INSTITUT DE STATISTIQUE

BIOSTATISTIQUE ET

SCIENCES ACTUARIELLES

(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

2013/30

Semiparametric Gaussian Copula Models: Geometry and Efficient Rank-based Estimation

SEGERS, J., VAN DEN AKKER, R. and B. WERKER

SEMIPARAMETRIC GAUSSIAN COPULA MODELS: GEOMETRY AND EFFICIENT RANK-BASED ESTIMATION

By Johan Segers*, Ramon van den Akker† and Bas J.M. Werker†,‡

*Université catholique de Louvain**
Econometrics Group, CentER, Tilburg University†
Finance Group, CentER, Tilburg University‡

For multivariate Gaussian copula models with unknown margins and structured correlation matrices, a rank-based, semiparametrically efficient estimator is proposed for the Euclidean copula parameter. This estimator is defined as a one-step update of a rank-based pilot estimator in the direction of the efficient influence function, which is calculated explicitly. Moreover, finite-dimensional algebraic conditions are given that completely characterize adaptivity of the model with respect to the unknown marginal distributions and of efficiency of the pseudo-likelihood estimator. For correlation matrices structured according to a factor model, the pseudo-likelihood estimator turns out to be semiparametrically efficient. On the other hand, for Toeplitz correlation matrices, the asymptotic relative efficiency of the pseudo-likelihood estimator with respect to our one-step estimator can be as low as 20%. These findings are confirmed by Monte Carlo simulations.

1. Introduction. Let $\boldsymbol{X} = (X_1, \ldots, X_p)'$ be a p-dimensional random vector with absolutely continuous marginal distribution functions $F_1, \ldots, F_p$ and joint distribution function F. The copula, C, of F is the joint distribution function of the vector $\boldsymbol{U} = (U_1, \ldots, U_p)'$ with $U_j = F_j(X_j)$, uniformly distributed on $(0, 1)$. By Sklar's theorem,

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

yielding a separation of F into its margins $F_1, \ldots, F_p$ and its copula C. The copula remains unchanged if strictly increasing transformations are applied to the p components of $\boldsymbol{X}$.

A semiparametric copula model for the law of the random vector $\boldsymbol{X}$ is a model where F is allowed to have arbitrary, absolutely continuous

AMS 2000 subject classifications: Primary 62F12, 62G20; secondary 62B15, 62H20

Keywords and phrases: adaptivity, correlation matrix, influence function, quadratic form, ranks, score function, tangent space

margins and a copula C_θ which belongs to a finite-dimensional parametric family. An important inference problem is the development of an efficient estimator of the copula parameter θ on the basis of a random sample $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})'$, $i = 1, \ldots, n$. The marginal distributions $F_1, \ldots, F_p$ are thus considered as infinite-dimensional nuisance parameters. In accordance to the structure of the model, a desirable property for the estimator of θ is that it is invariant with respect to strictly increasing transformations applied to the p components. This is equivalent to the requirement that the estimator is measurable with respect to the vectors of ranks, $\boldsymbol{R}_i^{(n)} = (R_{i1}^{(n)}, \ldots, R_{ip}^{(n)})'$, with $R_{ij}^{(n)}$ the rank of X_{ij} within the jth marginal sample $X_{1j}, \ldots, X_{nj}$; see also Hoff (2007, Section 5), who shows that the ranks are G-sufficiency.

There exist a number of rank-based estimation strategies for θ, none of them guaranteed to be semiparametrically efficient. The most common estimators are method-of-moment type estimators (Oakes, 1986; Genest and Rivest, 1993; Klüppelberg and Kuhn, 2009; Brahimi and Necir, 2012; Liu et al., 2012), the pseudo-likelihood estimator (Oakes, 1994; Genest, Ghoudi and Rivest, 1995), and minimum-distance estimators (Tsukahara, 2005; Liebscher, 2009). An expectation–maximization algorithm for a Gaussian copula mixture model is proposed in Li et al. (2011). For vine copulas, the pseudo-likelihood estimator and variants thereof are studied in Hobæk Haff (2013).

Conditions for the efficiency of the pseudo-likelihood estimator are derived in Genest and Werker (2002), where it is concluded that efficiency is the exception rather than the rule. One notable exception is the bivariate Gaussian copula model (see below), where the pseudo-likelihood estimator is asymptotically equivalent to the normal scores rank correlation coefficient, shown to be efficient in Klaassen and Wellner (1997).

A semiparametrically efficient estimator for the copula parameter is proposed in Chen, Fan and Tsyrennikov (2006). However, the estimator is based on parametric sieves for the unknown margins, so that the estimator is not invariant under increasing transformations of the component variables, i.e., the estimator is not rank-based. Moreover, it requires the choice of the orders of the sieves as tuning parameters.

Besides the already mentioned paper by Klaassen and Wellner (1997), the issue of efficient, rank-based estimation is taken up in Hoff, Niu and Wellner (2012) for the important class of semiparametric Gaussian copula models with structured correlation matrices. The derive the semiparametric lower bound to the asymptotic variance of regular estimators for the copula parameter and demonstrate that this bound could be attained by a rank-based estimator. However, they do not provide such an estimator. They also construct a specific Gaussian copula model for which the pseudo-likelihood estimator is not efficient. They conclude their paper with the suggestion that the maximum rank-likelihood estimator in Hoff (2007) may be efficient.

Following Klaassen and Wellner (1997) and Hoff, Niu and Wellner (2012), we put the focus in this paper on semiparametric Gaussian copula models. The Gaussian copula C_θ is the copula of a p-variate Gaussian distribution $N_p(0, R(\theta))$ with $p \times p$ positive definite correlation matrix $R(\theta)$:

$$C_\theta(\boldsymbol{u}) = \Phi_\theta(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)), \quad \boldsymbol{u} = (u_1, \ldots, u_p) \in (0,1)^p, \tag{1.1}$$

with Φ_θ the $N_p(0, R(\theta))$ cumulative distribution function and Φ^{-1} the standard normal quantile function. In the unrestricted Gaussian copula model, $R(\theta)$ can be any $p \times p$ positive definite correlation matrix. Submodels arise by considering structured correlation matrices. In that case, the dimension, k, of the parameter set is smaller than the number of pairs of variables, $p(p-1)/2$. As model for one observation, we thus consider

$$\mathcal{P} = \left(\mathrm{P}_{\theta, F_1, \ldots, F_p} \mid \theta \in \Theta,\, F_1, \ldots, F_p \in \mathcal{F}_{\mathrm{ac}}\right), \qquad \Theta \subset \mathbb{R}^k, \tag{1.2}$$

where $\mathcal{F}_{\mathrm{ac}}$ denotes the set of absolutely continuous distributions on the real line and $\mathrm{P}_{\theta, F_1, \ldots, F_p}$ denotes the law of the random vector $\boldsymbol{X}$ which has copula (1.1) and margins $F_1, \ldots, F_p$.

For the model $\mathcal{P}$, we compute the efficient score and influence functions and the efficient information matrix for the vector-valued parameter θ of a p-variate structured correlation matrix. The computations are based on a detailed analysis of the tangent space structure of the model. The practical interest of this analysis is that it yields a semiparametrically efficient, rank-based estimator for θ. Starting from an initial, rank-based, $\sqrt{n}$-consistent estimator, our estimator is defined as a one-step update in the direction of the estimated efficient influence function. Existence of such an initial estimator is usually no problem: just take the pseudo-likelihood estimator or use a minimum-distance estimator (Klüppelberg and Kuhn, 2009). The update step is easy to calculate, relying just on some simple matrix algebra. We thereby provide a positive answer to the conjecture formulated in Hoff, Niu and Wellner (2012) whether it is possible to attain the semiparametric lower bound using a rank-based estimator. We wish to stress the fact that, in contrast to the estimator in Chen, Fan and Tsyrennikov (2006), our estimator is rank-based and thus invariant with respect to increasing transformations of the component variables, in accordance with the group structure of the model. Note, however, that the methodology in Chen, Fan and Tsyrennikov (2006) also applies to general, not necessarily Gaussian, copula models.

Moreover, by restricting attention to estimators with influence functions of a certain form, we construct an algebraic framework which allows for particularly simple, finite-dimensional conditions on the parametrization $\theta \mapsto R(\theta)$ for efficiency of the pseudo-likelihood estimator and for adaptivity of the model. These are detailed for a large number of examples. The pseudo-likelihood estimator turns out to be efficient not only in the unrestricted model, confirming a remark in Klaassen and Wellner (1997), but also in the often-used class of factor models. On the other hand, for correlation matrices with a Toeplitz structure, the pseudo-likelihood estimator can be quite inefficient, with an asymptotic relative efficiency as low as 20%. Although Hoff, Niu and Wellner (2012) already identified a Gaussian copula model for which the pseudo-likelihood estimator is inefficient, the asymptotic relative efficiency of the pseudo-likelihood estimator in their example was still not far from 100%. The asymptotic results are complemented by Monte Carlo simulations, confirming the above theoretical findings for small samples, even in high dimensions.

The outline of the paper is as follows. In Section 2, we study the model's tangent space, culminating in the calculation of the efficient score function and information matrix for θ. These serve to define the one-step estimator in Section 3, where its semiparametric efficiency is proved. Simple criteria for efficiency of estimators and of adaptivity of the model are established in Section 4. Examples and numerical illustrations are provided in Section 5. Section 6 concludes. Detailed proofs are collected in the appendices.

2. Tangent space and efficient score. The purpose of this section is to compute the semiparametric lower bound for estimating the Gaussian copula parameter θ. Main keys to obtain this lower bound are the tangent space of the semiparametric Gaussian copula model $\mathcal{P}$ and the efficient score function for θ; see Bickel et al. (1993, Chapters 2–3) and Van der Vaart (2000, Chapter 25) for detailed expositions on these notions. Readers mainly interested in our rank-based efficient estimator for θ may want to jump to Section 3 on a first reading.

Section 2.1 states our assumptions and introduces notation that will be used throughout. Section 2.2 shortly discusses the Gaussian copula model with known marginals. In Sections 2.3 and 2.4 we determine the tangent space and efficient score function, respectively. The tangent space theory in Section 2.3 is inspired upon the one for bivariate semiparametric copula models in Bickel et al. (1993, Section 4.7).

2.1. *Assumptions and notations.* The log density of the Gaussian copula (1.1) with $p \times p$ correlation matrix $R(\theta)$ is given by

$$\ell(\boldsymbol{u};\theta) = \log c_\theta(\boldsymbol{u}) = -\frac{1}{2}\log(\det R(\theta)) - \frac{1}{2}\boldsymbol{z}'(R^{-1}(\theta) - I_p)\boldsymbol{z}, \tag{2.1}$$

for $\boldsymbol{u} \in (0,1)^p$, where I_p is the $p \times p$ identity matrix and where $\boldsymbol{z} = (z_1, \ldots, z_p)'$ with $z_j = \Phi^{-1}(u_j)$. Non-singularity of the correlation matrix is part of the following assumption.

ASSUMPTION 2.1. *Suppose $\Theta \subset \mathbb{R}^k$ is open and:*

(i) the mapping $\theta \mapsto R(\theta)$ is one-to-one;
(ii) for all $\theta \in \Theta$, the inverse $S(\theta) = R^{-1}(\theta)$ exists;
(iii) for all $\theta \in \Theta$, the matrices of partial derivatives $\dot{R}_1(\theta), \ldots, \dot{R}_k(\theta)$, defined by $\dot{R}_{m,ij}(\theta) = \partial R_{ij}(\theta)/\partial\theta_m$, for $m = 1, \ldots, k$ and $i, j = 1, \ldots, p$, exist and are continuous in θ;
(iv) for all $\theta \in \Theta$, the matrices $\dot{R}_1(\theta), \ldots, \dot{R}_k(\theta)$ are linearly independent.

Let us also define $p \times p$ matrices $\dot{S}_m(\theta)$ by $\dot{S}_{m,ij}(\theta) = \partial S_{ij}(\theta)/\partial\theta_m$. These derivatives satisfy $\dot{S}_m(\theta) = -S(\theta)\,\dot{R}_m(\theta)\,S(\theta)$, which follows from differentiating $R(\theta)\,S(\theta) = I_p$ (Magnus and Neudecker, 1999, Section 8.4).

The p-dimensional vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ denotes, as in the introduction, a random vector with copula (1.1) and margins $F_1, \ldots, F_p \in \mathcal{F}_{\mathrm{ac}}$. Its law is denoted by $\mathrm{P}_{\theta,F_1,\ldots,F_p}$ and expectations with respect to this law are denoted by $\mathrm{E}_{\theta,F_1,\ldots,F_p}$. In case all margins are uniform on $[0,1]$, notation $\mathrm{Un}[0,1]$, we use $\boldsymbol{U}$, P_θ and E_θ as notations. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$ be defined by $Z_j = \Phi^{-1}(U_j)$ and note that $\boldsymbol{Z} \sim \Phi_\theta$ under P_θ.

Moreover, we consider a measurable space $(\Omega, \mathcal{F})$ supporting probability measures $\mathbb{P}_{\theta,F_1,\ldots,F_p}$, for $\theta \in \Theta$ and $F_1, \ldots, F_p \in \mathcal{F}_{\mathrm{ac}}$, and i.i.d. random vectors $\boldsymbol{X}_i$, $i \in \mathbb{N}$, each with law $\mathrm{P}_{\theta,F_1,\ldots,F_p}$ under $\mathbb{P}_{\theta,F_1,\ldots,F_p}$. Expectations with respect to $\mathbb{P}_{\theta,F_1,\ldots,F_p}$ are denoted by $\mathbb{E}_{\theta,F_1,\ldots,F_p}$. Furthermore, $\mathbb{P}_\theta$ is shorthand for $\mathbb{P}_{\theta,\mathrm{Un}[0,1],\ldots,\mathrm{Un}[0,1]}$, for which expectations are denoted by $\mathbb{E}_\theta$.

2.2. *The Gaussian copula model with known margins.* Starting point of the analysis is the case that the margins $F_1, \ldots, F_p \in \mathcal{F}_{\mathrm{ac}}$ are known. In particular, we compute the Fisher information matrix for θ in this case. Due to the transformation structure of the model, it suffices to consider uniform margins, that is, to consider the model $\mathcal{P}_{\mathrm{km}} = (\mathrm{P}_\theta \mid \theta \in \Theta)$.

Under Assumption 2.1, the score $\dot{\ell}_\theta(\boldsymbol{u};\theta) = (\dot{\ell}_{\theta,m}(\boldsymbol{u};\theta))_{m=1}^k$ is given by

$$\dot{\ell}_{\theta,m}(\boldsymbol{u};\theta) = \frac{\partial}{\partial\theta_m}\ell(\boldsymbol{u};\theta) = -\frac{1}{2}\operatorname{tr}\bigl(S(\theta)\dot{R}_m(\theta)\bigr) - \frac{1}{2}\boldsymbol{z}'\dot{S}_m(\theta)\boldsymbol{z}, \tag{2.2}$$

for $\boldsymbol{u} \in (0,1)^p$, where the partial derivative of $\det R(\theta)$ follows from Jacobi's formula (Magnus and Neudecker, 1999, Section 8.3). The $k \times k$ Fisher information matrix is defined by $I(\theta) = \mathrm{E}_\theta[\dot{\ell}_\theta \dot{\ell}_\theta'(\boldsymbol{U};\theta)]$.

To obtain a convenient representation of $I(\theta)$, we introduce an inner product on the linear space $\mathrm{Sym}(p)$ of real symmetric $p \times p$ matrices. For $A, B \in \mathrm{Sym}(p)$, put

$$\langle A, B\rangle_\theta = \mathrm{cov}_\theta\left(\frac{1}{2}\boldsymbol{Z}'A\boldsymbol{Z}, \frac{1}{2}\boldsymbol{Z}'B\boldsymbol{Z}\right) = \frac{1}{2}\,\mathrm{tr}\big(A\,R(\theta)\,B\,R(\theta)\big), \tag{2.3}$$

the covariance being calculated for $\boldsymbol{Z} \sim \Phi_\theta$. It is easily verified that $\langle\,\cdot\,,\,\cdot\,\rangle_\theta$ defines an inner product on $\mathrm{Sym}(p)$. In particular, if $A \in \mathrm{Sym}(p)$ is such that $\langle A, A\rangle_\theta = 0$, then $\boldsymbol{Z}'A\boldsymbol{Z}$ is almost surely equal to a constant and thus, $R(\theta)$ being nonsingular (Assumption 2.1), $A = 0$.

From (2.2) and (2.3) we obtain, for $m, m' = 1, \ldots, k$,

$$I_{mm'}(\theta) = \langle -\dot{S}_m(\theta), -\dot{S}_{m'}(\theta)\rangle_\theta, \tag{2.4}$$

which is a continuous function of θ. Note that $I(\theta)$ is the Gram matrix associated to the matrices $-\dot{S}_1(\theta), \ldots, -\dot{S}_k(\theta)$, using (2.3) as inner product. Since Assumption 2.1 implies linear independence of $-\dot{S}_1(\theta), \ldots, -\dot{S}_k(\theta)$ (see Part A of the proof of Proposition 2.8 below), the information matrix $I(\theta)$ is non-singular.

From these observations, it follows that the Gaussian copula model with known margins is regular (Bickel et al., 1993, Definition 2.1.1 and Proposition 2.1.1). For ease of reference, we state this fact in the following lemma.

LEMMA 2.2. *If the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1, then the parametric Gaussian copula model with known, uniform margins is regular.*

REMARK 2.3. Regularity of $\mathcal{P}_{\mathrm{km}}$ has some useful consequences.

(i) The score function has zero expectation, $\mathrm{E}_\theta[\dot{\ell}_\theta(\boldsymbol{U};\theta)] = 0$. This can also be seen from (2.2) and $\mathrm{E}_\theta(\boldsymbol{Z}'A\boldsymbol{Z}) = \mathrm{tr}(A\,R(\theta))$.
(ii) By the Hájek–Le Cam convolution theorem, the inverse of the Fisher information matrix, $I^{-1}(\theta)$, constitutes a lower bound to the asymptotic variance of regular estimators of θ in the model $\mathcal{P}_{\mathrm{km}}$ (Van der Vaart, 2000, Chapter 8).

2.3. *Tangent space.* In this section we derive the tangent space of the semiparametric Gaussian copula model $\mathcal{P}$. The structure of the space is similar to the one of the bivariate semiparametric Gaussian copula model in

Klaassen and Wellner (1997). In Section 2.4 we will use this tangent space to calculate the efficient score for the copula parameter θ. In turn, the efficient score determines the semiparametric lower bound for the asymptotic variance of regular estimators of θ.

Informally, the tangent space at $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$ is given by the collection of score functions of parametric submodels of $\mathcal{P}$. Such score functions can be thought of as functions on $\mathbb{R}^p$ of the form

$$\boldsymbol{x} \mapsto \frac{\partial}{\partial \eta} \log p_{\theta+\eta\alpha,F_{1,\eta},\ldots,F_{p,\eta}}(\boldsymbol{x}) \bigg|_{\eta=0}. \tag{2.5}$$

Here $\alpha \in \mathbb{R}^k$, while $p_{\theta+\eta\alpha,F_{1,\eta},\ldots,F_{p,\eta}}$ is the density of $\mathrm{P}_{\theta+\eta\alpha,F_{1,\eta},\ldots,F_{p,\eta}} \in \mathcal{P}$, depending on a real parameter η taking values in a neighbourhood of 0. The marginal distributions are parametrized through paths $\eta \mapsto F_{j,\eta}$ in $\mathcal{F}_{\mathrm{ac}}$ that pass through F_j at $\eta = 0$.

The tangent space falls apart into two subspaces:

- a parametric part, arising from score functions of parametric submodels for which the margins are constant, $F_{j,\eta} = F_j$;
- a nonparametric part, arising from score functions of parametric submodels for which the copula parameter is constant, $\alpha = 0$.

The parametric part corresponds in fact to the linear span of the score functions in the parametric Gaussian copula model with known margins. The nonparametric part describes the additional part of the model stemming from the margins being unknown.

Formally, the tangent space is a subspace of $\mathrm{L}_2(\mathrm{P}_{\theta,F_1,\ldots,F_p})$, the space of square-integrable functions with respect to $\mathrm{P}_{\theta,F_1,\ldots,F_p}$. The pointwise derivatives in (2.5) will be replaced by derivatives in quadratic mean. The nonparametric part of the tangent space is described most conveniently as the image of a bounded linear operator, called score operator. It is the description and the analysis of this score operator that constitutes the gist of this section.

The density of $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$ is given by

$$p_{\theta,F_1,\ldots,F_p}(\boldsymbol{x}) = c_\theta\big(F_1(x_1),\ldots,F_p(x_p)\big) \prod_{j=1}^{p} f_j(x_j), \qquad \boldsymbol{x} \in \mathbb{R}^p,$$

with $f_1,\ldots,f_p$ the densities of $F_1,\ldots,F_p$, respectively. If $\eta \mapsto F_{j,\eta}(x_j)$ is differentiable at $\eta = 0$, we obtain

$$\frac{\partial}{\partial \eta} \log p_{\theta, F_{1,\eta}, \ldots, F_{p,\eta}}(\boldsymbol{x}) \bigg|_{\eta=0}$$
$$= \sum_{j=1}^{p} \left\{ \frac{\partial}{\partial \eta} \log f_{j,\eta}(x_j) \bigg|_{\eta=0} + \dot{\ell}_j(F_1(x_1), \ldots, F_p(x_p); \theta) \, \frac{\partial}{\partial \eta} F_{j,\eta}(x_j) \bigg|_{\eta=0} \right\},$$

with, for $j = 1, \ldots, p$ and $\boldsymbol{u} \in (0,1)^p$,

$$\dot{\ell}_j(\boldsymbol{u}; \theta) = \frac{\partial}{\partial u_j} \ell(\boldsymbol{u}; \theta) = \frac{z_j}{\varphi(z_j)} - \sum_{i=1}^{p} S_{ij}(\theta) \frac{z_i}{\varphi(z_j)}, \tag{2.6}$$

where φ denotes the standard normal density. Note that, for $u_j \in (0,1)$,

$$\mathrm{E}_\theta \big[\dot{\ell}_j(\boldsymbol{U}; \theta) \,\big|\, U_j = u_j \big] = \frac{z_j}{\varphi(z_j)} \left(1 - \sum_{i=1}^{p} S_{ij}(\theta) R_{ij}(\theta) \right) = 0. \tag{2.7}$$

The above formulas motivate the introduction of the following linear operators, which together will constitute the above-mentioned score operator. Let $\mathrm{L}_2^0[0,1]$ be the subspace of $\mathrm{L}_2[0,1] = \mathrm{L}_2([0,1], \mathcal{B}_{[0,1]}, \mathrm{d}\lambda)$ resulting from the restriction $\int h(\lambda) \, \mathrm{d}\lambda = 0$ for $h \in \mathrm{L}_2[0,1]$. For $j = 1, \ldots, p$, we introduce linear operators $\mathcal{O}_{\theta,j} : \mathrm{L}_2^0[0,1] \to \mathrm{L}_2(\mathrm{P}_\theta)$ by

$$\mathcal{O}_{\theta,j} h = [\mathcal{O}_{\theta,j} h](\boldsymbol{U}) = h(U_j) + \dot{\ell}_j(\boldsymbol{U}; \theta) \, H(U_j), \tag{2.8}$$

where $H(u) = \int_0^u h(\lambda) \mathrm{d}\lambda$ and where $\boldsymbol{U}$ is the identity mapping on $(0,1)^p$. The claim that the random variable on the right-hand side has a finite variance for $\boldsymbol{U} \sim \mathrm{P}_\theta$ is part of Lemma 2.4.

The score operator itself, $\mathcal{O}_\theta$, has domain $(\mathrm{L}_2^0[0,1])^p$ and is defined by

$$\mathcal{O}_\theta \boldsymbol{h} = \sum_{j=1}^{p} \mathcal{O}_{\theta,j} h_j, \qquad \boldsymbol{h} = (h_1, \ldots, h_p) \in (\mathrm{L}_2^0[0,1])^p.$$

Lemma 2.4 will present basic properties of $\mathcal{O}_{\theta,j}$ and $\mathcal{O}_\theta$. A formal description of the tangent space via the score operator will be given in Proposition 2.6.

To this end, we first need to introduce some additional notation. For $i, j = 1, \ldots, p$ and $\boldsymbol{u} \in (0,1)^p$, we define

$$\ddot{\ell}_{ij}(\boldsymbol{u}; \theta) = \frac{\partial}{\partial u_i} \dot{\ell}_j(\boldsymbol{u}; \theta)$$
$$= \begin{cases} \dfrac{z_j^2 + 1 - S_{jj}(\theta)}{\varphi^2(z_j)} - \displaystyle\sum_{t=1}^{p} S_{tj}(\theta) \frac{z_t z_j}{\varphi^2(z_j)}, & \text{if } i = j; \\[2ex] -\dfrac{S_{ij}(\theta)}{\varphi(z_i)\,\varphi(z_j)}, & \text{if } i \neq j. \end{cases}$$

For $u_j \in (0,1)$, we have

$$I_{jj}(u_j;\theta) = \mathrm{E}_\theta\big[\dot\ell_j^2(\boldsymbol{U};\theta) \,\big|\, U_j = u_j\big] = -\,\mathrm{E}_\theta\big[\ddot\ell_{jj}(\boldsymbol{U};\theta) \,\big|\, U_j = u_j\big] = \frac{S_{jj}(\theta)-1}{\varphi^2(z_j)}. \tag{2.9}$$

From well-known results on Mill's ratio (Gordon, 1941), we obtain the bound $1/\varphi(z_j) \le M\{u_j(1-u_j)\}^{-1}$, for all $u_j \in (0,1)$ and some constant $M > 0$. Hence, under Assumption 2.1, there exists a constant $M_\theta > 0$ such that

$$I_{jj}(u_j;\theta) \le \frac{M_\theta}{\{u_j(1-u_j)\}^2}, \qquad u_j \in (0,1). \tag{2.10}$$

This bound is exploited in the proof of the following lemma, which states that $\mathcal{O}_\theta$ is a continuously invertible operator from $(\mathrm{L}_2^0[0,1])^p$ into $\mathrm{L}_2^0(\mathrm{P}_\theta)$. Here we equip $(\mathrm{L}_2^0[0,1])^p$ with the inner product $\langle \boldsymbol{g},\boldsymbol{h}\rangle = \sum_{j=1}^p \int_0^1 g_j(\lambda)\, h_j(\lambda)\,\mathrm{d}\lambda$ for $\boldsymbol{g},\boldsymbol{h} \in (\mathrm{L}_2^0[0,1])^p$, while $\mathrm{L}_2^0(\mathrm{P}_\theta)$ is the subspace of $\mathrm{L}_2(\mathrm{P}_\theta)$ resulting from the restriction $\mathrm{E}_\theta[f(\boldsymbol{U})] = 0$ for $f \in \mathrm{L}_2(\mathrm{P}_\theta)$.

LEMMA 2.4. *Let $\theta \mapsto R(\theta)$ be a parametrization that satisfies Assumption 2.1 and let $\theta \in \Theta$.*

(a) The map $\mathcal{O}_{\theta,j}$, $j = 1,\dots,p$, is a bounded operator from $\mathrm{L}_2^0[0,1]$ into $\mathrm{L}_2^0(\mathrm{P}_\theta)$. The map $\mathcal{O}_\theta$ is a bounded operator from $(\mathrm{L}_2^0[0,1])^p$ into $\mathrm{L}_2^0(\mathrm{P}_\theta)$.
(b) The operator $\mathcal{O}_\theta$ is continuously invertible on its range, $\mathrm{R}\mathcal{O}_\theta$, with inverse ${\mathcal{O}_\theta}^{-1} : \mathrm{L}_2^0(\mathrm{P}_\theta) \to (\mathrm{L}_2^0[0,1])^p$ given by

$$[{\mathcal{O}_\theta}^{-1} f]_j(u_j) = \mathrm{E}_\theta[f(\boldsymbol{U}) \mid U_j = u_j], \qquad u_j \in (0,1),\ j = 1,\dots,p.$$

The proof is given in Appendix A. Here we just note that the proof of part (a) follows the one of Proposition 4.7.2 in Bickel et al. (1993).

REMARK 2.5. An application of Banach's theorem (see, e.g., Bickel et al., 1993, Proposition A.1.7) implies that $\mathrm{R}\mathcal{O}_\theta$ is closed.

We proceed with the construction of the tangent space. First we define the "local paths" through $\mathcal{F}_{\mathrm{ac}}$. To this end, fix $F_1,\dots,F_p \in \mathcal{F}_{\mathrm{ac}}$ with densities $f_1,\dots,f_p$. Let $h_1,\dots,h_p \in \mathrm{L}_2^0[0,1]$ and introduce univariate densities $f_{j,\eta}(\,\cdot\,;h_j)$, for $\eta \in (-1,1)$ and $j = 1,\dots,p$, by

$$f_{j,\eta}(x;h_j) = d(\eta;f_j,h_j)\, g(\eta\, h_j(F_j(x)))\, f_j(x), \qquad x \in \mathbb{R},$$

where $g(z) = 2/(1+e^{-2z})$ and where $d(\eta;f_j,h_j)$ is a positive constant such that $f_{j,\eta}(\cdot;h_j)$ integrates to one. Let $F_{j,\eta}^{h_j} \in \mathcal{F}_{\mathrm{ac}}$ be the induced distribution functions. The path $\eta \mapsto (F_{1,\eta}^{h_1}, \dots, F_{p,\eta}^{h_p})$, with values in the space $(\mathcal{F}_{\mathrm{ac}})^p$, passes through $(F_1, \dots, F_p)$ at $\eta = 0$.

The following proposition describes the score function at $\eta = 0$ of the parametric submodel

$$\Big(\mathrm{P}_{\theta+\eta\alpha, F_{1,\eta}^{h_1}, \dots, F_{p,\eta}^{h_p}} \,\Big|\, -\varepsilon < \eta < \varepsilon\Big) \subset \mathcal{P}$$

for fixed $\alpha \in \mathbb{R}^k$ and $h_1, \dots, h_p \in \mathrm{L}_2^0[0,1]$ and some $\varepsilon > 0$. The collection of all such score functions is, by definition, the tangent set of the semiparametric Gaussian copula model $\mathcal{P}$ at $\mathrm{P}_{\theta, F_1, \dots, F_p}$.

PROPOSITION 2.6. *Consider a parametrization $\theta \mapsto R(\theta)$ for which Assumption 2.1 holds and let $\mathrm{P}_{\theta, F_1, \dots, F_p} \in \mathcal{P}$. Let $\boldsymbol{h} = (h_1, \dots, h_p) \in (\mathrm{L}_2^0[0,1])^p$ and let $\alpha \in \mathbb{R}^k$. Then the path $\eta \mapsto (\theta + \eta\alpha, F_{1,\eta}^{h_1}, \dots, F_{p,\eta}^{h_p})$, in $\Theta \times (\mathcal{F}_{ac})^p$, yields the following score at $\eta = 0$,*

$$\dot{\ell}^{\alpha, \boldsymbol{h}}(\boldsymbol{x}) = \alpha' \dot{\ell}_\theta(F_1(x_1), \dots, F_p(x_p); \theta) + [\mathcal{O}_\theta \boldsymbol{h}](F_1(x_1), \dots, F_p(x_p))$$

for $\boldsymbol{x} \in \mathbb{R}^p$, that is,

$$\lim_{\eta \to 0} \int_{\mathbb{R}^p} \left(\frac{\sqrt{p_\eta(\boldsymbol{x})} - \sqrt{p_0(\boldsymbol{x})}}{\eta} - \frac{1}{2} \dot{\ell}^{\alpha, \boldsymbol{h}}(\boldsymbol{x}) \sqrt{p_0(\boldsymbol{x})} \right)^2 \mathrm{d}\boldsymbol{x} = 0,$$

where $p_\eta = p_{\theta+\eta\alpha, F_{1,\eta}^{h_1}, \dots, F_{p,\eta}^{h_p}}$. The tangent set

$$\big\{ \dot{\ell}^{\alpha, \boldsymbol{h}} \,\big|\, \alpha \in \mathbb{R}^k,\ \boldsymbol{h} \in (\mathrm{L}_2^0[0,1])^p \big\}, \tag{2.11}$$

is a closed subspace of $\mathrm{L}_2^0(\mathrm{P}_{\theta, F_1, \dots, F_p})$.

The tangent set (2.11) being a closed subspace, it is called the *tangent space* of $\mathcal{P}$ at $\mathrm{P}_{\theta, F_1, \dots, F_p}$. Apart from the statement on closedness, which we discuss below, the proof of Proposition 2.6 is analogous to the proof of Proposition 4.7.4 in Bickel et al. (1993) and is omitted.

The nonparametric part $\mathcal{T}_{\mathrm{P}_{\theta, F_1, \dots, F_p}}$ of the tangent space at $\mathrm{P}_{\theta, F_1, \dots, F_p}$ corresponds, by definition, to the subset of (2.11) resulting from the restriction $\alpha = 0$, that is,

$$\begin{aligned} \mathcal{T}_{\mathrm{P}_{\theta, F_1, \dots, F_p}} &= \big\{ \dot{\ell}^{0, \boldsymbol{h}} \,\big|\, \boldsymbol{h} \in (\mathrm{L}_2^0[0,1])^p \big\} \\ &= \big\{ \boldsymbol{x} \mapsto [\mathcal{O}_\theta \boldsymbol{h}](F_1(x_1), \dots, F_p(x_p)) \,\big|\, \boldsymbol{h} \in (\mathrm{L}_2^0[0,1])^p \big\}. \end{aligned}$$

Since $\mathcal{T}_{\mathrm{P}_{\theta, F_1, \dots, F_p}}$ is isometric to $\mathcal{T}_{\mathrm{P}_\theta}$ and since $\mathcal{T}_{\mathrm{P}_\theta} = \mathrm{R}\mathcal{O}_\theta$, which is closed (see Remark 2.5), it follows that $\mathcal{T}_{\mathrm{P}_{\theta, F_1, \dots, F_p}}$ is a closed subspace of $\mathrm{L}_2^0(\mathrm{P}_{\theta, F_1, \dots, F_p})$. As the tangent space (2.11) is the sum of $\mathcal{T}_{\mathrm{P}_{\theta, F_1, \dots, F_p}}$ and a finite-dimensional space, the tangent space is closed as well.

2.4. *Efficient score.* The semiparametric lower bound for regular estimators of the copula parameter θ is determined by the efficient score, $\dot\ell^*_\theta(\boldsymbol{X};\mathrm{P}_{\theta,F_1,\dots,F_p})$. (As before, we identify square-integrable functions with random variables; formally, view $\boldsymbol{X}$ as the identity map on $\mathbb{R}^p$.) This efficient score is, by definition, given by

$$\begin{aligned}\dot\ell^*_\theta(\boldsymbol{X};\mathrm{P}_{\theta,F_1,\dots,F_p}) &= \dot\ell_\theta(F_1(X_1),\dots,F_p(X_p);\theta)\\ &\qquad - \Pi(\dot\ell_\theta(F_1(X_1),\dots,F_p(X_p);\theta)\mid \mathcal{T}_{\mathrm{P}_{\theta,F_1,\dots,F_p}}),\end{aligned}$$

where $\Pi(\,\cdot\,|\mathcal{T}_{\mathrm{P}_{\theta,F_1,\dots,F_p}})$ is the (coordinate-wise) projection operator from $\mathrm{L}_2(\mathrm{P}_{\theta,F_1,\dots,F_p})$ onto the closed subspace $\mathcal{T}_{\mathrm{P}_{\theta,F_1,\dots,F_p}}$. Note that $\dot\ell_\theta$ and hence $\dot\ell^*_\theta$ are vectors of length k, the length of the copula parameter θ.

For Gaussian copula models, the projection can be calculated explicitly, which will lead eventually to our one-step estimator in Section 3. This situation is in contrast to the one for most other copula models, even for bivariate copulas indexed by a real-valued parameter, where the calculation of the efficient score requires the solution of a pair of coupled Sturm–Liouville differential equations (Bickel et al., 1993, Section 4.7).

From the isometry between $\mathcal{T}_{\mathrm{P}_{\theta,F_1,\dots,F_p}}$ and $\mathcal{T}_{\mathrm{P}_\theta}$, via the mapping $\boldsymbol{X}\mapsto(F_1(X_1),\dots,F_p(X_p))$, we obtain the important identity

$$\dot\ell^*_\theta(\boldsymbol{X};\mathrm{P}_{\theta,F_1,\dots,F_p}) = \dot\ell^*_\theta(F_1(X_1),\dots,F_p(X_p);\theta),\qquad \mathrm{P}_{\theta,F_1,\dots,F_p}\text{-a.s.,}\tag{2.12}$$

where $\dot\ell^*_\theta(\,\cdot\,;\theta)$ is shorthand notation for $\dot\ell^*_\theta(\,\cdot\,;\mathrm{P}_\theta)$. As a consequence, the $k\times k$ efficient information matrix for θ at $\mathrm{P}_{\theta,F_1,\dots,F_p}$, given by

$$I^*(\theta) = \mathrm{E}_{\theta,F_1,\dots,F_p}[\dot\ell^*_\theta\dot\ell^{*\prime}_\theta(\boldsymbol{X}|\mathrm{P}_{\theta,F_1,\dots,F_p})] = \mathrm{E}_\theta[\dot\ell^*_\theta\dot\ell^{*\prime}_\theta(\boldsymbol{U};\theta)],$$

does not depend on the marginals $F_1,\dots,F_p$.

Because $\mathcal{T}_{\mathrm{P}_\theta}=\mathrm{R}\mathcal{O}_\theta$ and $\mathcal{O}_\theta$ is one-to-one (see Proposition 2.4), there exist unique elements $\boldsymbol{h}^\theta_m=(h^\theta_{1,m},\dots,h^\theta_{p,m})\in(\mathrm{L}^0_2[0,1])^p$, $m=1,\dots,k$, such that

$$\Pi(\dot\ell_{\theta,m}(\boldsymbol{U};\theta)\mid\mathcal{T}_{\mathrm{P}_\theta}) = [\mathcal{O}_\theta\boldsymbol{h}^\theta_m](\boldsymbol{U}).\tag{2.13}$$

These "generators of the efficient score" are completely determined by the orthogonality conditions

$$0=\mathrm{E}_\theta\big[\big(\dot\ell_{\theta,m}(\boldsymbol{U};\theta)-[\mathcal{O}_\theta\boldsymbol{h}^\theta_m](\boldsymbol{U})\big)\,[\mathcal{O}_\theta\boldsymbol{h}](\boldsymbol{U})\big],\qquad \boldsymbol{h}\in(\mathrm{L}^0_2[0,1])^p.\tag{2.14}$$

Before we solve these equations for $\boldsymbol{h}^\theta_m$, $m=1,\dots,k$, we discuss a result that provides a necessary and sufficient condition for orthogonality of quadratic forms in the Gaussianized variables to the space $\mathcal{T}_{\mathrm{P}_\theta}$. Its proof is given in Appendix A.

LEMMA 2.7. *Consider a parametrization $\theta \mapsto R(\theta)$ for which Assumption 2.1 holds and let $\theta \in \Theta$ and $A \in \mathrm{Sym}(p)$. The following two conditions are equivalent:*

(a) the function $\boldsymbol{Z}'A\boldsymbol{Z}$ in $\mathrm{L}_2(\mathrm{P}_\theta)$ is orthogonal to $\mathcal{T}_{\mathrm{P}_\theta}$;
(b) $(R(\theta)A)_{jj} = 0$ for $j = 1, \ldots, p$.

The following proposition presents the solution $\boldsymbol{h}_m^\theta$ to (2.14) and the resulting efficient score and efficient information matrix. To formulate these results, we introduce the notation, for $\boldsymbol{b} \in \mathbb{R}^p$ and $\theta \in \Theta$,

$$D_\theta(\boldsymbol{b}) = S(\theta)\,\mathrm{diag}(\boldsymbol{b}) + \mathrm{diag}(\boldsymbol{b})\,S(\theta), \tag{2.15}$$

where $\mathrm{diag}(\boldsymbol{b})$ is the diagonal matrix with diagonal $\boldsymbol{b}$. Let $\boldsymbol{\iota}_p$ denote the p-dimensional vector of ones and let $A \circ B$ denote the Hadamard product of conformable matrices A and B. Recall the inner product introduced in (2.3) and recall the convention $z = \Phi^{-1}(u)$ for $u \in (0, 1)$.

PROPOSITION 2.8. *Consider a parametrization $\theta \mapsto R(\theta)$ for which Assumption 2.1 holds and let $\theta \in \Theta$. Then, for $m = 1, \ldots, k$, the vector $\boldsymbol{h}_m^\theta \in (\mathrm{L}_2^0[0,1])^p$ in (2.13) is given by*

$$h_{j,m}^\theta(u_j) = g_{j,m}(\theta)(1 - z_j^2), \qquad u_j \in (0,1),\ j = 1, \ldots, p, \tag{2.16}$$

where $\boldsymbol{g}_m(\theta) = (g_{1,m}(\theta), \ldots, g_{p,m}(\theta))'$ is given by

$$\boldsymbol{g}_m(\theta) = -\big(I_p + R(\theta) \circ S(\theta)\big)^{-1}\big(\dot{R}_m(\theta) \circ S(\theta)\big)\,\boldsymbol{\iota}_p. \tag{2.17}$$

Moreover, the efficient score, $\dot{\ell}_\theta^*$, is given by

$$\dot{\ell}_{\theta,m}^*(\boldsymbol{u};\theta) = \frac{1}{2}\boldsymbol{z}'\big(D_\theta(\boldsymbol{g}_m(\theta)) - \dot{S}_m(\theta)\big)\boldsymbol{z} \tag{2.18}$$

for $m = 1, \ldots, k$. Finally, the efficient information matrix, $I^*(\theta)$, is given by

$$I_{mm'}^*(\theta) = \langle D_\theta(\boldsymbol{g}_m(\theta)) - \dot{S}_m(\theta), D_\theta(\boldsymbol{g}_{m'}(\theta)) - \dot{S}_{m'}(\theta)\rangle_\theta, \tag{2.19}$$

for $m, m' = 1, \ldots, k$, and is non-singular.

OUTLINE OF THE PROOF. The proof is decomposed into four parts. Here, we just give an outline. For a detailed proof, please see Appendix A.

In Part A we show that the $k + p$ matrices

$$-\dot{S}_1(\theta), \ldots, -\dot{S}_k(\theta), D_\theta(\boldsymbol{e}_1), \ldots, D_\theta(\boldsymbol{e}_p),$$

with $\boldsymbol{e}_i$ the ith canonical unit vector in $\mathbb{R}^p$, are linearly independent. Part B exploits this result to demonstrate non-singularity of $I_p+R(\theta)\circ S(\theta)$, thereby showing that the vector $\boldsymbol{g}_m(\theta)$ in (2.17) is well-defined. Part C shows that equation (2.16) together with the definition

$$\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta) = \dot{\ell}_{\theta,m}(\boldsymbol{u};\theta) - [\mathcal{O}_\theta \boldsymbol{h}^\theta_m](\boldsymbol{u}) \tag{2.20}$$

lead to (2.18)–(2.19) and also demonstrates non-singularity of $I^*(\theta)$. Finally, Part D proves that the orthogonality conditions (2.14) hold by applying Lemma 2.7, and thus shows that (2.18) is the efficient score. □

REMARK 2.9. The $k \times k$ positive semidefinite matrix $I(\theta) - I^*(\theta)$ represents the loss of information due to not knowing the marginals. In Section 4.4 we provide conditions for adaptivity, i.e. $I(\theta) = I^*(\theta)$.

REMARK 2.10. Klaassen and Wellner (1997) derived the efficient score for the bivariate ($p = 2$) unrestricted Gaussian copula model by solving a system of Sturm–Liouville equations. The functions (2.16) solve a p-variate analogue for the component $m \in \{1, \dots, k\}$ of the efficient score, see Appendix 2.

3. An efficient rank-based estimator.

In this section we use the efficient score $\dot{\ell}^*_\theta$ and the efficient information matrix $I^*(\theta)$, as obtained in Proposition 2.8, to construct a rank-based, semiparametrically efficient estimator of θ. Recall (Van der Vaart, 2000, Sections 25.3–25.4) that $\hat{\theta}_n$ is an efficient estimator of θ in model $\mathcal{P}$ at $\mathrm{P}_{\theta,F_1,\dots,F_p} \in \mathcal{P}$ if and only if, under $\mathbb{P}_{\theta,F_1,\dots,F_p}$,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I^{*-1}(\theta)\, \dot{\ell}^*_\theta(F_1(X_{i1}), \dots, F_p(X_{ip}); \theta) + o_P(1). \tag{3.1}$$

Moreover, $\hat{\theta}_n$ is called an efficient estimator of θ in the model $\mathcal{P}$ if it is efficient at all $\mathrm{P}_{\theta,F_1,\dots,F_p} \in \mathcal{P}$. The limiting distribution of an efficient estimator is thus given by $N_k(0, I^{*-1}(\theta))$. By the Hájek–Le Cam convolution theorem (Van der Vaart, 2000, Section 25.3), the limiting distribution of any regular estimator is given by the convolution of $N_k(0, I^{*-1}(\theta))$ and another, estimator specific, distribution. As a consequence, $I^{*-1}(\theta)$ provides a lower bound to the asymptotic variance of regular estimators. See Section 4.2 for an insightful characterization of regularity for estimators in structured Gaussian copula models.

The vector-valued function

$$\boldsymbol{x} \mapsto I^{*-1}(\theta)\, \dot{\ell}^*_\theta(F_1(x_1), \dots, F_p(x_p); \theta) \tag{3.2}$$

is called the *efficient influence function*. According to (3.1), an estimator sequence $\hat{\theta}_n$ is efficient for θ if and only if $\sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically linear in the efficient influence function. By Proposition 2.8, each component of the efficient influence function is a centered quadratic form in the Gaussianized vector $\boldsymbol{z} \in \mathbb{R}^p$, where $z_j = \Phi^{-1}(u_j)$ and $u_j = F_j(x_j)$. This fact will be extensively used in Section 4.

We construct an efficient one-step estimator (OSE) by updating an initial $\sqrt{n}$-consistent estimator of θ. Since we want to construct a rank-based estimator of θ, we require that the initial estimator is rank-based too. We summarize these requirements in the following assumption. Recall that we consider a measurable space equipped with probability measures $\mathbb{P}_{\theta,F_1,\ldots,F_p}$ and carrying random vectors $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})'$, for $i \geq 1$, which, under $\mathbb{P}_{\theta,F_1,\ldots,F_p}$, are i.i.d. with common law $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$. Also recall that $\boldsymbol{R}_i^{(n)} = (R_{i1}^{(n)}, \ldots, R_{ip}^{(n)})'$ denotes a vector of ranks, with $R_{ij}^{(n)}$ the rank of X_{ij} within the jth marginal sample $X_{1j}, \ldots, X_{nj}$.

ASSUMPTION 3.1. *There exists an estimator $\tilde{\theta}_n^* = t_n(\boldsymbol{R}_1^{(n)}, \ldots, \boldsymbol{R}_n^{(n)})$ such that, for all $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$, we have $\sqrt{n}(\tilde{\theta}_n^* - \theta) = O_P(1)$ under $\mathbb{P}_{\theta,F_1,\ldots,F_p}$.*

An obvious candidate for the initial $\sqrt{n}$-consistent estimator is the pseudo-likelihood estimator. If the copula parameter θ can be expressed as a smooth function of the correlation matrix $R(\theta)$, an alternative is to construct a minimum distance type estimator of θ using the normal scores rank correlations.

In what follows, $\tilde{\theta}_n$ denotes a discretized version of $\tilde{\theta}_n^*$, obtained by rounding $\tilde{\theta}_n^*$ to the grid $n^{-1/2}\mathbb{Z}^k$. This discretization, of course, does not disturb the $\sqrt{n}$-consistency; see also Remark 3.3(ii).

From Proposition 2.8, recall the efficient score function $\dot{\ell}_\theta^*$ and the efficient information matrix $I^*(\theta)$. Further, rescaled versions of the marginal empirical distribution functions are provided by

$$\hat{F}_{n,j}(x) = \frac{1}{n+1} \sum_{i=1}^{n} 1\{X_{ij} \leq x\}, \qquad x \in \mathbb{R},\ j = 1, \ldots, p.$$

The *one-step estimator* is then defined by

$$\hat{\theta}_n^{\mathrm{OSE}} = \tilde{\theta}_n + \frac{1}{n} \sum_{i=1}^{n} I^{*-1}(\tilde{\theta}_n)\, \dot{\ell}_\theta^*\big(\hat{F}_{n,1}(X_{i1}), \ldots, \hat{F}_{n,p}(X_{ip}); \tilde{\theta}_n\big). \tag{3.3}$$

The initial estimator being rank-based, the one-step estimator is rank-based too. In particular, $\hat{\theta}_n^{\mathrm{OSE}}$ is invariant with respect to strictly increasing transformations applied to each of the p variables.

The following theorem states that the proposed one-step estimator is efficient. This gives a positive answer to the question raised in Hoff, Niu and Wellner (2012) whether it is possible to develop a rank-based, semiparametrically efficient estimator for Gaussian copula models.

THEOREM 3.2. *Suppose that the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1. Also assume that the initial estimator $\tilde{\theta}_n^*$ satisfies Assumption 3.1. Then $\hat{\theta}_n^{\mathrm{OSE}}$ is an efficient, rank-based estimator of θ in the semiparametric Gaussian copula model $\mathcal{P}$, i.e., for all $F_1, \ldots, F_p \in \mathcal{F}_{ac}$ and $\theta \in \Theta$ we have (3.1) with $\hat{\theta}_n$ replaced by $\hat{\theta}_n^{\mathrm{OSE}}$ and thus, as $n \to \infty$,*

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{OSE}} - \theta) \xrightarrow{d} N_k(0, I^{*-1}(\theta)).$$

OUTLINE OF THE PROOF. We give a detailed proof in Appendix B and restrict ourselves here to a short outline. First we show that it suffices to prove the theorem for uniform marginals. Let $\boldsymbol{U}_i$, $i \in \mathbb{N}$, be i.i.d. random vectors with law P_θ under $\mathbb{P}_\theta$. Following the lines of the proof of the efficiency of *parametric* one-step estimators (Bickel et al., 1993, Theorem 2.5.2), we show that (3.1) holds if

(P1) for any sequence $\theta_n = \theta + h_n/\sqrt{n}$, with $h_n \in \mathbb{R}^k$ bounded, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_\theta^*(\boldsymbol{U}_i; \theta_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_\theta^*(\boldsymbol{U}_i; \theta) - I^*(\theta) h_n + o(1; \mathbb{P}_\theta);$$

(P2) for any sequence $\theta_n = \theta + h_n/\sqrt{n}$, with $h_n \in \mathbb{R}^k$ bounded, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_\theta^*\big(\hat{F}_{n,1}(U_{i1}), \ldots, \hat{F}_{n,p}(U_{ip}); \theta_n\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_\theta^*(\boldsymbol{U}_i; \theta_n) + o(1; \mathbb{P}_\theta).$$

We show that (P1) holds by exploiting the smoothness of the efficient score and Proposition A.10 in Van der Vaart (1988). Statement (P2) follows from a modification of Corollary 3.1 in Hoff, Niu and Wellner (2012). □

REMARK 3.3.

(i) The special structure (2.12) of the efficient score allows us to follow the lines of the proof of the efficiency of *parametric* one-step estimators. We thus do not need to use sample-splitting as in, e.g., Klaassen (1987).
(ii) Discretization of the initial estimator is needed in the efficiency proof but has little to no practical implications. Certainly, one should not bother to discretize estimators in practice: see pages 125 or 188 in Le Cam and Yang (1990) for a discussion of this point.

(iii) The update step (3.3) is simple to implement in practice. Regarding the calculation of the efficient information matrix in (2.19), recall that the inner product $\langle \cdot, \cdot \rangle_\theta$ is defined in (2.3) and that $\dot{S}_m(\theta) = -S(\theta)\,\dot{R}_m(\theta)\,S(\theta)$.

4. Quadratic influence functions. Structured Gaussian copula models are completely specified by the parametrization $\theta \mapsto R(\theta)$. This implies that conditions for, e.g., efficiency of a given estimator or adaptivity of the model at a certain value of θ can be expressed in terms of this parametrization. In addition, all relevant score and influence functions turn out to be quadratic in the Gaussianized observations in the sense of (4.2). In this section, we will give finite-dimensional algebraic conditions under which an estimator with such an influence function is regular (Proposition 4.6) or even efficient (Proposition 4.7). In addition, Proposition 4.8 gives a simple condition for adaptivity of the model at some value of θ.

For practice, the most relevant result in this section concerns the characterization of those Gaussian copula models for which the pseudo-likelihood estimator is efficient. Recall that the pseudo-likelihood estimator $\hat{\theta}_n^{\mathrm{PLE}}$ is defined in Genest, Ghoudi and Rivest (1995) as the maximizer of the function $\theta \mapsto \sum_{i=1}^n \log c_\theta(\hat{F}_{n,1}(X_{i1}), \ldots, \hat{F}_{n,p}(X_{ip}))$.

THEOREM 4.1. *Suppose that the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1. The pseudo-likelihood estimator $\hat{\theta}_n^{\mathrm{PLE}}$ is semiparametrically efficient at $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$ in the sense of (3.1) if and only if, for every $m = 1, \ldots, k$, the matrix*

$$L_m(\theta) - \frac{1}{2}\big(\mathrm{diag}(L_m(\theta))R(\theta) + R(\theta)\,\mathrm{diag}(L_m(\theta))\big) \tag{4.1}$$

with

$$L_m(\theta) = R(\theta)\ \mathrm{diag}\big(\dot{R}_m(\theta)S(\theta)\big)\ R(\theta)$$

belongs to the linear span of $\dot{R}_1(\theta), \ldots, \dot{R}_k(\theta)$.

The proof of Theorem 4.1 is given in Appendix C, after the proofs of the other results in this section, on which it depends. Theorem 4.1 immediately implies that the pseudo-likelihood estimator is efficient in the unrestricted model, in which each of the $p(p-1)/2$ off-diagonal entries of the correlation matrix is a parameter. Indeed, in the unrestricted model, the matrices $\dot{R}_1(\theta), \ldots, \dot{R}_k(\theta)$, with $k = p(p-1)/2$, span the space of symmetric matrices with zero diagonal, to which the matrix in (4.1) clearly belongs. In Section 5, efficiency of the pseudo-likelihood estimator will also be established exchangeable correlation matrices (Example 5.3) and for the versatile class of factor models (Example 5.5).

In the context of the present section, it is convenient to assume that all marginal distributions are uniform, so that the relevant nonparametric part of the tangent space becomes $\mathcal{T}_{\mathrm{P}_\theta}$. We maintain the notation $\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$ with $Z_j = \Phi^{-1}(U_j)$ for $j = 1, \ldots, p$, where $\boldsymbol{U}$ is the identity mapping on the space $(0,1)^p$, which is equipped with the probability measure P_θ induced by the Gaussian copula C_θ. The distribution of $\boldsymbol{Z}$ is then $N_p(0, R(\theta))$. Recall that $\mathrm{Sym}(p)$ is the space of real, symmetric $p \times p$ matrices.

4.1. *Quadratic functions and the tangent space.* For $A \in \mathrm{Sym}(p)$, define the function $q_A \in \mathrm{L}_2^0(\mathrm{P}_\theta)$ via

$$q_A(\boldsymbol{U}) = \frac{1}{2}\left(\boldsymbol{Z}'A\boldsymbol{Z} - \mathrm{E}_\theta[\boldsymbol{Z}'A\boldsymbol{Z}]\right). \tag{4.2}$$

We call a score or influence function of this form "quadratic" and "generated by A". For instance, the parametric score (2.2) for θ_m is generated by $A = -\dot{S}_m(\theta)$, while the semiparametrically efficient score (2.18) is generated by $A = D_\theta(\boldsymbol{g}_m(\theta)) - \dot{S}_m(\theta)$ with $D_\theta(\,\cdot\,)$ defined in (2.15). By Theorem 3.2, the one-step estimator has a quadratic influence function, and in the proof of Theorem 4.1, we will see that the pseudo-likelihood estimator has a quadratic influence function too.

First we characterize the matrices $A \in \mathrm{Sym}(p)$ that generate quadratic forms $q_A(\boldsymbol{U})$ belonging to $\mathcal{T}_{\mathrm{P}_\theta}$, the nonparametric part of the tangent space.

Lemma 4.2. *Suppose that the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1. The matrix $A \in \mathrm{Sym}(p)$ generates an element $q_A(\boldsymbol{U})$ of $\mathcal{T}_{\mathrm{P}_\theta}$ if and only if there exists a vector $\boldsymbol{b} \in \mathbb{R}^p$ such that $A = D_\theta(\boldsymbol{b})$.*

Lemma 4.2 motivates us to define a linear subspace of $\mathrm{Sym}(p)$,

$$\dot{T}_\theta = \left\{ D_\theta(\boldsymbol{b}) \,\middle|\, \boldsymbol{b} \in \mathbb{R}^p \right\},$$

whose elements generate quadratic scores in the nonparametric part, $\mathcal{T}_{\mathrm{P}_\theta}$, of the tangent space. It follows from Part A in the proof of Proposition 2.8 that the dimension of $\dot{T}_\theta$ is p.

Recall that we had endowed the space $\mathrm{Sym}(p)$ with the inner product $\langle \cdot, \cdot \rangle_\theta$ in (2.3). In the present notation, the inner product can be written as

$$\langle A, B \rangle_\theta = \mathrm{cov}_\theta\big(q_A(\boldsymbol{U}), q_B(\boldsymbol{U})\big), \qquad A, B \in \mathrm{Sym}(p). \tag{4.3}$$

It follows that the map $A \mapsto q_A(\boldsymbol{U})$ constitutes an isometry between $\mathrm{Sym}(p)$ and the linear subspace $\mathcal{Q}_{\mathrm{P}_\theta} = \{q_A(\boldsymbol{U}) \mid A \in \mathrm{Sym}(p)\}$ of $\mathrm{L}_2^0(\mathrm{P}_\theta)$. Lemma 4.2 then states that the intersection of $\mathcal{Q}_{\mathrm{P}_\theta}$ and $\mathcal{T}_{\mathrm{P}_\theta}$ is isometric to $\dot{T}_\theta$.

The construction of the one-step estimator (3.3) was based on the efficient score function, which in turn was found through a projection of the parametric score on (the orthocomplement of) $\mathcal{T}_{\mathrm{P}_\theta}$ (Proposition 2.8). For quadratic forms, the calculation of such projections is particularly simple.

COROLLARY 4.3. *Suppose that the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1 and let $\theta \in \Theta$ and $A \in \mathrm{Sym}(p)$. Statements (a) and (b) in Lemma 2.7 are both equivalent to*

(c) the matrix A is orthogonal to $\dot{T}_\theta$.

COROLLARY 4.4. *If the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1, the projection of $A \in \mathrm{Sym}(p)$ on $\dot{T}_\theta$ is equal to $D_\theta(\boldsymbol{b})$ and the projection of $q_A(\boldsymbol{U})$ on $\mathcal{T}_{\mathrm{P}_\theta}$ is equal to $q_{D_\theta(\boldsymbol{b})}(\boldsymbol{U})$, where $\boldsymbol{b} \in \mathbb{R}^p$ is the unique solution of*

$$\mathrm{diag}\big(R(\theta)\, A - \mathrm{diag}(\boldsymbol{b}) - R(\theta)\ \mathrm{diag}(\boldsymbol{b})\, S(\theta)\big) = \mathbf{0}. \tag{4.4}$$

Corollary 4.4 sheds new light on Proposition 2.8. The projection of the parametric score $\dot{\ell}_{\theta,m}(\boldsymbol{U};\theta) = q_{-\dot{S}_m(\theta)}(\boldsymbol{U})$ on $\mathcal{T}_{\mathrm{P}_\theta}$ is given by $\dot{\ell}_{\theta,m}(\boldsymbol{U};\theta) - \dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta) = q_{D_\theta(\boldsymbol{g}_m(\theta))}(\boldsymbol{U})$ with the vector $\boldsymbol{g}_m(\theta)$ given by (2.17). But the latter equation says that $-\boldsymbol{g}_m(\theta)$ is equal to the solution to (4.4) with the matrix A equal to $-\dot{S}_m(\theta)$.

Corollaries 4.3 and 4.4 greatly simplify all tangent space projection calculations. Any quadratic score or influence function $q_A(\boldsymbol{U})$ generated by a matrix A such that $\mathrm{diag}(R(\theta)A) = 0$ is automatically orthogonal to the infinite-dimensional space $\mathcal{T}_{\mathrm{P}_\theta}$ in $\mathrm{L}_2^0(\mathrm{P}_\theta)$. This property will be used extensively below in the investigation of regularity and efficiency of estimators and of adaptivity of the model.

REMARK 4.5. By studying rank-based likelihoods, Hoff, Niu and Wellner (2012) conclude that, with Gaussian marginals, the p unknown marginal variances generate the least-favorable directions. Indeed, using the calculations in their Theorem 4.1, one may verify directly that these marginal variances generate scores of the form $q_{D_\theta(\boldsymbol{b})}(\boldsymbol{U})$ with $\boldsymbol{b} \in \mathbb{R}^p$.

4.2. *Regularity.* The study of (semi)parametric efficiency is usually limited to *regular* estimators, which are estimators with the property that their limiting distribution, after proper centering and rescaling, is the same under any sequence of local alternatives (Van der Vaart, 2000, Section 23.5, p. 365). Consider an estimator $\hat{\theta}_n$ whose components are asymptotically linear with quadratic influence functions, i.e., for all $\theta \in \Theta$ and all $m = 1, \ldots, k$ there exists $A_m(\theta) \in \mathrm{Sym}(p)$ such that

$$\sqrt{n}(\hat{\theta}_{n,m} - \theta_m) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} q_{A_m(\theta)}(\boldsymbol{U}_i) + o(1; \mathbb{P}_\theta). \tag{4.5}$$

(Recall that we focus on rank-based estimators, so that we may, without loss of generality, assume that the margins are uniform.) For an asymptotically linear estimator, regularity is equivalent to the statement that the orthogonal projection of its influence function on the full tangent set is equal to the efficient influence function (Bickel et al., 1993, Proposition 3.3.1). Since the full tangent set (2.11) of the Gaussian copula model is spanned by the nonparametric part $\mathcal{T}_{\mathrm{P}_\theta}$ and the parametric scores $\dot{\ell}_{\theta,m}(\boldsymbol{U};\theta)$, $m = 1, \ldots, k$, regularity of $\hat{\theta}_n$ in (4.5) is equivalent to

$$q_{A_m(\theta)}(\boldsymbol{U}) \perp \mathcal{T}_{\mathrm{P}_\theta}, \tag{4.6}$$

$$\mathrm{cov}_\theta\big(q_{A_m(\theta)}(\boldsymbol{U}), \dot{\ell}_{\theta,m'}(\boldsymbol{U})\big) = \delta_{m=m'}, \tag{4.7}$$

for all $m, m' \in \{1, \ldots, k\}$. These equations pose restrictions on $A_m(\theta)$, characterized by the following proposition.

Proposition 4.6. *Suppose that the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1. Let $\hat{\theta}_n$ be an estimator sequence satisfying (4.5) for every $\theta \in \Theta$ and $m = 1, \ldots, k$. Then $\hat{\theta}_n$ is regular at P_θ if and only if, for all $m, m' \in \{1, \ldots, k\}$,*

$$A_m(\theta) \perp \dot{T}_\theta, \tag{4.8}$$

$$\langle A_m(\theta), -\dot{S}_{m'}(\theta)\rangle_\theta = \delta_{m=m'}, \tag{4.9}$$

or equivalently, if and only if, for all $m, m' \in \{1, \ldots, k\}$,

$$\mathrm{diag}\big(R(\theta)\, A_m(\theta)\big) = \mathbf{0}, \tag{4.10}$$

$$\mathrm{tr}\big(A_m(\theta)\, \dot{R}_{m'}(\theta)\big) = 2\, \delta_{m=m'}. \tag{4.11}$$

A matrix $A_m(\theta)$ satisfying the two conditions (4.10) and (4.11) is called a *regular influence matrix* for θ_m at θ. For the one-step estimator, for instance, the influence matrices are

$$A^*_m(\theta) = \sum_{m'=1}^{k} (I^{*-1}(\theta))_{mm'} \big(D(\boldsymbol{g}_{m'}(\theta)) - \dot{S}_{m'}(\theta)\big), \tag{4.12}$$

which can be seen from the right-hand side of (3.1) and the expression for the efficient score in Proposition 2.8. By construction, the matrices $A^*_m(\theta)$ in (4.12) are regular influence matrices. Of course, this property also follows more generally from the above description of regularity and the fact that the one-step estimator is asymptotically linear in the efficient influence function. In the proof of Theorem 4.1, we will check that the pseudo-likelihood estimator is regular too.

For each $m = 1, \ldots, k$ and $\theta \in \Theta$, equations (4.8) and (4.9) pose $p + k$ independent linear restrictions on $A_m(\theta)$. Indeed, the dimension of $\dot{T}_p(\theta)$ is p while the matrices $-\dot{S}_1(\theta), \ldots, -\dot{S}_m(\theta)$ are linearly independent and are not contained in $\dot{T}_p(\theta)$; see Part A of the proof of Proposition 2.8. It follows that the set of regular influence matrices for a given component of θ is an affine subspace of $\mathrm{Sym}(p)$ of dimension $p(p+1)/2-(p+k) = p(p-1)/2-k$.

4.3. *Efficiency.* In the unrestricted model, each pairwise correlation being a parameter, there are $k = p(p-1)/2$ parameters, so that there is a unique regular influence matrix for each component θ_m, which then must be equal to the efficient one, $A^*_m(\theta)$. In Example 5.1, this matrix will be identified with the one generating the influence function of the rank correlation estimator, proving efficiency of the latter.

In structured Gaussian copula models, i.e., those models where the matrices $\dot{S}_m(\theta)$, $m = 1, \ldots, k$, span a subspace of dimension $k = \dim(\Theta)$ less than $p(p-1)/2$, multiple regular quadratic influence functions exist. Within this set of regular influence matrices, the efficient influence matrices $A^*_m(\theta)$ admit the following characterization.

PROPOSITION 4.7. *Suppose that the parametrization $\theta \mapsto R(\theta)$ satisfies Assumption 2.1. Consider a regular, rank-based estimator $\hat{\theta}_n$ for which there exist matrices $B_1(\theta), \ldots, B_k(\theta) \in \mathrm{Sym}(p)$ such that the influence functions of all components $\hat{\theta}_{n,1}, \ldots, \hat{\theta}_{n,k}$ are quadratic and are given by matrices belonging to the linear span of $B_1(\theta), \ldots, B_k(\theta)$. Write $B_{m,R}(\theta) = R(\theta)B_m(\theta)R(\theta)$. Then $\hat{\theta}_n$ is efficient at P_θ in the sense of (3.1) if and only if, for each m, the matrix*

$$B_{m,R}(\theta) - \frac{1}{2}\big(\mathrm{diag}(B_{m,R}(\theta))R(\theta) + R(\theta)\,\mathrm{diag}(B_{m,R}(\theta))\big) \tag{4.13}$$

belongs to the linear span of $\dot{R}_1(\theta), \ldots, \dot{R}_k(\theta)$.

The characterization is based essentially on the fact that, by Proposition 4.6, the efficient influence matrices $A^*_m(\theta)$ are the only regular influence matrices that belong to the space

$$\operatorname{span}\bigl(\dot{T}_\theta \cup \{\dot{S}_1(\theta), \ldots, \dot{S}_k(\theta)\}\bigr). \tag{4.14}$$

Moreover, the projection of any other regular influence matrix $A_m(\theta)$ on the space (4.14) is equal to $A_m^*(\theta)$. The efficiency criterion for the pseudo-likelihood estimator in Theorem 4.1 is essentially a particular case of Proposition 4.7.

4.4. *Adaptivity.* If the efficient information matrix $I^*(\theta)$ is equal to the Fisher information matrix $I(\theta)$ in the parametric model with known margins, the fact of not knowing the margins does not make a difference asymptotically for the efficient estimation of θ. For semiparametric Gaussian copula models, there is a simple criterion for the occurrence of this phenomenon, called adaptivity.

PROPOSITION 4.8. *The semiparametric Gaussian copula model is adaptive at* $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$ *if and only if*

$$\operatorname{diag}\bigl(R(\theta)\dot{S}_m(\theta)\bigr) = 0, \qquad m = 1, \ldots, k. \tag{4.15}$$

Obviously, adaptivity always occurs at the independence copula (assuming it belongs to the model) as in that case $R(\theta)$ and $S(\theta)$ equal the identity matrix and $\operatorname{diag} \dot{S}_m(\theta) = -\operatorname{diag} \dot{R}_m(\theta) = 0$. See Example 5.6 for a model which is adaptive at a copula different from the independence one. Still, adaptivity is the exception rather than the rule: most of the time, not knowing the margins makes inference on the copula parameter θ more difficult.

5. Examples and simulations. This section presents some analytical and numerical results for the one-step and pseudo-likelihood estimators for a number of correlation structures. For some cases, the pseudo-likelihood estimator is efficient (unrestricted model, exchangeable model, factor model, Toeplitz model in $p = 3$), sometimes it is almost efficient (circular model), and sometimes it is quite inefficient (Toeplitz model in $p = 4$). The results of a Monte Carlo study indicate that the asymptotic approximations to the finite-sample distributions are excellent for the one-step estimator. Adaptivity almost never occurs, except at independence and at a contrived example. The simulation study is implemented in MATLAB 2012a and the code is available upon request.

EXAMPLE 5.1 (Unrestricted model). In the full, unrestricted model, there are $k = p(p-1)/2$ parameters, which can be identified with the correlations r_{ij} between Z_i and Z_j for $i = 2, \ldots, p$ and $j = 1, \ldots, i-1$. Efficiency of the normal scores rank correlation estimator for the unrestricted Gaussian copula model was already observed in Klaassen and Wellner (1997). Here we obtain this result within our general algebraic analysis of (possibly) structured Gaussian copula models.

First one can check that for arbitrary Gaussian copula models, the pseudo-score equations $\sum_{i=1}^n \dot{\ell}_{\theta,m}(\hat{F}_{n,1}(X_{i1}),\ldots,\hat{F}_{n,p}(X_{ip});\theta)=0$ are equivalent to

$$\operatorname{tr}\big(\dot{S}_m(\theta)\,(R(\theta)-\hat{R}_n)\big)=0, \qquad m=1,\ldots,k, \tag{5.1}$$

where $\hat{R}_n$ is the $p\times p$ matrix

$$\hat{R}_n=\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{Z}}_{n,i}\,\hat{\boldsymbol{Z}}'_{n,i} \tag{5.2}$$

and $\hat{Z}_{n,ij}=\Phi^{-1}(\hat{F}_{n,j}(X_{ij}))$ for $i=1,\ldots,n$ and $j=1,\ldots,p$. It follows that for the unrestricted model, the pseudo-likelihood estimator is given by $\hat{R}_n$ itself. The normal scores rank correlation estimator is just $\hat{R}_n/\sigma_n^2$ with $\sigma_n^2=n^{-1}\sum_{i=1}^n[\Phi^{-1}(i/(n+1))]^{-2}=1+O(n^{-1}\log n)$, and is therefore asymptotically equivalent to the pseudo-likelihood estimator. But the latter was already shown to be efficient in the paragraph following Theorem 4.1.

The fact that for the unrestricted model, the normal scores rank correlation estimator, the pseudo-likelihood estimator and the one-step estimator are all asymptotically equivalent is due to the uniqueness of the regular influence matrices for the correlation parameters. Indeed, fix $i=2,\ldots,p$ and $j=1,\ldots,i-1$ and consider a regular influence matrix $A_{ij}\in\operatorname{Sym}(p)$ for r_{ij}. The derivative matrix $\dot{R}_{ij}$ of R with respect to r_{ij} equals the $p\times p$ matrix whose elements are all zero, except for the (i,j)th and (j,i)th elements, which equal unity. In view of (4.11), regularity implies, for every $i'=2,\ldots,p$ and $j'=1,\ldots,i'-1$,

$$2\,\delta_{i=i',\,j=j'}=\operatorname{tr}\big(A_{ij}\,\dot{R}_{i'j'}\big)=2\,(A_{ij})_{i'j'}\,.$$

Hence, the regularity conditions (4.11) imply that all off-diagonal elements of A_{ij} are zero, but for $(A_{ij})_{ij}=(A_{ij})_{ji}=1$. Subsequently, (4.10) implies $(A_{ij})_{ii}=(A_{ij})_{jj}=-r_{ij}$, while the other diagonal elements are zero. Consequently, in the unrestricted model, there is only one quadratic regular influence function for each pair (i,j), which, therefore, must be equal to the efficient influence function in (3.2).

Example 5.2 (Toeplitz model). A Toeplitz model for the correlation matrix arises if there are $k=p-1$ parameters $\theta_1,\ldots,\theta_{p-1}$ such that $r_{ij}=$

$\theta_{|i-j|}$, where r_{ij} is the correlation between Z_i and Z_j. In dimension $p = 3$, for instance, the model is $R_{12}(\theta_1, \theta_2) = R_{23}(\theta_1, \theta_2) = \theta_1$ and $R_{13}(\theta_1, \theta_2) = \theta_2$, and a brute-force calculation using the computer algebra system Maxima (version 5.29.1, http://maxima.sourceforge.net) shows that the inverse of the efficient information matrix is equal to the asymptotic covariance matrix of the pseudo-likelihood estimator, calculated from (C.1)–(C.2). The explicit formula for $I^{*-1}(\theta)$ is given in (D.2).

In dimension $p = 4$, however, the pseudo-likelihood estimator is no longer efficient. The analytic expressions for the asymptotic variance matrices of the one-step and pseudo-likelihood estimators are too long to display, but for specific values of $\theta = (\theta_1, \theta_2, \theta_3)'$, they can easily be computed numerically. This allows to search for values for θ in which the asymptotic relative efficiency of the pseudo-likelihood estimator is particularly low. At $\theta_* = (0.4945460, -0.4592764, -0.8462492)'$, for instance, the asymptotic relative efficiencies of the pseudo-likelihood estimator with respect to the information bound are equal to $(18.3\%, 19.8\%, 96.9\%)$. The finite-sample performance of the one-step and pseudo-likelihood estimators were compared using 15,000 Monte Carlo samples of sizes $n = 50$ and $n = 250$. The boxplots of the estimation errors, shown in Figure 1, confirm that for the components θ_1 and θ_2, the pseudo-likelihood estimator is quite inefficient at $\theta = \theta_*$.

Example 5.3 (Exchangeable model). The exchangeable Gaussian copula model is a one-parameter model in which all off-diagonal entries of the $p \times p$ matrix $R(\theta)$ are equal to the same value of θ between $-1/(p-1)$ and 1. Efficiency of the pseudo-likelihood estimator for θ in dimension $p = 4$ was established in Hoff, Niu and Wellner (2012) and can, for general p, be verified easily using Theorem 4.1; see Appendix D for some algebraic details. Using the computer algebra system Maxima, we calculated the optimal asymptotic variance for regular estimators of θ in dimensions three and four:

$$I^{*-1}(\theta) = \begin{cases} \frac{1}{3}(\theta-1)^2(2\theta+1)^2 & \text{if } p = 3, \\ \frac{1}{6}(\theta-1)^2(3\theta+1)^2 & \text{if } p = 4. \end{cases}$$

In contrast, if the margins are known, the optimal asymptotic variance reduces to

$$I^{-1}(\theta) = \begin{cases} I^{*-1}(\theta)/(1+2\theta^2) & \text{if } p = 3, \\ I^{*-1}(\theta)/(1+3\theta^2) & \text{if } p = 4, \end{cases}$$

so that adaptivity occurs at independence ($\theta = 0$) only.

We assessed the finite-sample performance of the one-step and pseudo-likelihood estimators by 15,000 Monte Carlo samples of sizes $n = 50$ and

$n = 250$ in dimension $p = 3$ for θ in a grid of values between $-1/2$ and 1; see Figure 2. Even for $n = 50$, the finite-sample variance of the one-step estimator is well approximated by its limit. For the pseudo-likelihood estimator, the convergence is slower and its variance in finite samples is a bit larger. The biases of the both estimators are of comparable order and are generally negligible relative to the variances.

In order to see the impact of the dimension, we also compared the one-step and pseudo-likelihood estimators in dimension $p = 100$ for 15,000 Monte Carlo samples of size $n = 50$ at a true parameter value $\theta = .25$. Boxplots of the estimation errors are shown in Figure 3. Although the variances of both estimators are about the same, the pseudo-likelihood estimator suffers from a large bias, whereas the one-step estimator remains centered around the true value.

EXAMPLE 5.4 (Circular model). The circular model is presented in Hoff, Niu and Wellner (2012) as a one-parameter Gaussian copula model where the pseudo-likelihood estimator is not efficient. It is defined in dimension $p = 4$ by

$$R(\theta) = \begin{pmatrix} 1 & \theta & \theta^2 & \theta \\ \theta & 1 & \theta & \theta^2 \\ \theta^2 & \theta & 1 & \theta \\ \theta & \theta^2 & \theta & 1 \end{pmatrix}, \qquad -1 < \theta < 1.$$

The optimal asymptotic variance if margins are unknown, the asymptotic variance of the pseudo-likelihood estimator, and the inverse Fisher information if margins are known can be computed explicitly:

$$\begin{aligned} I^{*-1}(\theta) &= \frac{1}{4}(1-\theta^2)^2, \\ \sigma^2_{\mathrm{PLE}} &= I^{*-1}(\theta)\left(1 + \frac{2\theta^6}{(1+2\theta^2)^2}\right), \\ I^{-1}(\theta) &= I^{*-1}(\theta)/(1+2\theta^2). \end{aligned}$$

Even though the pseudo-likelihood estimator is not efficient, its asymptotic relative efficiency is close to 100%, except for θ close to 1 or -1. Adaptivity occurs at independence ($\theta = 0$) only.

We assessed the finite-sample performance of the one-step and pseudo-likelihood estimators by 15,000 Monte Carlo samples of sizes $n = 50$ and $n = 250$ for θ in a grid of values between -1 and 1; see Figure 4. The results are comparable to the ones for the exchangeable model.

EXAMPLE 5.5 (Factor models). Factor models are a popular tool for dimension reduction. In dimension $p \geq 3$, set

$$\boldsymbol{Z} = \theta \boldsymbol{W} + \left(I_p - \operatorname{diag}\left(\theta\theta'\right)\right)^{1/2} \boldsymbol{\varepsilon}, \tag{5.3}$$

where θ denotes a $p \times q$ matrix, $q < p$, and where $\boldsymbol{W}$ and $\boldsymbol{\varepsilon}$ are independent random vectors of dimensions q and p, respectively, such that $(\boldsymbol{W}', \boldsymbol{\varepsilon}')' \sim N_{q+p}(\mathbf{0}, I_{p+q})$. The parameter space Θ is an open subset of $\{\theta \in \mathbb{R}^{p\times q} \mid (\theta\theta')_{jj} < 1,\ j = 1, \ldots, p\}$; in particular, $I_p - \operatorname{diag}(\theta\theta')$ is a diagonal matrix with positive elements on the diagonal. As the variance matrix of $\boldsymbol{Z}$ is a correlation matrix, (5.3) defines a Gaussian copula model with

$$R(\theta) = \theta\theta' + \left(I_p - \operatorname{diag}\left(\theta\theta'\right)\right) = I_p + \operatorname{rd}\left(\theta\theta'\right) \tag{5.4}$$

in terms of the diagonal-removal operator $\operatorname{rd}(A) = A - \operatorname{diag}(A)$ for $A \in \mathbb{R}^{p\times p}$. The parameter θ is not identifiable: if O is an orthogonal $q \times q$ matrix, then $R(\theta O) = R(\theta)$. Still, by Corollary D.2 in Appendix D, we may study efficiency of the pseudo-likelihood estimator for the parameter ν in any reparametrization $\nu \mapsto \theta(\nu)$ satisfying Assumption D.1, which makes the model identified, by the criterion in Theorem 4.1. After some calculations, which are detailed in the appendix, the criterion can be shown to be satisfied, confirming the efficiency of the pseudo-likelihood estimator for Gaussian factor copula models.

EXAMPLE 5.6 (Adaptivity). Proposition 4.8 gives a necessary and sufficient criterion for adaptivity of a Gaussian copula model at a certain value of θ. Adaptivity always occurs at the independence copula but, apart from this, is the exception rather than the rule. With some trial and error, other (artificial) examples can be constructed. For instance, the one-parameter model in dimension $p = 3$ given by $R_{12}(\theta) = R_{13}(\theta) = \theta^2 + .5$ and $R_{23}(\theta) = \theta + .25$, for θ in a neighbourhood of 0, can be verified to be adaptive at $\theta = 0$.

6. Conclusion.

The present paper provides a semiparametrically efficient, rank-based estimator for the copula parameter in structured Gaussian copula models under mild, straightforward conditions on the parametrization $\theta \mapsto R(\theta)$ of the correlation matrix. This gives a positive answer to the conjecture formulated in Hoff, Niu and Wellner (2012) that in Gaussian copula models, semiparametrically efficient, rank-based estimators do exist.

The estimator is based on the analysis of the tangent space structure of the model and the explicit calculation of the efficient score function. We expect that similar techniques can be developed for non-Gaussian copula models too. Simulations indicate that the large-sample distribution provides an accurate approximation to the finite-sample distribution of the estimator, even in large dimensions (up to $p = 100$).

Moreover, we show that inference in structured Gaussian copula models can be studied using a convenient finite-dimensional algebraic representation of relevant scores and influence functions. This leads to straightforward conditions to verify the regularity or the efficiency of existing estimators. In particular, we provide a convenient necessary and sufficient condition for the semiparametric efficiency of the pseudo-likelihood estimator. It follows, e.g., that this estimator is efficient in models that exhibit a suitable factor structure. However, we also provide examples where its relative efficiency can be as low as 20%. Several other concrete examples complement the analysis.

Acknowledgments. The authors gratefully acknowledge extensive discussions with John H.J. Einmahl (Tilburg University) and Christian Genest (McGill University). Johan Segers' research was supported by contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" and by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy).

APPENDIX A: PROOFS FOR SECTION 2

Recall Hardy's inequality (Hardy, Littlewood and Pólya, 1934): for integrable $h : (0, 1) \to \mathbb{R}$,

$$\int_0^1 \left(\frac{1}{\lambda} \int_0^\lambda h(\lambda) \, \mathrm{d}\lambda \right)^2 \mathrm{d}\lambda \le 4 \int_0^1 h^2(\lambda) \, \mathrm{d}\lambda. \tag{A.1}$$

The following version also allows for explosions at 1.

Proposition A.1 (two-sided Hardy). *For integrable $h : (0, 1) \to \mathbb{R}$ such that $\int_0^1 h(\lambda) \, \mathrm{d}\lambda = 0$, we have*

$$\int_0^1 \left(\frac{1}{u(1-u)} \int_0^u h(\lambda) \, \mathrm{d}\lambda \right)^2 \mathrm{d}u \le 16 \int_0^1 h^2(\lambda) \, \mathrm{d}\lambda. \tag{A.2}$$

Proof. Write $\frac{1}{u(1-u)} = \frac{1}{u} + \frac{1}{1-u}$, note that $\int_0^u h(\lambda) \, \mathrm{d}\lambda = -\int_u^1 h(\lambda) \, \mathrm{d}\lambda$, and apply Hardy's inequality (A.1) twice. □

Proof of Lemma 2.4. *Proof of (a).* The result for $\mathcal{O}_\theta$ follows if we show the result for the operators $\mathcal{O}_{\theta,j}$ for $j = 1, \dots, p$. For $h \in \mathrm{L}_2^0[0,1]$, we have

$$\begin{aligned}
&\mathrm{E}_\theta\big[\big([\mathcal{O}_{\theta,j}h](\boldsymbol{U})\big)^2\big] \\
&= \mathrm{E}_\theta[h^2(U_j)] + \mathrm{E}_\theta[\dot\ell_j^2(\boldsymbol{U};\theta)\, H^2(U_j)] + 2\,\mathrm{E}_\theta[h(U_j)\,\dot\ell_j(\boldsymbol{U};\theta)\, H(U_j)] \\
&= \int_0^1 h^2(u)\,\mathrm{d}u + \mathrm{E}_\theta[I_{jj}(U_j;\theta)\, H^2(U_j)],
\end{aligned}$$

where we used (2.7). From the bound (2.10) and Proposition A.1 we obtain

$$\begin{aligned}
\mathrm{E}_\theta[I_{jj}(U_j;\theta)\, H^2(U_j)] &\le M_\theta \int_0^1 \left(\frac{1}{u(1-u)}\int_0^u h(\lambda)\,\mathrm{d}\lambda\right)^2 \mathrm{d}u \\
&\le 16 M_\theta \int_0^1 h^2(\lambda)\,\mathrm{d}\lambda.
\end{aligned}$$

Hence

$$\mathrm{E}_\theta\big[\big([\mathcal{O}_{\theta,j}h](\boldsymbol{U})\big)^2\big] \le (1 + 16M_\theta)\int_0^1 h^2(\lambda)\,\mathrm{d}\lambda,$$

i.e. $\|\mathcal{O}_{\theta,j}\| \le (1+16M_\theta)^{1/2}$, which shows that $\mathcal{O}_{\theta,j} : \mathrm{L}_2^0[0,1] \to \mathrm{L}_2(\mathrm{P}_\theta)$ is a bounded operator. The claim that the range of $\mathcal{O}_{\theta,j}$ is a subset of $\mathrm{L}_2^0(\mathrm{P}_\theta)$ follows from (2.7).

Proof of (b). Let $\boldsymbol{h} = (h_1, \dots, h_p) \in (\mathrm{L}_2^0[0,1])^p$, $j = 1, \dots, p$ and $u_j \in (0,1)$. We will show that, for $i = 1, \dots, p$,

$$\mathrm{E}_\theta\big[[\mathcal{O}_{\theta,i}h_i](\boldsymbol{U}) \,\big|\, U_j = u_j\big] = \begin{cases} h_j(u_j) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{A.3}$$

This will imply that

$$\mathrm{E}_\theta\big[[\mathcal{O}_\theta\boldsymbol{h}](\boldsymbol{U}) \,\big|\, U_j = u_j\big] = \sum_{i=1}^p \mathrm{E}_\theta\big[[\mathcal{O}_{\theta,i}h_i](\boldsymbol{U}) \,\big|\, U_j = u_j\big] = h_j(u_j)$$

and thus that $\mathcal{O}_\theta$ is one-to-one with inverse ${\mathcal{O}_\theta}^{-1}$ as stated in the lemma. Boundedness of ${\mathcal{O}_\theta}^{-1}$ then follows from

$$\int_0^1 \big(\mathrm{E}_\theta[f(\boldsymbol{U}) \mid U_j = u_j]\big)^2\,\mathrm{d}u_j \le \int_0^1 \mathrm{E}_\theta[f(\boldsymbol{U})^2 \mid U_j = u_j]\,\mathrm{d}u_j = \mathrm{E}_\theta[f(\boldsymbol{U})^2]$$

for $f \in \mathrm{L}_2^0(\mathrm{P}_\theta)$ and $j = 1, \dots, p$.

We show (A.3). The case $i = j$ is immediate from (2.7):

$$\begin{aligned}
\mathrm{E}_\theta\big[[\mathcal{O}_{\theta,j}h_j](\boldsymbol{U}) \,\big|\, U_j = u_j\big] &= \mathrm{E}_\theta\big[h(U_j) + \dot\ell_j(\boldsymbol{U};\theta)\, H_j(U_j) \,\big|\, U_j = u_j\big] \\
&= h(u_j) + \mathrm{E}_\theta[\dot\ell_j(\boldsymbol{U};\theta) \mid U_j = u_j]\, H_j(u_j) \\
&= h(u_j).
\end{aligned}$$

Next suppose $i \neq j$. Since the marginal distribution of U_j is just the uniform distribution on $(0, 1)$, conditional expectations given $U_j = u_j$ can be written as $(p-1)$-dimensional integrals

$$\mathrm{E}_\theta[f(\boldsymbol{U}) \mid U_j = u_j] = \int_{(0,1)^{p-1}} f(\boldsymbol{u})\, c_\theta(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}_{-j},$$

where $\boldsymbol{u}_{-j} = (u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_p)'$. We find

$$\mathrm{E}_\theta\big[[\mathcal{O}_{\theta,i} h_i](\boldsymbol{U}) \,\big|\, U_j = u_j\big] = \int_{(0,1)^{p-1}} \{h_i(u_i) + \dot{\ell}_i(\boldsymbol{u};\theta)\, H_i(u_i)\}\, c_\theta(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}_{-j}.$$

By Fubini's theorem, the integral over $\boldsymbol{u}_{-j}$ can be written as a double integral, the inner integral being over u_i and the outer integral over the $p-2$ remaining variables (if $p = 2$, there is no outer integral). It suffices to show that the inner integral is zero, i.e. that

$$\int_0^1 \{h_i(u_i) + \dot{\ell}_i(\boldsymbol{u};\theta)\, H_i(u_i)\}\, c_\theta(\boldsymbol{u})\, \mathrm{d}u_i = 0. \tag{A.4}$$

To proceed, note that for fixed $\boldsymbol{u}_{-i} \in (0, 1)^{p-1}$, the function $u_i \mapsto c_\theta(\boldsymbol{u})$ converges to a finite value as u_i tends to 0 or 1. Indeed, check the formula for c_θ in (2.1) and observe that the ith diagonal element of $S(\theta)$ is larger than or equal to one. [The ith diagonal element of $S(\theta)$ is equal to $1/(1 - r_i(\theta)^2)$, where $r_i(\theta)^2$ is the squared multiple correlation of Z_i with the other variables.] If the ith diagonal element is larger than one, then the limits of $c_\theta(\boldsymbol{u})$ as $u_i \to 0$ or $u_i \to 1$ are equal to zero, while if the ith diagonal element is equal to one, then U_i is independent of the other variables and $c_\theta(\boldsymbol{u})$ is constant in u_i.

This property of the Gaussian copula density allows for a simple proof of (A.4) using Fubini's theorem:

$$\begin{aligned}
\int_0^1 \dot{\ell}_i(\boldsymbol{u};\theta)\, H_i(u_i)\, c_\theta(\boldsymbol{u})\, \mathrm{d}u_i &= \int_{u_i=0}^1 \dot{c}_i(\boldsymbol{u};\theta) \int_{t=0}^{u_i} h_i(t)\, \mathrm{d}t\, \mathrm{d}u_i \\
&= \int_{t=0}^1 h_i(t) \int_{u_i=t}^1 \dot{c}_i(\boldsymbol{u};\theta)\, \mathrm{d}u_i \mathrm{d}t \\
&= \int_0^1 h_i(t)\, \{c_\theta(\boldsymbol{u})|_{u_i=1} - c_\theta(\boldsymbol{u})|_{u_i=t}\}\, \mathrm{d}t \\
&= -\int_0^1 h_i(u_i)\, c_\theta(\boldsymbol{u})\, \mathrm{d}u_i.
\end{aligned}$$

where we used the property $\int_0^1 h_i(t)\, \mathrm{d}t = 0$. This concludes the proof of (A.4) and thus of (A.3) for $i \neq j$. □

PROOF OF LEMMA 2.7. *(a) implies (b).* Fix $j \in \{1, \dots, p\}$ and consider the vector $\boldsymbol{h} \in (\mathrm{L}_2^0[0,1])^p$ given by $h_j(u) = 1 - z^2$, with $z = \Phi^{-1}(u)$ and $u \in (0,1)$, while $h_i(u) = 0$ if $i \neq j$. The primitive function of h_j is $H_j(u) = z\,\varphi(z)$. Then, using (2.6),

$$\begin{aligned}
[\mathcal{O}_\theta \boldsymbol{h}](\boldsymbol{U}) = [\mathcal{O}_{\theta,j} h_j](\boldsymbol{U}) &= 1 - Z_j^2 + \dot{\ell}_j(\boldsymbol{U};\theta)\, Z_j\, \varphi(Z_j) \\
&= 1 - \sum_{i=1}^p S_{ij}(\theta)\, Z_i\, Z_j \\
&= 1 - \boldsymbol{Z}'\, S(\theta)\, \mathrm{diag}(\boldsymbol{e}_j)\, \boldsymbol{Z}.
\end{aligned} \tag{A.5}$$

Orthogonality implies

$$\begin{aligned}
0 &= \mathrm{cov}_\theta\big(\boldsymbol{Z}' A \boldsymbol{Z},\, \boldsymbol{Z}'\, S(\theta)\, \mathrm{diag}(\boldsymbol{e}_j)\, \boldsymbol{Z}\big) \\
&= 2\, \mathrm{tr}\big(R(\theta)\, A\, R(\theta)\, S(\theta)\, \mathrm{diag}(\boldsymbol{e}_j)\big) = 2\big(R(\theta)\, A\big)_{jj},
\end{aligned}$$

yielding (b).

(b) implies (a). Let $h \in \mathrm{L}_2^0[0,1]$ and $j = 1, \dots, p$. Writing $a(\boldsymbol{U}) = \boldsymbol{Z}' A \boldsymbol{Z}$, it suffices to show that

$$\mathrm{E}_\theta\big[a(\boldsymbol{U})\, [\mathcal{O}_{\theta,j} h](\boldsymbol{U})\big] = 0.$$

The expectation on the left-hand side is equal to

$$\mathrm{E}_\theta\big[a(\boldsymbol{U})\, [\mathcal{O}_{\theta,j} h](\boldsymbol{U})\big] = \mathrm{E}_\theta\big[a(\boldsymbol{U})\, h(U_j)\big] + \mathrm{E}_\theta\big[a(\boldsymbol{U})\, \dot{\ell}_j(\boldsymbol{U};\theta)\, H(U_j)\big],$$

with $H(u) = \int_0^u h(\lambda)\, \mathrm{d}\lambda$. Let

$$\dot{a}_j(\boldsymbol{u}) = \frac{\partial}{\partial u_j} a(\boldsymbol{u}) = \frac{2\boldsymbol{z}'\, A_{\cdot j}}{\varphi(z_j)},$$

with $A_{\cdot j}$ the jth column of A. Statement (b) implies that

$$\mathrm{E}_\theta[\dot{a}_j(\boldsymbol{U}) \mid U_j = u_j] = \frac{2z_j}{\varphi z_j} \sum_{i=1}^p R_{ij} A_{ij} = \frac{2z_j}{\varphi z_j} (R(\theta) A)_{jj} = 0.$$

Writing $\boldsymbol{u}_{-j} = (u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_p) \in (0,1)^{p-1}$, it follows that

$$\begin{aligned}
&\frac{\partial}{\partial u_j} \int_{(0,1)^{p-1}} a(\boldsymbol{u})\, c(\boldsymbol{u};\theta)\, \mathrm{d}\boldsymbol{u}_{-j} \\
&= \int_{(0,1)^{p-1}} \frac{\partial}{\partial u_j} \{a(\boldsymbol{u})\, c(\boldsymbol{u};\theta)\}\, \mathrm{d}\boldsymbol{u}_{-j} \\
&= \int_{(0,1)^{p-1}} \dot{a}_j(\boldsymbol{u})\, c(\boldsymbol{u};\theta)\, \mathrm{d}\boldsymbol{u}_{-j} + \int_{(0,1)^{p-1}} a(\boldsymbol{u})\, \dot{c}_j(\boldsymbol{u};\theta)\, \mathrm{d}\boldsymbol{u}_{-j} \\
&= \int_{(0,1)^{p-1}} a(\boldsymbol{u})\, \dot{\ell}_j(\boldsymbol{u};\theta)\, c(\boldsymbol{u};\theta)\, \mathrm{d}\boldsymbol{u}_{-j}.
\end{aligned}$$

The interchanging of the derivative and the integral in the first equality sign can be justified, for instance, by switching to standard normal marginals. Alternatively, since $a(\boldsymbol{u})$ is quadratic in $\boldsymbol{z}$ and since $\dot{\ell}_j(\boldsymbol{u};\theta)$ is linear in $\boldsymbol{z}$, all integrals can be calculated explicitly, confirming their equality.

By partial integration and by the equality in the previous display,

$$\begin{aligned}\mathrm{E}_\theta\big[a(\boldsymbol{U})\,h(U_j)\big] &= \int h(u_j)\left(\int a(\boldsymbol{u})\,c_\theta(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}_{-j}\right)\mathrm{d}u_j \\ &= -\int H(u_j)\int a(\boldsymbol{u})\,\dot{\ell}_j(\boldsymbol{u};\theta)\,c_\theta(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}_{-j}\,\mathrm{d}u_j \\ &= -\,\mathrm{E}_\theta\big[a(\boldsymbol{U})\,\dot{\ell}_j(\boldsymbol{U};\theta)\,H(U_j)\big],\end{aligned}$$

yielding (a). Here we used that $\left[H(u_j)\int a(\boldsymbol{u})\,c_\theta(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}_{-j}\right]\Big|_{u_j=0}^{1} = 0$, which is justified by the fact that, for general $B \in \mathrm{Sym}(p)$,

$$\mathrm{E}_\theta[\boldsymbol{Z}'B\boldsymbol{Z} \mid U_j = u_j] = \mathrm{tr}(RB) + (RBR)_{jj}(Z_j^2 - 1)$$

together with the elementary bound

$$|H(u)| \le \min\{u^{1/2}, (1-u)^{1/2}\}\left(\int_0^1 h^2(\lambda)\,\mathrm{d}\lambda\right)^{1/2} \tag{A.6}$$

for $u \in (0,1)$, which in turn follows from Cauchy–Schwartz. □

Proof of Proposition 2.8. The proof consists of four parts.

Part A. Consider the map

$$\mathrm{Sym}(p) \to \mathrm{Sym}(p) : A \mapsto R(\theta)\,A\,R(\theta), \tag{A.7}$$

which is a linear isomorphism of $\mathrm{Sym}(p)$ with inverse $A \mapsto S(\theta)\,A\,S(\theta)$. Since the map (A.7) is linear and invertible, it is sufficient to prove that the matrices $-R(\theta)\,\dot{S}_m(\theta)\,R(\theta) = \dot{R}_m(\theta)$, for $m = 1,\dots,k$, and $R(\theta)\,D_\theta(\boldsymbol{e}_j)\,R(\theta) = R(\theta)\,\mathrm{diag}(\boldsymbol{e}_j) + \mathrm{diag}(\boldsymbol{e}_j)\,R(\theta)$, for $j = 1,\dots,p$, are all linearly independent. But the latter follows from Assumption 2.1(iv) and the observation that the diagonal of $\dot{R}_m(\theta)$ is zero while the diagonal of $R(\theta)\,\mathrm{diag}(\boldsymbol{e}_j)+\mathrm{diag}(\boldsymbol{e}_j)\,R(\theta)$ is equal to $2\boldsymbol{e}_j$.

Part B. A simple calculation shows $\langle D_\theta(\boldsymbol{e}_i), D_\theta(\boldsymbol{e}_j)\rangle_\theta = (I_p + R(\theta) \circ S(\theta))_{ij}$. As $D_\theta(\boldsymbol{e}_1),\dots,D_\theta(\boldsymbol{e}_p)$ are linearly independent, we can conclude that the Gram matrix $I_p + R(\theta) \circ S(\theta)$ is non-singular.

Part C. Observe that indeed $h_{j,m}^\theta \in \mathrm{L}_2^0[0,1]$ and note that the primitive of $h_{j,m}^\theta$ is given by $H_{j,m}^\theta(u_j) = \int_0^{u_j} h_{j,m}^\theta(\lambda)\,\mathrm{d}\lambda = g_{j,m}(\theta)\,z_j\,\varphi(z_j)$.

By (2.17), we have

$$g_{j,m}(\theta) + \sum_{i=1}^{p} R_{ij}(\theta)\, S_{ij}(\theta)\, g_{i,m}(\theta) = -\sum_{i=1}^{p} \dot{R}_{ij,m}(\theta)\, S_{ij}(\theta). \tag{A.8}$$

Taking the sum over $j = 1, \ldots, p$ yields

$$\sum_{j=1}^{p} g_{j,m}(\theta) = -\frac{1}{2} \operatorname{tr}\big(\dot{R}_m(\theta)\, S(\theta)\big).$$

Inserting these expressions and the one for $\dot{\ell}_j$ in (2.6) into the one for $\mathcal{O}_{\theta,j}$ in (2.8) yields

$$[\mathcal{O}_\theta \boldsymbol{h}^\theta](\boldsymbol{U}) = -\frac{1}{2} \operatorname{tr}\big(\dot{R}_m(\theta)\, S(\theta)\big) - \frac{1}{2} \boldsymbol{z}'\, D_\theta(\boldsymbol{g}_m(\theta))\, \boldsymbol{z} \tag{A.9}$$

A combination of this display with the expression for $\dot{\ell}_\theta$ in (2.2) shows that (2.18) holds, with $\dot{\ell}^*_{\theta,m}$ defined as in (2.20). From here, (2.19) is immediate.

Since $I^*(\theta)$ is the Gram matrix associated to the matrices $D_\theta(\boldsymbol{g}_m(\theta)) - \dot{S}_m(\theta)$, for $m = 1, \ldots, k$, nonsingularity of $I^*(\theta)$ follows if we can show that these matrices are linearly independent. Let $\alpha \in \mathbb{R}^k$ be such that $\sum_{m=1}^{k} \alpha_m \{D_\theta(\boldsymbol{g}_m(\theta)) - \dot{S}_m(\theta)\} = 0$. Then

$$\sum_{m=1}^{k} \alpha_m(-\dot{S}_m(\theta)) + \sum_{j=1}^{p} \left(\sum_{m=1}^{k} \alpha_m \boldsymbol{g}_{j,m}(\theta) \right) D_\theta\,(\boldsymbol{e}_j) = 0,$$

which yields $\alpha = 0$ because of Part A.

Part D. In view of Part C, all that remains to be shown is that the function $\dot{\ell}^*_{\theta,m}$ in (2.18) is orthogonal to $\mathcal{T}_{\mathrm{P}_\theta}$. We establish orthogonality by an application of Lemma 2.7 to the matrix $A = D_\theta(\boldsymbol{g}_m(\theta)) - \dot{S}_m(\theta)$. The jth diagonal element of the matrix $R(\theta)A$ is equal to

$$\sum_{i=1}^{p} [\{g_{i,m}(\theta) + g_{j,m}(\theta)\} S_{ij}(\theta) - \dot{S}_{ij,m}(\theta)] R_{ij}(\theta).$$

This expression is zero by linear equation (A.8) for the weights $\boldsymbol{g}_m(\theta)$ together with the identities $R(\theta)\, S(\theta) = I_p$ and $\dot{S}(\theta)\, R(\theta) = -S(\theta)\, \dot{R}(\theta)$. □

Details for Remark 2.10. In analogy to the bivariate case (Klaassen and Wellner, 1997), we derive the Sturm–Liouville equations to which the functions in (2.16) are the solution. Fix $\lambda \in (0, 1)$ and $j = 1, \ldots, p$, and consider the orthogonality relation (2.14) at the vector $\boldsymbol{h}$ defined by $h_j(u) = 1\{u \leq \lambda\} - \lambda$

and $h_i(u) = 0$ for $i \neq j$ and $u \in (0,1)$. We have $[\mathcal{O}_\theta \boldsymbol{h}](\boldsymbol{U}) = 1\{U_j \leq \lambda\} - \lambda + \dot{\ell}_j(\boldsymbol{U};\theta)(U_j \wedge \lambda - U_j\lambda)$. Writing $H^\theta_{j,m}(u_j) = \int_0^{u_j} h^\theta_{j,m}(t)\,\mathrm{d}t$, equation (2.14) yields

$$\text{(A.10)}\quad H^\theta_{j,m}(\lambda) + \int_0^1 I_{jj}(u_j;\theta) H^\theta_{j,m}(u_j)(u_j \wedge \lambda - u_j\lambda)\mathrm{d}u_j - \sum_{i:\, i\neq j} \int \ddot{\ell}_{ij}(\boldsymbol{u};\theta) H^\theta_{i,m}(u_i)(u_j \wedge \lambda - u_j\lambda) c_\theta(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \int_0^1 I_{\theta j,m}(u_j;\theta)(u_j \wedge \lambda - u_j\lambda)\,\mathrm{d}u_j,$$

with

$$I_{\theta j,m}(u_j;\theta) = \mathbb{E}_\theta\big[\dot{\ell}_{\theta,m}\dot{\ell}_j(\boldsymbol{U};\theta)\,\big|\,U_j = u_j\big] = -\frac{z_j}{\varphi(z_j)}\sum_{i=1}^p \dot{R}_{ij,m}(\theta) S_{ij}(\theta).$$

Differentiating (A.10) twice with respect to λ yields

$$\frac{\partial}{\partial\lambda} h^\theta_{j,m}(\lambda) - I_{jj}(\lambda;\theta)\, H^\theta_{j,m}(\lambda) = - I_{\theta j,m}(\lambda;\theta) - \sum_{i:\, i\neq j} \mathbb{E}_\theta\big[\ddot{\ell}_{ij}(\boldsymbol{U};\theta) H^\theta_{i,m}(U_i)\,\big|\,U_j = \lambda\big],$$

for $\lambda \in (0,1)$ and $j = 1, \ldots, p$. For $p = 2$ (and $k = 1$), this system of p coupled Sturm–Liouville differential equations in the p functions $h^\theta_{1,m}, \ldots, h^\theta_{p,m}$ reduces to equations (4.57)–(4.58) in Klaassen and Wellner (1997). □

APPENDIX B: PROOF OF THEOREM 3.2

We prove efficiency of $\hat{\theta}_n^{\mathrm{OSE}}$ at $\mathrm{P}_\theta \in \mathcal{P}$ by demonstrating asymptotic linearity in the efficient influence function, i.e.

$$\text{(B.1)}\qquad D_n = \sqrt{n}(\hat{\theta}_n^{\mathrm{OSE}} - \theta) - \frac{1}{\sqrt{n}}\sum_{i=1}^n I^{*-1}(\theta)\,\dot{\ell}^*_\theta(\boldsymbol{U}_i;\theta) = o(1;\mathbb{P}_\theta).$$

Since $\hat{\theta}_n^{\mathrm{OSE}}$ is rank-based, the previous display straightforwardly yields

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{OSE}} - \theta) - \frac{1}{\sqrt{n}}\sum_{i=1}^n I^{*-1}(\theta)\,\dot{\ell}^*_\theta\big(F_1(X_{i1}), \ldots, F_p(X_{ip});\theta\big) = o(1;\mathbb{P}_{\theta,F_1,\ldots,F_p}),$$

which shows that $\hat{\theta}_n^{\mathrm{OSE}}$ is efficient at all $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$.

Fix $\epsilon > 0$. Let B be such that $\mathbb{P}_\theta\big(|\sqrt{n}(\tilde{\theta}_n - \theta)| \geq B\big) < \epsilon$ for all $n \in \mathbb{N}$. As $\tilde{\theta}_n$ takes values in $n^{-1/2}\mathbb{Z}^k$ we have

$$\mathbb{P}_\theta(|D_n| > \epsilon) \leq \epsilon + \sum_{\substack{\theta_n \in n^{-1/2}\mathbb{Z}^k \\ |\sqrt{n}(\theta_n - \theta)| \leq B}} \mathbb{P}_\theta\big(|D_n| > \epsilon,\, \tilde{\theta}_n = \theta_n\big)$$

Since the number of terms in the sum can be bounded by a constant not depending on n, we can conclude that (B.1) holds if, for all sequences $\theta_n = \theta + O(n^{-1/2})$,

$$I^{*-1}(\theta_n)\,\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta^*\big(\hat{F}_{n,1}(U_{i1}),\ldots,\hat{F}_{n,p}(U_{ip});\theta_n\big) - \frac{1}{\sqrt{n}}\sum_{i=1}^n I^{*-1}(\theta)\,\dot{\ell}_\theta^*(\boldsymbol{U}_i;\theta) = -\sqrt{n}(\theta_n - \theta) + o(1;\mathbb{P}_\theta).$$

Because of the continuity of $\theta \mapsto I^*(\theta)$ [see (2.19)], it follows that it is sufficient to show the following two properties:

(P1) for any sequence $\theta_n = \theta + h_n/\sqrt{n}$, with h_n bounded, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta^*(\boldsymbol{U}_i;\theta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta^*(\boldsymbol{U}_i;\theta) - I^*(\theta)h_n + o(1;\mathbb{P}_\theta);$$

(P2) for any sequence $\theta_n = \theta + h_n/\sqrt{n}$, with h_n bounded, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta^*\big(\hat{F}_{n,1}(U_{i1}),\ldots,\hat{F}_{n,p}(U_{ip});\theta_n\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_\theta^*(\boldsymbol{U}_i;\theta_n) + o(1;\mathbb{P}_\theta).$$

Proof of (P1). We prove the result for the mth component using the following lemma, which is an application of Proposition A.10 in Van der Vaart (1988).

LEMMA B.1. *Let $\theta_n = \theta + h_n/\sqrt{n}$ with h_n bounded. Suppose that for sequences $\kappa^{(n)} \in \mathrm{L}_2^0(\mathrm{P}_{\theta_n})$ and $\zeta^{(n)} \in \mathrm{L}_2^0(\mathrm{P}_\theta)$, the following three conditions hold:*

$$\text{(B.2)} \qquad \int \Big(\kappa^{(n)}(\boldsymbol{u})\sqrt{c_{\theta_n}(\boldsymbol{u})} - \zeta^{(n)}(\boldsymbol{u})\sqrt{c_\theta(\boldsymbol{u})}\Big)^2 \mathrm{d}\boldsymbol{u} \to 0,$$

$$\text{(B.3)} \qquad \sup_{n\in\mathbb{N}} \mathrm{E}_\theta\big[\big(\zeta^{(n)}(\boldsymbol{U})\big)^2\big] < \infty,$$

$$\text{(B.4)} \qquad \lim_{n\to\infty} \mathrm{E}_{\theta_n}\big[\big(\kappa^{(n)}(\boldsymbol{U})\big)^2\, 1\{|\kappa^{(n)}(\boldsymbol{U})| > \epsilon\sqrt{n}\}\big] = 0 \textit{ for every } \epsilon > 0.$$

Then we have, with $\boldsymbol{U}_i$, $i = 1, \ldots, n$, i.i.d. with law P_θ,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \kappa^{(n)}(\boldsymbol{U}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta^{(n)}(\boldsymbol{U}_i) - h_n' \, \mathrm{E}_\theta\big[\dot{\ell}_\theta(\boldsymbol{U};\theta)\, \zeta^{(n)}(\boldsymbol{U})\big] + o(1;\mathbb{P}_\theta).$$

PROOF. This follows by an application of Proposition A.10 in Van der Vaart (1988). Let us comment on the precise application. We set $Q_{nj} = \mathrm{P}_{\theta_n}$, $P_{nj} = \mathrm{P}_\theta$, and $g_{nj} = h_n' \dot{\ell}_\theta(\boldsymbol{U}_j;\theta)$. Next, note that the conditions to Proposition A.8 in Van der Vaart (1988) are satisfied, because $\mathcal{P}_{\mathrm{km}}$ is a regular parametric model (see Lemma 2.3). □

We apply the Lemma B.1 with $\kappa^{(n)}(\boldsymbol{U}) = \dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta_n) \in \mathrm{L}_2^0(\mathrm{P}_{\theta_n})$ and $\zeta^{(n)}(\boldsymbol{U}) = \dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta) \in \mathrm{L}_2^0(\mathrm{P}_\theta)$. As $\zeta^{(n)}$ is constant in n, condition (B.3) is immediate.

First we demonstrate (B.2). We have, for $\boldsymbol{u} \in (0,1)^p$,

$$\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta_n)\, \sqrt{c_{\theta_n}}(\boldsymbol{u}) \to \dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta)\, \sqrt{c_\theta}(\boldsymbol{u}),$$

and, by continuity of $\theta \mapsto I^*(\theta)$,

$$\begin{aligned}\lim_{n\to\infty} \int \Big(\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta_n)\, \sqrt{c_{\theta_n}}(\boldsymbol{u})\Big)^2 \,\mathrm{d}\boldsymbol{u} &= \lim_{n\to\infty} I^*_{mm,\theta_n} \\ &= I^*_{mm,\theta} = \int \Big(\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta)\, \sqrt{c_\theta}(\boldsymbol{u})\Big)^2 \,\mathrm{d}\boldsymbol{u}.\end{aligned}$$

An application of the L_r-convergence theorem (Loève, 1977, p. 165) yields

$$\lim_{n\to\infty} \int \Big(\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta_n)\, \sqrt{c_{\theta_n}}(\boldsymbol{u}) - \dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta)\, \sqrt{c_\theta}(\boldsymbol{u})\Big)^2 \,\mathrm{d}\boldsymbol{u} = 0, \tag{B.5}$$

which shows that (B.2) holds.

Next, we show that (B.4) holds. Let $\epsilon > 0$. Define $F_n = \{\boldsymbol{u} \in (0,1)^p : |\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta_n)| > \epsilon\sqrt{n}\}$. Notice first that

$$\begin{aligned}\mathrm{P}_\theta(F_n) \le \mathrm{P}_\theta&\Big(|\dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta_n) - \dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta)| > \epsilon\Big) \\ &+ \mathrm{P}_\theta\Big(|\dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta)| > \epsilon(1+\sqrt{n})\Big) \to 0,\end{aligned}$$

which implies

$$\lim_{n\to\infty} \mathrm{E}_\theta\big[1_{F_n}(\boldsymbol{U})\, \big(\dot{\ell}^*_{\theta,m}(\boldsymbol{U};\theta)\big)^2\big] = 0.$$

A combination with (B.5) yields

$$\begin{aligned}
\mathrm{E}_{\theta_n}\big[\big(\dot\ell^*_{\theta,m}(\boldsymbol{U};\theta_n)\big)^2\,1_{F_n}(\boldsymbol{U})\big] &= \int_{F_n}\Big(\sqrt{c_{\theta_n}}(\boldsymbol{u})\,\dot\ell^*_{\theta,m}(\boldsymbol{u};\theta_n)\Big)^2\,\mathrm{d}\boldsymbol{u}\\
&\le 2\,\mathrm{E}_\theta\big[1_{F_n}(\boldsymbol{U})\,\big(\dot\ell^*_{\theta,m}(\boldsymbol{U};\theta)\big)^2\big]\\
&\quad + 2\int\Big(\sqrt{c_{\theta_n}}(\boldsymbol{u})\,\dot\ell^*_{\theta,m}(\boldsymbol{u};\theta_n) - \sqrt{c_\theta}(\boldsymbol{u})\,\dot\ell^*_{\theta,m}(\boldsymbol{u};\theta)\Big)^2\,\mathrm{d}\boldsymbol{u}\to 0,
\end{aligned}$$

which demonstrates (B.4). So all conditions to Lemma B.1 hold.

Property (P1) now follows from the observation

$$\mathrm{E}_\theta\big[\dot\ell^*_{\theta,m}(\boldsymbol{U};\theta)\,h_n'\,\dot\ell_\theta(\boldsymbol{U};\theta)\big] = I^*_{m\cdot,\theta}h_n,$$

where $I^*_{m\cdot,\theta}$ denotes the mth row of $I^*(\theta)$. Indeed, the components of the difference vector $\dot\ell_\theta - \dot\ell^*_\theta$ are orthogonal with respect to those of $\dot\ell^*_\theta$, so we can replace $\dot\ell_\theta$ by $\dot\ell^*_\theta$ in the above display.

Proof of (P2). Recall $Z_{ij} = \Phi^{-1}(U_{ij})$ and write $\tilde Z_{ij} = \Phi^{-1}(R^{(n)}_{ij}/(n+1))$. We have

$$\begin{aligned}
d_n &= \frac{1}{\sqrt n}\sum_{i=1}^n \dot\ell^*_{\theta,m}\big(\hat F_{n,1}(U_{i1}),\ldots,\hat F_{n,p}(U_{ip});\theta_n\big) - \frac{1}{\sqrt n}\sum_{i=1}^n \dot\ell^*_{\theta,m}(\boldsymbol{U}_i;\theta_n)\\
&= \frac{1}{2\sqrt n}\sum_{i=1}^n\big(\tilde{\boldsymbol{Z}}_i' A_{m,n}\tilde{\boldsymbol{Z}}_i - \boldsymbol{Z}_i' A_{m,n}\boldsymbol{Z}_i\big),
\end{aligned}$$

where the $p\times p$ symmetric matrix $A_{m,n}$ is given by (see Proposition 2.8)

$$A_{m,n} = D_{\theta_n}(\boldsymbol{g}_m(\theta_n)) - \dot S_m(\theta_n).$$

Note that the elements of $A_{m,n}$ are bounded in n.

To show that $d_n = o(1;\mathbb{P}_\theta)$, we follow the line of argumentation in Hoff, Niu and Wellner (2012, Theorem 3.1). First we recall the following lemma, due to de Wet and Venter (1972, Theorem 1).

LEMMA B.2. *Let Y_i, $i\in\mathbb{N}$, be i.i.d. $N(0,1)$, let $R_i^{(n)}$ be the rank of Y_i among $Y_1,\ldots,Y_n$, and write $\tilde Y_i^{(n)} = \Phi^{-1}(\frac{R_i^{(n)}}{n+1})$. Then*

$$\frac{1}{\sqrt n}\sum_{i=1}^n\big(Y_i - \tilde Y_i^{(n)}\big)^2 = o_P(1),\qquad \text{as } n\to\infty.$$

Decompose $2d_n = Q_n + 2L_n$ with

$$Q_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{\boldsymbol{Z}}_i - \boldsymbol{Z}_i)' A_{m,n} (\tilde{\boldsymbol{Z}}_i - \boldsymbol{Z}_i)$$

and

$$\begin{aligned} L_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{\boldsymbol{Z}}_i - \boldsymbol{Z}_i)' A_{m,n} \boldsymbol{Z}_i \\ &= \sum_{j=1}^{p} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{Z}_{ij} - Z_{ij}) a'_{m,n;\cdot j} \boldsymbol{Z}_i \right] = \sum_{j=1}^{p} L_{nj}, \end{aligned}$$

where $a_{m,n;\cdot j}$ is the jth column of $A_{m,n}$. Since the elements of $A_{m,n}$ are bounded, an application of Lemma B.2 yields $Q_n = o(1; \mathbb{P}_\theta)$. To conclude the proof, it suffices to show that $L_{n,j} = o(1; \mathbb{P}_\theta)$ for every $j = 1, \ldots, p$.

The vector $a_{m,n;\cdot j}$ is orthogonal to the jth row of $R(\theta_n)$, denoted by $r_{n,\cdot j}$. Indeed, since $R\dot{S}_m = -\dot{R}_m S$, we have, by (A.8),

$$\begin{aligned} &\operatorname{diag}(R(\theta_n) A_{m,n}) \\ &\quad = \boldsymbol{g}_m(\theta_n) + \operatorname{diag}(R(\theta_n) \operatorname{diag}(\boldsymbol{g}_m(\theta_n)) S(\theta_n)) - \operatorname{diag}(R(\theta_n) \dot{S}_m(\theta_n)) = 0. \end{aligned}$$

Hence, under $\mathbb{P}_{\theta_n}$,

$$a'_{m,n;\cdot j} \boldsymbol{Z}_i \mid Z_{ij} \sim N\left(0, a'_{m,n;\cdot j} \left(R(\theta_n) - r_{n,\cdot j} r'_{n,\cdot j}\right) a_{m,n;\cdot j}\right).$$

This yields $\mathbb{E}_{\theta_n}[L_{nj} \mid Z_{1j}, \ldots, Z_{nj}] = 0$. The conditional variance is

$$\begin{aligned} &\operatorname{var}_{\theta_n}[L_{nj} \mid Z_{1j}, \ldots, Z_{nj}] \\ &\quad = \frac{1}{n} a'_{m,n;\cdot j} \left(R(\theta_n) - r_{n,\cdot j} r'_{n,\cdot j}\right) a_{m,n;\cdot j} \sum_{i=1}^{n} \left(\tilde{Z}_{ij} - Z_{ij}\right)^2 = o(1; \mathbb{P}_{\theta_n}), \end{aligned}$$

by another application of Lemma B.2 and by the fact that $|a_{m,n;\cdot j}|$, $R(\theta_n)$ and $r_{n,\cdot j}$ are bounded. The conditional version of the Cauchy-Schwarz inequality thus yields, for $\varepsilon > 0$,

$$\mathbb{P}_{\theta_n}(|L_{nj}| > \varepsilon \mid Z_{1j}, \ldots, Z_{nj}) = o_P(1; \mathbb{P}_{\theta_n}),$$

By the bounded convergence theorem (the conditional probabilities are, of course, bounded by 1), we have $L_{nj} = o_P(1; \mathbb{P}_{\theta_n})$. By a contiguity argument [recall that $\mathcal{P}_{\mathrm{km}}$ is a regular parametric model by Lemma 2.2 and use Proposition 2.2.3 in Bickel et al. (1993)] we obtain $L_{nj} = o(1; \mathbb{P}_\theta)$, which completes the verification of (P2). □

APPENDIX C: PROOFS FOR SECTION 4

PROOF OF LEMMA 4.2. Using Lemma 2.4, the function $q_A(\boldsymbol{U})$ belongs to $\mathcal{T}_{\mathrm{P}_\theta}$ if and only if it is of the form $q_A(\boldsymbol{U}) = [\mathcal{O}_\theta \boldsymbol{h}](\boldsymbol{U}) = \sum_{j=1}^p [\mathcal{O}_{\theta,j} h_j](U_j)$ for some $\boldsymbol{h} = (h_1, \ldots, h_p) \in (\mathrm{L}_2^0[0,1])^p$. The component functions h_j are determined by

$$h_j(U_j) = \mathrm{E}_\theta\big[q_A(\boldsymbol{U}) \,\big|\, U_j\big] = \frac{1}{2}(RAR)_{jj}\left(Z_j^2 - 1\right).$$

Defining $\boldsymbol{b} \in \mathbb{R}^p$ by $b_j = \frac{1}{2}(RAR)_{jj}$ and using the definition of $\mathcal{O}_{\theta,j}$ in (2.8), we find, by a similar calculation as in (A.5),

$$q_A(\boldsymbol{U}) = \boldsymbol{Z}'\, S(\theta)\ \mathrm{diag}(\boldsymbol{b})\, \boldsymbol{Z} - \frac{1}{2}\,\mathrm{tr}(RAR),$$

from which we obtain $A = D_\theta(\boldsymbol{b})$ by symmetrization. □

PROOF OF COROLLARY 4.3. *(b) and (c) are equivalent.* The matrix $A \in \mathrm{Sym}(p)$ is orthogonal to $\dot{T}_\theta$ if and only if, for all $\boldsymbol{b} \in \mathbb{R}^p$, we have

$$\begin{aligned} 0 &= \langle S(\theta)\ \mathrm{diag}(\boldsymbol{b}) + \mathrm{diag}(\boldsymbol{b})\, S(\theta), A\rangle_\theta \\ &= \mathrm{tr}\big(R(\theta)\, S(\theta)\ \mathrm{diag}(\boldsymbol{b})\, R(\theta)\, A + R(\theta)\ \mathrm{diag}(\boldsymbol{b})\, S(\theta)\, R(\theta)\, A\big) \\ &= \mathrm{tr}\big(\mathrm{diag}(\boldsymbol{b})\, \{R(\theta)\, A + A\, R(\theta)\}\big). \end{aligned}$$

As $\boldsymbol{b} \in \mathbb{R}^p$ is arbitrary, this occurs if and only if $\mathrm{diag}(R(\theta)A) = 0$. □

PROOF OF COROLLARY 4.4. By Corollary 4.3, the vector $\boldsymbol{b}$ is to be found by solving $\mathrm{diag}(R(\theta)(A - D_\theta(\boldsymbol{b}))) = \mathbf{0}$. But this is equation (4.4). □

PROOF OF PROPOSITION 4.6. Thanks to Corollary 4.3, equations (4.6), (4.8) and (4.10) are equivalent. Further, since $\dot{\ell}_{\theta,m'} = q_{-\dot{S}_{m'}(\theta)}$, we have

$$\begin{aligned} \mathrm{cov}_\theta\big(q_{A_m(\theta)}(\boldsymbol{U}),\, \dot{\ell}_{\theta,m'}(\boldsymbol{U})\big) &= \frac{1}{2}\,\mathrm{tr}\big(R(\theta)\, A_m(\theta)\, R(\theta)\, (-\dot{S}_{m'}(\theta))\big) \\ &= \frac{1}{2}\,\mathrm{tr}\big(A_m(\theta)\, \dot{R}_{m'}(\theta)\big), \end{aligned}$$

and therefore equations (4.7), (4.9) and (4.11) are equivalent too. □

PROOF OF PROPOSITION 4.7. First assume that $\hat{\theta}_n$ is efficient. By assumption, the efficient influence matrices $A_1^*(\theta), \ldots, A_k^*(\theta)$ belong to the linear span of $B_1(\theta), \ldots, B_k(\theta)$. Since the former matrices are linearly independent, their Gram matrix $I^*(\theta)$ being non-singular by Proposition 2.8, the matrices $B_1(\theta), \dots, B_k(\theta)$ can be written as linear combinations of the matrices $A_1^*(\theta), \dots, A_k^*(\theta)$. By (4.12), it follows that every matrix $B_m(\theta)$ is of the form

$$B_m(\theta) = D_\theta(\boldsymbol{b}_m(\theta)) - \sum_{m'=1}^{k} \lambda_{mm'}(\theta)\dot{S}_{m'}(\theta)$$

for some vector $\boldsymbol{b}_m(\theta) \in \mathbb{R}^p$ and some coefficients $\lambda_{mm'}(\theta) \in \mathbb{R}$, $m' = 1, \dots, k$. Pre- and postmultiplication with $R(\theta)$ yields

$$R(\theta)\, B_m\, R(\theta) = R(\theta)\, \operatorname{diag}(\boldsymbol{b}_m(\theta)) + \operatorname{diag}(\boldsymbol{b}_m(\theta))\, R(\theta) + \sum_{m'=1}^{k} \lambda_{mm'}(\theta)\dot{R}_{m'}(\theta).$$

As all diagonal elements of $R(\theta)$ are unity while all diagonal elements of $\dot{R}_{m'}(\theta)$ are zero, an inspection of the diagonals of the matrices at the left- and right-hand sides of the previous display yields $\boldsymbol{b}_m(\theta) = \frac{1}{2}\operatorname{diag}(R(\theta)B_mR(\theta))$. We find (4.13).

Conversely, assume that $\hat{\theta}_n$ satisfies (4.5) for regular influence matrices $A_1(\theta), \dots, A_k(\theta)$ belonging to the linear span of matrices $B_1(\theta), \dots, B_k(\theta)$ satisfying (4.13). By pre- and postmultiplication with $S(\theta)$, it follows that, for every $m = 1, \dots, k$, the matrix

$$B_m(\theta) - \frac{1}{2}\big\{S(\theta)\operatorname{diag}\big(R(\theta)B_m(\theta)R(\theta)\big) + \operatorname{diag}\big(R(\theta)B_m(\theta)R(\theta)\big)S(\theta)\big\}$$

belongs to the linear span of $\dot{S}_1(\theta), \dots, \dot{S}_k(\theta)$. Therefore, the matrices $A_m(\theta)$ belong to the subspace of $\operatorname{Sym}(p)$ defined by (4.14). Moreover, the regularity conditions (4.8)–(4.9) fix the inner products of each $A_m(\theta)$ with each matrix in the space (4.14). The efficient influence matrices $A_m^*(\theta)$ belong to the space (4.14) and are regular too. It follows that $A_m(\theta) = A_m^*(\theta)$ for all $m = 1, \dots, k$, as required. □

PROOF OF THEOREM 4.1. Since $\hat{\theta}_n^{\mathrm{PLE}}$ is rank-based, we can, without loss of generality, assume that the marginal distributions $F_1, \dots, F_p$ are uniform on $(0, 1)$. By Hoff, Niu and Wellner (2012, Section 5), the influence function of the pseudo-likelihood estimator is quadratic and is generated by

$$A_m^{\mathrm{PLE}}(\theta) = \sum_{m'=1}^{k} \big(I^{-1}(\theta)\big)_{m,m'} B_{m'}^{\mathrm{PLE}}(\theta), \tag{C.1}$$

with

$$B_m^{\mathrm{PLE}}(\theta) = -\dot{S}_m(\theta) + \operatorname{diag}\big(R(\theta)\dot{S}_m(\theta)\big), \qquad m = 1, \dots, k. \tag{C.2}$$

Regularity of the pseudo-likelihood estimator in the sense of (4.8)–(4.9) is easily established. First, $A_m^{\mathrm{PLE}}(\theta)$ is orthogonal to $\dot{T}_\theta$ since the same is true for each $B_m(\theta)$. Indeed, note that $\mathrm{diag}(R(\theta)\, B_m^{\mathrm{PLE}}(\theta)) = 0$ and apply Corollary 4.3. Second, as each $-\dot{S}_m(\theta)$ is orthogonal to any diagonal matrix, we have, by (2.4), $\langle B_m^{\mathrm{PLE}}(\theta), -\dot{S}_{m'}(\theta)\rangle_\theta = (I_\theta)_{m,m'}$, which implies (4.9).

The efficiency criterion now follows from Proposition 4.7 upon noting that

$$\begin{aligned} R(\theta)\, B_m^{\mathrm{PLE}}(\theta)\, R(\theta) &= -R(\theta)\, \dot{S}_m(\theta) R(\theta) + R(\theta)\, \mathrm{diag}\big(R(\theta)\dot{S}_m(\theta)\big)\, R(\theta) \\ &= \dot{R}_m(\theta) + L_m(\theta) \end{aligned}$$

and that $\dot{R}_m(\theta)$ has zero diagonal. ☐

Proof of Proposition 4.8. The model is adaptive at $\theta \in \Theta$ if and only if the parametric scores $\dot{\ell}_{\theta,m}(\boldsymbol{U}) = q_{-\dot{S}_m(\theta)}(\boldsymbol{U})$ are orthogonal to $\mathcal{T}_{\mathrm{P}_\theta}$, the nonparametric part of the tangent space in $\mathrm{L}_2^0(\mathrm{P}_\theta)$. An application of Corollary 4.3 then gives the result. ☐

APPENDIX D: PROOFS FOR SECTION 5

The efficiency criterion in Theorem 4.1 for the pseudo-likelihood estimator even holds if the parameter θ is itself not identifiable and a reparametrization $\nu \mapsto \theta(\nu)$ in terms of a lower-dimensional parameter ν is needed to enforce identifiability. This situation occurs, for instance, in the factor model of Example 5.5. Also in such cases, the efficiency criterion in Theorem 4.1 formulated in terms of the original θ then still yields efficiency of the pseudo-likelihood estimator for the new parameter ν. We formalize this in the present appendix.

Assumption D.1. *Let $1 \le d \le k$ and let both sets $N \subset \mathbb{R}^d$ and $\Theta \subset \mathbb{R}^k$ be open. The maps $N \to \Theta : \nu \mapsto \theta(\nu)$ and $\Theta \to \mathbb{R}^{p\times p} : \theta \mapsto R(\theta)$ are such that:*

- *(i) for every $\theta \in \Theta$, the matrix $R(\theta)$ is an invertible correlation matrix with inverse $S(\theta) = R^{-1}(\theta)$;*
- *(ii) the map $\nu \mapsto R(\theta(\nu))$ is one-to-one;*
- *(iii) the maps $\nu \mapsto \theta(\nu)$ and $\theta \mapsto R(\theta)$ are continuously differentiable;*
- *(iv) for every $\nu \in N$, the $p \times p$ matrices $\partial R(\theta(\nu))/\partial \nu_i$, for $i = 1, \ldots, d$, are linearly independent;*
- *(v) for every $\nu \in N$, we have*

$$\mathrm{span}\{\partial R(\theta(\nu))/\partial \nu_i \mid i = 1, \ldots, d\} = \mathrm{span}\{\partial R(\theta)/\partial \theta_m \mid m = 1, \ldots, k\}.$$

Note that the assumptions imply that the parametrization $\nu \mapsto R(\theta(\nu))$ satisfies Assumption 2.1.

COROLLARY D.2. *Suppose Assumption D.1 holds and consider the semi-parametric Gaussian copula model induced by $\nu \mapsto R(\theta(\nu))$. The pseudo-likelihood estimator $\hat{\nu}_n^{\mathrm{PLE}}$ is efficient at $\nu \in N$ in the sense of (3.1) if, for every $m = 1, \ldots, k$, the matrix*

$$L_m(\theta) - \frac{1}{2}\big(\operatorname{diag}(L_m(\theta))R(\theta) + R(\theta)\operatorname{diag}(L_m(\theta))\big)$$

with

$$L_m(\theta) = R(\theta)\ \operatorname{diag}\big((\partial R(\theta)/\partial \theta_m)\, S(\theta)\big)\ R(\theta)$$

belongs to the linear span of $\{\partial R(\theta)/\partial\theta_m \mid m = 1, \ldots, k\}$.

PROOF. Write $\dot{\theta}_{m,i}(\nu) = \partial\theta_m(\nu)/\partial\nu_i$. By the chain rule,

$$\partial R(\theta(\nu))/\partial\nu_i = \sum_{m=1}^{k} \dot{\theta}_{m,i}(\nu)\, (\partial R(\theta)/\partial\theta_m)|_{\theta=\theta(\nu)}. \tag{D.1}$$

According to Theorem 4.1, we have to check that, for every $i = 1, \ldots, d$, the matrix

$$L_{\nu,i}(\nu) - \frac{1}{2}\big(\operatorname{diag}(L_{\nu,i}(\nu))R(\theta(\nu)) + R(\theta(\nu))\operatorname{diag}(L_{\nu,i}(\nu))\big)$$

with

$$L_{\nu,i}(\nu) = R(\theta(\nu))\ \operatorname{diag}\big((\partial R(\theta(\nu))/\partial\nu_i)\, S(\theta(\nu))\big)\ R(\theta(\nu))$$

belongs to $\operatorname{span}\{\partial R(\theta(\nu))/\partial\nu_i \mid i = 1, \ldots, d\}$. But this follows from the criterion in the corollary and Assumption D.1(v) upon noting that, by (D.1), $L_{\nu,i}(\nu) = \sum_{m=1}^{k} \dot{\theta}_{m,i}(\nu)\, L_m(\theta(\nu))$. □

The crux of Assumption D.1 is item (v), which says that, locally at least, the ν-parametrization reconstructs the full θ-parametrization. It also requires that the image $\{\theta(\nu) \mid \nu \in N\} \subset \Theta$ is such that for each point θ in the image, the dimension of the linear span of the matrices $\partial R(\theta)/\partial\theta_m$, for $m = 1, \ldots, k$, is the same, namely d.

DETAILS FOR EXAMPLE 5.2. For the Toeplitz model in dimension $p = 3$, the inverse of the efficient information matrix is given by

$$\begin{cases} [I^{*-1}(\theta)]_{11} &=& \frac{1}{4}\big(\theta_1^2\,\theta_2^2 - 4\,\theta_1^2\,\theta_2 + 2\,\theta_2 + 4\,\theta_1^4 - 5\,\theta_1^2 + 2\big), \\ [I^{*-1}(\theta)]_{22} &=& (1-\theta_2^2)^2, \\ [I^{*-1}(\theta)]_{12} &=& \frac{1}{2}\,\theta_1\ (\theta_2 - 1)\ \big(\theta_2^2 - \theta_2 + 2\,\theta_1^2 - 2\big)\,. \end{cases} \tag{D.2}$$

□

Details for Example 5.3. Using the criterion in Theorem 4.1, we verify efficiency of the pseudo-likelihood estimator for the exchangeable model in Example 5.3. We have $k = 1$ and $\dot{R}_1(\theta)$ has unit elements everywhere, except on the diagonal, which is zero. Since $R(\theta)\,\boldsymbol{\iota}_p = [1 + (p-1)\theta]\,\boldsymbol{\iota}_p$, we find $S(\theta)\,\boldsymbol{\iota}_p = [1 + (p-1)\theta]^{-1}\,\boldsymbol{\iota}_p$, meaning that all row and column sums of $S(\theta)$ are identical. As all diagonal elements of $S(\theta)$ are the same as well, it follows that all diagonal elements of $\dot{R}_1(\theta)\,S(\theta)$ are the same, and thus that $\operatorname{diag}(\dot{R}_1(\theta)\,S(\theta))$ is a multiple of I_p. By Theorem 4.1, it suffices therefore to show that the matrix

$$R(\theta)^2 - \frac{1}{2}\big(\operatorname{diag}(R(\theta)^2)\,R(\theta) + R(\theta)\,\operatorname{diag}(R(\theta)^2)\big)$$

is a multiple of $\dot{R}_1(\theta)$, i.e., that it has zero diagonal and that all its off-diagonal elements are the same. But these two properties are easily verified; note, e.g., that all diagonal elements of $R(\theta)^2$ are equal. □

Details for Example 5.5. For the factor model, the pseudo-likelihood estimator of the parameter ν in any reparametrization $\nu \mapsto \theta(\nu)$ satisfying Assumption D.1 is semiparametrically efficient. We prove this by verifying the criterion in Corollary D.2.

For $m = (m_1, m_2) \in K = \{1, \ldots, p\} \times \{1, \ldots, q\}$, let $\dot{R}_m(\theta)$ be the partial derivative of $R(\theta)$ in (5.4) with respect to θ_{m_1,m_2} and put

$$L_m(\theta) = R(\theta)\,\operatorname{diag}\big(\dot{R}_m(\theta)\,S(\theta)\big)\,R(\theta).$$

We need to show that for each $m \in K$, the matrix

$$L_m(\theta) - \frac{1}{2}\big\{\operatorname{diag}(L_m(\theta))\,R(\theta) + R(\theta)\,\operatorname{diag}(L_m(\theta))\big\} \tag{D.3}$$

belongs to the space $\operatorname{Sp}(\theta) = \operatorname{span}\{\dot{R}_m(\theta) : m \in K\}$.

The space $\operatorname{Sp}(\theta)$ admits a simple description. By linearity of the diagonal-removal operator, we have $\dot{R}_m(\theta) = \operatorname{rd}(E_m\theta' + \theta E_m')$, with $E_m \in \mathbb{R}^{p\times q}$ having zero elements everywhere, except for a unit element at position (m_1, m_2). As a consequence,

$$\operatorname{Sp}(\theta) = \{\operatorname{rd}(B\theta' + \theta B') \mid B \in \mathbb{R}^{p\times q}\}.$$

To show that the matrix in (D.3) belongs to $\operatorname{Sp}(\theta)$, note that, by (5.4), it can be written as

$$\operatorname{rd}(L_m(\theta)) - \frac{1}{2}\operatorname{rd}\big(\operatorname{diag}(L_m(\theta))\,\theta\theta' + \theta\theta'\,\operatorname{diag}(L_m(\theta))\big). \tag{D.4}$$

Here we used the property $\Delta\,\mathrm{rd}(A) = \mathrm{rd}(\Delta A)$ whenever Δ is a diagonal matrix. The second term in the previous display clearly belongs to $\mathrm{Sp}(\theta)$.

It remains to be shown that $\mathrm{rd}(L_m(\theta))$ belongs to $\mathrm{Sp}(\theta)$ as well. Abbreviating $\Lambda = \mathrm{diag}(\dot{R}_m(\theta)\,S(\theta)) \in \mathbb{R}^{p\times p}$, we have

$$L_m(\theta) = \big(I_p + \theta\theta' - \mathrm{diag}(\theta\theta')\big)\,\Lambda\,\big(I_p + \theta\theta' - \mathrm{diag}(\theta\theta')\big).$$

The right-hand side can be expanded as a sum of nine terms. Four of these terms are diagonal matrices, and they are killed by the diagonal-removal operator. It follows that $\mathrm{rd}(L_m(\theta))$ only involves the five remaining terms, yielding

$$\begin{aligned}\mathrm{rd}(L_m(\theta)) = \mathrm{rd}\big(\Lambda\theta\theta' + \theta\theta'\Lambda\big) + \mathrm{rd}\big(\theta\theta'\Lambda\theta\theta'\big)\\ - \mathrm{rd}\big(\theta\theta'\Lambda\,\mathrm{diag}(\theta\theta') + \mathrm{diag}(\theta\theta')\Lambda\theta\theta'\big).\end{aligned}$$

This matrix clearly belongs to $\mathrm{Sp}(\theta)$, as required. □

REFERENCES

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press, Baltimore.

Brahimi, B. and Necir, A. (2012). A semiparametric estimation of copula models based on the method of moments. *Statistical Methodology* **9** 467–477.

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association* **101** 1228–1240.

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

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

Genest, C. and Rivest, L. P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* **88** 1034–1043.

Genest, C. and Werker, B. J. M. (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In *Proceedings of the Conference on Distributions With Given Marginals and Statistical Modelling* (C. M. Cuadras and J. A. R. Lallena, eds.) 103–112. Kluwer Academic Publishers, Dordrecht.

Gordon, R. D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *The Annals of Mathematical Statistics* **12** 364–366.

Hardy, G. W., Littlewood, J. E. and Pólya, G. (1934). *Inequalities*. Cambridge at the University Press, Cambridge.

Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli* **19** 462–491.

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics* **1** 265–283.

Hoff, P. D., Niu, X. and Wellner, J. A. (2012). Information bounds for Gaussian copulas. *Bernoulli* forthcoming.

Klaassen, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *The Annals of Statistics* **15** 1548–1562.

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

Klüppelberg, C. and Kuhn, G. (2009). Copula structure analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71** 737–753.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics—Some Basic Concepts*. Springer, New York.

Li, Q., Brown, J. B., Huang, H. and Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* **5** 1752–1779.

Liebscher, E. (2009). Semiparametric estimation of the parameters of multivariate copulas. *Kybernetika* **45** 972–991.

Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics* **40** 2293–2326.

Loève, M. (1977). *Probability Theory I*. Springer-Verlag, New York.

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd ed. *Wiley series in probability and statistics*. J. Wiley & Sons, Chichester.

Oakes, D. (1986). Semi-parametric inference in a model for association in bivariate survival data. *Biometrika* **73** 353–361.

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

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics* **33** 357–375.

Van der Vaart, A. W. (1988). *Statistical Estimation in Large Parameter Spaces*. *CWI tract* **44**. CWI, Amsterdam.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Université catholique de Louvain
ISBA, Voie du Roman Pays 20
B-1348 Louvain-la-Neuve
Belgium
E-mail: Johan.Segers@UCLouvain.be

Tilburg University
PO Box 90153
5000 LE Tilburg
The Netherlands
E-mail: R.vdnAkker@TilburgUniversity.edu
E-mail: B.J.M.Werker@TilburgUniversity.edu

Fig 1: Simulation results, based on 15,000 replications, for the Toeplitz model in Example 5.2 in dimension $p = 4$ at $\theta = (0.4945, -0.4593, -0.8462)'$ and sample size $n = 50$ (left) and $n = 250$ (right). Boxplots of $\hat{\theta}^{\mathrm{PLE}}_{n,m} - \theta_m$ and $\hat{\theta}^{\mathrm{OSE}}_{n,m} - \theta_m$ for $m = 1$ (top), $m = 2$ (middle) and $m = 3$ (bottom).

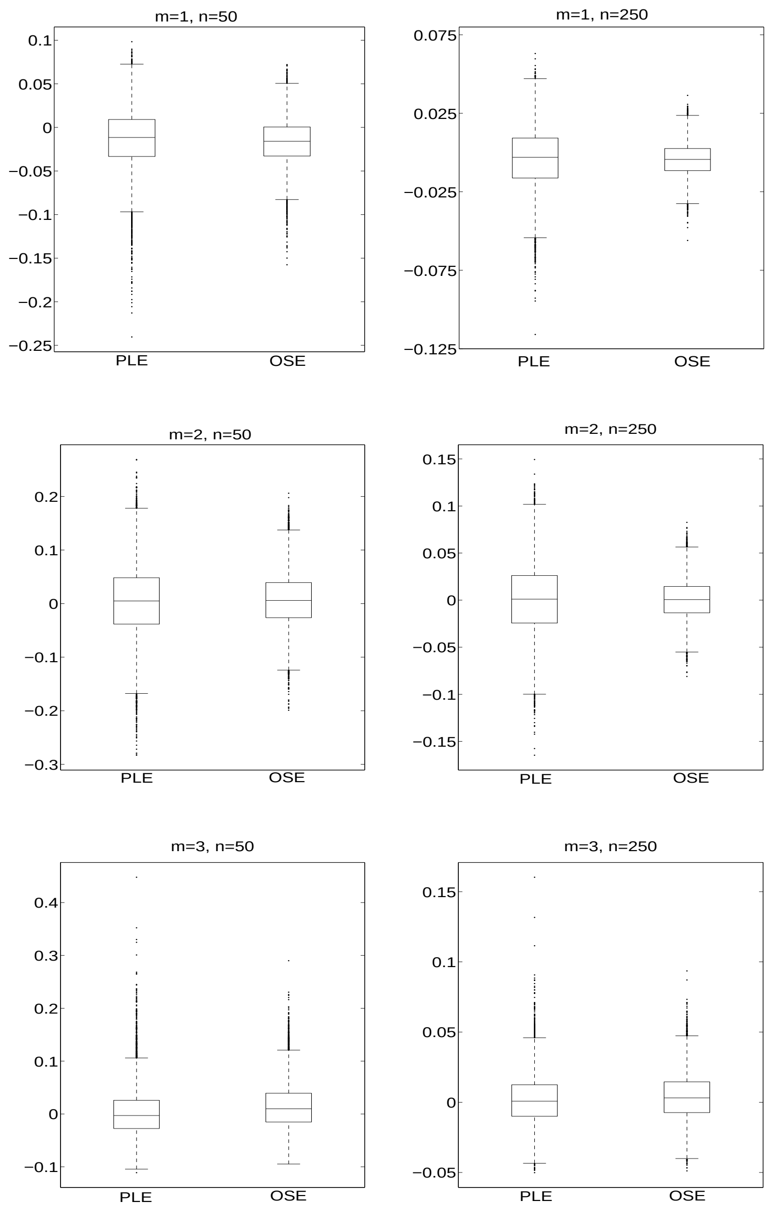
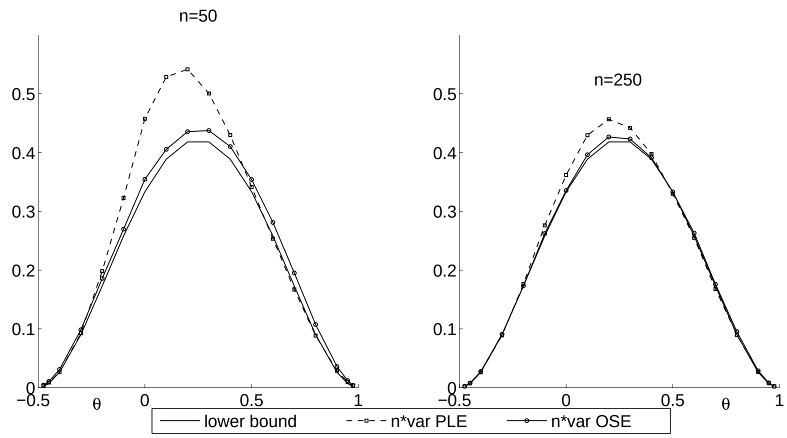
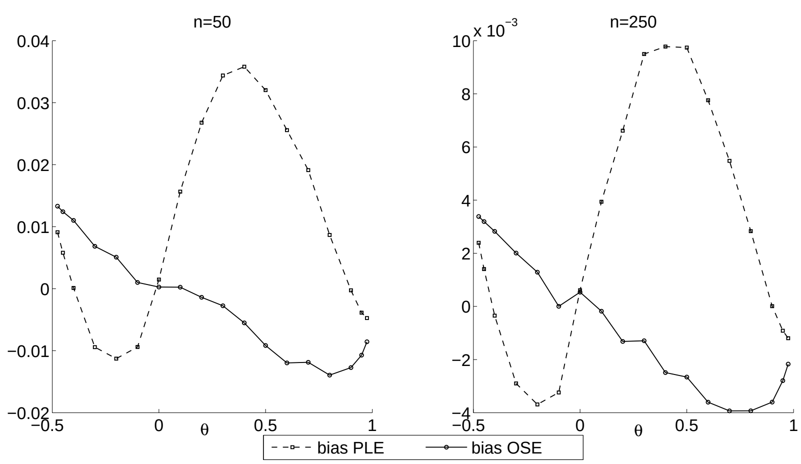


Fig 2: Simulation results for the exchangeable model of Example 5.3 in dimension $p = 3$ for $\theta \in \{-.475, -.45\} \cup \{k/10 \mid k = -4, \ldots, 9\} \cup \{.95, .975\}$ and $n \in \{50, 250\}$.



(a) The semiparametric lower bound $I^{*-1}(\theta)$ and approximations (based on 15,000 replications) to $n\,\mathrm{var}_\theta(\hat{\theta}_n^{\mathrm{PLE}})$ and $n\,\mathrm{var}_\theta(\hat{\theta}_n^{\mathrm{OSE}})$ as a function of θ.



(b) Approximations (based on 15,000 replications) to the biases $\mathbb{E}_\theta[\hat{\theta}_n^{\mathrm{PLE}}] - \theta$ and $\mathbb{E}_\theta[\hat{\theta}_n^{\mathrm{OSE}}] - \theta$ as a function of θ.

Fig 3: Boxplots of $\hat{\theta}_n^{\mathrm{PLE}} - \theta$ and $\hat{\theta}_n^{\mathrm{OSE}} - \theta$ (15,000 replications) for the exchangeable model of Example 5.3 in dimension $p = 100$ at $\theta = .25$ and $n = 50$.

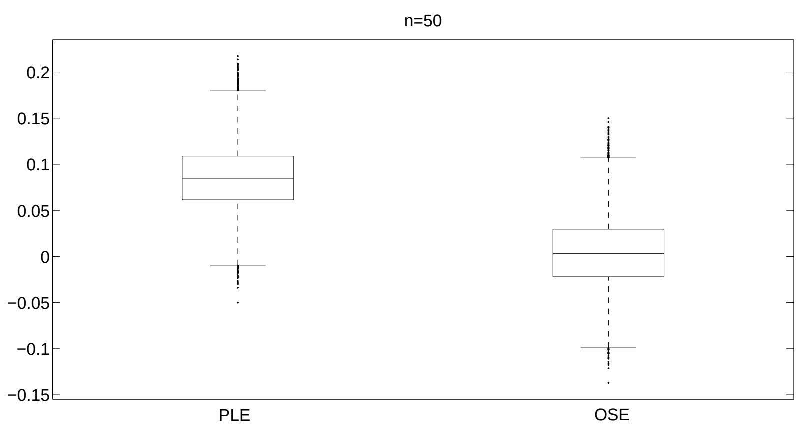
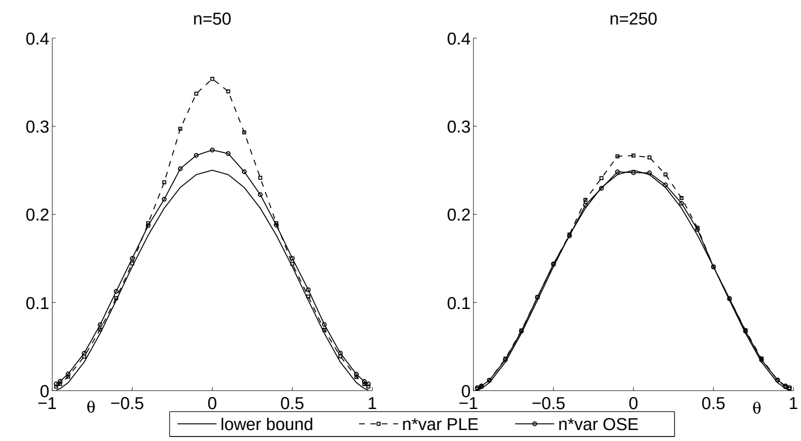
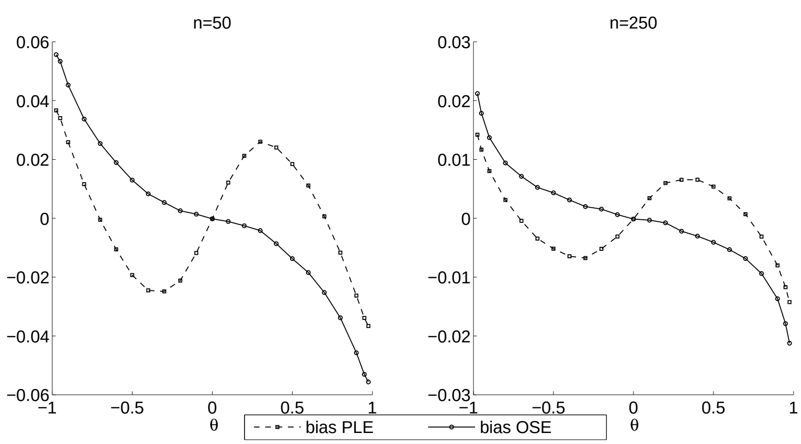


Fig 4: Simulation results for the circular model of Example 5.4 in dimension $p = 4$ for $\theta \in \{-.975, -.95\} \cup \{k/10 \mid k = -9, \ldots, 9\} \cup \{.95, .975\}$ and $n \in \{50, 250\}$.



(a) The semiparametric lower bound $I^{*-1}(\theta)$ and approximations (based on 15,000 replications) to $n\,\mathrm{var}_\theta(\hat{\theta}_n^{\mathrm{PLE}})$ and $n\,\mathrm{var}_\theta(\hat{\theta}_n^{\mathrm{OSE}})$ as a function of θ.



(b) Approximations (based on 15,000 replications) to the biases $\mathbb{E}_\theta[\hat{\theta}_n^{\mathrm{PLE}}] - \theta$ and $\mathbb{E}_\theta[\hat{\theta}_n^{\mathrm{OSE}}] - \theta$ as a function of θ.