Modelling dependence in point fluctuation with Archimedean copulas

Tomáš Bacigál

Department of Mathematics and Descriptive Geometry
Faculty of Civil Engineering
Slovak University of Technology

Bratislava 2005

PhD supervisor: Professor Magda Komorníková

Content

Data review

Two univarite time series tied with time variable

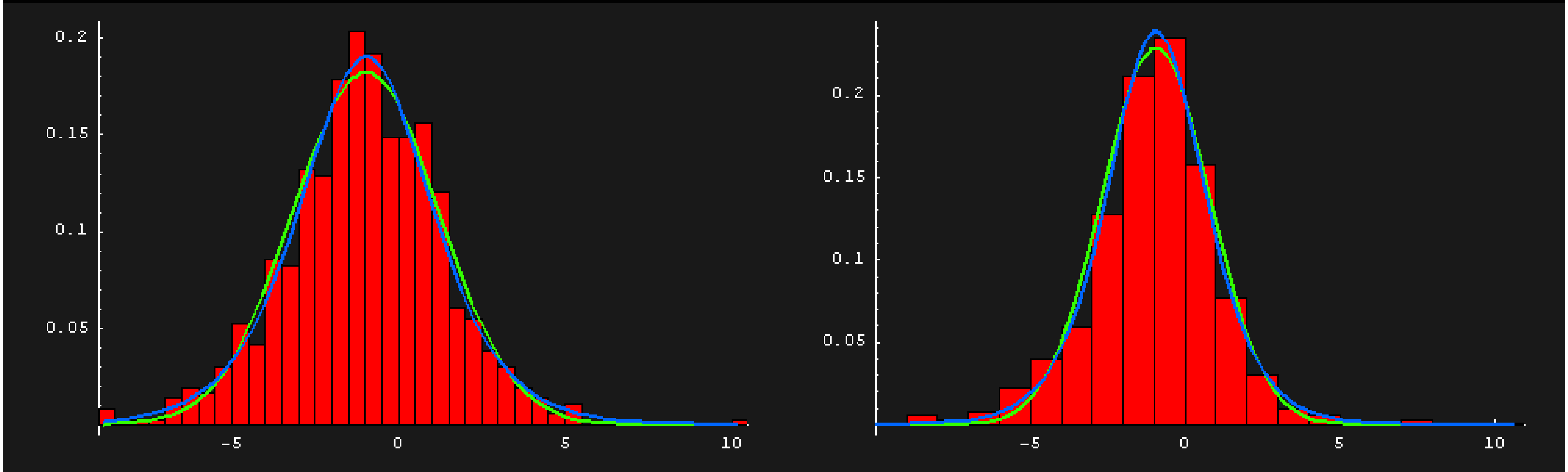
→ bivariate vector of observations

- Satellite based global positioning system (GPS)
- Daily observations performed on EUREF permanent station MOPI
- Sample length: 728 days
- Co-ordinate system of the raw data: geocentric oriented with Earth rotation
- Co-ordinate system of the modelled data: horizontal topocentric (North, East, Vertical)

Data review

Univariate analysis of probability distribution:

- Observe empirical distribution of data with histogram
- Estimate the best fitting continuous probability density function



- histogram - normal distribution - logistic distribution

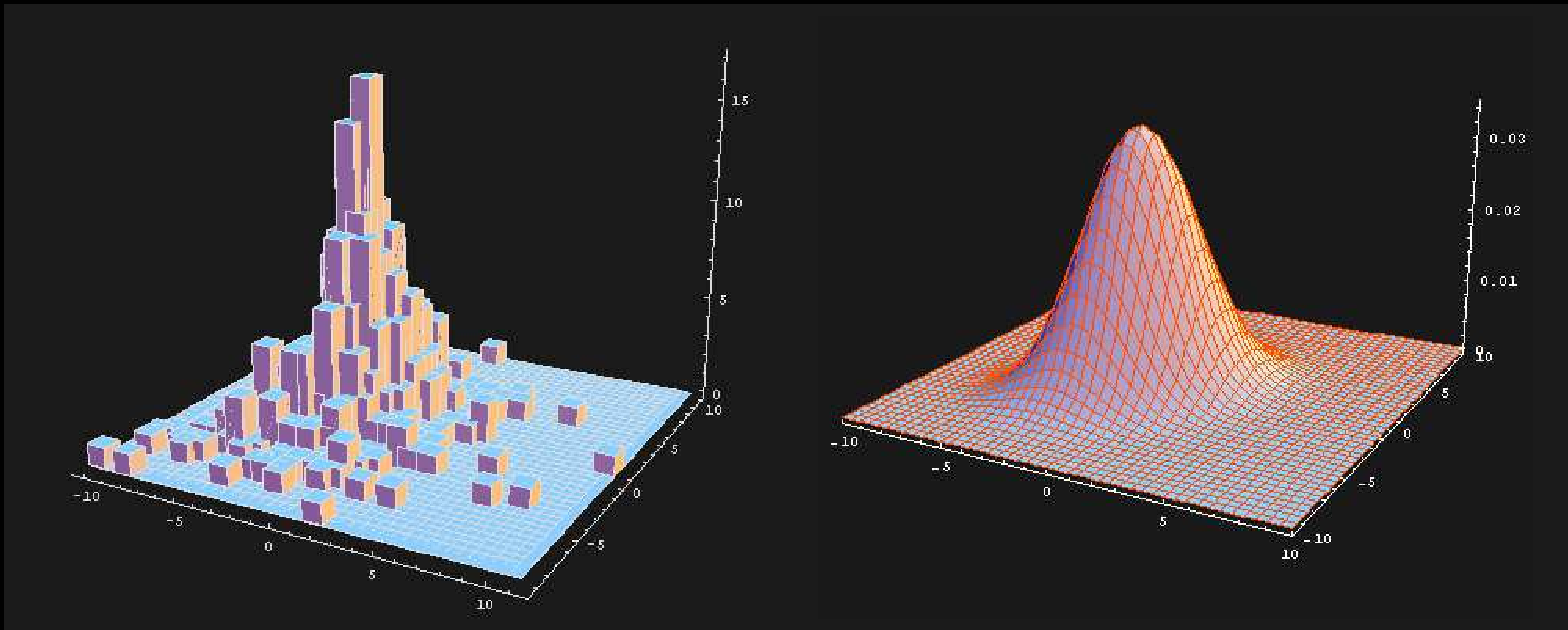
$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f(x)=\frac{\partial}{\partial x}\frac{1}{1+e^{-\frac{(x-\mu)^2}{\beta}}}$$

Data review

Multivariate analysis of joined probability distribution:

- Histogram of multivariate data
- Joined probability distribution function



Is multivariate normal distribution the most appropriate?

Introduction to copula theory

Definition

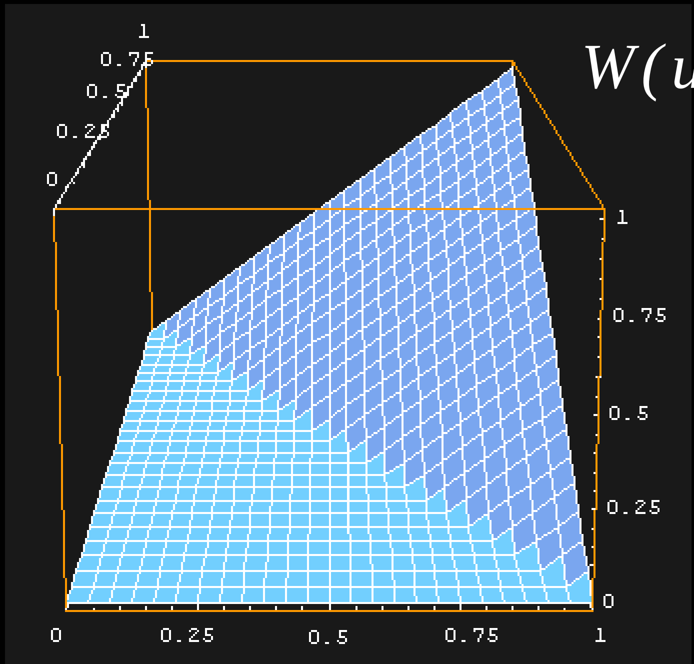
- Copulas are functions, that link univariate marginals to their joint distribution function

$$H(X,Y)=C[F(X),Y(G)]$$

where H is joint, F and G marginal ditribution functions and C a copula

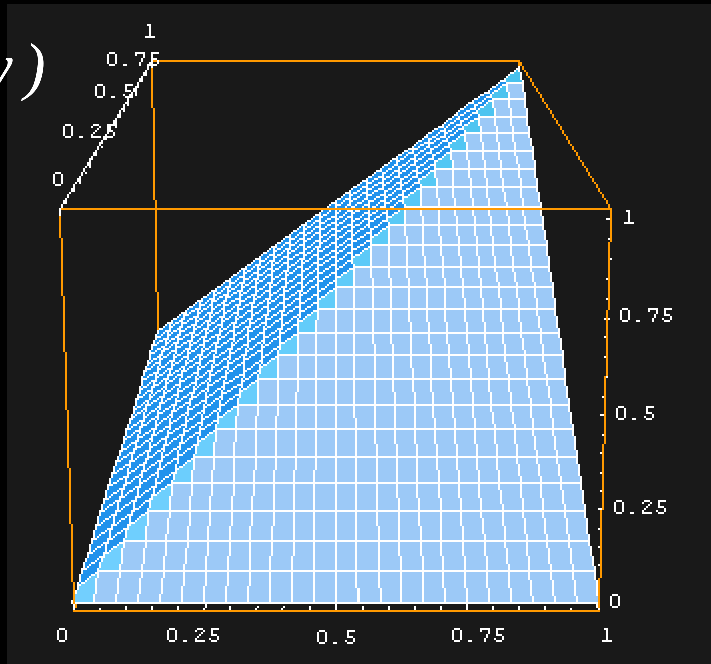
- Thus, copula captures solely the relationships among individual variables, not their distinctiveness

$$W(u,v)\leq C(u,v)\leq M(u,v)$$



perfect negative dependence

Fréchet-Hoeffding bounds



perfect positive dependence

Archimedean copulas

- Archimedean class of copulas allows us to reduce the study of a multivariate copula to a single univariate function ϕ, i.e.,

$$C_\phi(u,v)=\phi^{-1}[\phi(u)+\phi(v)]$$

where ϕ is convex, decreasing function $(0,1] \rightarrow [0,\infty)$ called *generator* of copula, and ϕ^{-1} its inverse

- As a generator uniquely determines an Archimedean copula, different choices of generator yield many families of copulas

one-parameter families :

Family of copulas	Generator $\phi(t)$	Dependence parameter	Bivariate copula $C_\phi(u,v)$	Special cases
Independence	$-\ln t$		uv	$C=\Pi$
Gumbel	$(-\ln t)^\theta$	$\theta \geq 1$	$e^{-[(-\ln u)^\theta+(-\ln v)^\theta]^{-1/\theta}}$	$C_1=\Pi,\ C_\infty=M$
Clayton	$t^{-\theta}-1$	$\theta > 0$	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$C_0=\Pi,\ C_\infty=M$
Frank	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$\theta \in \Re$	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right)$	$C_0=\Pi$ $C_{-\infty}=W,\ C_\infty=M$

Estimation of copula parameter

a) Nonparametric estimation (Genest and Rivest 1993)

is based on estimating empirical distribution function of unobserved random variable $Z_i = H(X_i, Y_i)$, that is $K_n(z) = Prob(Z_i < z)$, and its parametric version conditional to generator of a copula family, that is

$$K_\phi(z) = z - \frac{\phi(z)}{\phi'(z)}$$

and their comparing either by graphical or numerical way.

b) Semi-parametric estimation (Genest, Ghoudi and Rivest 1995)

uses functional expression of copulas to look for its parameter(s). “Semi“ means, that the empirical marginal distribution function is employed rather than estimated continuous d.f. of particular form.

- Maximization of log-likelihood function $L(\theta) = \sum_{i=1}^{n} ln\left(c_\theta[F(x_i), G(y_i)]\right)$

  where c denotes copula density (partial derivative of $C(u,v)$ with respect to u and v)

- Least-squares fit to empirical copula

Estimation of copula parameter

a) Nonparametric estimation - procedure

1. Find Kendall's tau using the usual nonparametric estimate

$$\tau_n = \binom{n}{2}^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} Sign[(X_i, X_j)(Y_i, Y_j)]$$

2. Construct a nonparametric estimate of K

$$K_n(z) = \frac{1}{n} \sum_{i=1}^{n} If\,[Z_i \le z, 1, 0]$$

where

$$Z_i = \frac{1}{n-1} \sum_{j=1}^{n} If\,[(X_j < X_i) \wedge (Y_j < Y_i), 1, 0] \qquad \text{(pseudo-observations)}$$

3. Construct a parametric estimate K_ϕ using

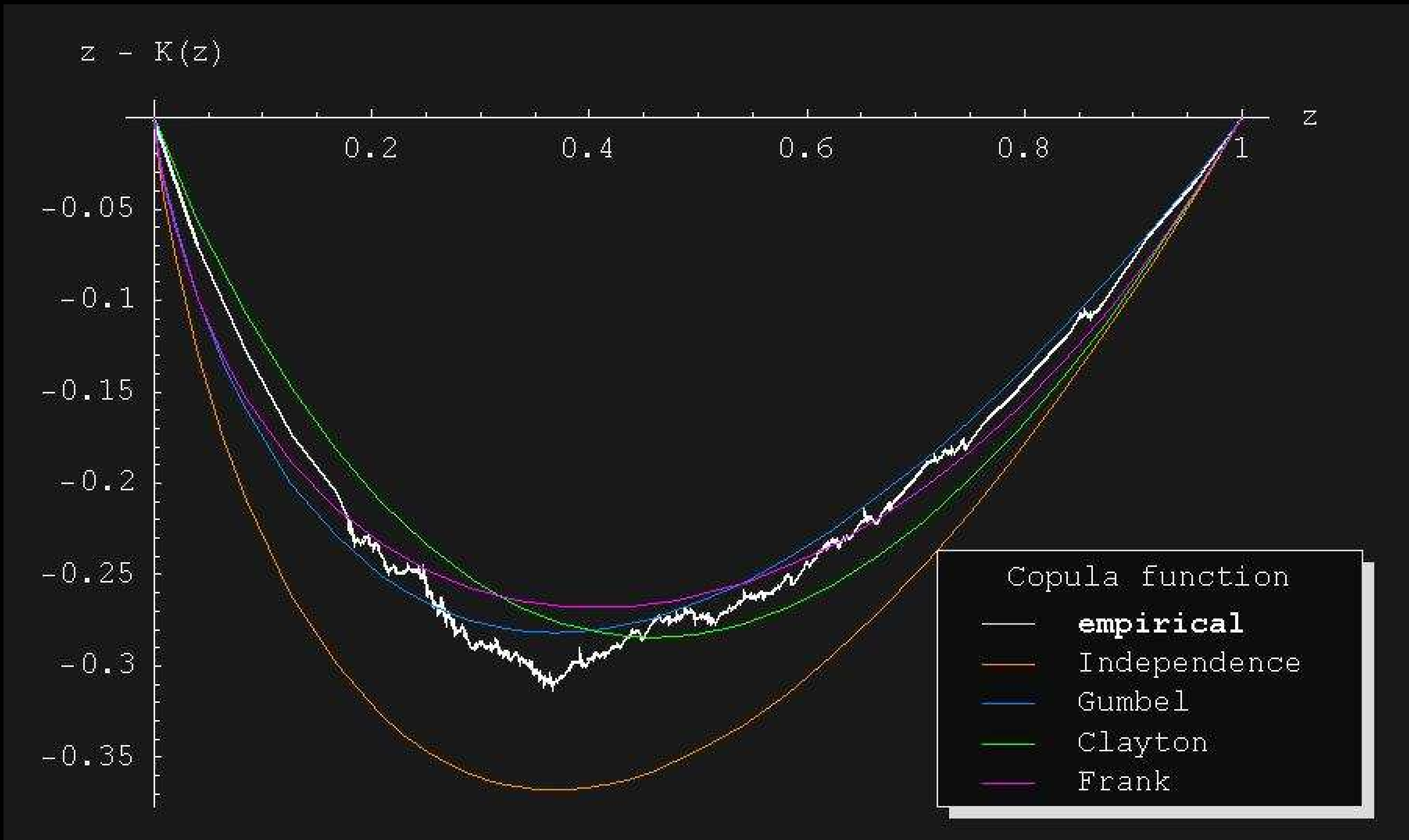
$$K_\phi(z) = z - \frac{\phi(z)}{\phi'(z)}$$

where generator ϕ (through its parameter θ) can be obtained by solving

$$\tau_n = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt$$

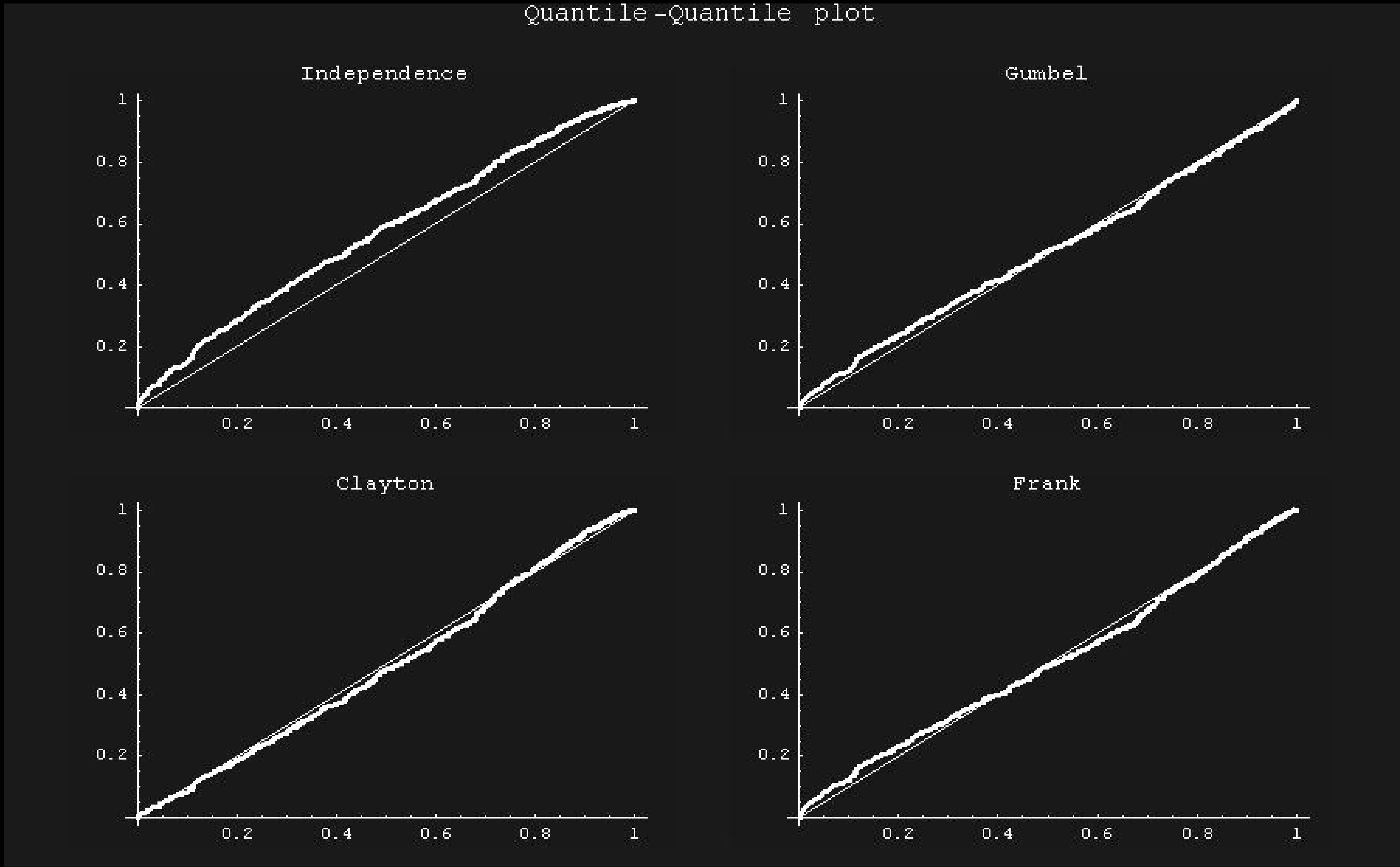
Estimation of copula parameter

a) Nonparametric estimation – graphical method



Estimation of copula parameter

a) Nonparametric estimation – graphical method



Q-Q plots are used to determine whether two data sets come from populations with a common distribution. If the points of the plot, which are formed from the quantiles of the data, are roughly on a line with a slope of 1, then the distributions are the same.

Estimation of copula parameter

b) Semi-parametric estimation – procedure

Simply looking for copula parameter $\hat{\theta}$ that maximizes the pseudo log-likelihood function

$$L(\theta)=\sum_{i=1}^{n} ln\left(c_{\theta}[F(x_i),G(y_i)]\right)$$

in which F_n, G_n stands for re-scaled empirical marginal distribution functions, i.e.,

$$F_n(x)=\frac{1}{n+1}\sum_{i=1}^{n} If[X_i \leq x,1,0]$$

(and similarly $G_n(y)$ for variable Y). The copula density c_{θ} for Archimedean class can be acquired from

$$c_{\theta}(u,v)=\frac{\partial^2 C_{\theta}(u,v)}{\partial u \partial v}=\frac{-\phi''(C_{\theta}(u,v))\phi'(u)\phi'(v)}{\phi'(C_{\theta}(u,v))^3}$$

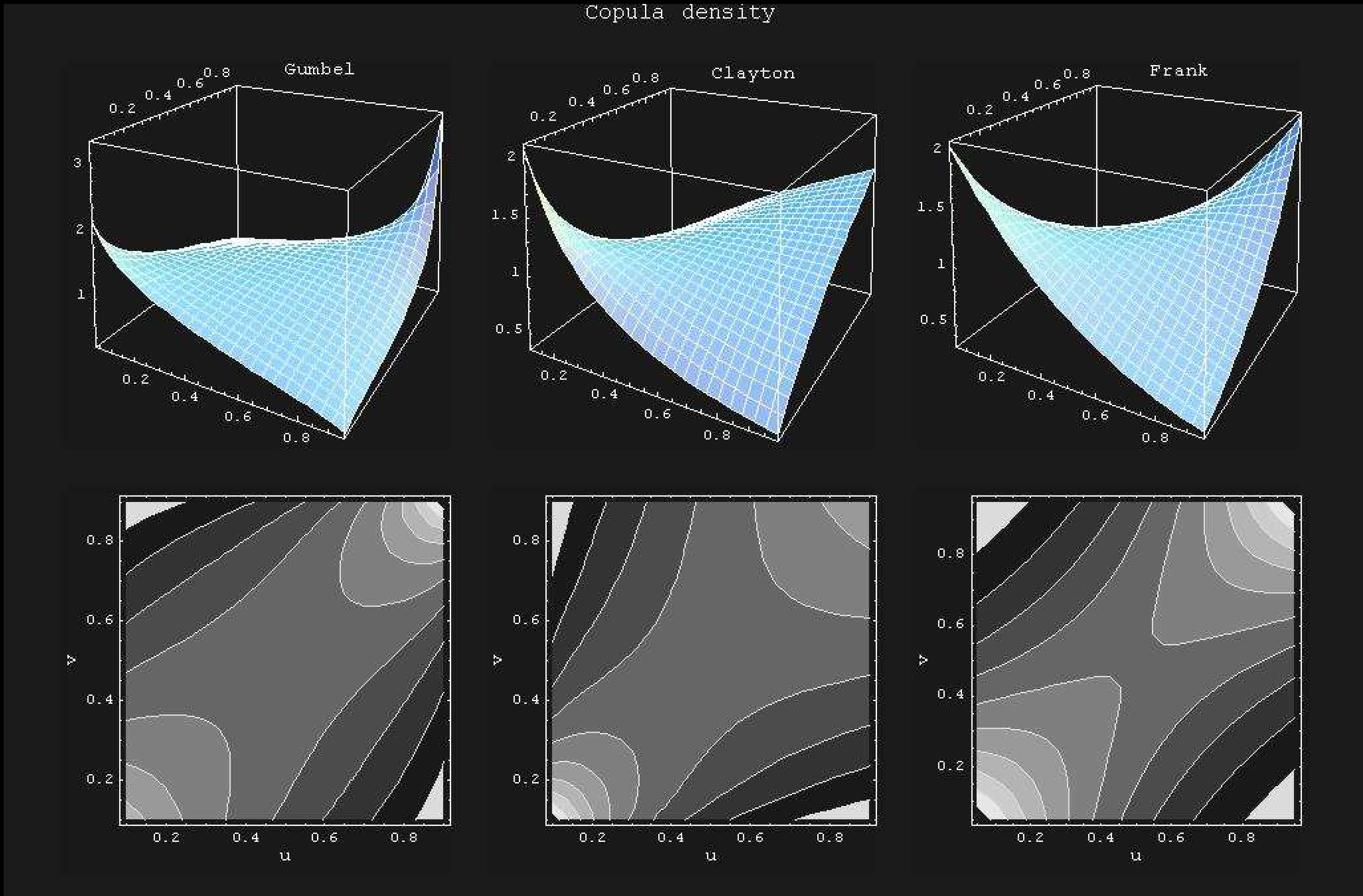
To examine goodness of estimation, there is a modification of the well known information criterion available: $AIC=-2L(\hat{\theta})+2k, \quad k=1.$

Estimation of copula parameter

Estimated parameters and measures of their appropriateness:

Family:	Gumbel	Clayton	Frank
Nonparametric procedure			
θ	1.3060	0.6120	2.2083
$d(K_\phi, K_n)$	0.445	0.542	0.492
Log-Likelihood procedure (semi-parametric)			
θ	1.3044	0.5638	2.3153
AIC	-106.2	-109.0	-90.7
$d(C_\theta, C_n)$	3.700	4.127	3.806
Nonlinear Fit procedure (semi-parametric)			
θ	1.3031	0.5595	2.1030
$d(C_\theta, C_n)$	3.700	4.127	3.598

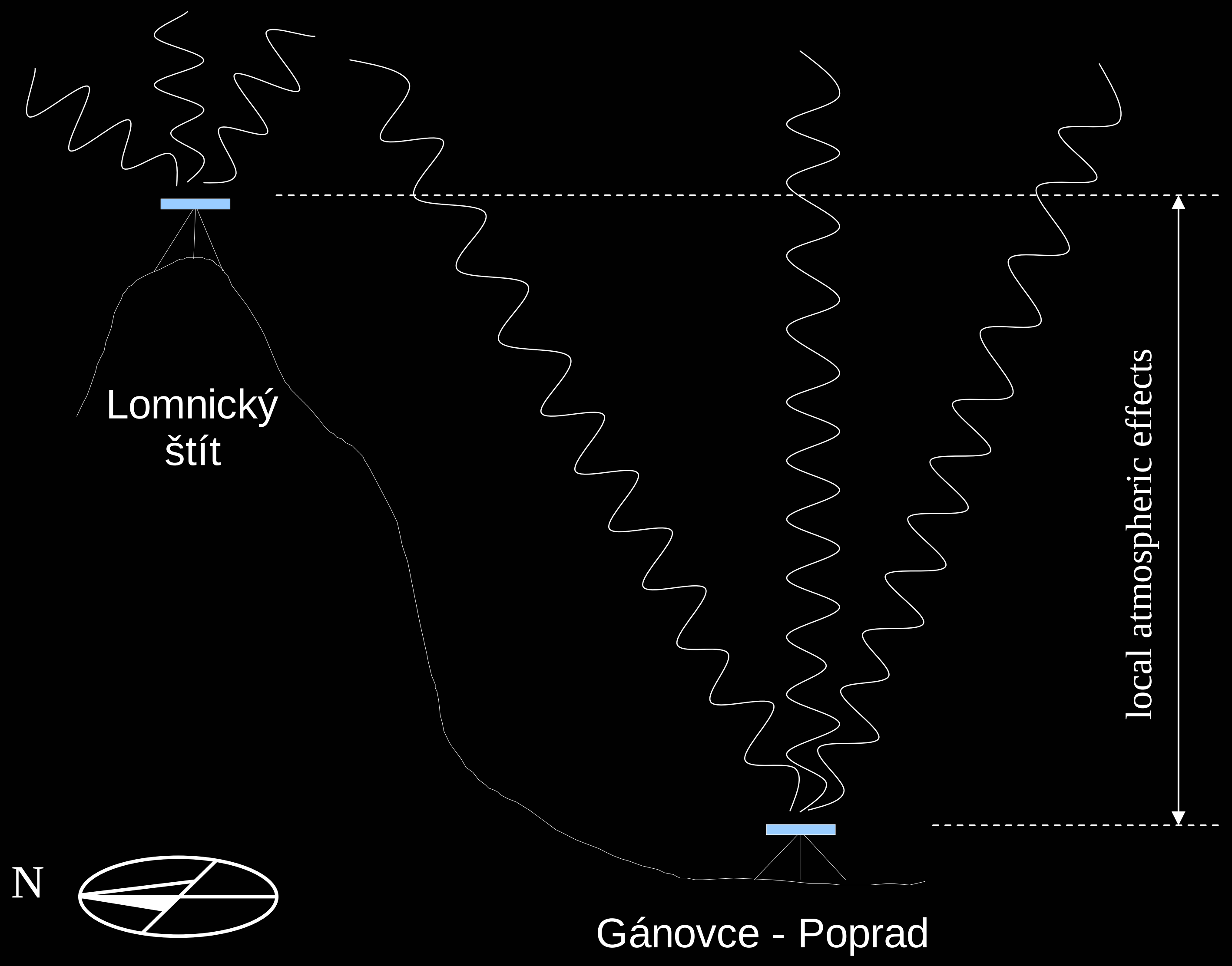
Linear convex combination:	Clayton - Gumbel	Clayton - Frank	Frank -Gumbel
Nonlinear Fit procedure (semi-parametric)			
α	1.3060	0.6120	2.2083
$d\left(C_{\theta 1} + (\alpha - 1) C_{\theta 2},\, C_n\right)$	0.445	0.542	0.492


Copula density
Gumbel
Clayton
Frank
0.2
0.4
0.6
0.8
u
v

Conclusion

- To be aware of distribution of dependence in data – tails, overall shape...
- To consider computational intensity.
- Improvement of fit with linear convex combination of copulas

Further applications



permanent observations
-- GPS coordinates
and time
-- temperature
-- atmosph. pressure

calculation
-- total zenit delay
-- precipitable water
vapour

Thank you