A copula goodness-of-fit approach based on the conditional probability integral transform

Daniel Berg[1] Henrik Bakken[2]

[1]Norwegian Computing Center (NR) & University of Oslo (UiO)

[2]Norwegian University of Science and Technology (NTNU)

Workshop on quantitative risk management,
Norwegian Computing Center, Oslo, Norway, 24 April, 2007

Outline

- 1. Preliminaries
- 2. Copula goodness-of-fit (GoF) testing
  - 2.0 GoF approaches
  - 2.1 Proposed GoF approaches
  - 2.2 Cpit-approach
  - 2.3 Cpit2-approach - a generalization and extension
  - 2.4 Testing procedure
- 3. Size/power results from Monte Carlo simulations
- 4. Application to daily stock return data
- 5. Summary

1. Preliminaries

- Empirical Margins
  - Nuisance parameters
  - n samples of d-variate vector $\boldsymbol{X} = (X_1, \ldots, X_d)$
  - Transform $\boldsymbol{X}$ into pseudo-vector $\boldsymbol{Z}$ through empirical marginal df's:

$$\begin{aligned} \boldsymbol{Z}_j &= (Z_{j1}, \ldots, Z_{jd}) \\ &= (\widehat{F}_1(X_{j1}), \ldots, \widehat{F}_d(X_{jd})) \\ &= \left(\frac{R_{j1}}{n+1}, \ldots, \frac{R_{jd}}{n+1} \right), \; j = 1, \ldots, n, \end{aligned}$$

where R_{ji} is the rank of X_{ji} among $(X_{1i}, \ldots, X_{ni})$.

1. Preliminaries

- Anderson-Darling statistic
  - Random vector $W = (w_1, \ldots, w_n)$, iid $U(0,1)^d$ and cdf of W is $F(w) = w$.
  - AD statistic defined as:

$$\mathcal{T} = n \int \frac{\left\{\widehat{F}(w) - w\right\}^2}{w(1-w)} \mathrm{d}w, \quad w \in [0,1].$$

  - Strongly weights deviations close to $w = 0$ and $w = 1$.

1. Preliminaries

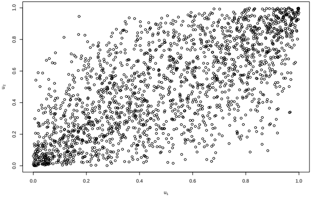
- Conditional probability integral transform
  - Introduced by Rosenblatt (1952)
  - D'Agostino and Stephens (1986): conditional probability integral transform (cpit)
  - Transforms dependent variables into independent variables, given the multivariate distribution
  - Cpit of $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ defined as $T(\boldsymbol{Z}) = (T_1(Z_1), \ldots, T_d(Z_d))$ where

$$
\begin{aligned}
T_1(Z_1) &= F_1(z_1), \\
T_2(Z_2) &= F_{2|1}(z_2|z_1), \\
&\vdots \\
T_d(Z_d) &= F_{d|1\ldots d-1}(z_d|z_1, \ldots, z_{d-1}),
\end{aligned}
$$

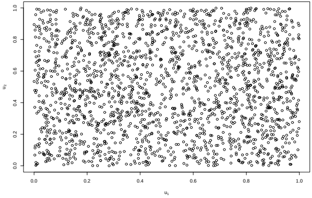
  - Rv's $V_i = T_i(Z_i),\ i = 1, \ldots, d$ iid $U(0,1)^d$.

1. Preliminaries

- ▷ Conditional probability integral transform



(a) Z



(b) V=T(Z)

$$\begin{matrix} z_{11} & \cdots & z_{1d} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nd} \end{matrix} \quad \overset{\text{cpit}}{\Longrightarrow} \quad \begin{matrix} v_{11} & \cdots & v_{1d} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nd} \end{matrix}$$

1. Preliminaries

- Parameter estimation
  - Semi-parametric method based on pseudo-vector $\boldsymbol{Z}$
  - Elliptical copulae - pairwise invert sample Kendall's tau
  - Student't t copula - numerically maximize likelihood wrt ν, given estimated scale matrix
  - Archimedean copulae: exchangeable construction with one parameter - estimate parameter by ML

2. Copula GoF testing

- Determine if a copula family appropriately fits data
- Univariate distributions ⇒ e.g. Anderson-Darling test or QQ-plot
- Multivariate domain ⇒ fewer alternatives
- Copula GoF complicated due to empirical margins and parameter estimation
- p-value estimates commonly found through parametric bootstrap procedures.

2.0 GoF approaches:

- Several approaches proposed lately:
  - Binning of the probability space
  - Multivariate smoothing
  - **Dimension reduction techniques**
  - Constructions to test specific copulae
  - **General tests for any copula**

2.1 Proposed GoF approaches:

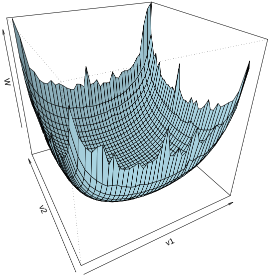
- **Breymann et al. (2003) - based on the cpit and tests of independence (cpit-approach)**
- Genest and Rémillard (2005) - based on the copula df ($\widehat{C} - C_{\widehat{\theta}}$)
- Savu and Trede (2004) and Genest et al. (2006) - based on the cdf of the copula df ($\widehat{K} - K_{\widehat{\theta}};\ K(t) = P[C \leq t]$)
- Genest et al. (2007) - based on the cpit and the copula df ($\widehat{C} - C_{\perp}$)
- Quessy et al. (2007) - based on Spearman's process ($\widehat{L} - L_{\widehat{\theta}};\ L(t) = P[\prod_{i=1}^{d} Z_i \leq t]$)

2.2 Cpit-approach

- Proposed by Breymann et al. (2003)
- Test procedure corrected by Dobrić and Schmid (2007)
- $\boldsymbol{Z}$ is the $U(0,1)^d$ pseudo-vector from empirical margin transformation of $\boldsymbol{X}$
- $\boldsymbol{V} = T(\boldsymbol{Z})$ is the cpit vector, iid $U(0,1)^d$ under the null
- Dimension reduction: $W_G = \sum_{i=1}^{d} \Phi^{-1}(V_i)^2$
- Test observator $G(w) = P[F_{\chi^2_d}(W_G) \leq w]$
- Under $\mathcal{H}_0$, $G(w) = w$ and $g(w) = 1$.
- Empirical version $\widehat{G}(w)$ plugged in for $\widehat{F}(w)$ in expression for AD statistic

2.2 Cpit-approach

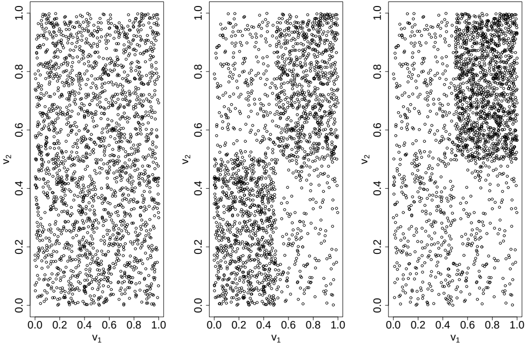
- Problem 1:
  - Extreme weight to corners and edges of unit hypercube
  - Low power for small sample sizes



- Solution 1 - generalization: allows for any weight function

2.2 Cpit-approach

- Problem 2:
  - Some deviations cancel out, $\widehat{\mathcal{T}}^{AD} = \{0.411, 0.411, 0.411\}$.



- Solution 2 - extension: additional cpit based on order statistics detects radial asymmetry

2.3 Cpit2-approach: a generalization and extension

- $\boldsymbol{V} = T(\boldsymbol{Z})$ is the cpit vector, iid $U(0,1)^d$ under the null
- Denote order statistics of $\boldsymbol{V}$ by $V_{(1)} \leq V_{(2)} \leq \ldots \leq V_{(d)}$
- Known results from Deheuvels (1984) and iid uniformity of $\boldsymbol{V}$ gives for the additional, order statistic cpit:

$$H_i = F_{V_{(i)}|V_{(i-1)}}(v_{(i)}) = 1 - \left(\frac{1 - v_{(i)}}{1 - v_{(i-1)}} \right)^{d-(i-1)}, \ i = 1, \ldots, d, \ v_{(0)} = 0.$$

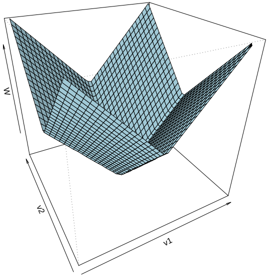
- Poor fit in d-space is indicated by extreme values of H.

2.3 Cpit2-approach: a generalization and extension

- Dimension reduction: $W_B = \sum_{i=1}^{d} \Gamma_V(V_{(i)}; \boldsymbol{\alpha}) \cdot \Gamma_H(H_i; \boldsymbol{\alpha})$
- Γ_V and Γ_H are weight functions and $\boldsymbol{H}$ respectively and $\boldsymbol{\alpha}$ is the set of weight parameters.
- Any weight functions can be used, for example:
  - (i) $\Phi^{-1}(X)^2$
  - (ii) $|X - 0.5|$
  - (iii) $(X - 0.5)^{\alpha}, \ \alpha = (2, 4, \ldots)$
- Special case (cpit-approach): $\Gamma_V(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2$ and $\Gamma_H(X; \boldsymbol{\alpha}) = 1$

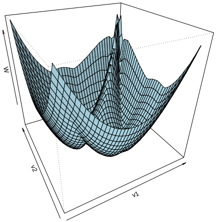
2.3 Cpit2-approach: a generalization and extension

- Solution to problem 1: Any weight can now be used
- Example 1: $\Gamma_V(X;\alpha) = |X - 0.5|$ and $\Gamma_H(X;\alpha) = 1$



2.3 Cpit2-approach: a generalization and extension

- Solution to problem 1: Any weight can now be used
- Example 2: $\Gamma_V(X;\alpha) = |X - 0.5|$ and $\Gamma_H(X;\alpha) = |X - 0.5|$



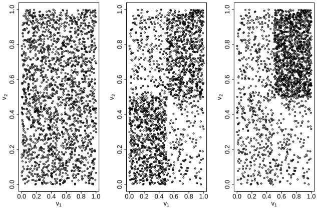
- Note the Γ_H term emerging as weight to the diagonal, detecting radial asymmetry

2.3 Cpit2-approach: a generalization and extension

- Solution to problem 2: Radial asymmetry is detected by the Γ_H term:

$$\left.\begin{array}{l}\Gamma_V(X;\alpha) = \Phi^{-1}(X)^2 \\ \Gamma_H(X;\alpha) = 1\end{array}\right\} \widehat{\mathcal{T}}^{AD} = \{0.411, 0.411, 0.411\}$$

$$\left.\begin{array}{l}\Gamma_V(X;\alpha) = 1 \\ \Gamma_H(X;\alpha) = \Phi^{-1}(X)^2\end{array}\right\} \widehat{\mathcal{T}}^{AD} = \{0.409, 65.455, 109.523\}$$



2.4 Testing procedure

- Assume that W_G is χ^2_d distributed
- Only close to, but not exactly true
- Empirical margins and parameter estimation introduce dependence between observations
- Parametric bootstrap procedures are needed
- Asymptotics of our bootstrap procedure are not proven

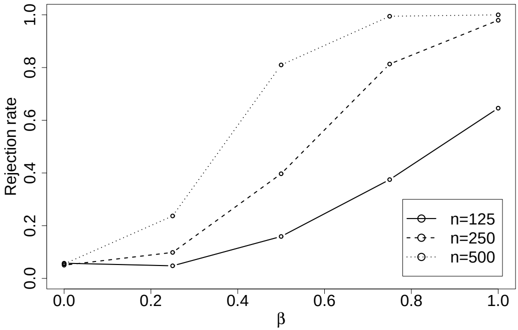
3. Monte Carlo study

Setup:

- Mixing tests: $C^{Mix} = (1-\beta)\cdot C^{Ga} + \beta \cdot C^{Alt}$
- Null hypothesis copula: Gaussian copula
- Alternative copulae: Student's t, Clayton, Gumbel
- Dependency parameter: $\rho_\tau = 0.20$, $\nu = 4$
- $d = \{2, 5\}$, $n = \{125, 250, 500\}$
- All combinations of (i)-(iii) for Γ_V and Γ_H
- Repeat 2000 times to obtain rejection rates (5% significance level)

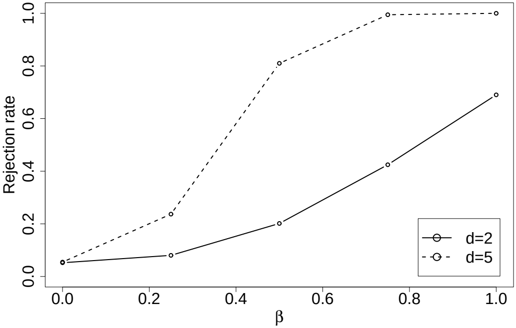
3. Monte Carlo results

Effect of sample size:
Cpit2-approach, $d = 5$, $\Gamma_V(X;\alpha) = |X - 0.5|$, $\Gamma_H(X;\alpha) = 1$



3. Monte Carlo results

Effect of copula dimension:
Cpit2-approach, $n = 500$, $\Gamma_V(X;\alpha) = |X - 0.5|$, $\Gamma_H(X;\alpha) = 1$



3. Monte Carlo results

General size/power conclusions:

- Nominal levels kept, indicating validity of bootstrap procedure
- Cpit-approach performs rather poor, particularly for small sample sizes
- Best performing cpit2-approach combination: $\Gamma_V(X;\alpha) = |X - 0.5|$, $\Gamma_H(X;\alpha) = 1$
- Γ_H term adds power for high dimension ($d = 5$), small sample size ($n = 125$) and skewed alternative copula (Clayton, Gumbel)

4. Application

Setup:

- 1000 samples (daily log-returns) of 45 large cap stocks from NYSE
- Collections of 2 and 5 stocks randomly selected
- Gaussian, Student's t, Clayton and Gumbel copulae fitted
- Repeated 2000 times to obtain rejection rates
- Raw returns and GARCH(1,1) filtered returns

4. Application

Results:

- ▷ Exchangeable Clayton- and Gumbel copulae strongly rejected
- ▷ Gaussian copula rejected in many cases
- ▷ Student's t copula rarely rejected
- ▷ Rejection rates lower for filtered returns

5. Summary

- We generalize and extend the cpit-approach
- Cpit2-approach solve two issues with the cpit-approach
- Radial assymmetry detected by Γ_H term
- Combination $\Gamma_V(X;\alpha) = |X - 0.5|,\ \Gamma_H(X;\alpha) = 1$ best, good power for small sample size
- Nominal levels kept - bootstrap procedure valid
- Applied to daily log-returns the Student's t copula is superior

5 Further work

- Can we further generalize to base the dimension reduction on the original data, $\Gamma_Z(Z_i)$?
- This way we can weight any region of the original copula
- Compare with other proposed approaches (work in progress)
- Examine local power and sensitivity of various copula GoF approaches (work in progress)

References

Breymann, W., A. Dias, and P. Embrechts (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance 1*, 1–14.

D'Agostino, R. B. and M. A. Stephens (1986). *Goodness-of-Fit Techniques*. New York: Marcel Dekker Inc.

Deheuvels, P. (1984). The characterization of distributions by order statistics and record values: A unified approach. *Annals of Eugenics 21*(2), 326–334.

Dobrić, J. and F. Schmid (2007). A goodness of fit test for copulas based on rosenblatt's transformation. Forthcoming in Computational Statistics and Data Analysis.

Genest, C., J.-F. Quessy, and B. Rémillard (2006). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian Journal of Statistics 33*.

Genest, C. and B. Rémillard (2005). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Les Cahiers du GERAD G-2005-51*.

Genest, C., B. Rémillard, and D. Beaudoin (2007). Omnibus goodness-of-fit tests for copulas: A review and a power study. *To appear in Insurance: Mathematics and Economics*.

Quessy, J.-F., M. Mesfioui, and M.-H. Toupin (2007). A goodness-of-fit test based on spearmans dependence function. *Working paper*.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*, 470–472.

Savu, C. and M. Trede (2004). Goodness-of-fit tests for parametric families of archimedean copulas. CAWM discussion paper, No. 6.

Contact details

Daniel Berg,
Norwegian Computing Center and University of Oslo,
daniel@danielberg.no
www.danielberg.no