Price fluctuations: Lecture 9

Zempléni, András

Department of Probability Theory
Faculty of Sciences
ELTE

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 $1, \dots, d$
- 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

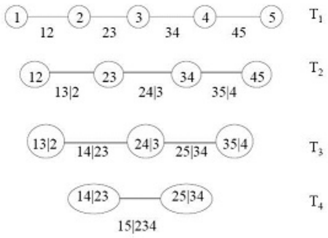


Figure: A graph of a 5-dimensional D-vine

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

Goodness of fit

- 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)
- 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.

Random matrices: main questions

- $X : \Omega \to M_{N \times N}$ is a measurable transformation (considering the set of matrices as a Euclidean space)
- N is large, so the limit $N \to \infty$ is interesting
- Example
  - Let X_{ij} be symmetric ($X_{ij} = X_{ji}$), independent, identically, normally distributed with mean 0 (Wigner matrix)
  - $X^T X / T$ (X is a matrix of size $N \times T$): Wishart matrix, this corresponds to the covariance matrix estimator (it is of size $N \times N$)
- Eigenvalues are random variables as well. It is not easy to determine them one by one, but the spectrum (the set of all eigenvalues) can be investigated

Limit theorem

- Let X be an $N \times N$ symmetric matrix having elements, which are independent and have standard normal distribution (Wigner matrix). The spectrum of X has N elements: $\lambda_i (i = 1, \ldots, N)$.
- The weak limit of the spectrum in case of the Wigner matrix is the so-called Wigner (semicircle) distribution (Wigner, 1950)

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i / (2\sqrt{N})} = \nu$$

where the density of ν is $\frac{2}{\pi}\sqrt{1 - \lambda^2}$ for $-1 < \lambda < 1$.

The idea of the proof

- Let

$$\nu_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\frac{\lambda_i}{2\sqrt{N}}}, \quad Y_{Nk} = \int x^k d\nu_N$$

  random variable (ν_N is a random measure)
- It can be seen that $E(Y_{Nk}) \to \int x^k d\nu$ where ν is the Wigner-distribution and $Var(Y_{Nk}) \leq \frac{c_k}{N^2}$
- The Borel Cantelli lemma implies that

$$P\left(|Y_{Nk} - E(Y_{Nk})| > \frac{1}{N^{1/4}} \text{ i.o.} \right) = 0$$

- Thus the sequence ν_N is tight, the moments of the limit for any of its subsequence are unique

Carleman condition

- If
$$\sum_{k=1}^{\infty} \frac{1}{\mu_k^{1/2k}} = \infty$$
there can be at most one random variable, which has moments μ_k.
- Proof: in this case the characteristic function is given by its Taylor expansion
- This condition holds for the Wigner distribution
- Generalizations
  - The proof can be carried out for non normally distributed matrix elements as well
  - If the elements of the matrix are stable distributions with infinite variance, then the limit distribution of the spectrum is not bounded (it has a tail of polynomial order)

Wishart matrix

- Let X be an $N \times T$ matrix, with elements that are
  - independent, identically distributed
  - with expected value 0
  - and variance 1
  - $\kappa = E(X_{ij}^4)$
- Then for $W = X^T X / T$ we have
  - $E(W_{ij}) = 0$, if $i \neq j$ and 1, if $i = j$ (notation: δ_{ij})
  - $Var(W_{ij}) = (1 + (\kappa - 2)\delta_{ij})/T$

Degenerate limit distribution

- $Var(W_{ij}) \to 0$, if N is constant, $T \to \infty$ and so $W_{ij} \to I$.
- This remains true if $T \to \infty$ and $N \to \infty$ so that $N/T \to 0$
- If $T < N$, W_{ij} is degenerate, the weight of the 0 eigenvalue is $(N - T)/N$. Thus if $N \to \infty$ so that T is constant, then the spectrum tends to 0 (enough: $N/T \to \infty$).
- Thus nontrivial limit distribution can only be expected if $N/T \to c$

Marchenko-Pastur theorem (1967)

Theorem:

- Let X be a random matrix of size $N \times T$ ($T > N$), having independent, identically distributed elements with expected value 0 and variance 1,
- $W = X^T X / T$ is the corresponding Wishart-matrix.
- Let the spectral density of W be denoted by $\rho_{N;T}(\lambda)$. Then if $N \to \infty$ and $N/T = r \leq 1$

$$\frac{1}{N}\rho_{N;T}(\lambda) \to \frac{1}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{r\lambda} I_{\{\lambda_- < \lambda < \lambda_+\}}$$

  where $\lambda_\pm = (1 \pm \sqrt{r})^2$. This is the Marchenko-Pastur distribution
- Thus similarly to the Wigner-matrices, the spectrum of the Wishart-matrices also tends to a deterministic function.

Properties

- The density is concentrated to the interval $[\lambda_-, \lambda_+]$
- In case of $r \to 0$
$$\lambda_\pm = (1 \pm \sqrt{r})^2 \to 1$$
so it tends to the degenerate distribution concentrated to 1(in accordance to the previous results).
- If $r > 1$ then the limit is the mixture of the degenerate at 0 and the distribution given in the above theorem
- If $r \to 1$ $(T \to N + 0)$ then
$$\lambda_- = (1 - \sqrt{r})^2 \to 0$$
as if $N > T$ then the matrix has not full rank, so 0 eigenvalues also turn up.

Applications to the covariance estimator

- The estimator:

$$\hat{\sigma}_{ij} = \frac{1}{T-1}\sum_{t=1}^{T} Z_{it}Z_{jt}$$

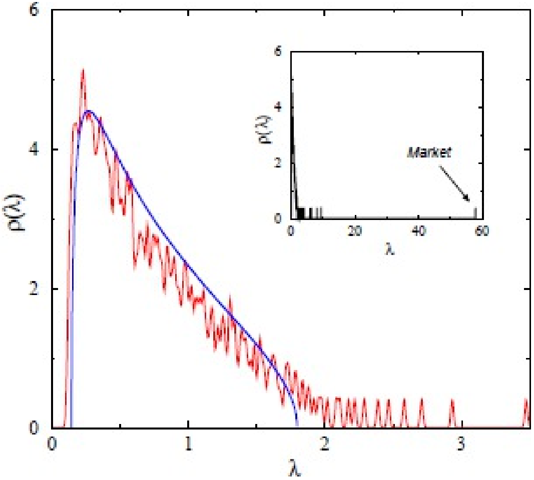
where

$$Z_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}$$

- The asymptotics of the Marchenko- Pastur theorem can be applied here
- But in practice the elements of the matrix are not independent!

Real data

- Empirical spectrum for the 406 stocks in the S& P 500 index and the fit of MP distribution
- The largest eigenvalue is seen on the separate figure
- After having removed the intermediate eigenvalues (describing the sectoral effects) there is still systematic deviation from the MP distribution



Risk measures

- Possibilities:
- *sd* (standard deviation, possibly only for the losses)
- VaR (a given high quantile of the loss)
- cVaR (expected loss, if we loose more than the VaR)
- mVaR (modified formula, the higher moments are also used)

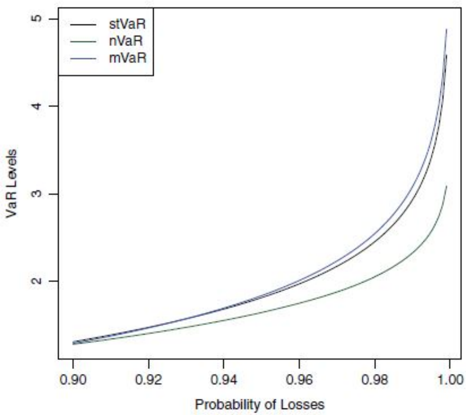


Figure: mVaR, actual VaR and VaR based on Gaussian dist., when the actual loss has a skewed t-dist.

Properties of a coherent risk measure

- Additivity: $R(X + c) = R(X) + c$
- Homogeneity: $R(aX) = aR(X)$
- Monotonicity: if $X \leq Y$ then $R(X) \leq R(Y)$
- Subadditivity (convexity):
  $R(wX + (1 - w)Y) \leq wR(X) + (1 - w)R(Y)$
  Shows that the diversification is preferable!
- VaR does not fulfill this. Example: $P(Z = 1) = 0.91$,
  $P(Z = 90) = 0.04$, $P(Z = 100) = 0.01$, $P(Z = 200) = 0.01$.
  $Z = X + Y$: $X = \{Z : Z < 100\}$, $Y = \{Z : Z \geq 100\}$
  $VaR_{0.95}(X) = 1$, $VaR_{0.95}(Y) = 0$, but $VaR_{0.95}(Z) = 90$

Properties of a coherent risk measure/2

- If the loss-distribution is elliptic, then VaR is already convex
- It can be estimated: in case of VaR several nonparametric methods can also be applied
  - By empirical quantile (not robust)
  - Weighted average of quantiles
- It is important that the authorities should accept the chosen method

CVaR

- Possible definitions:

$$C_{\alpha-} = E(X|X \geq VaR_\alpha)$$

$$C_{\alpha+} = E(X|X > VaR_\alpha)$$

- These are identical for continuous distributions
- Properties:

$$C_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\beta d\beta = \min_C \left\{\frac{1}{C} + \frac{1}{\alpha}E(X-C)^+\right\}$$

- Further properties of the CVaR:
  - It is a coherent risk measure
  - It is closed: for $X_n \to X$ and $R(X_n) \leq 0$ we have $R(X) \leq 0$
  - On the other hand, it is not easy to estimate it if we do not have a reliable parametric model

Comparing the models

- The estimated risk measure depends on the chosen model
- Example: daily logreturn of stocks, the value is 10,000; annual volatility is 20% (normal vs t_4-this is a heavy tailed distribution)

α	0.9	0.95	0.975	0.99	0.995
VaR_α (norm. dist.)	162.1	208.1	247.9	294.3	325.8
VaR_α (t_4 dist.)	137.1	190.7	248.3	335.1	411.8
$cVaR_\alpha$ (norm. dist.)	222.0	260.9	295.7	337.2	365.8
$cVaR_\alpha$ (t_4 dist.)	223.4	286.3	356.7	465.8	563.5

Portfolio-optimization

- Aim: to reach the largest return
- But: the risk has to be as small as possible
- Compromise: the expected value should reach a given value
- Approximation: we may have any real number from a stock (negative values corresponds to short selling)
- We assume that the market is liquid, the prices are not influenced by our trade
- We assume that there are no transaction costs

Portfolio-optimization/2

- Let us assume that we intend to find the minimal variance portfolio:

$$\min_{w\in\mathbb{R}^N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}w_iw_j$$

- The condition: $\sum_{i=1}^N w_i = 1$
- The solution:

$$w_i^* = \frac{\sum_{j=1}^N \sigma_{ij}^{-1}}{\sum_{j=1}^N \sum_{k=1}^N \sigma_{jk}^{-1}}$$

- Markowitz-problem:
  - To find the portfolio to the return μ with the minimal risk: $\min_{w\in\Pi} R(w)$
  - The conditions: $\sum_{i=1}^N E(w_iY_i) = \mu$, $\sum_{i=1}^N w_i = 1$

Alternatives

- Efficient mean-variance portfolios are to be found on the hyperbola
- A riskless asset may also be held (with expected return μ_r)
- More general utility functions may be used
- The Sharpe-ratio may also be maximized: $E(\mu_p - \mu_r)/\sigma_p$, where μ_p is the portfolio return and σ_p is the portfolio standard deviation

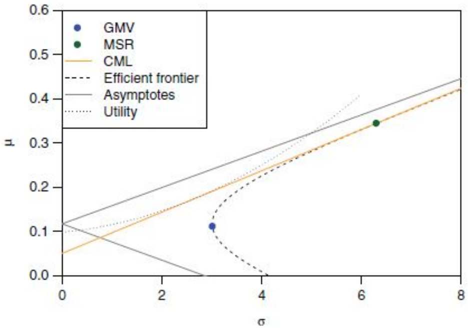


Figure: The variance-optimal portfolio and the one with the maximal Sharpe-ratio.

Problems in real life

- The return distribution is unknown, even its parameters (expectation, standard deviation) are not known
- The parameter estimations differ from their theoretical values, thus the resulted optimum is different from the real one
- The weights have random error
- The risk will be higher than the risk of the real optimum
- Question: how large is this difference?
- If we have an idea about the distributions, we may simulate data

Simulation

- We simulated a dataset of length T
- The estimated portfolio, based on the simulation: $\hat{w}$. (The real optimum: w.) Measure:

$$q_0 = \frac{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \hat{w}_i \hat{w}_j \sigma_{ij}}}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}}}$$

  the squared root of the ratio of the estimated and real optimums
- $q_0 \geq 1$, it shows the increase of risk due to the estimation error
- q_0 is a random variable, so its expectation, standard deviation is important

The properties of q_0

- Let N/T be constant and $N \to \infty$. Then $sd(q_0) \to 0$, $q_0 \to E(q_0)$.

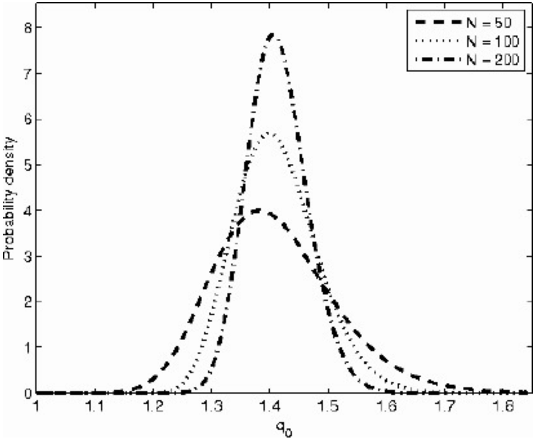


Figure: The distribution of q_0 for different N values. $N/T = 0.5$

The properties of q_0

- If N is constant and T is decreasing, then the expectation and standard deviation of q_0 both are increasing
- If $N/T > 1$, there is no solution, since the covariance matrix cannot be inverted
- If $N < T$, the problem can be solved, but if $N/T \to 1$ then $q_0 \to \infty$
- If N is constant and $T \to \infty$, then $q_0 \sim (1 - N/T)^{-1/2}$, independent of the expectations and standard deviations

The properties of q_0

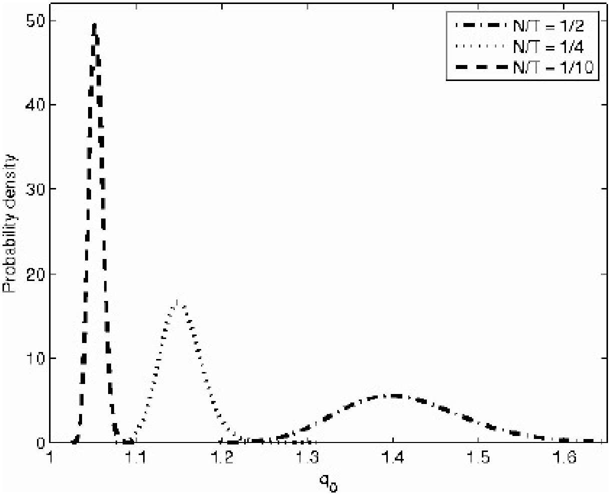


Figure: q_0 as a function of N/T, $N = 100$

Practical applications

- Thus if N is large enough, ($N > 100$ is enough in general), then the error can be estimated
- $q_0 \sim (1 - N/T)^{-1/2}$
- Thus the needed sample size can be calculated, to a given error: $T = N/(1 - 1/q_0^2)$
- E.g. if $N = 100$ and $q_0 = 1.2$ then $T = 328$, it is increasing linearly in N
- But if we can tolerate only a smaller error, or if there are more stocks, then one may need a much longer data sequence
- On the other hand there is no stationarity for such a long time period
- Thus the portfolio optimization shows a substantial instability even under the used idealistic conditions
- If we consider not only the risk, but the weights themselves, then the situation is even worse, the fluctuations are well over 100 percent. So this problem is almost hopeless.

Practical observations

- The empirical covariance matrix has typically one large eigenvalue
- The basis portfolio, corresponding to it has mostly positive weights
- Thus this is the joint fluctuation of the economy
- Theoretical background: Frobenius-Perron theorem. It states that a positive definite matrix consisting only positive elements has a largest positive eigenvalue which has multiplicity one and for which the corresponding eigenvector has positive elements
- Although the empirical covariance matrix has not always only positive elements, most of them are positive, and the theorem is true in this case, too
- The medium eigenvalues, which represent 90-95% of the total variance, correspond to the sectoral effects
- They bear substantial information
- The rest (typically 90-95% of the eigenvalues) is the noise
- Filtering is needed in order to suppress this noise, especially as the eigenvalues of the inverse matrix are the reciprocal of the original eigenvalues

Further eigenvalues

- Let us search for the lower dimensional subspace that is responsible for most of the variability
- If it is k-dimensional, then its basis consists of the the eigenvectors corresponding to the k largest eigenvalues
- The number of principal components can be determined by fitting a Marchenko-Pastur distribution to the empirical spectrum and to use those, which are outside the accepted range
- Instead of the further eigenvalues we may take their average. Thus the new covariance matrix is

$$\overline{\sigma}_{ij} = \sum_{l=1}^{k} \lambda_l v_i^{(l)} v_j^{(l)} + \overline{\lambda} \sum_{l=k+1}^{N} v_i^{(l)} v_j^{(l)}$$

  where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ are the eigenvalues, $v^{(1)}, \ldots, v^{(N)}$ are the corresponding eigenvectors
- Thus the smallest eigenvalues have increased

Conclusion, further ideas

- The problem will be solvable for $T < N$ as well, as the 0 eigenvalues were replaced by positive ones
- The divergence vanishes at the point $N/T = 1$, the value of q_0 will not be too large, even when N/T is small.
- But the fluctuation of the portfolio weights does not decrease substantially
- There are other methods, like Lasso or robust models
- Based on the robust models, confidence intervals might be constructed, and the worst cases of the conf.int. might be used in the portfolio optimisation
- mean-VaR or mean-cVaR portfolios might also be constructed
- Instead of Pearson-correlation, pairwise tail-dependence might be used (with 1's in the main diagonal)

References

- M. Potters, J.-P. Bouchaud, and L. Laloux. Financial applications of random matrix theory: old laces and new pieces. 2005.
- J. Fan, Y. Liao, and M. Mincheva: High dimensional covariance matrix estimation in approximate factor models. 2011.
- J.P. Bouchaud, M. Potters: Financial Applications of Random Matrix Theory, a short review. 2009.
- B. Valko's lectures: `http://www.math.wisc.edu/~valko/courses/833/2009f/lec_4_5.pdf,.../lec_6_7.pdf`
- A.J. McNeil, R. Frey and P. Embrechts: Quantitative Risk Management: Concepts, Techniques and Tools, 2005.
- Pfaff, B.: Financial Risk Modelling and Portfolio Optimization...
- Dowd, K.-Blake, D.: After VaR. The theory and estimation of quantile-based risk measures, 2006.
- A. Kempf and C. Memmel. On the estimation of the global minimum variance portfolio, 2003.
- Varga-Haszonits I.:Instability of Risk Measures PhD thesis, 2009