Bootstrap methods for copulas, with applications to stock index data

László Varga
András Zempléni
László Sajtos

Eötvös Loránd University (ELTE)
Faculty of Science
Institute of Mathematics
Department of Probability Theory and Statistics
Budapest, Hungary

Applied Stochastic Models and Data Analysis (ASMDA)
7 - 10 June 2011

Outline

Copulas

- C is a copula, if it is a d-dimensional random vector with marginals $\sim$ Unif[0,1]
- Existence (Sklar's Theorem): to any d-dimensional random variable X with c.d.f. F and marginals F_i (i=1,...,d) there exists a copula C : $F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$
- Uniqueness: if F_i are continuous (i=1,...,d)
- Separation of the marginal model and the dependence

Archimedean copulas

- Copula generator function: $\phi : [0, 1] \rightarrow [0, \infty]$
  where ϕ is continuous, strictly decreasing and $\phi(1)$=0
- d-variate Archimedean copula: $C_\phi(\underline{u}) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right)$
- Gumbel: $\phi(u) = (-\log(u))^\theta$ where $\theta \in [1, \infty]$
- Clayton: $\phi(u) = u^{-\theta} - 1$ where $\theta > 0$

Copulas

- C is a copula, if it is a d-dimensional random vector with marginals $\sim$ Unif[0,1]
- Existence (Sklar′s Theorem): to any d-dimensional random variable X with c.d.f. F and marginals F_i (i=1,...,d) there exists a copula C : $F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$
- Uniqueness: if F_i are continuous (i=1,...,d)
- Separation of the marginal model and the dependence

Archimedean copulas

- Copula generator function: $\phi : [0,1] \to [0,\infty]$
  where ϕ is continuous, strictly decreasing and ϕ(1)=0
- d-variate Archimedean copula: $C_\phi(\underline{u}) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right)$
- Gumbel: $\phi(u) = (-\log(u))^\theta$ where $\theta \in [1,\infty]$
- Clayton: $\phi(u) = u^{-\theta} - 1$ where $\theta > 0$

Copulas

- C is a copula, if it is a d-dimensional random vector with marginals $\sim$ Unif[0,1]
- Existence (Sklar′s Theorem): to any d-dimensional random variable X with c.d.f. F and marginals F_i (i=1,...,d) there exists a copula C : $F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$
- Uniqueness: if F_i are continuous (i=1,...,d)
- Separation of the marginal model and the dependence

Archimedean copulas

- Copula generator function: $\phi : [0, 1] \to [0, \infty]$
  where ϕ is continuous, strictly decreasing and ϕ(1)=0
- d-variate Archimedean copula: $C_\phi(\underline{u}) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right)$
- Gumbel: $\phi(u) = (-\log(u))^\theta$ where $\theta \in [1, \infty]$
- Clayton: $\phi(u) = u^{-\theta} - 1$ where $\theta > 0$

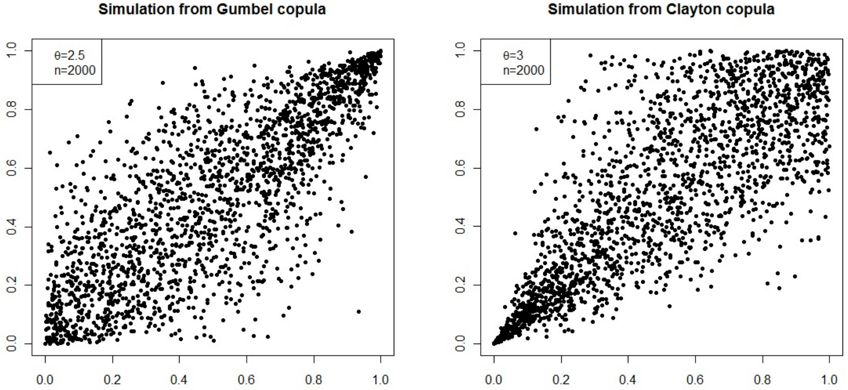
Copulas

- C is a copula, if it is a d-dimensional random vector with marginals $\sim$ Unif[0,1]
- Existence (Sklar′s Theorem): to any d-dimensional random variable X with c.d.f. F and marginals F_i (i=1,...,d) there exists a copula C : $F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$
- Uniqueness: if F_i are continuous (i=1,...,d)
- Separation of the marginal model and the dependence

Archimedean copulas

- Copula generator function: $\phi : [0,1] \to [0,\infty]$
  where ϕ is continuous, strictly decreasing and ϕ(1)=0
- d-variate Archimedean copula: $C_\phi(\underline{u}) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right)$
- Gumbel: $\phi(u) = (-\log(u))^\theta$ where $\theta \in [1,\infty]$
- Clayton: $\phi(u) = u^{-\theta} - 1$ where $\theta > 0$

Copulas - examples



Outline

Goodness-of-fit (GoF) tests in general

- Let $X_1, \ldots, X_n$ be random vectors, with c.d.f. $F : \mathbb{R}^d \to \mathbb{R}$
- We want to test the following hypothesis:
  $H_0 : F \in \mathcal{F} = \{F_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^p$ is an open set
- Estimate of the parameter: $\theta_n = T_n(X_1, \ldots, X_n)$
- Cramér-von Mises statistics based on the empirical process or its functional
  $\mathbb{G}_n^F = \sqrt{n}(F_n - F_{\theta_n})$
  $S_n = \phi(\mathbb{G}_n^F)$
- Problem: the asymptotic null distribution of S_n is rarely known, but if known, usually it is rather complicated

Goodness-of-fit (GoF) tests in general

- Let $X_1, \ldots, X_n$ be random vectors, with c.d.f. $F : \mathbb{R}^d \to \mathbb{R}$
- We want to test the following hypothesis:
  $H_0 : F \in \mathcal{F} = \{F_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^p$ is an open set
- Estimate of the parameter: $\theta_n = T_n(X_1, \ldots, X_n)$
- Cramér-von Mises statistics based on the empirical process or its functional
  $$\mathbb{G}_n^F = \sqrt{n}(F_n - F_{\theta_n})$$
  $$S_n = \phi(\mathbb{G}_n^F)$$
- Problem: the asymptotic null distribution of S_n is rarely known, but if known, usually it is rather complicated

Goodness-of-fit (GoF) tests in general

- Solution: *parametric bootstrap* - generation of N bootstrap repetitions: ($k \in 1, \ldots, N$)
  1. Estimate of the parameter: θ_n
  2. Generate n independent observations from distribution F_{θ_n}: $X^*_{1,k}, \ldots, X^*_{n,k}$
  3. Estimate the parameter and the empirical c.d.f. from the sample:
     $$\theta^*_{n,k} = T_n(X^*_{1,k}, \ldots, X^*_{n,k})$$
     $$F^*_{n,k}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X^*_{i,k} \leq x)$$
  4. $\mathbb{G}^{F^*}_{n,k} = \sqrt{n}(F^*_{n,k} - F_{\theta^*_{n,k}})$
     $S^*_{n,k} = \phi(\mathbb{G}^{F^*}_{n,k})$
  5. Calculate the test statistics and the critical values
- Instead of F_θ an abstract quantity A_θ depending on P_θ
- Let's suppose that $\Phi_n = \sqrt{n}(\theta_n - \theta) \rightsquigarrow \Phi$ and
  $$\mathbb{A}_n = \sqrt{n}(A_n - A) \rightsquigarrow \mathbb{A}$$
- For the convergence of the bootstrap alteregos we need further regularity assumptions.

Goodness-of-fit (GoF) tests in general

- Solution: *parametric bootstrap* - generation of N bootstrap repetitions: $(k \in 1, \ldots, N)$
  1. Estimate of the parameter: θ_n
  2. Generate n independent observations from distribution F_{θ_n}: $X^*_{1,k}, \ldots, X^*_{n,k}$
  3. Estimate the parameter and the empirical c.d.f. from the sample:
     $$\theta^*_{n,k} = T_n(X^*_{1,k}, \ldots, X^*_{n,k})$$
     $$F^*_{n,k}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X^*_{i,k} \leq x)$$
  4. $\mathbb{G}^{F^*}_{n,k} = \sqrt{n}(F^*_{n,k} - F_{\theta^*_{n,k}})$
     $S^*_{n,k} = \phi(\mathbb{G}^{F^*}_{n,k})$
  5. Calculate the test statistics and the critical values
- Instead of F_θ an abstract quantity A_θ depending on P_θ
- Let's suppose that $\Phi_n = \sqrt{n}(\theta_n - \theta) \rightsquigarrow \Phi$ and
  $$\mathbb{A}_n = \sqrt{n}(A_n - A) \rightsquigarrow \mathbb{A}$$
- For the convergence of the bootstrap alteregos we need further regularity assumptions.

GoF tests - regularity assumptions

- Def.: A $\mathcal{P} = \{\theta \in \Theta\}$ is said to belong to the class $S(\lambda)$ (given λ reference measure), if
  1. The measure P_θ is abs. continuous with respect to λ for all $\theta \in \Theta$.
  2. The density $p_\theta = dP_\theta / d\lambda$ admits first and second order derivatives with respect to $\theta \in \Theta$. Gradient vector: $\dot{p}_\theta$, Hessian matrix: $\ddot{p}_\theta$
  3. For arbitrary $u \in R^d$ and every $\theta \in \Theta$, the mappings $\theta \to \dot{p}_\theta(u)/p_\theta(u)$ and $\theta \to \ddot{p}_\theta(u)/p_\theta(u)$ are continuous at θ_0, P_{θ_0}-a.s.
  4. For every $\theta_0 \in \Theta$, there exist a neighborhood $\mathcal{N}$ of θ_0 and a λ-integrable function $h : R^d \to R$ such that for all $u \in R^d$, $\sup_{\theta \in \mathcal{N}} ||\ddot{p}_\theta(u)|| = h(u)$.
  5. For every $\theta_0 \in \Theta$, there exist a neighborhood $\mathcal{N}$ of θ_0 and P_{θ_0}-integrable functions $h_1, h_2 : R^d \to R$ such that for every $u \in R^d$, $\sup\limits_{\theta \in \mathcal{N}} ||\frac{\dot{p}_\theta(u)}{p_\theta(u)}||^2 \leq h_1(u)$ and $\sup\limits_{\theta \in \mathcal{N}} ||\frac{\ddot{p}_\theta(u)}{p_\theta(u)}|| \leq h_2(u)$.
- Coroll.: Let $\mathcal{P} \in S(\lambda)$, if $P \in \mathcal{P}$ and $n \to \infty$ then

$$\mathbb{W}_{P,n} := n^{-1/2} \sum_{i=1}^{n} \frac{\dot{p}^T(X_i)}{p(X_i)} \rightsquigarrow \mathbb{W}_P \sim N(0, I_P)$$

  where I_P: Fisher information matrix

GoF tests - regularity assumptions

- Def.: A $\mathcal{P} = \{\theta \in \Theta\}$ is said to belong to the class $S(\lambda)$ (given λ reference measure), if
  1. The measure P_θ is abs. continuous with respect to λ for all $\theta \in \Theta$.
  2. The density $p_\theta = dP_\theta / d\lambda$ admits first and second order derivatives with respect to $\theta \in \Theta$. Gradient vector: $\dot{p}_\theta$, Hessian matrix: $\ddot{p}_\theta$
  3. For arbitrary $u \in R^d$ and every $\theta \in \Theta$, the mappings $\theta \to \dot{p}_\theta(u)/p_\theta(u)$ and $\theta \to \ddot{p}_\theta(u)/p_\theta(u)$ are continuous at θ_0, P_{θ_0}-a.s.
  4. For every $\theta_0 \in \Theta$, there exist a neighborhood $\mathcal{N}$ of θ_0 and a λ-integrable function $h : R^d \to R$ such that for all $u \in R^d$, $\sup_{\theta \in \mathcal{N}} ||\ddot{p}_\theta(u)|| = h(u)$.
  5. For every $\theta_0 \in \Theta$, there exist a neighborhood $\mathcal{N}$ of θ_0 and P_{θ_0}-integrable functions $h_1, h_2 : R^d \to R$ such that for every $u \in R^d$, $\sup\limits_{\theta \in \mathcal{N}} ||\frac{\dot{p}_\theta(u)}{p_\theta(u)}||^2 \leq h_1(u)$ and $\sup\limits_{\theta \in \mathcal{N}} ||\frac{\ddot{p}_\theta(u)}{p_\theta(u)}|| \leq h_2(u)$.
- Coroll.: Let $\mathcal{P} \in S(\lambda)$, if $P \in \mathcal{P}$ and $n \to \infty$ then

$$\mathbb{W}_{P,n} := n^{-1/2} \sum_{i=1}^{n} \frac{\dot{p}^T(X_i)}{p(X_i)} \rightsquigarrow \mathbb{W}_P \sim N(0, I_P)$$

where I_P: Fisher information matrix

Convergence of bootstrap alteregos

- Def.: Let $U_1, \ldots, U_n$ be a random sample from $P = P_{\theta_0}$.
  A sequence A_n is said to be P_{θ_0}-regular for $A = A_{\theta_0}$, if the process $(\mathbb{A}_n, \mathbb{W}_{P,n}) \rightsquigarrow (\mathbb{A}, \mathbb{W}_P)$ and $\dot{A}(t) = E(\mathbb{A}(t)\mathbb{W}_P^T)$ for every t.
- Def.: The A_n sequence is said to be $\mathcal{P}$-regular for $\mathcal{A}$ ($\mathcal{A} = \{A_{\theta_0} : \theta_0 \in \Theta\}$) if it is P_{θ_0}-regular at all $A = A_{\theta_0}$.
- Theorem: Assume that
  - $\mathcal{P} \in S(\lambda)$
  - if $P \in \mathcal{P}$ and $n \to \infty$, $(\mathbb{A}_n, \Phi_n) \rightsquigarrow (\mathbb{A}, \Phi)$, where the limit is a centered Gaussian process
  - (A_n, θ_n) is $\mathcal{P}$-regular for $\mathcal{A} \times \Theta$

  Then $\quad (\mathbb{G}_n^A, \mathbb{G}_n^{A^*}) \rightsquigarrow (\mathbb{G}^A, \mathbb{G}^{A^*})$
  where $\mathbb{G}^{A^*}$ is an independent copy of $\mathbb{G}^A$.
- Prop.: The assumptions hold for the empirical copula process (Genest&Rémillard (2005))

Convergence of bootstrap alteregos

- Def.: Let $U_1, \ldots, U_n$ be a random sample from $P = P_{\theta_0}$.
  A sequence A_n is said to be P_{θ_0}-regular for $A = A_{\theta_0}$, if the process $(\mathbb{A}_n, \mathbb{W}_{P,n}) \rightsquigarrow (\mathbb{A}, \mathbb{W}_P)$ and $\dot{A}(t) = E(\mathbb{A}(t)\mathbb{W}_P^T)$ for every t.
- Def.: The A_n sequence is said to be $\mathcal{P}$-regular for $\mathcal{A}$ $(\mathcal{A} = \{A_{\theta_0} : \theta_0 \in \Theta\})$ if it is P_{θ_0}-regular at all $A = A_{\theta_0}$.
- Theorem: Assume that
  - $\mathcal{P} \in S(\lambda)$
  - if $P \in \mathcal{P}$ and $n \to \infty$, $(\mathbb{A}_n, \Phi_n) \rightsquigarrow (\mathbb{A}, \Phi)$, where the limit is a centered Gaussian process
  - (A_n, θ_n) is $\mathcal{P}$-regular for $\mathcal{A} \times \Theta$

  Then $(\mathbb{G}_n^A, \mathbb{G}_n^{A^*}) \rightsquigarrow (\mathbb{G}^A, \mathbb{G}^{A^*})$
  where $\mathbb{G}^{A^*}$ is an independent copy of $\mathbb{G}^A$.
- Prop.: The assumptions hold for the empirical copula process (Genest&Rémillard (2005))

Convergence of bootstrap alteregos

- Def.: Let $U_1, \ldots, U_n$ be a random sample from $P = P_{\theta_0}$.
  A sequence A_n is said to be P_{θ_0}-regular for $A = A_{\theta_0}$, if the process $(\mathbb{A}_n, \mathbb{W}_{P,n}) \rightsquigarrow (\mathbb{A}, \mathbb{W}_P)$ and $\dot{A}(t) = E(\mathbb{A}(t)\mathbb{W}_P^T)$ for every t.
- Def.: The A_n sequence is said to be $\mathcal{P}$-regular for $\mathcal{A}$ $(\mathcal{A} = \{A_{\theta_0} : \theta_0 \in \Theta\})$ if it is P_{θ_0}-regular at all $A = A_{\theta_0}$.
- Theorem: Assume that
  - $\mathcal{P} \in S(\lambda)$
  - if $P \in \mathcal{P}$ and $n \to \infty$, $(\mathbb{A}_n, \Phi_n) \rightsquigarrow (\mathbb{A}, \Phi)$, where the limit is a centered Gaussian process
  - (A_n, θ_n) is $\mathcal{P}$-regular for $\mathcal{A} \times \Theta$

  Then $\quad (\mathbb{G}_n^A, \mathbb{G}_n^{A^*}) \rightsquigarrow (\mathbb{G}^A, \mathbb{G}^{A^*})$
  where $\mathbb{G}^{A^*}$ is an independent copy of $\mathbb{G}^A$.
- Prop.: The assumptions hold for the empirical copula process (Genest&Rémillard (2005))

Convergence of bootstrap alteregos

- Def.: Let $U_1, \ldots, U_n$ be a random sample from $P = P_{\theta_0}$.
  A sequence A_n is said to be P_{θ_0}-regular for $A = A_{\theta_0}$, if the process $(\mathbb{A}_n, \mathbb{W}_{P,n}) \rightsquigarrow (\mathbb{A}, \mathbb{W}_P)$ and $\dot{A}(t) = E(\mathbb{A}(t)\mathbb{W}_P^T)$ for every t.
- Def.: The A_n sequence is said to be $\mathcal{P}$-regular for $\mathcal{A}$ $(\mathcal{A} = \{A_{\theta_0} : \theta_0 \in \Theta\})$ if it is P_{θ_0}-regular at all $A = A_{\theta_0}$.
- Theorem: Assume that
  - $\mathcal{P} \in S(\lambda)$
  - if $P \in \mathcal{P}$ and $n \to \infty$, $(\mathbb{A}_n, \Phi_n) \rightsquigarrow (\mathbb{A}, \Phi)$, where the limit is a centered Gaussian process
  - (A_n, θ_n) is $\mathcal{P}$-regular for $\mathcal{A} \times \Theta$

  Then $\quad (\mathbb{G}_n^A, \mathbb{G}_n^{A^*}) \rightsquigarrow (\mathbb{G}^A, \mathbb{G}^{A^*})$
  where $\mathbb{G}^{A^*}$ is an independent copy of $\mathbb{G}^A$.
- Prop.: The assumptions hold for the empirical copula process (Genest&Rémillard (2005))

Outline

Limit distribution of the empirical copula process

- Let $X_1, \dots, X_n$ be i.i.d. bivariate random vectors with continuous c.d.f. F, marginal distribution functions F_1, F_2 and copula C:
  $$C(u_1, u_2) = F(F_1^{-}(u_1), F_2^{-}(u_2)) \text{ (Sklar's theorem)}$$
- Empirical copula:
  $C_n(u_1, u_2) = F_n(F_{n1}^{-}(u_1), F_{n2}^{-}(u_2))$,
  where $F_n(x) = F_n(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\}$,
  $$F_{np}(x_p) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_{ip} \le x_p\},\ p = 1, 2.$$
- $\rightsquigarrow$ denote weak convergence in space $l^\infty([0,1]^2)$ of all uniformly bounded functions on $[0,1]^2$

The limit distribution of the empirical copula process and the pdm method

- Theorem: If the copula C possesses continuous partial derivatives $\partial_1 C, \partial_2 C$ on $[0,1]^2$ then the empirical copula process converges weakly:

$$\alpha_n := \sqrt{n}(C_n - C) \rightsquigarrow \mathbb{G}_C$$

  - where $\mathbb{G}_C(u_1, u_2) = \mathbb{B}_C(u_1, u_2) - \partial_1 C(u_1, u_2)\mathbb{B}_C(u_1, 1) - \partial_2 C(u_1, u_2)\mathbb{B}_C(1, u_2)$
  - where $\mathbb{B}_C$ is a centered Gaussian field with the following covariance structure: $Cov(\mathbb{B}_C(u_1, u_2), \mathbb{B}_C(v_1, v_2)) = C(u_1 \wedge v_1, u_2 \wedge v_2) - C(u_1, u_2)C(v_1, v_2)$.
- The **pdm method** (Bücher&Dette (2010)):
  - Let $Z_1, \ldots, Z_n$ be centered i.i.d. random variables; $D^2 Z_i = 1$; independent of $X_1, \ldots, X_n$; $\int_0^\infty \sqrt{P(|Z_1| > x)}dx < \infty$.
  - $C_n^*(u) = \frac{1}{n}\sum_{i=1}^{n} Z_i \mathbb{I}\{F_{n1}(X_{i1}) \le u_1, F_{n2}(X_{i2}) \le u_2\}$

The limit distribution of the empirical copula process and the pdm method

- Theorem: If the copula C possesses continuous partial derivatives $\partial_1 C, \partial_2 C$ on $[0,1]^2$ then the empirical copula process converges weakly:

$$\alpha_n := \sqrt{n}(C_n - C) \rightsquigarrow \mathbb{G}_C$$

  - where $\mathbb{G}_C(u_1,u_2) = \mathbb{B}_C(u_1,u_2) - \partial_1 C(u_1,u_2)\mathbb{B}_C(u_1,1) - \partial_2 C(u_1,u_2)\mathbb{B}_C(1,u_2)$
  - where $\mathbb{B}_C$ is a centered Gaussian field with the following covariance structure: $Cov(\mathbb{B}_C(u_1,u_2),\mathbb{B}_C(v_1,v_2)) = C(u_1 \wedge v_1, u_2 \wedge v_2) - C(u_1,u_2)C(v_1,v_2)$.
- The **pdm method** (Bücher&Dette (2010)):
  - Let $Z_1,\ldots,Z_n$ be centered i.i.d. random variables; $D^2 Z_i = 1$; independent of $X_1,\ldots,X_n$; $\int\limits_0^\infty \sqrt{P(|Z_1| > x)}dx < \infty$.
  - $C_n^*(u) = \frac{1}{n}\sum\limits_{i=1}^n Z_i \mathbb{I}\{F_{n1}(X_{i1}) \le u_1, F_{n2}(X_{i2}) \le u_2\}$

The pdm method

- Approximate the partial derivatives:

$$\widehat{\partial_1 C}(u,v) := \frac{C_n(u+h,v)-C_n(u-h,v)}{2h}$$
$$\widehat{\partial_2 C}(u,v) := \frac{C_n(u,v+h)-C_n(u,v-h)}{2h}$$
$$\text{where } h = n^{-1/2} \to 0$$

- Estimate of $\mathbb{B}_C$: $\beta_n := \sqrt{n}(C_n^* - \overline{Z}_n C_n)$
- Estimate of $\mathbb{G}_C$: $\alpha_n^{pdm}(u_1,u_2) :=$

$$\beta_n(u_1,u_2) - \widehat{\partial_1 C}(u_1,u_2)\beta_n(u_1,1) - \widehat{\partial_2 C}(u_1,u_2)\beta_n(1,u_2)$$

- Theorem: Using the foregoing notations

$$(\alpha_n, \alpha_n^{pdm}) \rightsquigarrow (\mathbb{G}_C, \mathbb{G}'_C) \text{ in } \quad l^\infty([0,1]^2)^2$$

where $\mathbb{G}'_C$ is an independent copy of $\mathbb{G}_C$.

Cramér-von Mises teststatistics:

- $L_n = \int_{[0,1]^2} \alpha_n^2(\underline{x})d\underline{x}$
- $L_n^{pdm} = \int_{[0,1]^2} (\alpha_n^{pdm})^2(\underline{x})d\underline{x}$

The pdm method

- Approximate the partial derivatives:

$$\widehat{\partial_1 C}(u,v) := \frac{C_n(u+h,v)-C_n(u-h,v)}{2h}$$
$$\widehat{\partial_2 C}(u,v) := \frac{C_n(u,v+h)-C_n(u,v-h)}{2h}$$
$$\text{where } h = n^{-1/2} \to 0$$

- Estimate of $\mathbb{B}_C$: $\beta_n := \sqrt{n}(C_n^* - \overline{Z}_n C_n)$
- Estimate of $\mathbb{G}_C$: $\alpha_n^{pdm}(u_1,u_2) :=$

$$\beta_n(u_1,u_2) - \widehat{\partial_1 C}(u_1,u_2)\beta_n(u_1,1) - \widehat{\partial_2 C}(u_1,u_2)\beta_n(1,u_2)$$

- Theorem: Using the foregoing notations

$$(\alpha_n, \alpha_n^{pdm}) \rightsquigarrow (\mathbb{G}_C, \mathbb{G}'_C) \text{ in } \quad l^\infty([0,1]^2)^2$$

where $\mathbb{G}'_C$ is an independent copy of $\mathbb{G}_C$.

Cramér-von Mises teststatistics:

- $L_n = \int_{[0,1]^2} \alpha_n^2(\underline{x})d\underline{x}$
- $L_n^{pdm} = \int_{[0,1]^2} (\alpha_n^{pdm})^2(\underline{x})d\underline{x}$

The pdm method

- Approximate the partial derivatives:

$$\widehat{\partial_1 C}(u,v) := \frac{C_n(u+h,v)-C_n(u-h,v)}{2h}$$
$$\widehat{\partial_2 C}(u,v) := \frac{C_n(u,v+h)-C_n(u,v-h)}{2h}$$
where $h = n^{-1/2} \to 0$

- Estimate of $\mathbb{B}_C$: $\beta_n := \sqrt{n}(C_n^* - \overline{Z}_n C_n)$
- Estimate of $\mathbb{G}_C$: $\alpha_n^{pdm}(u_1,u_2) :=$

$$\beta_n(u_1,u_2) - \widehat{\partial_1 C}(u_1,u_2)\beta_n(u_1,1) - \widehat{\partial_2 C}(u_1,u_2)\beta_n(1,u_2)$$

- Theorem: Using the foregoing notations

$$(\alpha_n, \alpha_n^{pdm}) \rightsquigarrow (\mathbb{G}_C, \mathbb{G}'_C) \text{ in } \quad l^\infty([0,1]^2)^2$$

where $\mathbb{G}'_C$ is an independent copy of $\mathbb{G}_C$.

Cramér-von Mises teststatistics:

- $L_n = \int_{[0,1]^2} \alpha_n^2(\underline{x})d\underline{x}$
- $L_n^{pdm} = \int_{[0,1]^2} (\alpha_n^{pdm})^2(\underline{x})d\underline{x}$

Outline

Kendall's process

- Kendall's transform (K function):
$$K(\theta, t) = P(C_\theta(F_1(X_1), ..., F_1(X_1)) \le t)$$
advantage: one-dimensional
- For Archim. copulas $K(\theta, t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!} \left[\phi_\theta(t)^i\right] f_i(\theta, t)$
where $f_i(\theta, t) = \frac{d^i}{dx^i} \phi_\theta^{-1}(x) \Big|_{x=\phi_\theta(t)}$
- Kendall's process: $\kappa_n(t) = \sqrt{n}(K(\theta_n, t) - K_n(t))$
where $K_n(t)$: empirical version of $K(\theta, t)$
- Cramér von Mises statistics:
$$S_n = \int_0^1 (\kappa_n(t))^2 \Phi(t) dt$$

Focused Regions	$\Phi(t)$
Global	1
Upper Tail	$1 - K(\theta_n, t)$
Lower Tail	$K(\theta_n, t)$
Lower and Upper Tail	$K(\theta_n, t)(1 - K(\theta_n, t))$

Kendall's process

- Kendall's transform (K function):
  $$K(\theta, t) = P(C_\theta(F_1(X_1), ..., F_1(X_1)) \leq t)$$
  advantage: one-dimensional
- For Archim. copulas $K(\theta, t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!} \left[\phi_\theta(t)^i\right] f_i(\theta, t)$
  where $f_i(\theta, t) = \frac{d^i}{dx^i} \phi_\theta^{-1}(x) \Big|_{x=\phi_\theta(t)}$
- Kendall's process: $\kappa_n(t) = \sqrt{n}(K(\theta_n, t) - K_n(t))$
  where $K_n(t)$: empirical version of $K(\theta, t)$
- Cramér von Mises statistics:
  $$S_n = \int_0^1 (\kappa_n(t))^2 \Phi(t) dt$$

Focused Regions	$\Phi(t)$
Global	1
Upper Tail	$1 - K(\theta_n, t)$
Lower Tail	$K(\theta_n, t)$
Lower and Upper Tail	$K(\theta_n, t)(1 - K(\theta_n, t))$

Kendall's process

- Kendall's transform (K function):
$$K(\theta, t) = P(C_\theta(F_1(X_1), ..., F_1(X_1)) \leq t)$$
advantage: one-dimensional
- For Archim. copulas $K(\theta, t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!} \left[\phi_\theta(t)^i\right] f_i(\theta, t)$
where $f_i(\theta, t) = \frac{d^i}{dx^i} \phi_\theta^{-1}(x)\Big|_{x=\phi_\theta(t)}$
- Kendall's process: $\kappa_n(t) = \sqrt{n}(K(\theta_n, t) - K_n(t))$
where $K_n(t)$: empirical version of $K(\theta, t)$
- Cramér von Mises statistics:
$$S_n = \int_0^1 (\kappa_n(t))^2 \Phi(t) dt$$

Focused Regions	$\Phi(t)$
Global	1
Upper Tail	$1 - K(\theta_n, t)$
Lower Tail	$K(\theta_n, t)$
Lower and Upper Tail	$K(\theta_n, t)(1 - K(\theta_n, t))$

Analysis of stock index data

2 stock index daily maxima for 9 years (2002-2011):
NASDAQ Composite & Dow Jones Industrial Average

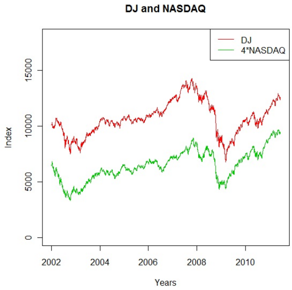
- Method: copula fitting and goodness-of-fit testing
- Analysed copulas: Gumbel, Clayton
- Tests:
  1. Cramér-von Mises teststatistics based on Kendall's process
  2. Cramér-von Mises teststatistics based on empirical copula process estimated with parametric bootstrap
  3. Cramér-von Mises teststatistics based on empirical copula process estimated using α_n
  4. Testing correlation structure with multivariate Bartlett test

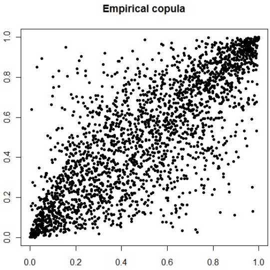
Analysis of stock index data

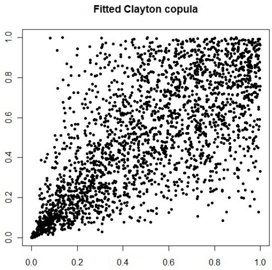
2 stock index daily maxima for 9 years (2002-2011):
NASDAQ Composite & Dow Jones Industrial Average

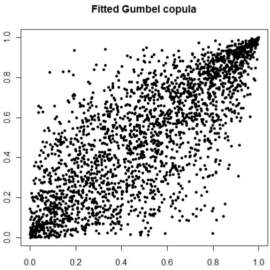
- Method: copula fitting and goodness-of-fit testing
- Analysed copulas: Gumbel, Clayton
- Tests:
    1. Cramér-von Mises teststatistics based on Kendall's process
    2. Cramér-von Mises teststatistics based on empirical copula process estimated with parametric bootstrap
    3. Cramér-von Mises teststatistics based on empirical copula process estimated using α_n
    4. Testing correlation structure with multivariate Bartlett test

Stock index data and fitted copulas

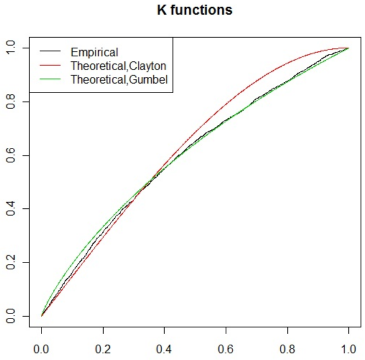




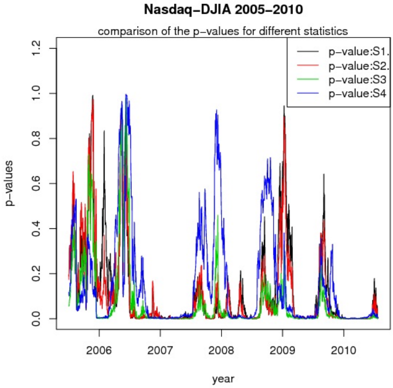




CvM test based on Kendall's process

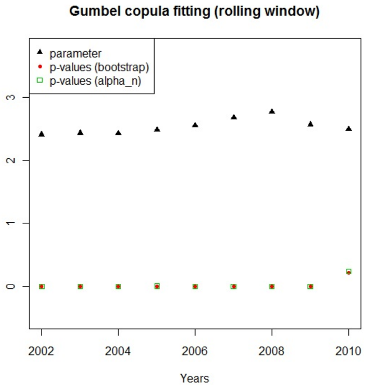


Gumbel is better

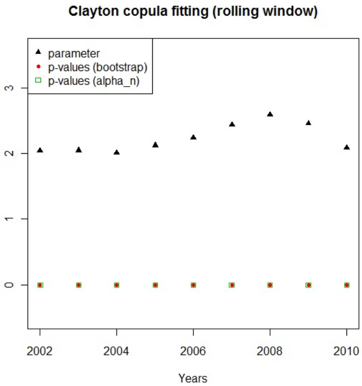


For some periods Gumbel can't be rejected

CvM test based on empirical copula process estimated with parametric bootstrap and with α_n



mostly rejected



always rejected

Testing correlation structure – multivariate Bartlett test

- Multivariate Bartlett test: $H_0 : \Sigma_{\alpha_n} = \Sigma$
- Critical values and teststatistics with simulation
- Analysed pairs: (0.1,0.11); (0.1,0.9); (0.89,0.9)
- Results (Gumbel copula):

Years	(0.1,0.11)	(0.1,0.9)	(0.89,0.9)
2010-2011	0.1263	0.0421	0.1003
2009-2011	0.0021	0.0000	0.0040
2008-2011	0.0001	0.0000	0.0026
2007-2011	0.0017	0.0009	0.0011
2006-2011	0.0091	0.0057	0.0153
2005-2011	0.0007	0.0002	0.0009
2004-2011	0.0001	0.0000	0.0000
2003-2011	0.0000	0.0000	0.0000
2002-2011	0.0000	0.0000	0.0000

Testing correlation structure – multivariate Bartlett test

- Multivariate Bartlett test: $H_0 : \Sigma_{\alpha_n} = \Sigma$
- Critical values and teststatistics with simulation
- Analysed pairs: (0.1,0.11); (0.1,0.9); (0.89,0.9)
- Results (Gumbel copula):

Years	(0.1,0.11)	(0.1,0.9)	(0.89,0.9)
2010-2011	0.1263	0.0421	0.1003
2009-2011	0.0021	0.0000	0.0040
2008-2011	0.0001	0.0000	0.0026
2007-2011	0.0017	0.0009	0.0011
2006-2011	0.0091	0.0057	0.0153
2005-2011	0.0007	0.0002	0.0009
2004-2011	0.0001	0.0000	0.0000
2003-2011	0.0000	0.0000	0.0000
2002-2011	0.0000	0.0000	0.0000

Final remarks

- Conclusions
  - K process based tests are the fastest but not the best
  - The null hypotesis was not rejected between 2010 and 2011
- Future work
  - CvM test based on empirical copula process estimated with pdm method
  - Working with Gaussian and t-copulas (problem: estimating the partial derivatives)

- Acknowledgement
  - András Zempléni
  - László Sajtos
  - Doctoral School of Mathematics of ELTE for supporting my participation at ASMDA2011 Conference
  - András Zempléni's work was supported by the European Union Social Fund (Grant Agreement No.TÁMOP 4.2.1/B-09/1/KMR-2010-0003)

Final remarks

- Conclusions
  - K process based tests are the fastest but not the best
  - The null hypotesis was not rejected between 2010 and 2011
- Future work
  - CvM test based on empirical copula process estimated with pdm method
  - Working with Gaussian and t-copulas (problem: estimating the partial derivatives)
- Acknowledgement

  - András Zempléni
  - László Sajtos
  - Doctoral School of Mathematics of ELTE for supporting my participation at ASMDA2011 Conference
  - András Zempléni's work was supported by the European Union Social Fund (Grant Agreement No.TÁMOP 4.2.1/B-09/1/KMR-2010-0003)

References I

S.N. Lahiri:
Resampling Methods for Dependent Data.
Springer, 2003.

P. Rakonczai, L. Varga, A. Zempléni:
Copula Fitting to Autocorrelated Data, with Applications to Wind Speed Modelling.
Working paper, november 11., 2010.

P. Rakonczai, A. Zempléni:
Copulas and goodness of fit tests. Recent advances in stochastic modeling and data analysis.
World Scientific, pp. 198-206, 2007.

References II

A. Bücher, H. Dette:
A note on bootstrap approximations for the empirical copula process.
German Research Foundation July 29., 2010.

C. Genest, B. Rémillard:
Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models.
Annales de lInstitut Henri Poincaré - Probabilités et Statistiques, Vol. 44, No. 6, 10961127, 2008.

C. Genest, J-F. Quessy, B. Rémillard:
Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation.
Scandinavian Journal of Statistics, Vol. 33, 337-366, 2006.

References III

C. Genest, B. Rémillard:
Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models.
Les Cahiers du GERAD, July 2005.

R.B.V. Silva, D.F. Ferreira, D.A. Nogueira:
Robustness of asymptotic and bootstrap tests for multivariate homogeneity of covariance matrices.
Cienc. agrotec., Lavras, Vol. 32, No. 1, 157-166, jan./febr. 2008.