Dependency models with bivariate case study

Master's thesis presentation

Karolina Wojciechowska

TU Delft

July 26, 2007

Overview

1. Copula function
   - Basic facts
   - Archimedean class of copulas
   - Statistical inference
   - Case study
2. Bivariate conditional models
   - Transformation
   - Constant Spread Model
   - Variable Spread Model
   - Constant Symmetric Spread Model
   - Case study
3. Rotation Model
   - Construction
   - Properties
   - Case study
4. Conclusions and recommendations

Copula function - basic facts

- A copula is a distribution function defined on the unit square with uniform marginal distributions.
- Sklar's theorem $H(x, y) = C(F(x), G(y))$.
- If F and G are continuous then C is unique.
- $h(x, y) = c(F(x), G(y))f(x)g(y)$
- A copula describes margin-free dependence between random variables.

Archimedean class of copulas - definition

Definition

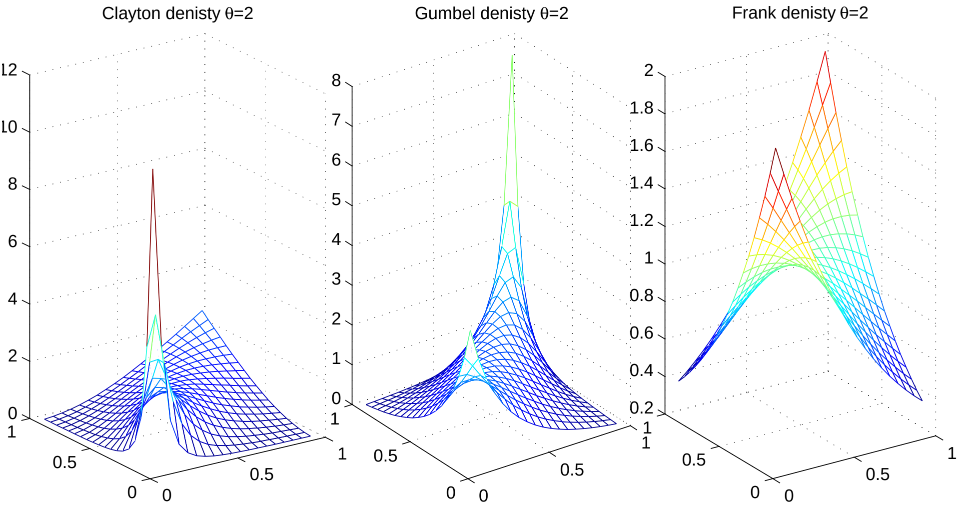
A bivariate copula C is called Archimedean with generator ϕ if:

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

a generator $\phi : [0, 1] \rightarrow [0, \infty]$ is convex, strictly decreasing $\phi(1) = 0$ and its pseudo-inverse $\phi^{[-1]}$ is:

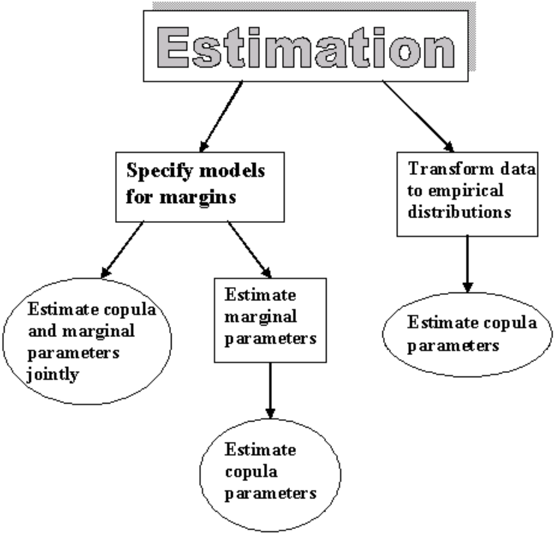
$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t \leq \phi(0) \\ 0 & \phi(0) \leq t \leq \infty \end{cases}$$

Archimedean class of copulas - examples



Estimation - Maximum Likelihood Inference

- $\prod_{i=1}^{n}\{c(F(X_i), G(Y_i))f(X_i)g(Y_i)\}$



Pseudo-graphical method for Archimedean class

- Assume that copula C is Archimedean and consider a sample of n bivariate observations $\{(X_i, Y_i)\}_{i=1}^{i=n}$. The method is based on comparison of parametric and non-parametric estimations of function $K(t) = P\{C(U, V) \leq t\}$, for $t \in [0, 1]$.

The non-parametric estimation of $K(t)$

$$W_i = \sum_{j=1}^{n} 1(X_j \leq X_i, Y_j \leq Y_i)/n \quad \Rightarrow \quad K_n(t) = \sum_{i=1}^{n} 1(W_i \leq t)/n$$

The parametric estimation of $K(t)$

$$\hat{\theta} \quad \Rightarrow \quad K(t; \hat{\theta}) = P\{C(U, V; \hat{\theta}) \leq t\} = t - \frac{\phi(t, \hat{\theta})}{\phi'(t^+, \hat{\theta})}$$

Pseudo-graphical method for Archimedean class

- Plot $K_n(t)$ and $K(t;\hat{\theta})$ on one graph - if the plots "agree" then the copula fits well.
- Compute distance E:

$$E = \int_0^1 |K_n(t) - K(t;\hat{\theta})|^2 dt$$

  The best copula minimizes this distance.
- Perform a statistical test with the test statistic:

$$S_n = n \int_0^1 |K_n(t) - K(t;\hat{\theta})|^2 k(t;\hat{\theta}) dt$$

  The Bootstrap method is used to obtain the critical value.

Method of percentile lines

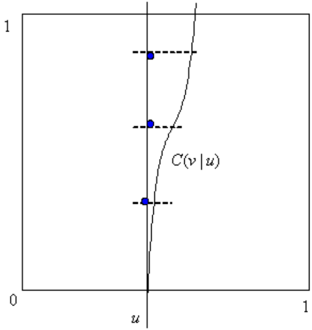
- (X, Y) - physical space
- $(F(X), G(Y))$ - copula space

Definition

The p-percentile line in the copula space is:

$$v = f(u;p), \qquad u \in [0,1]$$

where p is a percentage and $f(u;p)$ solves the equation $C(v|u) = p$.



- p-percentile line in the physical space:

$$\{(u, f(u;p)) : u \in [0,1]\} \quad \Rightarrow \quad \{(F^{-1}(u), G^{-1}(f(u;p))) : u \in [0,1]\}$$

Case study - dataset

- Covariance matrix:

$$\Sigma = \begin{bmatrix} 0.26 & 0.61 \\ 0.61 & 7.50 \end{bmatrix}$$

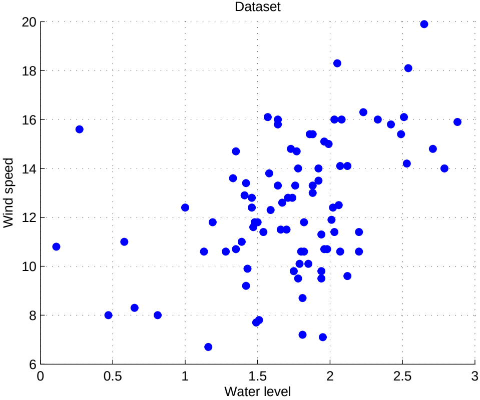


Figure: 89 observations of water level and wind speed.

Case study - marginal distributions

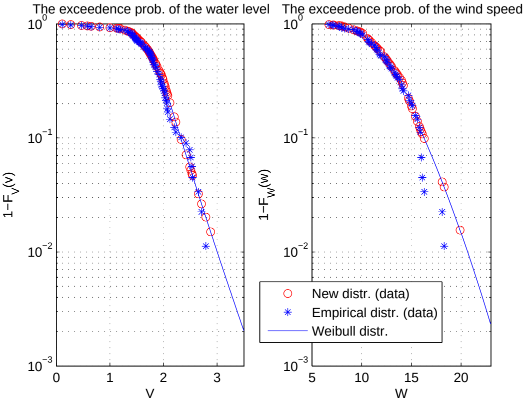


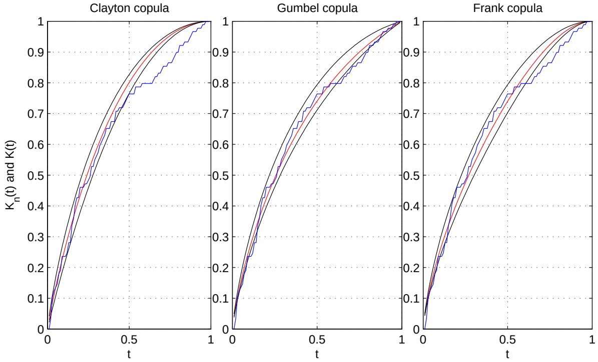
Figure: The functions $1 - F_V(v)$ and $1 - F_W(w)$ on a logarithmic scale.

Fitting copula

- Consider three bivariate Archimedean copulas: Clayton, Gumbel and Frank.
- Estimated parameters and 95% confidence intervals:

Copula	$\hat{\theta}$	95% confidence interval
Clayton	0.42	$[0.10, 0.74]$
Gumbel	1.44	$[1.21, 1.67]$
Frank	2.82	$[1.50, 4.14]$

Evaluation of the fit - comparison of $K_n(t)$ and $K(t;\hat{\theta})$



Clayton copula

E	p-value
0.00195	0.106

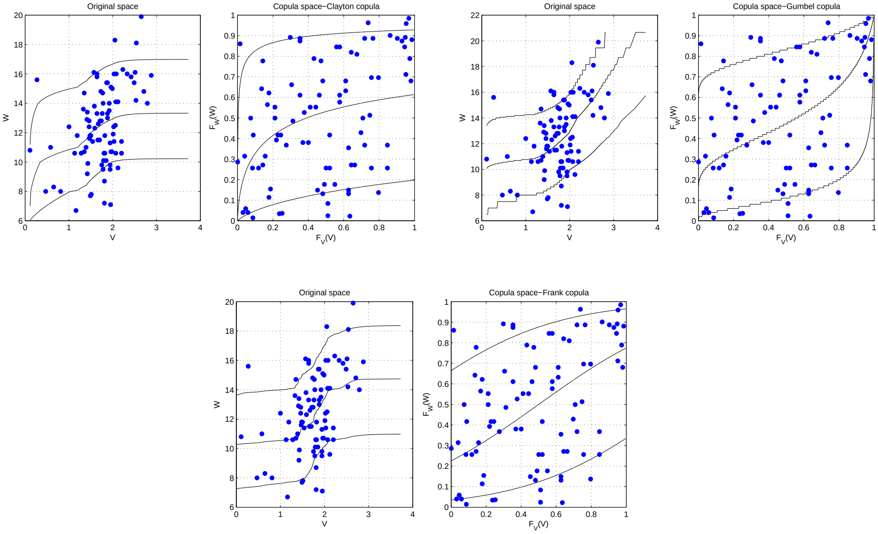
Gumbel copula

E	p-value
0.00094	**0.65**

Frank copula

E	p-value
0.0017	0.214

Evaluation of the fit - percentile lines



Transformation

- Consider random vector (V, W) with known F_V and F_W, and unknown $f_{V,W}$.
- Consider model (X, Y) with known F_X, F_Y and $f_{X,Y}$.
- Transform (V, W) to (X, Y):

Transformation

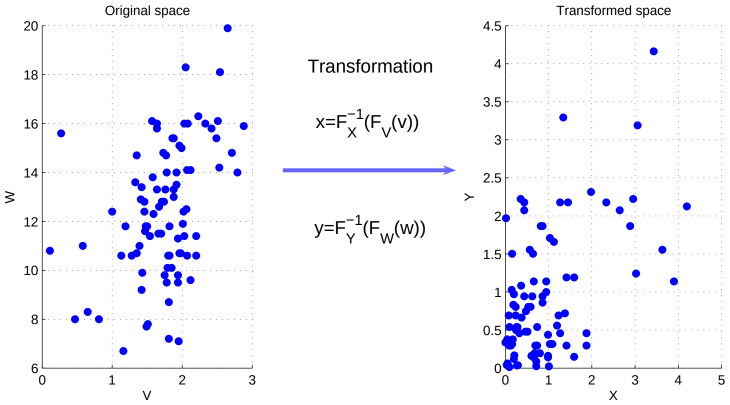
$$x(v) = F_X^{-1}(F_V(v)) \text{ and } y(w) = F_Y^{-1}(F_W(w))$$

- Then model $f_{V,W}$ is:

$$f_{V,W}(v, w) = \frac{f_{X,Y}(x(v), y(w)) f_V(v) f_W(w)}{f_X(x(v)) f_Y(y(w))}$$

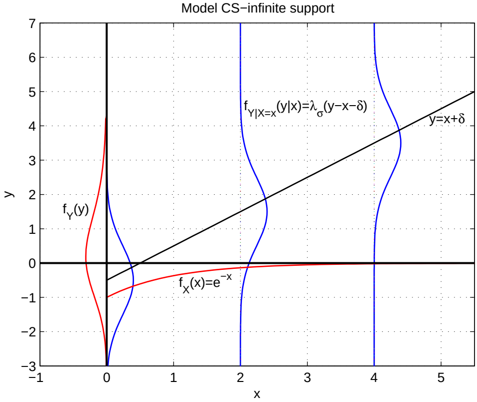
Transformation - example

- Consider model (X, Y) with $F_X(x) = 1 - e^{-x}$ and $F_Y(y) = 1 - e^{-y}$ and some $f_{X,Y}$.



Constant Spread Model (Model CS) - construction

- Model (X, Y)
- $f_X(x) = e^{-x}, x \geq 0$
- $f_{Y|X=x}(y|x) = \lambda_\sigma(y - x - \delta)$
- λ_σ - density function
- $E(Y|X = x) = x + \delta$
- $\sigma > 0$ - standard deviation
- $f_{X,Y}(x, y) = e^{-x}\lambda_\sigma(y - x - \delta)$
- $f_Y(y)$ - "exponential"



Model CS - tail dependence and related copula

Lower tail dependence coefficient

$$\lambda^L = \lim_{u \downarrow 0} P\{Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)\}$$

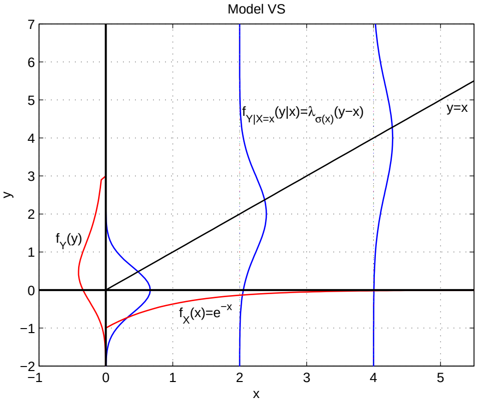
Upper tail dependence coefficient

$$\lambda^U = \lim_{u \uparrow 1} P\{Y > F_Y^{-1}(u) | X > F_X^{-1}(u)\}$$

- $\lambda^L = 0$
- If λ_σ is a normal density then $\lambda^U = 2\Phi(-\sigma/2)$, Φ - cdf of standard normal variable.
- If λ_σ is a normal density then the related copula does not belong to the Archimedean class.

Variable Spread Model (Model VS) - construction

- Model (X, Y)
- $f_X(x) = e^{-x}, x \geq 0$
- $f_{Y|X=x}(y|x) = \lambda_{\sigma_s(x)}(y - x)$
- $\lambda_{\sigma_s(x)}$ - density function
- $E(Y|X = x) = x$
- $\sigma_s(x) > 0$ - "spread function"
- $f_{X,Y}(x, y) = e^{-x}\lambda_{\sigma_s(x)}(y - x)$



Model VS - tail dependence and related copula

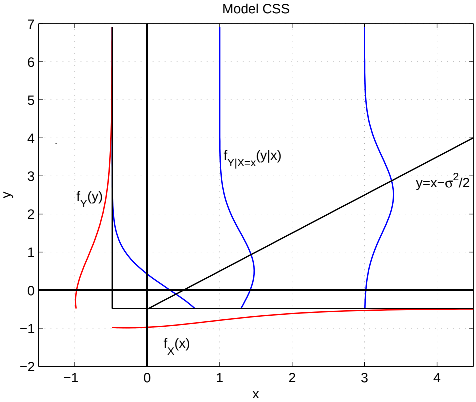
- If $\lambda_{\sigma_s(x)}$ is a normal density and $\sigma_s(x) > 0$ is increasing and $\lim_{x\to\infty} \sigma_s(x) = \sigma_1 < \infty$:

$$\lambda^L = 0 \quad \text{and} \quad \lambda^U = 2\Phi(-\sigma_1/2)$$

- If $\lambda_{\sigma_s(x)}$ is a normal density and $\sigma_s(x) = 2 + 0.5x$ then the related copula does not belong to the Archimedean class.

Constant Symmetric Spread Model - construction

- Model (X, Y)
- $f_X(x)$, $f_Y(y)$ - assym. exp.
- $f_{Y|X=x}(y|x)$ - modified normal
- $\sigma > 0$ - spread



Model CSS - tail dependence and related copula

- The lower tail independence occurs, because $\lambda^L = 0$.
- The upper tail dependence occurs, because $\lambda^U = 2\Phi(-\sigma/2)$.
- The related copula function is "close" to the Archimedean class (numerical experiments).

Model CS - case study

- Assume that λ_σ is a normal density, $\delta = 0$, the unknown parameter is σ.

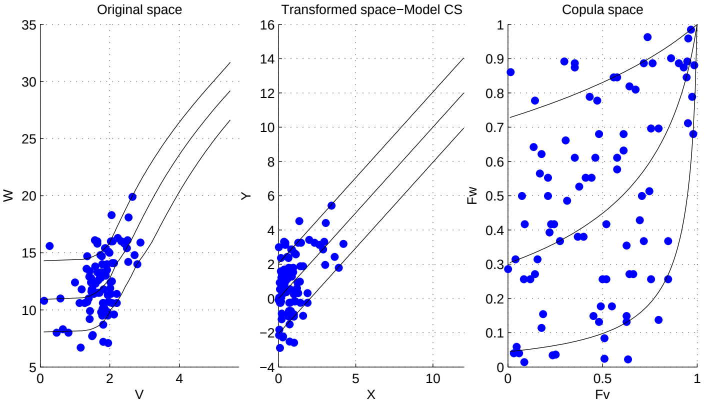


Figure: Model CS, $\hat{\sigma} = 1.6$

Model VS - case study

- Assume that $\lambda_{\sigma_s(x)}$ is a normal density and the spread function $\sigma_s(x)$ is:

$$\sigma_s(x) = \begin{cases} \frac{\sigma}{5}x + \sigma & \text{for } x \in [0,5] \\ 2\sigma & \text{for } x > 5 \end{cases}$$

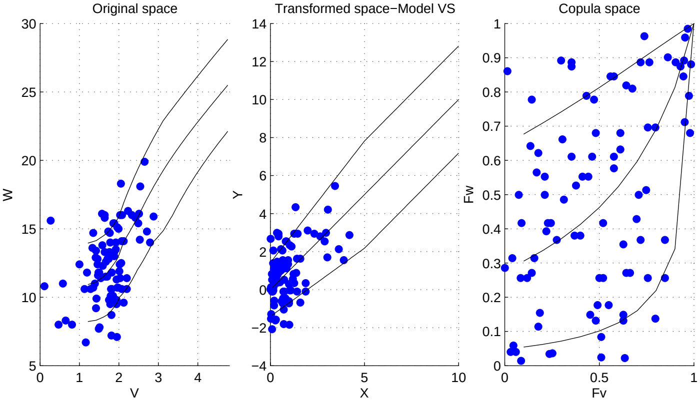


Figure: Model VS, $\hat{\sigma} = 1.1$.

Model CSS - evaluation of the fit

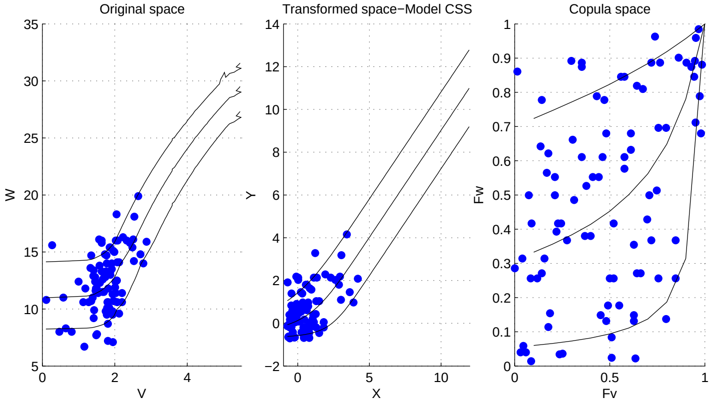
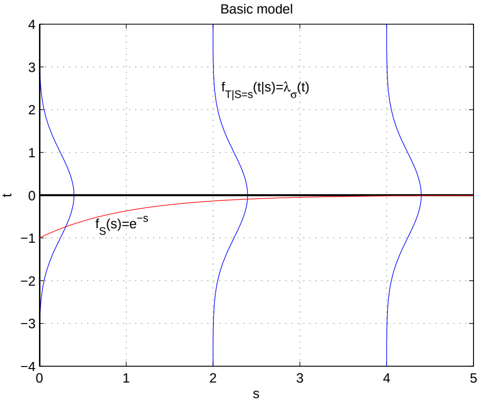


Figure: Model CSS, $\hat{\sigma} = 1.4$.

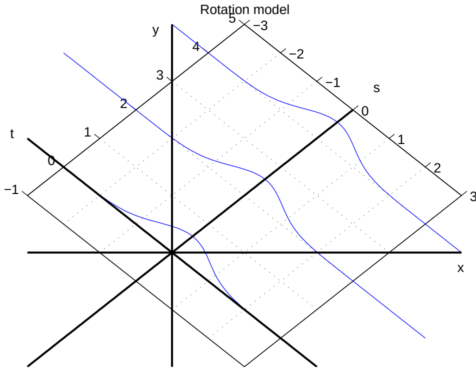
Rotation Model - construction

- Model (S, T)
- $f_{S,T}(s,t) = e^{-s}\lambda_\sigma(t), s \geq 0$
- λ_σ - density function
- $\sigma > 0$ - standard deviation



Rotation Model - construction

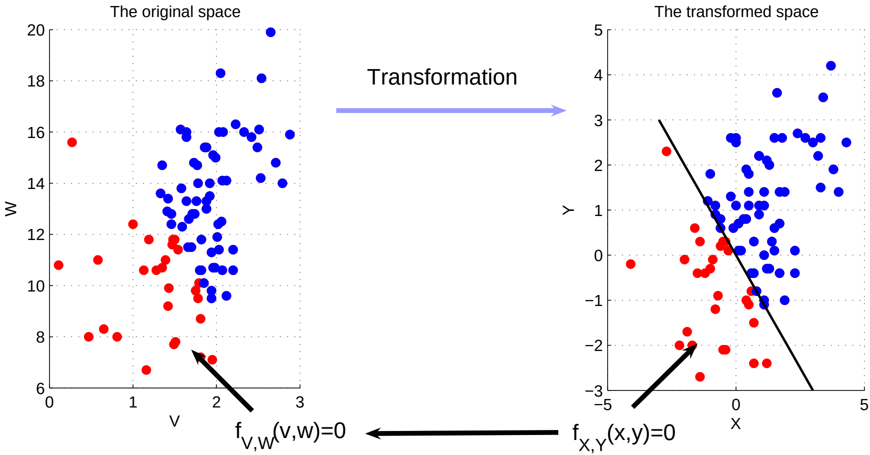
- Rotate space (X, Y) to (S, T)
- $f_{X,Y}(x,y) = e^{-\frac{x+y}{\sqrt{2}}} \lambda_\sigma\left(\frac{-x+y}{\sqrt{2}}\right)$
- λ_σ - density function
- $\sigma > 0$ - standard deviation



Rotation Model - properties

- If λ_σ has finite support then $f_X(x)$ and $f_Y(y)$ become shifted exponential.
- If λ_σ is a normal density then $f_{Y|X=x}(y|x)$ is a modified normal density.
- The definition of the model always entails lower tail independence $\lambda^L = 0$.
- If λ_σ is a normal density then $\lambda^U = 2\Phi(-\sigma)$.

Case study - problem



Conclusions and recommendations

- Gumbel copula is a good model to the considered dataset.
- It is a bit difficult to judge the fit in the extreme region in the copula space using the percentile lines method.
- The lower tail independence and upper tail dependence occur for the considered conditional models.
- It is relatively easy to judge the fit in the extreme region in the model space using the percentile lines method.
- The Rotation Model is not always suitable.
- Further work with evaluation of the fit (AIC coefficient for models?).
- Further work with the Rotation Model (Some additional conditions?).

Thank you!