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

Johan Segers[1] Ramon van den Akker[2] Bas J.M. Werker[2]

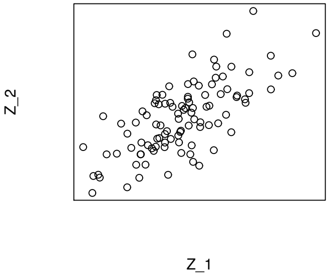
[1] Université catholique de Louvain (BE)
Institut de statistique, biostatistique et sciences actuarielles

[2] Tilburg University (NL)
CentER

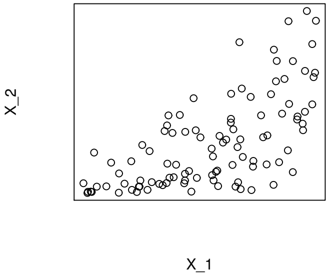
Conference on Copulas and Dependence: Theory and Applications
11–12 October 2013, Columbia University, New York City

How to recover the correlation matrix of latent Gaussian variables?

latent



observable



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

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \eta_1(Z_1) \\ \eta_2(Z_2) \end{pmatrix}$$

Increasing transformations of a latent Gaussian vector with standard margins and unknown correlation matrix

Observables: p-variate sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$
Model: $\boldsymbol{X}_i$ are iid $\boldsymbol{X} = (X_1, \ldots, X_p)$ where

$$X_j = \eta_j(Z_j), \qquad j = 1, \ldots, p, \qquad \text{observable}$$
$$\boldsymbol{Z} = (Z_1, \ldots, Z_p) \sim N_p(\mathbf{0}, R(\theta)) \qquad \text{latent}$$

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 θ in the presence of infinite-dimensional nuisance parameters $\eta_1, \ldots, \eta_p$

Higher dimensions: structured correlation matrices

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

- *Full model*: e.g. if $p = 3$,

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

  The pairwise normal scores rank correlations are efficient.

  [Klaassen & Wellner (1997)]

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

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

- *Exchangeable models, circular matrices, factor models, ...*

Invariance suggests rank-based inference

Applying arbitrary increasing transformations T_j produces

$$T_j(X_j) = (T_j \circ \eta_j)(Z_j)$$

The parameter of interest, θ, remains the same.

Requirement

The estimator $\hat{\theta}_n$ is invariant w.r.t. increasing transformations:

$$\hat{\theta}_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \hat{\theta}_n(\boldsymbol{T}(\boldsymbol{X}_1), \ldots, \boldsymbol{T}(\boldsymbol{X}_n)), \qquad \textit{all } \boldsymbol{T} = (T_1, \ldots, T_p)$$

$\Rightarrow$ $\hat{\theta}_n$ depends only on the ranks

$$\hat{\theta}_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \hat{\theta}_n(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n),$$
$$R_{ij} = \text{rank of } X_{ij} \text{ among } X_{1j}, \ldots, X_{nj}$$

The latent-variable model is a copula model

Recall $\boldsymbol{X} = (X_1, \ldots, X_p)$ and $X_j = \eta_j(Z_j)$ with

- $\boldsymbol{Z} \sim N_p(\mathbf{0}, R(\theta))$
- $\eta_1, \ldots, \eta_p$ increasing functions

Then

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

with C_θ the Gaussian copula with correlation matrix $R(\theta)$:

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

- $\Phi_{R(\theta)}$ the $N_p(\mathbf{0}, R(\theta))$ joint cdf
- Φ^{-1} the $N(0, 1)$ quantile function

Finite-dimensional parameter of interest, infinite-dimensional nuisance parameters

Semiparametric model:

$$(X_1, \ldots, X_p) = \big(\eta_1(Z_1), \ldots, \eta_p(Z_p)\big) \qquad \text{where} \qquad \boldsymbol{Z} \sim N_p(\mathbf{0}, R(\theta))$$

$$F(x_1, \ldots, x_p) = C_\theta(F_1(x_1), \ldots, F_p(x_p)) \qquad \text{where} \qquad C_\theta \text{ is Gaussian } R(\theta)\text{-copula}$$

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

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

Questions

Information bound for θ?

- 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? When does not knowing the margins does not matter?

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

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

Densities of latent and observable variables

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

1. Density of $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$: Gaussian (latent)

$$\varphi_\theta(\boldsymbol{z}) = \frac{1}{\sqrt{(2\pi)^p \det R(\theta)}} \exp\{-\frac{1}{2} \boldsymbol{z}' \, S(\theta) \, \boldsymbol{z}\}$$

2. Density of $\boldsymbol{U} = (\Phi(Z_1), \ldots, \Phi(Z_p))$: Uniform (latent)

$$c(\boldsymbol{u}; \theta) = \frac{\varphi_\theta(z_1, \ldots, z_p)}{\varphi(z_1) \cdots \varphi(z_p)}, \qquad z_j = \Phi^{-1}(u_j)$$

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

$$f(\boldsymbol{x}) = c(F_1(x_1), \ldots, F_p(x_p); \theta) \, f_1(x_1) \cdots f_p(x_p)$$

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 θ:

$$f(\boldsymbol{x}) = c(F_1(x_1), \ldots, F_p(x_p); \theta) \; f_1(x_1) \cdots f_p(x_p)$$

Maximum likelihood estimator:

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

Under regularity conditions on $\theta \mapsto R(\theta)$,
the MLE behaves as expected, see below.

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

Pseudo-likelihood estimator for θ

1. Estimate F_j by the empirical distribution function

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

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

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

- 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 be semiparametrically efficient

  [Genest, Ghoudi, Rivest (1995), Genest & Werker (2002)]

- For multivariate Gaussian copula models, the PLE is efficient for some models and not efficient for some other ones.

  [Hoff, Niu & Wellner (2013)]

Efficient scores and their covariance matrix

For $\theta \in \Theta \subset \mathbb{R}^k$ and $m = 1, \ldots, k$, let

$$A_m(\theta) = [\text{easily computable matrix in terms of } R(\theta) \\ \text{and its partial derivatives w.r.t. } \theta_m] \qquad \in \mathbb{R}^{p \times p}$$

Efficient score function

For each component $m = 1, \ldots, k$ of θ:

$$\dot{\ell}^*_{\theta,m}(\boldsymbol{u}; \theta) = \frac{1}{2} \boldsymbol{z}' \, A_m(\theta) \, \boldsymbol{z}, \qquad z_j = \Phi^{-1}(u_j)$$

Efficient information matrix

For $m, m' = 1, \ldots, k$:

$$I^*_{mm'}(\theta) = \frac{1}{2} \operatorname{tr}\{R(\theta) \, A_m(\theta) \, R(\theta) \, A_{m'}(\theta)\}$$

So what is this mysterious matrix?

Verify that the following objects can be readily computed:

$$\boldsymbol{g}_m(\theta) = -\big(I_p + R(\theta) \circ S(\theta)\big)^{-1} \big(\dot{R}_m(\theta) \circ S(\theta)\big)\, \boldsymbol{\iota}_p \qquad \mathbb{R}^{p\times 1}$$

$$A_m(\theta) = S(\theta)\, \mathrm{diag}(\boldsymbol{g}_m(\theta)) + \mathrm{diag}(\boldsymbol{g}_m(\theta))\, S(\theta) - \dot{S}_m(\theta) \qquad \mathbb{R}^{p\times p}$$

- $S(\theta) = R(\theta)^{-1}$
- '$\circ$' the elementwise product of matrices
- I_p the $p \times p$ identity matrix
- $\boldsymbol{\iota}_p = (1, \dots, 1)' \in \mathbb{R}^{p\times 1}$

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 $A_m(\tilde{\theta}_{n,m})$ for $m = 1, \ldots, k$
4. Compute $\dot{\ell}^*_{\theta,m}(\,\cdot\,; \tilde{\theta}_n)$ and $I^*_{mm'}(\tilde{\theta}_n)$ for $m, m' = 1, \ldots, k$
5. Compute the one-step update estimator:

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

Getting some feeling for the one-step estimator

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

- 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

Intermezzo: the Fréchet–Cramér–Rao inequality

Consider a parametric model $\{f_\theta : \theta \in \mathbb{R}\}$. Let

$$\dot{\ell}_\theta(X) = \frac{\partial}{\partial\theta} \log f_\theta(X) \qquad \text{score function}$$

$$I(\theta) = \operatorname{var}_\theta\{\dot{\ell}_\theta(X)\} \qquad \text{Fisher information}$$

Lower bound for the variance of a statistic $T(X)$:

$$\operatorname{var}_\theta\{T(X)\} \geqslant I(\theta)^{-1} \{\partial \mathrm{E}_\theta[T(X)]/\partial\theta\}^2$$

Exercise: Proof by Cauchy–Schwarz and differentation under integral sign.

Efficiency of estimators in semiparametric models: look at worst-case parametric submodels

- Estimation of θ in the semiparametric model is at least as hard as in any 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), . . .]

Semiparametric Gaussian copula model

Let

$$\mathcal{F}_{\mathrm{ac}} = \{\text{absolutely continuous distributions on } \mathbb{R}\}$$

$$\mathrm{P}_{\theta,F_1,\ldots,F_p} = \text{law of } \boldsymbol{X} \text{ with copula } C_\theta \text{ and margins } F_1,\ldots,F_p$$

Model for one observation $\boldsymbol{X}$:

$$\mathcal{P} = \left(\mathrm{P}_{\theta,F_1,\ldots,F_p} \mid \underbrace{\theta \in \Theta}_{\text{parametric}}, \underbrace{F_1,\ldots,F_p \in \mathcal{F}_{\mathrm{ac}}}_{\text{nonparametric}}\right),$$

Data-generating process: $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ iid $\boldsymbol{X}$.

Tangent space of the model at a distribution: collection of score functions of parametric submodels

Tangent space at $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$:
collection of scores functions of local parametric submodels

$$\frac{\partial}{\partial\eta} \log p_{\theta+\eta\alpha,F_{1,\eta},\ldots,F_{p,\eta}}(\boldsymbol{x})\bigg|_{\eta=0}, \quad \boldsymbol{x} \in \mathbb{R}^p,$$

- $\eta \mapsto F_{j,\eta}$ is a path in $\mathcal{F}_{\mathrm{ac}}$ that passes through F_j at $\eta = 0$
- $p_{\theta+\eta\alpha,F_{1,\eta},\ldots,F_{p,\eta}}$ is the density of $\mathrm{P}_{\theta+\eta\alpha,F_{1,\eta},\ldots,F_{p,\eta}}$

Local description of the model $\mathcal{P}$ in $L^2(\mathrm{P}_{\theta,F_1,\ldots,F_p})$:
how do small changes to the parameters affect the joint density?

The tangent space is the sum of a parametric and a nonparametric part

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

- Parametric scores: only θ changes. Spanned by

$$\boldsymbol{u} \mapsto \frac{\partial}{\partial \theta_m} \log c(\boldsymbol{u};\theta), \qquad m = 1,\ldots,k$$

- Nonparametric scores: only the margins change. Spanned by

$$\boldsymbol{u} \mapsto h(u_j) + \frac{\partial}{\partial u_j} \log c(\boldsymbol{u};\theta) \int_0^{u_j} h(v)\,dv, \qquad j = 1,\ldots,p$$

  where $h \in L^2([0,1])$ and $\int_0^1 h(v)\,dv = 0$

The efficient score function is a projection of the parametric score function

Efficient score function $\dot{\ell}^*_\theta(\boldsymbol{u};\theta)$:
orthogonal projection in $L^2(\mathrm{P}_\theta)$ of parametric scores
on the orthocomplement 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 *Gaussian copula models*, the projections can be computed explicitly, leading to the expression stated earlier:

$$\dot{\ell}^*_{\theta,m}(\boldsymbol{u};\theta) = \frac{1}{2}\boldsymbol{z}'\,A_m(\theta)\,\boldsymbol{z}, \qquad z_j = \Phi^{-1}(u_j)$$

where $A_m(\theta)$ depends on $R(\theta)$ and its partial derivatives w.r.t. the k components of θ

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

Summary of tangent space geometry

Parametric and nonparametric scores quantify
how the distribution changes if θ and the margins change.

If parametric and nonparametric scores are correlated,
not knowing the margins makes identifying changes in θ harder.

Adaptivity: if parametric and nonparametric scores are uncorrelated,
not knowing the margins does not matter for estimation of θ.

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

Assumption on the correlation matices

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 θ.

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

$\Rightarrow$ The parametric model in θ with known margins in $\mathcal{F}_{\text{ac}}$ is regular.

The one-step estimator is efficient

Theorem

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

$$\tilde{\theta}_n = \theta + O_p(1/\sqrt{n}) \qquad \textit{under every } \mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$$

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

$$\sqrt{n}\,\left(\hat{\theta}_{n,OSE} - \theta\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I^{*-1}(\theta)\,\dot{\ell}^*_\theta(F_1(X_{i1}), \ldots, F_p(X_{ip}); \theta) + o_P(1)$$
$$\xrightarrow{d} N_k(\mathbf{0}, I^*(\theta)^{-1})$$

Moreover, the one-step estimator is an <u>efficient</u> estimator of θ in the semiparametric Gaussian copula model $\mathcal{P}$.

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

For the MLE for θ if margins are known:

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

For the one-step estimator:

$$I^*(\theta)^{-1} \geqslant I(\theta)^{-1}$$

For the pseudo-likelihood estimator:

$$\Sigma_{\mathrm{PLE}}(\theta) \geqslant I^*(\theta)^{-1}$$

Notation: $A \geqslant B$ iff $A - B$ is positive semi-definite

An efficiency criterion for the PLE

Theorem

The PLE is semiparametrically efficient at $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$ *if and only if, for every* $m = 1, \ldots, k$*, the matrix*

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

with

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

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

Adaptivity is the exception rather than the rule

The semiparametric Gaussian copula model is said to be adaptive at $\mathrm{P}_{\theta,F_1,\ldots,F_p} \in \mathcal{P}$ if

$$I^*(\theta) = I(\theta)$$

i.e. knowing the margins or not does not make a difference.

Theorem

A necessary and sufficient condition for adaptivity is

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

Apart from independence, this does not usually seem to occur:
See the next few examples.

The full model: The PLE is efficient

Without restrictions, there are $p(p-1)/2$ parameters, e.g. if $p = 3$,

$$R(\theta) = \begin{pmatrix} 1 & \theta_{12} & \theta_{13} \\ \cdot & 1 & \theta_{23} \\ \cdot & \cdot & 1 \end{pmatrix}$$

- The OSE, PLE and normal scores rank correlations are asymptotically equivalent and semiparametrically efficient.
- Adaptivity only occurs at independence.

[Klaassen & Wellner (1997)]

Exchangeable correlation matrices: The PLE is still efficient

Determined by a single parameter $\theta \in (-1/p, 1)$, for instance if $p = 3$:

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

The PLE is efficient, and the inverse Fisher information for θ is

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

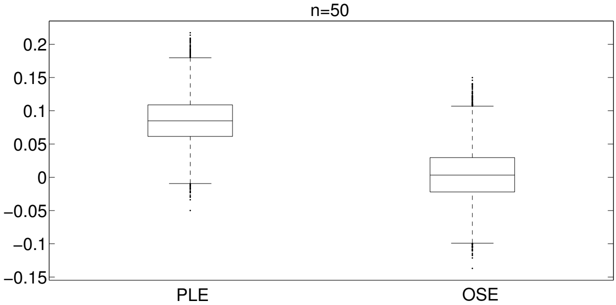
If the margins were known, the minimal asymptotic variance would reduce to

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

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

[HOFF, NIU & WELLNER (2013)]

In high dimensions, the OSE seems less biased



$$R(\theta) = \begin{pmatrix} 1 & \theta & \dots & \dots & \theta \\ \cdot & 1 & \theta & \dots & \theta \\ \cdot & \dots & \dots & \dots & \dots \\ \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \qquad p = 100, n = 50$$

Circulation correlation matrices: The PLE is nearly efficient

The circular model has a single parameter $\theta \in (-1, 1)$:

$$R(\theta) = \begin{pmatrix} 1 & \theta & \theta^2 & \theta \\ \cdot & 1 & \theta & \theta^2 \\ \cdot & \cdot & 1 & \theta \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

The PLE is not efficient, but still nearly.
Adaptivity occurs at independence only.

$$I^{*-1}(\theta) = \frac{1}{4}(1 - \theta^2)^2 \qquad \text{unknown margins, OSE}$$

$$\sigma^2_{\text{PLE}} = I^{*-1}(\theta) \left(1 + \frac{2\theta^6}{(1 + 2\theta^2)^2}\right) \qquad \text{unknown margins, PLE}$$

$$I^{-1}(\theta) = I^{*-1}(\theta)/(1 + 2\theta^2) \qquad \text{known margins, MLE}$$

[Hoff, Niu & Wellner (2013)]

Factor models: the PLE is efficient

Suppose $p \geqslant 3$ and if there are q factors, $1 \leqslant q < p$, then

$$R(\theta) = \theta\theta' - \mathrm{diag}(\theta\theta') + I_p, \qquad \theta \in \mathbb{R}^{p \times q}$$

Identifiability issue: resolve by reparametrization $\nu \mapsto \theta(\nu)$.

The efficiency criterion can be shown to be fulfilled
$\Rightarrow$ the PLE is efficient.

Toeplitz models: The PLE can be quite inefficient

The Toeplitz model has $p - 1$ parameters, e.g. in $p = 4$:

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

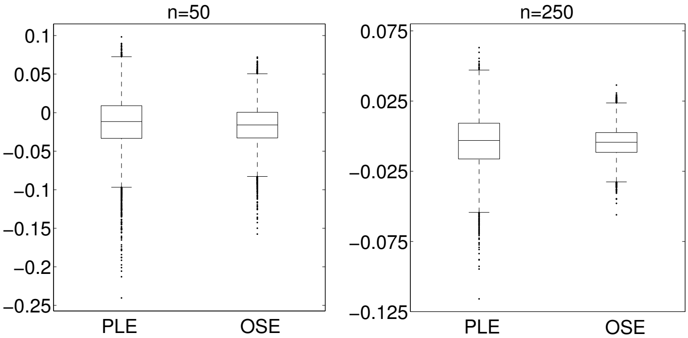
$p = 3$ The PLE is still semiparametrically efficient.

$p = 4$ Not anymore! See next plots.

The OSE may do much better than the PLE

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

Monte Carlo, 15 000 samples of size $n = 50$ and $n = 250$



$$R(\theta) = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \theta_3 \\ \cdot & 1 & \theta_1 & \theta_2 \\ \cdot & \cdot & 1 & \theta_1 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix} \qquad \theta = (0.4945, -0.4593, -0.8462)$$

Contribution: Efficient inference for Gaussian copulas

- Inference in semiparametric Gaussian copula models
  - structured correlation matrices
  - unknown, continuous margins
- One-step estimator
  - rank-based
  - semiparametrically efficient
  - outperforms the pseudo-likelihood estimator
- Adaptivity usually occurs only at independence

`http://arxiv.org/abs/1306.6658`

Next: Efficient, rank-based inference in general semiparametric copula models

- Efficient score function and information matrix?
- One-step estimator?
- Efficient estimation of the margins as well?
- Time series?
- Discrete margins? E.g. multivariate probit models
- …

Thank you!