

Copula goodness-of-fit testing: An overview and power comparison

DANIEL BERG

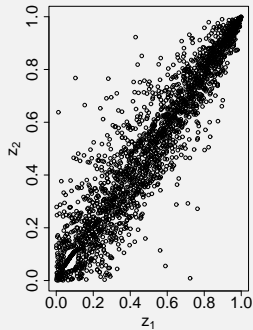
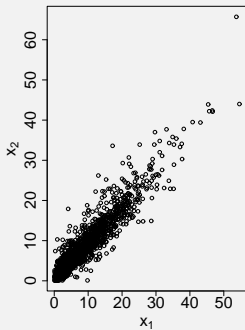
University of Oslo & Norwegian Computing Center

Conference on copulae and multivariate probability distributions in finance –
theory, applications, opportunities and problems.

Warwick Business School – September 2007

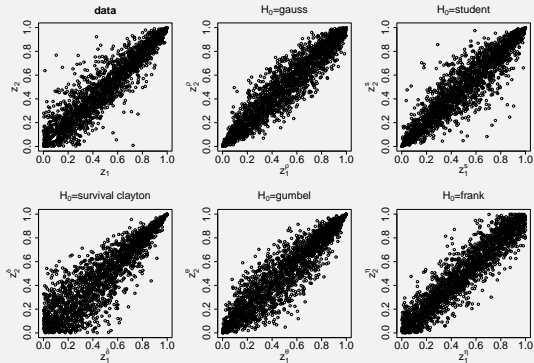
Motivation

- ▷ Does our model fit the data?



Motivation

▷ Does our model fit the data?



Outline

- ▶ 1. Copula GoF testing
 - 1.1. Introduction
 - ▶ 1.2. Preliminaries
 - 1.3. Proposed approaches
- ▶ 2. MC simulation results
 - 2.1. Test procedure
 - 2.2. Experimental setup
 - 2.3. Testing the Gaussian hypothesis
 - 2.4. Testing the Student-t hypothesis
 - 2.5. Testing the Clayton hypothesis
 - 2.6. Testing the Gumbel hypothesis
 - 2.7. Conclusions and recommendations
- ▶ 3. Discussion

1.1. Copula GoF testing – Introduction

- ▶ $\mathcal{H}_0 : C \in \mathcal{C} = \{C_\theta; \theta \in \Theta\}$ vs. $\mathcal{H}_1 : C \notin \mathcal{C} = \{C_\theta; \theta \in \Theta\}$
- ▶ Univariate \Rightarrow Anderson-Darling or QQ-plot,
 Multivariate \Rightarrow fewer alternatives
- ▶ Pseudo-observations (normalized ranks):

$$z_j = (z_{j1}, \dots, z_{jd}) = \left(\frac{R_{j1}}{n+1}, \dots, \frac{R_{jd}}{n+1} \right), j = 1, \dots, n$$
- ▶ Limiting distribution of copula GoF test depends on parameter value
- ▶ p -value estimation via parametric bootstrap procedures
- ▶ Focus in literature almost exclusively bivariate
- ▶ NOT model selection!

1.1. Copula GoF testing – Introduction contd.

- ▷ Several techniques proposed:
 - Discretization of unit hypercube (binning)
 - Multivariate smoothing
 - **'Dimension reduction'**
 - Approaches for the testing of one specific copula family
 - **General approaches for the testing of any copula family**

1.2. Copula GoF testing – Preliminaries

Rosenblatt's transform:

- ▷ Dependent variables \Rightarrow independent $U[0, 1]$ variables, given multivariate distribution
- ▷ $\mathbf{v} = \mathcal{R}(\mathbf{z}) = (\mathcal{R}_1(z_1), \dots, \mathcal{R}_d(z_d))$:

$$v_1 = \mathcal{R}_1(z_1) = F_1(z_1) = z_1,$$

$$v_2 = \mathcal{R}_2(z_2) = F_{2|1}(z_2|z_1),$$

$$\vdots$$

$$v_d = \mathcal{R}_d(z_d) = F_{d|1\dots d}(z_d|z_1, \dots, z_d).$$

- ▷ Inverse of simulation (conditional inversion)
- ▷ GoF: $\mathbf{v} = \mathcal{R}(\mathbf{z}) \Rightarrow$ test \mathbf{v} for independence
- ▷ $d!$ different permutation orders

1.3.1 Proposed approaches – approach \mathcal{A}_1 (1/9)

- ▷ $\mathbf{v} = \mathcal{R}(z)$, $\mathbf{h} = \mathcal{R}(\mathbf{v})$
- ▷ $W_1 = \sum_{i=1}^d \Gamma_V\{V_{(i)}; \alpha\} \cdot \Gamma_H\{H_i; \alpha\}$
- ▷ Special case (i): $\sum_{i=1}^d \Phi^{-1}(V_i)^2$
- ▷ Special case (ii): $\sum_{i=1}^d |V_i - 0.5|$
- ▷ $S_1(t) = P\{F_1(W_1) \leq t\}$, $t \in [0, 1]$
- ▷ CvM statistic:

$$\hat{T}_1 = n \int_0^1 \left\{ \hat{S}_1(t) - S_1(t) \right\}^2 dS_1(t)$$

1.3.2. Proposed approaches – approach \mathcal{A}_2 (2/9)

- ▷ Empirical copula:

$$\widehat{C}(\mathbf{u}) = \frac{1}{n+1} \sum_{j=1}^n I\{Z_{j1} \leq u_1, \dots, Z_{jd} \leq u_d\}$$

- ▷ CvM statistic:

$$\widehat{T}_2 = n \int_{[0,1]^d} \left\{ \widehat{C}(z) - C_{\widehat{\theta}}(z) \right\}^2 d\widehat{C}(z)$$

1.3.3. Proposed approaches – approach \mathcal{A}_3 (3/9)

- ▷ Approach \mathcal{A}_2 on $\mathbf{v} = \mathcal{R}(\mathbf{z})$
- ▷ CvM statistic:

$$\hat{T}_3 = n \int_{[0,1]^d} \left\{ \hat{C}(\mathbf{v}) - C_{\perp}(\mathbf{v}) \right\}^2 d\hat{C}(\mathbf{v})$$

1.3.4. Proposed approaches – approach \mathcal{A}_4 (4/9)

- ▷ Cdf of empirical copula (Kendall's dependence function):

$$S_4(t) = P\{C(z) \leq t\}$$

- ▷ CvM statistic:

$$\widehat{T}_4 = n \int_0^1 \left\{ \widehat{S}_4(t) - S_{4, \widehat{\theta}}(t) \right\}^2 dS_{4, \widehat{\theta}}(t)$$

1.3.5. Proposed approaches – approach \mathcal{A}_5 (5/9)

- ▷ Spearman's dependence function:

$$S_5(t) = P\{C_{\perp}(z) \leq t\}$$

- ▷ CvM statistic:

$$\widehat{T}_5 = n \int_0^1 \left\{ \widehat{S}_5(t) - S_{5, \widehat{\theta}}(t) \right\}^2 dS_{5, \widehat{\theta}}(t)$$

1.3.6. Proposed approaches – approach \mathcal{A}_6 (6/9)

- ▶ Shih's test for bivariate gamma frailty model (Clayton):

$$\hat{T}_{Shih} = \sqrt{n} \left\{ \hat{\theta}_\tau - \hat{\theta}_W \right\}$$

- ▶ Extension to arbitrary dimension:

$$\hat{T}_6 = \sum_{i=1}^{d-1} \sum_{j=i+1}^d \left\{ \hat{\theta}_{\tau,ij} - \hat{\theta}_{W,ij} \right\}^2$$

1.3.7. Proposed approaches – approach \mathcal{A}_7 (7/9)

- ▷ Inner product of two vectors = 0 iff from the same family

$$Q(z) = \langle z - z_{\hat{\theta}} | \kappa_d | z - z_{\hat{\theta}} \rangle$$

- ▷ κ a symmetric kernel, e.g. the gaussian kernel:

$$\kappa_d(z, z_{\hat{\theta}}) = \exp \left\{ -\|z - z_{\hat{\theta}}\|^2 / (2dh^2) \right\}$$

- ▷ Statistic becomes:

$$\begin{aligned} \hat{T}_7 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \kappa_d(z_i, z_j) - \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \kappa_d(z_i, z_{\hat{\theta},j}) \\ &+ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \kappa_d(z_{\hat{\theta},i}, z_{\hat{\theta},j}) \end{aligned}$$

1.3.8. Proposed approaches – approach \mathcal{A}_8 (8/9)

- ▷ Approach \mathcal{A}_7 on $\mathbf{v} = \mathcal{R}(\mathbf{z})$
- ▷ Statistic:

$$\begin{aligned} \widehat{T}_8 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \kappa_d(\mathbf{v}_i, \mathbf{v}_j) - \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \kappa_d(\mathbf{v}_i, \mathbf{v}_{\widehat{\theta}_j}) \\ &+ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \kappa_d(\mathbf{v}_{\widehat{\theta}_i}, \mathbf{v}_{\widehat{\theta}_j}) \end{aligned}$$

1.3.9. Proposed approaches – approach \mathcal{A}_9 (9/9)

- ▶ Each approach may detect deviations from \mathcal{H}_0 differently
- ▶ Average of all approaches:

$$\widehat{T}_9 = \frac{1}{9} \left\{ \widehat{T}_1^{(i)} + \widehat{T}_1^{(ii)} + \sum_{t=2}^8 \widehat{T}_t \right\}$$

1.3.10. Proposed approaches – summary

- ▶ \mathcal{A}_1 : Rosenblatt's transform; special cases – Breyermann, Dias and Embrechts (2003) and Malevergne and Sornette (2003).
- ▶ \mathcal{A}_2 : Copula distribution function – empirical vs. null
- ▶ \mathcal{A}_3 : Rosenblatt's transform and \mathcal{A}_2 – empirical vs. independence
- ▶ \mathcal{A}_4 : Cdf of copula distribution function – empirical vs. null
- ▶ \mathcal{A}_5 : Spearman's GoF process
- ▶ \mathcal{A}_6 : Generalized Shih test – for testing the Clayton copula
- ▶ \mathcal{A}_7 : Panchenko's approach – inner product with gaussian kernel
- ▶ \mathcal{A}_8 : Rosenblatt's transform and \mathcal{A}_7
- ▶ \mathcal{A}_9 : An average approach – average of \mathcal{A}_1 – \mathcal{A}_8

2.1. MC simulations – Test procedure

- 1) $\mathbf{x} \sim n$ samples from the d -dimensional \mathcal{H}_1 copula with $\theta(\tau)$.
- 2) $\mathbf{z} \sim$ pseudo-observations (normalized ranks)
- 3) $\hat{\theta} \sim$ estimated parameter of the \mathcal{H}_0 copula
- 4) $\hat{T}_i \sim$ test statistic i computed under the \mathcal{H}_0 copula using $\hat{\theta}$.
- 5) Repeat steps 1-4 M times with $\mathcal{H}_1 = \mathcal{H}_0$ and $\theta = \hat{\theta} \Rightarrow \hat{T}_{i,m}^0$
- 6) $\hat{p} = \frac{1}{M} \sum_{m=1}^M \mathbf{1}(\hat{T}_{i,m}^0 > \hat{T}_i)$
- 7) $\hat{p} < 5\% \Rightarrow$ reject \mathcal{H}_0

2.2. MC simulations – Experimental setup

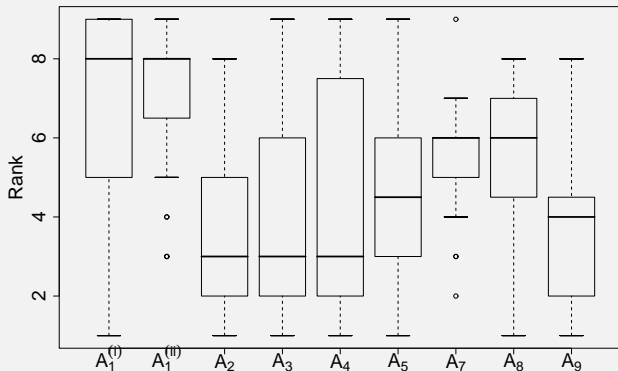
- ▶ \mathcal{H}_0 copula (4 choices: Gaussian, Student-t, Clayton, Gumbel),
- ▶ \mathcal{H}_1 copula (4 choices: Gaussian, Student-t ($\nu = 6$), Clayton, Gumbel,
- ▶ Kendall's tau (2 choices: $\tau = \{0.2, 0.4\}$),
- ▶ Dimension (3 choices: $d = \{2, 4, 8\}$),
- ▶ Sample size (2 choices: $n = \{100, 500\}$)
- ▶ For each of these $4 \times 4 \times 2 \times 3 \times 2 = 192$ cases \Rightarrow 10,000 repetitions \Rightarrow size/power

2.3. MC simulations – Testing the Gaussian copula

Table: Percentage rejections (5% level) of the Gaussian copula ($\tau = 0.2$).

d	n	\mathcal{H}_1	$\mathcal{A}_1^{(T)}$	$\mathcal{A}_1^{(H)}$	\mathcal{A}_2	\mathcal{A}_3	\mathcal{A}_4	\mathcal{A}_5	\mathcal{A}_6	\mathcal{A}_7	\mathcal{A}_8	\mathcal{A}_9	
2	100	Ga	4.8	5.5	5.0	4.7	5.2	5.4	–	4.8	4.8	5.1	
		St	1.1	4.6	7.9	8.7	6.4	5.9	–	5.6	6.0	3.6	
		Cl	2.9	5.1	22.1	21.1	21.0	16.3	–	7.7	7.1	11.3	
		Gu	2.1	4.9	11.9	3.4	12.4	9.0	–	6.0	5.9	5.2	
	500	Ga	4.5	4.7	5.0	4.6	4.6	5.3	–	4.4	5.0	4.9	
		St	18.2	16.2	10.2	17.5	8.0	8.4	–	9.3	9.8	20.2	
		Cl	2.1	5.3	73.0	72.2	71.5	56.9	–	21.4	20.4	55.2	
		Gu	2.5	6.1	32.9	8.9	33.0	25.7	–	11.8	12.2	21.1	
	4	100	Ga	5.2	5.1	5.2	4.8	4.9	5.1	–	4.8	4.9	4.9
			St	5.1	6.3	9.3	16.0	7.9	6.5	–	6.9	6.0	7.8
			Cl	1.4	5.0	46.4	31.5	52.5	18.8	–	9.4	7.2	20.9
			Gu	1.3	3.4	12.7	0.9	41.7	56.3	–	13.6	8.2	13.2
500		Ga	4.8	5.9	5.8	5.8	6.2	5.4	–	5.2	4.8	5.9	
		St	98.4	74.4	18.1	49.9	12.8	13.3	–	15.0	13.1	96.8	
		Cl	4.1	8.9	99.0	94.9	98.1	88.8	–	38.7	20.0	94.8	
		Gu	6.5	6.8	85.2	50.2	98.4	98.5	–	69.8	32.1	93.8	

2.3. MC simulations – Testing the Gaussian copula

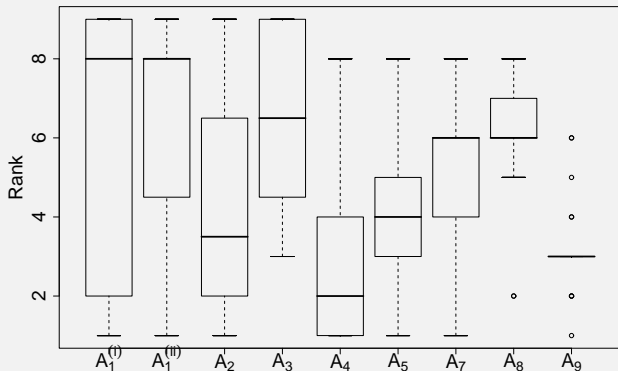


2.4. MC simulations – Testing the Student copula

Table: Percentage rejections (5% level) of the Student copula ($\tau = 0.2$).

d	n	\mathcal{H}_1	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	\mathcal{A}_2	\mathcal{A}_3	\mathcal{A}_4	\mathcal{A}_5	\mathcal{A}_6	\mathcal{A}_7	\mathcal{A}_8	\mathcal{A}_9	
2	100	Gaussian	5.6	5.7	4.5	3.9	5.0	5.3	–	5.3	4.7	5.4	
		Student	4.4	4.9	4.7	3.9	4.9	5.2	–	5.1	4.9	4.7	
		Clayton	5.3	5.2	19.0	10.5	20.4	17.6	–	7.5	7.1	16.0	
		Gumbel	4.7	5.2	9.7	5.1	10.5	7.3	–	6.2	6.2	7.8	
	500	Gaussian	5.7	5.6	5.3	4.9	5.4	5.2	–	5.4	5.2	5.7	
		Student	4.9	4.9	4.7	4.7	5.0	5.4	–	5.0	5.4	4.9	
		Clayton	6.0	5.2	70.3	60.4	72.6	61.4	–	21.9	19.6	66	
		Gumbel	5.1	5.3	28.0	18.2	29.0	18.6	–	10.4	9.6	22.2	
	4	100	Gaussian	6.3	6.0	4.5	4.5	5.3	5.1	–	5.3	5.2	5.9
			Student	4.3	4.7	4.5	4.5	5.0	5.3	–	4.8	5.0	4.2
			Clayton	5.5	5.1	45.0	24.5	49.9	21.1	–	9.5	6.8	36.8
			Gumbel	5.5	5.2	7.2	0.7	39.6	56.0	–	13.2	7.6	28.3
500		Gaussian	6.6	5.8	5.1	4.8	5.1	5.3	–	5.4	5.3	6.1	
		Student	4.9	5.0	4.9	4.5	5.2	5.6	–	5.1	5.1	5.3	
		Clayton	9.0	5.5	99.1	91.3	97.7	90.2	–	35.9	21.0	97.5	
		Gumbel	8.0	6.7	77.1	52.6	97.6	98.0	–	65.8	31.0	96.7	

2.4. MC simulations – Testing the Student copula

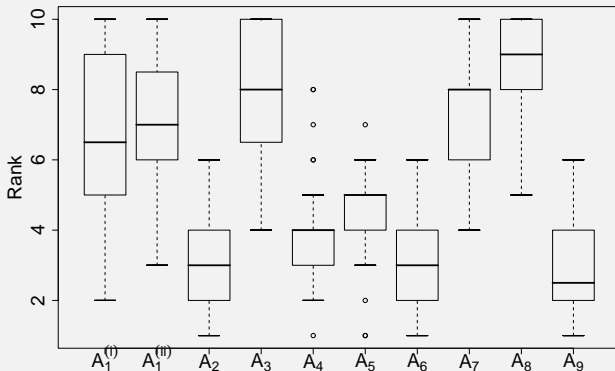


2.5. MC simulations – Testing the Clayton copula

Table: Percentage rejections (5% level) of the Clayton copula ($\tau = 0.2$).

d	n	\mathcal{H}_1	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	\mathcal{A}_2	\mathcal{A}_3	\mathcal{A}_4	\mathcal{A}_5	\mathcal{A}_6	\mathcal{A}_7	\mathcal{A}_8	\mathcal{A}_9	
2	100	Ga	7.1	7.6	21.3	6.9	22.3	14.0	20.7	7.7	6.9	20.6	
		St	8.0	8.6	24.0	8.3	23.6	16.1	17.0	8.2	7.0	21.3	
		Cl	5.3	5.0	5.3	5.1	4.7	5.0	5.3	5.2	4.7	4.9	
		Gu	6.3	8.8	45.8	12.5	45.6	31.1	41.0	12.9	10.2	40.5	
	500	Ga	18.9	13.0	77.3	46.2	69.6	54.1	85.1	23.6	20.7	68.6	
		St	25.5	22.8	80.9	35.6	73.8	64.1	68.9	24.9	22.2	76.2	
		Cl	4.7	5.2	4.8	5.4	4.8	5.9	4.8	5.2	5.0	5.4	
		Gu	12.9	23.4	99.2	85.9	98.1	94.3	99.1	58.6	52.6	97.4	
	4	100	Ga	11.0	10.5	37.7	2.9	39.7	39.3	50.4	9.8	7.2	48.3
			St	26.6	20.5	48.3	17.2	37.8	41.1	37.5	10.5	8	55.9
			Cl	4.9	4.4	4.7	4.8	5.2	4.8	4.2	5.4	5.1	4.0
			Gu	9.0	11.5	62.9	2.5	90.5	93.6	81.2	30.8	14.9	87.6
500		Ga	89.8	37.7	99.5	17.3	96.5	91.1	99.9	39.1	23.6	99.4	
		St	93.5	75.7	99.9	88.8	95.1	94.0	97.6	42.6	29.7	99.9	
		Cl	5.3	4.5	5.2	5.1	5.2	5.6	4.8	5.2	5.1	4.8	
		Gu	71.8	37	100	79.6	100	100	100	97.7	82.9	100	

2.5. MC simulations – Testing the Clayton copula

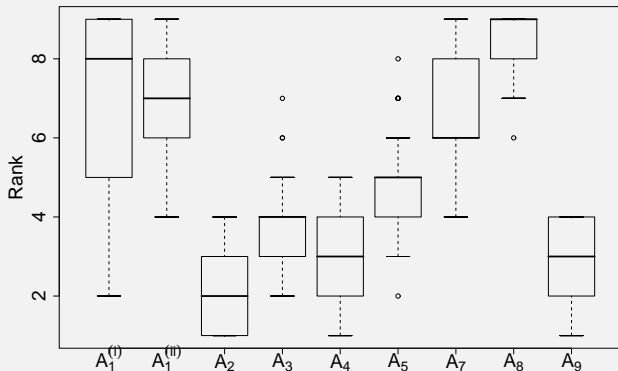


2.6. MC simulations – Testing the Gumbel copula

Table: Percentage rejections (5% level) of the Gumbel copula ($\tau = 0.2$).

d	n	\mathcal{H}_1	$\mathcal{A}_1^{(I)}$	$\mathcal{A}_1^{(II)}$	\mathcal{A}_2	\mathcal{A}_3	\mathcal{A}_4	\mathcal{A}_5	\mathcal{A}_6	\mathcal{A}_7	\mathcal{A}_8	\mathcal{A}_9
2	100	Ga	7.8	6.6	9.2	6.9	9.3	9.6	–	6.3	6.1	9.8
		St	6.6	6.8	11.1	9.9	8.9	7.0	–	5.9	5.7	9.3
		Cl	5.8	6.6	45.0	29.9	43.5	34.4	–	12.1	10.3	33.0
		Gu	5.2	5.2	5.2	5.3	5.0	5.0	–	5.1	5.2	5.2
	500	Ga	20.9	10.2	37.5	23.2	31.1	26.9	–	12.8	11.4	33.7
		St	17.2	11.9	40.1	33.1	26.8	17.1	–	11.4	10.6	30.2
		Cl	8.9	10.6	99.5	98.0	98.4	95.5	–	58.0	52.1	97.1
		Gu	5.1	4.5	5.1	5.0	5.4	4.8	–	5.3	5.1	4.9
4	100	Ga	7.1	13.4	54.1	44.3	52.6	25.1	–	16.1	6.7	42.2
		St	25.8	25.1	56.2	57.0	53.2	21.7	–	15.2	7.5	59.0
		Cl	3.2	15.5	89.4	85.1	97.1	82.3	–	31.7	9.5	90.4
		Gu	5.1	5.0	5.3	5.3	5.2	5.1	–	5.5	5.0	4.5
	500	Ga	76.0	58.7	99.5	98.5	98.4	95.9	–	69.4	21.0	99.4
		St	92.1	87.9	99.1	99.8	97.8	94.4	–	67.3	29.6	100
		Cl	34.6	65.1	100	100	100	100	–	98.1	55.3	100
		Gu	5.0	4.6	4.7	5.2	5.4	5.4	–	4.3	5.5	4.9

2.6. MC simulations – Testing the Gumbel copula



2.7. MC simulations – Conclusions and recommendations

- ▷ Nominal levels are well matched
- ▷ $\mathcal{A}_2 : \sqrt{n}\{\widehat{C} - C_{\hat{\theta}}\}$ – very good almost always
- ▷ \mathcal{A}_2 – not so good to detect heavy tails under Gaussian null
- ▷ \mathcal{A}_1 – good to detect heavy tails under Gaussian null, otherwise poor
- ▷ \mathcal{A}_6 – very good for testing Clayton null
- ▷ \mathcal{A}_9 – recommended, in particular if we don't know what kind of deviation we are trying to detect

3. Discussion

- ▶ 9 approaches considered, including a proposed generalization of Shih's test and two new approaches
- ▶ Extensive MC simulations – size/power
- ▶ Easier to detect deviations in higher dimensions and with increasing sample size, as expected
- ▶ No 'best' approach in general although \mathcal{A}_2 comes close
- ▶ An average approach such as \mathcal{A}_9 recommended
- ▶ Disadvantage of $\mathcal{A}_2, \mathcal{A}_4, \mathcal{A}_5$: computationally intensive when double bootstrap is called for
- ▶ Advantage of approaches based on Rosenblatt's transform: computationally efficient
- ▶ What about constructions of multivariate dependence?

Contact details

Daniel Berg
Norwegian Computing Center & University of Oslo
E-mail: [daniel at danielberg.no](mailto:daniel@anielberg.no)
Web: www.danielberg.no