Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Johan Segers\textsuperscript{1}  Ramon van den Akker\textsuperscript{2}  Bas J.M. Werker\textsuperscript{2}

\textsuperscript{1}Université catholique de Louvain (BE)  
Institut de statistique, biostatistique et sciences actuarielles

\textsuperscript{2}Tilburg University (NL)  
CentER

45e Journées de Statistique  
Toulouse, May 27–31, 2013
How to recover the correlation matrix of latent Gaussian variables?

\[
\begin{pmatrix}
Z_1 \\
Z_2
\end{pmatrix} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \right)
\]

\[
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix} = \begin{pmatrix}
\eta_1(Z_1) \\
\eta_2(Z_2)
\end{pmatrix}
\]
Increasing transformations of a latent Gaussian vector with standard margins and unknown correlation matrix

Observables: $p$-variate sample $X_1, \ldots, X_n$

Model: $X_i$ are iid $X = (X_1, \ldots, X_p)'$ where

$$X_j = \eta_j(Z_j), \quad j = 1, \ldots, p,$$

$$Z = (Z_1, \ldots, Z_p)' \sim N_p(0, R(\theta))$$

where

- $R(\theta)$ is a $p \times p$ correlation matrix indexed by $\theta \in \Theta \subset \mathbb{R}^k$
- $p$ unknown strictly increasing functions $\eta_j : \mathbb{R} \to \mathbb{R}$

Contribution

Efficient inference on parameter vector $\theta$ in the presence of infinite-dimensional nuisance parameters $\eta_1, \ldots, \eta_p$
Solution in the bivariate case: normal scores rank correlation

Suppose $p = 2$. Normal scores rank correlation coefficient:

1. Compute component-wise ranks: for $i = 1, \ldots, n$ and $j = 1, 2$

   $$ R_{ij} = R_{ij}^{(n)} = \text{rank of } X_{ij} \text{ among } X_{1j}, \ldots, X_{nj} $$

2. Compute van der Waerden scores:

   $$ \hat{Z}_{ij} = \Phi^{-1}(R_{ij}/(n + 1)) $$

3. Compute their correlation:

   $$ \hat{\theta}_n = \frac{\frac{1}{n} \sum_{i=1}^{n} \hat{Z}_{i1} \hat{Z}_{i2}}{\frac{1}{n} \sum_{i=1}^{n} \{\Phi^{-1}(i/(n + 1))\}^2} $$

Semiparametrically efficient [KLAASSEN & WELLNER (1997)]
Higher dimensions: structured correlation matrices

Some $k$-dimensional models for $p \times p$ correlation matrices $R(\theta)$:

- **Full model:** if $p = 3$,

$$R(\theta) = \begin{pmatrix} 1 & \theta_{12} & \theta_{13} \\ \cdot & 1 & \theta_{23} \\ \cdot & \cdot & 1 \end{pmatrix}, \quad k = p(p - 1)/2$$

Pairwise normal scores rank correlations still efficient

[KLAASSEN & WELLNER (1997)]

- **Toeplitz matrices:** if $p = 4$:

$$R(\theta) = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \theta_3 \\ \cdot & 1 & \theta_1 & \theta_2 \\ \cdot & \cdot & 1 & \theta_1 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad k = p - 1$$

- **Exchangeable models, circular matrices, factor models, . . .**
Invariance suggests rank-based inference

Applying arbitrary increasing transformations $T_1, \ldots, T_p$ produces

$$T(X) = (T_1(X_1), \ldots, T_p(X_p))$$

$$= (T_1 \circ \eta_1(Z_1), \ldots, T_p \circ \eta_p(Z_p)), \quad Z \sim N_p(0, R(\theta))$$

The parameter of interest, $\theta$, remains the same.

Requirement

*The estimator $\hat{\theta}_n$ is invariant w.r.t. increasing transformations:* $\hat{\theta}_n(X_1, \ldots, X_n) = \hat{\theta}_n(T(X_1), \ldots, T(X_n)), \quad \text{all } T$

Therefore, the estimator $\hat{\theta}_n$ must depend on the data only through the vectors of component-wise ranks

$$\hat{\theta}_n(X_1, \ldots, X_n) = \hat{\theta}_n(R_1, \ldots, R_n),$$

$$R_i = (R_{i1}, \ldots, R_{ip})'$$
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
- The infeasible MLE
- The PLE
- The one-step update estimator (new)

Asymptotics
- Asymptotic distribution
- Efficiency comparisons

Tangent space geometry
The Gaussian copula

To link up with the literature, use copulas.
Joint and marginal distributions of \( \mathbf{Z} = (Z_1, \ldots, Z_p)' \sim N_p(\mathbf{0}, R(\theta)) \):

\[
\Phi_{R(\theta)}(z_1, \ldots, z_p) = \Pr(Z_1 \leq z_1, \ldots, Z_p \leq z_p),
\]

\[
\Phi(z_j) = \Pr(Z_j \leq z_j), \quad N(0, 1)
\]

Probability integral transform:

\[
U_j = \Phi(Z_j) \sim \text{Uniform}(0, 1), \quad j = 1, \ldots, p
\]

The joint distribution of \( \mathbf{U} = (U_1, \ldots, U_j)' \) is the Gaussian copula:

\[
C_{R(\theta)}(u_1, \ldots, u_p) = \Pr(U_1 \leq u_1, \ldots, U_p \leq u_p)
\]

\[
= \Phi_{R(\theta)}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p))
\]

From now on, abbreviate \( \Phi_\theta = \Phi_{R(\theta)} \) and \( C_\theta = C_{R(\theta)} \) etc.
The transformation model is a copula model

Recall $X_j = \eta_j(Z_j)$ with $Z \sim N_p(0, R(\theta))$.

Marginal distribution functions:

$$F_j(x_j) = \Pr(X_j \leq x_j) = \Pr[\eta_j(Z_j) \leq x_j] = \Phi(\eta_j^{-1}(x_j))$$

Joint distribution function:

$$F(x_1, \ldots, x_p) = \Pr(X_1 \leq x_1, \ldots, X_p \leq x_p)$$
$$= \Pr[\eta_1(Z_1) \leq x_1, \ldots, \eta_p(Z_p) \leq x_p]$$
$$= \Pr[\Phi(Z_1) \leq F_1(x_1), \ldots, \Phi(Z_p) \leq F_p(x_p)]$$
$$= C_\theta(F_1(x_1), \ldots, F_p(x_p))$$

with $C_\theta$ the Gaussian copula with correlation matrix $R(\theta)$.

Decomposition is a particular case of Sklar’s theorem.
Finite-dimensional parameter of interest, infinite-dimensional nuisance parameters

Semiparametric model:

$$(X_1, \ldots, X_p) = (\eta_1(Z_1), \ldots, \eta_p(Z_p))$$

where $Z \sim N_p(0, R(\theta))$

$$F(x_1, \ldots, x_p) = C_\theta(F_1(x_1), \ldots, F_p(x_p))$$

where $C_\theta$ is Gaussian $R(\theta)$-copula

Parameter of interest: correlation parameter $\theta \in \Theta \subset \mathbb{R}^k$ in dimension $k \leq p(p - 1)/2$

Nuisance parameters: functions $\eta_1, \ldots, \eta_p$ or, alternatively, the margins $F_1, \ldots, F_p$ (infinite-dimensional)
Questions

Information bound for $\theta$?

- Minimal asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta)$ for regular estimators?
- Compare with information bounds based on rank likelihood
  
  [Hoff, Niu & Wellner (2013)]

Efficient, rank-based estimators?

- Estimator achieving the minimal asymptotic variance?
- Finite-sample performance?
- Compare with pseudo-likelihood estimator  [Genest, Ghoudi & Rivest (1995)]
- Efficient sieve estimator for semiparametric copula models: not rank-based  [Chen, Fan & Tsyrennikov (2006)]

Information loss?

- Price to pay for not knowing the margins?
- Adaptivity?
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
- The infeasible MLE
- The PLE
- The one-step update estimator (new)

Asymptotics
- Asymptotic distribution
- Efficiency comparisons

Tangent space geometry
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
- The infeasible MLE
- The PLE
- The one-step update estimator (new)

Asymptotics
- Asymptotic distribution
- Efficiency comparisons

Tangent space geometry
Densities of latent and observable variables

Assumption

- \( R(\theta) \) is of full rank; put \( S(\theta) = R(\theta)^{-1} \)
- \( F_1, \ldots, F_p \) possess Lebesgue densities \( f_1, \ldots, f_p \)

1. Density of \( Z = (Z_1, \ldots, Z_p) \):

   \[
   \varphi(\theta)(z) = \frac{1}{\sqrt{(2\pi)^p \det R(\theta)}} \exp \left\{ -\frac{1}{2} z' S(\theta) z \right\}
   \]

2. Density of \( U = (\Phi(Z_1), \ldots, \Phi(Z_p))' \):

   \[
   c(u; \theta) = \frac{\varphi(\theta)(Z_1, \ldots, Z_p)}{\varphi(Z_1) \cdots \varphi(Z_p)}, \quad z_j = \Phi^{-1}(u_j)
   \]

3. Density of \( X = (F_1^{-1}(U_1), \ldots, F_p^{-1}(U_p))' \):

   \[
   f(x) = c(F_1(x_1), \ldots, F_p(x_p); \theta) f_1(x_1) \cdots f_p(x_p)
   \]
If margins were known, we could estimate the correlation parameter by maximum likelihood

If margins $f_1, \ldots, f_p$ are known, the model is parametric in $\theta$. Estimate $\theta$ by, for instance, maximum likelihood:

$$
\hat{\theta}_{n,\text{MLE}} = \arg \max_{\theta \in \Theta} \sum_{i=1}^{n} \left( \log c(F_1(X_{i1}), \ldots, F_p(X_{ip}); \theta) + \sum_{j=1}^{p} \log f_j(X_{ij}) \right)
$$

Under regularity conditions on $\theta \mapsto R(\theta)$, the MLE behaves as expected, see below.
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
- The infeasible MLE
- The PLE
- The one-step update estimator (new)

Asymptotics
- Asymptotic distribution
- Efficiency comparisons

Tangent space geometry
If margins are unknown, estimate them nonparametrically and pretend they are known

Pseudo-likelihood estimator for $\theta$

1. Estimate $F_j$ by the empirical distribution function

\[
\hat{F}_{n,j}(x_j) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbf{1}(X_{ij} \leq x_j)
\]

2. Pretend these are the true margins and use MLE:

\[
\hat{\theta}_{n,PLE} = \arg \max_{\theta \in \Theta} \sum_{i=1}^{n} \log c(\hat{F}_{n,1}(X_{i1}), \ldots, \hat{F}_{n,p}(X_{ip}); \theta)
\]

- The estimator is rank-based: \( \hat{F}_{n,j}(X_{ij}) = \frac{1}{n+1} R_{ij} \)
- Pseudo-likelihood: margins are ignored
Although not necessarily efficient, the PLE works quite well in practice

- Estimation strategy applies to general copula models, but the PLE need not semiparametrically efficient
  \[\text{Genest, Ghoudi, Rivest (1995), Genest \& Werker (2002)}\]
- For multivariate Gaussian copula models, \( R(\theta) \), compare with information bound from rank likelihood:
  - For some models, the PLE is efficient, e.g. full model
  - For some other models, the PLE is not efficient, although it still performs quite well, e.g. circular model
  \[\text{Hoff, Niu \& Wellner (2013)}\]
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
The infeasible MLE
The PLE
The one-step update estimator (new)

Asymptotics
Asymptotic distribution
Efficiency comparisons

Tangent space geometry
Building blocks of the estimator: some linear algebra

Linear algebra conventions:

\[ l_p = p \times p \text{ identity matrix} \]

\[ \nu_p = (1, \ldots, 1)' \in \mathbb{R}^p \]

\[ A \circ B = (A_{ij}B_{ij})_{ij} \quad \text{elementwise product of matrices} \]

\[ \text{diag}(b) = \text{diagonal matrix with diagonal } b \]

Partial derivatives of \( R(\theta) \) and \( S(\theta) = R(\theta)^{-1} \):

\[ \dot{R}_m(\theta) = \partial R(\theta)/\partial \theta_m \]

\[ \dot{S}_m(\theta) = \partial S(\theta)/\partial \theta_m \]
Efficient scores and their covariance matrix

Verify that the following quantities can be readily computed:

\[ g_m(\theta) = -\left( I_p + R(\theta) \circ S(\theta) \right)^{-1} \left( \dot{R}_m(\theta) \circ S(\theta) \right) \nu_p \]

\[ D_\theta(b) = S(\theta) \text{diag}(b) + \text{diag}(b) S(\theta) \]

\[ A_m(\theta) = D_\theta(g_m(\theta)) - \dot{S}_m(\theta) \]

Efficient score function

For each component \( m = 1, \ldots, k \) of \( \theta \):

\[ \ell_{\theta,m}(u; \theta) = \frac{1}{2} z' A_m(\theta) z, \quad z_j = \Phi^{-1}(u_j) \]

Efficient information matrix

For \( m, m' = 1, \ldots, k \):

\[ I_{mm'}^*(\theta) = \frac{1}{2} \text{tr}\{ R(\theta) A_m(\theta) R(\theta) A_{m'}(\theta) \} \]
Description of the one-step estimator: updating an initial estimator

1. Compute $\hat{F}_{n,j}(X_{ij}) = R_{ij}/(n + 1)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$

2. Compute an initial, rank-based estimate $\tilde{\theta}_n$
   - Should be $\sqrt{n}$-consistent.
   - For instance take the PLE.
   - In theory, needs to discretized to a grid in $\mathbb{Z}^k$ of mesh $n^{-1/2}$.

3. Compute $A_m(\tilde{\theta}_m)$ for $m = 1, \ldots, k$

4. Compute $\ell_{\tilde{\theta},m}(\cdot; \tilde{\theta}_n)$ and $l_{mm'}^*(\tilde{\theta}_n)$ for $m, m' = 1, \ldots, k$

5. Compute the one-step update estimator:

$$\hat{\theta}_{n,\text{OSE}} = \tilde{\theta}_n + \frac{1}{\sqrt{n}} \sum_{i=1}^n l^*(\tilde{\theta}_n)^{-1} \ell_{\tilde{\theta}}^*(\hat{F}_{n,1}(X_{i1}), \ldots, \hat{F}_{n,p}(X_{ip}); \tilde{\theta}_n)$$
Getting some feeling for the one-step estimator

\[ \hat{\theta}_{n,\text{OSE}} = \tilde{\theta}_n + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^*(\tilde{\theta}_n)^{-1} \dot{l}_{\theta}^*(\hat{F}_{n,1}(X_{i1}), \ldots, \hat{F}_{n,p}(X_{ip}); \tilde{\theta}_n) \]

- Reminiscent of one-step update estimators in parametric models
  - The “efficient score” replaces the ordinary score function
- If initial estimator is rank-based, so is one-step estimator
- Update step is easy to implement – linear algebra only

Q: So where does it come from?
A: Tangent space calculations…

Q: Cute, but does it really work?
A: Yes!
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
   The infeasible MLE
   The PLE
   The one-step update estimator (new)

Asymptotics
   Asymptotic distribution
   Efficiency comparisons

Tangent space geometry
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
The infeasible MLE
The PLE
The one-step update estimator (new)

Asymptotics
Asymptotic distribution
Efficiency comparisons

Tangent space geometry
Semiparametric Gaussian copula model

Let

\[ \mathcal{F}_{ac} = \{ \text{absolutely continuous distributions on } \mathbb{R} \} \]

\[ P_{\theta,F_1,\ldots,F_p} = \text{law of } \boldsymbol{X} \text{ with copula } C_\theta \text{ and margins } F_1, \ldots, F_p \]

Model for one observation \( \boldsymbol{X} \):

\[ \mathcal{P} = \left( P_{\theta,F_1,\ldots,F_p} \mid \theta \in \Theta, \ F_1, \ldots, F_p \in \mathcal{F}_{ac} \right), \]

Data-generating process: \( \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \) iid \( \boldsymbol{X} \).
Assumption on the correlation matrices

Suppose $\Theta \subset \mathbb{R}^k$ is open and for all $\theta \in \Theta$:

(i) The inverse $S(\theta) = R^{-1}(\theta)$ exists.

(ii) The matrices of partial derivatives $\dot{R}_m(\theta)$, for $m = 1, \ldots, k$, exist and are continuous in $\theta$.

(iii) The matrices $\dot{R}_1(\theta), \ldots, \dot{R}_k(\theta)$ are linearly independent.

Under this assumption, the parametric model in $\theta$ with known margins in $\mathcal{F}_{ac}$ is regular.
The one-step estimator is efficient

**Theorem**

Suppose there exists a rank-based estimator \( \tilde{\theta}_n \) such that

\[
\tilde{\theta}_n = \theta + O_p(1/\sqrt{n}) \quad \text{under every } P_{\theta,F_1,\ldots,F_p} \in \mathcal{P}
\]

Then for all \( F_1, \ldots, F_p \in \mathcal{F}_{ac} \) and \( \theta \in \Theta \),

\[
\sqrt{n} \left( \hat{\theta}_{n,OSE} - \theta \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I^*_{-1}(\theta) \ell^{*}_{\theta}(F_1(X_{i1}), \ldots, F_p(X_{ip}); \theta) + o_P(1)
\]

\[\xrightarrow{d} N_k(0, I^*(\theta)^{-1})\]

Moreover, the one-step estimator is an efficient estimator of \( \theta \) in the semiparametric Gaussian copula model \( \mathcal{P} \).
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
The infeasible MLE
The PLE
The one-step update estimator (new)

Asymptotics
Asymptotic distribution
Efficiency comparisons

Tangent space geometry
Asymptotic covariance matrices: The OSE is at least as efficient as the PLE

For the MLE for \( \theta \) if margins are known:

\[
I(\theta)^{-1} \quad \text{where} \quad I_{mm'}(\theta) = \frac{1}{2} \text{tr}\{R(\theta) \dot{S}_m(\theta) R(\theta) \dot{S}_{m'}(\theta)\}
\]

For the one-step estimator:

\[
I^*(\theta)^{-1} \geq I(\theta)^{-1}
\]

For the pseudo-likelihood estimator:

\[
\Sigma_{\text{PLE}}(\theta) \geq I^*(\theta)^{-1}
\]

Notation: \( A \succeq B \) iff \( A - B \) is positive semi-definite
Still, the PLE is often (almost) as efficient as the OSE

Asymptotic relative efficiency of pseudo-likelihood estimator: (nearly) 1

\[ R(\theta) = \begin{pmatrix}
1 & \theta & \theta \\
\cdot & 1 & \theta \\
\cdot & \cdot & 1
\end{pmatrix} \]

\[ R(\theta) = \begin{pmatrix}
1 & \theta & \theta^2 & \theta \\
\cdot & 1 & \theta & \theta^2 \\
\cdot & \cdot & 1 & \theta \\
\cdot & \cdot & \cdot & 1
\end{pmatrix} \]
Even when the PLE is efficient
the OSE does still a bit better in finite samples

Exchangeable model: Finite-sample variances

\[
R(\theta) = \begin{pmatrix}
1 & \theta & \theta \\
\cdot & 1 & \theta \\
\cdot & \cdot & 1
\end{pmatrix}, \quad -1/2 < \theta < 1
\]
Even when the PLE is efficient, the OSE does still a bit better in finite samples.

Exchangeable model: Finite-sample bias

\[ R(\theta) = \begin{pmatrix} 1 & \theta & \theta \\ \cdot & 1 & \theta \\ \cdot & \cdot & 1 \end{pmatrix}, \quad -1/2 < \theta < 1 \]
Even when the PLE is nearly efficient the OSE does still a bit better in finite samples

Circular model: Finite-sample variances

\[
R(\theta) = \begin{pmatrix}
1 & \theta & \theta^2 & \theta \\
\cdot & 1 & \theta & \theta^2 \\
\cdot & \cdot & 1 & \theta \\
\cdot & \cdot & \cdot & 1 \\
\end{pmatrix}, \quad -1 < \theta < 1
\]
Even when the PLE is nearly efficient the OSE does still a bit better in finite samples.

Circular model: Finite-sample bias

\[ R(\theta) = \begin{pmatrix} 1 & \theta & \theta^2 & \theta \\ \cdot & 1 & \theta & \theta^2 \\ \cdot & \cdot & 1 & \theta \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad -1 < \theta < 1 \]
In high dimensions, the OSE seems less biased

\[ R(\theta) = \begin{pmatrix} 1 & \theta & \cdots & \cdots & \theta \\ \cdot & 1 & \theta & \cdots & \theta \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & 1 \end{pmatrix} \]

\[ p = 100, \quad n = 50 \]
Sometimes, the PLE is quite inefficient

Toeplitz model in $p = 4$: boxplots for $\hat{\theta}_{n,1} - \theta_1$

$$R(\theta) = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \theta_3 \\ \cdot & 1 & \theta_1 & \theta_2 \\ \cdot & \cdot & 1 & \theta_1 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

$\theta = (0.4945, -0.4593, -0.8462)$
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Semiparametric Gaussian copula models

Estimators
- The infeasible MLE
- The PLE
- The one-step update estimator (new)

Asymptotics
- Asymptotic distribution
- Efficiency comparisons

Tangent space geometry
Intermezzo: the Fréchet–Cramér–Rao inequality

Parametric model \( \{ f_\theta : \theta \in \mathbb{R} \} \), statistic \( T(X) \). Score

\[
\dot{\ell}_\theta(X) = \frac{\partial}{\partial \theta} \log f_\theta(X)
\]

By Cauchy–Schwarz:

\[
(cov_\theta \{ T(X), \dot{\ell}_\theta(X) \})^2 \leq var_\theta \{ T(X) \} \cdot \left[ var_\theta \{ \dot{\ell}_\theta(X) \} \right] = I(\theta)
\]

On the other hand,

\[
cov_\theta \{ T(X), \dot{\ell}_\theta(X) \} = \ldots = \frac{\partial}{\partial \theta} \int T(x) f_\theta(x) \mu(dx) = E_\theta T(X)
\]

As a consequence, we find a lower bound for the variance of \( T(X) \):

\[
var_\theta \{ T(X) \} \geq I(\theta)^{-1} \left\{ \partial E_\theta T(X) / \partial \theta \right\}^2
\]
Efficiency via tangent spaces

- Estimation of $\theta$ in the semiparametric model is at least as hard as in a parametric submodel.
- For a parametric submodel, the inverse Fisher information gives a lower bound for the asymptotic variance of regular estimators.
- The largest such lower bound is a lower bound for the asymptotic variance of a regular estimator in the semiparametric model.
- This lower bound can be found via the geometry of tangent spaces and the theory of limits of experiments.

[Le Cam & Yang (1990), Bickel, Ritov, Klaassen & Wellner (1993), van der Vaart (1998), ...]
Tangent space of the model at a distribution:
collection of score functions of parametric submodels

Recall

\[ \mathcal{F}_{ac} = \{ \text{absolutely continuous distributions on } \mathbb{R} \} \]
\[ P_{\theta,F_1,...,F_p} = \text{law of } X \text{ with copula } C_\theta \text{ and margins } F_1, \ldots, F_p \]

Tangent space at \( P_{\theta,F_1,...,F_p} \in \mathcal{P} \):
collection of scores functions of local parametric submodels

\[
\frac{\partial}{\partial \eta} \log p_{\theta+\eta\alpha,F_1,\eta,...,F_p,\eta}(x) \bigg|_{\eta=0}, \quad x \in \mathbb{R}^p,
\]

- \( \eta \mapsto F_{j,\eta} \) is a path in \( \mathcal{F}_{ac} \) that passes through \( F_j \) at \( \eta = 0 \)
- \( p_{\theta+\eta\alpha,F_1,\eta,...,F_p,\eta} \) is the density of \( P_{\theta+\eta\alpha,F_1,\eta,...,F_p,\eta} \)

Local description of the model in \( L^2(P_{\theta,F_1,...,F_p}) \)
The tangent space is the sum of a parametric and a nonparametric part

Tangent space at \( P_\theta = P_{\theta, F_1, \ldots, F_p} \) for \( F_j \) Uniform(0, 1):

- **Parametric part**: only \( \theta \) changes, whereas the margins are fixed

  \[
  \text{linear span of } \ell_{\theta, m}(u; \theta) = \frac{\partial}{\partial \theta_m} \log c(u; \theta)
  \]

- **Nonparametric part**: only the margins change, whereas \( \theta \) is fixed

  \[
  \text{linear span of } h(u_j) + \ell_j(u; \theta) \int_0^{u_j} h(v) \, dv
  \]

where

- \( h \in L^2([0, 1]) \) and \( \int_0^1 h(v) \, dv = 0 \)
- \( \ell_j(u) = \frac{\partial}{\partial u_j} \log c(u; \theta) \)
The efficient score function is a projection of the parametric score function

Parametric and nonparametric scores quantify how the distribution changes if $\theta$ and the margins change.

If parametric and nonparametric scores are correlated, not knowing the margins makes identifying changes in $\theta$ harder. Otherwise: adaptivity – not knowing the margins does not matter.

Efficient score function $\dot{\ell}_\theta(u; \theta)$: orthogonal projection in $L^2(P_\theta)$ of parametric scores on the ortocomplement of the space of nonparametric scores.

Efficient information matrix $I^*(\theta)$: variance matrix of the efficient score function. Its inverse yields a lower bound for the variance of regular estimators.
For Gaussian copulas, the efficient score function can be explicitly computed.

For general copula models, computing the efficient score function amounts to a system of coupled Sturm–Liouville differential equations.

For Gaussian copula models, these equations can be solved explicitly, leading to the expression stated earlier:

\[
\dot{\ell}^*,m(u; \theta) = \frac{1}{2} z' A_m(\theta) z,
\]

where

\[
A_m(\theta) = D_\theta(g_m(\theta)) - \dot{S}_m(\theta)
\]

\[
D_\theta(b) = S(\theta) \text{ diag}(b) + \text{ diag}(b) S(\theta)
\]

\[
g_m(\theta) = - (I_p + R(\theta) \circ S(\theta))^{-1} \left( \dot{R}_m(\theta) \circ S(\theta) \right) \nu_p
\]

Whence the paper…
All relevant functions are quadratic forms

All relevant score and influence functions turn out to be centered quadratic forms in the Gaussianized observations $z$: 

$$q_A(z) = \frac{1}{2} z' A z - \frac{1}{2} \mathbb{E}_\theta [Z' A Z]$$

Identifying $q_A$ with $A$ (symmetric $p \times p$) leads to an inner product for matrices that also appeared in the efficient information matrix:

$$\langle A, B \rangle_\theta = \text{cov}_\theta \{ q_A(Z), q_B(Z) \} = \frac{1}{2} \text{tr} \{ R(\theta) A R(\theta) B \}$$

Recall the efficient information matrix:

$$I_{mm'}^*(\theta) = \langle A_m(\theta), A_{m'}(\theta) \rangle_\theta$$
Quadratic forms and symmetric matrices

- Statistical interpretation of reduction to quadratic forms: least favourable submodel is Gaussian with unknown variances
  \[ \text{[Hoff, Niu & Wellner (2013)]} \]

- Identification with matrices yields convenient criteria for
  - (in)efficiency of the pseudo-likelihood estimator
  - adaptivity: iff \( \text{diag}(R(\theta) \dot{S}_m(\theta)) = 0 \) for all \( m = 1, \ldots, k \)
Conclusion: efficient inference for (Gaussian) copulas

Contributions

- Inference in semiparametric Gaussian copula model with structured correlation matrices
- One-step estimator: rank-based and semiparametrically efficient
- Outperforms the popular pseudo-likelihood estimator both asymptotically and in finite samples
- Adaptivity occurs at independence and exceptionally in other cases

Next: general semiparametric copula models?

- Efficient score function and information matrix?
- One-step estimator?

Thank you!