Price fluctuations: Lecture 8

Zempléni, András

Department of Probability Theory
Faculty of Sciences
ELTE

Dependence measures for copulas

- Linear correlation: $R(X,Y) = \frac{E(X-EX)(Y-EY)}{D(X)D(Y)}$
- disadvantages:
  - It is sensitive to outliers
  - It changes if we transform the marginals
- Alternatíves: Kendall-τ:

$$\tau(X,Y) = P\left[(X-\tilde{X})(Y-\tilde{Y}) > 0\right] - P\left[(X-\tilde{X})(Y-\tilde{Y}) < 0\right].$$

Spearman-ρ:

$$\begin{aligned}\rho(X,Y) = & \quad 3P\left[(X-\tilde{X})(Y-Y') > 0\right] \\ - & \quad 3P\left[(X-\tilde{X})(Y-Y') < 0\right]\end{aligned}$$

where $(X,Y),\ (\tilde{X},\tilde{Y}),\ (X',Y')$ are independent, identically distributed

Properties

- These are so-called rank correlations (just the sequence of the values is interesting)
- These are not sensitive to outliers
- Their computation with the copula

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1$$

$$\rho(X, Y) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] \, du dv.$$

Further properties

- Both are invariant for monotonic transformations Let $\kappa = \rho$ or $\kappa = \tau$. Then
  - $-1 \leq \kappa \leq 1;\ \kappa_{X,X} = 1,\ \kappa_{X,-X} = -1.$
  - If X and Y are independent, then $\kappa_{X,Y} = 0$.
  - $\kappa_{X,-Y} = \kappa_{-X,Y} = -\kappa_{X,Y}.$
- The dependence measures for the copulas are functions of the parameter(s), so from their estimators we may get at the same time an estimator for the copula. Example for the Gumbel copula: $\tau = 1 - 1/\beta$.

Applications

- For the Gauss copula the pairwise correlations:

$$R_{ij} = \sin\left(\pi\tau(X_i, X_j)/2)\right)$$

- The choice between copula-types is important (e.g. by the tail dependence or by theoretical considerations). It is an empirical fact that for financial portfolios the extreme losses appear for every stock at the same time (stock exchange fall) - this implies the probable appearance of the tail dependence.
- The different models may rise to substantial differences between the estimators for the probabilities.

Time dependence

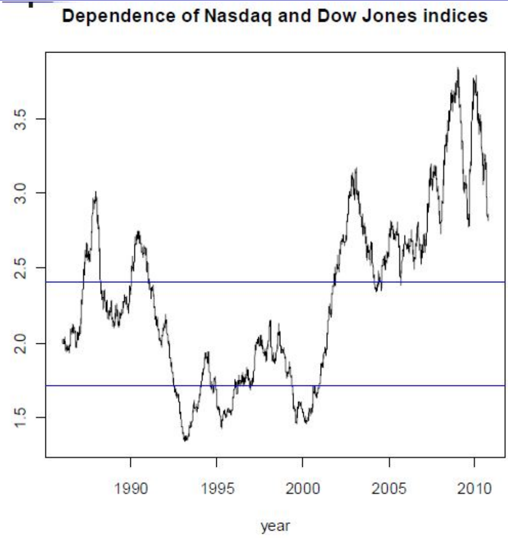


Figure: Time series of the estimated dependence parameter

- The figure shows the dependence parameter of the fitted Gumbel copula, based on windows of 251 days (one year).
- The blue lines represent the 0.003 and 0.997 bootstrap quantiles for the parameter of the first year.
- The dependence has obviously strengthened due to the financial crisis.

Goodness-of-fit

- In order to reduce the computational burden, the dimensionality has to be reduced. The K-function:

$$K(\vartheta, t) = P(F(\underline{X}) < t) = P\left(C_\vartheta(F_1(X_1), \ldots, F_d(X_d)) < t\right)$$

- It can be computed for Archimedean copulae as

$$K(\vartheta, t) = t + \sum_{i=1}^{d-1} \frac{(-1)^j}{i!} \left[\varphi_\vartheta(t)^j\right] f_i(\vartheta, t)$$

where

$$f_i(\vartheta, t) = \frac{d^i}{dx^i} \varphi_\vartheta(x)|_{x=\varphi_\vartheta(t)}.$$

- If it has no closed form, it can still be approximated by simulations

The test based on the K function

- Empirical version:

$$K_n(t) = \frac{1}{n}\sum_{j=1}^{n} \chi(E_j < t) \quad t \in [0,1]$$

ahol

$$E_j = \frac{1}{n}\sum_{i=1}^{n} \chi\left(U_{j,1} < U_{i,1}, \ldots, U_{j,d} < U_{i,d}\right)$$

Kendall process $\kappa_n(t) = \sqrt{n}\,(K(\vartheta_n, t) - K_n(t))$.

- Cramér-von Mises type statistics:

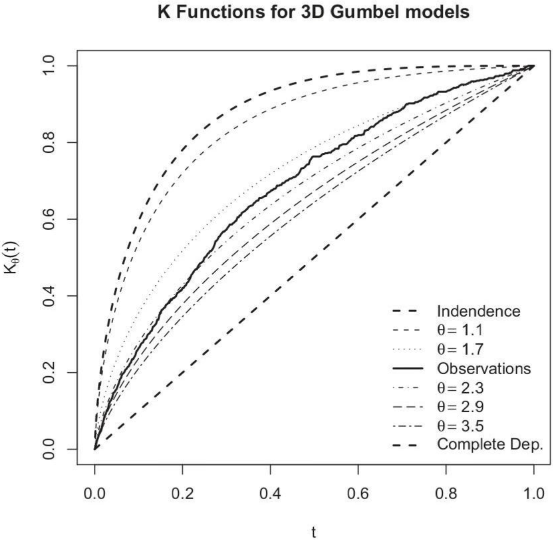
$$S_n = \int_0^1 (\kappa_n(t))^2 \Phi(t)\,dt$$

where Φ is the weight function

The test

- Formal test can also be got from the statistics S_n (if it is large, we reject the fit).
- The asymptotic distribution can be computed only in case of known copulae.
- In those realistic case, where C is estimated, we may get the critical values via simulations

Comparison of copulae



Rosenblatt-transform

- The test of Breyman (Breymann et al, Berg & Bakken) is based on the Rosenblatt transform $\mathcal{R} : (0,1)^d \to (0,1)^d$
$\mathcal{R}(\underline{u}) = (e_1, ..., e_d)$, where $e_1 = u_1$ and for $i \geq 2$

$$e_i = \frac{\partial^{i-1}\mathbf{C}(u_1, ..., u_i, 1, 1, ...1)}{\partial u_1 ... \partial u_{i-1}} \Big/ \frac{\partial^{i-1}\mathbf{C}(u_1, ..., u_{i-1}, 1, 1, ...1)}{\partial u_1 ... \partial u_{i-1}}.$$

- Property: the distribution of U is the C copula, if and only $R(U)$ is the independence copula.

Breymann-test: testing independence

- $Y_B = \sum_{i=1}^{d} \Phi^{-1}(E_i)^2$ has just the chi-squared distribution, with degree of freedom d.
- If it is substituted into its own distribution function, we get the uniform distribution.
- It can be tested e.g. by the Anderson-Darling test.
- Berg and Bakken developed the method further, achieving its consistency.

New test-procedure, based on the weighted bootstrap

- Bootstrap simulations are needed to determine the critical values for the goodness of fit tests
- But: the model has to be fitted for every bootstrap sample, which is very slow in high dimensions
- The difference between the empirical copula and the fitted parametric model is a natural statistics. Its limit distribution:

$$\sqrt{n}(C_n - C_{\vartheta_n}) = \sqrt{n}(C_n - C_\vartheta + C_\vartheta - C_{\vartheta_n}) \to \mathbf{C}_\vartheta - \mathbf{C}_{*\vartheta}$$

- Because of the limit theorem for the weighted bootstrap sample, it can be approximated without re-estimating the parameter(s).

Steps of the test procedure

- To calculate C_n and the determination of a suitable estimator for ϑ
- The calculation of the Cramer- von Mises statistics:

$$\int_{[0,1]} \left(C_n(u,v) - C_{\vartheta_n(u,v)}\right)^2 dC_n(u,v) =$$
$$\sum_{i=1}^{n} \left(C_n(U_{i,n}, V_{i,n}) - C_{\vartheta_n}(U_{i,n}, V_{i,n})\right)^2$$

- The calculation of the weighted bootstrap statistics
- These allow for estimation of the critical values (or the p-value)
- The method is much faster than the parametric bootstrap

Practical experiences

- The strength of the method depends on the estimation: the maximum-pseudo likelihood gives good results in general
- It can be calculated quickly in 3-5 dimensions
- It is already contained in the package copula

Fitting multivariate models

- We have learned
  - Multivariate stable distributions
  - Multivariate extreme-value distributions
  - Multivariate Pareto distributions
  - Copulas
- So it looks as we would only need to choose from the models
- However, bootstrap simulations are needed to determine the critical values for the goodness of fit tests
- But: the model has to be fitted for every bootstrap sample, which is very slow in general in high dimensions
- We shall see ideas for solution

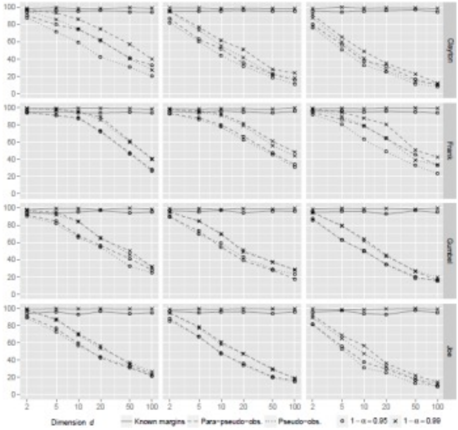
Models

- The structures are nonparametric, but to the estimations we need models
- Parameter estimation: e.g. by maximum likelihood method
- Number of parameters:
  - If too few (e.g. we have only one parameter for a typical Archimedean copula), in general the fit is not good enough
  - If too many (e.g. one to each pairs in case of a Gauss copula) then the estimators will not be reliable

Practical experiences in high dimensions

- The inference is usually based on the pseudolikelihood (here the margins are estimated nonparametrically)
- These are neither independent nor a sample with uniform distribution
- The error of the estimators (in case of single parameter Archimedean copulas) is decreasing if the dimension increases

Confidence intervals



- The coverage probability of the ML-based confidence interval for the copula parameter as a function of the dimensions
- It is OK for known margins, but in case of unknown marginals it dramatically decreases for high dimensions

Even higher

- The methods seen previously can be realized in the range of 2-4 dimensions
- For even higher dimensions the main problem is the lack of the needed sample size to a reliable analysis (it should grow exponentially with the dimensions)
- Some simplifications are needed:
  - Rarity conditions (Lasso and its versions), but it is not realistic for most of real data (e.g. at direct estimation of the covariance matrix)
  - Dimension reduction

LASSO method

- LASSO=least absolute shrinkage and selection operator
- Method for L_1-regularization: it optimizes under the condition of the L_1 norm of the coefficient vector (portfolio weights)
- Typical realization: addition of $\lambda \sum_{i=1}^{N} |\vartheta_i|$ to the target function to be minimized (e.g. in the least square method)
- λ determines the strength of the regularization, its choice is not trivial
- Application: portfolio optimization (typically we get less variable weights)

Estimation method in high dimensions: pairwise likelihood

- Its definition:

$$\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} f_2(x_i, x_j; \vartheta)$$

- So only the pairwise dependence is to be taken into account in the model
- It is easier to calculate (important in really high dimensions)
- Example: spatial model (point process)

$$\hat{\vartheta} = \operatorname{argmax} \sum_{t=1}^{T-\max k} \sum_{k \in K} \sum_{s_1=1}^{s} \sum_{s_2=1}^{s} \log f_2(z_{s_1,t}, z_{s_2,t+k}; \vartheta)$$

where z is the observed value, s denotes the sites, t the time and K is a chosen index set (e.g. 0,1,2,4,8,..)

Copulas in high dimension: vine copulas

- These structures may be used in high dimensions as well
- They are based on bivariate copulas
- The further structure is determined by a graph
- The density function of the copula: $c_{12}(x, y) = \frac{\partial^2 C(x,y)}{\partial x \partial y}$
- Thus the density function of the original distribution: $c_{12}(F_1(x), F_2(y))f_1(x)f_2(y)$
- Conditional density: $f(x|y) = c_{12}(F_1(x), F_2(y))f_1(x)$

Construction based on pair-copulas

- In 3 dimensions:

$$\begin{aligned} f(x_1, x_2, x_3) = & \quad f_{1|23}(x_1|x_2, x_3) f_{2|3}(x_2|x_3) f_3(x_3) \\ = & \quad c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3); x_3) c_{13}(F(x_1), F(x_3)) \\ \times & \quad c_{23}(F(x_2), F(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned}$$

- The decomposition is not unique: there are 3 decompositions in 3 dimensions, but in 5 dimensions there are already 240
- Simplification: we omit the dependence of the conditional copulas on the variables in the condition:

$$\begin{aligned} f(x_1, x_2, x_3) = & \quad f_{1|23}(x_1|x_2, x_3) f_{2|3}(x_2|x_3) f_3(x_3) \\ = & \quad c_{12|3}(F_1(x_1), F_2(x_2)) c_{13}(F(x_1), F(x_3)) \\ \times & \quad c_{23}(F(x_2), F(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned}$$

The graphs of vine copulas

- The description by pair-copulas can be characterized by graphs (trees)
- Property: there are $d - 1$ graphs for a d-dimensional vine
- T_1 is a tree over $1, \ldots, d$
- The vertices of the next graph are the edges of the previous one
- If there is an edge between two vertices in T_{j+1}, then the corresponding edges had a common vertex in the previous graph T_j
- The tree T_j has $d + 1 - j$ vertices and $d - j$ edges

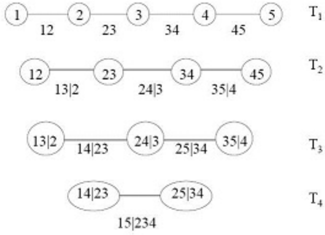


Figure: A graph of a 5-dimensional D-vine

Types and practical applications of the vine copulas

- C-vine: the graphs are star-shaped
- D-vine: the graphs are paths
- Estimation in practice, e.g. by the Kendall-τ: the most important pairs are estimated separately, then the others together - universally by the same copula (this is the so-called simplification)

Statistical methods

- Parameter estimation: by maximum likelihood, iteratively for the levels of the graph, first for the copulas of the first level
- How to choose the pair-copulas? By the previous tests the fit can be investigated
- Having estimated the copulas of the first level, the same may be carried out for the next level (after transforming the data)
- The iteration is continued until the remaining levels can be simplified as it was mentioned before

Practical applications

- It was possible to fit the whole model for a 16 dimensional data set
- In the first step the spanning tree is sought for which the sum of the Kendall-τ values over the edges is maximal
- Truncation: we assume every copula beyond a given level being independent
- Simplification: we assume every copula beyond a given level being identical

Remarks

- If the bivariate copulae are *t*-copulae, then the vine-copula is a submodel of the full *d*-dimensional *t*-copula
- The choice from nested models can be based on the loglikelihood
- For non nested models the so-called Vuong teststatistics can be applied, which is als based on the loglikelihood function and has information theoretical background (R package: CDVine)

Further possibilities for goodness-of-fit

- Information matrix-proportion tests
- White-type misspecification test
- The tests seen previously (K-funcion-based, Rosenblatt transforms) can be generalised – critical values can be based on the weighted bootstrap; here also the Cramér-von Mises type tests are the strongest

References

- Berg, D. (2009) Copula Goodness-of-fit testing: An overview and power comparison.
- Berg, D. and Bakken, H. (2006) Copula Goodness-of-fit Tests: A Comparative Study.
- C. Czado (2019) Analyzing Dependent Data with Vine Copulas.
- Embrechts-Hofer: Statistical inference for copulas in high dimension: a simulation study (2013)
- Varin-Reid-Firth: An overview of composite likelihood methods (2011)
- J. Dißmann, E. C. Brechmann, C. Czado, D. Kurowicka: Selecting and estimating regular vine copulae ... 2012.
- Schepsmeier: Efficient goodness-of-fit tests in multi-dimensional vine copula models (2013).
- Kojadinovic,I., Yan,J. and Holmes,M.: Fast large-sample goodness-of-fit tests for copulas. 2011.