23, 37] and this will also have consequences for the semiparametric efficiency of $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$; we will come back to this point in Section 4.1, below Proposition 4.1.

More generally, our result on $\widehat{\mathbf{R}}_n$ will be potentially useful when analyzing any quantity involving the summation over $i \in [n]$ of quadratic forms of the Gaussianized observations (2.7), because the resulting sum will involve components of $\widehat{\mathbf{R}}_n$. In addition to the pseudo-likelihood example shown here, this will also apply to the one-step estimator (see Remark 3).

From Section 3 alone, it appears that the estimator $\widehat{\mathbf{R}}_n^{\rho}$ via inversion of Spearman's rho requires weaker conditions than the normal scores rank correlation coefficient estimator $\widehat{\mathbf{R}}_n$, cf. Assumption 3.3 for the former and Assumptions 3.4 and 3.5 for the latter. On the other hand, as we have already mentioned above, the one-step estimator will involve the estimator $\widehat{\mathbf{R}}_n$.

4. Semiparametrically efficient estimation.

Recall that the copula parameter is $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$ and $\boldsymbol{\theta}$ belongs to some parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^d$. Our treatment of the (asymptotically) semiparametrically efficient estimation of $\boldsymbol{\theta}$ is rooted in the classical Hàjek-Le Cam theory of asymptotics of statistical experiments as adapted to the semiparametric setting. For textbook treatment of this subject, see, e.g., Chapter 3 in [3] or Chapter 25 in [41]. Section 4.1 is concerned with the semiparametric lower bound for estimating $\boldsymbol{\theta}$ in the presence of the nuisance parameters $\mathbf{B}$, $F_1, \dots, F_p$ and $F_{\mathbf{X}}$. In Section 4.2, relying on Theorem 3.4 developed earlier, we show that the one-step estimator in conjunction with the residual ranks 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 [37], 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) *regular* if it has the same asymptotic distribution under any sequence of local alternatives (e.g., p. 365 in [41]). Following Definition 2.8 and Lemma 2.9 in [40], under suitable regularity conditions (differentiability in quadratic mean of suitable local parametric submodels passing through P suffices), an estimator (sequence) $\widehat{\boldsymbol{\theta}}_n$ is *(asymptotically) semiparametrically efficient* at the law P (which has the built-in requirement that the sequence is regular at this law) for estimating $\boldsymbol{\theta}$ if and only if it is asymptotically linear in the *efficient influence function*. Following Lemma 25.25 in [41], 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*:

$$\mathbf{I}^*(\boldsymbol{\theta}) = \mathbb{E}[(\dot{\mathbf{l}}_{\boldsymbol{\theta}}^* \dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\top})(\mathbf{Y}, \mathbf{X}; \mathrm{P})]. \tag{4.1}$$

Then, semiparametric efficiency of $\widehat{\boldsymbol{\theta}}_n$ is equivalent to

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\big) = \frac{1}{\sqrt{n}} \sum_{i\in[n]} \mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\mathbf{Y}_i, \mathbf{X}_i; \mathrm{P}) + o_{\mathrm{P}}(1). \tag{4.2}$$

By the Hàjek-Le Cam convolution theorem (e.g., Theorem 25.20 in [41]), 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 (4.2), then

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \rightsquigarrow \mathcal{N}_d(\mathbf{0}, \mathbf{I}^{*-1}(\boldsymbol{\theta})) \tag{4.3}$$

where $\rightsquigarrow$ denotes weak convergence; 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 [37]. 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 $\Phi^{\leftarrow}_{\bullet} : [0,1]^p \to \mathbb{R}^p$ as

$$\Phi^{\leftarrow}_{\bullet}(\mathbf{u}) = (\Phi^{\leftarrow}(u_1), \ldots, \Phi^{\leftarrow}(u_p))^\top, \quad \mathbf{u} = (u_1, \ldots, u_p)^\top \in [0,1]^p. \tag{4.4}$$

Then specifically

$$\dot{l}^{*\mathrm{o}}_{\boldsymbol{\theta},m}(\mathbf{u};\boldsymbol{\theta}) = \frac{1}{2}\Phi^{\leftarrow}_{\bullet}(\mathbf{u})^\top \Big\{\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{g}_m(\boldsymbol{\theta})) - \dot{\mathbf{S}}_m(\boldsymbol{\theta})\Big\}\Phi^{\leftarrow}_{\bullet}(\mathbf{u}), \quad \mathbf{u} \in [0,1]^p, m \in [d]. \tag{4.5}$$

In (4.5), the vector $\mathbf{g}_m(\boldsymbol{\theta}) = (g_{1,m}(\boldsymbol{\theta}), \ldots, g_{p,m}(\boldsymbol{\theta}))^\top$ is given by

$$\mathbf{g}_m(\boldsymbol{\theta}) = -\left\{\mathbf{I}_p + \mathbf{R}(\boldsymbol{\theta}) \circ \mathbf{S}(\boldsymbol{\theta})\right\}^{-1} \Big\{\dot{\mathbf{R}}_m(\boldsymbol{\theta}) \circ \mathbf{S}(\boldsymbol{\theta})\Big\}\boldsymbol{\iota}_p \tag{4.6}$$

(in (4.6), $\boldsymbol{\iota}_p$ denotes a p-dimensional vector of all ones, and $\circ$ denotes the Hadamard product), and with diag$(\mathbf{b})$ denoting the diagonal matrix with the elements of $\mathbf{b} \in \mathbb{R}^p$ arranged on the diagonal, the matrix $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{b}) \in \mathbb{R}^{p\times p}$ is given by

$$\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{b}) = \mathbf{S}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{b}) + \mathrm{diag}(\mathbf{b})\mathbf{S}(\boldsymbol{\theta}). \tag{4.7}$$

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 [37], and under which the parametric Gaussian copula model for $\mathbf{E}$ with known, uniform margins is regular (Lemma 2.2 in [37]). 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* $\mathbb{E}\left[\|\mathbf{X}\|^2\right] \le \infty$*.*
(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,*

$$\int_{[a_k,b_k]} \frac{\dot{f}_k^2}{f_k}(t)\mathrm{d}t < \infty. \tag{4.8}$$

Moreover,

$$\begin{aligned} \limsup_{\epsilon\to 0} \Bigg\{ &\int_{(0,1/2]} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \sup_{\delta:|\delta|\le\epsilon} \left\{f_k \circ F_k^{\leftarrow}((1+\delta)u)\right\}^2 \mathrm{d}u \\ &+ \int_{(1/2,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \sup_{\delta:|\delta|\le\epsilon} \left\{f_k \circ F_k^{\leftarrow}(1-(1+\delta)(1-u))\right\}^2 \mathrm{d}u \Bigg\} < \infty. \end{aligned} \tag{4.9}$$

REMARK 1. *Under Assumption 4.1(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 [37]).*

REMARK 2. *Using the first part of Lemma C.3 involving Inequality (C.2) in the supplement to bound the factor* $1/\phi^2(\Phi^{\leftarrow}(u))$ *in (4.9), it can be seen that the left hand side of (4.9) is quite similar to* δ_n*, introduced in (3.24), in our analysis of the asymptotics of the normal scores rank correlation coefficient estimator* $\widehat{\mathbf{R}}_n$*. It is also straightforward to check that condition (4.9) is satisfied by the normal density and the densities with polynomial decay of the form (3.25) that were discussed following (3.24). The perturbation by* δ *in condition (4.9) 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.*

Introduce $F_\bullet : \mathbb{R}^p \to \mathbb{R}^p$ as

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

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

$$\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{x}; \mathrm{P}) = \dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_\bullet(\mathbf{y} - \mathbf{B}^\top\mathbf{x}); \boldsymbol{\theta}), \quad \forall(\mathbf{y}, \mathbf{x}) \in \mathbb{R}^p \times \mathbb{R}^q, \tag{4.10}$$

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 (4.5). Hence the efficient information matrix* $\mathbf{I}^*(\boldsymbol{\theta})$ *in (4.1) coincides with the the efficient information matrix in the ordinary Gaussian copula model, and has the explicit representation*

$$\begin{aligned} (\mathbf{I}^*(\boldsymbol{\theta}))_{mm'} = \frac{1}{2}\,\mathrm{tr}[&\{\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{g}_m(\boldsymbol{\theta})) - \dot{\mathbf{S}}_m(\boldsymbol{\theta})\}\mathbf{R}(\boldsymbol{\theta}) \\ &\times \{\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{g}_{m'}(\boldsymbol{\theta})) - \dot{\mathbf{S}}_{m'}(\boldsymbol{\theta})\}\mathbf{R}(\boldsymbol{\theta})], \quad m, m' \in [d]. \end{aligned} \tag{4.11}$$

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

The simple relationship (4.10) 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 $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ from Section 3.4 as an example. Suppose that under the ordinary Gaussian copula model the oracle pseudo-likelihood estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ is semiparametrically efficient. Then $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ is asymptotically linear in its efficient influence function, which by isometry (e.g., Eq. (55) in Section 4.7 in [3]) is given by $\mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(F_\bullet(\cdot);\boldsymbol{\theta})$. Thus

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

(That is, (4.12) given later holds with $\widehat{\boldsymbol{\theta}}_n$ replaced by $\boldsymbol{\theta}_n^{\mathrm{PLE}}$.) Now we impose the regression structure. If in addition (3.32) holds (which is in turn implied by (3.29) as discussed in Section 3.4), then by (1.5) and (4.10), Equation (4.2) with $\widehat{\boldsymbol{\theta}}_n$ replaced by $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ holds as well, implying the semiparametric efficiency of $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ in the regression setting. Therefore, we have essentially reduced the characterization of the semiparametric efficiency of $\widehat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ to that of $\boldsymbol{\theta}_n^{\mathrm{PLE}}$, and the latter has been extensively studied in, e.g., [27, 23, 37]. In particular, in the ordinary Gaussian copula model, [27, 23] established that $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ is semiparametrically efficient in the unrestricted model for $\mathbf{R}$ (see the discussion of Example 5.3 in [39]), and more recently Section 4 in [37] established conditions for semiparametric efficiency of $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ under general constrained parametrizations $\mathbf{R}(\boldsymbol{\theta})$.

4.2. *Semiparametrically efficient estimator.* We first recall from Section 4.1 that an estimator (sequence) $\widehat{\boldsymbol{\theta}}_n$ is semiparametrically efficient at the law P for estimating $\boldsymbol{\theta}$ if and only if it satisfies (4.2). Using Proposition 4.1, this is equivalent to, under P,

$$\begin{aligned}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\big)&=\frac{1}{\sqrt{n}}\sum_{i\in[n]}\mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(F_\bullet(\mathbf{Y}_i-\mathbf{B}^\top\mathbf{X}_i);\boldsymbol{\theta})+o_{\mathrm{p}}(1)\\&=\frac{1}{\sqrt{n}}\sum_{i\in[n]}\mathbf{I}^{*-1}(\boldsymbol{\theta})\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(F_\bullet(\mathbf{E}_i);\boldsymbol{\theta})+o_{\mathrm{p}}(1).\end{aligned}\tag{4.12}$$

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 [41] for a textbook treatment. In the ordinary Gaussian copula model, [37] constructed and established the semiparametric efficiency of an one-step estimator. [39] studied a different update technique; they start from an already semiparametrically efficient estimator in an unrestricted model (e.g., the oracle pseudo-likelihood estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ for Gaussian copulas) to produce an efficient estimator in a constrained parametrization. We will adopt the approach of [37]. Because we would like the one-step estimator to be based on the residual ranks, we require the initial estimator $\boldsymbol{\theta}_n^*$ to be constructed from the residual ranks as well. These requirements are formally summarized 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 residual ranks $\widehat{F}_{n,k}(\widehat{E}_{i,k})$ or 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

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}} = \widetilde{\boldsymbol{\theta}}_n + \mathbf{I}^{*-1}(\widetilde{\boldsymbol{\theta}}_n)\left\{\frac{\phi_n}{n}\sum_{i\in[n]}\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1}),\ldots,\widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p});\widetilde{\boldsymbol{\theta}}_n)\right\}, \tag{4.13}$$

where the second term on the right is the update term. The one-step estimator above is essentially obtained from (3.3) in [37] via substitution of the (for us, unobservable) oracle ranks $F^{\mathrm{r}}_{n,k}(E_{i,k})$ by the (rescaled) residual ranks $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$. Just as in the ordinary Gaussian copula model, by Proposition 4.2 below the one-step estimator above is semiparametrically efficient for $\boldsymbol{\theta}$ in the regression setting.

REMARK 3. *The update term in (4.13) is given in terms of the efficient score $\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}$ as in a usual representation of the one-step method, but can also be rewritten to make the connection with the normal scores rank correlation coefficient estimator $\widehat{\mathbf{R}}_n$ more transparent. The efficient score $\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}$, given in (4.5), is quadratic in its argument which further leads to quadratic forms of the Gaussianized observations (2.7); specifically the mth component, $m\in[d]$, of the term in the curly bracket in (4.13) can be written alternatively as*

$$\frac{1}{2}\,\mathrm{tr}\left(\left\{\mathbf{D}_{\widetilde{\boldsymbol{\theta}}_n}(\mathbf{g}_m(\widetilde{\boldsymbol{\theta}}_n)) - \dot{\mathbf{S}}_m(\widetilde{\boldsymbol{\theta}}_n)\right\}\widehat{\mathbf{R}}_n\right),$$

where $\mathbf{g}_m$ and $\mathbf{D}_{\boldsymbol{\theta}}$ are introduced in (4.6) and (4.7) respectively.

PROPOSITION 4.2. *Suppose that Assumptions 4.1 and 4.2 hold. Moreover assume the conditions in Theorem 3.4 as well as (3.28) (so $\Delta_n = o(1)$ and (3.29) hold by that theorem). Then (under P) the one-step estimator $\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$ satisfies (4.2) (or equivalently (4.12)) and (4.3) with $\widehat{\boldsymbol{\theta}}_n$ replaced by $\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$. In particular, $\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$ is a semiparametrically efficient estimator of $\boldsymbol{\theta}$ at P.*

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

The information for estimating $\boldsymbol{\theta}$ in our regression setting should be no larger than that in the ordinary Gaussian copula model. Interestingly, by Proposition 4.2 and under the extra conditions in Theorem 3.4, the one-step estimator $\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$ satisfies (4.12) 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 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 here, the analysis in Section B will not require a matching, efficient estimator such as the $\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}}$, and so in particular does not require the extra 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 accuracy of our estimation procedures based on the residual ranks. We consider two distinct parametrizations, namely an unrestricted model and a Toeplitz model, of the copula correlation matrix $\mathbf{R}(\boldsymbol{\theta})$.

5.1. *Unrestricted model.* For the first scenario, we consider an unrestricted model for $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$ with $p = 3$, that is, each of the $d = p(p-1)/2$ elements of the upper-triangular portion of $\mathbf{R}$ is a free parameter. Specifically $\mathbf{R}$ is generated as follows:

- The elements in the upper-triangular portion of $\mathbf{R}$ are drawn independently from a normal distribution with standard deviation 0.5; if this does not produce a positive definite $\mathbf{R}$, the procedure is repeated until a positive definite matrix is obtained. 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 either $q = 2$ or $q = 10$. For each case, the elements of $\mathbf{B}$ are drawn independently from a standard normal distribution. 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 distribution or the same 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 of the N repetitions, 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 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 coefficient estimator $\widehat{\mathbf{R}}_n$, which coincides with the pseudo-likelihood estimator in this scenario as discussed in Section 3.4, and its oracle version $\mathbf{R}_n$ 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 and we omit presenting the results related to $\widehat{\mathbf{R}}_n^{\rho}$.

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 distribution 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 distribution, the Cauchy distribution, the standard normal distribution 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 $\|\widehat{\mathbf{R}}_n - \mathbf{R}\|_F$. For this row we additionally consider a naive estimator $\mathbf{R}_n^{\mathbf{Y}}$. The estimator $\mathbf{R}_n^{\mathbf{Y}}$ is the normal scores rank correlation coefficient 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, 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 will also exclude simulations done with $\widehat{\mathbf{B}}$ 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 2 presents the results for the common value of $\rho = 0.5$ but for $(p, q) = (3, 2)$ and $(p, q) = (3, 10)$. The results for the same (p, q) specifications but for $\rho = 0.9$ are deferred to Section E.1 in the supplement. Collectively, these latter results show that, as long as a reasonable estimator $\widehat{\mathbf{B}}$ (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.

5.2. *Toeplitz model.* For the second scenario, we consider 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 [37], here the oracle pseudo-likelihood estimator $\boldsymbol{\theta}_n^{\mathrm{PLE}}$ of $\boldsymbol{\theta}$ in the ordinary Gaussian copula model can be quite inefficient. In particular, [37] 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), and the efficient information matrices in the ordinary Gaussian copula model and in our regression setting coincide (see Proposition 4.1). Therefore, the one-step estimator (discussed in Section 4.2), which is semiparametrically efficient, can be expected to substantially outperform the pseudo-likelihood estimator.

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 same Cauchy distribution.

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

Figures 3 and 4 present the results for the common value of $\rho = 0.5$ but for $(p, q) = (4, 2)$ and $(p, q) = (4, 10)$ respectively. The results for the same (p, q) specifications but for $\rho = 0.9$ are deferred to Section E.2 in the supplement. 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.

Similar to [37], 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 (i.e., when $q = 2$), the one-step estimators perform almost as well as the corresponding oracle one-step estimators. When the covariate dimension is high (i.e., when $q = 10$), 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.

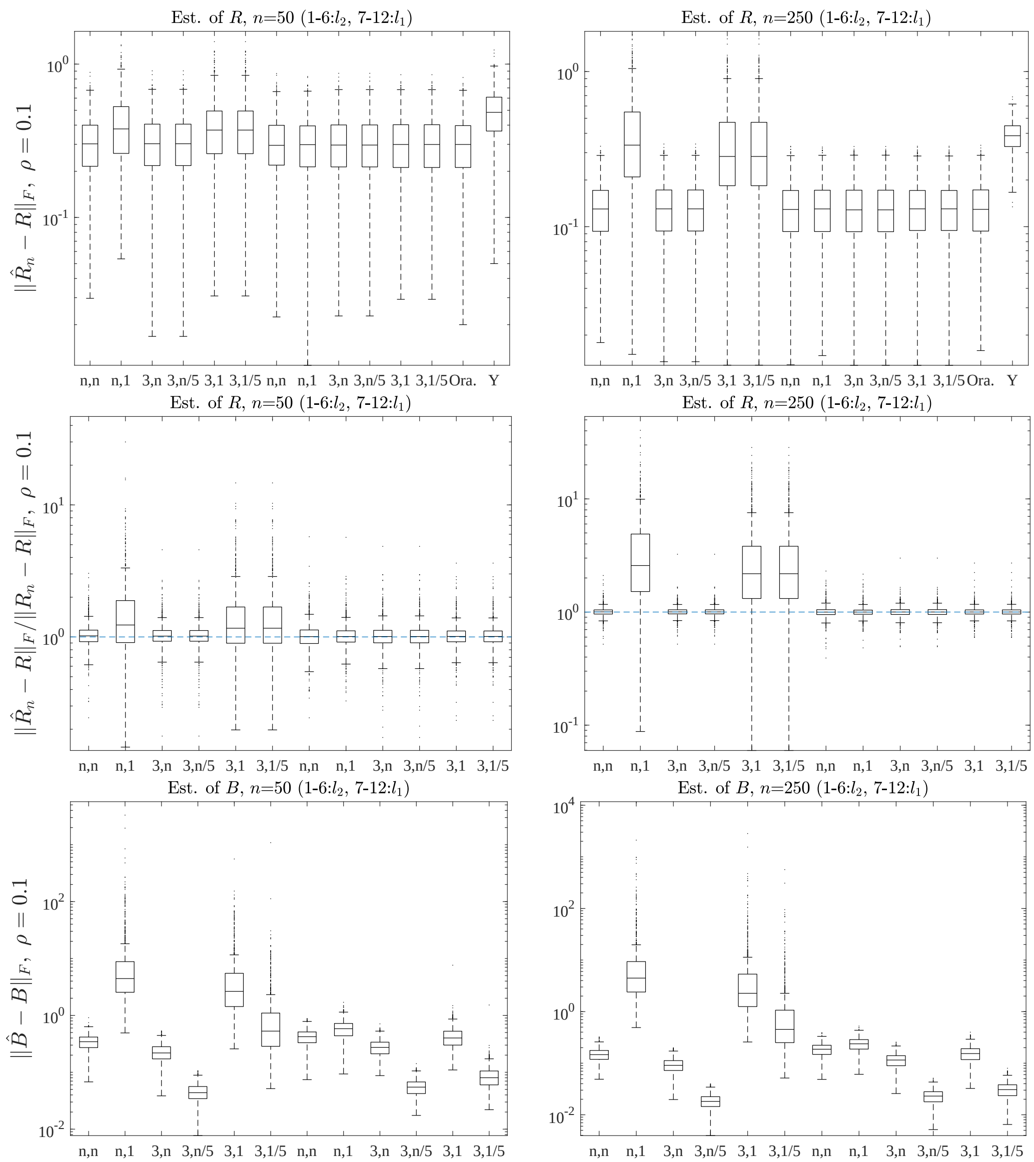

FIG 1. *Simulation results for the unrestricted model for* $(p, q) = (3, 2)$ *and* $\rho = 0.1$. *The meanings of the labels and of* ρ *are explained in the main text. All subfigures are produced with the y-axis in logarithm scale.*

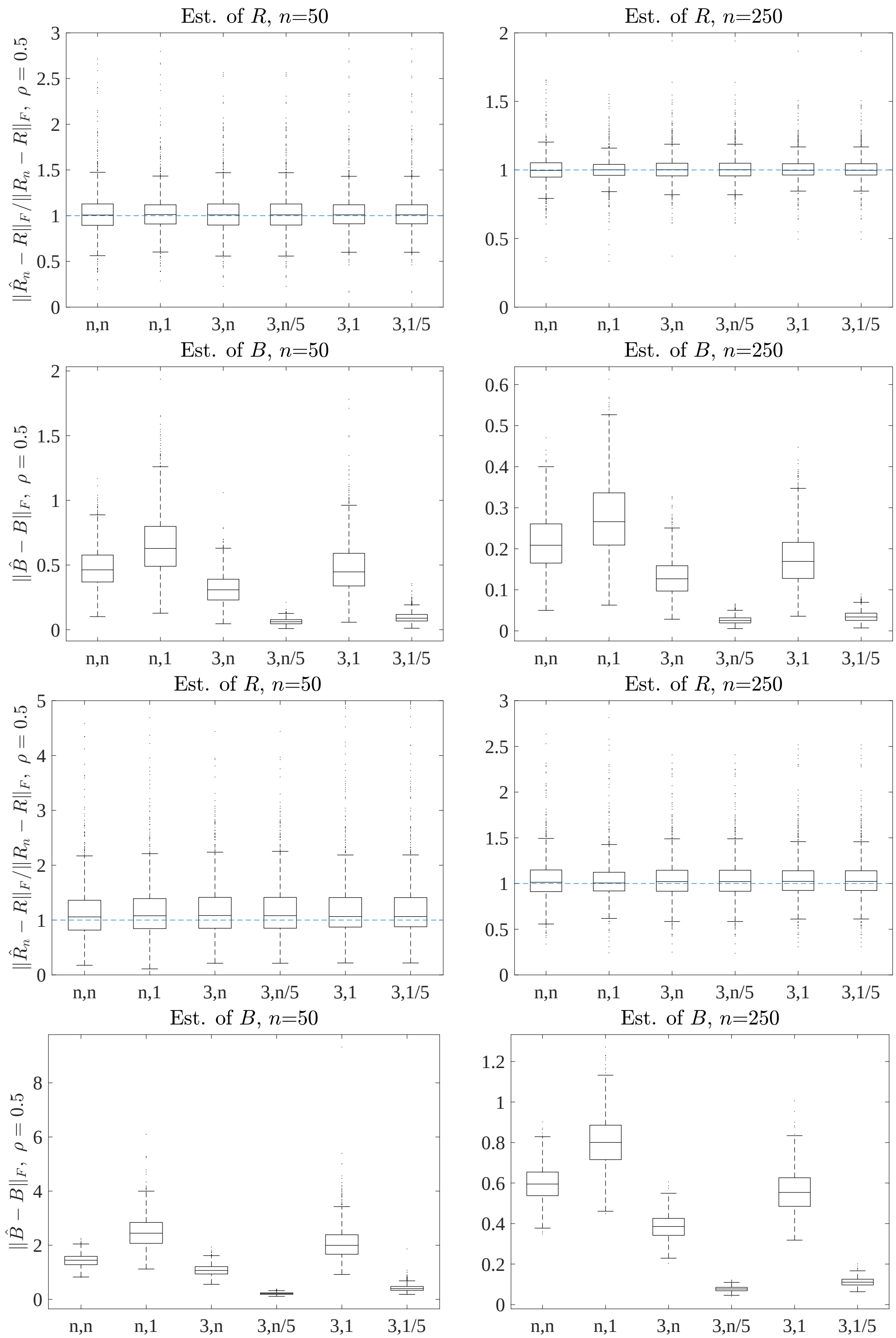

FIG 2. *Simulation results for the unrestricted model for* $(p, q) = (3, 2)$ *(two top rows) or* $(p, q) = (3, 10)$ *(bottom top rows) and the common value* $\rho = 0.5$. *The estimator* $\widehat{\mathbf{B}}$ *is produced by LAD. The meanings of the labels and of* ρ *are explained in the main text.*

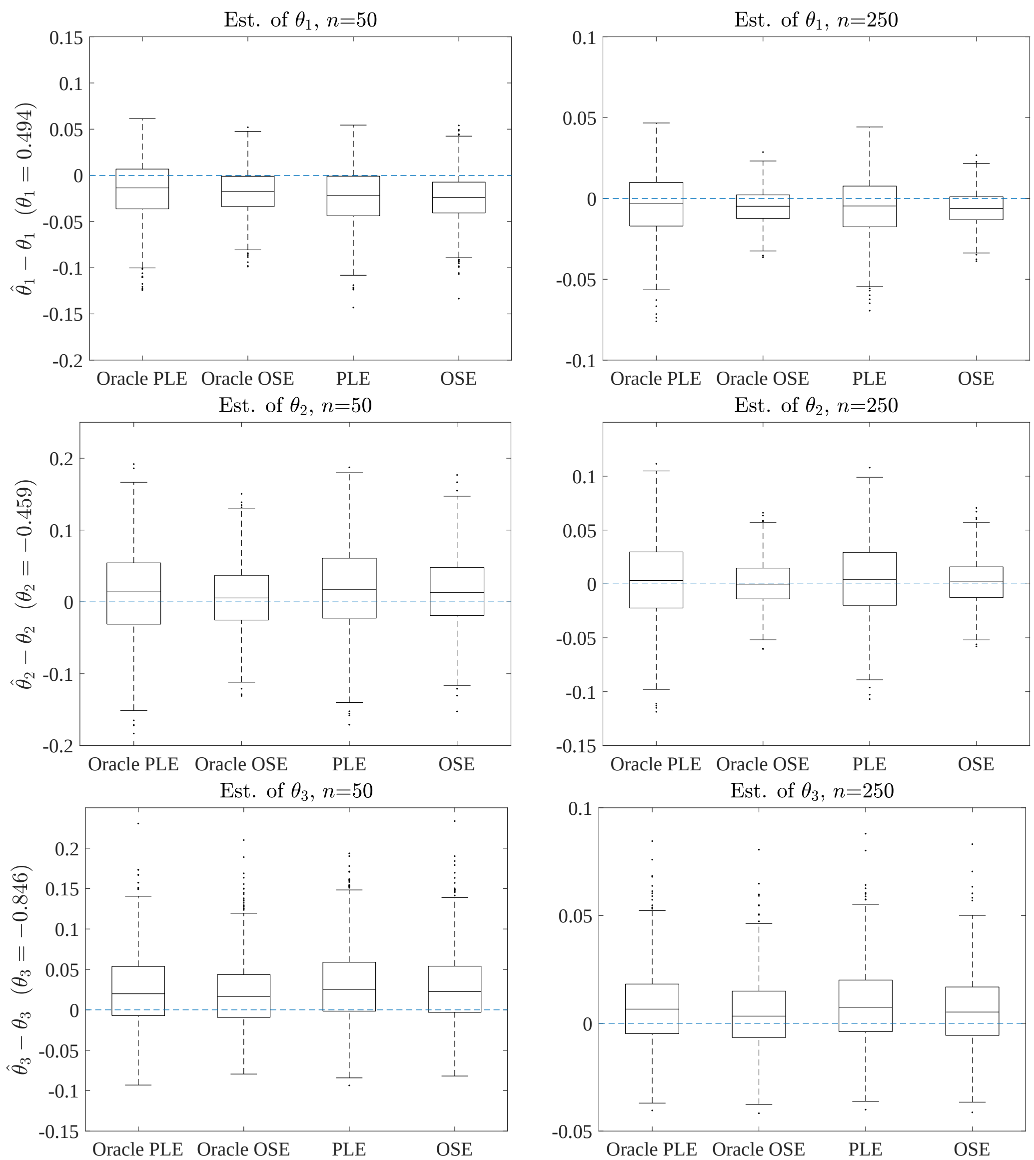

FIG 3. *Simulation results for the Toeplitz model under the parameters specified in the main text, for* $(p, q) = (4, 2)$ *and* $\rho = 0.5$.

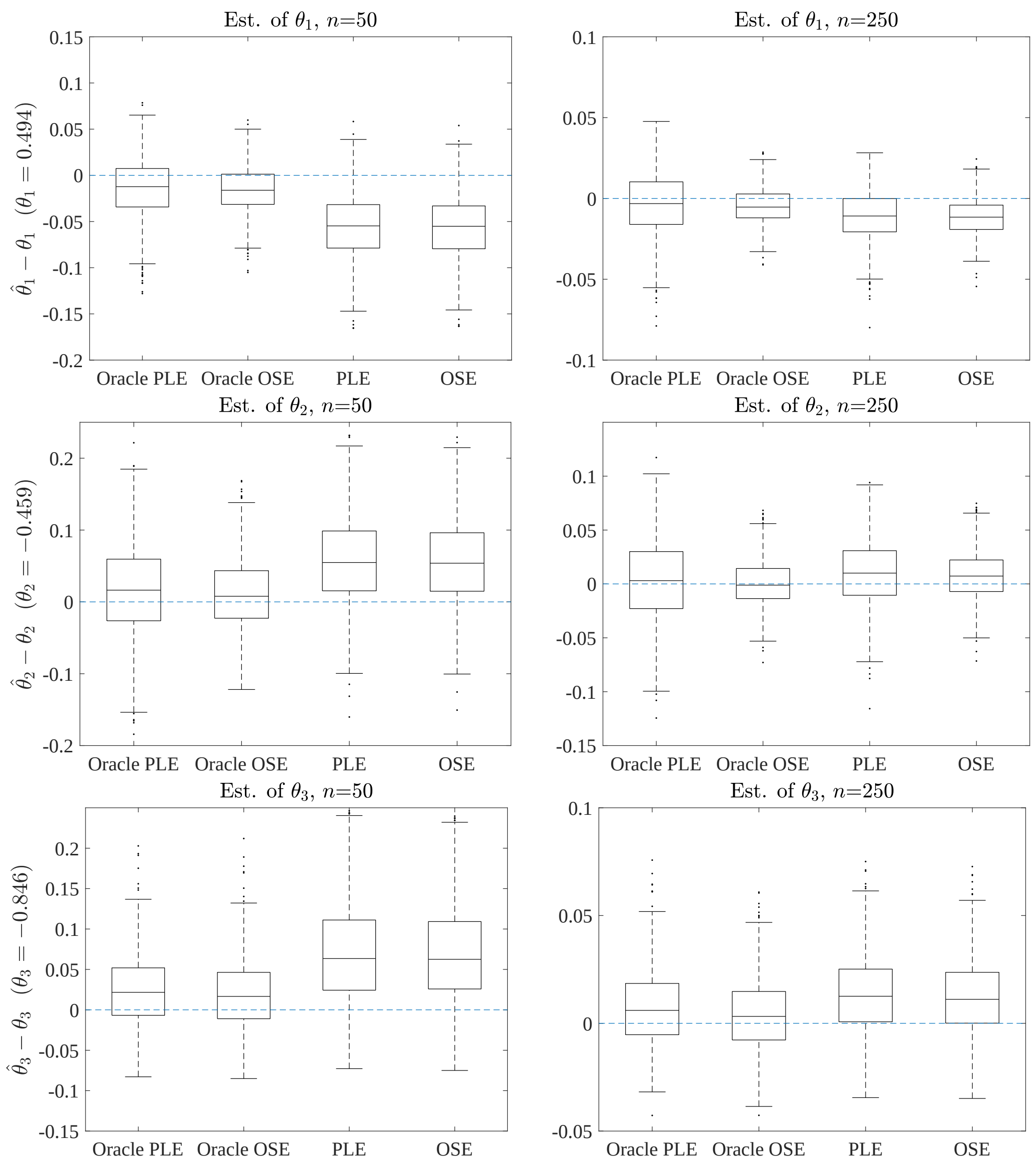

FIG 4. *Simulation results for the Toeplitz model under the exact same setting as Figure 3, except that now* $(p, q) = (4, 10)$.

6. Discussion on high-dimensional setting. In this paper we have studied the (semiparametrically efficient) estimation of the copula parameter in the covariate-adjusted Gaussian copula model. Our main conclusion is that asymptotically, and under mild conditions, we can find estimators based on the residual ranks (computed from an initial estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}$) that perform as well as their counterparts based on the oracle ranks. Our discussion so far has been confined to the fixed dimensional setting. In this section we briefly address the potential extension of our study to the high-dimensional setting, where the dimension p of the response and the dimension q of the covariate may no longer be fixed and may even be increasing with the sample size n.

We shall confine our discussion to the estimation of the copula correlation matrix $\mathbf{R}$ in the unrestricted model. In the high-dimensional setting, we need to strengthen the "in probability" statements in our analysis, such as those in Theorems 3.3 and 3.4, to precisely account for error accumulations over potentially increasing complexity of our model. For a generic estimator $\widetilde{\mathbf{R}}_n$ of $\mathbf{R}$ (in high dimensions), this is often achieved by first establishing a non-asymptotic deviation inequality of the form

$$\mathbb{P}\left\{|(\widetilde{\mathbf{R}}_n - \mathbf{R})_{kk'}| \geq f(n,\varepsilon)\right\} \leq \varepsilon, \quad \forall k,k' \in [p], \tag{6.1}$$

where the deviation $f(n,\varepsilon)$ is typically a decreasing function in both the sample size n and the exclusion probability ε. We would like (6.1) to be as tight as possible, so that we can find an exclusion probability ε small enough so that a simple union bound over all elements yields $p^2\varepsilon \ll 1$, yet also large enough so that the deviation $f(n,\varepsilon)$ is small. In words, it is desirable that with high probability the elements of $\widetilde{\mathbf{R}}_n$ are uniformly close to the corresponding elements in $\mathbf{R}$. (We can compare (6.1) to, e.g., Definition 1 in [34] on precision matrix estimation using the graphic LASSO. Compared to the latter, in (6.1) we consider the deviation $f(n,\varepsilon)$ as a function of the sample size n and the exclusion probability ε, which conforms better to the notation in (6.3) below.)

In the ordinary Gaussian copula model, (6.1) can usually be established without sparsity assumptions related to $\mathbf{R}$ or $\mathbf{S} = \mathbf{R}^{-1}$. As a concrete example, estimation of $\mathbf{R}$ through $\mathbf{R}_n^\rho$ via inversion of Spearman's rho has been well studied [31, 43]. For instance, Lemma 1 in [43] establishes (6.1) with $\widetilde{\mathbf{R}}_n$ replaced by $\mathbf{R}_n^\rho$ and with $f(n,\varepsilon)$ on the order of $\sqrt{\log(1/\varepsilon)/n}$; precisely, they derived, for some absolute constant c_0,

$$\mathbb{P}\left\{|(\mathbf{R}_n^\rho - \mathbf{R})_{kk'}| \geq \sqrt{\frac{1}{c_0}\frac{\log\left(2\varepsilon^{-1}\right)}{n}}\right\} \leq \varepsilon, \quad \forall k,k' \in [p]. \tag{6.2}$$

On the other hand, once (6.1) has been established, it can be combined with appropriate sparsity assumptions and correspondingly tailored sparsity-inducing procedures to produce accurate estimators of $\mathbf{R}$ or $\mathbf{S}$ in high dimensions. For instance, we may assume that $\mathbf{R}$ itself is sparse (that is, most of its off-diagonal elements are zero), and a sparsity-inducing estimator of $\mathbf{R}$ can be produced by thresholding $\widetilde{\mathbf{R}}_n$ at a level determined by $f(n,\varepsilon)$ [4]. Alternatively, as we have already mentioned in Section 1, in the context of high-dimensional graphical models, we may assume that $\mathbf{S} = \mathbf{R}^{-1}$ is sparse, and feed $\widetilde{\mathbf{R}}_n$ into procedures such as the graphic LASSO [34] to produce a sparse estimator $\widetilde{\mathbf{S}}_n$ of $\mathbf{S}$. The distance between $\widetilde{\mathbf{S}}_n$ and $\mathbf{S}$ is usually determined by $f(n,\varepsilon)$ as well.

In our regression setting, we need a deviation inequality analogous to (6.1) but with $\widetilde{\mathbf{R}}_n$ replaced by residual rank-based estimators such as $\widehat{\mathbf{R}}_n^\rho$ or $\widehat{\mathbf{R}}_n$. For simplicity we focus on the former case. Given (6.2), to obtain a deviation inequality for $\widehat{\mathbf{R}}_n^{\rho} - \mathbf{R}$, we simply need to complement (6.2) by a deviation inequality for $\widehat{\mathbf{R}}_n^{\rho} - \mathbf{R}_n^{\rho}$. This essentially calls for a refinement of our Theorem 3.3, which (by the decomposition of the residual ranks (3.11)) in turn requires establishing deviation inequalities for the remainder terms $r_{an,k}$, $a \in \{1, 2, 3, 4\}$, whose "in probability" bounds are stated in Propositions 3.1 and 3.2.

We will focus on the term $r_{1n,k}$, which in our opinion is the most difficult to deal with. In the high-dimensional setting, the dimension q of the covariate may be large. In this case, a complete treatment of $r_{1n,k}$ will likely require sparsity assumptions on $\mathbf{B}$ and a sparsity-inducing estimator of (the individual columns of) $\mathbf{B}$, to ensure the accuracy of the residual ranks whose calculation involves an estimator of (the individual columns of) $\widehat{\mathbf{B}}$. We will measure the sparsity of $\mathbf{B}$ by the maximum number s of nonzero elements in the individual columns of $\mathbf{B}$, and will simply assume that a good estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}$ is given. Specifically, we require that $\widehat{\mathbf{B}}$ satisfies the following properties:

- The events $A_{n,k} = \left\{\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| \le \delta_{\mathbf{B},n}\right\}$, $k \in [p]$, $n \ge 1$ all hold with high probability, for some sequence $\delta_{\mathbf{B},n}$, $n \ge 1$ that depends on the sample size and the sparsity of $\mathbf{B}$ and possibly mildly on the ambient dimensions p and q.
- There exists another set of events $B_{n,k}$, $k \in [p]$, $n \ge 1$, all holding with high probability, such that on the event $B_{n,k}$, $\operatorname{supp}(\widehat{\mathbf{B}}_k) \subset \operatorname{supp}(\mathbf{B}_k)$, and so in particular $\|\widehat{\mathbf{B}}_k\|_{\ell_0} \le s$. In the words of [13], the estimator $\widehat{\mathbf{B}}$ enjoys an oracle property in variable selection.

A condition similar to the first one above has already appeared in our earlier Assumption 3.1. Note that the dimension q of the covariate only explicitly enters our earlier analysis of $r_{1n,k}$ through the covering number bound in (A.6) in the supplement. Our second condition above ensures that, with high probability, the relevant covering number for our problem is controlled through the sparsity index s, rather than through the ambient dimension q; in particular, on the event $B_{n,k}$, we can rely on (A.6) in the supplement with q replaced by s. For a concrete example of how an estimator $\widehat{\mathbf{B}}$ satisfying the above conditions can work in conjunction with the residual empirical process theory in [9] (which as explained following Proposition 3.1 differs from ours) in high dimensions, we refer the readers to [8].

From the proof of Proposition 3.1, our bound (3.8) on $r_{1n,k}$ is essentially based on a non-asymptotic expectation bound. Thus, to establish a deviation inequality for $r_{1n,k}$, we only need to complement the expectation bound by a probability bound on the deviation from expectation. We will content ourselves here with deriving an unweighted version of the deviation inequality. We again assume that Assumptions 3.1, 3.2, 3.3 hold. Using the classical bound on the expectation of the suprema of an empirical process, which we have already seen in the proof of Proposition 3.1, we can establish the following expectation bound:

$$\mathbb{E}\left[\sup_{t\in\mathbb{R}} |r_{1n,k}(t)| \mathbb{1}_{A_{n,k}} \mathbb{1}_{B_{n,k}}\right] \lesssim \log^{1/2}\left(\delta_{\mathbf{B},n}^{-1}\right) \delta_{\mathbf{B},n}^{1/2} \sqrt{\frac{s}{n}} + \log\left(\delta_{\mathbf{B},n}^{-1}\right) \frac{s}{n} \equiv \bar{\mathbb{E}}_n.$$

Then, essentially complementing the above expectation bound by Talagrand's inequality for empirical processes (e.g., Theorem 3.3.9 in [21]), we arrive at

$$\mathbb{P}\left\{\sup_{t\in\mathbb{R}} |r_{1n,k}(t)| \gtrsim \bar{\mathbb{E}}_n + (\bar{\mathbb{E}}_n + \delta_{\mathbf{B},n})^{1/2} \sqrt{\frac{\log(\varepsilon^{-1})}{n}} + \frac{\log(\varepsilon^{-1})}{n}\right\}$$

$$\leq \varepsilon + \mathbb{P}(A_{n,k}^{\complement}) + \mathbb{P}(B_{n,k}^{\complement}). \tag{6.3}$$

We can note the similarity between (6.2) and (6.3). Under growth conditions on the sparsity level s and the ambient dimensions (both of which are typically required to establish a rate $\delta_{\mathbf{B},n}$ fast enough), the deviation of $r_{1n,k}$, $k \in [p]$ can be uniformly controlled at a level similar to or smaller than that of $\mathbf{R}_n^{\rho} - \mathbf{R}$ in (6.2). Analogous deviation inequalities can be established for all other remainder terms in the decomposition of the residual ranks (3.11). Therefore the residual rank-based $\widehat{\mathbf{R}}_n^{\rho}$ in the regression context can be regarded as good an estimator as the oracle rank-based estimator $\mathbf{R}_n^{\rho}$ in the ordinary Gaussian copula model in the high-dimensional setting.

SUPPLEMENT

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

REFERENCES

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

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

[3] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag New York, Inc., 1993.

[4] Bickel, P. J., and Levina, E. Covariance regularization by thresholding. *Ann. Statist. 36*, 6 (2008), 2577–2604.

[5] Cai, T. T., Li, H., Liu, W., and Xie, J. Covariate-adjusted precision matrix estimation with an application in genetical genomics. *Biometrika 100* (2013), 139–156.

[6] Cai, T. T., and Zhang, L. High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Stat. Sinica. abs/1512.02487* (2015). to appear.

[7] Chan, N.-H., Chen, J., Chen, X., Fan, Y., and Peng, L. Statistical inference for multivariate residual copula of garch models. *Stat. Sinica. 19* (2009), 53–70.

[8] Chatterjee, A., Gupta, S., and Lahiri, S. N. On the residual empirical process based on the ALASSO in high dimensions and its functional oracle property. *Journal of Econometrics 186*, 2 (2015), 317–324.

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

[10] Chen, M., Ren, Z., Zhao, H., and Zhou, H. Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model. *J. Amer. Statist. Assoc. 111*, 513 (2016), 394–406.

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

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

[13] Fan, J., and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties, 2001.

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

[15] Folland, G. B. *Real Analysis: Modern Techniques and Their Applications*, second ed. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs and Tracts. John Wiley & Sons Inc., 1999.

[16] Genest, C., Ghoudi, K., and Rivest, L.-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82* (1995), 543–552.

[17] Genest, C., Quessy, J.-F., and Rémillard, B. Testing the Gaussian copula hypothesis. Tech. rep.

[18] Genest, C., and Werker, B. J. M. Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In *Distributions with Given Marginals and Statistical Modelling*, C. M. Cuadras and J. A. R. Lallena, Eds. Kluwer Academic, Dordrecht, 2002, pp. 103–112.

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

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

[21] Giné, E., and Nickl, R. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge University Press,, 2016.

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

[23] Hoff, P. D., Niu, X., and Wellner, J. A. Information bounds for Gaussian copulas. *Bernoulli 20*, 2 (2014), 604–622.

[24] Hult, H., and Lindskog, F. Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. Appl. Probab. 34* (2002), 587–608.

[25] Kendall, M. G., and Gibbons, J. D. *Rank Correlation Methods*, 5th ed. Oxford University Press, U.S.A., 1990.

[26] Kim, J., and Pollard, D. Cube root asymptotics. *Ann. Statist. 18*, 1 (1990), 191–219.

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

[28] Kruskal, W. H. Ordinal measures of association. *J. Amer. Statist. Assoc. 53* (1958), 814–861.

[29] Lauritzen, S. L. *Graphical Models.* Clarendon Press, New York, 1996.

[30] Lee, S., and Wei, C.-Z. On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist. 27*, 1 (1999), 237–261.

[31] Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. High dimensional semiparametric Gaussian copula graphical models. *Ann. Statist. 40* (2012), 2293–2326.

[32] Masarotto, G., and Varin, C. Gaussian copula marginal regression. *Electron. J. Statist. 6* (2012), 1517–1549.

[33] Nelsen, R. B. *An Introduction to Copulas*, 2nd ed. Springer, New York, 2006.

[34] Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Statist. 5* (2011), 935–980.

[35] Rousseeuw, P. J. Least median of squares regression. *J. Amer. Statist. Assoc. 79* (1984), 871–880.

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

[37] Segers, J., van den Akker, R., and Werker, B. J. M. Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *Ann. Statist. 42*, 5 (2014), 1911–1940.

[38] Song, P. X.-K., Li, M., and Yuan, Y. Joint regression analysis of correlated data using Gaussian copulas. *Biometrics 65*, 1 (2009), 60–68.

[39] Susyanto, N., and Klaassen, C. A. J. Semiparametrically efficient estimation of constrained Euclidean parameters. *Electron. J. Statist. 11*, 2 (2017), 3120–3140.

[40] van der Vaart, A. *Semiparametric Statistics.* No. 1781 in Lecture Notes in Math. Springer, 2002, pp. 331–457. MR1915446.

[41] van der Vaart, A. W. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[42] van der Vaart, A. W., and Wellner, J. A. *Weak Convergence and Empirical Processes.* Springer, 1996.

[43] Xue, L., and Zou, H. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann. Statist. 40* (2012), 2541–2571.

I. Gijbels
Department of Mathematics
and Leuven Statistics
Research Center (LStat)
KU Leuven
Celestijnenlaan 200B
B-3001 Leuven (Heverlee)
Belgium
E-mail: irene.gijbels@wis.kuleuven.be

I. Van Keilegom
Y. Zhao
Research Centre for Operations Research
and Business Statistics (ORSTAT)
KU Leuven
Naamsestraat 69
3000 Leuven
Belgium
E-mail: ingrid.vankeilegom@kuleuven.be
yue.zhao@kuleuven.be

Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

SUPPLEMENT TO THE PAPER "INFERENCE FOR COVARIATE-ADJUSTED SEMIPARAMETRIC GAUSSIAN COPULA MODEL USING RESIDUAL RANKS"*

BY IRÈNE GIJBELS†, INGRID VAN KEILEGOM† AND YUE ZHAO†

KU Leuven†

A. Proofs for Section 3.

A.1. *Proofs for Section 3.1.*

PROPOSITION A.1. *Under the law* P, *further assume that Assumptions 3.1 and 3.2 hold. Then, for all* $k \in [p]$, (3.8) *holds.*

PROOF OF PROPOSITION A.1. Our proof is inspired by the proof of Lemma A.1 in [19] and Lemma 1 in [1] that treat residuals from a nonparametric regression problem, but here we rely on covering numbers instead of bracketing numbers, and also employ non-asymptotic moment bound for (weighted) empirical processes. Throughout the proof we fix arbitrary $k \in [p]$. We only prove the first half of (3.8) involving (3.1); the second half involving (3.2) then easily follows because

$$\sup_{t\in\mathbb{R}} |F^{\mathrm{r}}_{n,k}(t) - F_{n,k}(t)| \le 1/(n+1). \tag{A.1}$$

Let ε be an arbitrarily small but fixed constant, and recall the sequence $\delta_{\mathbf{B},n}$, $n \ge 1$ introduced in Assumption 3.1. Let $M = M(\varepsilon)$ be a constant large enough so that the events $A_{n,k} = \left\{\|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| \le M\delta_{\mathbf{B},n}\right\}$, $n \ge 1$ satisfy $\mathbb{P}(A_{n,k}) \ge 1 - \varepsilon$ for all n. We define the function $g_{t,\boldsymbol{\delta}}(\cdot_1, \cdot_2) : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ indexed by $(t, \boldsymbol{\delta}) \in \mathbb{R} \times \mathbb{R}^q$ as

$$g_{t,\boldsymbol{\delta}}(\cdot_1, \cdot_2) = \mathbb{1}\left\{\cdot_1 \le t + \boldsymbol{\delta}^\top \cdot_2\right\}, \tag{A.2}$$

and define the classes of functions $\mathcal{F}_n$, $n \ge 1$ as

$$\mathcal{F}_n = \{f_{t,\boldsymbol{\delta}}(\cdot_1, \cdot_2) \equiv g_{t,\boldsymbol{\delta}}(\cdot_1, \cdot_2) - g_{t,\mathbf{0}}(\cdot_1, \cdot_2) : t \in \mathbb{R}, \|\boldsymbol{\delta}\| \le M\delta_{\mathbf{B},n}\},$$

where M is the constant chosen as before.

Define the function $p_t : \mathbb{R}^q \to \mathbb{R}$ indexed by $t \in \mathbb{R}$ as

$$p_t(\mathbf{b}) = \mathbb{E}\left[\mathbb{1}\left\{E_k \le t + \mathbf{b}^\top \mathbf{X}\right\}\right]. \tag{A.3}$$

*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.

MSC 2010 subject classifications: Primary 62F12, 62G20; secondary 62G30, 62H20

Keywords and phrases: Gaussian copula, normal scores rank correlation coefficient, residual rank, Spearman's rho, semiparametric efficiency

Recall that $\widehat{\mathbf{B}}_k - \mathbf{B}_k$ is $\mathcal{X}_n$-measurable, while E_k and $\mathbf{X}$ are independent of $\mathcal{X}_n$. Then, by property 12(b) in Section 10.3 in [13], and (A.3),

$$\mathbb{E}\left[\mathbb{1}\left\{E_k \le t + (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}\right\} | \mathcal{X}_n\right] = p_t(\widehat{\mathbf{B}}_k - \mathbf{B}_k). \tag{A.4}$$

We let S denote the joint distribution of $(E_k, \mathbf{X})$, and let S_n be the corresponding empirical distribution constructed from $(E_{i,k}, \mathbf{X}_i)$, $i \in [n]$. We write $\mathbb{1}_{A_{n,k}}$ for the indicator function of the event $A_{n,k}$. Then, starting from (3.1) and (A.4) we have

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} |r_{1n,k}(t)| \mathbb{1}_{A_{n,k}} \\
&= \sup_{t\in\mathbb{R}} \left| \frac{1}{n} \sum_{i\in[n]} \left[\mathbb{1}\left\{E_{i,k} \le t + (\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i\right\} - \mathbb{1}\left\{E_{i,k} \le t\right\} - p_t(\widehat{\mathbf{B}}_k - \mathbf{B}_k) + F_k(t) \right] \right| \mathbb{1}_{A_{n,k}} \\
&\le \sup_{t\in\mathbb{R}, \|\boldsymbol{\delta}\| \le M\delta_{\mathbf{B},n}} \left| \frac{1}{n} \sum_{i\in[n]} \left[\mathbb{1}\left\{E_{i,k} \le t + \boldsymbol{\delta}^\top \mathbf{X}_i\right\} - \mathbb{1}\left\{E_{i,k} \le t\right\} - p_t(\boldsymbol{\delta}) + F_k(t) \right] \right| \\
&= \sup_{t\in\mathbb{R}, \|\boldsymbol{\delta}\| \le M\delta_{\mathbf{B},n}} \left| \frac{1}{n} \sum_{i\in[n]} \left[\mathbb{1}\left\{E_{i,k} \le t + \boldsymbol{\delta}^\top \mathbf{X}_i\right\} - \mathbb{1}\left\{E_{i,k} \le t\right\} - \mathbb{P}(E_k \le t + \boldsymbol{\delta}^\top \mathbf{X}) + F_k(t) \right] \right| \\
&= \sup_{f\in\mathcal{F}_n} |(\mathrm{S}_n - \mathrm{S}) f|.
\end{aligned} \tag{A.5}$$

The covering number $N(\mu, \mathcal{F}, \|\cdot\|_{\mathrm{g}})$ is the minimal number of balls of radius μ as measured by the (generic) norm $\|\cdot\|_{\mathrm{g}}$ needed to cover the set $\mathcal{F}$. We now bound the covering number of the classes $\mathcal{F}_n$. For $n \ge 1$, we define the classes

$$\mathcal{G}_n = \{g_{t,\boldsymbol{\delta}}(\cdot_1, \cdot_2) : t \in \mathbb{R}, \|\boldsymbol{\delta}\| \le M\delta_{\mathbf{B},n}\}, \quad \mathcal{H}_n = \{f - g : f, g \in \mathcal{G}_n\}.$$

Clearly, $\mathcal{F}_n \subset \mathcal{H}_n$. Thus, it suffices to bound the covering number of the classes $\mathcal{H}_n$, which can be done via bounding the covering number of the classes $\mathcal{G}_n$.

Let $\langle \cdot, \cdot \rangle$ denote the usual Euclidean inner product in $\mathbb{R}^{q+1}$. Then,

$$\begin{aligned}
\mathcal{G}_n \subset \{g_{t,\boldsymbol{\delta}}(\cdot_1, \cdot_2) : t \in \mathbb{R}, \boldsymbol{\delta} \in \mathbb{R}^q\} &= \left\{ \mathbb{1}\left\{\cdot_1 \le t + \boldsymbol{\delta}^\top \cdot_2\right\} : t \in \mathbb{R}, \boldsymbol{\delta} \in \mathbb{R}^q \right\} \\
&= \left\{ \mathbb{1}\left\{ \left\langle (1, \boldsymbol{\delta}^\top)^\top, (\cdot_1, -\cdot_2^\top)^\top \right\rangle \le t \right\} : t \in \mathbb{R}, \boldsymbol{\delta} \in \mathbb{R}^q \right\} \\
&\subset \left\{ \mathbb{1}\left\{ \left\langle \mathbf{u}, (\cdot_1, -\cdot_2^\top)^\top \right\rangle \le t \right\} : t \in \mathbb{R}, \mathbf{u} \in \mathbb{R}^{q+1} \right\}.
\end{aligned}$$

The last set above is simply the collection of all indicator functions of half planes in $\mathbb{R}^{q+1}$, and the collection of such half planes has VC dimension $q+3$ (see Theorem B in [7]). Then, from Theorem 2.6.4 in [18], we conclude that there exists an absolute constant $K > 0$ such that for a constant K_q depending only on K and q (we can take $K_q = K(q+3)(4e)^{q+3}$),

$$N(\mu, \mathcal{G}_n, L_2(Q)) \le K_q (1/\mu)^{2(q+2)}, \quad \forall Q \in \mathcal{B},\ 0 < \mu < 1.$$

Here $\mathcal{B}$ is the set of all Borel probability measures on $\mathbb{R}^{q+1}$. Then, by simple calculation,

$$N(\mu, \mathcal{F}_n, L_2(Q)) \le N(\mu, \mathcal{H}_n, L_2(Q)) \le N(\mu/2, \mathcal{G}_n, L_2(Q))^2$$

$$\leq K_q^2(2/\mu)^{4(q+2)}, \quad \forall Q \in \mathcal{B},\ 0 < \mu < 2. \tag{A.6}$$

Now, clearly the constant function $F = 1$ with $\|F\|_{L_2(\mathrm{S})} = 1$ is an envelope for all the classes $\mathcal{F}_n$, $n \geq 1$. In addition, for all $n \geq 1$ and all $f_{t,\boldsymbol{\delta}} \in \mathcal{F}_n$, we have

$$\begin{aligned}
\|f_{t,\boldsymbol{\delta}}\|^2_{L_2(\mathrm{S})} &= \mathbb{E}\left[\left|\mathbb{1}\left\{E_k \leq t + \boldsymbol{\delta}^\top \mathbf{X}\right\} - \mathbb{1}\left\{E_k \leq t\right\}\right|\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left|\mathbb{1}\left\{E_k \leq t + \boldsymbol{\delta}^\top \mathbf{X}\right\} - \mathbb{1}\left\{E_k \leq t\right\}\right| |\mathbf{X}\right]\right] \\
&\leq \mathbb{E}\left[f_k(t)|\boldsymbol{\delta}^\top \mathbf{X}| + L_k|\boldsymbol{\delta}^\top \mathbf{X}|^2\right] \\
&\leq f_k(t)M\mathbb{E}\|\mathbf{X}\|\delta_{\mathbf{B},n} + L_k M^2 \mathbb{E}[\|\mathbf{X}\|^2]\delta^2_{\mathbf{B},n} \equiv \sigma^2_{n,f_{t,\boldsymbol{\delta}}}.
\end{aligned} \tag{A.7}$$

Here the third step follows by the assumed Lipschitz continuity of f_k which implies that

$$|F_k(b) - F_k(a) - f_k(a)(b-a)| \leq L_k(b-a)^2, \quad \forall a, b \in \mathbb{R}. \tag{A.8}$$

To control $\sup_{f\in\mathcal{F}_n} |(\mathrm{S}_n - \mathrm{S})f|$, we partition $\mathcal{F}_n$ into two sets depending on the size of $\sigma_{n,f_{t,\boldsymbol{\delta}}}$. We first introduce some quantities from [8] which treats weighted empirical processes; this will make the distinction between large and small $\sigma_{n,f_{t,\boldsymbol{\delta}}}$ clear. For the function $g_q(r)$ introduced in the second equation display on p. 18 in [8], we choose $q = e$ and set $g_e(r) = e/r$. We let the sequence of constants r_n, $n \geq 1$ be defined as in Theorem 9 in [8], that is,

$$r_n = \sup\left\{r > 0 : r \leq \sqrt{\frac{\log(e/r) \vee \log\log(r^{-1})}{n}}\right\}.$$

Let the sequence of constants s_n, $n \geq 1$ be

$$s_n = \log^{1/2}(n)\, n^{-1/2}. \tag{A.9}$$

It is easy to calculate that $r_n = k_n s_n$ for s_n in (A.9) and another sequence of constants k_n, $n \geq 1$ satisfying $k_n = \sqrt{1/2} + o(1)$.

We first treat the case when $\sigma_{n,f_{t,\boldsymbol{\delta}}}$ is large enough, precisely, when $\sigma_{n,f_{t,\boldsymbol{\delta}}} > r_n$. Define the function $w : [0,1] \to \mathbb{R}$ as in Theorem 9 in [8]:

$$w(r) = r\sqrt{\log(e/r) \vee \log\log(r^{-1})}.$$

Note that $w(r)$ is increasing in r and $w(r)/r$ is decreasing in r on $[0,1]$. Then, we can follow the proof of Theorem 9 in [8] to verify the conditions of Theorem 4 in [8], and arrive at the last inequality in the proof of Theorem 4 in [8], which in our case reads

$$\mathbb{P}\left\{\sup_{f_{t,\boldsymbol{\delta}}\in\mathcal{F}_n:\sigma_{n,f_{t,\boldsymbol{\delta}}}>r_n} \frac{n^{1/2}|(\mathrm{S}_n - \mathrm{S})f_{t,\boldsymbol{\delta}}|}{w(\sigma_{n,f_{t,\boldsymbol{\delta}}})} \geq K_1 t + K_2\right\} \leq 2e^{-t}$$

for some absolute constants K_1 and K_2. As commented in the proof of Theorem 4 in [8], integrating the above tail bound yields that

$$\sup_{n\geq 1} \mathbb{E}\left[\sup_{f_{t,\boldsymbol{\delta}}\in\mathcal{F}_n:\sigma_{n,f_{t,\boldsymbol{\delta}}}>r_n} \frac{n^{1/2}|(\mathrm{S}_n - \mathrm{S})f_{t,\boldsymbol{\delta}}|}{w(\sigma_{n,f_{t,\boldsymbol{\delta}}})}\right] < \infty.$$

By the form of the function $w(r)$ and the fact that $\sup_{f_{t,\boldsymbol{\delta}}\in\mathcal{F}_n}\sigma_{n,f_{t,\boldsymbol{\delta}}}\to 0$, we can simplify the above to arrive at

$$\sup_{n\geq 1}\mathbb{E}\left[\sup_{f_{t,\boldsymbol{\delta}}\in\mathcal{F}_n:\sigma_{n,f_{t,\boldsymbol{\delta}}}>r_n}\frac{n^{1/2}|(\mathrm{S}_n-\mathrm{S})f_{t,\boldsymbol{\delta}}|}{\log^{1/2}(\delta_{\mathbf{B},n}^{-1})\sigma_{n,f_{t,\boldsymbol{\delta}}}}\right]<\infty. \tag{A.10}$$

Next we treat the case when $\sigma_{n,f_{t,\boldsymbol{\delta}}}\leq r_n$. By the classical bound on the expectation of the suprema of an empirical process via entropy integral (e.g., Theorem 3.5.4 in [9]),

$$\mathbb{E}\sup_{f_{t,\boldsymbol{\delta}}\in\mathcal{F}_n:\sigma_{n,f_{t,\boldsymbol{\delta}}}\leq r_n}|(\mathrm{S}_n-\mathrm{S})f_{t,\boldsymbol{\delta}}|\lesssim\log(n)\,n^{-1}. \tag{A.11}$$

Then Proposition A.1 follows from first combining (A.10) and (A.11), and then (A.5). □

LEMMA A.2. *Under the law* P, *further assume that Assumptions 3.1, 3.2 and 3.3 hold. Then for all* $k\in[p]$, (3.13) *holds.*

PROOF. We fix arbitrary $k\in[p]$, and let ε be an arbitrarily small but fixed constant. Then, let the constant M and the event $A_{n,k}$ be chosen exactly as in the proof of Proposition A.1, and again write $\mathbb{1}_{A_{n,k}}$ for the indicator function of the event $A_{n,k}$. In addition, define the sequence of random variables

$$T_{n,i}=f_k(E_{i,k})M\|\mathbf{X}_i\|\delta_{\mathbf{B},n}+L_kM^2\|\mathbf{X}_i\|^2\delta_{\mathbf{B},n}^2, \tag{A.12}$$

and the events $B_{n,k,i}=\{T_{n,i}\leq 1/2\}$ (the constant $1/2$ in the definition of $B_{n,k,i}$ can be replaced by any number on the open interval $(0,1)$), and write $\mathbb{1}_{B_{n,k,i}}$ for the indicator function of the event $B_{n,k,i}$.

Let $F_{n,k}^{\mathrm{U}}$ be the empirical distribution function of the i.i.d. Unif$(0,1)$ random variables $F_k(E_{i,k})$, $i\in[n]$; then, let the function $w_{n,k}$ be the modulus of continuity of the standard empirical process $\sqrt{n}(F_{n,k}^{\mathrm{U}}-I)$ (e.g., Eq. (1) in Section 14.2 in [15]). Recall that $F_k^{\leftarrow}$ defined as in (1.2) is the left-continuous inverse of F_k. Note that $\mathbb{P}(F_k^{\leftarrow}\circ F_k(E_{i,k})\neq E_{i,k})=1$ (e.g., Proposition 3 in Section 1.1 in [15]). Thus, with probability one,

$$\begin{aligned}F_{n,k}(\widehat{E}_{i,k})&=\frac{1}{n}\sum_{j\in[n]}\mathbb{1}\{E_{j,k}\leq\widehat{E}_{i,k}\}=\frac{1}{n}\sum_{j\in[n]}\mathbb{1}\{F_k^{\leftarrow}\circ F_k(E_{j,k})\leq\widehat{E}_{i,k}\}\\&=\frac{1}{n}\sum_{j\in[n]}\mathbb{1}\{F_k(E_{j,k})\leq F_k(\widehat{E}_{i,k})\}=F_{n,k}^{\mathrm{U}}(F_k(\widehat{E}_{i,k})),\end{aligned} \tag{A.13}$$

where the third step follows by, e.g., (21) in Section 1.1 in [15]. Similarly, with probability one, $F_{n,k}(E_{i,k})=F_{n,k}^{\mathrm{U}}(F_k(E_{i,k}))$.

Thus for $r_{3n,k,i}$, with probability one,

$$\begin{aligned}|r_{3n,k,i}|&=\left|F_{n,k}^{\mathrm{U}}(F_k(\widehat{E}_{i,k}))-F_k(\widehat{E}_{i,k})-F_{n,k}^{\mathrm{U}}(F_k(E_{i,k}))+F_k(E_{i,k})\right|\\&=\left|(F_{n,k}^{\mathrm{U}}-I)(F_k(\widehat{E}_{i,k}))-(F_{n,k}^{\mathrm{U}}-I)(F_k(E_{i,k}))\right|\\&\leq n^{-1/2}\,w_{n,k}(|F_k(\widehat{E}_{i,k})-F_k(E_{i,k})|)\\&\leq n^{-1/2}\,w_{n,k}(f_k(E_{i,k})|(\widehat{\mathbf{B}}_k-\mathbf{B}_k)^\top\mathbf{X}_i|+L_k|(\widehat{\mathbf{B}}_k-\mathbf{B}_k)^\top\mathbf{X}_i|^2),\end{aligned}$$

where in the last step we have invoked (A.8) and (2.5). Therefore, recalling $T_{n,i}$ from (A.12) and s_n from (A.9),

$$\begin{aligned}|r_{3n,k,i}|\mathbb{1}_{A_{n,k}}\mathbb{1}_{B_{n,k,i}} &\leq n^{-1/2}\, w_{n,k}(T_{n,i})\mathbb{1}_{B_{n,k,i}}\\ &\leq n^{-1/2}\, w_{n,k}((T_{n,i}\vee s_n^2)\wedge(1/2)).\end{aligned} \tag{A.14}$$

Note that we first take the maximum of $T_{n,i}$ and s_n^2, and then take the minimum of this and $1/2$ (as controlled by $\mathbb{1}_{B_{n,k,i}}$) in (A.14), so that the argument of the function $w_{n,k}$ is neither too small nor too large, and as a consequence we can always properly apply deviation bound for $w_{n,k}$ (e.g., Inequality 1 in Section 14.2 in [15], copied as Lemma C.5).

Now let $C_n = C\log^{1/2}(n)$, $n \geq 1$ be a sequence with yet unspecified constant $C > 0$, and let the function ψ below be as in Lemma C.5. First by (A.14) when conditioning on $\mathbf{X}_i$, and then by Lemma C.5 with $a = \delta = (T_{n,i}\vee s_n^2)\wedge(1/2)$, we obtain

$$\begin{aligned}&\mathbb{P}\left(n^{1/2}|r_{3n,k,i}|\mathbb{1}_{A_{n,k}}\mathbb{1}_{B_{n,k,i}} \geq C_n\sqrt{(T_{n,i}\vee s_n^2)\wedge(1/2)}\,|\,\mathbf{X}_i\right)\\ &\leq \mathbb{P}\left(w_{n,k}((T_{n,i}\vee s_n^2)\wedge(1/2)) \geq C_n\sqrt{(T_{n,i}\vee s_n^2)\wedge(1/2)}\,|\,\mathbf{X}_i\right)\\ &\leq \frac{20}{\{(T_{n,i}\vee s_n^2)\wedge(1/2)\}^4}\exp\Bigg(-[1-\{(T_{n,i}\vee s_n^2)\wedge(1/2)\}]^4\\ &\qquad\times\frac{C^2\log(n)}{2}\psi\left(\frac{C\log^{1/2}(n)}{\sqrt{n\{(T_{n,i}\vee s_n^2)\wedge(1/2)\}}}\right)\Bigg)\\ &\leq \frac{20}{s_n^8}\exp\left(-\left(\frac{1}{2}\right)^5 C^2\log(n)\,\psi\left(\frac{C\log^{1/2}(n)}{\sqrt{n\{(T_{n,i}\vee s_n^2)\wedge(1/2)\}}}\right)\right),\end{aligned} \tag{A.15}$$

where the last step follows by

$$s_n^2 \leq (T_{n,i}\vee s_n^2)\wedge(1/2) \leq 1/2, \quad \forall n\geq 1, i\in[n]. \tag{A.16}$$

By the first half of (A.16), the argument to the function ψ in (A.15) is bounded above by C, and thus by, e.g., Proposition 1(10) in Section 11.1 in [15], the function ψ itself is always bounded below by $1/(1+C/3)$ (all uniformly over $n \geq 1$ and $i \in [n]$). Then, if the positive constant C is large enough, the rightmost term in (A.15) is bound above by $C'n^{-(1+c)}$ for some constants $c, C' > 0$ independent of $n \geq 1$ and $i \in [n]$. We fix this sequence $C_n = C\log^{1/2}(n)$. Then, first using the union bound, then conditioning on $(\mathbf{X}_1,\ldots,\mathbf{X}_n)$, and finally invoking the bound $C'n^{-(1+c)}$ on the rightmost term in (A.15) for the fixed sequence C_n, we obtain

$$\begin{aligned}&\mathbb{P}\left(\max_{i\in[n]}\frac{|r_{3n,k,i}|\mathbb{1}_{A_{n,k}}\mathbb{1}_{B_{n,k,i}}}{\log^{1/2}(n)n^{-1/2}\sqrt{T_{n,i}\vee s_n^2}} \geq C\right)\\ &\leq \sum_{i\in[n]}\mathbb{P}\left(\frac{|r_{3n,k,i}|\mathbb{1}_{A_{n,k}}\mathbb{1}_{B_{n,k,i}}}{\log^{1/2}(n)n^{-1/2}\sqrt{T_{n,i}\vee s_n^2}} \geq C\right)\\ &= \sum_{i\in[n]}\mathbb{E}\left[\mathbb{P}\left(n^{1/2}|r_{3n,k,i}|\mathbb{1}_{A_{n,k}}\mathbb{1}_{B_{n,k,i}} \geq C_n\sqrt{T_{n,i}\vee s_n^2}\,|\,\mathbf{X}_i\right)\right]\end{aligned}$$

$$\leq \sum_{i\in[n]} \mathbb{E}\left[\mathbb{P}\left(n^{1/2}|r_{3n,k,i}|\mathbb{1}_{A_{n,k}}\mathbb{1}_{B_{n,k,i}} \geq C_n\sqrt{(T_{n,i}\vee s_n^2)\wedge(1/2)}\,|\,\mathbf{X}_i\right)\right]$$

$$= \mathcal{O}(n^{-c}) = o(1). \tag{A.17}$$

We rely on the following elementary lemma to lower bound the probability of the event $\cap_{i\in[n]}B_{n,k,i}$.

LEMMA A.3. *Under Assumption 3.3, for any fixed positive constant L,*

$$\mathbb{P}\left(\max_{i\in[n]}\|\mathbf{X}_i\|\delta_{\mathbf{B},n} > L\right) \to 0. \tag{A.18}$$

PROOF. First suppose that $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$. Then for some large enough constant K, $\delta_{\mathbf{B},n} \leq Kn^{-1/2}$ for all $n \geq 1$. This and the union bound gives

$$\mathbb{P}\left(\max_{i\in[n]}\|\mathbf{X}_i\|\delta_{\mathbf{B},n} > L\right) \leq n\mathbb{P}(\|\mathbf{X}\|\delta_{\mathbf{B},n} > L) \leq n\mathbb{P}(\|\mathbf{X}\| > (L/K)n^{1/2}) \to 0,$$

where the last convergence step holds because $\|\mathbf{X}\|$ has bounded second moment by Assumption 3.3; see., e.g., p. 161 in [13]. Thus we have shown (A.18).

Next, suppose instead it is not true that $\delta_{\mathbf{B},n} = \mathcal{O}(n^{-1/2})$, then (A.18) holds by the second half of Assumption 3.3. □

Now, by Lemma A.3, we have

$$\mathbb{P}(\cup_{i\in[n]}B^{\complement}_{n,k,i}) = \mathbb{P}\left(\max_{i\in[n]}T_{n,i} > 1/2\right)$$
$$\leq \mathbb{P}\left(\max_{i\in[n]}M\|f_k\|_{L_\infty}\|\mathbf{X}_i\|\delta_{\mathbf{B},n} > 1/4\right) + \mathbb{P}\left(\max_{i\in[n]}L_kM^2\|\mathbf{X}_i\|^2\delta^2_{\mathbf{B},n} > 1/4\right) \to 0.$$

Therefore, $\mathbb{P}(\cap_{i\in[n]}B_{n,k,i}) \to 1$ as $n\to\infty$. This together with the bound $\mathbb{P}(A_{n,k}) \geq 1-\varepsilon$ and (A.17) yield (3.13). □

LEMMA A.4. *Under the law P, further assume that Assumptions 3.1, 3.2 and 3.3 hold. Then for all $k\in[p]$, (3.9) and (3.14) hold.*

PROOF. We first prove (3.9). Conditioning on $\mathbf{X}$, it is straightforward to show from (A.3) that

$$p_t(\mathbf{b}) = F_k(t) + f_k(t)\mathbf{b}^\top\mathbb{E}[\mathbf{X}] + d(t,\mathbf{b}) \tag{A.19}$$

where we have further introduced

$$d(t,\mathbf{b}) = \mathbb{E}c(t,\mathbf{b},\mathbf{X}), \quad c(t,\mathbf{b},\mathbf{x}) = F_k(t+\mathbf{b}^\top\mathbf{x}) - F_k(t) - f_k(t)\mathbf{b}^\top\mathbf{x}.$$

Using Taylor expansion and (A.8),

$$|c(t,\mathbf{b},\mathbf{x})| \leq L_k(\mathbf{b}^\top\mathbf{x})^2,$$

and thus

$$\sup_{t\in\mathbb{R}} |d(t,\mathbf{b})| \le L_k \mathbf{b}^\top \mathbb{E}\left[\mathbf{X}\mathbf{X}^\top\right]\mathbf{b}. \tag{A.20}$$

From (3.3), using (A.4) and (A.19) for $t \in \mathbb{R}$, we have

$$\begin{aligned} r_{2n,k}(t) &= p_t(\widehat{\mathbf{B}}_k - \mathbf{B}_k) - F_k(t) - f_k(t)(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}] \\ &= d(t, \widehat{\mathbf{B}}_k - \mathbf{B}_k). \end{aligned}$$

Then (3.9) easily follows from the above equality, (A.20) and the assumptions.

Next we prove (3.14). From (3.5), by (A.8) and the Lipschitz continuity of f_k,

$$\begin{aligned} |r_{4n,k,i}| &\le |F_k(\widehat{E}_{i,k}) - F_k(E_{i,k}) + f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i| \\ &\quad + |(f_k(\widehat{E}_{i,k}) - f_k(E_{i,k}))(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]| \\ &\le L_k((\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i)^2 + L_k|(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i||(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbb{E}[\mathbf{X}]|. \end{aligned}$$

Then (3.14) follows from the above and the imposed assumptions. □

A.2. *Proof of Theorem 3.3.*

PROOF. First, when $k = k'$, $\widehat{\rho}_{n,kk'} = \rho_{n,kk'} = 1$ and so (3.17) clearly holds. Thus we focus on the case $k \ne k'$. We have the decomposition

$$\begin{aligned} &\sqrt{n}\left(\widehat{\rho}_{n,kk'} - \rho_{n,kk'}\right) \\ &= -6\frac{n^{3/2}}{n^2-1} 2\sum_{i\in[n]} \left[\{\widehat{F}_{n,k}(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k})\}\{F_{n,k}(E_{i,k}) - F_{n,k'}(E_{i,k'})\}\right] \\ &\quad + 6\frac{n^{3/2}}{n^2-1} 2\sum_{i\in[n]} \left[\{\widehat{F}_{n,k'}(\widehat{E}_{i,k'}) - F_{n,k'}(E_{i,k'})\}\{F_{n,k}(E_{i,k}) - F_{n,k'}(E_{i,k'})\}\right] \\ &\quad - 6\frac{n^{3/2}}{n^2-1} \sum_{i\in[n]} \{\widehat{F}_{n,k}(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k}) - \widehat{F}_{n,k'}(\widehat{E}_{i,k'}) + F_{n,k'}(E_{i,k'})\}^2 \\ &\equiv 6\frac{n^{3/2}}{n^2-1}(-2\,A + 2\,B - C). \end{aligned} \tag{A.21}$$

We first bound the term A. Invoking (3.11) in Proposition 3.2, we rewrite A as

$$\begin{aligned} A &= \sum_{i\in[n]} \{-f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top(\mathbf{X}_i - \mathbb{E}[\mathbf{X}])\}\{F_k(E_{i,k}) - F_{k'}(E_{i,k'})\} \\ &\quad + \sum_{i\in[n]} \{-f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top(\mathbf{X}_i - \mathbb{E}[\mathbf{X}])\} \\ &\qquad \times \{F_{n,k}(E_{i,k}) - F_k(E_{i,k}) - F_{n,k'}(E_{i,k'}) + F_{k'}(E_{i,k'})\} \\ &\quad + \sum_{i\in[n]} \{r_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i}\}\{F_{n,k}(E_{i,k}) - F_{n,k'}(E_{i,k'})\} \\ &\equiv A_1 + A_2 + A_3. \end{aligned} \tag{A.22}$$

Now we analyze the term A_1 in (A.22). We further rewrite A_1 as

$$\begin{aligned} A_1 &= -(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \sum_{i\in[n]} (\mathbf{X}_i - \mathbb{E}[\mathbf{X}])\, f_k(E_{i,k})\{F_k(E_{i,k}) - F_{k'}(E_{i,k'})\} \\ &\equiv -(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{A}_{11}. \end{aligned} \tag{A.23}$$

Note that the random variable $\mathbf{X}_i - \mathbb{E}[\mathbf{X}]$ has expectation zero, is independent of $E_{i,k}$ and $E_{i,k'}$, and each of its components has finite variance (by Assumption 3.3). Therefore the i.i.d. summands in the term $\mathbf{A}_{11}$ have expectation zero. Now, the variance of the ℓth coordinate of each individual summand in $\mathbf{A}_{11}$ is

$$\mathbb{E}\left[(X_\ell - \mathbb{E}[X_\ell])^2\, f_k^2(E_k)\,(F_k(E_k) - F_{k'}(E_{k'}))^2\right] \le \|f_k\|_{L_\infty}^2 \mathbb{E}[\|\mathbf{X}\|^2]. \tag{A.24}$$

Therefore we conclude that $\|\mathbf{A}_{11}\| = \mathcal{O}_{\mathrm{p}}(n^{1/2})$ and hence, from (A.23),

$$|A_1| \le \|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| \|\mathbf{A}_{11}\| = \mathcal{O}_{\mathrm{p}}(n^{1/2}\delta_{\mathbf{B},n}). \tag{A.25}$$

Next we consider the term A_2 in (A.22). By the Dvoretzky-Kiefer-Wolfowitz inequality (e.g., [12]),

$$\begin{aligned} &\max_{i\in[n]} |F_{n,k}(E_{i,k}) - F_k(E_{i,k}) - F_{n,k'}(E_{i,k'}) + F_{k'}(E_{i,k'})| \\ &\le \|F_{n,k} - F_k\|_{L_\infty} + \|F_{n,k'} - F_{k'}\|_{L_\infty} = \mathcal{O}_{\mathrm{p}}(n^{-1/2}). \end{aligned}$$

Then, together with Assumption 3.3 which implies that $\sum_{i\in[n]} \|\mathbf{X}_i\| = \mathcal{O}_{\mathrm{p}}(n)$,

$$|A_2| = \mathcal{O}_{\mathrm{p}}\left(n^{-1/2} \sum_{i\in[n]} \|f_k\|_\infty \|\widehat{\mathbf{B}}_k - \mathbf{B}_k\| \left(\|\mathbf{X}_i\| + \|\mathbb{E}[\mathbf{X}]\|\right)\right) = \mathcal{O}_{\mathrm{p}}\left(n^{1/2}\delta_{\mathbf{B},n}\right). \tag{A.26}$$

Next we consider the term A_3 in (A.22). By (3.8) (or the simplified version (3.10)) and (3.9) in Proposition 3.1, (3.13) and (3.14) in Proposition 3.2, Assumption 3.1 which implies

$$\log(\delta_{\mathbf{B},n}^{-1}) \lesssim \log(n), \tag{A.27}$$

and Assumption 3.3 which implies that $\sum_{i\in[n]} \|\mathbf{X}_i\|^{1/2}$, $\sum_{i\in[n]} \|\mathbf{X}_i\|$ and $\sum_{i\in[n]} \|\mathbf{X}_i\|^2$ are all $\mathcal{O}_{\mathrm{p}}(n)$, we obtain

$$\sum_{i\in[n]} \left| r_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i} \right| = \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\, n^{1/2}\delta_{\mathbf{B},n}^{1/2} + n\delta_{\mathbf{B},n}^2\right).$$

Therefore,

$$A_3 = \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\, n^{1/2}\delta_{\mathbf{B},n}^{1/2} + n\delta_{\mathbf{B},n}^2\right). \tag{A.28}$$

From (A.22), (A.25), (A.26) and (A.28), we conclude that $A = \mathcal{O}_{\mathrm{p}}(\log^{1/2}(n)\, n^{1/2}\delta_{\mathbf{B},n}^{1/2} + n\delta_{\mathbf{B},n}^2)$. Similarly, B satisfies the same bound. We now deal with the term C. We have

$$|C| \le 2\left[\sum_{i\in[n]} \{\widehat{F}_{n,k}(\widehat{E}_{i,k}) - F_{n,k}(E_{i,k})\}^2\right] + 2\left[\sum_{i\in[n]} \{\widehat{F}_{n,k'}(\widehat{E}_{i,k'}) - F_{n,k'}(E_{i,k'})\}^2\right] \equiv 2\,C_1 + 2\,C_2.$$

Using (3.11) again, (3.8) and (3.9) in Proposition 3.1, (3.13) and (3.14) in Proposition 3.2, and (A.27), we have

$$
\begin{aligned}
C_1 &\le 2 \sum_{i\in[n]} f_k(E_{i,k})^2 \|\widehat{\mathbf{B}}_k - \mathbf{B}_k\|^2 \|\mathbf{X}_i\|^2 \\
&\quad + 8 \sum_{i\in[n]} \left\{ r_{1n,k}(\widehat{E}_{i,k})^2 + r_{2n,k}(\widehat{E}_{i,k})^2 + r_{3n,k,i}^2 + r_{4n,k,i}^2 \right\} \\
&= \mathcal{O}_{\mathrm{p}}\left(n\delta_{\mathbf{B},n}^2\right) + \mathcal{O}_{\mathrm{p}}\Bigg(\log(n)\delta_{\mathbf{B},n} + \log^2(n)\, n^{-1} + n\delta_{\mathbf{B},n}^4 + \log(n)\, n^{-1}\delta_{\mathbf{B},n} \\
&\quad \times \sum_{i\in[n]} (\|\mathbf{X}_i\| + \delta_{\mathbf{B},n}\|\mathbf{X}_i\|^2) + \delta_{\mathbf{B},n}^4 \sum_{i\in[n]} \|\mathbf{X}_i\|^2 (\|\mathbf{X}_i\| + \|\mathbb{E}[\mathbf{X}]\|)^2 \Bigg).
\end{aligned}
\tag{A.29}
$$

By Lemma A.3, $\max_{i\in[n]}(\delta_{\mathbf{B},n}^2\|\mathbf{X}_i\|^2) = o_{\mathrm{p}}(1)$, and thus for the last term in (A.29) we have

$$
\begin{aligned}
\delta_{\mathbf{B},n}^4 \sum_{i\in[n]} \|\mathbf{X}_i\|^2 (\|\mathbf{X}_i\| + \|\mathbb{E}[\mathbf{X}]\|)^2 &\le \left(\max_{i\in[n]} \delta_{\mathbf{B},n}^2 \|\mathbf{X}_i\|^2 \right) \delta_{\mathbf{B},n}^2 \sum_{i\in[n]} (\|\mathbf{X}_i\| + \|\mathbb{E}[\mathbf{X}]\|)^2 \\
&= o_{\mathrm{p}}\left(n\delta_{\mathbf{B},n}^2\right).
\end{aligned}
$$

Plugging this into (A.29), and further simplifying gives

$$
C_1 = \mathcal{O}_{\mathrm{p}}\left(n\delta_{\mathbf{B},n}^2\right).
$$

Similarly, C_2 satisfies the same bound.

Collecting terms, and recalling the overall scaling by $n^{-1/2}$ of the terms A, B and C in (A.21), we conclude that (3.17) holds. □

A.3. *Proof of Theorem 3.4.*

Proof. First, when $k = k'$, $\widehat{r}_{n,kk'} = r_{n,kk'} = 1$ and so (3.27) clearly holds. Thus we focus on the case $k \ne k'$. We have the decomposition

$$
\begin{aligned}
\sqrt{n}(\widehat{r}_{n,kk'} - r_{n,kk'}) &= \frac{1}{\sqrt{n}} \sum_{i\in[n]} \left\{ \Phi^{\leftarrow}(\widehat{F}_{n,k}^{\mathrm{r}}(\widehat{E}_{i,k})) - \Phi^{\leftarrow}(F_{n,k}^{\mathrm{r}}(E_{i,k})) \right\} \Phi^{\leftarrow}(F_{n,k'}^{\mathrm{r}}(E_{i,k'})) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i\in[n]} \Phi^{\leftarrow}(F_{n,k}^{\mathrm{r}}(E_{i,k})) \left\{ \Phi^{\leftarrow}(\widehat{F}_{n,k'}^{\mathrm{r}}(\widehat{E}_{i,k'})) - \Phi^{\leftarrow}(F_{n,k'}^{\mathrm{r}}(E_{i,k'})) \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i\in[n]} \left\{ \Phi^{\leftarrow}(\widehat{F}_{n,k}^{\mathrm{r}}(\widehat{E}_{i,k})) - \Phi^{\leftarrow}(F_{n,k}^{\mathrm{r}}(E_{i,k})) \right\} \\
&\quad \times \left\{ \Phi^{\leftarrow}(\widehat{F}_{n,k'}^{\mathrm{r}}(\widehat{E}_{i,k'})) - \Phi^{\leftarrow}(F_{n,k'}^{\mathrm{r}}(E_{i,k'})) \right\} + \mathcal{O}(\log(n)\, n^{-1/2}),
\end{aligned}
\tag{A.30}
$$

where the last $\mathcal{O}(\log(n)\, n^{-1/2})$ term comes from the factor ϕ_n given in (2.2).

We first show that the first term on the right hand side of (A.30) satisfies

$$\frac{1}{\sqrt{n}} \sum_{i \in [n]} \left\{ \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})) - \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k})) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) = \mathcal{O}_{\mathrm{p}}(\Delta_n). \tag{A.31}$$

We divide the summation on the left hand side of (A.31) into two cases, corresponding to whether $F_k(E_{i,k}) \in A_1$ or $F_k(E_{i,k}) \in A_2$, for A_1, A_2 introduced in (3.23).

We first consider the case $F_k(E_{i,k}) \in A_1$. By the Chernoff bound (e.g., (6) in [11], with $m = n(2n^{-\gamma})$) and the assumption that $\gamma < 1$ (see (3.21)), the number of indices $i \in [n]$ satisfying $F_k(E_{i,k}) \in A_1$ is upper bounded by $2n^{1-\gamma}(1+o_{\mathrm{p}}(1))$. Using first this observation and then Lemma C.1 give

$$\begin{aligned}
&\left| \frac{1}{\sqrt{n}} \sum_{i \in [n]: F_k(E_{i,k}) \in A_1} \left\{ \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})) - \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k})) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \right| \\
&= \mathcal{O}_{\mathrm{p}} \left(n^{1/2-\gamma} \max_{i \in [n]} \left[\left\{ |\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))| + |\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))| \right\} |\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))| \right] \right) \\
&= \mathcal{O}_{\mathrm{p}}(\log(n)\, n^{1/2-\gamma}).
\end{aligned} \tag{A.32}$$

Next we consider the case $F_k(E_{i,k}) \in A_2$. For brevity sometimes we set $\mathbf{X}_i^{\mathrm{c}} = \mathbf{X}_i - \mathbb{E}[\mathbf{X}]$. First, adding $(F_{n,k} - F^{\mathrm{r}}_{n,k})(E_{i,k})$ to both sides of (3.12) in Proposition 3.2 yields

$$\begin{aligned}
\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k}) &= (F_{n,k} - F^{\mathrm{r}}_{n,k})(E_{i,k}) - f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i^{\mathrm{c}} \\
&\quad + r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i}.
\end{aligned} \tag{A.33}$$

In what follows, we abbreviate the summation $\sum_{i \in [n]: F_k(E_{i,k}) \in A_2}$ by $\sum_{F_k(E_{i,k}) \in A_2}$. By using first-order Taylor expansion twice, Lemma C.2, and invoking (A.33), the summation on the left hand side of (A.31) over $F_k(E_{i,k}) \in A_2$ becomes

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \left\{ \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})) - \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k})) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \\
&= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{1}{\phi(\Phi^{\leftarrow}(\overline{F}^{\mathrm{r}}_{n,k,i}))} \left\{ \widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k}) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \\
&= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \left\{ \widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k}) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \left\{ \frac{1}{\phi(\Phi^{\leftarrow}(\overline{F}^{\mathrm{r}}_{n,k,i}))} - \frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \right\} \left\{ \widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k}) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \\
&= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \left\{ -f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i^{\mathrm{c}} \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \left\{ (F_{n,k} - F^{\mathrm{r}}_{n,k})(E_{i,k}) + r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i} \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \left\{ \frac{1}{\phi(\Phi^{\leftarrow}(\overline{F}^{\mathrm{r}}_{n,k,i}))} - \frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \right\} \left\{ \widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k}) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \left\{ f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i^{\mathrm{c}} \right\} \Phi^{\leftarrow}(F_{k'}(E_{i,k'})) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \left\{ f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i^{\mathrm{c}} \right\} \left\{ \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) - \Phi^{\leftarrow}(F_{k'}(E_{i,k'})) \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \left\{ (F_{n,k} - F^{\mathrm{r}}_{n,k})(E_{i,k}) + r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i} \right\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \left\{ \frac{\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i})}{\phi^2(\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i}))} \right\} \left\{ \widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k}) \right\} \left\{ \overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k}) \right\} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \\
&\equiv A + B + C + D.
\end{aligned}
\tag{A.34}
$$

In the above, for each i, the quantity $\overline{F}^{\mathrm{r}}_{n,k,i}$ comes from the first Taylor expansion and is a random number on the open interval between $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$ and $F^{\mathrm{r}}_{n,k}(E_{i,k})$, and the quantity $\widetilde{F}^{\mathrm{r}}_{n,k,i}$ comes from the second Taylor expansion and is a random number on the open interval between $\overline{F}^{\mathrm{r}}_{n,k,i}$ and $F_k(E_{i,k})$. The terms A, B, C, D in the last line of (A.34) are all $\mathcal{O}_{\mathrm{p}}(\Delta_n)$ — which when combined with the earlier summation over $F_k(E_{i,k}) \in A_1$ in (A.32) yields that (A.31) holds — as we now show in sequence.

The term A

We rewrite A as $A = -\left\{ n^{-1/2}(\widehat{\mathbf{B}}_k - \mathbf{B}_k) \right\}^\top \mathbf{A}$ where

$$
\mathbf{A} = \sum_{i\in[n]} \mathbb{1}\{F_k(E_{i,k}) \in A_2\} \left\{ \frac{f_k(E_{i,k})\, \Phi^{\leftarrow}(F_{k'}(E_{i,k'}))}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))} \right\} \mathbf{X}_i^{\mathrm{c}}. \tag{A.35}
$$

By Assumption 3.1, the term $n^{-1/2}(\widehat{\mathbf{B}}_k - \mathbf{B}_k)$ in A is $\mathcal{O}_{\mathrm{p}}(n^{-1/2}\delta_{\mathbf{B},n})$. Next we treat $\mathbf{A}$.

In $\mathbf{A}$, the random variable $\mathbf{X}_i^{\mathrm{c}}$ has expectation zero, and is independent of $E_{i,k}$ and $E_{i,k'}$. Therefore the i.i.d. summands in $\mathbf{A}$ have expectation zero. Next, the variance of the ℓth coordinate of each individual summand in $\mathbf{A}$ is

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\} \left\{ \frac{f_k(E_k)\, \Phi^{\leftarrow}(F_{k'}(E_{k'}))}{\phi(\Phi^{\leftarrow}(F_k(E_k)))} (X_\ell - \mathbb{E}[X_\ell]) \right\}^2 \right] \\
&= \mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\} \left\{ \frac{f_k(E_k)\, \Phi^{\leftarrow}(F_{k'}(E_{k'}))}{\phi(\Phi^{\leftarrow}(F_k(E_k)))} \right\}^2 \right] \mathbb{E}\left[(X_\ell - \mathbb{E}[X_\ell])^2 \right] \\
&\lesssim \mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\} \left\{ \frac{f_k(E_k)}{\phi(\Phi^{\leftarrow}(F_k(E_k)))} \right\}^2 \mathbb{E}\left[\left\{ \Phi^{\leftarrow}(F_{k'}(E_{k'})) \right\}^2 | E_k \right] \right].
\end{aligned}
\tag{A.36}
$$

Here the first step follows by the independence of $\mathbf{X}$ and $\mathbf{E}$, and the second step follows because the second moment of $\|\mathbf{X}\|$ is bounded. Now, the pair $(\Phi^{\leftarrow}(F_k(E_k)), \Phi^{\leftarrow}(F_{k'}(E_{k'})))$ is bivariate Gaussian with marginal variances equal to one and with correlation $r_{kk'}$, and so conditional on E_k, $\Phi^{\leftarrow}(F_{k'}(E_{k'}))$ is a normal random variable with mean $r_{kk'}\,\Phi^{\leftarrow}(F_k(E_k))$ and variance $1-r_{kk'}^2$. Thus

$$\mathbb{E}\left[\{\Phi^{\leftarrow}(F_{k'}(E_{k'}))\}^2\,|E_k\right] = r_{kk'}^2\Phi^{\leftarrow}(F_k(E_k))^2 + (1-r_{kk'}^2).$$

Plugging the above into (A.36), then invoking (C.2) in Lemma C.3, and finally invoking (3.24) and the fact that $F_k(E_k) \sim \mathrm{Unif}(0,1)$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\left\{\frac{f_k(E_k)\,\Phi^{\leftarrow}(F_{k'}(E_{k'}))}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}(X_\ell-\mathbb{E}[X_\ell])\right\}^2\right]\\
&\lesssim \mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\left\{\frac{f_k(E_k)}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right\}^2\left\{r_{kk'}^2\Phi^{\leftarrow}(F_k(E_k))^2+(1-r_{kk'}^2)\right\}\right]\\
&\lesssim \mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\left\{\frac{f_k(E_k)}{(F_k\wedge(1-F_k))(E_k))}\right\}^2\{|\Phi^{\leftarrow}(F_k(E_k))|\vee 1\}^2\right]=\mathcal{O}(\delta_n).
\end{aligned}$$

So the variance of each coordinate of $\mathbf{A}$ in (A.35) is $\mathcal{O}(n\,\delta_n)$, implying that

$$A = \mathcal{O}_{\mathrm{p}}(\delta_{\mathbf{B},n}\delta_n^{1/2}). \tag{A.37}$$

This concludes our treatment of the term A.

The term B

As for the term A, first rewrite B as $B = -\left\{n^{-1/2}(\widehat{\mathbf{B}}_k-\mathbf{B}_k)\right\}^\top \mathbf{B}$ where

$$\mathbf{B} = \sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\{f_k(E_{i,k})\mathbf{X}_i^{\mathrm{c}}\}\left\{\Phi^{\leftarrow}(F_{n,k'}^{\mathrm{r}}(E_{i,k'}))-\Phi^{\leftarrow}(F_{k'}(E_{i,k'}))\right\}.$$

Again as for the term A, it suffices to treat $\mathbf{B}$. Using Hölder's inequality, the magnitude of the ℓth coordinate of $\mathbf{B}$ is bounded by

$$\begin{aligned}
&\left[\sum_{i\in[n]}\mathbb{1}\{F_k(E_{i,k})\in A_2\}\left\{\frac{f_k(E_{i,k})(X_{i,\ell}-\mathbb{E}[X_{i,\ell}])}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\right\}^2\right]^{1/2}\\
&\quad\times\left[\sum_{i\in[n]}\left\{\Phi^{\leftarrow}(F_{n,k'}^{\mathrm{r}}(E_{i,k'}))-\Phi^{\leftarrow}(F_{k'}(E_{i,k'}))\right\}^2\right]^{1/2}.
\end{aligned}$$

By Theorem 1 in [6], the sum in the second square bracket on the right hand side of the last line above is $o_{\mathrm{p}}(\sqrt{n})$. For the sum in the first square bracket,

$$\begin{aligned}
&\sum_{i\in[n]}\mathbb{1}\{F_k(E_{i,k})\in A_2\}\left\{\frac{f_k(E_{i,k})(X_{i,\ell}-\mathbb{E}[X_{i,\ell}])}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\right\}^2\\
&=\mathcal{O}_{\mathrm{p}}\left(n\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\left\{\frac{f_k(E_k)(X_\ell-\mathbb{E}[X_\ell])}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right\}^2\right]\right)
\end{aligned}$$

$$= \mathcal{O}_{\mathrm{p}}\left(n\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\left\{\frac{f_k(E_k)}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right\}^2\right]\mathbb{E}\left[(X_\ell-\mathbb{E}[X_\ell])^2\right]\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(n\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{f_k^2(E_k)}{(F_k\wedge(1-F_k))^2(E_k)}\right]\right) = \mathcal{O}_{\mathrm{p}}\left(n\,\delta_n\right).$$

Here the second step follows by the independence of $\mathbf{X}$ and $\mathbf{E}$, the third step follows by (C.2) in Lemma C.3 and Assumption 3.3, and the last step follows by (3.24). Collecting terms, we have

$$B = o_{\mathrm{p}}(n^{1/4}\delta_{\mathbf{B},n}\delta_n^{1/2}). \tag{A.38}$$

This concludes our treatment of the term B.

The term C

We first deal with the terms involving $(F_{n,k}-F^{\mathrm{r}}_{n,k})(E_{i,k})$. Because $1/\phi(\Phi^{\leftarrow}(\cdot))$ is the derivative of $\Phi^{\leftarrow}(\cdot)$ (see Lemma C.2),

$$\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right] = \int_{A_2}\frac{1}{\phi(\Phi^{\leftarrow}(u))}\mathrm{d}u$$
$$= \Phi^{\leftarrow}(1-n^{-\gamma})-\Phi^{\leftarrow}(n^{-\gamma}) \lesssim \log^{1/2}(n), \tag{A.39}$$

where the last step follows by (C.3) in Lemma C.3. Then, first by (A.1), Lemma C.1 and then by (A.39),

$$\left|\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\Big\{(F_{n,k}-F^{\mathrm{r}}_{n,k})(E_{i,k})\Big\}\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|$$
$$\lesssim \log^{1/2}(n)n^{-3/2}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)n^{-1/2}\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right]\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log(n)n^{-1/2}\right).$$

Next we deal with the terms involving $r^{\mathrm{r}}_{1n,k}$ and the $r_{3n,k,i}$'s. The latter quantities are bounded as in (3.8) in Proposition 3.1 and (3.13) in Proposition 3.2 respectively; we also recall (A.27). Then, simply replacing $f_k(t)$ by $\|f_k\|_{L_\infty}<\infty$ in (3.8) and (3.13), we have

$$\left|\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\Big\{r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k})+r_{3n,k,i}\Big\}\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)n^{-1/2}\sum_{F_k(E_{i,k})\in A_2}\frac{(1+\|\mathbf{X}_i\|^{1/2}+\|\mathbf{X}_i\|\delta^{1/2}_{\mathbf{B},n})\log^{1/2}(n)\,n^{-1/2}\delta^{1/2}_{\mathbf{B},n}}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log(n)n^{-1}\delta^{1/2}_{\mathbf{B},n}\sum_{F_k(E_{i,k})\in A_2}\frac{1+\|\mathbf{X}_i\|^{1/2}+\|\mathbf{X}_i\|\delta^{1/2}_{\mathbf{B},n}}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\right)$$

$$= \mathcal{O}_{\mathrm{p}}\left(\log(n)\delta_{\mathbf{B},n}^{1/2}\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{1+\|\mathbf{X}\|^{1/2}+\|\mathbf{X}\|\delta_{\mathbf{B},n}^{1/2}}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right]\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log(n)\delta_{\mathbf{B},n}^{1/2}\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right]\mathbb{E}\left[1+\|\mathbf{X}\|^{1/2}+\|\mathbf{X}\|\delta_{\mathbf{B},n}^{1/2}\right]\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log^{3/2}(n)\delta_{\mathbf{B},n}^{1/2}\right).$$

Next we deal with the terms involving $r_{2n,k}$ and the $r_{4n,k,i}$'s. The later quantities are bounded as in (3.9) in Proposition 3.1 and (3.14) in Proposition 3.2 respectively. Then,

$$\left|\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\left\{r_{2n,k}(\widehat{E}_{i,k})+r_{4n,k,i}\right\}\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)n^{-1/2}\sum_{F_k(E_{i,k})\in A_2}\frac{\delta_{\mathbf{B},n}^2+\|\mathbf{X}_i\|(\|\mathbf{X}_i\|\vee\|\mathbb{E}[\mathbf{X}]\|)\delta_{\mathbf{B},n}^2}{\phi(\Phi^{\leftarrow}(F_k(E_{i,k})))}\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)n^{1/2}\delta_{\mathbf{B},n}^2\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{1+\|\mathbf{X}_i\|(\|\mathbf{X}_i\|\vee\|\mathbb{E}[\mathbf{X}]\|)}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right]\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)n^{1/2}\delta_{\mathbf{B},n}^2\,\mathbb{E}\left[\mathbb{1}\{F_k(E_k)\in A_2\}\frac{1}{\phi(\Phi^{\leftarrow}(F_k(E_k)))}\right]\mathbb{E}\left[1+\|\mathbf{X}\|(\|\mathbf{X}\|\vee\|\mathbb{E}[\mathbf{X}]\|)\right]\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\log(n)n^{1/2}\delta_{\mathbf{B},n}^2\right).$$

Collecting terms, we have

$$C = \mathcal{O}_{\mathrm{p}}\left(\log^{3/2}(n)\delta_{\mathbf{B},n}^{1/2}+\log(n)n^{1/2}\delta_{\mathbf{B},n}^2\right). \tag{A.40}$$

This concludes our treatment of the term C.

The term D

Recall that the term D in the last line of (A.34) is

$$\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\left\{\frac{\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i})}{\phi^2(\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i}))}\right\}\left\{\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})-F^{\mathrm{r}}_{n,k}(E_{i,k})\right\}\left\{\overline{F}^{\mathrm{r}}_{n,k,i}-F_k(E_{i,k})\right\}\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})).$$

We first show that

$$\max_{i\in[n]:F_k(E_{i,k})\in A_2}\frac{|\widetilde{F}^{\mathrm{r}}_{n,k,i}-F_k(E_{i,k})|}{(F_k\wedge(1-F_k))(E_{i,k})}\le\max_{i\in[n]:F_k(E_{i,k})\in A_2}\frac{|\overline{F}^{\mathrm{r}}_{n,k,i}-F_k(E_{i,k})|}{(F_k\wedge(1-F_k))(E_{i,k})}=o_{\mathrm{p}}(1). \tag{A.41}$$

Recall that, for each i, the quantity $\overline{F}^{\mathrm{r}}_{n,k,i}$ is a random number on the open interval between $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$ and $F^{\mathrm{r}}_{n,k}(E_{i,k})$, and hence

$$|\overline{F}^{\mathrm{r}}_{n,k,i}-F^{\mathrm{r}}_{n,k}(E_{i,k})|\le|\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})-F^{\mathrm{r}}_{n,k}(E_{i,k})|; \tag{A.42}$$

recall also that the quantity $\widetilde{F}^{\mathrm{r}}_{n,k,i}$ is a random number on the open interval between $\overline{F}^{\mathrm{r}}_{n,k,i}$ and $F_k(E_{i,k})$, and hence

$$|\widetilde{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})| \le |\overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})|. \tag{A.43}$$

Then, the first half of (A.41) follows from (A.43). Next, applying the triangle inequality twice, and using (A.42) and (A.33) give

$$\begin{aligned}
|\overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})| &\le |\overline{F}^{\mathrm{r}}_{n,k,i} - F^{\mathrm{r}}_{n,k}(E_{i,k})| + |F^{\mathrm{r}}_{n,k}(E_{i,k}) - F_k(E_{i,k})| \\
&\le |\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k})| + |F^{\mathrm{r}}_{n,k}(E_{i,k}) - F_{n,k}(E_{i,k})| + |F_{n,k}(E_{i,k}) - F_k(E_{i,k})| \\
&\le 2|(F^{\mathrm{r}}_{n,k} - F_{n,k})(E_{i,k})| + |(F_{n,k} - F_k)(E_{i,k})| \\
&\quad + |-f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}^{\mathrm{c}}_i + r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{2n,k}(\widehat{E}_{i,k}) + r_{3n,k,i} + r_{4n,k,i}|.
\end{aligned} \tag{A.44}$$

We treat the terms in the last line of (A.44) one by one. First, by (A.1) and the fact that $\gamma < 1$ (see (3.21)),

$$\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{|(F^{\mathrm{r}}_{n,k} - F_{n,k})(E_{i,k})|}{(F_k \wedge (1-F_k))(E_{i,k})} = o_{\mathrm{p}}(1). \tag{A.45}$$

Next, by the second half of Lemma C.4 involving (C.4) on the weighted empirical processes,

$$\sup_{t:F_k(t)\in(1/n,1-1/n)} \frac{|(F_{n,k} - F_k)(t)|}{t_n\sqrt{(F_k \wedge (1-F_k))(t)}} = \mathcal{O}_{\mathrm{p}}(1) \tag{A.46}$$

for $t_n = \sqrt{\dfrac{\log\log(n)}{n}}$. Then, consecutively by (A.46) and $\gamma < 1$,

$$\begin{aligned}
&\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{|(F_{n,k} - F_k)(E_{i,k})|}{(F_k \wedge (1-F_k))(E_{i,k})} \\
&= \mathcal{O}_{\mathrm{p}}\left(\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{t_n}{\sqrt{(F_k \wedge (1-F_k))(E_{i,k})}}\right) = o_{\mathrm{p}}(1).
\end{aligned} \tag{A.47}$$

Next, consecutively by Assumptions 3.4 and 3.5, we have

$$\begin{aligned}
&\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{|f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}^{\mathrm{c}}_i|}{(F_k \wedge (1-F_k))(E_{i,k})} \\
&= \mathcal{O}_{\mathrm{p}}\left(\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{|f_k(F_k^{\leftarrow}(F_k(E_{i,k})))|}{(F_k \wedge (1-F_k))(E_{i,k})} n^{-\tau}\right) = o_{\mathrm{p}}(1).
\end{aligned} \tag{A.48}$$

Next, consecutively by (3.8) in Proposition 3.1, (3.13) in Proposition 3.2, (A.27) and s_n from (A.9), then the Lipschitz continuity of f_k implying

$$f_k^{1/2}(\widehat{E}_{i,k}) \le f_k^{1/2}(E_{i,k}) + L_k^{1/2}|(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}_i|^{1/2}, \tag{A.49}$$

then Assumption 3.4 (implying (3.20)), and finally the second inequality in (3.21) (implying $\gamma < 1/2 + \tau$), we have

$$\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{|r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k}) + r_{3n,k,i}|}{(F_k \wedge (1-F_k))(E_{i,k})}$$

$$= \mathcal{O}_{\mathrm{p}}\left(\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{s_n\left\{f_k^{1/2}(\widehat{E}_{i,k})\delta_{\mathbf{B},n}^{1/2} + \delta_{\mathbf{B},n}\right\} + s_n\left\{f_k^{1/2}(E_{i,k})\|\mathbf{X}_i\|^{1/2}\delta_{\mathbf{B},n}^{1/2} + \|\mathbf{X}_i\|\delta_{\mathbf{B},n}\right\} + s_n^2}{(F_k\wedge(1-F_k))(E_{i,k})}\right)$$

$$= \mathcal{O}_{\mathrm{p}}\left(\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{s_n f_k^{1/2}(E_{i,k})\delta_{\mathbf{B},n}^{1/2}(1+\|\mathbf{X}_i\|^{1/2}) + s_n\delta_{\mathbf{B},n}(1+\|\mathbf{X}_i\|) + L_k^{1/2}s_n\delta_{\mathbf{B},n}\|\mathbf{X}_i\|^{1/2} + s_n^2}{(F_k\wedge(1-F_k))(E_{i,k})}\right)$$

(A.50)

$$= \mathcal{O}_{\mathrm{p}}(s_n n^{\gamma/2} + s_n n^{-\tau+\gamma} + s_n^2 n^{\gamma}) = o_{\mathrm{p}}(1).$$

Finally, consecutively by (3.9) in Proposition 3.1 and (3.14) in Proposition 3.2, Assumption 3.4 (implying (3.20)), and finally the first half of the second inequality in (3.21),

$$\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{|r_{2n,k}(\widehat{E}_{i,k}) + r_{4n,k,i}|}{(F_k\wedge(1-F_k))(E_{i,k})} = \mathcal{O}_{\mathrm{p}}\left(n^{\gamma}\delta_{\mathbf{B},n}^2 \max_{i\in[n]:F_k(E_{i,k})\in A_2}\{1+\|\mathbf{X}_i\|(\|\mathbf{X}_i\|+\|\mathbb{E}[\mathbf{X}]\|)\}\right)$$

(A.51)

$$= \mathcal{O}_{\mathrm{p}}(n^{-2\tau+\gamma}) = o_{\mathrm{p}}(1).$$

Therefore, collecting the results from (A.44), (A.45), (A.47), (A.48), (A.50) and (A.51), we conclude that the second half of (A.41) also holds. Now, using first (C.2) in Lemma C.3, and then (A.41), we have

$$\begin{aligned}|D| &\le \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\left|\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i})\right|}{\phi^2(\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i}))} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right| \\ &\quad \times \left|\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k})\right| \left|\overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right| \\ &\lesssim \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\left|\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i})\right|}{\left\{\widetilde{F}^{\mathrm{r}}_{n,k,i}\wedge(1-\widetilde{F}^{\mathrm{r}}_{n,k,i})\right\}^2} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right| \\ &\quad \times \left|\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k})\right| \left|\overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right| \\ &= \mathcal{O}_{\mathrm{p}}\Bigg(\frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\left|\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i})\right|}{(F_k\wedge(1-F_k))^2(E_{i,k})} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right| \\ &\quad \times \left|\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k})\right| \left|\overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right| \Bigg). \qquad \text{(A.52)}\end{aligned}$$

Next, for the term $\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i})$ in (A.52), first by the mean value theorem,

$$\max_{i\in[n]:F_k(E_{i,k})\in A_2}\left|\Phi^{\leftarrow}(\widetilde{F}^{\mathrm{r}}_{n,k,i}) - \Phi^{\leftarrow}(F_k(E_{i,k}))\right| = \max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{\left|\widetilde{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right|}{\phi(\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k,i}))}$$

$$\lesssim \max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{\left|\widetilde{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right|}{\widehat{F}^{\mathrm{r}}_{n,k,i}\wedge(1-\widehat{F}^{\mathrm{r}}_{n,k,i})} \lesssim \mathcal{O}_{\mathrm{p}}\left(\max_{i\in[n]:F_k(E_{i,k})\in A_2} \frac{\left|\widetilde{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right|}{(F_k\wedge(1-F_k))(E_{i,k})}\right) = o_{\mathrm{p}}(1).$$

Here, for each i with $F_k(E_{i,k}) \in A_2$, $\widehat{F}^{\mathrm{r}}_{n,k,i}$ is a random number on the open interval between $\widetilde{F}^{\mathrm{r}}_{n,k,i}$ and $F_k(E_{i,k})$; hence (A.41) holds with the replacement of $\widetilde{F}^{\mathrm{r}}_{n,k,i}$ by $\widehat{F}^{\mathrm{r}}_{n,k,i}$, which further implies the second to last step above. (The second step follows by (C.2) in Lemma C.3 and the last step follows by (A.41).) Applying this result to (A.52), and using (A.44) to bound both $\left|\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}) - F^{\mathrm{r}}_{n,k}(E_{i,k})\right|$ (see (A.33)) and $\left|\overline{F}^{\mathrm{r}}_{n,k,i} - F_k(E_{i,k})\right|$, we conclude that to bound D, it suffices to bound

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|}{(F_k \wedge (1-F_k))^2(E_{i,k})} \Big[\left\{(F^{\mathrm{r}}_{n,k} - F_{n,k})(E_{i,k})\right\}^2 \\
&\quad + \{(F_{n,k} - F_k)(E_{i,k})\}^2 + \left\{f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}^{\mathrm{c}}_i\right\}^2 \\
&\quad + r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k})^2 + r_{2n,k}(\widehat{E}_{i,k})^2 + r^2_{3n,k,i} + r^2_{4n,k,i}\Big] \\
&\equiv D_0 + D_1 + D_2 + D_3 + D_4 + D_5 + D_6. \qquad \text{(A.53)}
\end{aligned}$$

We treat the D_i's one by one. First, by (C.3) in Lemma C.3,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\}\frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))(E_k)}\right] \\
&= \int_{A_2} \frac{|\Phi^{\leftarrow}(u)| \vee 1}{u \wedge (1-u)} \mathrm{d}u \lesssim \log^{3/2}(n), \qquad \text{(A.54)} \\
&\mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\}\, \frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))^2(E_k)}\right] \\
&= \int_{A_2} \frac{|\Phi^{\leftarrow}(u)| \vee 1}{(u \wedge (1-u))^2} \mathrm{d}u \lesssim \log^{1/2}(n)\, n^{\gamma}. \qquad \text{(A.55)}
\end{aligned}$$

Now, for the term D_0, using (A.1), Lemma C.1 and (A.55),

$$\begin{aligned}
D_0 &= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|}{(F_k \wedge (1-F_k))^2(E_{i,k})} \left\{(F^{\mathrm{r}}_{n,k} - F_{n,k})(E_{i,k})\right\}^2 \\
&= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\, n^{-3/2}\, \mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\}\frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))^2(E_k)}\right]\right) \\
&= \mathcal{O}_{\mathrm{p}}\left(\log(n)\, n^{-3/2+\gamma}\right). \qquad \text{(A.56)}
\end{aligned}$$

Next we treat D_1. Using (A.46), we have

$$\begin{aligned}
D_1 &= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k})\in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|}{(F_k \wedge (1-F_k))^2(E_{i,k})} \{(F_{n,k} - F_k)(E_{i,k})\}^2 \\
&= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\, \log\log(n)\, n^{-3/2} \sum_{F_k(E_{i,k})\in A_2} \frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))(E_{i,k})}\right) \\
&= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\, \log\log(n)\, n^{-1/2}\, \mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\}\, \frac{|\Phi^{\leftarrow}(F_k(E_k))| \vee 1}{(F_k \wedge (1-F_k))(E_k)}\right]\right) \\
&= \mathcal{O}_{\mathrm{p}}\left(\log^2(n)\, \log\log(n)\, n^{-1/2}\right).
\end{aligned}$$

Due to $\gamma < 1$ (see (3.21)), the bound on D_1 is of a larger order than that on D_0.

Next we treat D_2. We have

$$\begin{aligned}
D_2 &= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|}{(F_k \wedge (1-F_k))^2(E_{i,k})} \\
&\quad \times \left\{ f_k(E_{i,k})(\widehat{\mathbf{B}}_k - \mathbf{B}_k)^\top \mathbf{X}^{\mathrm{c}}_i \right\}^2 \\
&= \mathcal{O}_{\mathrm{p}}\left(\log^{1/2}(n)\, n^{-1/2} \delta^2_{\mathbf{B},n} \sum_{F_k(E_{i,k}) \in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\}}{(F_k \wedge (1-F_k))^2(E_{i,k})} f^2_k(E_{i,k}) \, \|\mathbf{X}^{\mathrm{c}}_i\|^2 \right) \\
&= \mathcal{O}_{\mathrm{p}}\Bigg(\log^{1/2}(n)\, n^{1/2} \delta^2_{\mathbf{B},n} \mathbb{E}\Bigg[\mathbb{1}\{F_k(E_k) \in A_2\} \left\{ \frac{f_k(E_k)}{(F_k \wedge (1-F_k))(E_k)} \right\}^2 \\
&\quad \times \{|\Phi^{\leftarrow}(F_k(E_k))| \vee 1\} \Bigg] \mathbb{E}\left[\|\mathbf{X}\|^2\right] \Bigg) \\
&= \mathcal{O}_{\mathrm{p}}(\log^{1/2}(n)\, n^{1/2} \delta^2_{\mathbf{B},n} \delta_n). \qquad \text{(A.57)}
\end{aligned}$$

Here the last step follows from (3.24).

Next we treat $D_3 + D_5$. Starting from (3.8), (3.13) and invoking s_n from (A.9), then invoking (A.49), and finally using Hölder's inequality and invoking δ_n from (3.24), we have

$$\begin{aligned}
&D_3 + D_5 \\
&= \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\} \left|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\right|}{(F_k \wedge (1-F_k))^2(E_{i,k})} \\
&\qquad \times \left\{ r^{\mathrm{r}}_{1n,k}(\widehat{E}_{i,k})^2 + r^2_{3n,k,i} \right\} \\
&= \mathcal{O}_{\mathrm{p}}\Bigg(s_n \sum_{F_k(E_{i,k}) \in A_2} \frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))^2(E_{i,k})} \\
&\qquad \times \left\{ s^2_n \delta_{\mathbf{B},n} \left\{ f_k(\widehat{E}_{i,k}) + f_k(E_{i,k}) \|\mathbf{X}_i\| \right\} + s^2_n \delta^2_{\mathbf{B},n}(1 + \|\mathbf{X}_i\|^2) + s^4_n \right\} \Bigg) \\
&= \mathcal{O}_{\mathrm{p}}\Bigg(s^3_n \sum_{F_k(E_{i,k}) \in A_2} \frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))^2(E_{i,k})} \\
&\qquad \times \left\{ f_k(E_{i,k}) \delta_{\mathbf{B},n}(1 + \|\mathbf{X}_i\|) + \delta^2_{\mathbf{B},n}(1 + \|\mathbf{X}_i\|^2) + \delta^2_{\mathbf{B},n}\|\mathbf{X}_i\| + s^2_n \right\} \Bigg) \\
&= \mathcal{O}_{\mathrm{p}}\Bigg(s^3_n \, n \, \mathbb{E}\Bigg[\mathbb{1}\{F_k(E_k) \in A_2\} \frac{|\Phi^{\leftarrow}(F_k(E_k))| \vee 1}{(F_k \wedge (1-F_k))^2(E_k)} \\
&\qquad \times \left\{ f_k(E_k) \delta_{\mathbf{B},n}(1 + \|\mathbf{X}\|) + \delta^2_{\mathbf{B},n}(1 + \|\mathbf{X}\|^2) + \delta^2_{\mathbf{B},n}\|\mathbf{X}\| + s^2_n \right\} \Bigg] \Bigg)
\end{aligned}$$

$$= \mathcal{O}_{\mathrm{p}}\Big(s_n^3\, n\, \delta_{\mathbf{B},n}\, \mathbb{E}\left[\mathbb{1}\{F_k(E_k) \in A_2\} \frac{\{|\Phi^{\leftarrow}(F_k(E_k))| \vee 1\}\, f_k(E_k)}{(F_k \wedge (1-F_k))^2(E_k)} \right] + \log^2(n)\, n^{-1/2+\gamma}\, \delta_{\mathbf{B},n}^2 + \log^3(n)\, n^{-3/2+\gamma} \Big)$$

$$= \mathcal{O}_{\mathrm{p}}\Big(s_n^3\, n\, \delta_{\mathbf{B},n} \Big\{ \mathbb{E}\left[\frac{\mathbb{1}\{F_k(E_k) \in A_2\}\, \{|\Phi^{\leftarrow}(F_k(E_k))| \vee 1\}^2\, f_k^2(E_k)}{(F_k \wedge (1-F_k))^2(E_k)} \right] \times \mathbb{E}\left[\frac{\mathbb{1}\{F_k(E_k) \in A_2\}}{(F_k \wedge (1-F_k))^2(E_k)} \right] \Big\}^{1/2} + \log^2(n)\, n^{-1/2+\gamma}\, \delta_{\mathbf{B},n}^2 + \log^3(n)\, n^{-3/2+\gamma} \Big)$$

$$= \mathcal{O}_{\mathrm{p}}(\log^{3/2}(n)\, n^{-1/2+\gamma/2} \delta_{\mathbf{B},n}\, \delta_n^{1/2} + \log^2(n)\, n^{-1/2+\gamma}\, \delta_{\mathbf{B},n}^2 + \log^3(n)\, n^{-3/2+\gamma}). \tag{A.58}$$

Finally we treat $D_4 + D_6$. Starting from (3.9) and (3.14), then using Assumption 3.4 (implying (3.20)),

$$D_4 + D_6 = \frac{1}{\sqrt{n}} \sum_{F_k(E_{i,k}) \in A_2} \frac{\{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1\} \left| \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \right|}{(F_k \wedge (1-F_k))^2(E_{i,k})} \left\{ r_{2n,k}(\widehat{E}_{i,k})^2 + r_{4n,k,i}^2 \right\}$$

$$= \mathcal{O}_{\mathrm{p}}\Big(s_n \sum_{F_k(E_{i,k}) \in A_2} \frac{|\Phi^{\leftarrow}(F_k(E_{i,k}))| \vee 1}{(F_k \wedge (1-F_k))^2(E_{i,k})} \delta_{\mathbf{B},n}^4 \left\{ 1 + \|\mathbf{X}_i\|^2 (\|\mathbf{X}_i\| \vee \|\mathbb{E}[\mathbf{X}]\|)^2 \right\} \Big)$$

$$= \mathcal{O}_{\mathrm{p}}\left(\log(n)\, n^{1/2+\gamma-2\tau}\, \delta_{\mathbf{B},n}^2 \right) \tag{A.59}$$

which, due to $\gamma - 2\tau < 0$ (see (3.21)), is of smaller order than (A.57).

Collecting terms, and further simplifying using (3.20) and (3.21), we conclude that

$$\begin{aligned} D &= \mathcal{O}_{\mathrm{p}}(D_0 + D_1 + D_2 + D_3 + D_4 + D_5 + D_6) \\ &= \mathcal{O}_{\mathrm{p}}\big(\log^2(n) \log\log(n)\, n^{-1/2} + \log^{1/2}(n)\, n^{1/2} \delta_{\mathbf{B},n}^2 \delta_n \\ &\quad + \log^{3/2}(n)\, n^{-1/2+\gamma/2} \delta_{\mathbf{B},n}\, \delta_n^{1/2} \big). \end{aligned} \tag{A.60}$$

This concludes our treatment of the term D.

Finally, (A.31) follows from the bound (A.32), the decomposition (A.34), and the bounds (A.37), (A.38), (A.40) and (A.60) on A, B, C and D respectively, after further simplifying using (3.20) and (3.21).

The second term on the right hand side of (A.30) can be treated similarly as in the above proof of (A.31) to arrive at an analogous result:

$$\frac{1}{\sqrt{n}} \sum_{i \in [n]} \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k})) \left\{ \Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'})) - \Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'})) \right\} = \mathcal{O}_{\mathrm{p}}(\Delta_n). \tag{A.61}$$

By (A.30), (A.31) and (A.61), all that's left to establish (3.27) in Theorem 3.4 is to show that the third term on the right hand side of (A.30), which is a second-order term, satisfies

$$\begin{aligned}&\Bigg|\frac{1}{\sqrt{n}}\sum_{i\in[n]}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Big\}\\ &\times\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\Big\}\Bigg|=\mathcal{O}_{\mathrm{p}}(\Delta_n).\end{aligned}\tag{A.62}$$

Using Hölder's inequality we bound the left hand side of (A.62) as

$$\begin{aligned}&\Bigg|\frac{1}{\sqrt{n}}\sum_{i\in[n]}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Big\}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\Big\}\Bigg|\\ &\le\left[\frac{1}{\sqrt{n}}\sum_{i\in[n]}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Big\}^2\right]^{1/2}\\ &\quad\times\left[\frac{1}{\sqrt{n}}\sum_{i\in[n]}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k'}(\widehat{E}_{i,k'}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k'}(E_{i,k'}))\Big\}^2\right]^{1/2}.\end{aligned}$$

In the last line above, the term in the first square bracket is $\mathcal{O}_{\mathrm{p}}(\Delta_n)$ by Lemma A.5 below; that the term in the second square bracket is also $\mathcal{O}_{\mathrm{p}}(\Delta_n)$ follows analogously. Thus (A.62) holds. This completes the proof of (3.27). The remaining part of Theorem 3.4 is trivial.

LEMMA A.5. *Assume the same conditions as in Theorem 3.4. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i\in[n]}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Big\}^2=\mathcal{O}_{\mathrm{p}}(\Delta_n).\tag{A.63}$$

PROOF. For the left hand side of (A.63), as when treating the left hand side of (A.31), we again decompose the sum over $i\in[n]$ into the cases $F_k(E_{i,k})\in A_1$ and $F_k(E_{i,k})\in A_2$; we then apply the mean value theorem to each term in the sum in the second case. We arrive at

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i\in[n]}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Big\}^2\\ &=\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_1}\Big\{\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))-\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))\Big\}^2\\ &\quad+\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi^2(\Phi^{\leftarrow}(\overline{F}^{\mathrm{r}}_{n,k,i}))}\Big\{\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})-F^{\mathrm{r}}_{n,k}(E_{i,k})\Big\}^2.\end{aligned}$$

In the above, for each i, the quantity $\overline{F}^{\mathrm{r}}_{n,k,i}$ is identical to the one appearing in (A.34) and is a random number on the open interval between $\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})$ and $F^{\mathrm{r}}_{n,k}(E_{i,k})$. Reasoning similarly as in the derivation of (A.32), the first term in the last line above is $\mathcal{O}_{\mathrm{p}}(\Delta_n)$. For the second term, using a derivation similar to that of (A.52), now using the second half of (A.41), we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\phi^2(\Phi^{\leftarrow}(\overline{F}^{\mathrm{r}}_{n,k,i}))}\left\{\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})-F^{\mathrm{r}}_{n,k}(E_{i,k})\right\}^2\\&=\mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{n}}\sum_{F_k(E_{i,k})\in A_2}\frac{1}{\left(F_k\wedge(1-F_k)\right)^2(E_{i,k})}\left\{\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k})-F^{\mathrm{r}}_{n,k}(E_{i,k})\right\}^2\right).\end{aligned}$$

By expanding the term in the curly bracket above using (A.33), and then reasoning similarly as when deriving the bounds (A.56), (A.57), (A.58) and (A.59) on the terms D_0, D_2, D_3+D_5 and D_4+D_6 respectively, the right hand side of the above equation is also $\mathcal{O}_{\mathrm{p}}(\Delta_n)$. This is enough to conclude. □

□

B. Detailed expositions for Section 4.

B.1. *Distribution, density functions and parametric score functions.* In this section we review some distribution and density functions, and parametric score functions for the copula parameter $\boldsymbol{\theta}$ related to the law $\mathrm{P}=\mathrm{P}_{\boldsymbol{\theta},\mathbf{B},F_1,\ldots,F_p,F_{\mathbf{X}}}$. The qualifier "parametric" means we treat $\boldsymbol{\theta}$ as the only variable parameter and all other parameters are fixed. (In the literature, e.g. in [17], sometimes the qualifier "ordinary" is used, but we choose "parametric" to avoid confusion with the ordinary Gaussian copula model.)

B.1.1. *Distribution and density functions.* Recall that under the law P, the copula of $\mathbf{E}$ is $C_{\boldsymbol{\theta}}=C_{\mathbf{R}(\boldsymbol{\theta})}$. Let $c_{\boldsymbol{\theta}}$ be the density of $C_{\boldsymbol{\theta}}$ and denote its logarithm by $l^{\mathrm{c}}(\cdot;\boldsymbol{\theta}):[0,1]^p\to\mathbb{R}$. Then, $l^{\mathrm{c}}(\cdot;\boldsymbol{\theta})$ is given by, e.g., Eq. (2.1) in [14] as

$$\begin{aligned}l^{\mathrm{c}}(\mathbf{u};\boldsymbol{\theta})&=\log c_{\boldsymbol{\theta}}(\mathbf{u})=-\frac{p}{2}\log(2\pi)-\frac{1}{2}\log(\det\mathbf{R}(\boldsymbol{\theta}))\\&\quad-\frac{1}{2}\Phi^{\leftarrow}_{\bullet}(\mathbf{u})^{\top}(\mathbf{S}(\boldsymbol{\theta})-\mathbf{I}_p)\Phi^{\leftarrow}_{\bullet}(\mathbf{u}),\quad \mathbf{u}=(u_1,\ldots,u_p)^{\top}\in[0,1]^p.\end{aligned}\tag{B.1}$$

Let $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ denote the conditional density of $\mathbf{Y}$ at $\mathbf{y}$ given $\mathbf{X}=\mathbf{x}$. Under the law P, using the observation that the conditional density at $\mathbf{Y}=\mathbf{y}$ given $\mathbf{X}=\mathbf{x}$ is in fact the density of $\mathbf{E}$ at $\mathbf{y}-\mathbf{B}^{\top}\mathbf{x}$, it is easy to deduce that the joint density $f_{\mathbf{Y},\mathbf{X}}(\cdot_1,\cdot_2)$ of $(\mathbf{Y},\mathbf{X})$ at $\mathbf{Y}=\mathbf{y}$ and $\mathbf{X}=\mathbf{x}$ is given by

$$f_{\mathbf{Y},\mathbf{X}}(\mathbf{y},\mathbf{x})=f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})=c_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{y}-\mathbf{B}^{\top}\mathbf{x}))\Big\{\prod_{k\in[p]}f_k(y_k-\mathbf{B}_k^{\top}\mathbf{x})\Big\}f_{\mathbf{X}}(\mathbf{x}).\tag{B.2}$$

(We recall from Section 2.2.2 that $\mathbf{B}_k$ denotes the kth column of $\mathbf{B}$.) Denote the logarithm of the above joint density $f_{\mathbf{Y},\mathbf{X}}(\cdot_1,\cdot_2)$ by $l(\cdot_1,\cdot_2;\mathrm{P}):\mathbb{R}^p\times\mathbb{R}^q\to\mathbb{R}$. By (B.2) and (B.1), $l(\cdot_1,\cdot_2;\mathrm{P})$ is given by

$$\begin{aligned}l(\mathbf{y},\mathbf{x};\mathrm{P})&=l^{\mathrm{c}}(F_{\bullet}(\mathbf{y}-\mathbf{B}^{\top}\mathbf{x});\boldsymbol{\theta})\\&\quad+\sum_{k\in[p]}\log(f_k(y_k-\mathbf{B}_k^{\top}\mathbf{x}))+\log(f_{\mathbf{X}}(\mathbf{x})),\quad \mathbf{y}=(y_1,\ldots,y_p)^{\top},\mathbf{x}\in\mathbb{R}^q.\end{aligned}\tag{B.3}$$

In the above expression of $l(\cdot_1, \cdot_2; \mathrm{P})$, we have explicitly emphasized that this log density is evaluated under the law P, because later we will also need to evaluate this log density under perturbations of the law P. Also for later purpose, we denote by $F_{\mathbf{Y},\mathbf{X}}$ the joint distribution function of $(\mathbf{Y}, \mathbf{X})$.

B.1.2. *Parametric score functions.* We first consider the ordinary Gaussian copula model. We recall that $l^{\mathrm{c}}(\cdot; \boldsymbol{\theta})$, specified in (B.1), is the logarithm of the copula density $c_{\boldsymbol{\theta}}$. Let $\dot{\mathbf{l}}^{\mathrm{c,s}}(\cdot; \boldsymbol{\theta}) = (\dot{l}^{\mathrm{c,s}}_m(\cdot; \boldsymbol{\theta}))_{m=1}^d : [0,1]^p \to \mathbb{R}^d$ be the corresponding parametric score function for the copula parameter $\boldsymbol{\theta}$; then

$$\begin{aligned}\dot{l}^{\mathrm{c,s}}_m(\mathbf{u}; \boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_m} l^{\mathrm{c}}(\mathbf{u}; \boldsymbol{\theta}) \\ &= -\frac{1}{2}\operatorname{tr}\Big\{\mathbf{S}(\boldsymbol{\theta})\dot{\mathbf{R}}_m(\boldsymbol{\theta})\Big\} - \frac{1}{2}\Phi^{\leftarrow}_{\bullet}(\mathbf{u})^{\top}\dot{\mathbf{S}}_m(\boldsymbol{\theta})\Phi^{\leftarrow}_{\bullet}(\mathbf{u}), \quad \mathbf{u} \in [0,1]^p,\ m \in [d].\end{aligned} \tag{B.4}$$

Also let $(\dot{l}^{\mathrm{c,m}}_k(\cdot; \boldsymbol{\theta}))_{k=1}^p : [0,1]^p \to \mathbb{R}^p$ be the spatial derivatives of l^{c}; then

$$\begin{aligned}\dot{l}^{\mathrm{c,m}}_k(\mathbf{u}; \boldsymbol{\theta}) &= \frac{\partial}{\partial u_k} l^{\mathrm{c}}(\mathbf{u}; \boldsymbol{\theta}) \\ &= \frac{\Phi^{\leftarrow}(u_k)}{\phi(\Phi^{\leftarrow}(u_k))} - \sum_{k' \in [p]} (\mathbf{S}(\boldsymbol{\theta}))_{kk'} \frac{\Phi^{\leftarrow}(u_{k'})}{\phi(\Phi^{\leftarrow}(u_k))}, \quad \mathbf{u} \in [0,1]^p,\ k \in [p].\end{aligned}$$

(For the above two equations, see, e.g., Eqs. (2.2) and (2.6) in [14] respectively.)

Now we turn to the covariate-adjusted Gaussian copula model. Under the law P, we recall the logarithm $l(\cdot_1, \cdot_2; \mathrm{P})$ of the joint density of $(\mathbf{Y}, \mathbf{X})$ given in (B.3), which in turn involves the logarithm l^{c} in (B.1) of the copula density $c_{\boldsymbol{\theta}}$. Let $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P}) = (\dot{l}^{\mathrm{s}}_m(\cdot_1, \cdot_2; \mathrm{P}))_{m=1}^d : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$ be the parametric score function for the copula parameter $\boldsymbol{\theta}$ corresponding to $l(\cdot_1, \cdot_2; \mathrm{P})$. Then

$$\begin{aligned}\dot{l}^{\mathrm{s}}_m(\mathbf{y}, \mathbf{x}; \mathrm{P}) &= \frac{\partial}{\partial \theta_m} l(\mathbf{y}, \mathbf{x}; \mathrm{P}) = \frac{\partial}{\partial \theta_m} l^{\mathrm{c}}(F_{\bullet}(\mathbf{y} - \mathbf{B}^{\top}\mathbf{x}); \boldsymbol{\theta}) \\ &= \dot{l}^{\mathrm{c,s}}_m(F_{\bullet}(\mathbf{y} - \mathbf{B}^{\top}\mathbf{x}); \boldsymbol{\theta}), \quad (\mathbf{y}, \mathbf{x}) \in \mathbb{R}^p \times \mathbb{R}^q,\ m \in [d],\end{aligned} \tag{B.5}$$

where $\mathbf{l}^{\mathrm{c,s}}(\cdot; \boldsymbol{\theta})$ is introduced in (B.4). Therefore, in fact

$$\dot{\mathbf{l}}^{\mathrm{s}}(\mathbf{y}, \mathbf{x}; \mathrm{P}) = \dot{\mathbf{l}}^{\mathrm{c,s}}(F_{\bullet}(\mathbf{y} - \mathbf{B}^{\top}\mathbf{x}); \boldsymbol{\theta}), \quad (\mathbf{y}, \mathbf{x}) \in \mathbb{R}^p \times \mathbb{R}^q. \tag{B.6}$$

The specific form of $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P})$ can be readily obtained from (B.6) and (B.4).

B.2. *Semiparametric lower bound.*

B.2.1. *Overview.* As can be seen from Section 4.1, a central object for obtaining the semiparametric lower bound for estimating $\boldsymbol{\theta}$ is the efficient score (function) $\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P})$. We very briefly review this and some related concepts here, while more detailed expositions will be given later in this section. For textbook treatment of these concepts, we refer the readers to Chapter 3 in [2] or Chapter 25 in [17].

Let $\widetilde{\mathcal{T}}$, given explicitly later in (B.12), be the *tangent set* associated with the nuisance parameters, and let $\mathrm{cl}(\cdot)$ denote the closure operator. It will be clear later on that the tangent set $\widetilde{\mathcal{T}}$ is in fact a linear subspace of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$, the latter being the collection of functions $f : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ that are square-integrable with respect to the law $F_{\mathbf{Y},\mathbf{X}}$ and that obey the additional restriction $\int_{\mathbb{R}^p\times\mathbb{R}^q} f(\mathbf{y}, \mathbf{x})\mathrm{d}F_{\mathbf{Y},\mathbf{X}}(\mathbf{y}, \mathbf{x}) = 0$. Then, following p. 369 in [17], the efficient score is the (coordinate-wise) projection of the parametric score function $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P})$ in (B.5) for the parameter of interest $\boldsymbol{\theta}$ onto the orthogonal complement of the closure $\mathrm{cl}(\widetilde{\mathcal{T}})$ of $\widetilde{\mathcal{T}}$. More explicitly, the efficient score is given later in (B.14).

To be technically precise, the notion of semiparametric efficiency introduced in Section 4.1 depends on the tangent set, whose choice may not be absolute. For instance, the discussion following Definition 2.8 in [16] says "In 'practice' one hunts for a pair of a tangent set and estimator sequence such that the tangent set is 'big enough' and the estimator sequence 'efficient enough' so that the latter is asymptotically efficient according to the preceding definition." (See also p. 76 to 77 in [2].) For this paper, the tangent set is $\widetilde{\mathcal{T}}$ in (B.12) and the efficient estimator (sequence) is the one-step estimator given by (4.13).

Based on our brief discussion above, the remainder of Section B.2 is further divided into the following subsections. First, in Section B.2.2, we derive $\mathrm{cl}(\widetilde{\mathcal{T}})$, the closure of the tangent set $\widetilde{\mathcal{T}}$. Then, in Section B.2.3, we calculate the efficient score.

B.2.2. *Tangent set.* In what follows, we fix some $\varepsilon > 0$ small enough. We first construct a parametric family of distributions $\{\mathrm{P}_\eta, |\eta| < \varepsilon\}$ indexed by η such that the path $\eta \to \mathrm{P}_\eta$ passes through P at $\eta = 0$. To do this, we first construct paths through the infinite-dimensional nuisance parameters F_k, $k \in [p]$ and $F_{\mathbf{X}}$. Such constructions are covered by Example 1 in Section 3.2 in [2] or Example 25.16 in [17], and the specific case for the marginals F_k, $k \in [p]$ has been worked out below Remark 2.5 in [14].

First, for each $k \in [p]$, we define the path through the univariate marginal F_k. We fix some arbitrary $h_k \in L_2^{0,\mathrm{d}}[0, 1]$, where $L_2^{0,\mathrm{d}}[0, 1]$ is defined in Section 1.5. Then, for $\eta \in (-\varepsilon, \varepsilon)$, define the univariate density $f_{k,\eta}^{h_k}$ as

$$f_{k,\eta}^{h_k}(t) = (1 + \eta\, h_k \circ F_k(t))\, f_k(t), \quad t \in \mathbb{R}. \tag{B.7}$$

Because h_k is bounded (by the choice $h_k \in L_2^{0,\mathrm{d}}[0, 1]$), we can and will choose $\varepsilon > 0$ small enough so that $f_{k,\eta}^{h_k}$ is non-negative. Furthermore it is easy to check that $f_{k,\eta}^{h_k}$ integrates to one over $\mathbb{R}$, and so it is indeed a density. We let $F_{k,\eta}^{h_k}$ be the absolutely continuous univariate distribution function induced by the density $f_{k,\eta}^{h_k}$. Then the path $\eta \to F_{k,\eta}^{h_k}$ passes through F_k at $\eta = 0$. Moreover, by Example 1 in Section 3.2 in [2], we have

$$\frac{\partial}{\partial\eta} \log f_{k,\eta}^{h_k}|_{\eta=0} = \frac{1}{f_k}\frac{\partial}{\partial\eta} f_{k,\eta}^{h_k}|_{\eta=0} = h_k \circ F_k, \tag{B.8}$$

and $h_k \circ F_k$ is the score function for the the parametric model $\{F_{k,\eta}^{h_k} : |\eta| < \epsilon\}$ at $\eta = 0$.

Analogously, we define the path through $F_{\mathbf{X}}$ as follows. We fix some arbitrary $h \in L_2^0(F_{\mathbf{X}})$, where $L_2^0(F_{\mathbf{X}})$ is defined in Section 1.5. Let the function $\Psi : \mathbb{R} \to (0, 2)$ be $\Psi(t) = 2(1 + e^{-2t})^{-1}$. Then, for $\eta \in (-\varepsilon, \varepsilon)$, define the density function $f_{\mathbf{X},\eta}^h : \mathbb{R}^q \to \mathbb{R}$ as

$$f_{\mathbf{X},\eta}^h(\mathbf{x}) = \psi(\eta; f_{\mathbf{X}}, h)\Psi(\eta h(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^q, \tag{B.9}$$

where the function $\psi(\cdot; f_{\mathbf{X}}, h) = \left\{\int_{\mathbb{R}^q} \Psi(\cdot h(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}\right\}^{-1}$. Thus $f_{\mathbf{X},\eta}^h$ integrates to one over $\mathbb{R}^q$, and so it is indeed a density. We let $F_{\mathbf{X},\eta}^h$ be the absolutely continuous distribution function in $\mathbb{R}^q$ induced by the density $f^h_{\mathbf{X},\eta}$. Then the path $\eta \to F^h_{\mathbf{X},\eta}$ passes through $F_{\mathbf{X}}$ at $\eta = 0$. Moreover, again by Example 1 in Section 3.2 in [2], we have

$$\frac{\partial}{\partial\eta} \log f^h_{\mathbf{X},\eta}|_{\eta=0} = h,$$

and h is the score function for the parametric model $\{F^h_{\mathbf{X},\eta} : |\eta| < \epsilon\}$ at $\eta = 0$.

Note that, restricting to only differentiable h_k in the construction of $f^{h_k}_{k,\eta}$ is due to the technical reason that the perturbation parameter η will appear not only directly in $f^{h_k}_{k,\eta}$ but also implicitly in its argument when evaluating the derivative of $f^{h_k}_{k,\eta}$ (see (D.1) in Section D). Because $L_2^{0,\mathrm{d}}$ is dense in $L_2^0[0,1]$, the closure of the collections of score functions generated by $F^{h_k}_{k,\eta}$ over $h_k \in L_2^{0,\mathrm{d}}$ and over $h_k \in L_2^0[0,1]$ are equal. Because in the calculation of the efficient score we project the parametric score function for $\boldsymbol{\theta}$ onto the closure of the tangent set $\widetilde{\mathcal{T}}$ (which in turn is the linear span of the score functions for the nuisance parameters), the aforementioned restriction will turn out to be irrelevant.

Now we are ready to construct the path $\eta \to \mathrm{P}_\eta$. We consider arbitrary but fixed $\boldsymbol{\alpha} \in \mathbb{R}^d$, $\boldsymbol{\beta} \in \mathbb{R}^{q\times p}$, $\mathbf{h} = (h_1, \dots, h_p) \in (L_2^{0,\mathrm{d}}[0,1])^p$ and $h \in L_2^0(F_{\mathbf{X}})$. For $\eta \in (-\varepsilon, \varepsilon)$, we denote $\mathrm{P}_\eta \equiv \mathrm{P}_{\boldsymbol{\theta}+\eta\boldsymbol{\alpha}, \mathbf{B}+\eta\boldsymbol{\beta}, F^{h_1}_{1,\eta}, \dots, F^{h_p}_{p,\eta}, F^h_{\mathbf{X},\eta}}$ (for the definition of the latter, see Section 2). The collection $\{\mathrm{P}_\eta, |\eta| < \varepsilon\}$ forms a parametric submodel that passes through $\mathrm{P} = \mathrm{P}_0$ at $\eta = 0$. We denote by $p_\eta(\cdot_1, \cdot_2)$ the joint probability density function of $(\mathbf{Y}, \mathbf{X})$ under the law P_η. Following Eq. (7.1) in [17], we say that the path $\eta \to \mathrm{P}_\eta$ is *differentiable in quadratic mean* at $\eta = 0$ with *score function* $g : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ if

$$\lim_{\eta\to 0} \int_{\mathbb{R}^p\times\mathbb{R}^q} \left(\frac{\sqrt{p_\eta(\mathbf{y},\mathbf{x})} - \sqrt{p_0(\mathbf{y},\mathbf{x})}}{\eta} - \frac{1}{2} g(\mathbf{y},\mathbf{x}) \sqrt{p_0(\mathbf{y},\mathbf{x})} \right)^2 \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{x} = 0. \tag{B.10}$$

By, e.g., Theorem 7.2 in [17], necessarily $g \in L_2^0(F_{\mathbf{Y},\mathbf{X}})$. Differentiability in quadratic mean requires certain regularity conditions that we collected earlier in Assumption 4.1.

Analogous to Proposition 2.6 in [14] (that treats the ordinary Gaussian copula model), the first part of Proposition B.1 below gives our result regarding the score function for the path $\eta \to \mathrm{P}_\eta$ in the regression setting. Just as $\mathbf{B}_k$ denotes the kth column of $\mathbf{B}$, we let $\boldsymbol{\beta}_k$ denote the kth column of $\boldsymbol{\beta}$. Then, introduce the function $\dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h}(\cdot_1, \cdot_2) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ as

$$\begin{aligned} \dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h}(\mathbf{y},\mathbf{x}) = \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{z});\boldsymbol{\theta}) - \sum_{k\in[p]} \boldsymbol{\beta}_k^\top \mathbf{x} \left\{ \dot{l}_k^{\mathrm{c,m}}\left(F_\bullet(\mathbf{z});\boldsymbol{\theta}\right) f_k(z_k) + (\dot{f}_k/f_k)(z_k) \right\} \\ + \sum_{k\in[p]} \left\{ h_k \circ F_k(z_k) + \dot{l}_k^{\mathrm{c,m}}\left(F_\bullet(\mathbf{z});\boldsymbol{\theta}\right) \int_0^{F_k(z_k)} h_k\left(u\right)\, \mathrm{d}u \right\} + h(\mathbf{x}), \\ \mathbf{z} = (z_1, \dots, z_p)^\top \equiv \mathbf{y} - \mathbf{B}^\top \mathbf{x}. \end{aligned} \tag{B.11}$$

In brief, by the first part of Proposition B.1, $\dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h} \in L_2^0(F_{\mathbf{Y},\mathbf{X}})$ is the score function along the path $\eta \to \mathrm{P}_\eta$ at $\eta = 0$. Clearly, $\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0}$ (obtained by fixing $\boldsymbol{\beta} = \mathbf{0}$, $\mathbf{h} = \mathbf{0}$, $h = 0$) is a linear combination of the components of the parametric score function for $\boldsymbol{\theta}$ (see (B.6)), while $\dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{h},h}$ (obtained by fixing $\boldsymbol{\alpha} = \mathbf{0}$) is the score function for the nuisance parameters. Following the convention in p. 369 in [17], the collection of the score functions $\dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{h},h}$ for the nuisance parameters over $\boldsymbol{\beta} \in \mathbb{R}^{q\times p}$, $\mathbf{h} = (h_1, \ldots, h_p) \in (L_2^{0,\mathrm{d}}[0,1])^p$ and $h \in L_2^0(F_{\mathbf{X}})$ is by definition the (nuisance) tangent set at P, which we denote by $\widetilde{\mathcal{T}}$. Explicitly,

$$\widetilde{\mathcal{T}} = \left\{ \dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{h},h} : \boldsymbol{\beta} \in \mathbb{R}^{q\times p}, \mathbf{h} = (h_1, \ldots, h_p) \in (L_2^{0,\mathrm{d}}[0,1])^p, h \in L_2^0(F_{\mathbf{X}}) \right\}. \tag{B.12}$$

In our case, $\widetilde{\mathcal{T}}$ is a linear subspace of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$.

To succinctly represent $\mathrm{cl}(\widetilde{\mathcal{T}})$, introduce the following collection of functions:

$$\begin{aligned}
\mathcal{T}_1 &= \left\{ \dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0} : \boldsymbol{\alpha} \in \mathbb{R}^d \right\} \\
&= \left\{ f : f(\mathbf{y},\mathbf{x}) = \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{y} - \mathbf{B}^\top \mathbf{x}); \boldsymbol{\theta}), \boldsymbol{\alpha} \in \mathbb{R}^d \right\}, \qquad \text{(B.13)} \\
\mathcal{T}_2 &= \Bigg\{ f_{\boldsymbol{\beta}} : f_{\boldsymbol{\beta}}(\mathbf{y},\mathbf{x}) \equiv - \sum_{k\in[p]} \boldsymbol{\beta}_k^\top (\mathbf{x} - \mathbb{E}[\mathbf{X}]) \\
&\qquad \times \left\{ \dot{l}_k^{\mathrm{c,m}} \left(F_\bullet(\mathbf{y} - \mathbf{B}^\top \mathbf{x}); \boldsymbol{\theta} \right) f_k(y_k - \mathbf{B}_k^\top \mathbf{x}) + (\dot{f}_k/f_k)(y_k - \mathbf{B}_k^\top \mathbf{x}) \right\}, \boldsymbol{\beta} \in \mathbb{R}^{q\times p} \Bigg\}, \\
\mathcal{T}_3 &= \left\{ \dot{l}^{\mathbf{0},\mathbf{0},\mathbf{h},0} : \mathbf{h} \in (L_2^0[0,1])^p \right\}, \quad \mathcal{T}_4 = \left\{ \dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h} : h \in L_2^0(F_{\mathbf{X}}) \right\}.
\end{aligned}$$

It is clear that $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_4$, consisting of different components of the score function $\dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h}$, are linear subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$; that $\mathcal{T}_2$ is as well will be part of Proposition B.1.

PROPOSITION B.1. *Suppose that Assumption 4.1 holds. Let $\boldsymbol{\alpha} \in \mathbb{R}^d$, $\boldsymbol{\beta} \in \mathbb{R}^{q\times p}$, $\mathbf{h} = (h_1, \ldots, h_p) \in (L_2^{0,\mathrm{d}}[0,1])^p$ and $h \in L_2^0(F_{\mathbf{X}})$. Then, the path $\eta \to \mathrm{P}_\eta$ is differentiable in quadratic mean at $\eta = 0$ with score function $\dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h}$, that is,* (B.10) *holds with $g(\mathbf{y},\mathbf{x})$ replaced by $\dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h}(\mathbf{y},\mathbf{x})$.*

Furthermore, the $\mathcal{T}_a$'s, $a \in \{1,2,3,4\}$, are closed linear subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$ and are pairwise orthogonal (in $L_2(F_{\mathbf{Y},\mathbf{X}})$) except between $\mathcal{T}_1$ and $\mathcal{T}_3$, that is, $\mathcal{T}_a \perp \mathcal{T}_b$ for all $a, b \in \{1,2,3,4\}$ and $a \neq b$ with the exception of $a = 1$ and $b = 3$. Moreover, the closure $\mathrm{cl}(\widetilde{\mathcal{T}})$ of the (nuisance) tangent set $\widetilde{\mathcal{T}}$ at P *is given by $\mathrm{cl}(\widetilde{\mathcal{T}}) = \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$.*

PROOF. The proof is deferred to Section B.3. □

B.2.3. *Efficient score.* Having determined that $\mathrm{cl}(\widetilde{\mathcal{T}}) = \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$ from Proposition B.1, we are now ready to calculate the efficient score. We recall that the efficient score $\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P}) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$ for the copula parameter $\boldsymbol{\theta}$ at P is the (coordinate-wise) projection of the parametric score function $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P}) = (\dot{l}^{\mathrm{s}}_m(\cdot_1, \cdot_2; \mathrm{P}))_{m=1}^d$ for $\boldsymbol{\theta}$ onto the orthogonal complement of $\mathrm{cl}(\widetilde{\mathcal{T}})$. Hence,

$$\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P}) = \dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P}) - \Pi_{F_{\mathbf{Y},\mathbf{X}}} \left(\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P}) | \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 \right), \tag{B.14}$$

where $\Pi_{F_{\mathbf{Y},\mathbf{X}}}(\cdot|\mathcal{T})$ denotes the coordinate-wise projection operator from $L_2(F_{\mathbf{Y},\mathbf{X}})$ onto $\mathcal{T}$ where $\mathcal{T}$ should be a closed linear subspace of $L_2(F_{\mathbf{Y},\mathbf{X}})$. The explicit expression of $\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P})$ is stated in Proposition 4.1 which determines that $\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1, \cdot_2; \mathrm{P})$ is closely related to the efficient score $\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_\bullet(\cdot); \boldsymbol{\theta})$ in the ordinary Gaussian copula model (with unknown margins).

B.3. *Proof of Proposition B.1.*

PROOF. We first verify differentiability in quadratic mean and calculate the score function. We wish to apply Proposition 2.1.1 in [2] or Lemma 7.6 in [17]. We essentially need to verify two conditions: first, $p_\eta^{1/2}(\mathbf{y},\mathbf{x})$ is continuously differentiable in η for every $(\mathbf{y},\mathbf{x})$, and second, the information "matrix" (which is in fact a scalar) I_η for η, introduced below in (B.17), is well-defined and is continuous at $\eta = 0$, that is,

$$I_\eta \to I_0. \tag{B.15}$$

Here and in what follows all convergences are taken along the limit $\eta \to 0$. (Note that, on surface Lemma 7.6 in [17] requires continuity of I_η in η, but a closer inspection of its proof reveals that (B.15) is precisely what's required to the deduce the differentiability in quadratic mean of $\eta \to \mathrm{P}_\eta$ at $\eta = 0$.) The first condition is easily checked.

Next we briefly discuss the continuity of I_η, deferring some detailed calculations to Section D. For notational brevity we denote $\boldsymbol{\theta}_\eta = \boldsymbol{\theta} + \eta\boldsymbol{\alpha}$, $\mathbf{B}_\eta = \mathbf{B} + \eta\boldsymbol{\beta}$ and let $\mathbf{B}_{k,\eta}$ be the kth column of $\mathbf{B}_\eta$. Also let $F^{\mathbf{h}}_{\bullet,\eta} : \mathbb{R}^p \to \mathbb{R}^p$ be

$$F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{z}) = \left(F^{h_1}_{1,\eta}(z_1), \ldots, F^{h_p}_{p,\eta}(z_p)\right)^\top, \quad \mathbf{z} = (z_1, \ldots, z_p)^\top.$$

Let $l(\mathbf{y},\mathbf{x};\mathrm{P}_\eta)$ denote the logarithm of the joint density p_η. Similar to (B.3),

$$\begin{aligned} l(\mathbf{y},\mathbf{x};\mathrm{P}_\eta) &= l^{\mathrm{c}}\left(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{y} - \mathbf{B}_\eta^\top\mathbf{x}); \boldsymbol{\theta}_\eta\right) \\ &\quad + \sum_{k\in[p]} \log(f^{h_k}_{k,\eta}(y_k - \mathbf{B}_{k,\eta}^\top\mathbf{x})) + \log(f^{h}_{\mathbf{X},\eta}(\mathbf{x})), \quad \mathbf{y} = (y_1,\ldots,y_p)^\top, \mathbf{x} \in \mathbb{R}^q. \end{aligned} \tag{B.16}$$

Then the information "matrix" I_η is given by

$$I_\eta = \int_{\mathbb{R}^p\times\mathbb{R}^q} \left\{\frac{\partial}{\partial\eta} l(\mathbf{y},\mathbf{x};\mathrm{P}_\eta)\right\}^2 p_\eta(\mathbf{y},\mathbf{x})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}, \tag{B.17}$$

where

$$\frac{\partial}{\partial\eta} l(\mathbf{y},\mathbf{x};\mathrm{P}_\eta) = \frac{1}{p_\eta(\mathbf{y},\mathbf{x})}\frac{\partial}{\partial\eta} p_\eta(\mathbf{y},\mathbf{x})$$

is understood to be defined arbitrarily when $p_\eta(\mathbf{y},\mathbf{x}) = 0$. By calculations detailed in Section D, when $p_\eta(\mathbf{y},\mathbf{x}) \neq 0$,

$$\begin{aligned} \frac{\partial}{\partial\eta} l(\mathbf{y},\mathbf{x};\mathrm{P}_\eta) &= \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{y} - \mathbf{B}_\eta^\top\mathbf{x}); \boldsymbol{\theta}_\eta) \\ &\quad - \sum_{k\in[p]} \boldsymbol{\beta}_k^\top\mathbf{x}\Bigg\{ \dot{l}^{\mathrm{c,m}}_k\left(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{y} - \mathbf{B}_\eta^\top\mathbf{x}); \boldsymbol{\theta}_\eta\right) f^{h_k}_{k,\eta}(y_k - \mathbf{B}_{k,\eta}^\top\mathbf{x}) \\ &\qquad + \eta\, \dot{h}_k \circ F_k(y_k - \mathbf{B}_{k,\eta}^\top\mathbf{x}) \frac{f_k^2}{f^{h_k}_{k,\eta}}(y_k - \mathbf{B}_{k,\eta}^\top\mathbf{x}) \\ &\qquad + (1 + \eta\, h_k \circ F_k(y_k - \mathbf{B}_{k,\eta}^\top\mathbf{x}))\frac{\dot{f}_k}{f^{h_k}_{k,\eta}}(y_k - \mathbf{B}_{k,\eta}^\top\mathbf{x}) \Bigg\} \end{aligned}$$

$$+\sum_{k\in[p]}\Bigg\{h_k\circ F_k\left(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}\right)\frac{f_k}{f_{k,\eta}^{h_k}}(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})$$
$$+\dot{l}_k^{\mathrm{c,m}}\left(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{y}-\mathbf{B}_\eta^\top\mathbf{x});\boldsymbol{\theta}_\eta\right)\int_0^{F_k\left(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}\right)}h_k(u)\mathrm{d}u\Bigg\}$$
$$+\left\{\dot{\psi}(\eta;f_{\mathbf{X}},h)\psi(\eta;f_{\mathbf{X}},h)^{-1}+\Psi(\eta h(\mathbf{x}))\exp(-2\eta h(\mathbf{x}))h(\mathbf{x})\right\},\tag{B.18}$$

where $\dot{\psi}(\cdot;f_{\mathbf{X}},h)$ is the derivative of $\psi(\cdot;f_{\mathbf{X}},h)$ as in (D.4).

Section D establishes that (B.15) holds. Thus, finally, by Proposition 2.1.1 in [2] or Lemma 7.6 in [17], we conclude that the path $\eta\to\mathrm{P}_\eta$ is differentiable in quadratic mean at $\eta=0$, and the score function is given by $(\partial/\partial\eta)l(\mathbf{y},\mathbf{x};\mathrm{P}_\eta)$ evaluated at $\eta=0$. Evaluating (B.18) at $\eta=0$ yields that (B.11) is the score function.

Next we prove the second half of the proposition. *We first establish that the $\mathcal{T}_a$'s, $a\in\{1,2,3,4\}$ are all closed linear subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$.*

Because $\mathcal{T}_1$, $\mathcal{T}_3$ and $\mathcal{T}_4$ are simply different collections of score functions $\dot{l}^{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},h}\in L_2^0(F_{\mathbf{Y},\mathbf{X}})$, they are clearly linear subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$. Next, $\mathcal{T}_1$ and $\mathcal{T}_2$ are both closed because they are finite dimensional, and $\mathcal{T}_4$ is closed because $L_2^0(F_{\mathbf{X}})$ is closed in $L_2(F_{\mathbf{X}})$.

Thus we just need to show that $\mathcal{T}_2$ is a linear subspace of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$, and that $\mathcal{T}_3$ is closed. We first consider $\mathcal{T}_3$. For brevity of presentation, we introduce the score operator as in Eq. (2.8) in [14]. Let $L_2^0(C_{\boldsymbol{\theta}})$ be the collection of functions $f:[0,1]^p\to\mathbb{R}$ that are square-integrable with respect to the distribution $C_{\boldsymbol{\theta}}$ and that obey the additional restriction $\int_{[0,1]^p}f(\mathbf{u})\mathrm{d}C_{\boldsymbol{\theta}}(\mathbf{u})=0$. For $k\in[p]$, define the bounded linear operator $\mathcal{O}_{\boldsymbol{\theta},k}:L_2^0[0,1]\to L_2^0(C_{\boldsymbol{\theta}})$ (see Lemma 2.4 in [14]) by

$$\mathcal{O}_{\boldsymbol{\theta},k}h=[\mathcal{O}_{\boldsymbol{\theta},k}h](\mathbf{u})=h(u_k)+\dot{l}_k^{\mathrm{c,m}}(\mathbf{u};\boldsymbol{\theta})H(u_k),\quad \mathbf{u}=(u_1,\ldots,u_p)^\top,\ h\in L_2^0[0,1],$$

where $H(u)=\int_0^u h(t)\mathrm{d}t$. Then, following the construction below Eq. (2.8) in [14], we further define the bounded score operator $\mathcal{O}_{\boldsymbol{\theta}}:(L_2^0[0,1])^p\to L_2^0(C_{\boldsymbol{\theta}})$ by

$$\mathcal{O}_{\boldsymbol{\theta}}\mathbf{h}=\sum_{k\in[p]}\mathcal{O}_{\boldsymbol{\theta},k}h_k,\quad \mathbf{h}=(h_1,\ldots,h_p)\in(L_2^0[0,1])^p.$$

It is readily checked that in terms of the operator $\mathcal{O}_{\boldsymbol{\theta}}\mathbf{h}$, $\mathcal{T}_3$ can be succinctly expressed as

$$\mathcal{T}_3=\left\{f:\mathbb{R}^p\times\mathbb{R}^q\to\mathbb{R},\ f(\cdot_1,\cdot_2)=[\mathcal{O}_{\boldsymbol{\theta}}\mathbf{h}](F_\bullet(\cdot_1-\mathbf{B}^\top\cdot_2)),\ \mathbf{h}\in(L_2^0[0,1])^p\right\}.\tag{B.19}$$

In addition, Remark 2.5 in [14] states that the range of $(L_2^0[0,1])^p$ under $\mathcal{O}_{\boldsymbol{\theta}}$, which we denote by

$$\mathcal{T}_{3,C_{\boldsymbol{\theta}}}\equiv\{\mathcal{O}_{\boldsymbol{\theta}}\mathbf{h}:\mathbf{h}\in(L_2^0[0,1])^p\},\tag{B.20}$$

is a closed subspace of $L_2^0(C_{\boldsymbol{\theta}})$. Consider the map from $\mathcal{T}_{3,C_{\boldsymbol{\theta}}}$ to $\mathcal{T}_3$ that explicitly maps $\mathcal{O}_{\boldsymbol{\theta}}\mathbf{h}$ to $[\mathcal{O}_{\boldsymbol{\theta}}\mathbf{h}](F_\bullet(\cdot_1-\mathbf{B}^\top\cdot_2))$. From (B.20) and (B.19), clearly this map is a bijection. In what follows, let $\mathbf{U}=F_\bullet(\mathbf{E})=F_\bullet(\mathbf{Y}-\mathbf{B}^\top\mathbf{X})$; then $\mathbf{U}$ has distribution function $C_{\boldsymbol{\theta}}$. It is then easy to see that this map is also an isometry. Hence, $\mathcal{T}_3$ is a closed linear subspace of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$.

Next we consider $\mathcal{T}_2$. By the facts that

$$\int_{[0,1]} \frac{\dot{f}_k}{f_k} \circ F_k^{\leftarrow}(u)\mathrm{d}u = \int_{[a_k,b_k]} \frac{\dot{f}_k}{f_k} \circ F_k^{\leftarrow} \circ F_k(t) f_k(t)\mathrm{d}t$$
$$= \int_{[a_k,b_k]} \dot{f}_k(t)\mathrm{d}t = f(b_k) - f(a_k) = 0$$

and, by (4.8),

$$\left\| \frac{\dot{f}_k}{f_k} \circ F_k^{\leftarrow} \right\|^2_{L_2[0,1]} = \int_{[0,1]} \left\{ \frac{\dot{f}_k}{f_k} \circ F_k^{\leftarrow}(u) \right\}^2 \mathrm{d}u = \int_{[a_k,b_k]} \left\{ \frac{\dot{f}_k}{f_k} \circ F_k^{\leftarrow} \circ F_k(t) \right\}^2 f_k(t)\mathrm{d}t$$
$$= \int_{[a_k,b_k]} \frac{\dot{f}_k^2}{f_k}(t)\mathrm{d}t < \infty,$$

we conclude that in fact $(\dot{f}_k/f_k) \circ F_k^{\leftarrow} \in L_2^0[0,1]$. Then it is readily checked that, for $h_{f_k} \equiv (\dot{f}_k/f_k) \circ F_k^{\leftarrow} \in L_2^0[0,1]$,

$$\begin{aligned} &\dot{l}_k^{\mathrm{c,m}}\left(F_\bullet(\mathbf{z});\boldsymbol{\theta}\right) f_k(z_k) + (\dot{f}_k/f_k)(z_k) \\ &= [\mathcal{O}_{\boldsymbol{\theta},k} h_{f_k}](\mathbf{u}), \quad \mathbf{u} = F_\bullet(\mathbf{z}),\, \mathbf{z} = (z_1,\ldots,z_p)^\top \in \mathbb{R}^p. \end{aligned} \tag{B.21}$$

Then we can write an arbitrary $f_{\boldsymbol{\beta}} \in \mathcal{T}_2$ as

$$\begin{aligned} f_{\boldsymbol{\beta}}(\mathbf{y},\mathbf{x}) &= \dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{0},0}(\mathbf{y},\mathbf{x}) + \sum_{k\in[p]} \boldsymbol{\beta}_k^\top \mathbb{E}[\mathbf{X}][\mathcal{O}_{\boldsymbol{\theta},k} h_{f_k}](F_\bullet(\mathbf{y} - \mathbf{B}^\top \mathbf{x})) \\ &= \dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{0},0}(\mathbf{y},\mathbf{x}) + [\mathcal{O}_{\boldsymbol{\theta}} \mathbf{h}_{\boldsymbol{\beta}}](F_\bullet(\mathbf{y} - \mathbf{B}^\top \mathbf{x})), \end{aligned} \tag{B.22}$$

where we have introduced $\mathbf{h}_{\boldsymbol{\beta}} = (\boldsymbol{\beta}_1^\top \mathbb{E}[\mathbf{X}] h_{f_1}, \ldots, \boldsymbol{\beta}_p^\top \mathbb{E}[\mathbf{X}] h_{f_p})^\top \in (L_2^0[0,1])^p$. Thus, because $\dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{0},0} \in L_2^0(F_{\mathbf{Y},\mathbf{X}})$ and $[\mathcal{O}_{\boldsymbol{\theta}} \mathbf{h}_{\boldsymbol{\beta}}](F_\bullet(\cdot_1 - \mathbf{B}^\top \cdot_2)) \in \mathcal{T}_3 \subset L_2^0(F_{\mathbf{Y},\mathbf{X}})$ (see (B.19)), clearly $f_{\boldsymbol{\beta}} \in L_2^0(F_{\mathbf{Y},\mathbf{X}})$, and we conclude that $\mathcal{T}_2$ is a closed linear subspace of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$.

We have now shown that the $\mathcal{T}_a$'s, $a \in \{1,2,3,4\}$ are all closed linear subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$.

Next we note another fact that will be useful later on. First, we let

$$\widetilde{\mathcal{T}}_3 = \left\{ \dot{l}^{\mathbf{0},\mathbf{0},\mathbf{h},0} : \mathbf{h} \in (L_2^{0,\mathrm{d}}[0,1])^p \right\}. \tag{B.23}$$

Given that $\mathcal{T}_3$ is closed and $L_2^{0,\mathrm{d}}[0,1]$ is dense in $L_2^0[0,1]$ (see the remark regarding $L_2^{0,\mathrm{d}}[0,1]$ in Section 1.5), it is straightforward to show that

$$\mathcal{T}_3 = \mathrm{cl}(\widetilde{\mathcal{T}}_3). \tag{B.24}$$

Now we establish the pairwise orthogonality of the $\mathcal{T}_a$'s.

We first show that $\mathcal{T}_1 \perp \mathcal{T}_2$, which is related to a remark in the second paragraph on p. 1937 in [14]. Note that the inner product between two elements $l_1, l_2 \in L_2(F_{\mathbf{Y},\mathbf{X}})$ can be written as $\mathbb{E}[l_1(\mathbf{Y},\mathbf{X})l_2(\mathbf{Y},\mathbf{X})]$, where for the remainder of this proof the expectation $\mathbb{E}$ is always taken under the law P. We take arbitrary $\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0} \in \mathcal{T}_1$ and $f_{\boldsymbol{\beta}} \in \mathcal{T}_2$. Then

$$\begin{aligned}
&\mathbb{E}[\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0}(\mathbf{Y},\mathbf{X})f_{\boldsymbol{\beta}}(\mathbf{Y},\mathbf{X})] \\
&= -\boldsymbol{\alpha}^\top \sum_{k\in[p]} \mathbb{E}\Big[\dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{Y}-\mathbf{B}^\top\mathbf{X});\boldsymbol{\theta})\boldsymbol{\beta}_k^\top(\mathbf{X}-\mathbb{E}[\mathbf{X}]) \\
&\quad \times \Big\{\dot{l}_k^{\mathrm{c,m}}\Big(F_\bullet(\mathbf{Y}-\mathbf{B}^\top\mathbf{X});\boldsymbol{\theta}\Big) f_k(Y_k-\mathbf{B}_k^\top\mathbf{X}) - (\dot{f}_k/f_k)(Y_k-\mathbf{B}_k^\top\mathbf{X})\Big\}\Big] \\
&= -\boldsymbol{\alpha}^\top \sum_{k\in[p]} \mathbb{E}\Big[\dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{E});\boldsymbol{\theta})\boldsymbol{\beta}_k^\top(\mathbf{X}-\mathbb{E}[\mathbf{X}])\Big\{\dot{l}_k^{\mathrm{c,m}}(F_\bullet(\mathbf{E});\boldsymbol{\theta})\, f_k(E_k) - (\dot{f}_k/f_k)(E_k)\Big\}\Big] \\
&= -\boldsymbol{\alpha}^\top \sum_{k\in[p]} \mathbb{E}\Big[\boldsymbol{\beta}_k^\top(\mathbf{X}-\mathbb{E}[\mathbf{X}])\Big]\,\mathbb{E}\Big[\dot{\mathbf{l}}^{\mathrm{c,s}}(\mathbf{U};\boldsymbol{\theta})\Big\{\dot{l}_k^{\mathrm{c,m}}(\mathbf{U};\boldsymbol{\theta})\, f_k(E_k) - (\dot{f}_k/f_k)(E_k)\Big\}\Big] = 0.
\end{aligned}$$

Here the third step follows by the independence of $\mathbf{X}$ and $\mathbf{U} = F_\bullet(\mathbf{E})$, and the last step follows because $\mathbf{X} - \mathbb{E}[\mathbf{X}]$ is centered. That $\mathcal{T}_2 \perp \mathcal{T}_3$ can be shown analogously (the convenient representation (B.19) makes this easier to check).

Now we show that $\mathcal{T}_1 \perp \mathcal{T}_4$. We take arbitrary $\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0} \in \mathcal{T}_1$ and $\dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h} \in \mathcal{T}_4$. Then

$$\begin{aligned}
&\mathbb{E}[\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0}(\mathbf{Y},\mathbf{X})\dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h}(\mathbf{Y},\mathbf{X})] = \mathbb{E}[\boldsymbol{\alpha}^\top\dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{Y}-\mathbf{B}^\top\mathbf{X});\boldsymbol{\theta})h(\mathbf{X})] \\
&= \mathbb{E}[\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0}(\mathbf{Y},\mathbf{X})]\mathbb{E}[\dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h}(\mathbf{Y},\mathbf{X})] = 0.
\end{aligned}$$

Here the second step follows again by the independence of $\mathbf{X}$ and $\mathbf{U} = F_\bullet(\mathbf{Y}-\mathbf{B}^\top\mathbf{X})$, and the last step follows because the score functions $\dot{l}^{\boldsymbol{\alpha},\mathbf{0},\mathbf{0},0}$ and $\dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h}$ have mean zero. That $\mathcal{T}_3 \perp \mathcal{T}_4$ can be shown analogously.

Now we show that $\mathcal{T}_2 \perp \mathcal{T}_4$. We take arbitrary $f_{\boldsymbol{\beta}} \in \mathcal{T}_2$ and $\dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h} \in \mathcal{T}_4$. Then, first by (B.21), and then by $h_{f_k} \in L_2^0[0,1]$ and so $\mathcal{O}_{\boldsymbol{\theta},k}h_{f_k} \in L_2^0(C_{\boldsymbol{\theta}})$ which in turn implies that $\mathbb{E}\big[[\mathcal{O}_{\boldsymbol{\theta},k}h_{f_k}](\mathbf{U})\big] = 0$,

$$\begin{aligned}
\mathbb{E}[f_{\boldsymbol{\beta}}(\mathbf{Y},\mathbf{X})\dot{l}^{\mathbf{0},\mathbf{0},\mathbf{0},h}(\mathbf{Y},\mathbf{X})] &= -\sum_{k\in[p]} \mathbb{E}\Big[\boldsymbol{\beta}_k^\top(\mathbf{X}-\mathbb{E}[\mathbf{X}])[\mathcal{O}_{\boldsymbol{\theta},k}h_{f_k}](\mathbf{U})h(\mathbf{X})\Big] \\
&= -\sum_{k\in[p]} \boldsymbol{\beta}_k^\top\mathbb{E}\big[[\mathcal{O}_{\boldsymbol{\theta},k}h_{f_k}](\mathbf{U})\big]\,\mathbb{E}[(\mathbf{X}-\mathbb{E}[\mathbf{X}])h(\mathbf{X})] = 0.
\end{aligned}$$

We have now established all of the orthogonality statements.

At last we establish that $\mathrm{cl}(\widetilde{\mathcal{T}}) = \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$. Let

$$\widetilde{\mathcal{T}}_2 = \{\dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{0},0} : \boldsymbol{\beta} \in \mathbb{R}^{q\times p}\}.$$

Also recall $\widetilde{\mathcal{T}}_3$ from (B.23). Then we can write (B.12) as $\widetilde{\mathcal{T}} = \widetilde{\mathcal{T}}_2 + \widetilde{\mathcal{T}}_3 + \mathcal{T}_4$. By (B.24), we have

$$\widetilde{\mathcal{T}}_2 + \mathrm{cl}(\widetilde{\mathcal{T}}_3) + \mathcal{T}_4 = \widetilde{\mathcal{T}}_2 + \mathcal{T}_3 + \mathcal{T}_4. \tag{B.25}$$

By (B.22), for arbitrary $\dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{0},0} \in \widetilde{\mathcal{T}}_2$ we have $\dot{l}^{\mathbf{0},\boldsymbol{\beta},\mathbf{0},0} \in \mathcal{T}_2 + \mathcal{T}_3$, yielding $\widetilde{\mathcal{T}}_2 \subset \mathcal{T}_2 + \mathcal{T}_3$. Using (B.22) in the opposite direction yields $\mathcal{T}_2 \subset \widetilde{\mathcal{T}}_2 + \mathcal{T}_3$. Thus we conclude $\mathcal{T}_2 + \mathcal{T}_3 = \widetilde{\mathcal{T}}_2 + \mathcal{T}_3$. This together with (B.25) yield

$$\widetilde{\mathcal{T}}_2 + \mathrm{cl}(\widetilde{\mathcal{T}}_3) + \mathcal{T}_4 = \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4. \tag{B.26}$$

Next,

$$\begin{aligned}\mathrm{cl}(\widetilde{\mathcal{T}}) &= \mathrm{cl}(\widetilde{\mathcal{T}}_2 + \widetilde{\mathcal{T}}_3 + \mathcal{T}_4) \subset \mathrm{cl}(\widetilde{\mathcal{T}}_2 + \mathrm{cl}(\widetilde{\mathcal{T}}_3) + \mathcal{T}_4) \\ &= \mathrm{cl}(\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4) = \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 \subset \mathrm{cl}(\widetilde{\mathcal{T}}_2 + \widetilde{\mathcal{T}}_3 + \mathcal{T}_4).\end{aligned} \tag{B.27}$$

Here, the third step follows by (B.26), the fourth step follows because, as $\mathcal{T}_2$, $\mathcal{T}_3$ and $\mathcal{T}_4$ are all closed and they are pairwise orthogonal, $\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$ is closed, and the last step follows by (B.26) (in the opposite direction) and the fact that the left hand side of (B.26) is clearly contained in $\mathrm{cl}(\widetilde{\mathcal{T}}_2+\widetilde{\mathcal{T}}_3+\mathcal{T}_4)$. Thus all inclusions in (B.27) must in fact be equalities. Finally we conclude that $\mathrm{cl}(\widetilde{\mathcal{T}}) = \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$. □

B.4. *Proof of Proposition 4.1.*

Proof. By Proposition B.1, $\mathcal{T}_2$, $\mathcal{T}_3$ and $\mathcal{T}_4$ are pairwise orthogonal closed linear subspaces in $L_2^0(F_{\mathbf{Y},\mathbf{X}})$. Recall that the projection onto a sum of two orthogonal linear subspaces of a Hilbert space equals the sum of the projections onto the two individual subspaces (e.g., Proposition 3.C in Appendix A.2 in [2]). Then, applying this fact twice on (B.14) yields

$$\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1,\cdot_2;\mathrm{P}) = \dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1,\cdot_2;\mathrm{P}) - \sum_{a\in\{2,3,4\}} \Pi_{F_{\mathbf{Y},\mathbf{X}}}\left(\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1,\cdot_2;\mathrm{P})|\mathcal{T}_a\right). \tag{B.28}$$

Now, by $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1,\cdot_2;\mathrm{P})$ given in (B.6) and $\mathcal{T}_1$ introduced in (B.13), $\mathcal{T}_1$ simply consists of all linear combinations of the components of $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1,\cdot_2;\mathrm{P})$. By the fact from Proposition B.1 that $\mathcal{T}_1 \perp \mathcal{T}_2$ and $\mathcal{T}_1 \perp \mathcal{T}_4$ and therefore the components of $\dot{\mathbf{l}}^{\mathrm{s}}$ are all orthogonal to $\mathcal{T}_2$ and $\mathcal{T}_4$, Equation (B.28) simplifies to

$$\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}(\cdot_1,\cdot_2;\mathrm{P}) = \dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1,\cdot_2;\mathrm{P}) - \Pi_{F_{\mathbf{Y},\mathbf{X}}}\left(\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1,\cdot_2;\mathrm{P})|\mathcal{T}_3\right). \tag{B.29}$$

To calculate the right hand side of (B.29), we again invoke an isometry argument between certain subspaces of $L_2^0(C_{\boldsymbol{\theta}})$ and $L_2^0(F_{\mathbf{Y},\mathbf{X}})$, as we have already done in the proof of Proposition B.1. Analogous to (B.20), we define

$$\mathcal{T}_{1,C_{\boldsymbol{\theta}}} = \left\{f : [0,1]^p \to \mathbb{R}, f(\cdot) = \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(\cdot;\boldsymbol{\theta}), \boldsymbol{\alpha}\in\mathbb{R}^d\right\} \tag{B.30}$$

where we recall $\dot{\mathbf{l}}^{\mathrm{c,s}}$ given in (B.4). Now, Proposition 2.6 in [14] states that $\mathcal{T}_{1,C_{\boldsymbol{\theta}}}$ and $\mathcal{T}_{3,C_{\boldsymbol{\theta}}}$ are closed subspaces of $L_2^0(C_{\boldsymbol{\theta}})$, and our Proposition B.1 states that $\mathcal{T}_1$, $\mathcal{T}_3$ are closed subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$. Moreover $\mathbf{U} = F_\bullet(\mathbf{Y}-\mathbf{B}^\top\mathbf{X}) = F_\bullet(\mathbf{E})$ is distributed according to $C_{\boldsymbol{\theta}}$. Comparing the representations (B.30), (B.20), (B.13) and (B.19) for $\mathcal{T}_{1,C_{\boldsymbol{\theta}}}$, $\mathcal{T}_{3,C_{\boldsymbol{\theta}}}$ (as subspaces of $L_2^0(C_{\boldsymbol{\theta}})$), $\mathcal{T}_1$, $\mathcal{T}_3$ (as subspaces of $L_2^0(F_{\mathbf{Y},\mathbf{X}})$) respectively, we again see that the bijection $f \to f(F_\bullet(\cdot_1 - \mathbf{B}^\top\cdot_2))$ from $\mathcal{T}_{1,C_{\boldsymbol{\theta}}} + \mathcal{T}_{3,C_{\boldsymbol{\theta}}}$ to $\mathcal{T}_1 + \mathcal{T}_3$ is an isometry.

Next, from (B.30), each of the d coordinates of $\dot{\mathbf{l}}^{\mathrm{c,s}}(\cdot;\boldsymbol{\theta})$ belongs to $\mathcal{T}_{1,C_{\boldsymbol{\theta}}}$. Let $\Pi_{C_{\boldsymbol{\theta}}}(\cdot|\mathcal{T}_{3,C_{\boldsymbol{\theta}}})$ be the coordinate-wise projection operator from $L_2(C_{\boldsymbol{\theta}})$ onto the closed linear subspace $\mathcal{T}_{3,C_{\boldsymbol{\theta}}}$. Then, Proposition 2.8 in [14] shows that

$$\dot{\mathbf{l}}^{\mathrm{c,s}}(\cdot;\boldsymbol{\theta}) - \Pi_{C_{\boldsymbol{\theta}}}(\dot{\mathbf{l}}^{\mathrm{c,s}}(\cdot;\boldsymbol{\theta})|\mathcal{T}_{3,C_{\boldsymbol{\theta}}}) = \dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(\cdot;\boldsymbol{\theta}). \tag{B.31}$$

For us, each of the d coordinates of $\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P})$ belongs to $\mathcal{T}_1$. Thus, by the isometry between $\mathcal{T}_{1,C_{\boldsymbol{\theta}}} + \mathcal{T}_{3,C_{\boldsymbol{\theta}}}$ and $\mathcal{T}_1 + \mathcal{T}_3$, analogous to (B.31), we conclude that

$$\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P}) - \Pi_{F_{\mathbf{Y},\mathbf{X}}}(\dot{\mathbf{l}}^{\mathrm{s}}(\cdot_1, \cdot_2; \mathrm{P})|\mathcal{T}_3) = \dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\cdot_1 - \mathbf{B}^{\top}\cdot_2); \boldsymbol{\theta}).$$

Finally, (4.10) follows from (B.29) and the above equation.

By (4.10), the efficient information matrix $\mathbf{I}^*(\boldsymbol{\theta})$ from (4.1) can now be written as

$$\begin{aligned}\mathbf{I}^*(\boldsymbol{\theta}) &= \mathbb{E}\left[(\dot{\mathbf{l}}^*_{\boldsymbol{\theta}}\dot{\mathbf{l}}^{*\top}_{\boldsymbol{\theta}})(\mathbf{Y}, \mathbf{X}; \mathrm{P})\right] \\ &= \mathbb{E}\left[(\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}\dot{\mathbf{l}}^{*\mathrm{o}\top}_{\boldsymbol{\theta}})(F_{\bullet}(\mathbf{Y} - \mathbf{B}^{\top}\mathbf{X}); \boldsymbol{\theta})\right] = \mathbb{E}\left[(\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}\dot{\mathbf{l}}^{*\mathrm{o}\top}_{\boldsymbol{\theta}})(\mathbf{U}; \boldsymbol{\theta})\right],\end{aligned}$$

and it corresponds exactly to the quantity $\mathbf{I}^*(\boldsymbol{\theta})$ introduced at the bottom of p. 1920 and explicitly given in (2.20) in [14]. Thus (4.11) follows. □

B.5. *Proof of Proposition 4.2.*

PROOF. All probabilities in this proof are stated under the law P. We show that (4.12), understood as having $\widehat{\boldsymbol{\theta}}_n$ replaced by $\hat{\boldsymbol{\theta}}^{\mathrm{OSE}}_n$, holds. We fix arbitrarily small $\epsilon > 0$. Then, for any constant $B > 0$,

$$\begin{aligned}&\mathbb{P}\left(\left|\sqrt{n}(\hat{\boldsymbol{\theta}}^{\mathrm{OSE}}_n - \boldsymbol{\theta}) - \mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i); \boldsymbol{\theta})\right\}\right| > \epsilon\right) \\ &\le \mathbb{P}\left(\sqrt{n}\|\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| > B\right) \\ &\quad + \sum_{\substack{\boldsymbol{\theta}_n:\boldsymbol{\theta}_n\in n^{-1/2}\mathbb{Z}^d,\\ \sqrt{n}\|\boldsymbol{\theta}_n - \boldsymbol{\theta}\|\le B}} \mathbb{P}\left(\left|\sqrt{n}(\hat{\boldsymbol{\theta}}^{\mathrm{OSE}}_n - \boldsymbol{\theta}) - \mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i); \boldsymbol{\theta})\right\}\right| > \epsilon, \widetilde{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n\right).\end{aligned}$$

By Assumption 4.2, the first term on the right hand side above can be made arbitrarily small by enlarging B. Then, for this fixed, large enough B, the number of terms in the summation in the second term on the right hand side above is bounded from above by a finite constant uniformly over $n \ge 1$. Thus, to establish (4.12), it suffices to show that each term in the summation is $o(1)$. For a single term in the summation we have

$$\begin{aligned}&\mathbb{P}\left(\left|\sqrt{n}(\hat{\boldsymbol{\theta}}^{\mathrm{OSE}}_n - \boldsymbol{\theta}) - \mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i); \boldsymbol{\theta})\right\}\right| > \epsilon, \widetilde{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n\right) \\ &\le \mathbb{P}\Bigg(\Bigg|\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}) + \mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\left\{\frac{\phi_n}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(\widehat{F}^{\mathrm{r}}_{n,1}(\widehat{E}_{i,1}), \ldots, \widehat{F}^{\mathrm{r}}_{n,p}(\widehat{E}_{i,p}); \boldsymbol{\theta}_n)\right\} \\ &\qquad - \mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i); \boldsymbol{\theta})\right\}\Bigg| > \epsilon\Bigg).\end{aligned}$$

By the above inequality, to establish (4.12), it now suffices to show that, for each deterministic sequence $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \mathcal{O}(n^{-1/2})$, we have

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

$$-\mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i);\boldsymbol{\theta})\right\}+\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}) = o_{\mathrm{p}}(1). \tag{B.32}$$

We start from the first term on the left hand side of (B.32). Define the matrices

$$\mathbf{A}_m(\boldsymbol{\theta}') = \frac{1}{2}\{\mathbf{D}_{\boldsymbol{\theta}'}(\mathbf{g}_m(\boldsymbol{\theta}')) - \dot{\mathbf{S}}_m(\boldsymbol{\theta}')\}, \quad \boldsymbol{\theta}' \in \boldsymbol{\Theta}, m \in [d]. \tag{B.33}$$

Note that in terms of $\mathbf{A}_{m,n} \equiv \mathbf{A}_m(\boldsymbol{\theta}_n)$, by the observation made in Remark 3, the first term on the left hand side of (B.32) can be written as

$$\sqrt{n}\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\begin{pmatrix}\mathrm{tr}(\mathbf{A}_{1,n}\widehat{\mathbf{R}}_n)\\ \vdots\\ \mathrm{tr}(\mathbf{A}_{d,n}\widehat{\mathbf{R}}_n)\end{pmatrix} = \sqrt{n}\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\begin{pmatrix}\mathrm{tr}(\mathbf{A}_{1,n}\mathbf{R}_n)\\ \vdots\\ \mathrm{tr}(\mathbf{A}_{d,n}\mathbf{R}_n)\end{pmatrix}$$
$$+\sqrt{n}\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\begin{pmatrix}\mathrm{tr}(\mathbf{A}_{1,n}(\widehat{\mathbf{R}}_n-\mathbf{R}_n))\\ \vdots\\ \mathrm{tr}(\mathbf{A}_{d,n}(\widehat{\mathbf{R}}_n-\mathbf{R}_n))\end{pmatrix}.$$

All components of $\mathbf{I}^{*-1}(\boldsymbol{\theta})$ and $\mathbf{A}_m(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$ (see (4.11) and (B.33) respectively) and hence all components of $\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)$ and $\mathbf{A}_{m,n}$ are $\mathcal{O}_{\mathrm{p}}(1)$. Together with (3.29) in Theorem 3.4, we conclude that the second term on the right hand side above is $o_{\mathrm{p}}(1)$. Thus, instead of (B.32), it now suffices to show that

$$\sqrt{n}\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\begin{pmatrix}\mathrm{tr}(\mathbf{A}_{1,n}\mathbf{R}_n)\\ \vdots\\ \mathrm{tr}(\mathbf{A}_{d,n}\mathbf{R}_n)\end{pmatrix} - \mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i);\boldsymbol{\theta})\right\}$$
$$+\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}) = o_{\mathrm{p}}(1). \tag{B.34}$$

Applying Remark 3 in the opposite direction to re-express $\mathbf{R}_n$ in (B.34) in terms of the efficient score $\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}$ to conclude that (B.34) is equivalent to

$$\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\left\{\frac{\phi_n}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F^{\mathrm{r}}_{n,1}(E_{i,1}),\ldots,F^{\mathrm{r}}_{n,p}(E_{i,p});\boldsymbol{\theta}_n)\right\}$$

$$-\mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F_{\bullet}(\mathbf{E}_i);\boldsymbol{\theta})\right\}+\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}) = o_{\mathrm{p}}(1). \tag{B.35}$$

The second and third terms on the left hand side of (B.35) are obvious $\mathcal{O}_{\mathrm{p}}(1)$, so we are free to get rid of the ϕ_n factor in the first term to conclude that it suffices to show

$$\mathbf{I}^{*-1}(\boldsymbol{\theta}_n)\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}^{*\mathrm{o}}_{\boldsymbol{\theta}}(F^{\mathrm{r}}_{n,1}(E_{i,1}),\ldots,F^{\mathrm{r}}_{n,p}(E_{i,p});\boldsymbol{\theta}_n)\right\}$$

$$-\mathbf{I}^{*-1}(\boldsymbol{\theta})\left\{\frac{1}{\sqrt{n}}\sum_{i\in[n]}\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}(F_{\bullet}(\mathbf{E}_i);\boldsymbol{\theta})\right\}+\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta})=o_{\mathrm{p}}(1). \tag{B.36}$$

Thus we have reduced the problem in (B.32) involving the residual ranks to the one in (B.36) involving the (unobservable) oracle ranks. In fact it is readily checked that Equation (B.36) is identical to the second equation display on page 8 in the supplementary material of [14] (their $\dot{\ell}^*_\theta$, $\mathbf{U}_i=(U_{i1},\ldots,U_{ip})^\top$ and $\widehat{F}_{n,k}(U_{ik})$ correspond to our $\dot{\mathbf{l}}_{\boldsymbol{\theta}}^{*\mathrm{o}}$, $F_{\bullet}(\mathbf{E}_i)$ and $F^{\mathrm{r}}_{n,k}(E_{i,k})$ respectively), and the latter is shown to hold there (under conditions (i) to (iv) in Assumption 4.1). Thus we have finished the proof of the proposition. □

C. Auxiliary lemmas.

LEMMA C.1. *For each $k\in[p]$, we have*

$$\max_{i\in[n]}\max\{|\Phi^{\leftarrow}(\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}))|,|\Phi^{\leftarrow}(F^{\mathrm{r}}_{n,k}(E_{i,k}))|\}\lesssim\log^{1/2}(n). \tag{C.1}$$

PROOF. Note that

$$\min_{i\in[n]}\min\{\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}),F^{\mathrm{r}}_{n,k}(E_{i,k})\}=1/(n+1),$$
$$\max_{i\in[n]}\max\{\widehat{F}^{\mathrm{r}}_{n,k}(\widehat{E}_{i,k}),F^{\mathrm{r}}_{n,k}(E_{i,k})\}=n/(n+1).$$

This fact and (C.3) in Lemma C.3 then yield (C.1). □

LEMMA C.2. *We have*

$$\frac{\mathrm{d}}{\mathrm{d}u}\Phi^{\leftarrow}(u)=\frac{1}{\phi(\Phi^{\leftarrow}(u))}\quad\text{and}\quad\frac{\mathrm{d}^2}{\mathrm{d}^2u}\Phi^{\leftarrow}(u)=\frac{\Phi^{\leftarrow}(u)}{\phi^2(\Phi^{\leftarrow}(u))}.$$

PROOF. The first equation is well known and follows easily from the derivative of an inverse function. For the second equation, we first note that

$$\frac{\mathrm{d}}{\mathrm{d}u}\phi(u)=\frac{\mathrm{d}}{\mathrm{d}u}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}=-u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}=-u\,\phi(u).$$

Then we have

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}^2u}\Phi^{\leftarrow}(u)&=\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\mathrm{d}}{\mathrm{d}u}\Phi^{\leftarrow}(u)\right)=\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{1}{\phi(\Phi^{\leftarrow}(u))}\right)=-\frac{1}{\phi^2(\Phi^{\leftarrow}(u))}\frac{\mathrm{d}}{\mathrm{d}u}\phi(\Phi^{\leftarrow}(u))\\&=-\frac{1}{\phi^2(\Phi^{\leftarrow}(u))}\left(-\Phi^{\leftarrow}(u)\right)\phi(\Phi^{\leftarrow}(u))\frac{\mathrm{d}}{\mathrm{d}u}\Phi^{\leftarrow}(u)=\frac{\Phi^{\leftarrow}(u)}{\phi^2(\Phi^{\leftarrow}(u))}.\end{aligned}$$ □

LEMMA C.3 (Bounds involving the standard normal quantile function).

$$\sup_{u\in(0,1)}\frac{u\wedge(1-u)}{\phi(\Phi^{\leftarrow}(u))}\leq\infty, \tag{C.2}$$

$$\sup_{u\in(0,1)}\frac{|\Phi^{\leftarrow}(u)|}{\sqrt{2\log\left(\frac{1}{2(u\wedge(1-u))}\right)}}\leq 1. \tag{C.3}$$

PROOF. Inequality (C.2) is well-known and follows by, e.g., Inequality (10) in [10]. Next, by Corollary 1 in [4], we have $\Phi(x) \le (1/2)\exp(-x^2/2)$, $\forall x \le 0$. Inequality (C.3) then follows by inverting this inequality, and symmetry around $u = 1/2$. □

For the next lemma, let F_n^{U} be the empirical distribution function of independent Unif(0, 1) random variables $\xi_1, \dots, \xi_n$, and let $\alpha_n \equiv \sqrt{n}(F_n^{\mathrm{U}} - I)$ be the standard empirical process. For $0 \le \kappa \le 1/2$, let the function $q_\kappa : (0,1) \to \mathbb{R}$ be defined as $q_\kappa(t) = (t \wedge (1-t))^\kappa$.

LEMMA C.4 (Theorem 4.2.1 of [5]). *We have*

$$\sup_{t \in (1/n, 1-1/n)} \left| \frac{\alpha_n(t)}{q_{1/2}(t)} \right| = \mathcal{O}_{\mathrm{p}}\left(\sqrt{\log\log(n)}\right). \tag{C.4}$$

PROOF. The lemma is well known and is a straightforward consequence of Theorem 4.2.2 in [5], by choosing the (EFKP upper-class) function w in that theorem to be

$$w(t) = q_{1/2}(t)\sqrt{\log\log\left(\frac{1}{t \wedge (1-t)}\right)},$$

and a few additional algebraic steps. □

For the next lemma, let the function w_n be the modulus of continuity of the standard empirical process α_n (where α_n is introduced above Lemma C.4). Define the function

$$\psi(\lambda) = \frac{2[(1+\lambda)\{\log(1+\lambda) - 1\} + 1]}{\lambda^2}.$$

This is exactly the ψ function introduced in Section 11.1 in [15], and we refer the readers to Proposition 1 in the same section for some of its properties.

LEMMA C.5 (Modulus of continuity of empirical processes; see for instance Inequality 1 in Section 14.2 in [15]). *Let $0 < a \le \delta \le 1/2$. Then for all $\lambda > 0$,*

$$\mathbb{P}(w_n(a) \ge \lambda\sqrt{a}) \le \frac{20}{a\delta^3}\exp\left(-(1-\delta)^4 \frac{\lambda^2}{2}\psi\left(\frac{\lambda}{\sqrt{na}}\right)\right).$$

D. Detailed calculations for differentiability in quadratic mean. We first establish $(\partial/\partial\eta) l(\mathbf{y}, \mathbf{x}; \mathrm{P}_\eta)$ as in (B.18). Starting from (B.16), and using the expression (B.1) for $l^{\mathrm{c}}(\cdot;\cdot)$, we have, when $p_\eta(\mathbf{y}, \mathbf{x}) \ne 0$ (so necessarily $y_k - \mathbf{B}_{k,\eta}^\top \mathbf{x} \in J_k \equiv (a_k, b_k)$),

$$\begin{aligned}
\frac{\partial}{\partial\eta} l(\mathbf{y}, \mathbf{x}; \mathrm{P}_\eta) = \frac{\partial}{\partial\eta}\Bigg\{ & -\frac{1}{2}\log(\det \mathbf{R}(\boldsymbol{\theta}_\eta)) \\
& - \frac{1}{2}\Phi_\bullet^{\leftarrow}\left(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{y} - \mathbf{B}_\eta^\top \mathbf{x})\right)^\top (\mathbf{S}(\boldsymbol{\theta}_\eta) - \mathbf{I}_p)\Phi_\bullet^{\leftarrow}\left(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{y} - \mathbf{B}_\eta^\top \mathbf{x})\right) \\
& + \sum_{k \in [p]} \log\left(f_{k,\eta}^{h_k}(y_k - \mathbf{B}_{k,\eta}^\top \mathbf{x})\right) + \log f_{\mathbf{X},\eta}^{h}(\mathbf{x})\Bigg\} \\
= \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}&(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{y} - \mathbf{B}_\eta^\top \mathbf{x}); \boldsymbol{\theta}_\eta)
\end{aligned}$$

$$+\sum_{k\in[p]} \dot{l}_k^{\mathrm{c,m}}\left(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{y}-\mathbf{B}_\eta^\top\mathbf{x});\boldsymbol{\theta}_\eta\right)\left\{\frac{\partial}{\partial\eta}F^{h_k}_{k,\eta}\left(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}\right)\right\}$$

$$+\sum_{k\in[p]}\frac{1}{f^{h_k}_{k,\eta}(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})}\left\{\frac{\partial}{\partial\eta}f^{h_k}_{k,\eta}(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})\right\}+\frac{1}{f^h_{\mathbf{X},\eta}(\mathbf{x})}\frac{\partial}{\partial\eta}f^h_{\mathbf{X},\eta}(\mathbf{x}). \tag{D.1}$$

We treat the remaining derivatives in (D.1) one by one. First, using differentiation under the integral sign and (B.7) give

$$\begin{aligned}\frac{\partial}{\partial\eta}F^{h_k}_{k,\eta}\left(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}\right)&=\frac{\partial}{\partial\eta}\int_{a_k}^{y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}} f^{h_k}_{k,\eta}(t)\mathrm{d}t\\&=-\boldsymbol{\beta}_k^\top\mathbf{x}\, f^{h_k}_{k,\eta}(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})+\int_{a_k}^{y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}} h_k\circ F_k(t)\, f_k(t)\mathrm{d}t\\&=-\boldsymbol{\beta}_k^\top\mathbf{x}\, f^{h_k}_{k,\eta}(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})+\int_0^{F_k\left(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}\right)} h_k(u)\mathrm{d}u.\end{aligned} \tag{D.2}$$

Next, first using (B.7) and then the differentiability of $h_k\circ F_k$ (due to the restriction $h_k\in L_2^{0,\mathrm{d}}[0,1]$) give

$$\begin{aligned}&\frac{\partial}{\partial\eta}f^{h_k}_{k,\eta}(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})\\&=\frac{\partial}{\partial\eta}\left\{(1+\eta\, h_k\circ F_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}))\, f_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})\right\}\\&=h_k\circ F_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})\, f_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})-\boldsymbol{\beta}_k^\top\mathbf{x}\,\eta\,\dot{h}_k\circ F_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})\\&\quad\times f_k^2(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x})-\boldsymbol{\beta}_k^\top\mathbf{x}(1+\eta\, h_k\circ F_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}))\dot{f}_k(y_k-\mathbf{B}_{k,\eta}^\top\mathbf{x}).\end{aligned} \tag{D.3}$$

Next, for the last term in (D.1), by differentiation under the integral sign,

$$\begin{aligned}\frac{\partial}{\partial\eta}\psi(\eta;f_{\mathbf{X}},h)&=-\psi(\eta;f_{\mathbf{X}},h)^2\int_{\mathbb{R}}\Psi^2(\eta h(\mathbf{x}))\exp(-2\eta h(\mathbf{x}))h(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}\\&\equiv\dot{\psi}(\eta;f_{\mathbf{X}},h);\end{aligned} \tag{D.4}$$

then, dominated convergence theorem (DCT) yields the first equation on p. 53 in [2]:

$$\lim_{\eta\to0}\dot{\psi}(\eta;f_{\mathbf{X}},h)=\int_{\mathbb{R}}h(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}=0.$$

Then, from the above and (B.9), we have

$$\begin{aligned}\frac{\partial}{\partial\eta}f^h_{\mathbf{X},\eta}(\mathbf{x})&=\left\{\frac{\partial}{\partial\eta}\psi(\eta;f_{\mathbf{X}},h)\right\}\Psi(\eta h(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x})+\psi(\eta;f_{\mathbf{X}},h)\left\{\frac{\partial}{\partial\eta}\Psi(\eta h(\mathbf{x}))\right\}f_{\mathbf{X}}(\mathbf{x})\\&=\dot{\psi}(\eta;f_{\mathbf{X}},h)\Psi(\eta h(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x})+\psi(\eta;f_{\mathbf{X}},h)\Psi^2(\eta h(\mathbf{x}))\exp(-2\eta h(\mathbf{x}))h(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\\&=\left\{\dot{\psi}(\eta;f_{\mathbf{X}},h)\psi(\eta;f_{\mathbf{X}},h)^{-1}\right\}f^h_{\mathbf{X},\eta}(\mathbf{x})+\Psi(\eta h(\mathbf{x}))\exp(-2\eta h(\mathbf{x}))h(\mathbf{x})f^h_{\mathbf{X},\eta}(\mathbf{x}).\end{aligned} \tag{D.5}$$

Plugging (D.2), (D.3) and (D.5) into (D.1) yields (B.18).

Next we establish the continuity of I_η as in (B.15). To make the distributional dependence on η more explicit, we shall denote the random triplet $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$ under the law P_η by $(\mathbf{Y}_\eta, \mathbf{X}_\eta, \mathbf{E}_\eta)$ and so in particular

$$\mathbf{E}_\eta = (E_{1,\eta}, \dots, E_{p,\eta})^\top = \mathbf{Y}_\eta - \mathbf{B}_\eta^\top \mathbf{X}_\eta, \tag{D.6}$$

while $(\mathbf{Y}, \mathbf{X}, \mathbf{E})$ is reserved when this triplet follows the law P. Note that $p_\eta(\cdot_1, \cdot_2) \to p_0(\cdot_1, \cdot_2)$ pointwise. Thus, by Scheffé's lemma and the Portemanteau theorem (e.g., Theorems 16.11 and 29.1 in [3] respectively), $(\mathbf{Y}_\eta, \mathbf{X}_\eta) \rightsquigarrow (\mathbf{Y}, \mathbf{X})$. By Skorokhod representation theorem (e.g., Theorem 29.6 in [3]), we can and will choose $(\mathbf{Y}_\eta, \mathbf{X}_\eta)$, $|\eta| < \epsilon$ with $(\mathbf{Y}_0, \mathbf{X}_0) = (\mathbf{Y}, \mathbf{X})$ so that the $(\mathbf{Y}_\eta, \mathbf{X}_\eta)$'s are all defined on a common probability space and $(\mathbf{Y}_\eta, \mathbf{X}_\eta) \xrightarrow{a.s.} (\mathbf{Y}, \mathbf{X})$, so $\mathbf{E}_\eta \xrightarrow{a.s.} \mathbf{E}$ as well.

For $k \in [p]$, $\ell \in \{1, 2, 3\}$, $\ell' \in \{1, 2\}$, define the functions $B_{k,\ell}(\cdot_1, \cdot_2), C_{k,\ell'}(\cdot_1, \cdot_2) : \mathbb{R}^p \times (-\varepsilon, \varepsilon) \to \mathbb{R}$, and also the function $D(\cdot_1, \cdot_2) : \mathbb{R}^q \times (-\varepsilon, \varepsilon) \to \mathbb{R}$ as: for $\mathbf{z} = (z_1, \dots, z_p)^\top \in \mathbb{R}^p$, $\mathbf{x} \in \mathbb{R}^q$ and $\eta \in (-\varepsilon, \varepsilon)$,

$$\begin{aligned}
B_{k,1}(\mathbf{z}, \eta) &= \dot{l}_k^{\mathrm{c,m}} \left(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{z}); \boldsymbol{\theta}_\eta \right) f_{k,\eta}^{h_k}(z_k), \quad B_{k,2}(\mathbf{z}, \eta) = \eta\, \dot{h}_k \circ F_k(z_k) \frac{f_k^2}{f_{k,\eta}^{h_k}}(z_k), \\
B_{k,3}(\mathbf{z}, \eta) &= (1 + \eta\, h_k \circ F_k(z_k)) \frac{\dot{f}_k}{f_{k,\eta}^{h_k}}(z_k), \\
C_{k,1}(\mathbf{z}, \eta) &= h_k \circ F_k\,(z_k)\, \frac{f_k}{f_{k,\eta}^{h_k}}(z_k), \quad C_{k,2}(\mathbf{z}, \eta) = \dot{l}_k^{\mathrm{c,m}} \left(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{z}); \boldsymbol{\theta}_\eta \right) \int_0^{F_k(z_k)} h_k(u) \mathrm{d}u, \\
D(\mathbf{x}, \eta) &= \dot{\psi}(\eta; f_{\mathbf{X}}, h) \psi(\eta; f_{\mathbf{X}}, h)^{-1} + \Psi(\eta h(\mathbf{x})) \exp(-2\eta h(\mathbf{x})) h(\mathbf{x}).
\end{aligned}$$

Then we can rewrite (B.18) as, when $p_\eta(\mathbf{y}, \mathbf{x}) \neq 0$,

$$\begin{aligned}
\frac{\partial}{\partial \eta} l(\mathbf{y}, \mathbf{x}; \mathrm{P}_\eta) = \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{y} - \mathbf{B}_\eta^\top \mathbf{x}); \boldsymbol{\theta}_\eta) - \sum_{k \in [p]} \boldsymbol{\beta}_k^\top \mathbf{x} \sum_{\ell \in \{1,2,3\}} B_{k,\ell}(\mathbf{y} - \mathbf{B}_\eta^\top \mathbf{x}, \eta) \\
+ \sum_{k \in [p]} \sum_{\ell \in \{1,2\}} C_{k,\ell}(\mathbf{y} - \mathbf{B}_\eta^\top \mathbf{x}, \eta) + D(\mathbf{x}, \eta).
\end{aligned}$$

Using the above expression and (D.6) in (B.17), we have

$$\begin{aligned}
I_\eta &= \mathbb{E}\left[\left\{ \frac{\partial}{\partial \eta} l(\mathbf{Y}_\eta, \mathbf{X}_\eta; \mathrm{P}_\eta) \right\}^2 \right] \\
&= \mathbb{E}\Bigg[\Bigg\{ \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{E}_\eta); \boldsymbol{\theta}_\eta) - \sum_{k \in [p]} \boldsymbol{\beta}_k^\top \mathbf{X}_\eta \sum_{\ell \in \{1,2,3\}} B_{k,\ell}(\mathbf{E}_\eta, \eta) \\
&\qquad + \sum_{k \in [p]} \sum_{\ell \in \{1,2\}} C_{k,\ell}(\mathbf{E}_\eta, \eta) + D(\mathbf{X}_\eta, \eta) \Bigg\}^2 \Bigg].
\end{aligned} \tag{D.7}$$

We first treat the squared terms (after expanding the square of the curly bracket) in (D.7). We start with $\mathbb{E}\left[\left\{ \boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{E}_\eta); \boldsymbol{\theta}_\eta) \right\}^2 \right]$. Without the scaling by $\boldsymbol{\alpha}$, this corresponds exactly to the information matrix $\mathbf{I}(\boldsymbol{\theta}_\eta)$ for the parametric Gaussian copula model with known margins and copula correlation matrix $\mathbf{R}(\boldsymbol{\theta}_\eta)$. Then, following (2.2) to (2.4) in [14], we have

$$\begin{aligned}\mathbb{E}\left[\left\{\boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{E}_\eta);\boldsymbol{\theta}_\eta)\right\}^2\right] &= \boldsymbol{\alpha}^\top \mathbb{E}\left[\dot{\mathbf{l}}^{\mathrm{c,s}}(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{E}_\eta);\boldsymbol{\theta}_\eta)\dot{\mathbf{l}}^{\mathrm{c,s}}(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{E}_\eta);\boldsymbol{\theta}_\eta)^\top\right]\boldsymbol{\alpha}\\ &= \boldsymbol{\alpha}^\top \frac{1}{2}\left[\mathrm{tr}\left(\dot{\mathbf{S}}_m(\boldsymbol{\theta}_\eta)\mathbf{R}(\boldsymbol{\theta}_\eta)\dot{\mathbf{S}}_{m'}(\boldsymbol{\theta}_\eta)\mathbf{R}(\boldsymbol{\theta}_\eta)\right)\right]_{m,m'\in[d]}\boldsymbol{\alpha}\\ &\to \boldsymbol{\alpha}^\top \frac{1}{2}\left[\mathrm{tr}\left(\dot{\mathbf{S}}_m(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\dot{\mathbf{S}}_{m'}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})\right)\right]_{m,m'\in[d]}\boldsymbol{\alpha}\\ &= \mathbb{E}\left[\left\{\boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{E});\boldsymbol{\theta})\right\}^2\right] < \infty.\end{aligned}$$

Here the convergence step follows by the continuities of $\mathbf{R}(\cdot)$ and $\dot{\mathbf{S}}_m(\cdot)$.

Next we treat the terms $\mathbb{E}\left[\left\{\boldsymbol{\beta}_k^\top \mathbf{X}_\eta B_{k,\ell}(\mathbf{E}_\eta,\eta)\right\}^2\right]$, $k \in [p]$, $\ell \in \{1,2,3\}$. First,

$$\begin{aligned}\mathbb{E}\left[\mathbf{X}_\eta\mathbf{X}_\eta^\top\right] &= \int_{\mathbb{R}^q} \mathbf{x}\mathbf{x}^\top f^h_{\mathbf{X},\eta}(\mathbf{x})\mathrm{d}\mathbf{x}\\ &= \int_{\mathbb{R}^q} \mathbf{x}\mathbf{x}^\top \psi(\eta; f_{\mathbf{X}}, h)\Psi(\eta h(\mathbf{x}))f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x} \to \mathbb{E}[\mathbf{X}\mathbf{X}^\top],\end{aligned}$$

where the convergence step follows by the DCT. Next,

$$\begin{aligned}\mathbb{E}\left[B^2_{k,1}(\mathbf{E}_\eta,\eta)\right] &= \mathbb{E}\left[\dot{l}^{\mathrm{c,m}}_k\left(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{E}_\eta);\boldsymbol{\theta}_\eta\right)^2 f^{h_k}_{k,\eta}(E_{k,\eta})^2\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\dot{l}^{\mathrm{c,m}}_k\left(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{E}_\eta);\boldsymbol{\theta}_\eta\right)^2 \middle| E_{k,\eta}\right] f^{h_k}_{k,\eta}(E_{k,\eta})^2\right]\\ &= \mathbb{E}\left[\frac{(\mathbf{S}(\boldsymbol{\theta}_\eta))_{kk}-1}{\phi^2(\Phi^{\leftarrow}(F^{h_k}_{k,\eta}(E_{k,\eta})))} f^{h_k}_{k,\eta}(E_{k,\eta})^2\right]\\ &= \{(\mathbf{S}(\boldsymbol{\theta}_\eta))_{kk}-1\}\,\mathbb{E}\left[\frac{1}{\phi^2(\Phi^{\leftarrow}(F^{h_k}_{k,\eta}(E_{k,\eta})))} f^{h_k}_{k,\eta}(E_{k,\eta})^2\right],\end{aligned}\tag{D.8}$$

where the third step follows by (2.9) in [14]. Now, with $F^{h_k\leftarrow}_{k,\eta}$ denoting the left-continuous inverse of $F^{h_k}_{k,\eta}$, and using the facts that $E_{k,\eta} = F^{h_k\leftarrow}_{k,\eta} \circ F^{h_k}_{k,\eta}(E_{k,\eta})$ with probability one and $F^{h_k}_{k,\eta}(E_{k,\eta}) \sim \mathrm{Unif}(0,1)$,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\phi^2(\Phi^{\leftarrow}(F^{h_k}_{k,\eta}(E_{k,\eta})))} f^{h_k}_{k,\eta}(E_{k,\eta})^2\right] &= \mathbb{E}\left[\frac{1}{\phi^2(\Phi^{\leftarrow}(F^{h_k}_{k,\eta}(E_{k,\eta})))}\left\{f^{h_k}_{k,\eta}\circ F^{h_k\leftarrow}_{k,\eta}\circ F^{h_k}_{k,\eta}(E_{k,\eta})\right\}^2\right]\\ &= \int_{(0,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))}\left\{f^{h_k}_{k,\eta}\circ F^{h_k\leftarrow}_{k,\eta}(u)\right\}^2 \mathrm{d}u.\end{aligned}\tag{D.9}$$

Under (vi) in Assumption 4.1, for all η small enough, both F_k and $F^{h_k}_{k,\eta}$ are strictly increasing on J_k, and we can define a strictly increasing function $v = v_\eta : (0,1) \to (0,1)$ such that

$$F^{h_k\leftarrow}_{k,\eta}(u) = F^{\leftarrow}_k \circ v(u), \quad \forall u \in (0,1). \tag{D.10}$$

Now we divide the integral in (D.9) into two parts over the intervals $(0, 1/2]$ and $(1/2, 1)$ respectively. We first consider the first interval. Using (B.7) and the boundedness of $h_k \in L_2^{\mathrm{d}}[0,1]$, we can define a function $c_\eta : \mathbb{R} \to \mathbb{R}$ satisfying

$$F_{k,\eta}^{h_k}(t) = \frac{1}{1 + c_\eta(t)} F_k(t), \quad \forall t \in \mathbb{R}, \tag{D.11}$$

and it is easy to show that $\|c_\eta\|_{L_\infty} \to 0$ (as $|\eta| \to 0$). Applying $F_{k,\eta}^{h_k}$ on both sides of (D.10) and then invoking (D.11) on the right hand side yield

$$v(u) = \{1 + c_\eta \circ F_k^{\leftarrow} \circ v(u)\} u. \tag{D.12}$$

Then, consecutively using (B.7) and the boundedness of h_k, (D.10) and (D.12), the square root of the numerator of the integrand in (D.9) is bounded by a constant multiple of

$$\begin{aligned} f_k \circ F_{k,\eta}^{h_k \leftarrow}(u) &= f_k \circ F_k^{\leftarrow} \circ v(u) = f_k \circ F_k^{\leftarrow}\left(\{1 + c_\eta \circ F_k^{\leftarrow} \circ v(u)\} u\right) \\ &\le \sup_{\delta : |\delta| \le \|c_\eta\|_{L_\infty}} f_k \circ F_k^{\leftarrow}\left((1+\delta) u\right). \end{aligned}$$

Then, continuing from (D.9), by the pointwise convergence of $f_{k,\eta}^{h_k} \circ F_{k,\eta}^{h_k \leftarrow}$ to $f_k \circ F_k^{\leftarrow}$, the above inequality, condition (4.9) in Assumption 4.1 and the DCT,

$$\begin{aligned} &\int_{(0,1/2]} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ f_{k,\eta}^{h_k} \circ F_{k,\eta}^{h_k \leftarrow}(u) \right\}^2 \mathrm{d}u \\ &\to \int_{(0,1/2]} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ f_k \circ F_k^{\leftarrow}(u) \right\}^2 \mathrm{d}u < \infty. \end{aligned} \tag{D.13}$$

Analogously, for the integral in (D.9) over the interval $(1/2, 1)$,

$$\begin{aligned} &\int_{(1/2,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ f_{k,\eta}^{h_k} \circ F_{k,\eta}^{h_k \leftarrow}(u) \right\}^2 \mathrm{d}u \\ &\to \int_{(1/2,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ f_k \circ F_k^{\leftarrow}(u) \right\}^2 \mathrm{d}u < \infty. \end{aligned} \tag{D.14}$$

Then, by (D.8), (D.9), (D.13), (D.14) and the continuity of $\mathbf{S}(\cdot)$, we conclude that

$$\mathbb{E}\left[B_{k,1}^2(\mathbf{E}_\eta, \eta)\right] \to \mathbb{E}\left[B_{k,1}^2(\mathbf{E}, 0)\right] < \infty.$$

Next,

$$\begin{aligned} \mathbb{E}\left[B_{k,2}^2(\mathbf{E}_\eta, \eta)\right] &= \eta^2 \mathbb{E}\left[\left\{ \dot{h}_k \circ F_k(E_{k,\eta}) \frac{f_k^2}{f_{k,\eta}^{h_k}}(E_{k,\eta}) \right\}^2\right] \\ &= \eta^2 \int_{J_k} \left\{ \dot{h}_k \circ F_k(t) \frac{f_k^2}{f_{k,\eta}^{h_k}}(t) \right\}^2 f_{k,\eta}^{h_k}(t) \mathrm{d}t \\ &= \eta^2 \int_{J_k} \left\{ \dot{h}_k \circ F_k(t) \right\}^2 \frac{f_k^4}{f_{k,\eta}^{h_k}}(t) \mathrm{d}t. \end{aligned}$$

By the boundedness of h_k, from (B.7) it is easy to see that for all η small enough, $\|f_k/f_{k,\eta}^{h_k}\|_{L_\infty} < 2$ (say). In addition, $\|\dot{h}_k \circ F_k\|_{L_\infty} < \infty$ because h_k is continuously differentiable. Thus, for all η small enough, the integrand in the integral above is bounded by a constant multiple of f_k^3 uniformly over J_k. By the boundedness of f_k on J_k, $\int_{J_k} f_k^3(t)\mathrm{d}t \lesssim \int_{J_k} f_k(t)\mathrm{d}t = 1$. Then, by the DCT, we conclude that

$$\mathbb{E}\left[B_{k,2}^2(\mathbf{E}_\eta, \eta)\right] \to 0 = \mathbb{E}\left[B_{k,2}^2(\mathbf{E}, 0)\right].$$

Next,

$$\begin{aligned}\mathbb{E}\left[B_{k,3}^2(\mathbf{E}_\eta, \eta)\right] &= \mathbb{E}\left[\left\{(1+\eta\, h_k \circ F_k(E_{k,\eta}))\frac{\dot{f}_k}{f_{k,\eta}^{h_k}}(E_{k,\eta})\right\}^2\right] \\ &= \int_{J_k} \{1+\eta\, h_k\circ F_k(t)\}^2\,\frac{\dot{f}_k^2}{f_{k,\eta}^{h_k}}(t)\mathrm{d}t \\ &\to \int_{J_k} \frac{\dot{f}_k^2}{f_k}(t)\mathrm{d}t = \mathbb{E}\left[B_{k,3}^2(\mathbf{E}, 0)\right] < \infty,\end{aligned}$$

where the convergence step follows by the boundedness of h_k, condition (4.8) in Assumption 4.1 and the DCT. By the independence between $\mathbf{X}_\eta$ and $\mathbf{E}_\eta$ and the separate convergences of $\mathbb{E}[\mathbf{X}_\eta\mathbf{X}_\eta^\top]$ and $\mathbb{E}\left[B_{k,\ell}^3(\mathbf{E}_\eta, \eta)\right]$, we conclude that, for all $k \in [p]$ and $\ell \in \{1,2,3\}$,

$$\mathbb{E}\left[\left\{\boldsymbol{\beta}_k^\top \mathbf{X}_\eta B_{k,\ell}(\mathbf{E}_\eta, \eta)\right\}^2\right] \to \mathbb{E}\left[\left\{\boldsymbol{\beta}_k^\top \mathbf{X} B_{k,\ell}(\mathbf{E}, 0)\right\}^2\right] < \infty.$$

Next we treat the terms $\mathbb{E}\left[C_{k,\ell}^2(\mathbf{E}_\eta, \eta)\right]$, $k \in [p]$, $\ell \in \{1, 2\}$. First, by the DCT,

$$\begin{aligned}\mathbb{E}\left[C_{k,1}^2(\mathbf{E}_\eta, \eta)\right] &= \mathbb{E}\left[\left\{h_k \circ F_k(E_{k,\eta})\frac{f_k}{f_{k,\eta}^{h_k}}(E_{k,\eta})\right\}^2\right] \\ &= \int_{J_k} \{h_k \circ F_k(t)\}^2\,\frac{f_k^2}{f_{k,\eta}^{h_k}}(t)\mathrm{d}t \\ &\to \|h\|_{L_2[0,1]}^2 = \mathbb{E}\left[C_{k,1}^2(\mathbf{E}, 0)\right] < \infty.\end{aligned}$$

Next, using a derivation similar to that of (D.8), the facts that $F_{k,\eta}^{h_k}(E_{k,\eta}) \sim \mathrm{Unif}(0,1)$ and $F_k \circ F_{k,\eta}^{h_k\leftarrow}(u) = v(u)$ (for $v(u)$ in (D.10)), we have

$$\begin{aligned}\mathbb{E}\left[C_{k,2}^2(\mathbf{E}_\eta, \eta)\right] &= \mathbb{E}\left[\left\{\dot{l}_k^{\mathrm{c,m}}\left(F_{\bullet,\eta}^{\mathbf{h}}(\mathbf{E}_\eta); \boldsymbol{\theta}_\eta\right)\int_0^{F_k(E_{k,\eta})} h_k(\lambda)\mathrm{d}\lambda\right\}^2\right] \\ &= \mathbb{E}\left[\frac{(\mathbf{S}(\boldsymbol{\theta}_\eta))_{kk} - 1}{\phi^2(\Phi^{\leftarrow}(F_{k,\eta}^{h_k}(E_{k,\eta})))}\left\{\int_0^{F_k(E_{k,\eta})} h_k(\lambda)\mathrm{d}\lambda\right\}^2\right]\end{aligned}$$

$$
\begin{aligned}
&= \{(\mathbf{S}(\boldsymbol{\theta}_\eta))_{kk} - 1\} \int_{(0,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_0^{F_k \circ F_{k,\eta}^{h_k \leftarrow}(u)} h_k(\lambda) \mathrm{d}\lambda \right\}^2 \mathrm{d}u \\
&= \{(\mathbf{S}(\boldsymbol{\theta}_\eta))_{kk} - 1\} \left(C_{k,2,1} + 2C_{k,2,2,1} + 2C_{k,2,2,2} + C_{k,2,3} \right),
\end{aligned}
$$

where

$$
\begin{aligned}
C_{k,2,1} &= \int_{(0,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_0^{u} h_k(\lambda) \mathrm{d}\lambda \right\}^2 \mathrm{d}u, \\
C_{k,2,2,1} &= \int_{(0,1/2]} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_0^{u} h_k(\lambda) \mathrm{d}\lambda \right\} \left\{ \int_u^{v(u)} h_k(\lambda) \mathrm{d}\lambda \right\} \mathrm{d}u, \\
C_{k,2,2,2} &= \int_{(1/2,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_0^{u} h_k(\lambda) \mathrm{d}\lambda \right\} \left\{ \int_u^{v(u)} h_k(\lambda) \mathrm{d}\lambda \right\} \mathrm{d}u, \\
C_{k,2,3} &= \int_{(0,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_u^{v(u)} h_k(\lambda) \mathrm{d}\lambda \right\}^2 \mathrm{d}u.
\end{aligned}
$$

Following the two-sided Hardy's inequality (e.g., Proposition A.1 on p. 1 in the supplementary material of [14]) and (C.2) in Lemma C.3, $C_{k,2,1} < \infty$. Next, using the same lemma, the boundedness of h_k, (D.12) to calculate $u - v(u)$, and finally the fact that $\|c_\eta\|_{L_\infty} \to 0$, we have

$$
\begin{aligned}
|C_{k,2,2,1}| &= \left| \int_{(0,1/2]} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_0^{u} h_k(\lambda) \mathrm{d}\lambda \right\} \left\{ \int_u^{v(u)} h_k(\lambda) \mathrm{d}\lambda \right\} \mathrm{d}u \right| \\
&\lesssim \int_{(0,1/2]} \frac{1}{u^2} u |c_\eta \circ F_k^{\leftarrow} \circ v(u)| u \,\mathrm{d}u \to 0.
\end{aligned}
$$

Analogous to the above, $C_{k,2,2,2}, C_{k,2,3} \to 0$.

Collecting terms, and invoking the continuity of $\mathbf{S}(\cdot)$, we conclude that

$$
\begin{aligned}
\mathbb{E}\left[C_{k,2}^2(\mathbf{E}_\eta, \eta) \right] &\to \{(\mathbf{S}(\boldsymbol{\theta}))_{kk} - 1\} \int_{(0,1)} \frac{1}{\phi^2(\Phi^{\leftarrow}(u))} \left\{ \int_0^{u} h_k(\lambda) \mathrm{d}\lambda \right\}^2 \mathrm{d}u \\
&= \mathbb{E}\left[C_{k,2}^2(\mathbf{E}, 0) \right] < \infty.
\end{aligned}
$$

Finally we treat the term $\mathbb{E}\left[D^2(\mathbf{X}_\eta, \eta) \right]$. By the DCT (precisely, with the integrand below bounded by a constant multiple of $(1 \vee h(\mathbf{x}))^2 f_{\mathbf{X}}(\mathbf{x})$ which is integrable because $h \in L_2^0(F_{\mathbf{X}})$, and by the pointwise convergence of the integrand to $h(\mathbf{x})^2 f_{\mathbf{X}}(\mathbf{x})$),

$$
\begin{aligned}
\mathbb{E}\left[D^2(\mathbf{X}_\eta, \eta) \right] &= \int_{\mathbb{R}^q} \left\{ \dot{\psi}(\eta; f_{\mathbf{X}}, h) \psi(\eta; f_{\mathbf{X}}, h)^{-1} + \Psi(\eta h(\mathbf{x})) \exp(-2\eta h(\mathbf{x})) h(\mathbf{x}) \right\}^2 f_{\mathbf{X},\eta}^h(\mathbf{x}) \mathrm{d}\mathbf{x} \\
&\to \int_{\mathbb{R}^q} h^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x} = \|h\|_{L_2(F_{\mathbf{X}})}^2 < \infty. \qquad \text{(D.15)}
\end{aligned}
$$

We have checked that the expectations of all the squared terms in (D.7) are continuous at $\eta = 0$, and so we only need to verify that the expectations of all the cross terms in

(D.7) are continuous at $\eta = 0$ as well. We first discuss the cross terms involving the random variable $D(\mathbf{X}_\eta, \eta)$. The expectations of these cross terms will all converge to their expectations at $\eta = 0$ by the independence of $\mathbf{E}_\eta$ and $\mathbf{X}_\eta$, an argument similar to (D.15) to treat the terms involving $\mathbf{X}_\eta$, and Pratt's lemma to treat the remaining terms involving $\mathbf{E}_\eta$. For the last part involving Pratt's lemma, we refer to a similar argument in the next paragraph; we omit the details.

For all other cross terms (i.e., those not involving the random variable $D(\mathbf{X}_\eta, \eta)$), note that all functions $\boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{z}); \boldsymbol{\theta}_\eta)$, $B_{k,\ell}(\mathbf{z}, \eta)$, $C_{k,\ell}(\mathbf{z}, \eta)$ are jointly continuous in $\mathbf{z}$ and η. Then, the convergence $(\mathbf{X}_\eta, \mathbf{E}_\eta) \xrightarrow{a.s.} (\mathbf{X}, \mathbf{E})$ and the continuity of the functions involved imply that the random variables $\boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F^{\mathbf{h}}_{\bullet,\eta}(\mathbf{E}_\eta); \boldsymbol{\theta}_\eta)$, $B_{k,\ell}(\mathbf{E}_\eta, \eta)$, $C_{k,\ell}(\mathbf{E}_\eta, \eta)$ converge almost surely to their respective limits $\boldsymbol{\alpha}^\top \dot{\mathbf{l}}^{\mathrm{c,s}}(F_\bullet(\mathbf{E}); \boldsymbol{\theta})$, $B_{k,\ell}(\mathbf{E}, 0)$, $C_{k,\ell}(\mathbf{E}, 0)$. Then, using the inequality $|ab| \le (a^2 + b^2)/2$ to bound the cross terms by the squared terms, the convergences of the expectations of the cross terms to their respective limits at $\eta = 0$ follows by Pratt's lemma (e.g., p. 164 in [13]) and the already-established convergences of the expectations of the squared terms. At last, we conclude that (B.15) regarding the continuity of I_η holds.

E. Additional simulation results. In this section we present some additional simulation results that were left out from Section 5 in the maintext due to space constraint.

E.1. *Additional results for Section 5.1.* First, for $q = 2$ and $q = 10$, the particular $\mathbf{B}$ generated are respectively

$$\mathbf{B} = \begin{pmatrix} 0.0796 & -1.0101 & -2.0212 \\ 0.8551 & -0.3415 & -0.8337 \end{pmatrix} \quad (\text{for } q = 2),$$

$$\mathbf{B} = \begin{pmatrix} -0.6796 & -0.8770 & -2.0306 \\ -0.9054 & 0.8499 & -1.0389 \\ -0.3481 & 1.3291 & -0.0503 \\ 0.8612 & -0.4938 & -0.7351 \\ 2.2292 & -0.0226 & -1.8954 \\ 0.2748 & 1.4955 & 0.6093 \\ -0.1586 & -0.2656 & -0.4795 \\ -1.4194 & 0.5680 & 0.7511 \\ 0.8065 & -0.3997 & -0.9803 \\ 0.5926 & -0.0021 & -2.2374 \end{pmatrix} \quad (\text{for } q = 10).$$

Next, Figure 1 in the supplement presents the simulation results for highly correlated covariates with $\rho = 0.9$.

E.2. *Additional results for Section 5.2.* First, for $q = 2$ and $q = 10$, the particular $\mathbf{B}$ generated are respectively

$$\mathbf{B} = \begin{pmatrix} 3.5784 & -1.3499 & 0.7254 & 0.7147 \\ 2.7694 & 3.0349 & -0.0631 & -0.2050 \end{pmatrix} \quad (\text{for } q = 2),$$

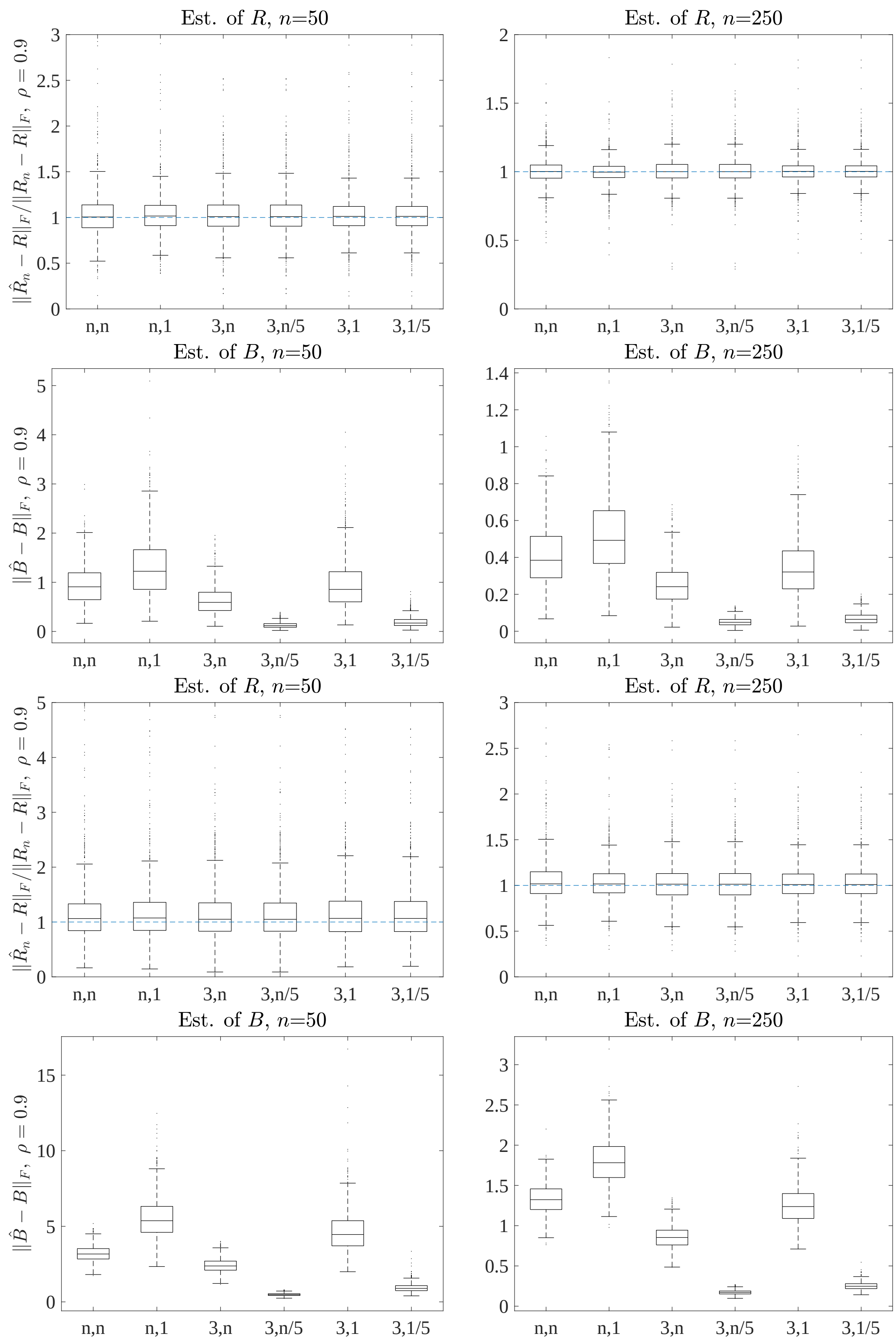

FIG 1. *Simulation results for the unrestricted model under the exact same setting as Figure 2 in the main text, except that now $\rho = 0.9$.*

$$\mathbf{B} = \begin{pmatrix} -0.1241 & 0.7269 & 0.3252 & -0.1649 \\ 1.4897 & -0.3034 & -0.7549 & 0.6277 \\ 1.4090 & 0.2939 & 1.3703 & 1.0933 \\ 1.4172 & -0.7873 & -1.7115 & 1.1093 \\ 0.6715 & 0.8884 & -0.1022 & -0.8637 \\ -1.2075 & -1.1471 & -0.2414 & 0.0774 \\ 0.7172 & -1.0689 & 0.3192 & -1.2141 \\ 1.6302 & -0.8095 & 0.3129 & -1.1135 \\ 0.4889 & -2.9443 & -0.8649 & -0.0068 \\ 1.0347 & 1.4384 & -0.0301 & 1.5326 \end{pmatrix} \quad \text{(for } q = 10\text{)}.$$

Next, Figures 2 and 3 in the supplement present the simulation results for highly correlated covariates with $\rho = 0.9$.

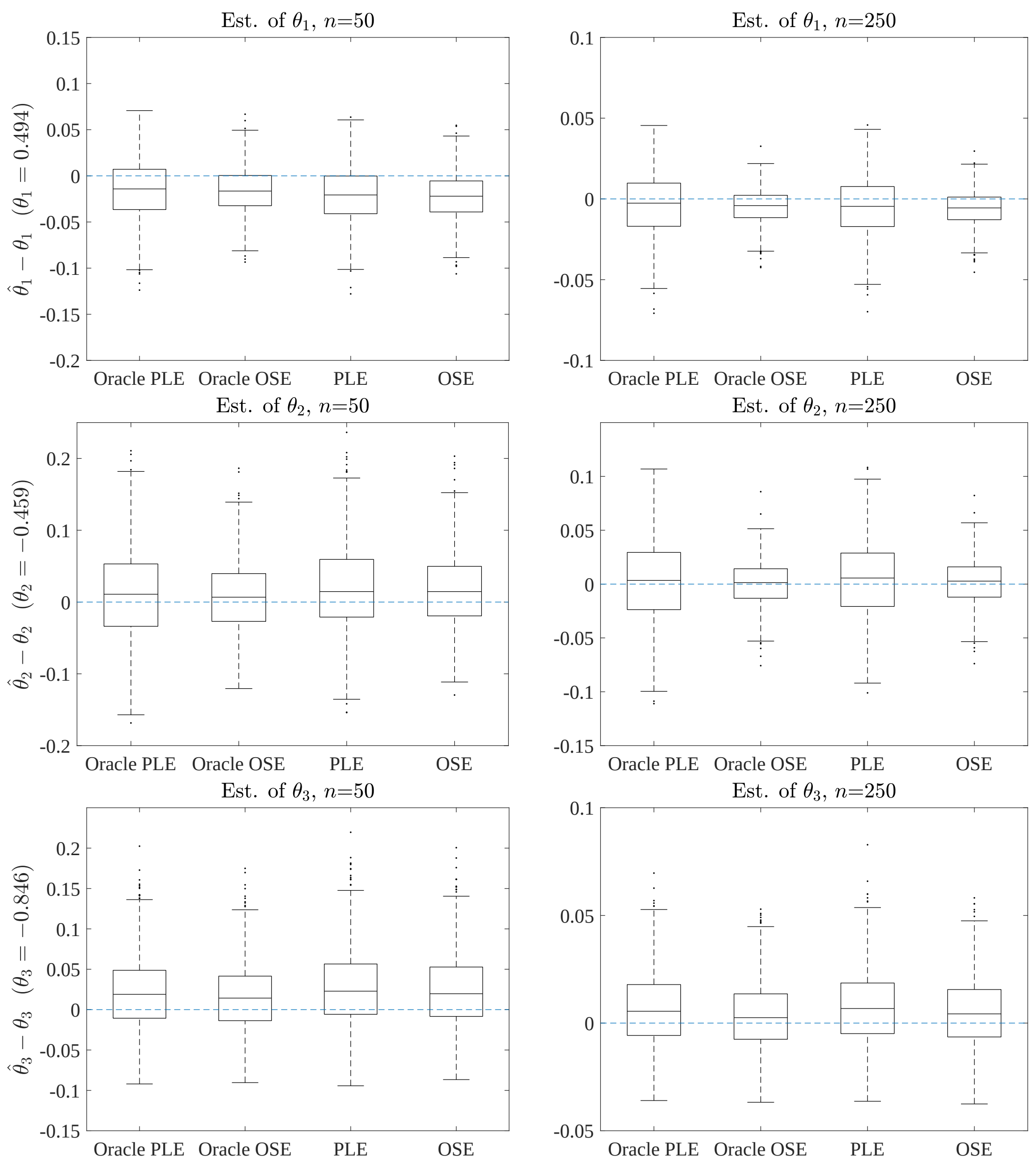

FIG 2. *Simulation results for the Toeplitz model under the exact same setting as Figure 3 in the main text, except that now* $\rho = 0.9$.

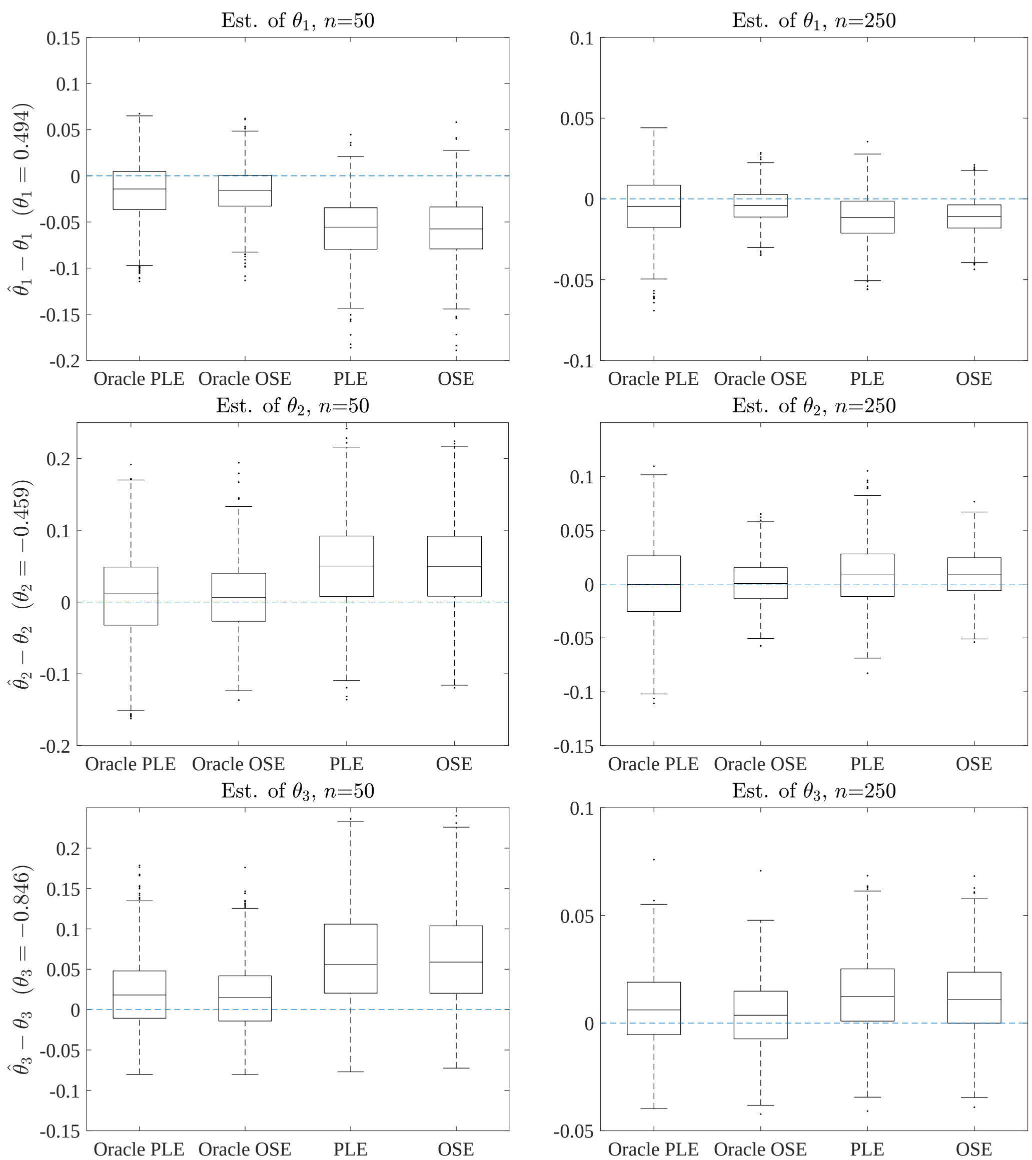

FIG 3. *Simulation results for the Toeplitz model under the exact same setting as Figure 4 in the main text, except that now* $\rho = 0.9$.

References.

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

[2] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag New York, Inc., 1993.

[3] Billingsley, P. *Probability and Measure*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., 1986.

[4] Chang, S.-H., Cosman, P. C., and Milstein, L. B. Chernoff-type bounds for the Gaussian error function. *IEEE Transactions on Communications 59*, 11 (2011), 2939–2944.

[5] Csörgő, M., Csörgő, S., Horváth, L., and Mason, D. M. Weighted empirical and quantile processes. *Ann. Probab. 14*, 1 (1986), 31–85.

[6] de Wet, T., and Venter, J. H. Asymptotic distributions of certain test criteria of normality. *South African Statistical Journal 6*, 2 (1972), 135–149.

[7] Dudley, R. Balls in $\mathbb{R}^k$ do not cut all subsets of $k+2$ points. *Advances in Mathematics 31*, 3 (1979), 306–308.

[8] Giné, E., Koltchinskii, V., and Wellner, J. A. Ratio limit theorems for empirical processes. In *Stochastic Inequalities and Applications*, E. Giné, C. Houdré, and D. Nualart, Eds. Birkhäuser, 2003, pp. 249–278.

[9] Giné, E., and Nickl, R. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press,, 2016.

[10] Gordon, R. D. Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Statist. 12*, 3 (1941), 364–366.

[11] Hagerup, T., and Rüb, C. A guided tour of Chernoff bounds. *Information Processing Letters 33*, 6 (1990), 305–308.

[12] Massart, P. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab. 18* (1990), 1269–1283.

[13] Resnick, S. *A Probability Path*. World Publishing Corporation, 1999.

[14] Segers, J., van den Akker, R., and Werker, B. J. M. Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *Ann. Statist. 42*, 5 (2014), 1911–1940.

[15] Shorack, G. R., and Wellner, J. A. *Empirical Processes with Applications to Statistics*. Society for Industrial and Applied Mathematics, 1986.

[16] van der Vaart, A. *Semiparametric Statistics*. No. 1781 in Lecture Notes in Math. Springer, 2002, pp. 331–457. MR1915446.

[17] van der Vaart, A. W. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[18] van der Vaart, A. W., and Wellner, J. A. *Weak Convergence and Empirical Processes*. Springer, 1996.

[19] Van Keilegom, I., and Akritas, M. G. Transfer of tail information in censored regression models. *Ann. Statist. 27*, 5 (1999), 1745–1784.

I. Gijbels
Department of Mathematics
and Leuven Statistics
Research Center (LStat)
KU Leuven
Celestijnenlaan 200B
B-3001 Leuven (Heverlee)
Belgium
E-mail: irene.gijbels@wis.kuleuven.be

I. Van Keilegom
Y. Zhao
Research Centre for Operations Research
and Business Statistics (ORSTAT)
KU Leuven
Naamsestraat 69
3000 Leuven
Belgium
E-mail: ingrid.vankeilegom@kuleuven.be
yue.zhao@kuleuven.be