Kiln

Society of Actuaries in Ireland

UCD DUBLIN

An Empirical copula-based approach to Estimating Tail Dependence in Insurance Loss Distributions

Dr Adrian O'Hagan and Mr Robert Mc Loughlin, MSc

Objectives

- Aim:
  - Analyse Kiln's manual method of imparting tail dependence in their simulated loss data
- Method:
  - Copulas (particularly the empirical copula).
- Why?
  - Reserving purposes.

Overview

- Data: Kiln Group simulated loss data.
- Introduction to Copulas and tail dependence.
- The Empirical Copula.
- Finding the Upper Tail Dependence Coefficient.
- Results, Conclusions, Further Work…

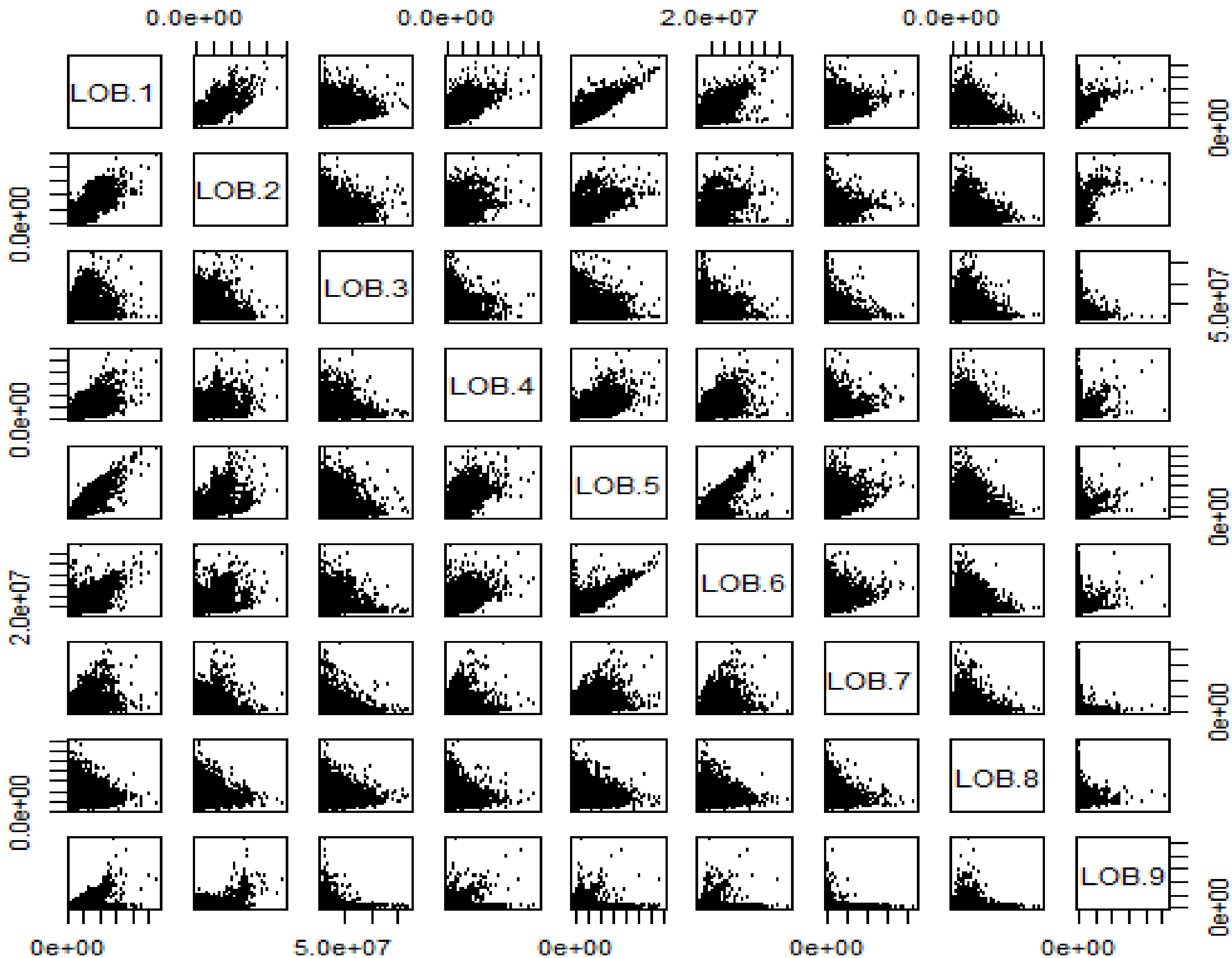
Data

- Data for this research provided by Kiln Group.
- 9 lines of business (LOBs).
- 50,000 randomly generated sample loss values for each line of business.
- Three loss layers:
  - Large(L), Attritional(A) and Catastrophe (CAT) losses.
- Each observation simulated using a convolution of distributions.

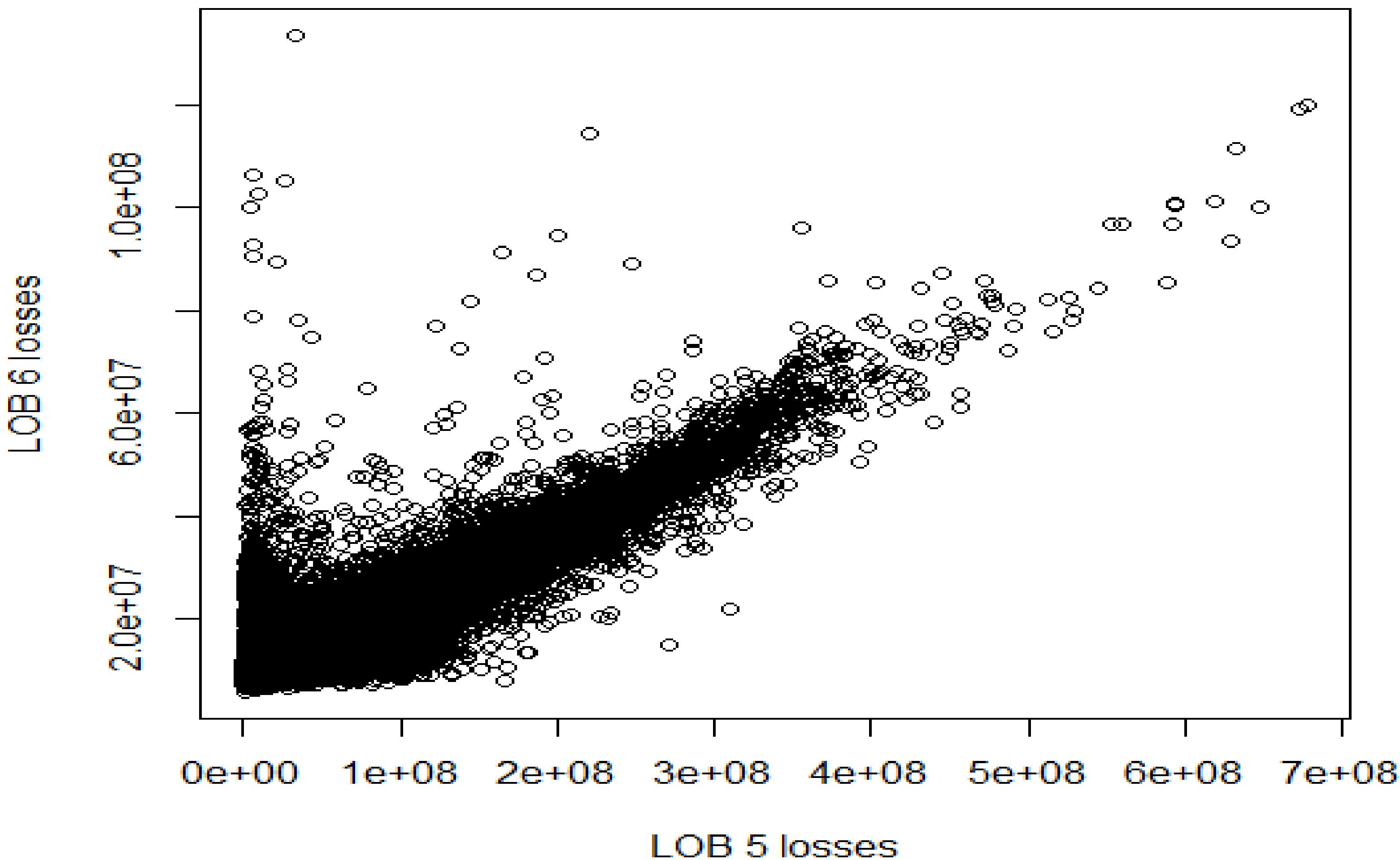
Data Simulation (hypothetical !)

- Construction of one observation in a line of business
- Large loss: Gamma(α, β)
- Attritional loss: Exp (γ)
- CAT loss: No. of CAT events X CAT loss size
  Pois (δ) X Exp(θ)
- **Total loss value = Large + Attritional + CAT**

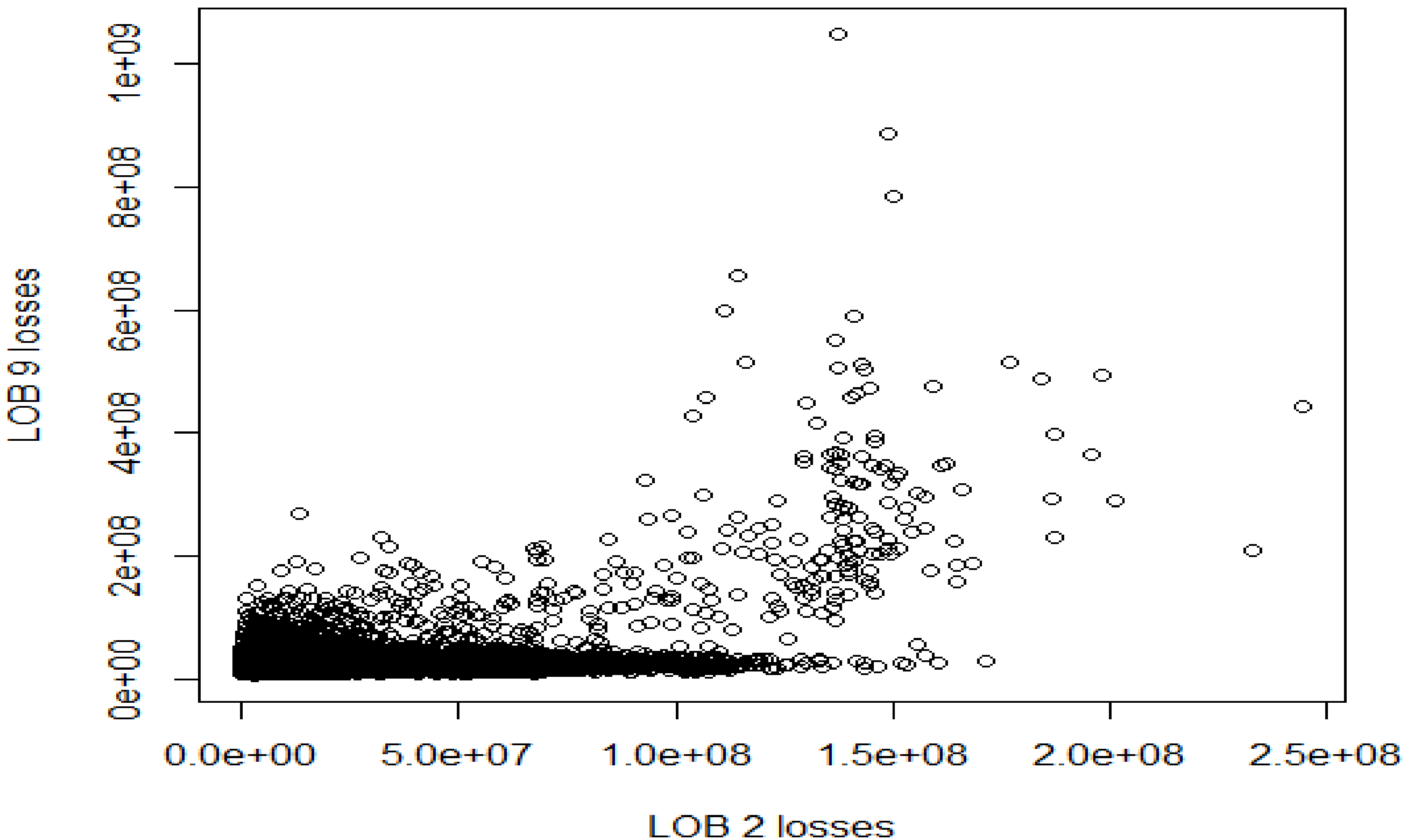
Data – Scatter Plots



Data – Lines of Business 5 and 6



Data – Lines of Business 2 and 9



Copulas

- First developed over 50 years ago (Sklar, 1959).

- Describe dependence between random variables.
- Model multivariate data via univariate marginals.

Sklar's Theorem

- $H(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$.
- $H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$.

- C is unique if all $(F_1, \dots, F_d)$ are continuous.

Copulas: the Gaussian Copula

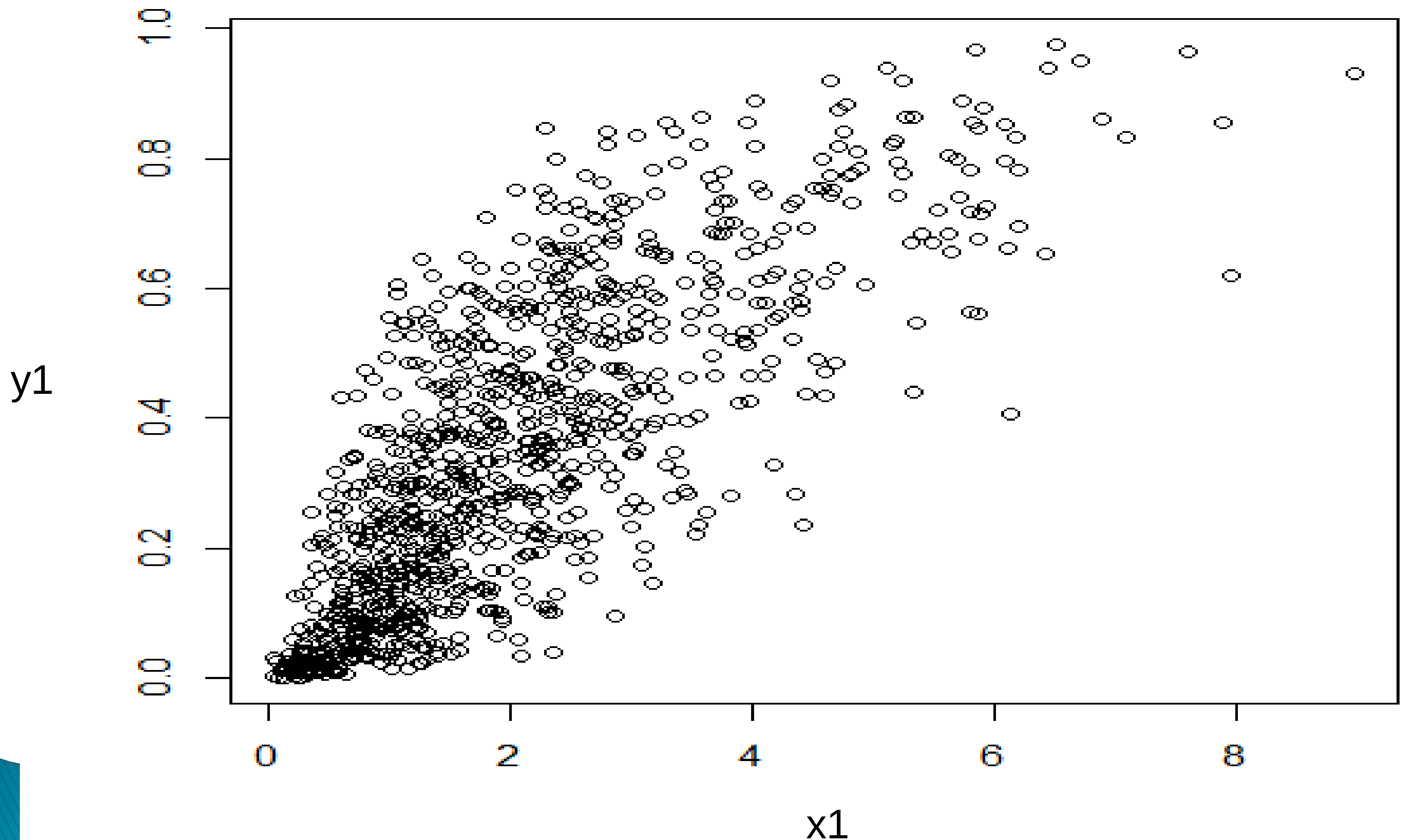
Became very prevalent in the decade leading up to the 2008 stock market crash.

- Gaussian copula (David Li):
- ${C_R}^{Gaussian}(\mathrm{u}) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$
- The correlation matrix R made the formula irresistible to many end users.
- Provides correlation but not tail dependence between the marginal distributions.

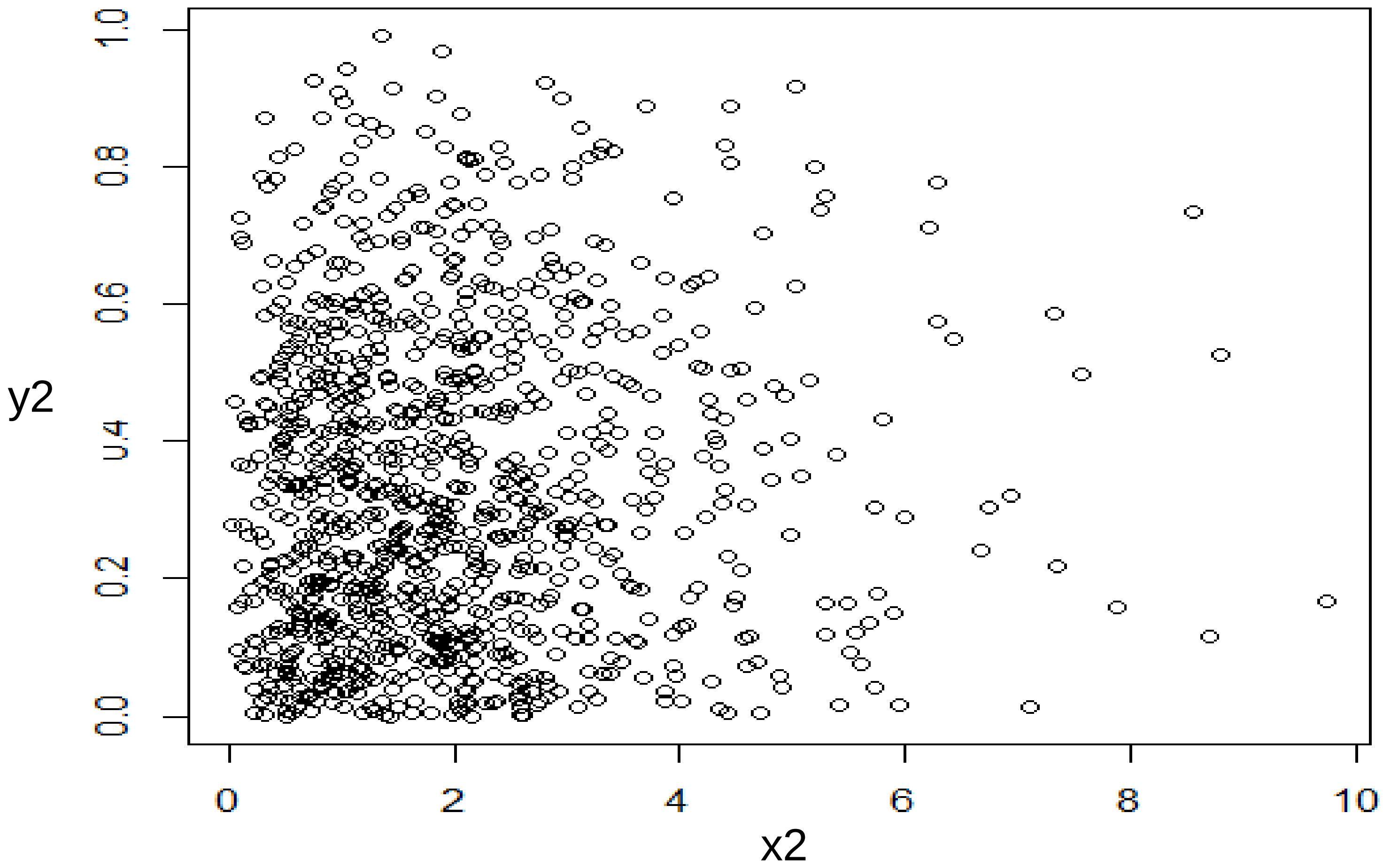
Gaussian Copula: simple example

```
S <- matrix(c(1,0.8,0.8,1),2,2)
#Correlation matrix

AB <- rmvnorm(mean=c(0,0),sig=S,n=1000)
#Our gaussian variables

U <- pnorm(AB)
#Now U has uniformly distributed columns

x1 <- qgamma(U[,1],2)     x2 <- rgamma(n=1000, 2, 1)
#x is gamma distributed

y1 <- qbeta(U[,2],1,2)    y2 <- rbeta(n=1000, 1, 2)
#y is beta distributed

plot(x1,y1) #They correlate!
plot(x2,y2) #They don't correlate!
```

x1 vs y1



x2 vs y2



Copulas: non-Gaussian Copulas

- Left tail
  - Clayton.
- Right tail
  - Joe, Gumbel.
- Two tailed
  - Student-t.

- Non-Parametric
  - Empirical.

Copulas and Tail Dependency

- Insurance companies mainly interested in upper tail dependence (worst case scenario).

- Upper Tail Dependence Coefficient λ_U (Schmidt, 2006):

$$\lambda_U = \lim_{v \to 1^-} \mathbb{P}\{G(X) > v | H(Y) > v\}$$

$$\lambda_U = \lim_{v \to 1^-} \frac{1-2v+C(v,v)}{1-v}$$

Fitting Joe and Gumbel Copulas

Heavy Right tail copulas: Joe (α) and Gumbel (θ).

- nacopula package in R fits these parametric copulas.

- nacopula works out the upper tail dependence coefficient λ_U from the fitted copula.

- Can also simulate data from the fitted copula.

- $\lambda_{U,Joe} = 2 - 2^{1/\alpha}$ $\lambda_{U,Gumbel} = 2 - 2^{1/\theta}$

Fitting the Student-t Copula

- Student-t copula:
  - Both upper and lower tail dependences are equal.

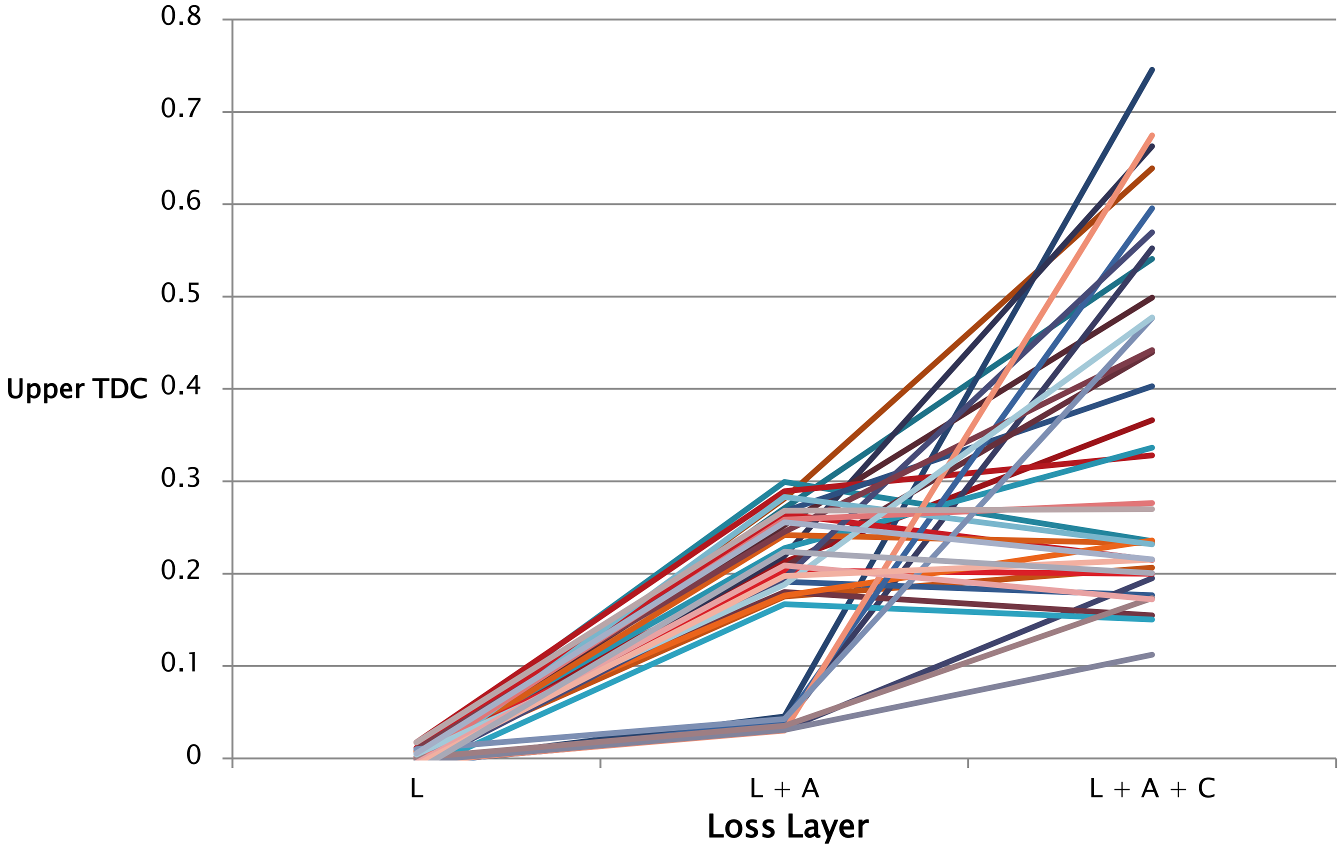
  For the bivariate case:

- $$\lambda_L = \lambda_U = 2t_{k+1}\left(-\sqrt{\frac{(k+1)(1-\rho)}{(1+\rho)}}\right)$$

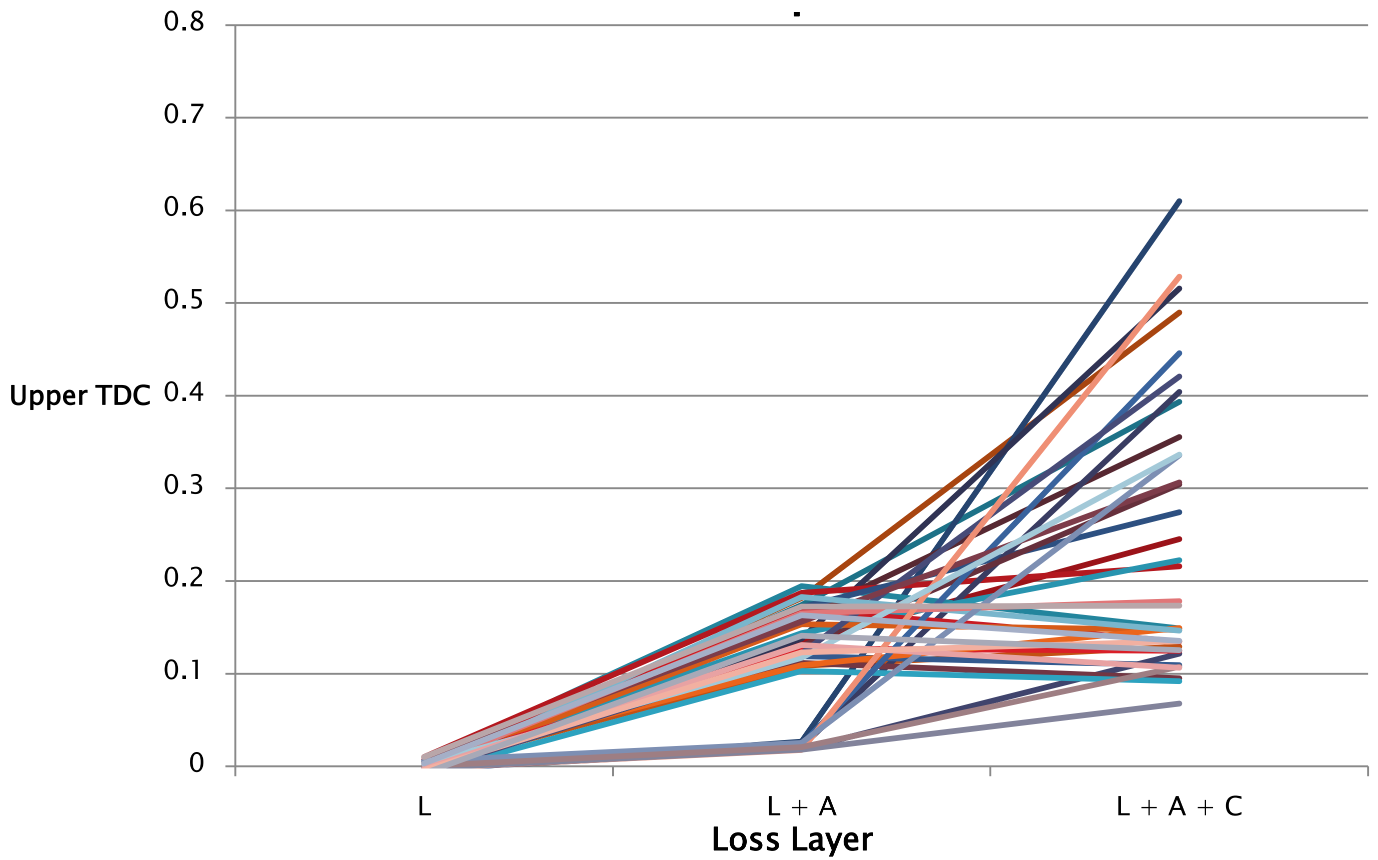
- t_{k+1} denotes the CDF of a t distribution with $(k+1)$ degrees of freedom; ρ is the correlation coefficient.

- Ellip-Copula R package automates this process.

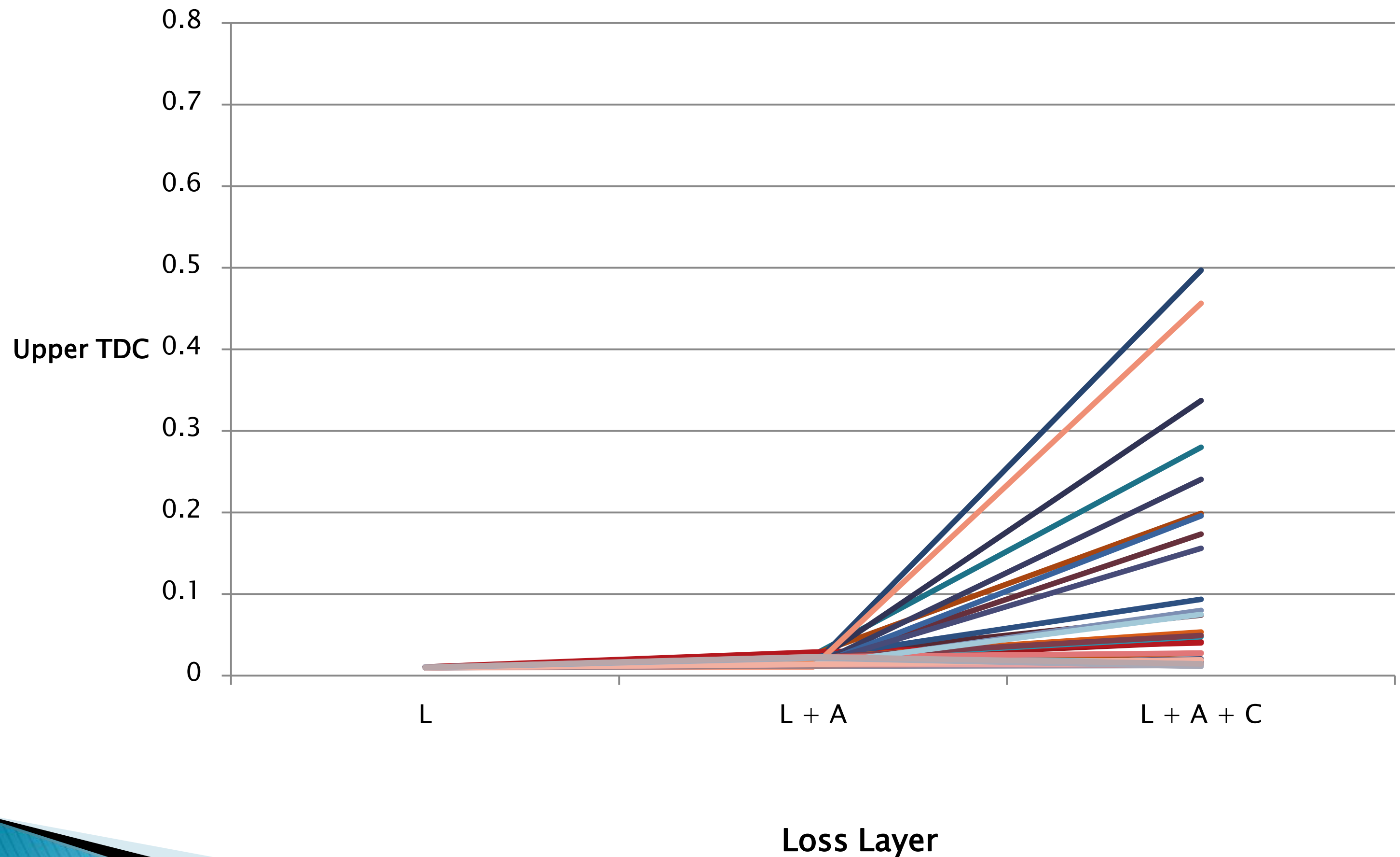
Scree Plot – Joe Copula



Scree Plot – Gumbel Copula



Scree Plot – Student-t Copula



Fitting an Empirical Copula

Uses the Empirical CDF of the data.

$$\lambda_U \approx 2 - \frac{lnC(1-\frac{t}{T},\ \ 1-\frac{t}{T})}{ln(1-\frac{t}{T})}$$

- T = number of observations.
- $t = \sqrt{T}$ "seems to work well" (Fischer and Dörflinger, 2006).

Fitting an Empirical Copula

- Can find *(X=x*)* and *(Y=y*)* yielding CDF values close to $(1-\frac{t}{T})$ for each of the marginal distributions…
- However the paired value *(x*, y*)* is unlikely to exist in the observed data.
- Extrapolate the joint empirical CDF to find a smoothed approximation to the intersection $C(1-\frac{t}{T},\ 1-\frac{t}{T})$.

Fitting an Empirical Copula

- Compute the marginal empirical CDFs for *X* and *Y*.
- Scale each set of marginal empirical CDF values by $T/(T+1)$. This prevents the maximum loss in each marginal having a CDF value of 1.
- Calculate the value $(1-\frac{t}{T})$ using $t = \sqrt{T}$, the joint CDF intersection point of interest.
- Determine a value *x'* such that $P(X < x') \approx (1-\frac{t}{T})$.
- Repeat previous step to identify y'.

Fitting an Empirical Copula

- Calculate the joint CDF for *X* and *Y*, smoothed across values of *X* and *Y* not in the data, *H*(x,y).*

  Currently use nonparametric kernel smoothing method in the *np* package in R.

- Use the smoothed joint CDF to find the value *H*(x',y')* and under Sklar's Theorem:

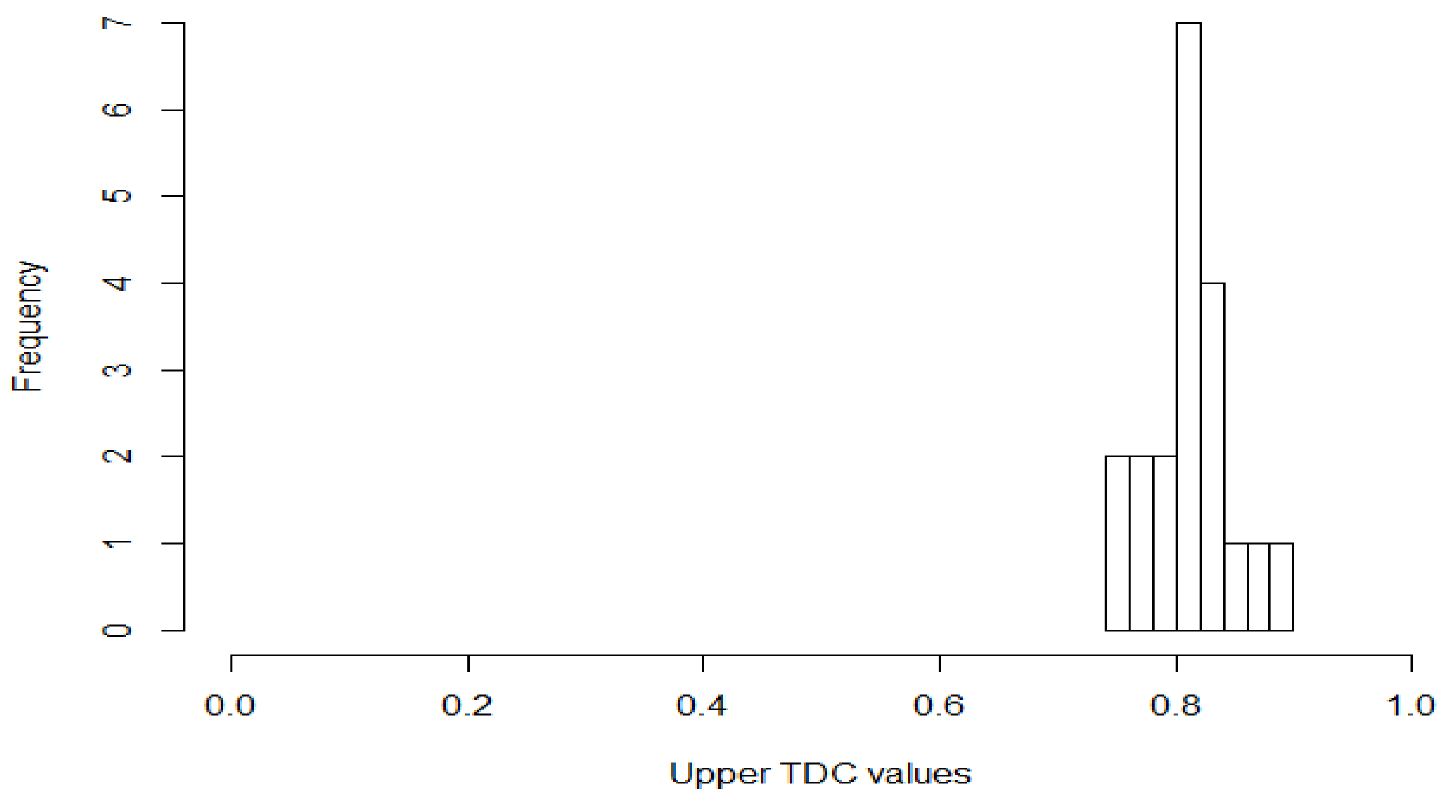
$$H^*(x',y') \approx \mathbb{P}(X<x',\ Y<y') \approx C\left(1-\frac{t}{T},\ 1-\frac{t}{T}\right)$$

- Calculate the estimate of the upper tail dependence coefficient λ_U' by substituting *H*(x',y')* for $C(1-\frac{t}{T},\ \ 1-\frac{t}{T})$.

Results – Empirical Copula

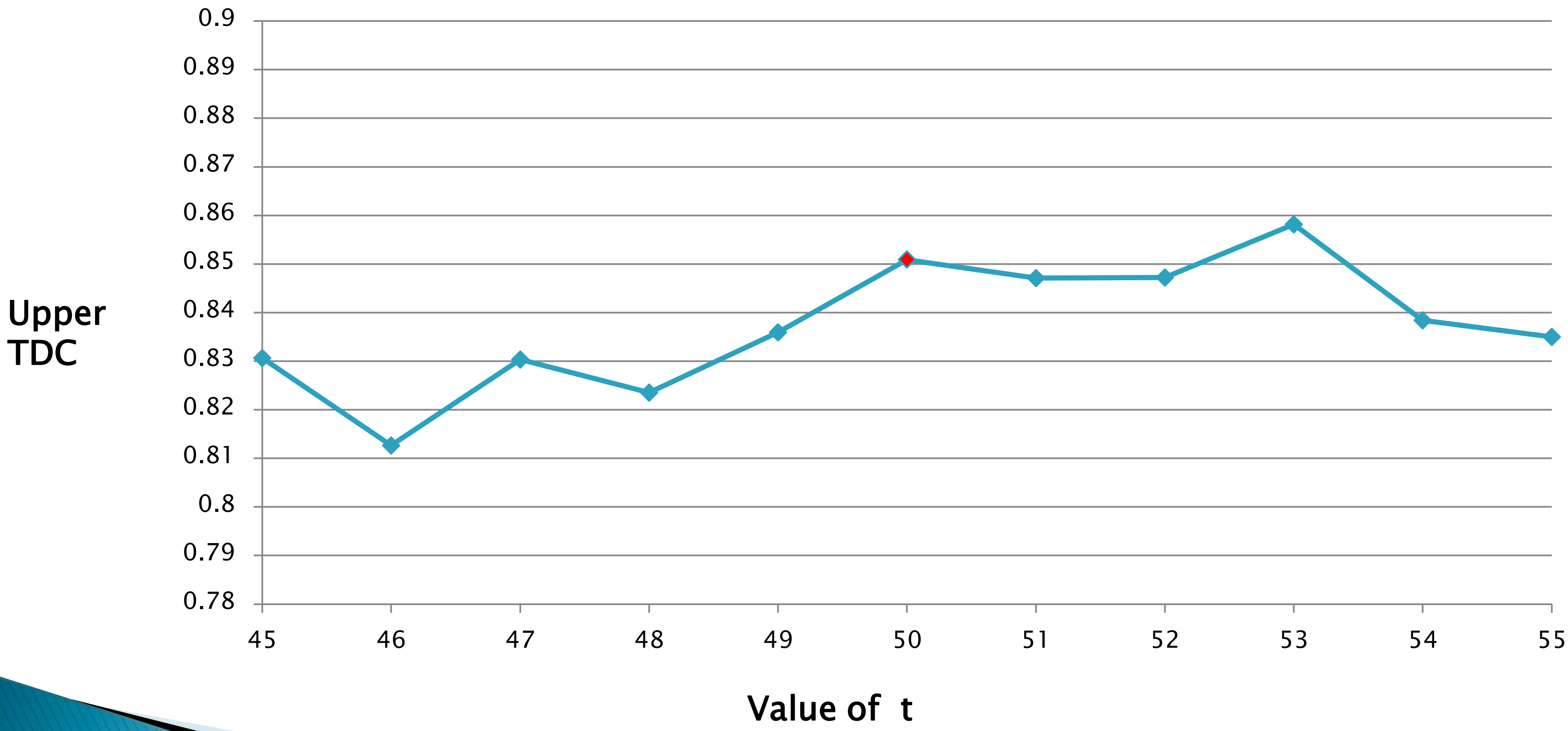
Pairing LOB(5,6), 20 samples of 2,500 losses.

- Mean = 0.84   Standard deviation = 0.05



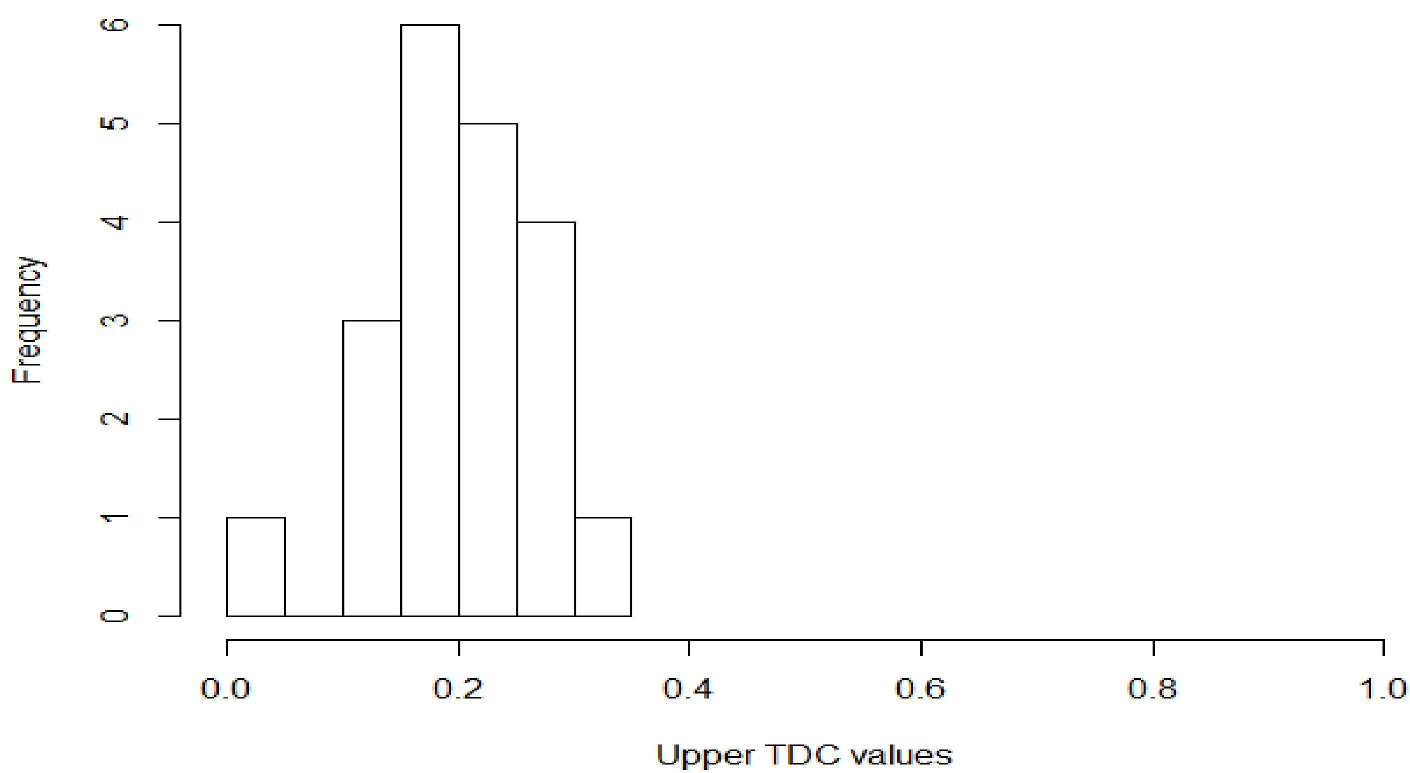
Results – Varying t in Empirical Copula

- Pairing LOB(5,6), 1 sample of 2,500 losses.
- t varied from 45 to 55.



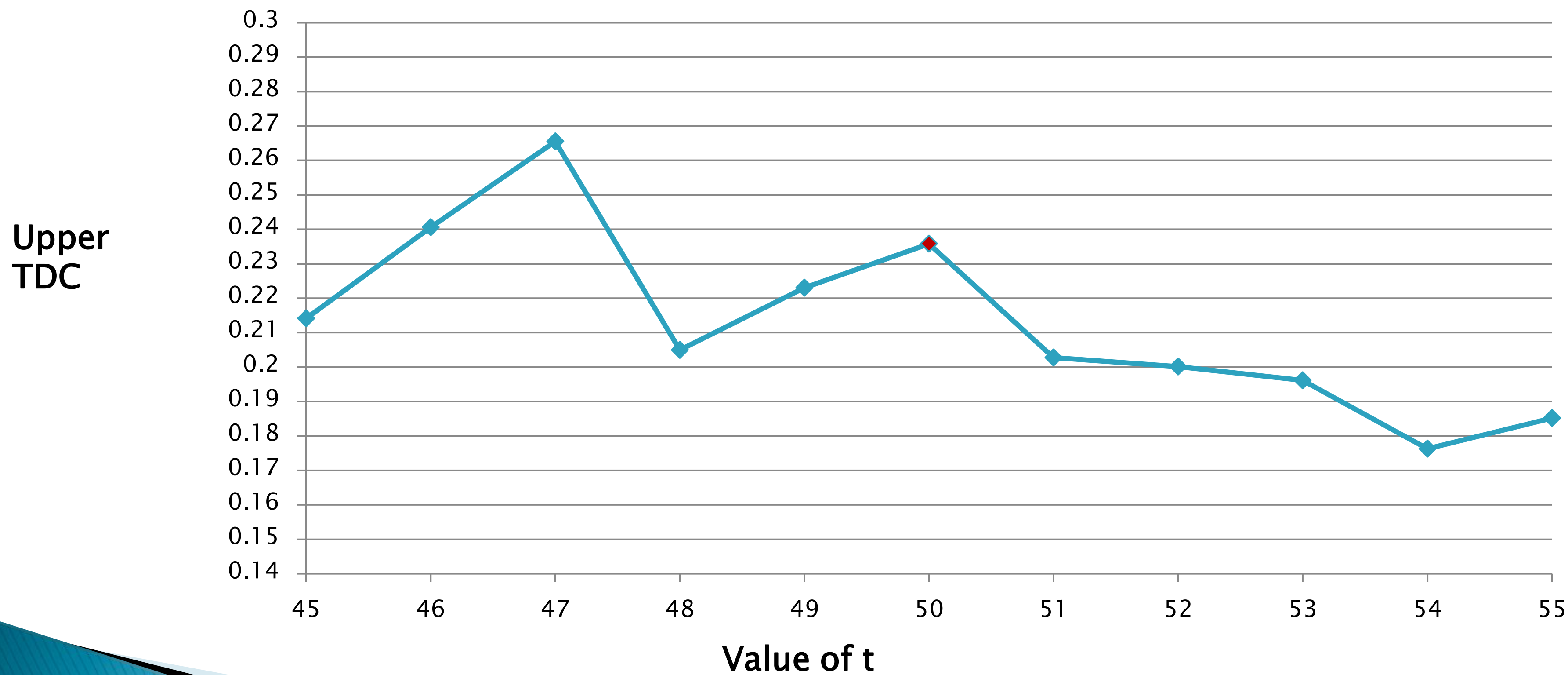
Results – Empirical Copula

- Pairing LOB(2,9). 20 samples of 2,500 losses.
  - Mean = 0.19   Standard deviation = 0.08

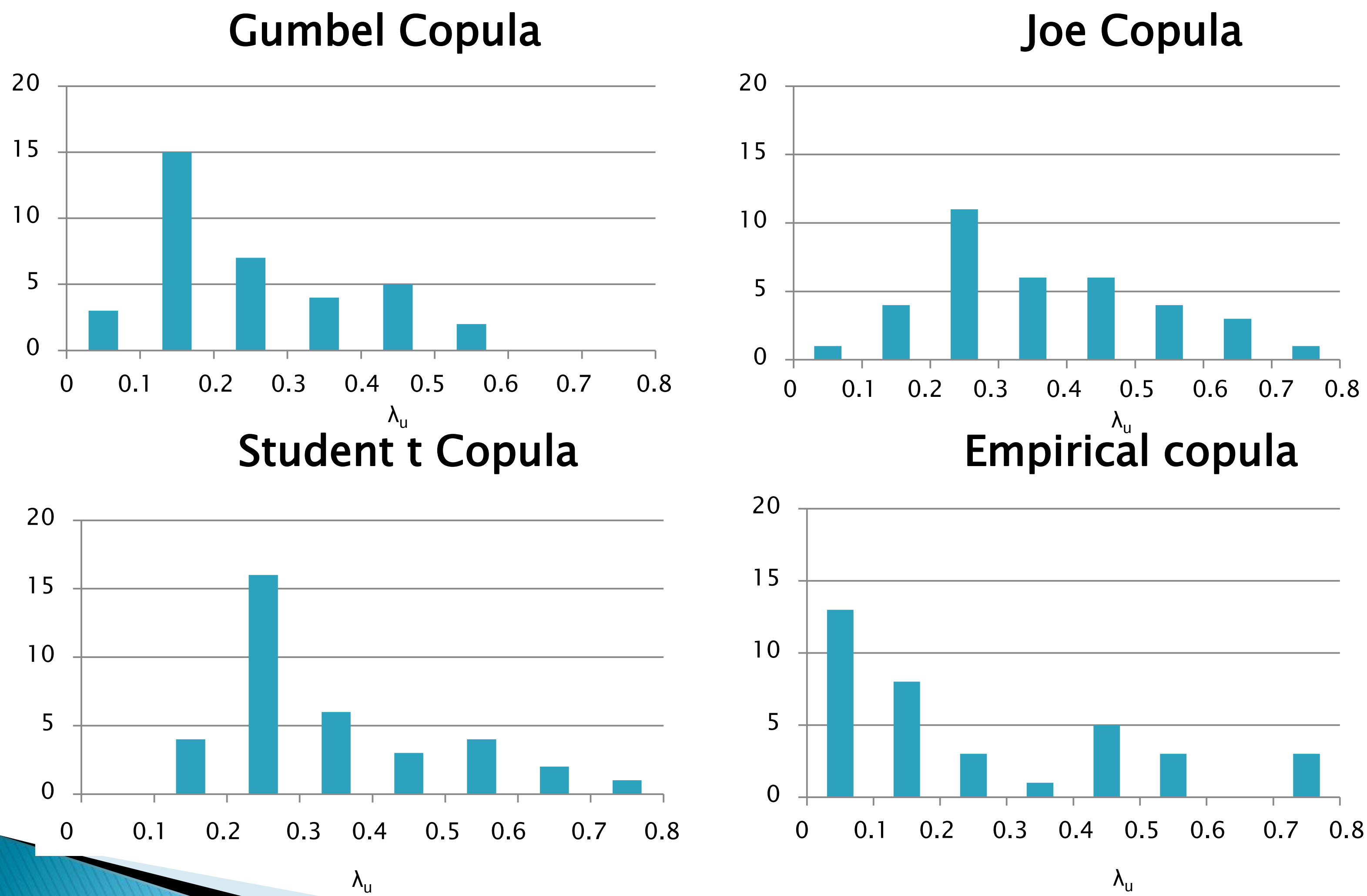


Results – Varying t in Empirical Copula

- Pairing LOB(2,9), 1 sample of 2500 losses.
- t varied from 45 to 55.



Histograms of Upper TDCs



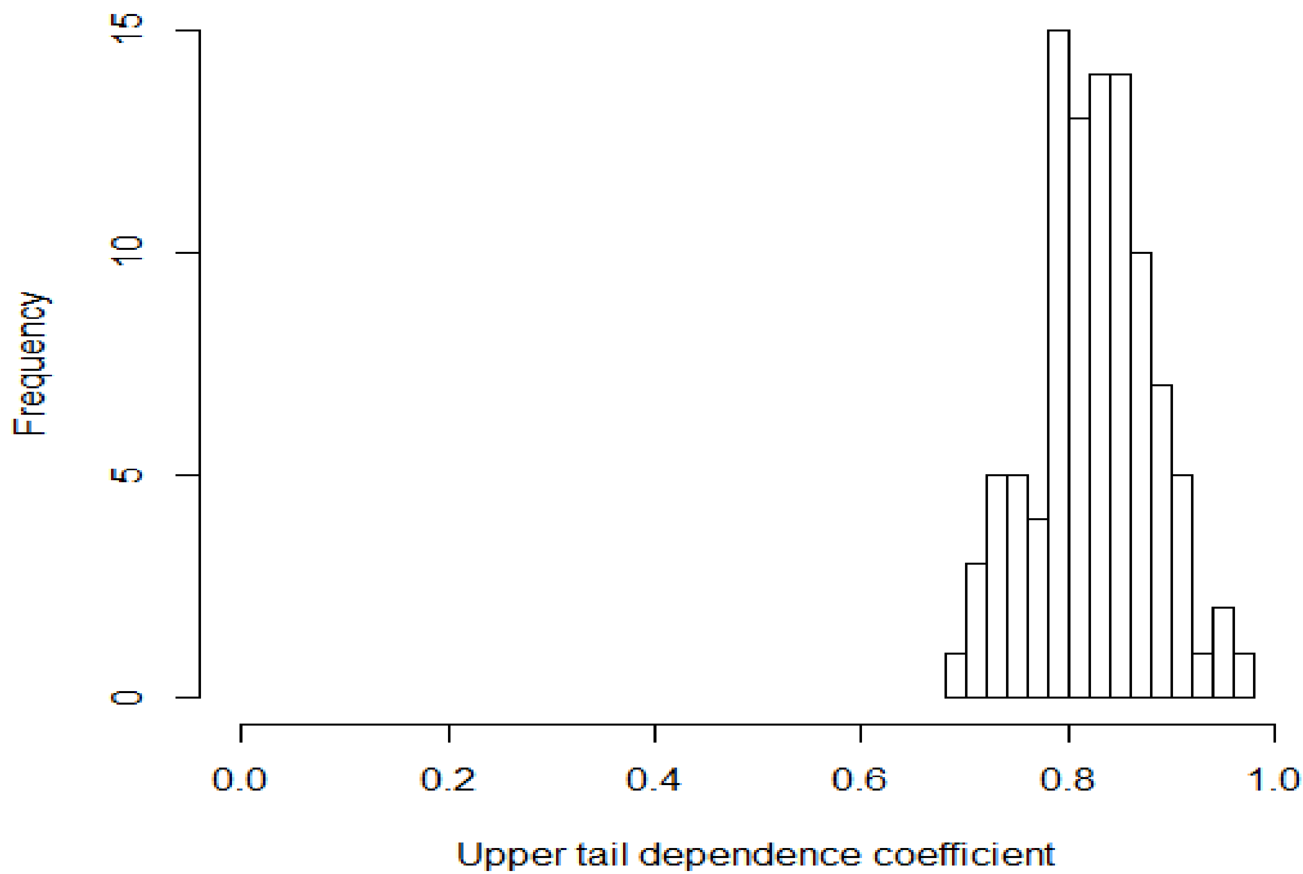
Performance for simulated data

- Simulated 100 sets of 1000 bivariate gamma and beta observations using the Joe, Gumbel, Student-t and Gaussian copulas.
- Compare known upper tail dependence coefficient for each versus the estimate using the empirical copula.
- Done for cases of both strong and weak upper tail dependence/correlation.

Joe Copula: strong upper tail dependence

- Joe (α=4) $\lambda_U = 0.824$

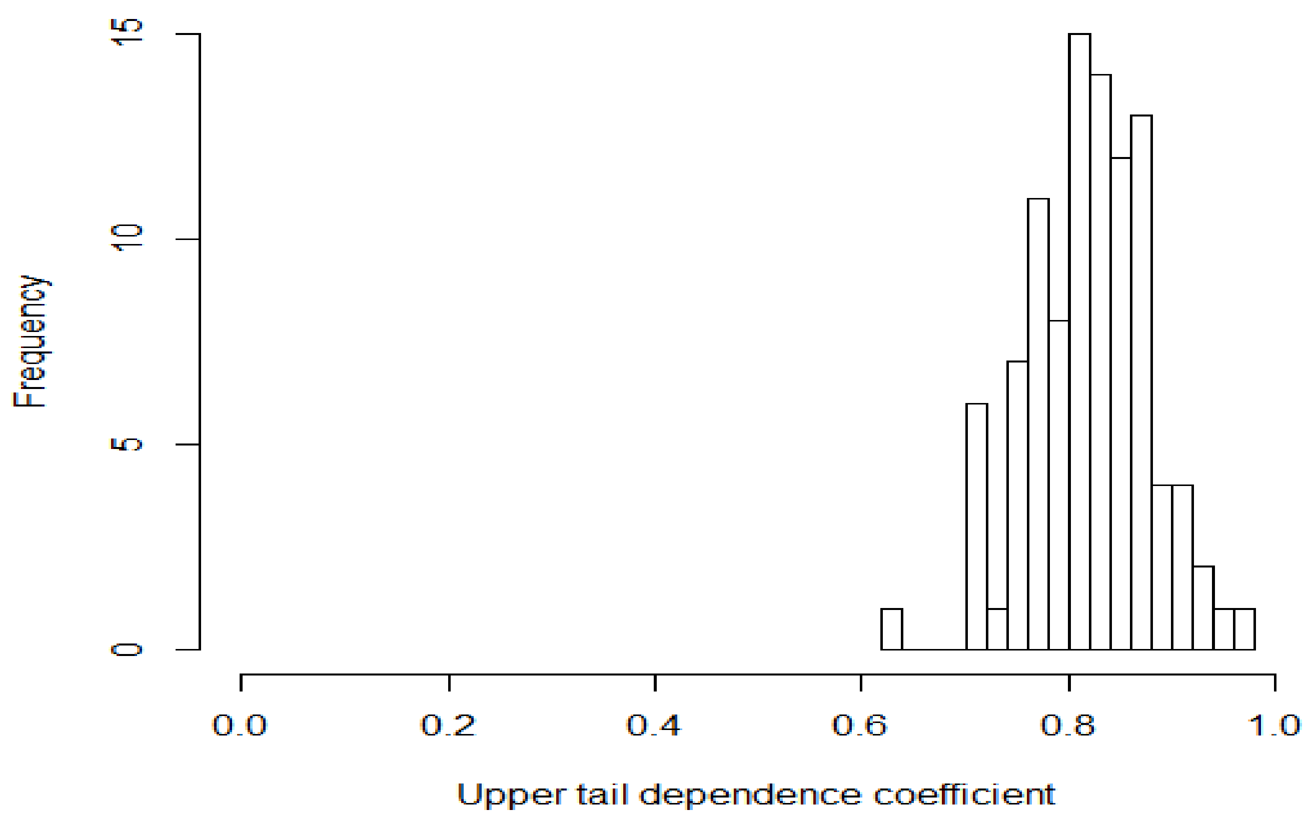
Mean = 0.811 Standard Deviation = 0.056



Gumbel Copula: strong upper tail dependence

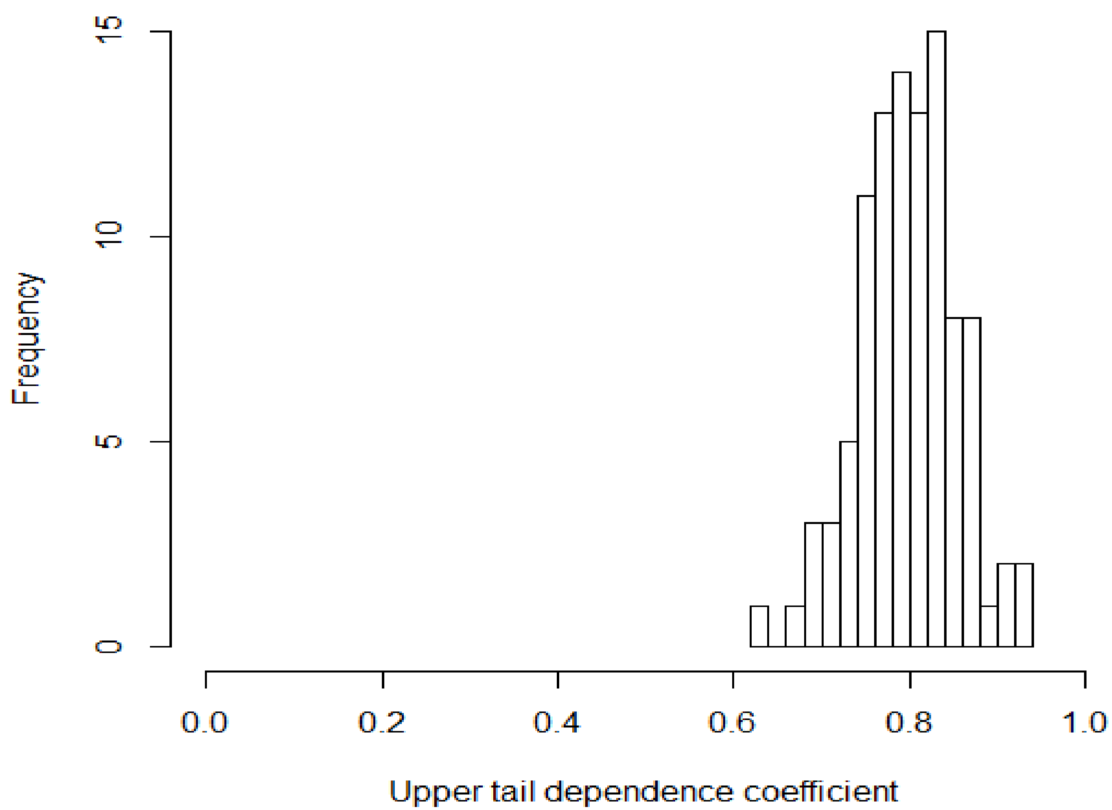
- Gumbel (θ=4) $\lambda_U = 0.824$

Mean = 0.818 Standard Deviation = 0.060



Student-t copula: strong upper tail dependence

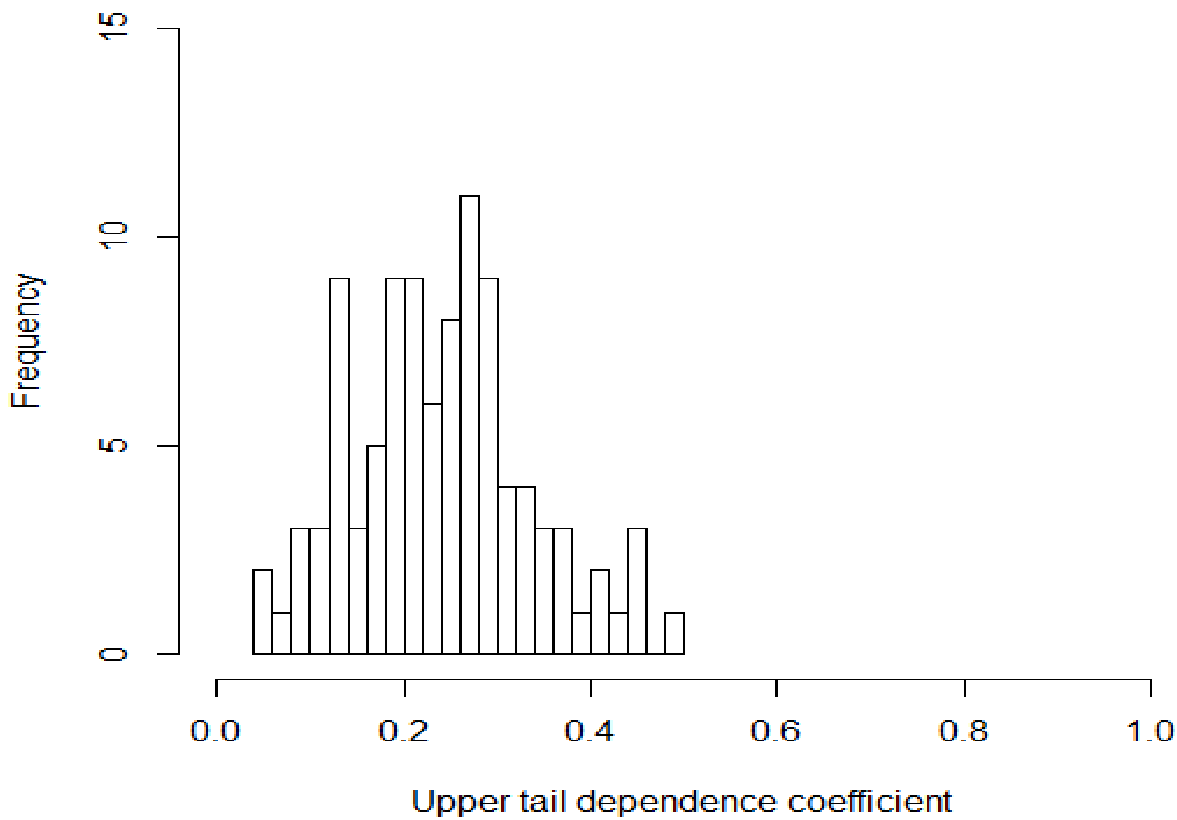
- Student-t (ρ=0.95) λ_u= 0.923
- Mean=0.821 Standard Deviation = 0.091



Joe Copula: weak upper tail dependence

- Joe (α=1.25) $\lambda_U = 0.259$

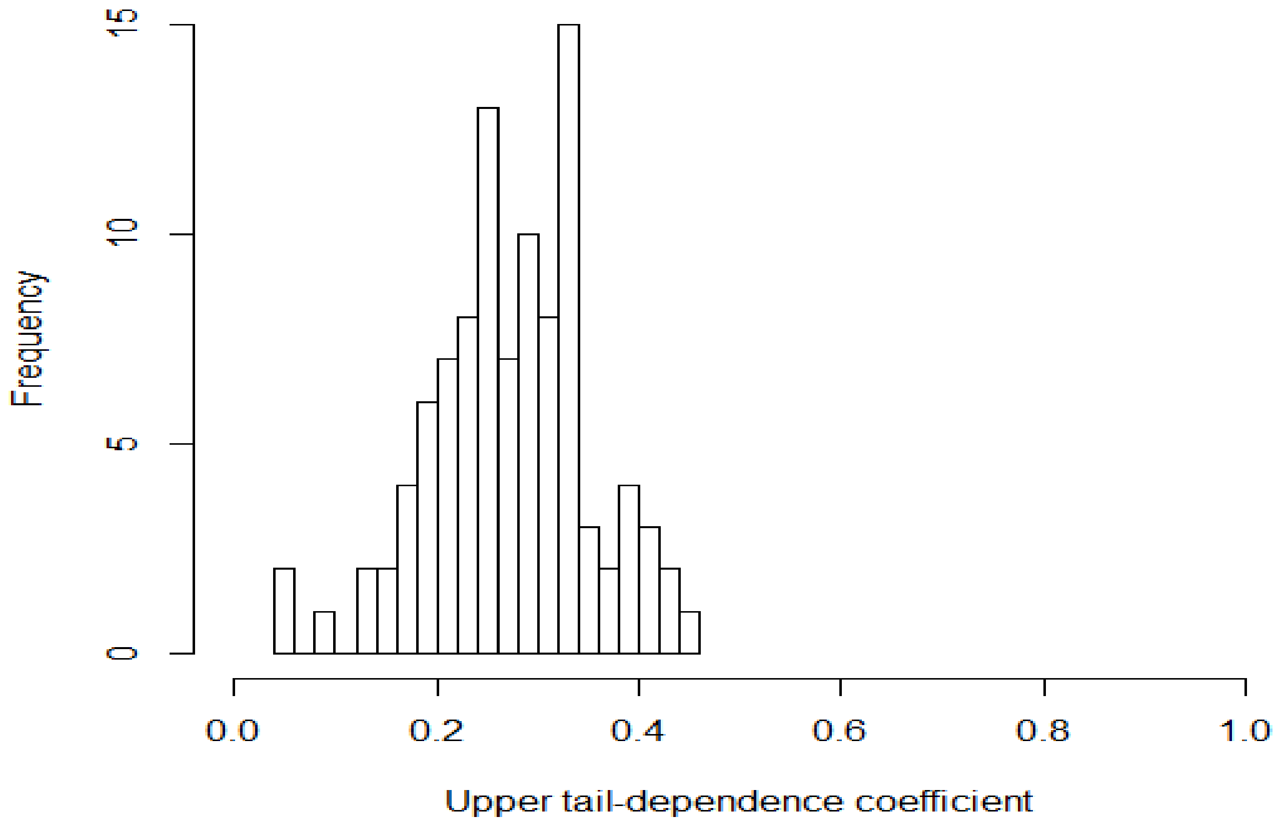
Mean = 0.240 Standard Deviation = 0.096



Gumbel Copula: weak upper tail dependence

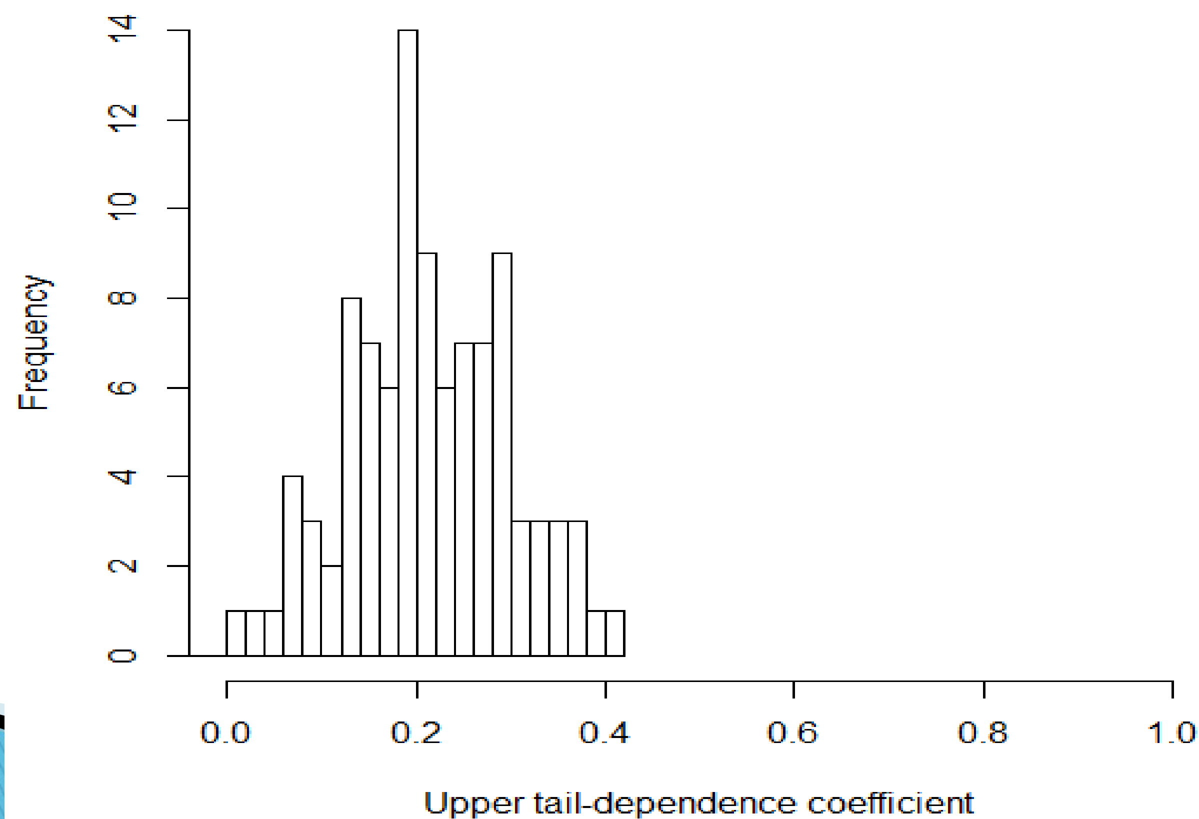
- Gumbel (θ=1.25) $\lambda_U = 0.259$

Mean = 0.272 Standard Deviation = 0.078



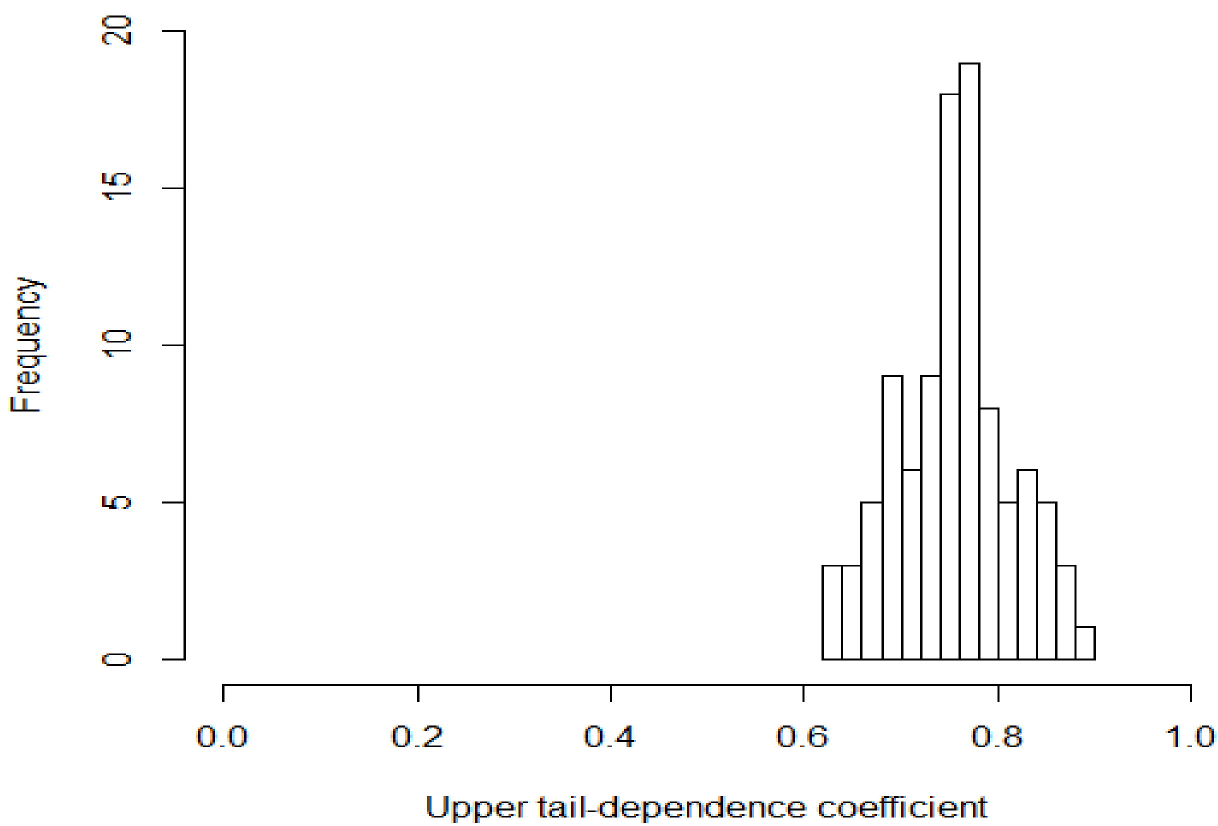
Student-t copula: weak upper tail dependence

- Student-t (correlation coefficient $\rho=0.2$) $\lambda_u=0.177$
- Mean = 0.209   Standard Deviation = 0.090



Gaussian copula: strong correlation

- Correlation coefficient $\rho = 0.95$, $\lambda_u = 0$
- Mean=0.734 Standard Deviation=0.058



Conclusions

Parametric copulas require distributional assumptions as to the nature of the data dependence structure.

Hence not suitable for data with a complex dependence structure already incorporating tail dependence.

Empirical copula is a useful post-modelling tool to assess upper tail dependence when tail dependence is present.

Further Work

- Statistical package EmpCop to automate calculation of empirical copula tail dependence coefficient.
- R package VS Shiny package ???
- GIRO 2014

Thank you !

Questions and Feedback ?