Goodness-of-Fit Test for Specification of Semiparametric Copula Dependence Models

Shulin Zhang, Ostap Okhrin

Qian M. Zhou, Peter X.-K. Song

Southwestern University of Finance and Economics

Dresden University of Technology

Simon Fraser University

University of Michigan

http://csr.swufe.edu.cn

http://osv.vkw.tu-dresden.de

http://stat.sfu.ca/

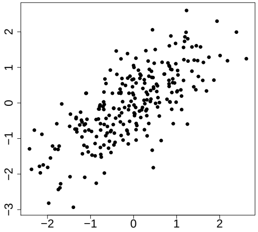
http://www.sph.umich.edu/

Applications

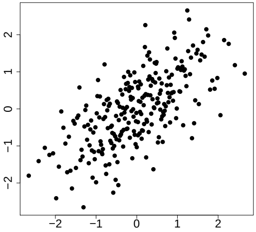
1. medicine (Vandenhende (2003), ...)
2. hydrology (Genest and Favre (2006), ...)
3. biometrics (Wang and Wells (2000, JASA), Chen and Fan (2006, CanJoS), ...)
4. economics
    - portfolio selection (Patton (2004, JoFE), Hennessy and Lapan (2002, MathFin), ...)
    - time series (Chen and Fan (2006a, 2006b, JoE), Fermanian and Scaillet (2003, JoR), Lee and Long (2005, JoE), ...)
    - risk management (Junker and May (2002, EJ), Breyman et. al. (2003, QF), ...)
5. ...

How to be sure, that one uses a proper copula?

Different tests ⇒ Different outcomes



(a) Gaussian copula



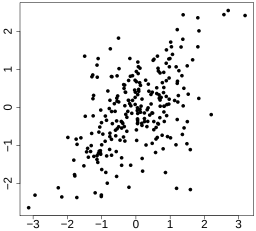
(b) year 2004

Figure 1: Sample from Gauss copula with $\mathrm{N}(0,1)$ margins, $\theta = 0.71$, $N = 250$ and residuals transformed to standard normal for Citygroup/BoA for 2004.

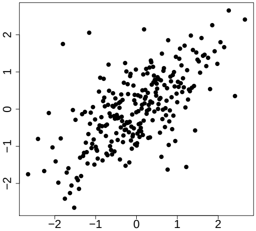
Visually - Gaussian copula

Test 1: Gumbel, Test 2: Gauss, Test 3: Gauss

Different tests ⇒ Different outcomes



(a) t-copula



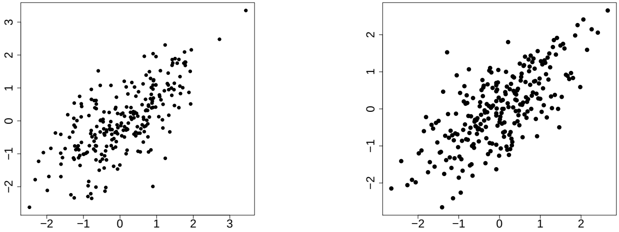
(b) year 2006

Figure 2: Sample from t-copula with $\mathrm{N}(0,1)$ margins, $\theta = 0.6$, $N = 250$ and residuals transformed to standard normal for Citygroup/BoA for 2006.

Visually - t-copula

Test 1: t-copula, Test 2: Gauss, Test 3: t-copula

Different tests ⇒ Different outcomes



(a) Gumbel copula (b) year 2009

Figure 3: Sample from Gumbel copula with N(0, 1) margins, $\theta = 2$, $N = 250$ and residuals transformed to standard normal for Citygroup/BoA for 2009.

Visually - Gumbel copula

Test 1: Gumbel, Test 2: Gumbel, Test 3: Gauss

Outline

1. Motivation ✓
2. Pseudo in-and-out-of-sample (PIOS) Test
3. Hybrid Test
4. Asymptotic Properties of PIOS Test
5. Extension of PIOS Test
6. Applications
7. Conclusion

PIOS test, I

$$\mathcal{H}_0 : C_0 \in \mathcal{C} \quad vs. \quad \mathcal{H}_1 : C_0 \notin \mathcal{C}$$

where $\mathcal{C} = \{C(\cdot;\theta) : \theta \in \Theta\}$.

- $X_1 = (X_{11}, \ldots, X_{1d})^\top, \ldots, X_n = (X_{n1}, \ldots, X_{nd})^\top$ random sample of size n drawn from multivariate distribution $H(x) = H(x_1, x_2, \ldots, x_d)$
- Continuous marginal cdf $F(x) = \{F_1(x_1), \ldots, F_d(x_d)\}$

$$H(x_1, x_2, \ldots, x_d) = C_0\{F(x)\} = C_0\{F_1(x_1), \ldots, F_d(x_d)\}.$$

PIOS test, II

Define $\ell\{\tilde{F}(X_i);\theta\} = \log c\{\tilde{F}_1(X_{i1}),\dots,\tilde{F}_d(X_{id});\theta\}$ and $\hat{\theta}$ be the two-step pseudo maximum likelihood method (PMLE) of θ given by

$$\hat{\theta} = \operatorname*{argmax}_{\theta\in\Theta} \sum_{i=1}^{n} \ell\{\tilde{F}(X_i);\theta\}.$$

Compute delete-one-block PLMEs $\hat{\theta}_{-b}$, $1 \le b \le B$:

$$\hat{\theta}_{-b} = \operatorname*{argmax}_{\theta\in\Theta} \sum_{b'\neq b}^{B} \sum_{i=1}^{m} \ell\{\tilde{F}(X_i^{b'});\theta\}, \quad b = 1,\dots,B,$$

where

$$\tilde{F}_k(x_k) = \frac{1}{n+1}\sum_{t=1}^{n} \mathbf{I}\left(X_{tk} \le x_k\right), \quad k = 1,\dots,d.$$

PIOS test, III

Comparing "in-sample" and "out-of-sample" pseudo-likelihoods with the following test statistic:

$$T_n(m) = \sum_{b=1}^{B} \sum_{i=1}^{m} \left[\ell\{\tilde{F}(X_i^b); \hat{\theta}\} - \ell\{\tilde{F}(X_i^b); \hat{\theta}_{-b}\} \right].$$

Challenge: needed $[\frac{n}{m}]$ *dependence parameters*

Solution: test statistic which is asymptotically equivalent.

PIOS test, IV

- Under suitable regularity conditions and under assumption, that $\exists\ \theta^* \in \Theta$ with $\hat{\theta} \xrightarrow{p} \theta^*$ for $n \to \infty$:

$$T_n(m) \xrightarrow{p} \operatorname{tr}\{S(\theta^*)^{-1} V(\theta^*)\}$$

with

$$S(\theta) = -\operatorname{E}_0 \left[\frac{\partial^2}{\partial\theta\partial\theta^\top} \ell\{F(X_1);\theta\} \right],$$

$$V(\theta) = \operatorname{E}_0 \left[\frac{\partial}{\partial\theta} \ell\{F(X_1);\theta\} \frac{\partial}{\partial\theta} \ell^\top\{F(X_1);\theta\} \right].$$

PIOS test, V

- Under a correct model specification, it holds: $V(\theta^*) = S(\theta^*)$.
- Then is $\mathrm{tr}\{S(\theta^*)^{-1}V(\theta^*)\} = p$.
- Asymptotic test statistic:

$$R_n = \mathrm{tr}\left\{\hat{S}(\hat{\theta})^{-1}\hat{V}(\hat{\theta})\right\}$$

where $\hat{S}(\hat{\theta})$ and $\hat{V}(\hat{\theta})$ are the empirical counterparts to $S(\theta)$ and $V(\theta)$.

Law of Large Numbers

Theorem

Under assumptions A1 and A2 hold

$$R_n \xrightarrow{p} tr\left\{S(\theta^*)^{-1}V(\theta^*)\right\}, \quad as \quad n \to \infty,$$

where θ^* is the limiting value of PMLE $\hat{\theta}$.

▸ Assumptions

Central Limit Theorem

Theorem

- *Under the null hypothesis, if A2 and B1 - B3 hold, then*

$$\sqrt{n}\,(R_n - p) \xrightarrow{d} N(0, \sigma_R^2), \quad \text{as } n \to \infty,$$

  where σ_R^2 is the asymptotic variance.

- *Under assumptions A2, B1 - B3 and C1,*

$$R_n - T_n(m) = o_p(n^{-1/2}).$$

▸ Assumption $A2$ ▸ Assumptions $B1 - B3$ ▸ Assumption $C1$

Simulation Study - Benchmark tests, I

- S_n from Genest, Rémillard and Beaudoin (2009, IME)
  - Cramér-von Mises statistic
$$S_n = n \int_{[0,1]^d} \{D_n(u) - C_\perp(u)\}^2 \, du.$$
  - Based on Rosenblatt's transform, with E_{tk} as pseudo observations:
$$E_{tk} = \frac{\partial^{k-1} C(U_{t,1}, \ldots, U_{t,k}, 1, \ldots, 1)/\partial U_{t,1} \cdots \partial U_{t,k-1}}{\partial^{k-1} C(U_{t,1}, \ldots, U_{t,k-1}, 1, \ldots, 1)/\partial U_{t,1} \cdots \partial U_{t,k-1}}, k = 1, 2, \ldots, d,$$
    with $D_n(u) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{I}(E_t \leq u)$ and $C_\perp(u) = u_1 \times u_2 \times \cdots \times u_d$.
  - test has on average one of the best performances among all the existing "blanket tests", see Genest el al. (2009).

Simulation Study - Benchmark tests, II

- J_n from Scaillet (2007, JoMA)
  - Kernel-based GoF test statistic with fixed smoothing parameter

$$J_n = \int_{[0,1]^d} \{\hat{c}(u) - K_H * c(u;\hat{\theta})\} w(u) du,$$

  - The copula density is estimated as

$$\hat{c}(u) = \frac{1}{n} \sum_{t=1}^{n} K_H[u - \{\tilde{F}_1(X_{t1}), \dots, \tilde{F}_d(X_{td})\}^\top].$$

Residual-based Bootstrap

Step 1. Generate bootstrap sample $\left\{\epsilon_t^{(k)}, t = 1, \ldots, n\right\}$ from copula $C(u; \hat{\theta})$ under H_0 with PMLE $\hat{\theta}$ and estimated marginal distribution $\check{F}$ obtained from original data;

Step 2. Based on $\left\{\epsilon_t^{(k)}, t = 1, \ldots, n\right\}$ from Step 1, estimate θ of the copula under H_0 by the two-step PMLE method, and compute R_n, denoted by R_n^k;

Step 3. Repeat Steps 1 - 2 N times and obtain N statistics $R_n^k, k = 1, \ldots, N$;

Step 4. Compute empirical p-value as $p_e = \frac{1}{N}\sum_{k=1}^{N} \mathbf{I}\left(|R_n^k| \geq |R_n|\right)$.

Simulation Study - Fixed true model setup

- Tests used in the study:
  - S_n
  - J_n
  - R_n
  - $T_n(1)$ and $T_n(3)$
- Copulae: Gaussian, t, Clayton and Gumbel
- $\tau \in \{0.25;\ 0.50;\ 0.75\}$
- $n \in \{100;\ 300\}$
- Rounds of simulation $N = 1000$
- Bootstrap sample paths in every simulation $M = 1000$

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	5.5	4.3	4.4	4.5	4.2
	t	t	4.3	5.1	5.5	4.6	6.2
	Cl.	Cl.	5.0	5.9	6.6	6.5	5.0
	Gu.	Gu.	4.5	3.3	5.2	5.2	5.2
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 1: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	**5.5**	**4.3**	**4.4**	**4.5**	**4.2**
	t	t	**4.3**	**5.1**	**5.5**	**4.6**	**6.2**
	Cl.	Cl.	**5.0**	**5.9**	**6.6**	**6.5**	**5.0**
	Gu.	Gu.	**4.5**	**3.3**	**5.2**	**5.2**	**5.2**
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 2: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
	Ga.	Ga.	**5.5**	**4.3**	**4.4**	**4.5**	**4.2**
	t	t	**4.3**	**5.1**	**5.5**	**4.6**	**6.2**
	Cl.	Cl.	**5.0**	**5.9**	**6.6**	**6.5**	**5.0**
	Gu.	Gu.	**4.5**	**3.3**	**5.2**	**5.2**	**5.2**
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
$n = 300$	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 3: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Simulation Study - Results

	True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$
$n = 300$	Ga.	Ga.	**5.5**	**4.3**	**4.4**	**4.5**	**4.2**
	t	t	**4.3**	**5.1**	**5.5**	**4.6**	**6.2**
	Cl.	Cl.	**5.0**	**5.9**	**6.6**	**6.5**	**5.0**
	Gu.	Gu.	**4.5**	**3.3**	**5.2**	**5.2**	**5.2**
	Ga.	t	5.1	12.4	66.0	61.7	22.4
	Ga.	Cl.	99.1	100.0	77.7	78.8	62.5
	Ga.	Gu.	60.2	36.3	7.3	6.9	6.3
	t	Ga.	65.7	12.3	95.6	96.3	88.1
	t	Cl.	98.3	100.0	98.0	98.0	86.5
	t	Gu.	88.3	24.7	71.4	72.6	52.7
	Cl.	Ga.	100.0	100.0	100.00	99.8	97.2
	Cl.	t	100.0	98.5	36.6	97.7	75.9
	Cl.	Gu.	100.0	100.0	100.0	100.0	100.0
	Gu.	Ga.	26.1	30.9	87.8	84.1	69.4
	Gu.	t	47.0	25.6	5.5	4.3	5.9
	Gu.	Cl.	100.0	100.0	100.0	100.0	97.5

Table 4: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Hybrid Test, I

- Different tests + different situations = Different power
- Hybrid test combines several test methods
- Consider q test statistics $T_n^{(1)}, T_n^{(2)}, \dots, T_n^{(q)}$
- Common H_0 hypothesis and given significance level α
- Hybrid test statistic, T_n^{hybrid}, will have p-value

$$p_n^{\text{hybrid}} = \min\{q \times \min\{p_n^{(1)}, \dots, p_n^{(q)}\}, 1\}$$

- Rejection rule: $p_n^{\text{hybrid}} \leq \alpha$

Hybrid Test, II

- Type I error:

$$P(p_n^{(hybrid)} \leq \alpha | H_0) \leq \alpha$$

- Type II error:

$$P(p_n^{hybrid} \leq \alpha | H_1) \geq \max \left\{ \beta_n^1(\alpha/q), \ldots, \beta_n^q(\alpha/q) \right\}$$

- Implication: If at least one test is consistent, hybrid test is consistent as well
- Simulation study shows that the Hybrid Test behaves more desirably than the individual tests

Simulation Study - cont.

- Bootstrap technique to numerically establish the null distribution of the test statistics

- Applied single tests:
  - S_n
  - J_n
  - R_n
  - $T_n(1)$ and $T_n(3)$

- Applied hybrid tests:
  - SR_n
  - $ST_n(1)$
  - JR_n
  - $JT_n(1)$
  - SJR_n
  - $SJT_n(1)$

Simulation Study - Results

$n = 300$

True	H_0	S_n	J_n	R_n	$T_n(1)$	$T_n(3)$	SR_n	$ST_n(1)$	JR_n	$JT_n(1)$	SJR_n	$SJT_n(1)$
Ga.	Ga.	5.5	4.3	4.4	4.5	4.2	4.7	4.2	4.3	4.1	5.6	5.7
t	*t*	4.3	5.1	5.5	4.6	6.2	5.6	4.5	4.7	5.1	5.1	4.7
Cl.	Cl.	5.0	5.9	6.6	6.5	5.0	5.5	5.5	3.5	3.5	3.2	3.2
Gu.	Gu.	4.5	3.3	5.2	5.2	5.2	4.4	4.3	4.5	4.3	5.1	5.1
Ga.	*t*	5.1	12.4	66.0	61.7	22.4	55.3	46.4	58.3	50.3	51.2	42.9
Ga.	Cl.	99.1	100.0	77.7	78.8	62.5	98.3	98.3	100.0	100.0	100.0	100.0
Ga.	Gu.	60.2	36.3	7.3	6.9	6.3	49.5	49.1	26.8	26.9	57.9	57.9
t	Ga.	65.7	12.3	95.6	96.3	88.1	92.9	93.7	93.2	94.0	91.9	92.5
t	Cl.	98.3	100.0	98.0	98.0	86.5	99.6	99.6	100.0	100.0	100.0	100.0
t	Gu.	88.3	24.7	71.4	72.6	52.7	88.3	88.3	67.9	68.1	83.1	83.1
Cl.	Ga.	100.0	100.0	100	99.8	97.2	100.0	100.0	100.0	100.0	100.0	100.0
Cl.	*t*	100.0	98.5	36.6	97.7	75.9	100.0	100.0	97.9	99.6	100.0	100.0
Cl.	Gu.	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Gu.	Ga.	26.1	30.9	87.8	84.1	69.4	83.1	80.0	82.8	82.1	79.7	78.4
Gu.	*t*	47.0	25.6	5.5	4.3	5.9	32.2	31.8	19.6	19.5	30.4	29.2
Gu.	Cl.	100.0	100.0	100.0	100.0	97.5	100.0	100.0	100.0	100.0	100.0	100.0

Table 5: Percentage of rejection of H_0 by various tests of size $n = 300$ from different copula models with $\tau = 0.75$, $N = 1000$, $M = 1000$.

Results - Summary, IV

1. No significant difference between $T_n(m)$ and R_n over τ, n and copula family;
2. $T_n(1)$ performs overall better or equal than $T_n(3)$;
3. Mostly when t-copula is true under H_0, R_n; performs much better than $T_n(1)$ (similar results for hybrid tests);
4. Almost no test has power in the case of low correlation;
5. PIOS tests superior to benchmarks to differentiate between t and Gaussian copula;
6. For $\tau = 0.5$ or $\tau = 0.75$ and $n = 300$ all tests behave very well and sometimes benchmark tests are superior;
7. Hybrid tests have overall superior performance.

Local Power, I

- Asymptotic power of R_n against a local alternative in the Pitman sense for a constant $\delta > 0$:

$$H_{1,n} : P_n^{C_1,\delta}(x) = C_0\{F(x);\theta_0\} + \frac{\delta}{\sqrt{n}}\left[C_1\{F(x)\} - C_0\{F(x);\theta_0\}\right]$$

- Assume $C_1\{F(x)\} \geq C_0\{F(x);\theta_0\}$ for all $x \in \mathbb{R}^d$
  - Ensures that $P_n^{C_1,\delta}(x)$ is a copula for $0 < \delta \leq n^{1/2}$ and the departure from the null $C_0\{F(x);\theta_0\}$ increases as δ increases.

Local Power, II

Theorem

Suppose D1 holds in addition to the assumptions A2 and B1 - B3. Then under $H_{1,n}$

$$\sqrt{n}\,(R_n - p) \xrightarrow{\mathcal{L}} \mathrm{N}\{\delta m(c_0, c_1), \sigma_R^2\}$$

where

$$m(c_0, c_1) = \mathrm{E}_{c_0}\left[W(X_t) g\left\{F(X_t); \theta_0\right\}\right],$$

and $\mathrm{E}_{c_0}(\cdot)$ denotes the expectation under the null distribution c_0 or P_0, and $W(\cdot)$ as a weighting function. That is, $m(c_0, c_1)$ is a weighted expectation of $g\{F(X_t); \theta_0\}$ under P_0.

▸ Assumption $A2$ ▸ Assumptions $B1 - B3$ ▸ Assumption $D1$

Local Power, III

- Implication: as long as $m(c_0, c_1) \neq 0$
  - R_n will yield power locally
  - The asymptotic local power increases to 1 as δ increases to infinity
  - $\rightarrow$ R_n is a consistent test
  - T_n has the same asymptotic local power function as R_n
  - $\rightarrow$ T_n is also a consistent test

Local power, Simulation Study I

- Asymptotic power of R_n under alternatives in the Pitman sense
- Two settings: Clayton copula under H_0, and Gaussian copula under H_0
- $n = 500$, $N = 1000$
- Margins $F(\cdot)$ uniform on $[0, 1]$
- $(\tau_1, \tau_2) = (0.4, 0.8)$
- $\delta \in [0.0; 0.5]$

Results

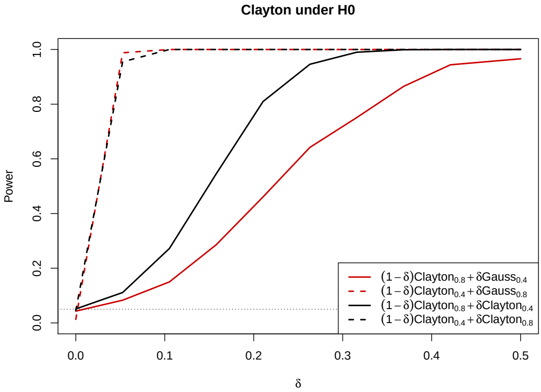


Figure 4: Local Power curves for the R_n test with Clayton copula being under H_0 and four different cases of true mixture copulas.

PIOS for the time series models, I

- Semi-Parametric Copula based Multivariate DYnamic model (SCOMDY), Chen and Fan (2006), for time series data

$$Y_t = \mu_t(\eta_1^0) + \Sigma_t^{1/2}(\eta^0)\epsilon_t,$$

- $Y_t = (Y_{t1}, \dots, Y_{td})^\top$
- $\mu_t(\eta_1^0) = \left\{\mu_{t1}(\eta_1^0), \dots, \mu_{td}(\eta_1^0)\right\}^\top = \mathsf{E}\left(Y_t | \mathcal{F}_{t-1}\right)$
- $\mathcal{F}_t$ is sigma-field generated by $(Y_{t-1}, Y_{t-2}, \dots; Z_t, Z_{t-1}, \dots)$, and Z_t is a vector of predetermined or exogenous variables.
- $\Sigma_t(\eta^0) = diag\left\{\Sigma_{t1}(\eta^0), \dots, \Sigma_{td}(\eta^0)\right\}$, where $\Sigma_{tj}(\eta^0) = \mathsf{E}\left[\left\{Y_{tj} - \mu_{tj}(\eta_1^0)\right\}^2 | \mathcal{F}_{t-1}\right]$, $j = 1, \dots, d$,
- $\epsilon_t = (\epsilon_{t1}, \dots, \epsilon_{td})^\top$, $t = 1, \dots, n$ with $\epsilon_t \overset{iid}{\sim} \mathcal{L}(0, 1)$

PIOS for the time series models, II

- Special cases of SCOMDY:
  - VAR
  - Multivariate ARMA
  - Multivariate GARCH
  - ...
- Estimation:
  - Performed with three-stage procedure
- Resulting residuals are used to construct PIOS test to test the specification of a parametric copula.

Estimation, I

1. Univariate quasi ML with $\epsilon \sim \mathrm{N}(0,1)$ to estimate $\eta = \left(\eta_1^\top, \eta_2^\top\right)^\top$:

$$\hat{\eta}_1 = \arg\min_{\eta_1 \in \Psi_1} \left[\frac{1}{n} \sum_{t=1}^{n} \left\{ Y_t - \mu_t(\eta_1) \right\}^\top \left\{ Y_t - \mu_t(\eta_1) \right\} \right]$$

and

$$\hat{\eta}_2 = \arg\min_{\eta_2 \in \Psi_2} \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{d} \left[\Sigma_{tj}^{-1}(\hat{\eta}_1, \eta_2) \left\{ Y_t - \mu_t(\hat{\eta}_1) \right\}^2 + \log \Sigma_{tj}(\hat{\eta}_1, \eta_2) \right] \right)$$

Estimation, II

2. Estimate marginal distribution $F_j(\cdot)$ of $\tilde{\epsilon}_{tj}$

$$\tilde{\epsilon}_{tj} = \Sigma_{tj}^{-1/2}(\hat{\eta})\left\{y_{tj} - \mu_{tj}(\hat{\eta}_1)\right\}, \quad j = 1, \ldots, d; \quad t = 1, \ldots, n$$

by

$$\check{F}_j(x) = \frac{1}{n+1}\sum_{t=1}^{n} \mathbf{I}\left\{\tilde{\epsilon}_{tj} \leq x\right\}, \; x \in \mathbb{R}, \; j = 1, \ldots, d.$$

Estimation, III

3. Estimate θ by

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \ell\{\check{F}(\tilde{\epsilon}_t); \theta\},$$

where $\ell(\cdot;\cdot) = \log c(\cdot;\cdot)$.

⊡ use residuals to estimate T_n and R_n

Theorem

(i) *Under conditions A1 - A2 and E1 - E4, we have*

$$\tilde{R}_n \xrightarrow{p} tr\left\{S(\theta^*)^{-1}V(\theta^*)\right\}, \quad as \quad n \to \infty.$$

(ii) *Under the null hypothesis, if A2, B1 - B3 and conditions E1 - E4 hold, we have*

$$\sqrt{n}\left(\tilde{R}_n - p\right) \xrightarrow{d} N(0, \tilde{\sigma}_R^2), \quad as \quad n \to \infty,$$

where $\tilde{\sigma}_R^2$ is the asymptotic variance.

(iii) *Under assumptions A2, B1 - B3, C1 and E1 - E4, we have*

$$\tilde{R}_n - \tilde{T}_n(m) = o_p(n^{-1/2}).$$

Assumption $A1 - A2$ | Assumptions $B1 - B3$ | Assumptions $C1$ | Assumption $E1 - E4$

PIOS for SCOMDY model

- True data-generating processes are GARCH(1,1):

$$
\begin{aligned}
x_{it} &= \sigma_{it}\varepsilon_{it} \\
\sigma^2_{it} &= \omega + \alpha x^2_{i,t-1} + \beta\sigma^2_{i,t-1}, \quad \text{for } i = 1, 2
\end{aligned}
$$

  with $\{\varepsilon_{1t}, \varepsilon_{2t}\} \sim C\{F_1(\cdot), F_2(\cdot); \theta\}$, $\varepsilon_{i,t} \perp \varepsilon_{i,t-1}$ for $i = 1, 2$.

- $\omega = 10^{-1}$, $\alpha = 0.8$ and $\beta = 0.1$

1. Simulated *iid* samples in bootstrap loop
2. Bootstrap loop with time series structure

Observation-based Bootstrap

Step 1. Generate time series $\{Y_t^{(k)}, t = 1, \ldots, n\}$ from SCOMDY model with $\hat{\eta}_1$ and $\hat{\eta}_2$ estimated from original data, and with innovation process generated from assumed copula under H_0 with $\hat{\theta}$ and marginal distribution $\check{F}$.

Step 2. Based on $\{Y_t^{(k)}, t = 1, \ldots, n\}$, estimate $\hat{\eta}_1^{(k)}$ and $\hat{\eta}_2^{(k)}$. Estimate residuals $\tilde{\epsilon}_{tj}^{(k)} = \{y_{tj}^{(k)} - \mu_{tj}(\hat{\eta}_1^{(k)})\}/\Sigma_{tj}^{1/2}(\hat{\eta}_2^{(k)})$.

Step 3. Based on $\{\tilde{\epsilon}_t^{(k)}, t = 1, \ldots, n\}$, estimate θ of copula under H_0 by two-step PMLE method and compute R_n^k;

Step 4. Repeat Steps 1- 3 N times and obtain N statistics $R_n^k, k = 1, \ldots, N$;

Step 5. Compute empirical p-value as $p_e = \frac{1}{N}\sum_{k=1}^{N} \mathbf{I}\left(|R_n^k| \geq |R_n|\right)$.

SCOMDY, I

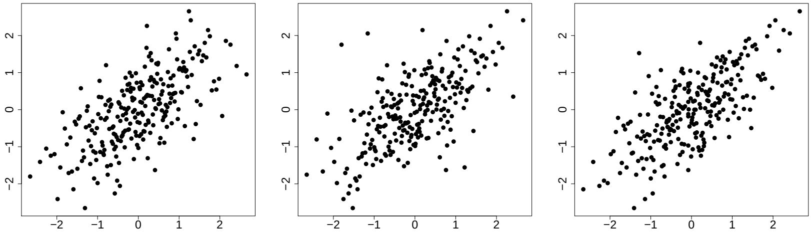
True	H_0	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
		R_n	$T_n(1)$	R_n	$T_n(1)$	R_n	$T_n(1)$
Ga	Ga	*0.062*	*0.059*	*0.058*	*0.066*	*0.085*	*0.088*
		0.058	0.061	0.046	0.043	0.042	0.041
Cl	Cl	*0.058*	*0.052*	*0.061*	*0.068*	*0.113*	*0.113*
		0.053	0.057	0.038	0.039	0.050	0.050
t	t	*0.054*	*0.053*	*0.048*	*0.044*	*0.062*	*0.043*
		0.042	0.043	0.052	0.060	0.049	0.046
Gu	Gu	*0.054*	*0.056*	*0.055*	*0.052*	*0.070*	*0.069*
		0.052	0.055	0.048	0.049	0.046	0.045

Table 6: Percentages of rejection of H_0 by various tests from different copula models for $n = 300$, $N = 300$, $M = 1000$ for the GARCH(1,1) dependent data. Type I errors were obtained using residual-based (in italic) and observation-based bootstrap procedures.

1. Application: Structural changes in the dependency

- Daily returns of Citigroup and Bank of America
- Period 2004 – 2013
- Apply GARCH(1,1) to each year separately
- Chosen is the copula dependency with the largest p-value for each year

Scatterplots



(a) year 2004 (b) year 2006 (c) year 2009

Figure 5: Scatterplots of residuals transformed to the standard normal for Citygroup/Bank of America for 2004, 2006 and 2009.

Results

	$T_n(1)$	R_n	S_n	J_n	$ST_n(1)$	SR_n	$JT_n(1)$	JR_n	$SJT_n(1)$	SJR_n
2004	Gu.	**Gu.**	**Ga.**	**Ga.**	Ga.	Ga.	Gu.	Gu.	**Ga.**	**Ga.**
2005	Gu.	Gu.	*t*	*t*	Gu.	Gu.	Gu.	Gu.	Gu.	Gu.
2006	*t*	**t**	**Ga.**	**t**	*t*	*t*	*t*	*t*	**t**	**t**
2007	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2008	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2009	Gu.	**Gu.**	**Gu.**	**Ga.**	Gu.	Gu.	Gu.	Gu.	**Gu.**	**Gu.**
2010	*t*	*t*	Gu.	*t*	Gu.	Gu.	*t*	*t*	Gu.	Gu.
2011	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2012	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
2013	*t*	*t*	*t*	Gu.	*t*	*t*	*t*	*t*	*t*	*t*

Table 7: Copulas that are preferred in each time period by each goodness-of-fit test for the Citigroup / Bank of America.

2. Application: ALAE

- Insurance dataset
- Losses and Allocated Loss Adjustment Expenses (ALAE)
- Dependence model for 1466 complete available data
- GoF tests for Gaussian, t, Gumbel and Clayton copula
- Dependence parameter estimated with PMLE

Results

copula	Clayton	Gumbel	Gauss	t
$\hat{\theta}$	0.511	1.428	0.456	0.466
$T_n(1)$	0.000 (1.316)	0.370 (0.954)	0.000 (1.223)	1.000 (0.998)
R_n	0.000 (1.323)	0.315 (0.959)	0.000 (1.274)	1.000 (1.654)
S_n	0.000 (0.407)	0.006 (0.072)	0.000 (0.118)	0.000 (0.163)
J_n	0.000 (0.095)	0.789 (0.023)	0.041 (0.038)	0.296 (0.033)
$ST_n(1)$	0.000	0.012	0.000	0.000
SR_n	0.000	0.012	0.000	0.000
$JT_n(1)$	0.000	0.740	0.000	0.592
JR_n	0.000	0.630	0.000	0.592
$SJT_n(1)$	0.000	0.018	0.000	0.000
SJR_n	0.000	0.018	0.000	0.000

Table 8: Summary of data analysis results obtained from the four copulas: Gaussian, Student's t, Clayton and Gumbel, including dependence parameter estimates, p-values with test statistics in brackets.

Conclusion

- New method based on pseudo likelihood of cross-validation
- Comparing "in-sample" and "out-of-sample"
- New tests provides a highly competitive performance
- Hybrid mechanism to combine several different tests
- Simulation show that tests perform satisfactorily in type I error control
- Comparable to best performer in Genest et al.
- Hybrid tests show superior performance

Goodness-of-Fit Test for Specification of Semiparametric Copula Dependence Models

Shulin Zhang, Ostap Okhrin

Qian M. Zhou, Peter X.-K. Song

Southwestern University of Finance and Economics

Dresden University of Technology

Simon Fraser University

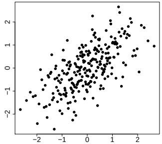
University of Michigan

http://csr.swufe.edu.cn

http://osv.vkw.tu-dresden.de

http://stat.sfu.ca/

http://www.sph.umich.edu/



Papers

Genest, C., Rémillard, B., and Beaudoin, D.
Goodness-of-fit tests for copulas: A review and a power study
Insurance: Mathematics and Economics, 44:199–213, 2009

Scaillet, O.
Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters.
Journal of Multivariate Analysis, 98:533–543, 2007

Assumptions - Law of Large Numbers

Back LLN | Back CLT | Back Local Power | Back Theorem

- Notation:
  - $\mathcal{N}(\theta^*)$ denote an open neighborhood of θ^*
  - $\ell_{\theta,j}(u_1, \ldots, u_d; \theta) = \frac{\partial \ell_\theta(u_1,\ldots,u_d;\theta)}{\partial u_j}$, $j = 1, \ldots, d$
  - $\ell_{\theta\theta,j}(u_1, \ldots, u_d; \theta) = \frac{\partial \ell_{\theta\theta}(u_1,\ldots,u_d;\theta)}{\partial u_j}$, $j = 1, \ldots, d$
- Assumptions:

A1: $\ell_\theta(u;\theta)$ and $\ell_{\theta\theta}(u;\theta)$ are continuous with respective to θ for any $u \in [0,1]^d$; there exist integrable functions $G_1(u)$ and $G_2(u)$ such that $\|\ell_\theta(u;\theta)\ell_\theta^\top(u;\theta)\| \le G_1(u)$, $\|\ell_{\theta\theta}(u;\theta)\| \le G_2(u)$ $\forall \theta \in \mathcal{N}(\theta^*)$

A2: Matrix $S(\theta^*) = -\mathsf{E}_0\left[\ell_{\theta\theta}\{F(X_1)\};\theta^*\right]$ is finite and nonsingular.

Assumptions - CLT I

Back CLT | Back Local Power | Back Theorem

B1: Denote $J_i(u) = const \times \prod_{k=1}^{d} \{u_k(1-u_k)\}^{-\xi_{ik}}$, where $\xi_{ik} \geq 0$, $i = 1, 2$, ξ_{ik} are some constants. Suppose that for all $\theta \in \mathbb{N}_{\theta^*}$, $\|\ell_\theta(u;\theta)\ell_\theta^\top(u;\theta)\| \leq J_1(u)$, $\|\ell_{\theta\theta}(u;\theta)\| \leq J_2(u)$, and $\mathrm{E}_0\left[J_i^2\{F(X_1)\}\right] < \infty$.

B2: Suppose that both $\ell_{\theta,k}(u;\theta)$ and $\ell_{\theta\theta,k}(u;\theta)$, $k = 1, 2, \ldots, d$ exist and are continuous. Denote $\tilde{J}_i^k(u) = const \times \{u_k(1-u_k)\}^{-\tilde{\xi}_{ik}} \prod_{j=1, j\neq k}^{d} \{u_j(1-u_j)\}^{-\xi_{ij}}$, where $\tilde{\xi}_{ij} > \xi_{ij}$ are some constants, such that for all $\theta \in \mathbb{N}(\theta^*)$, $\|\ell_{\theta,k}(u;\theta)\| \leq \tilde{J}_1^k(u)$ and $\|\ell_{\theta\theta,k}(u;\theta)\| \leq \tilde{J}_2^k(u)$, and furthermore, $\mathrm{E}_0\left[\tilde{J}_i\{F(X_1)\}\right] < \infty$, $i = 1, 2$ and $k = 1, 2, \ldots, d$.

Assumptions - CLT II

▸ Back CLT ▸ Back Theorem

B3: Suppose $\frac{\partial \ell_{\theta\theta}(u;\theta)}{\partial \theta_k}$, $k = 1, 2, \ldots, p$ exist and are continuous with $\theta \in \mathbb{N}(\theta^*)$, and there exists an integrable function $G_3(u)$ such that $\|\frac{\partial \ell_{\theta\theta}(u;\theta)}{\partial \theta_k}\| \leq G_3(u)$ for all $\theta \in \mathbb{N}(\theta^*)$, $k = 1, \ldots, d$.

C1: The block size m is of order $o(n^a)$ with $0 \leq a \leq \frac{1}{4}$.

Assumption - Local Power of Evaluation

▸ Back Local Power

D1: Both the copula $C_0(\cdot;\theta_0)$ and $C_1(\cdot)$ in $P_n^{C_1,\delta}(x)$ are absolutely continuous with respective to square integrable densities $c_0(\cdot;\theta_0)$ and $c_1(\cdot)$. Moreover

$$\int_{u\in[0,1]^d}\left[\sqrt{n}\left\{\sqrt{p_n^{c_1,\delta}(u)}-\sqrt{p_0(u)}\right\}-\frac{1}{2}\delta g(u)\sqrt{p_0(u)}\right]^2 du \to 0,$$

as $n\to\infty$, where $p_n^{c_1,\delta}(u)=(1-\frac{\delta}{\sqrt{n}})c_0(u;\theta_0)+\frac{\delta}{\sqrt{n}}c_1(u)$, $p_0(u)=c_0(u;\theta_0)$ and $g(u)=\frac{c_1(u)-c_0(u;\theta_0)}{c_0(u;\theta_0)}$.

Assumptions - Large sample properties I

▸ Back Theorem

E1. $\left\{ \left(Y_t^\top, Z_t^\top \right), t = 1, \ldots, n \right\}$ is stationary β-mixing with serial decay rate of order $O(t^{-\frac{\xi}{\xi-1}})$ for some $\xi > 1$

E2. $\hat{\eta}$ is a root-n consistent estimator of η_0

E3. For all $t \geq 1$ and $j = 1, \ldots, d$,
$\epsilon_{tj} = \Sigma_{tj}^{-1/2}(\eta^0) \left\{ Y_{tj} - \mu_{tj}(\eta_1^0) \right\}$ is continuously differentiable in the neighbood of η^0, and
$\omega_1 = \mathrm{E}_0 \left\{ \Sigma_{tj}^{-1/2}(\eta^0) \dot{\mu}_{tj}(\eta_1^0) \right\} < \infty$ and
$\omega_2 = \mathrm{E}_0 \left\{ \Sigma_{tj}^{-1}(\eta^0) \dot{\Sigma}_{tj}(\eta^0) \right\} < \infty$, where $\dot{\mu}_{tj}(\eta_1^0) = \frac{\partial \mu_{tj}(\eta_1^0)}{\partial \eta_1}$ and $\dot{\Sigma}_{tj}(\eta^0) = \frac{\partial \Sigma_{tj}(\eta^0)}{\partial \eta}$.

Assumptions - Large sample properties II

E4. The PMLE $\hat{\theta}$ has the following asymptotic expansion

$$\hat{\theta} - \theta^* = \frac{1}{n} \sum_{t=1}^{n} \varphi_\theta \left(U_t; \theta^* \right) + o_p(n^{-1/2}),$$

where $U_t = (U_{t1}, \ldots, U_{td})^\top$, $U_{tj} = F_j(\epsilon_{tj})$,
$j = 1, \ldots, d, t = 1, \ldots, n$ and

$$\varphi_\theta \left(U_t; \theta^* \right) = S(\theta^*)^{-1} (\ell_\theta \left(U_t; \theta^* \right)$$
$$+ \sum_{j=1}^{d} \mathrm{E}_0 \left[\ell_{\theta,j} \left(U_s; \theta^* \right) \{ \mathbf{I}(U_{tj} \leq U_{sj}) - U_{sj} \} | U_{tj} \right]).$$