Ovarian Extensions

Florian Stijven

11-4-2022

Fit Copula Models

Two types of models are considered. The first three are copula based models that impose time orderings. The last three are copula based models that do not impose time orderings. The following copulas are considered: Normal, Clayton and Frank.

In the following table, results from the fitted models are shown. The maximized likelihood value can be compared within the first three, and last three models, but not across. This is because the data are different across these two sets. The best fitting model is nonetheless the Frank copula in both sets because the corresponding log likelihood is largest. The worst model is the normal copula in the first set, and the Clayton copula in the second.

The normal and Frank copula yield similar estimates for Kendall's tau and Spearman's rho. However, the Clayton copula results in smaller values for these measures of association. Because the Frank copula certainly fits the best, the corresponding measures of association are more trustworthy.

model	loglik	copula_par0	copula_par1	kendall0	kendall1	spearman0	spearman1
normal (ord)	-4185.41	0.91	0.95	0.72	0.79	0.90	0.94
Clayton (ord)	-4113.97	5.84	8.71	0.74	0.81	0.91	0.95
Frank (ord)	-4119.60	14.20	18.64	0.75	0.80	0.92	0.95
gumbel (ord)	-4172.09	3.78	4.82	0.74	0.79	0.90	0.94

	model	loglik	copula_par0	copula_par1	kendall0	kendall1	spearman0	spearman1
5	normal	-4177.76	0.93	0.97	0.77	0.83	0.93	0.96
6	Clayton	-4132.61	7.81	12.19	0.80	0.86	0.94	0.97
7	Frank	-4132.84	18.61	24.00	0.80	0.84	0.95	0.97
8	gumbel (ord)	-4135.12	4.68	6.14	0.79	0.84	0.94	0.96

Because the Frank copula provides the best fit based on the likelihood, this fitted model is examined next in more detail.

Goodness of fit

The goodness of fit is checked visually for the fitted frank copula model with time ordering. The marginal goodness of fit is checked by a comparison of the KM-estimates and model based estimates for the survival functions of PFS and OS in both treatment groups. This is shown in the next Figure. Clearly, the model based estimates follow the KM-estimates very closely.

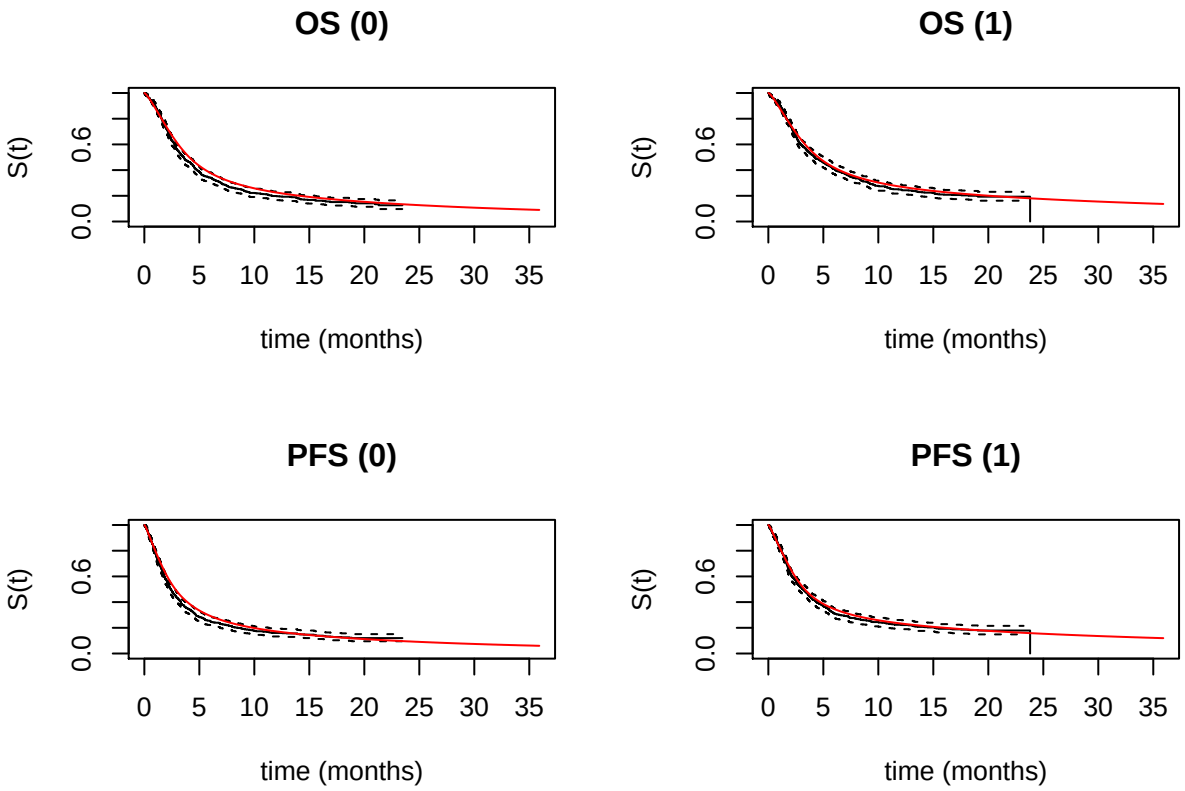


Figure 1: Marginal goodness of fit for Frank copula model with time ordering.

This is compared with the Frank copula model without time orderings. From the fitted models without time orderings, this is the best fitting model based on the likelihood. The fit is about equally good as the corresponding model with time orderings, which is certainly expected as the main difference between both models lies in the association, and not in the marginal survival functions.

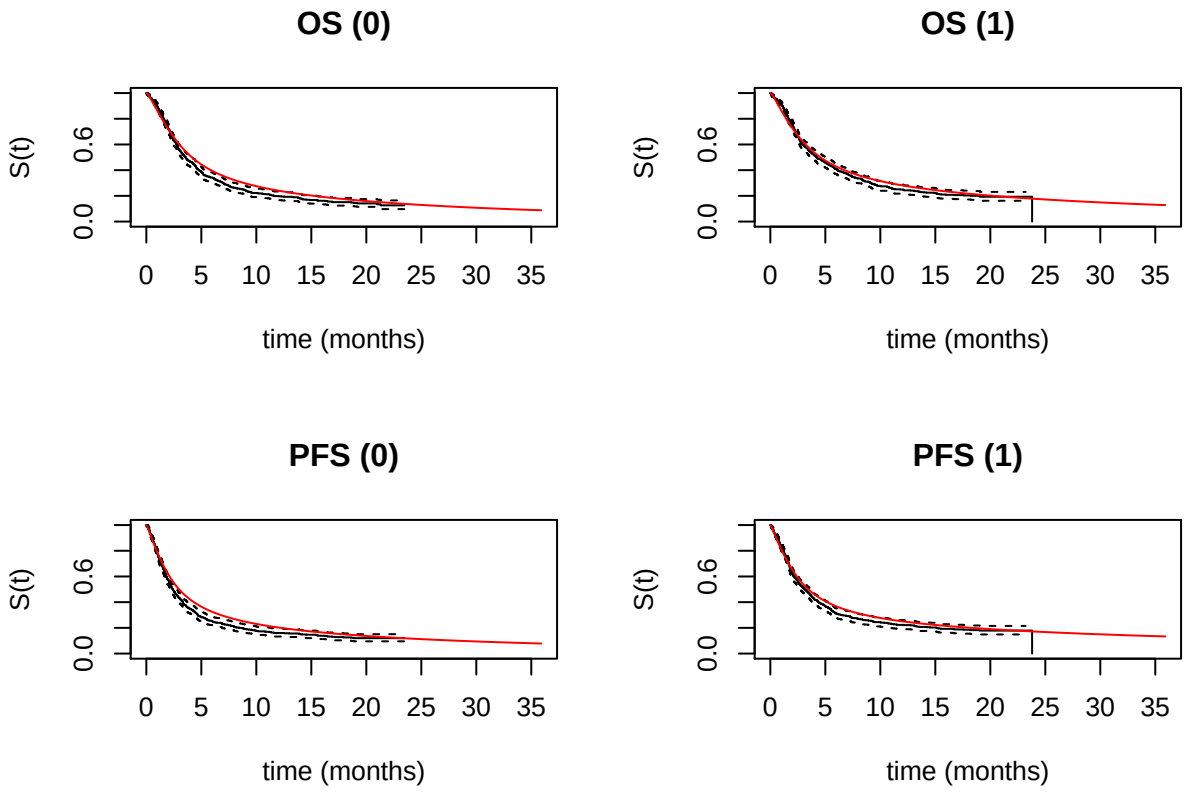


Figure 2: Marginal goodness of fit for Frank copula without time orderings

The issue with very long OS remains. This is less problematic if the measures of association further considered are based on ranks. Still, this remains an important issue. The existence of a cured fraction might further complicate the analysis.

The goodness of fit is also evaluated for the joint distribution of PFS and OS in both groups by sampling from the fitted model, and comparing this sample with the original data. However, the original data are censored which complicates a direct comparison. The distribution of censoring indicators is therefore first estimated. Next, censoring times are sampled from this distribution as to artificially censor the sample from the fitted model. In this way, the sample from the fitted model is comparable with the original data.

These plots indicate that the fit (at least the observable part) is good. Moreover, the fit is clearly better for the Frank copula model with time orderings than without time orderings. Explicitly taking into account the time-orderings thus results in a better fit. This does however not necessarily mean that the conclusions with respect to the strength of surrogacy are different.

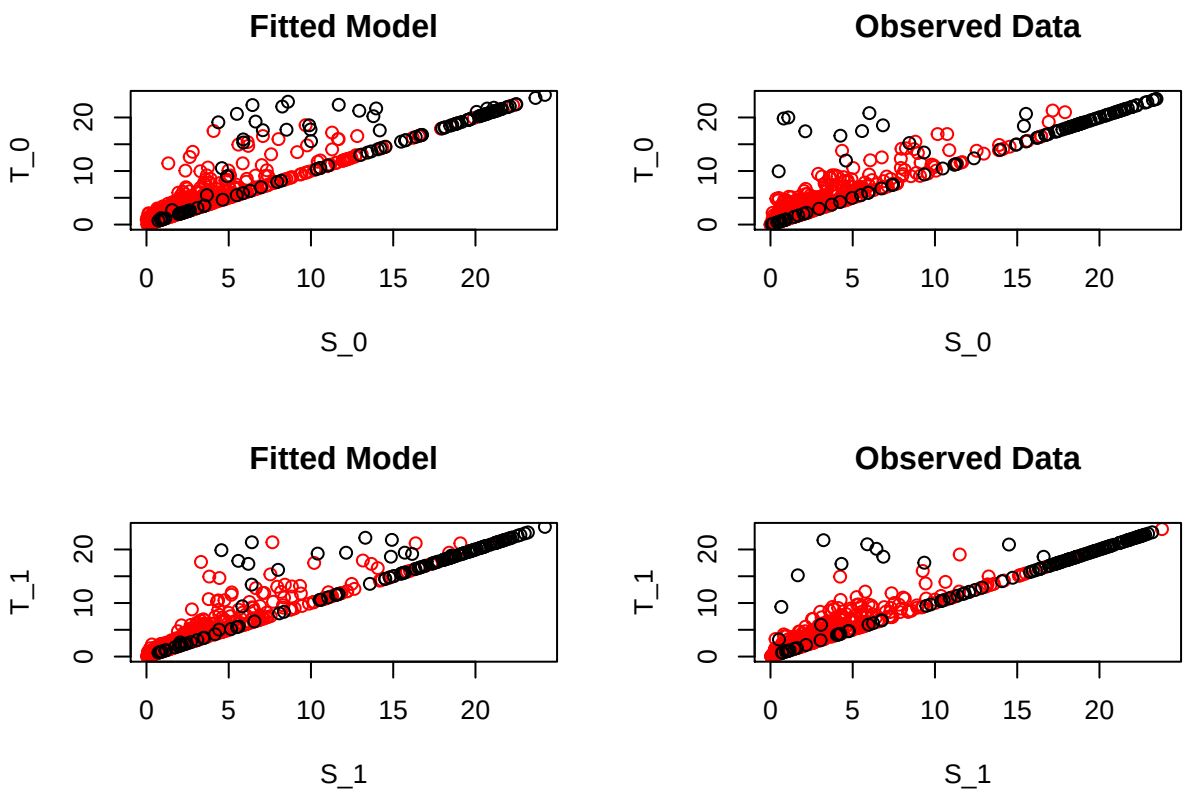


Figure 3: Sample from fitted model versus observed values. The fitted model is the Frank copula model with time orderings.

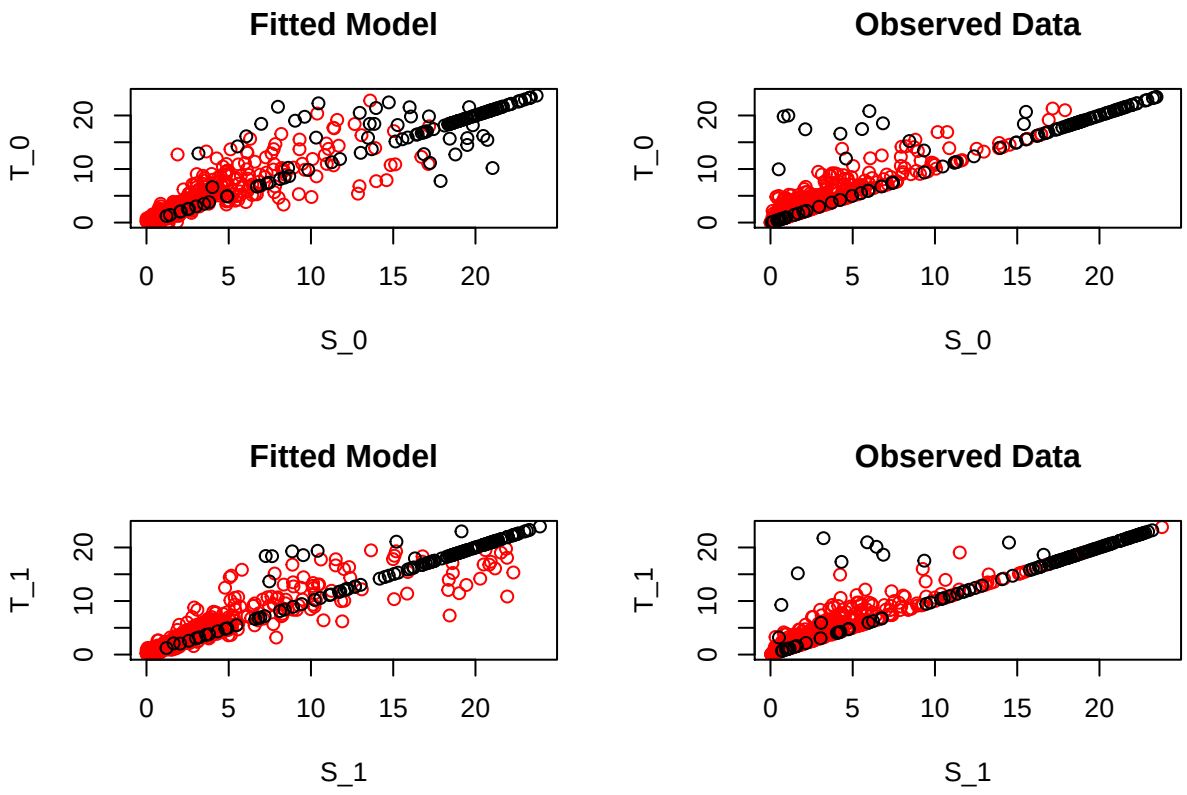
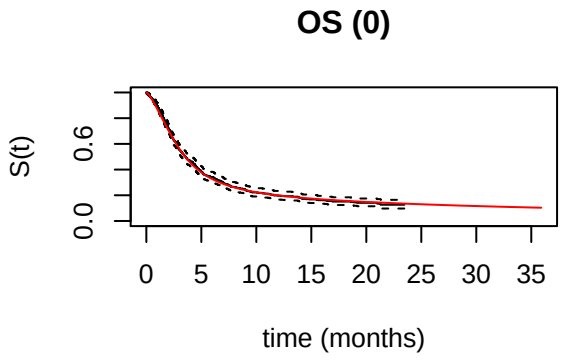
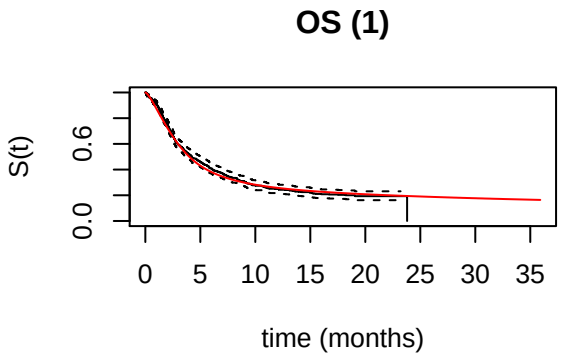
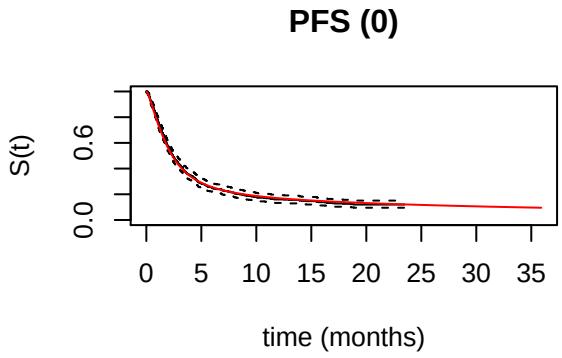


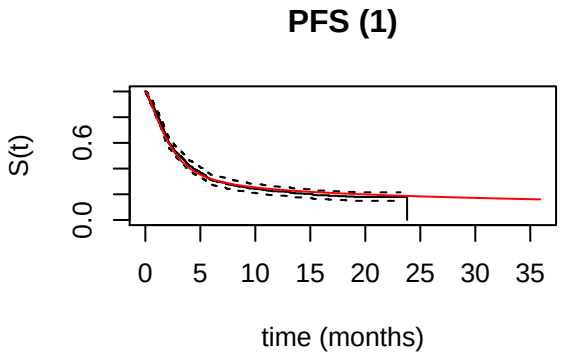
Figure 4: Sample from fitted model versus observed values. The fitted model is the Frank copula model without time orderings.

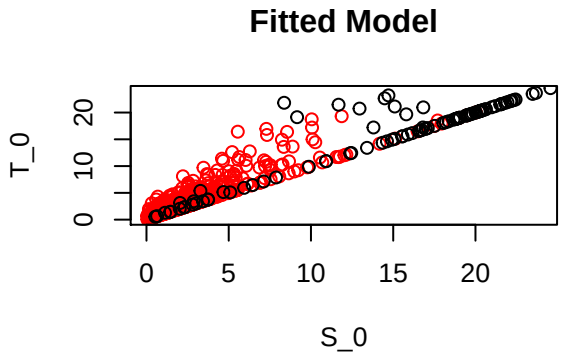
Goodness of fit for other models

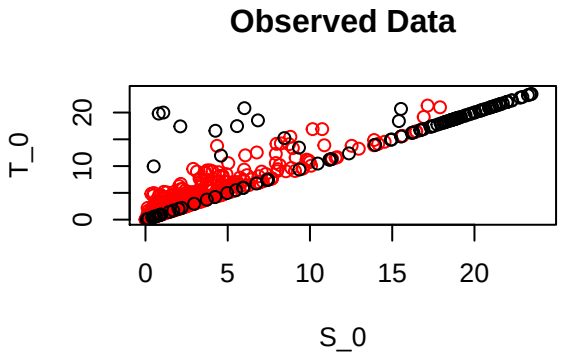


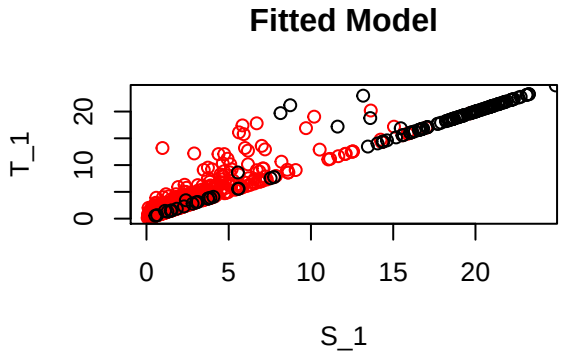


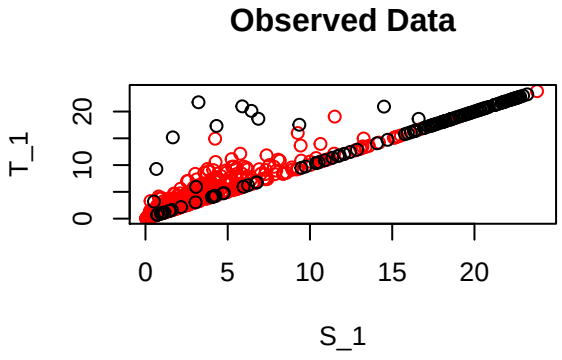


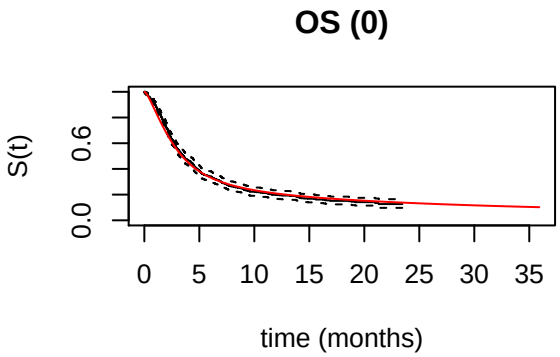


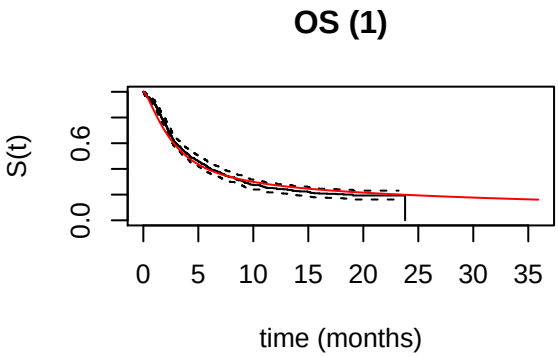


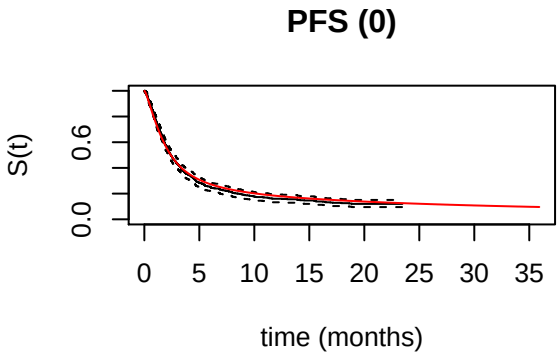


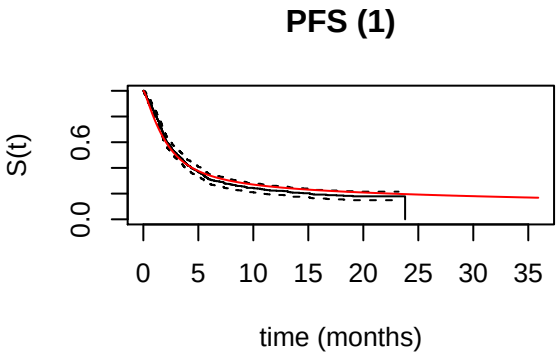


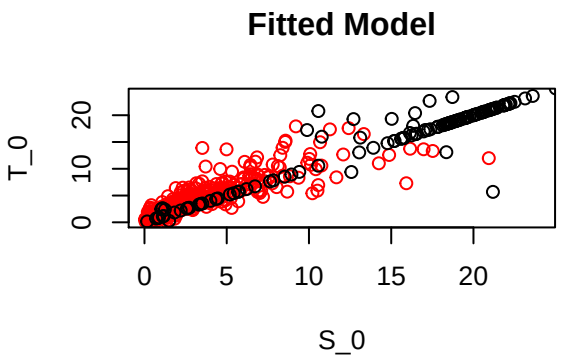


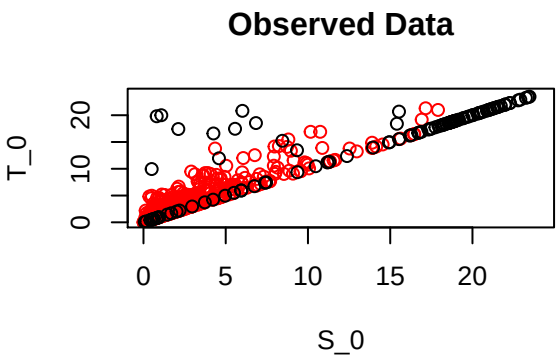

OS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

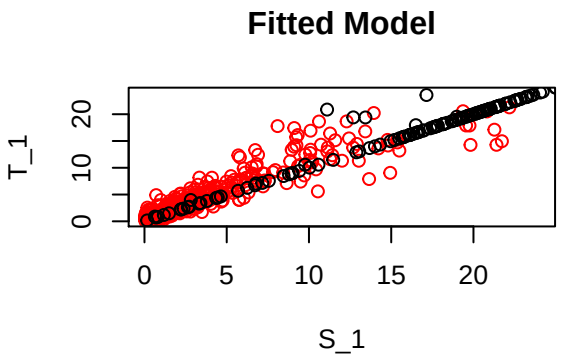

OS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

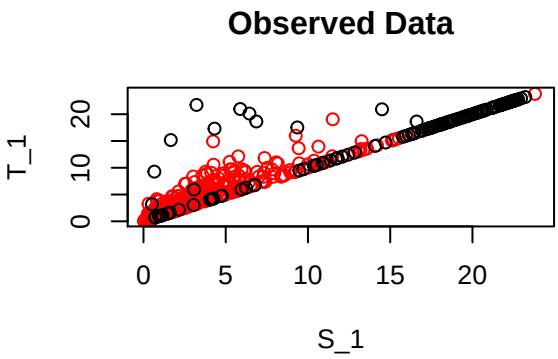

PFS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

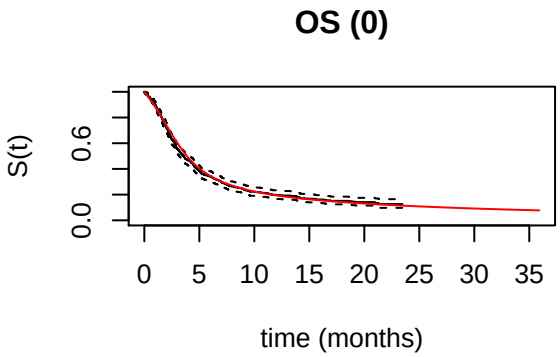

PFS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

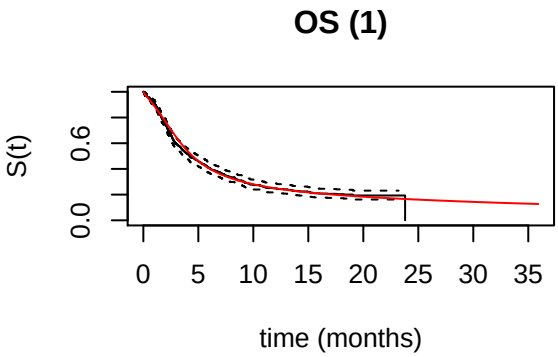

Fitted Model
T_0
S_0
0
5
10
15
20

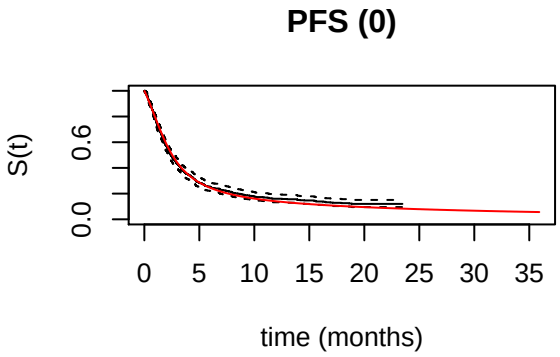

Observed Data
T_0
S_0
0
5
10
15
20

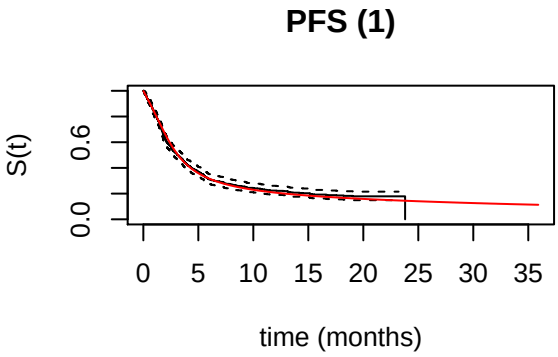

Fitted Model
T_1
S_1
0
5
10
15
20

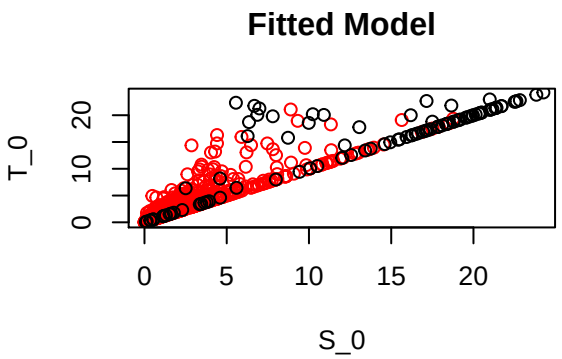

Observed Data
T_1
S_1
0
5
10
15
20

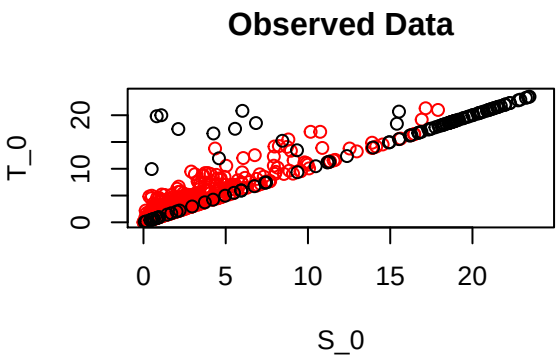

OS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

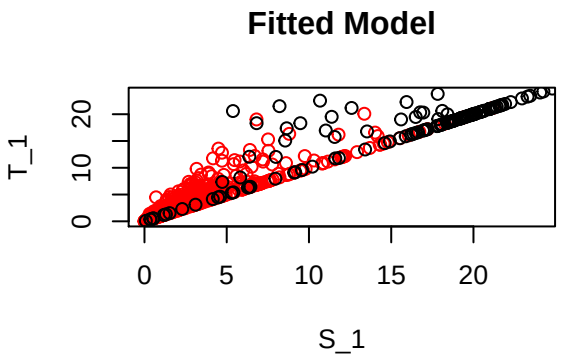

OS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

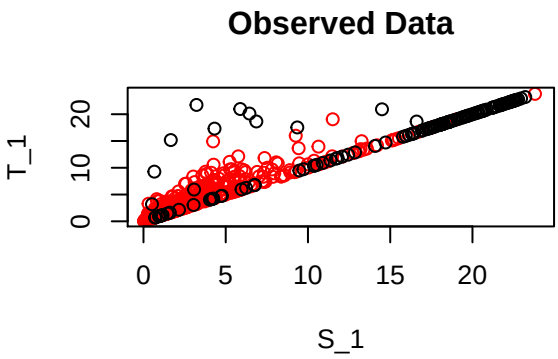

PFS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

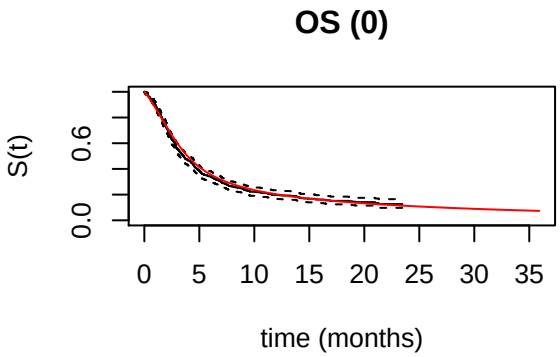

PFS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

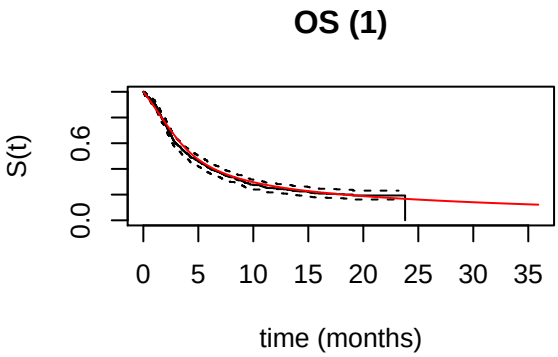

Fitted Model
T_0
0
10
20
0
5
10
15
20
S_0

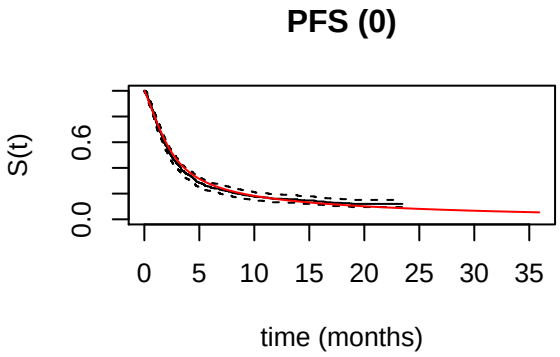

Observed Data
T_0
0
10
20
0
5
10
15
20
S_0

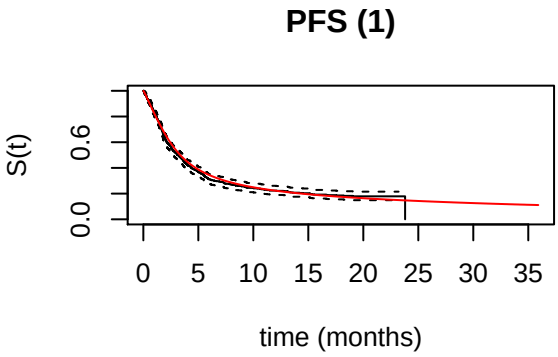

Fitted Model
T_1
0
10
20
0
5
10
15
20
S_1

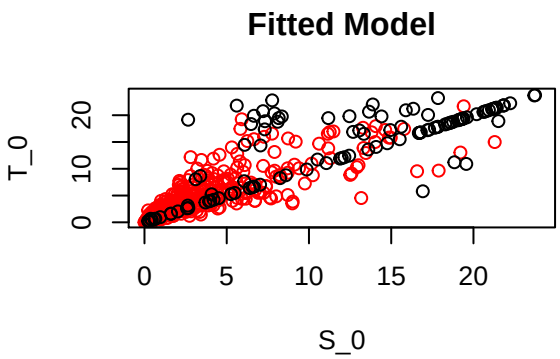

Observed Data
T_1
0
10
20
0
5
10
15
20
S_1

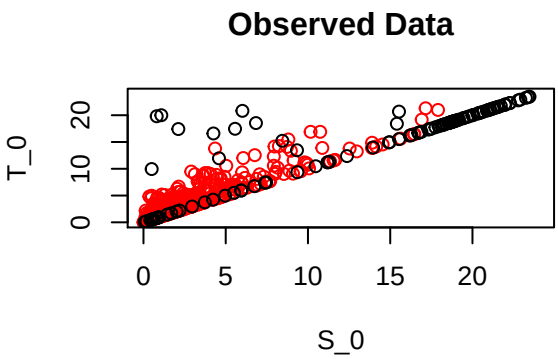

OS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

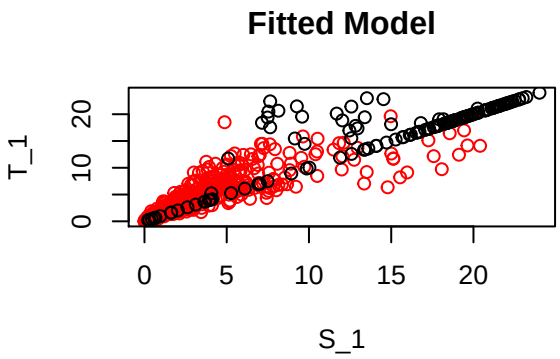

OS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

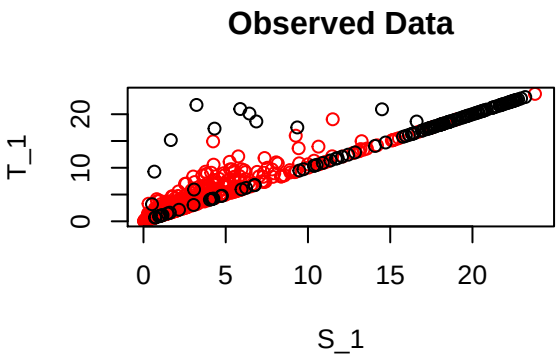

PFS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

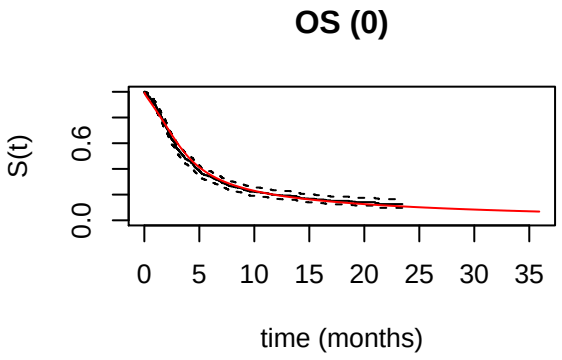

PFS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

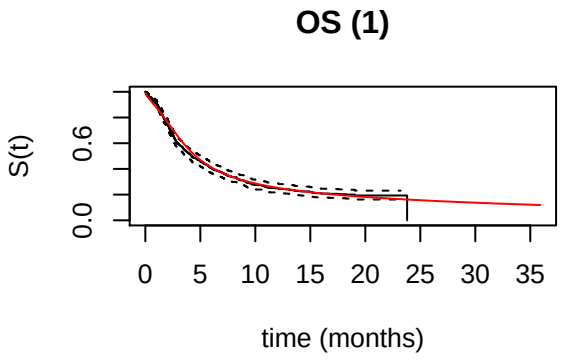

Fitted Model
T_0
S_0
0
5
10
15
20

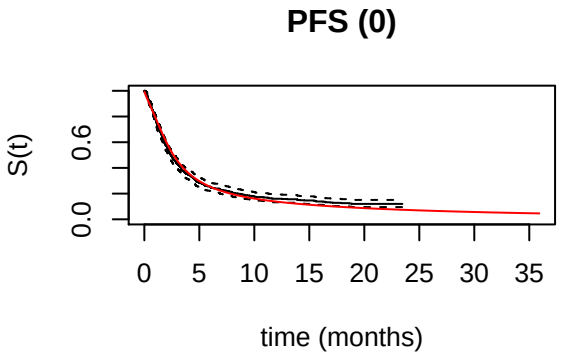

Observed Data
T_0
S_0
0
5
10
15
20

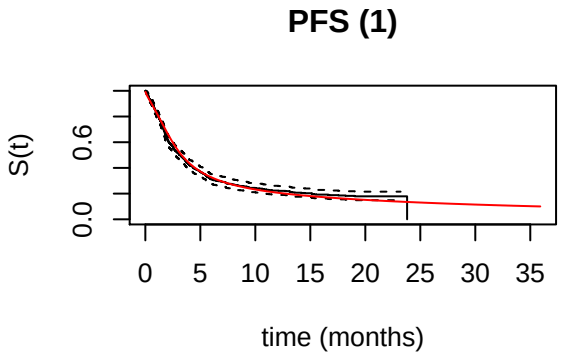

Fitted Model
T_1
S_1
0
5
10
15
20

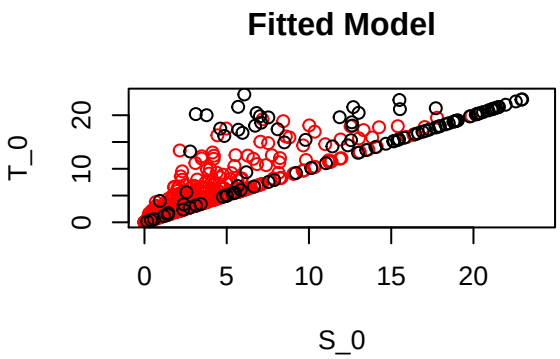

Observed Data
T_1
S_1
0
5
10
15
20

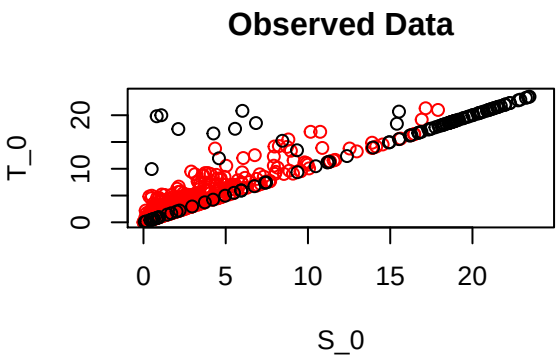

OS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

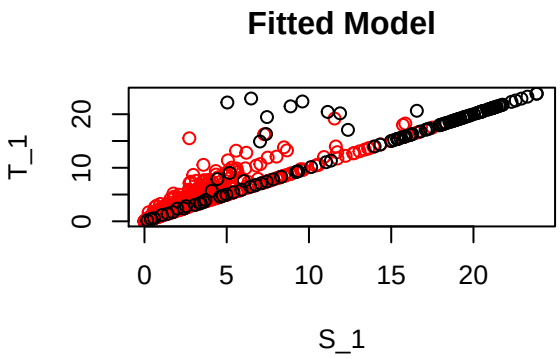

OS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

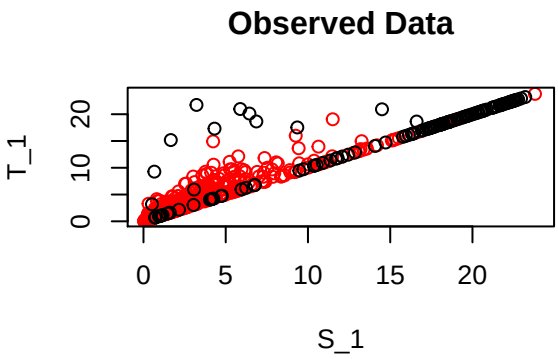

PFS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

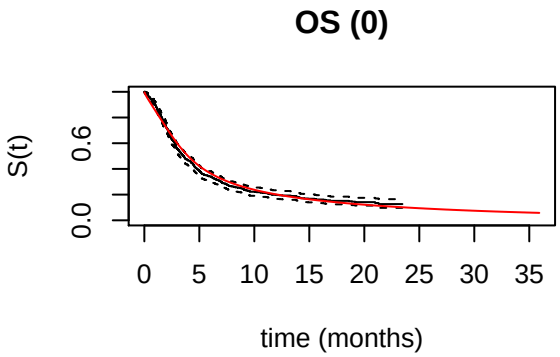

PFS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)

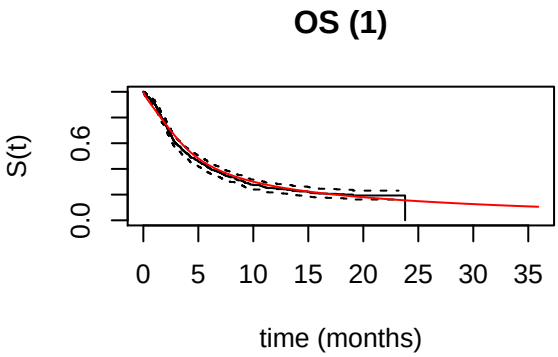

Fitted Model
T_0
S_0
0
5
10
15
20

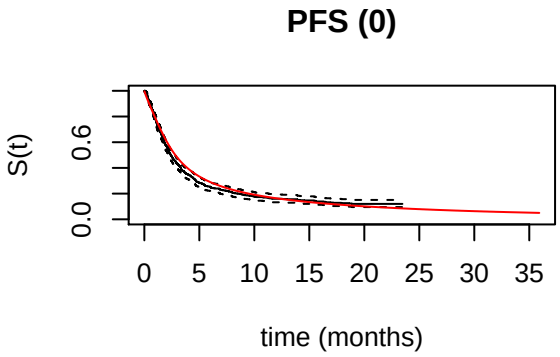

Observed Data
T_0
S_0
0
5
10
15
20

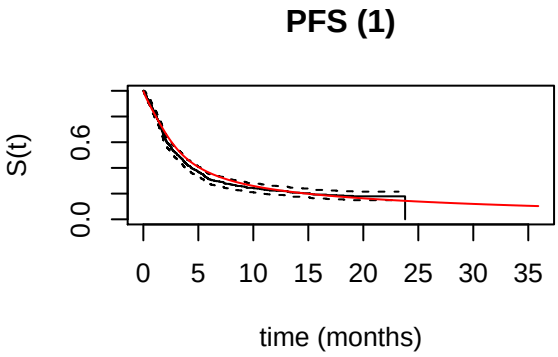

Fitted Model
T_1
S_1
0
5
10
15
20

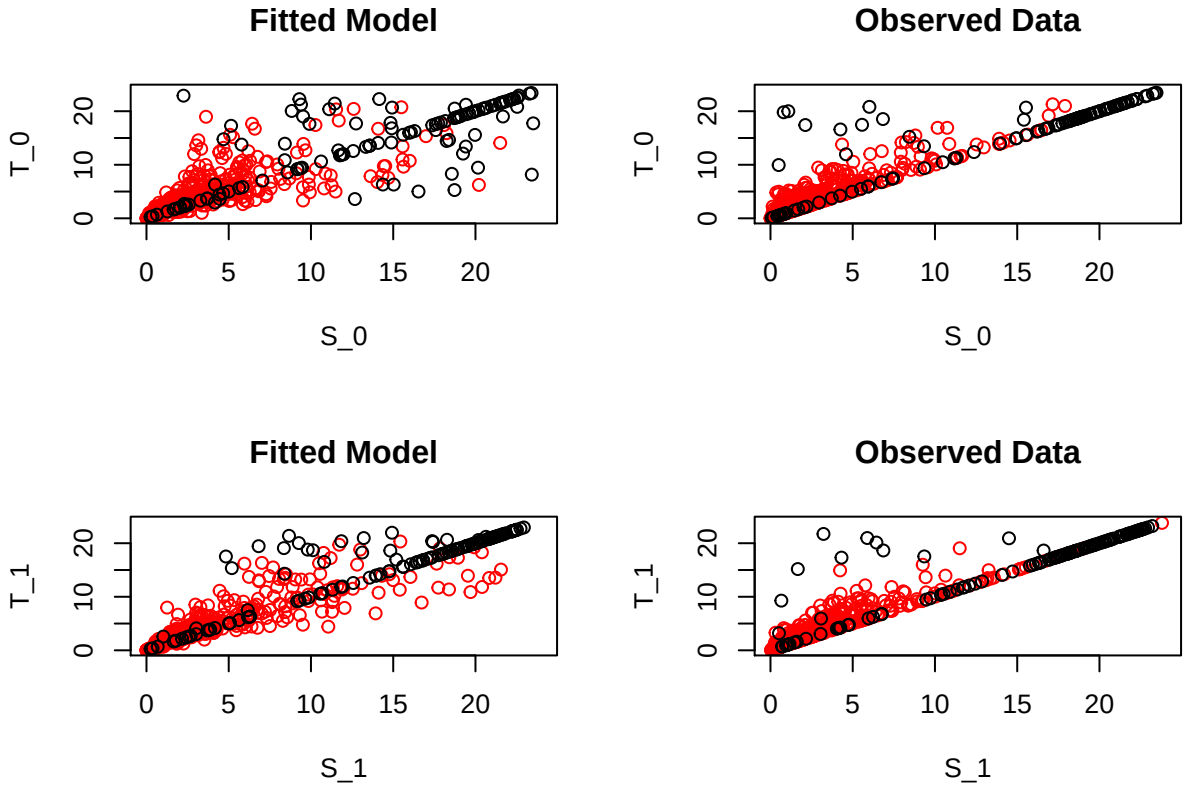

Observed Data
T_1
S_1
0
5
10
15
20


OS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)


OS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)


PFS (0)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)


PFS (1)
S(t)
0.0
0.6
0 5 10 15 20 25 30 35
time (months)



Measures of Surrogacy

The parameter for the Frank copula is defined on the entire real line. So sampling from a uniform distribution is not possible as can be done for the gaussian copula. This is addressed by sampling from a uniform distribution with as bounds (minus) the observed parameter. This is reasonable because we would not expect the association to be stronger between for example S_0 and T_1, than between S_0 and T_0.

Note that the Frank copula can model both negative and positive associations. This is also the case for the gaussian copula, but not for the Clayton copula. This is important because some of the constituent copulas in the vine copula are not identifiable. If we were to use a copula that can only model positive association in the construction of the vine copula, then we would (implicitly) assume that all observable associations are positive. Such an assumption is unverifiable, and possibly difficult to argue for based on scientific grounds. The is however an immediate solution for such copulas. Besides making the parameters of such unobservable copulas part of the sensitivity analysis, rotating the copulas by 90, 180 and 270 degrees is also part of the sensitivity analysis. If a copula that can only model positive association is rotated by 90 or 270 degrees, it can only model negative association. Such a sensitivity analysis would thus not anymore impose the constraint of a positive unobservable association.

Two definitions for the individual causal effects are considered. The first one is based on the difference: $\Delta S = S_1 - S_0$ and $\Delta T = T_1 - T_0$. This definition has been used in the literature for the gaussian-gaussian case. For a multivariate normal distribution, this is a natural and convenient definition. For survival outcomes, the following definition is also possible: $\Delta S = \frac{S_1}{S_0}$ and $\Delta T = \frac{T_1}{T_0}$ (or the log-ratios). This is related to the specifiation of effects in a accelerated failure time model.

Two measures of surrogacy are considered based on ranks: Kendall's tau and Spearman's correlation. Computing these measures for a bivariate distribution requires integration. However, the joint distribution of ΔS and ΔT is very complicated. Nonetheless, it is easy to sample from this distribution. These measures are

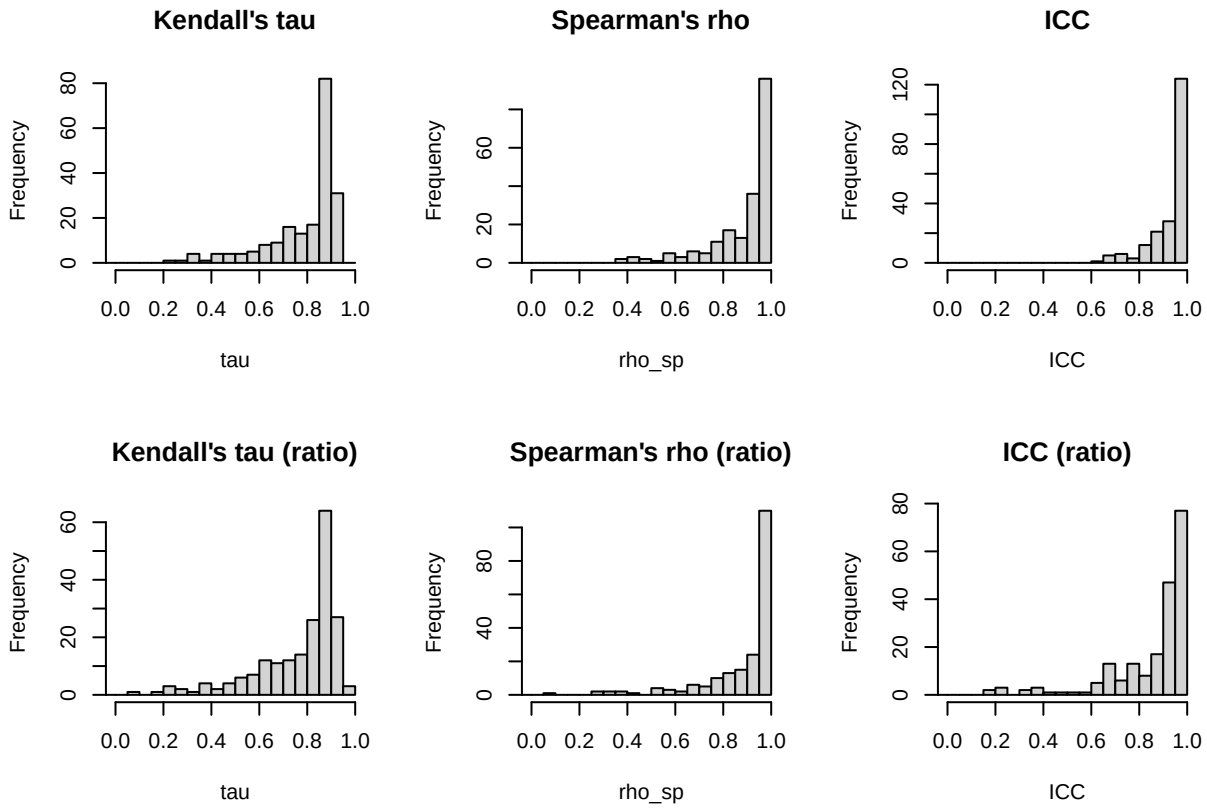
therefore computed by Monte Carlo integration. This is straightforward. For each sampled positive definite matrix, 10.000 samples are drawn from the corresponding model. Kendall's tau and spearman's rho are consequently estimated based on this sample.

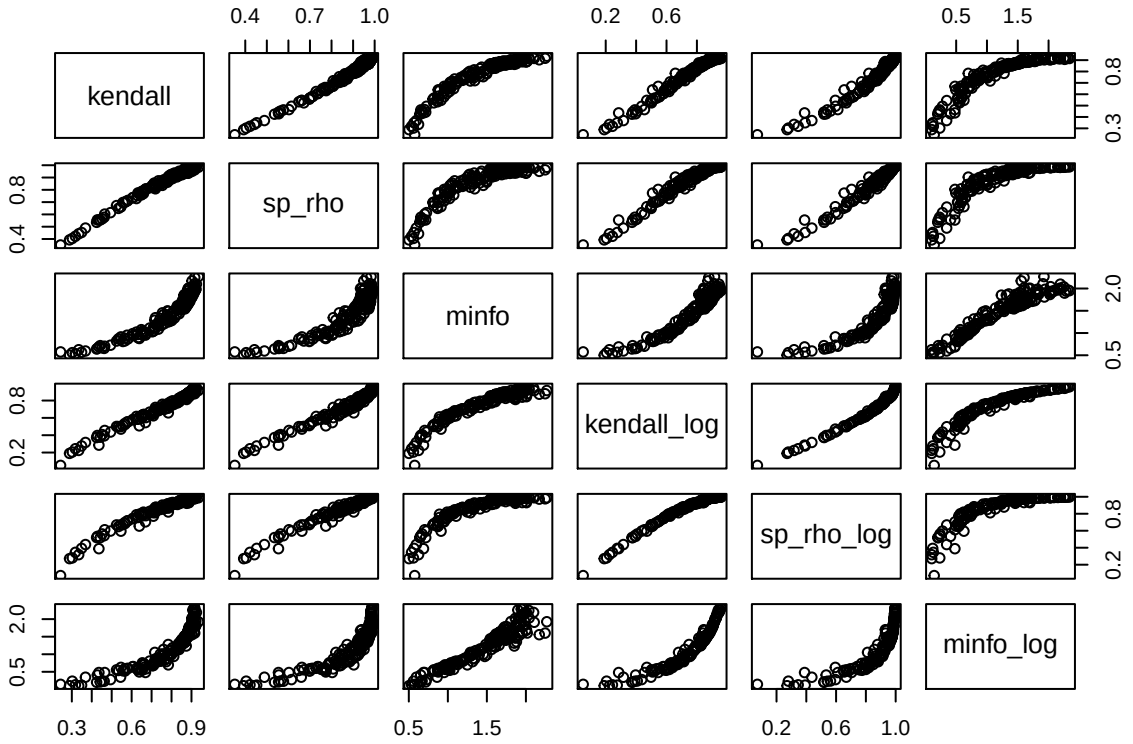
$$\rho_{sp} = cor(R(\Delta S), R(\Delta T))$$

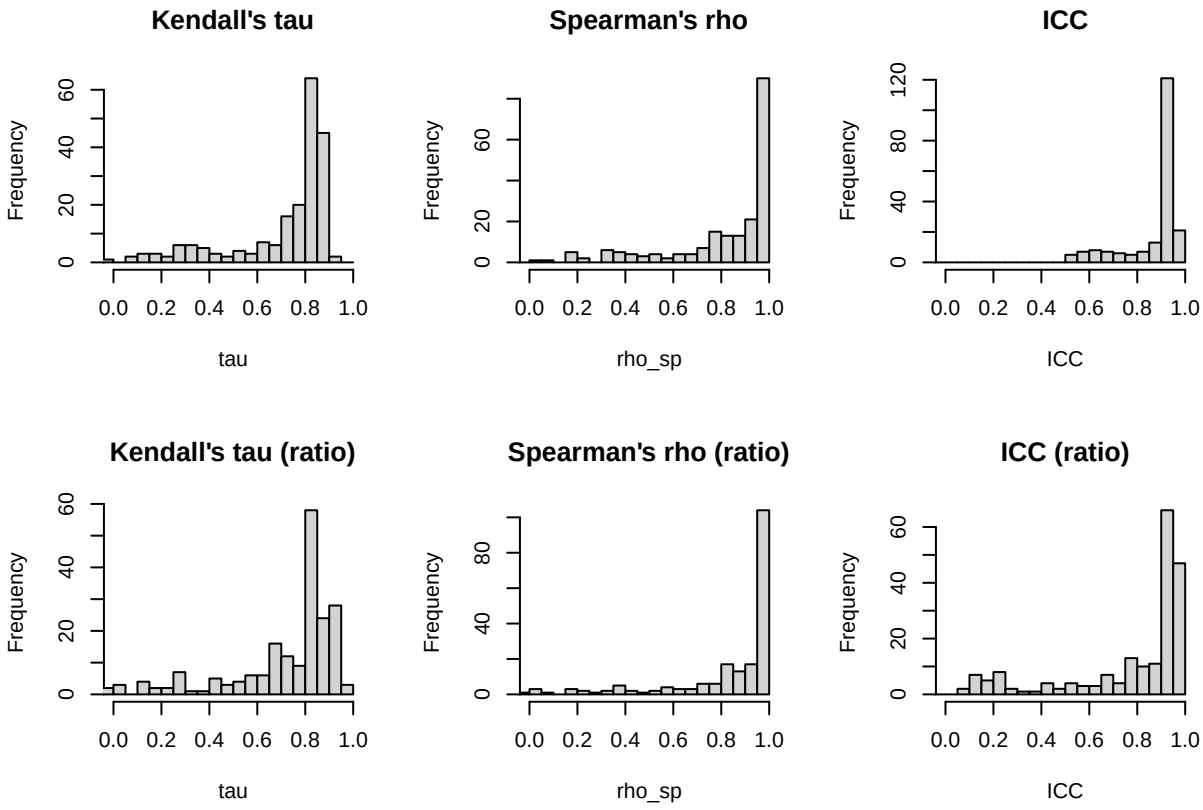
$$\tau = P((\Delta S_1 - \Delta S_2)(\Delta T_1 - \Delta T_2) > 0) - P((\Delta S_1 - \Delta S_2)(\Delta T_1 - \Delta T_2) < 0)$$

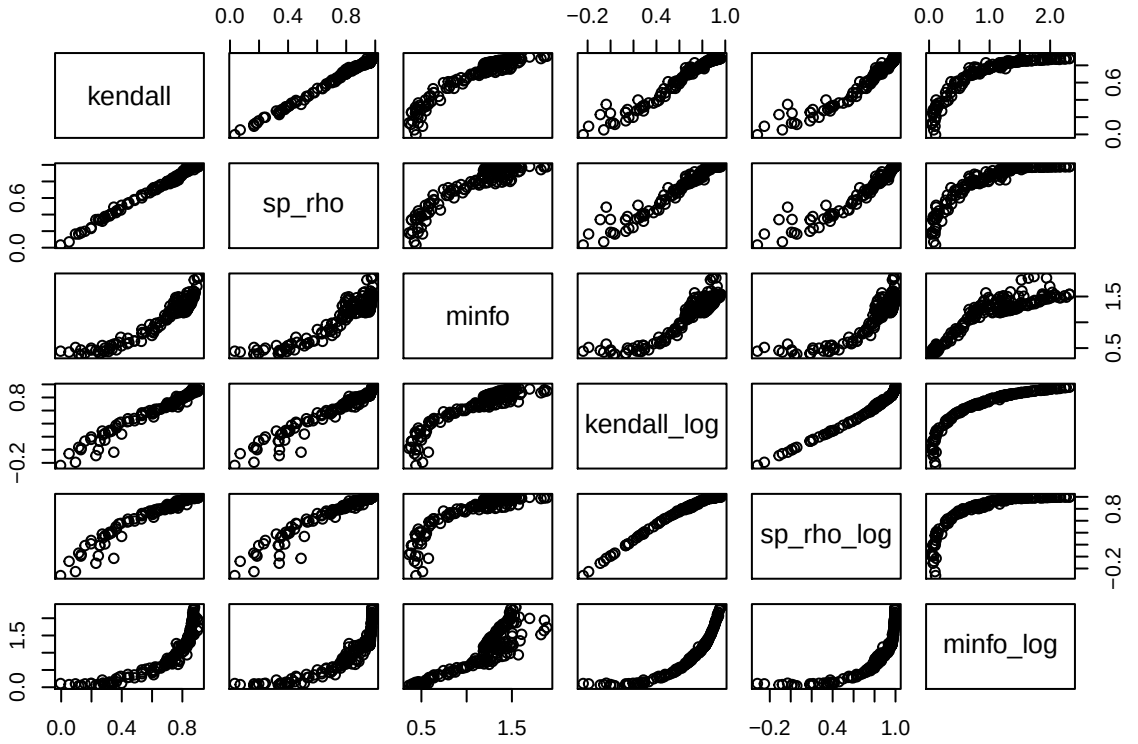
Frank copula model

The results of the sensitivity analysis of the Frank copula model with and without time orderings are shown in the following plots.



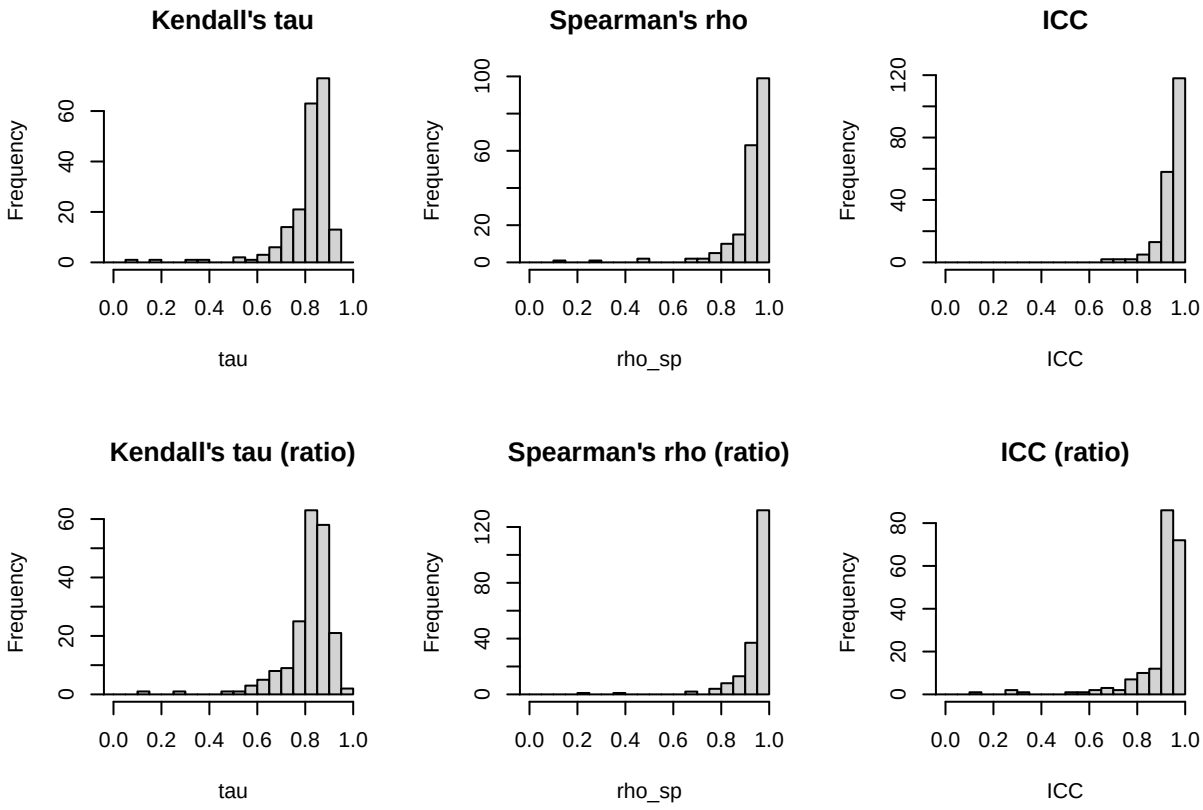

kendall
sp_rho
minfo
kendall_log
sp_rho_log
minfo_log
0.4 0.7 1.0
0.2 0.6
0.5 1.5
0.3 0.6 0.9
0.5 1.5
0.2 0.6 1.0
0.3 0.8
0.4 0.8
0.5 2.0
0.2 0.8
0.2 0.8
0.5 2.0

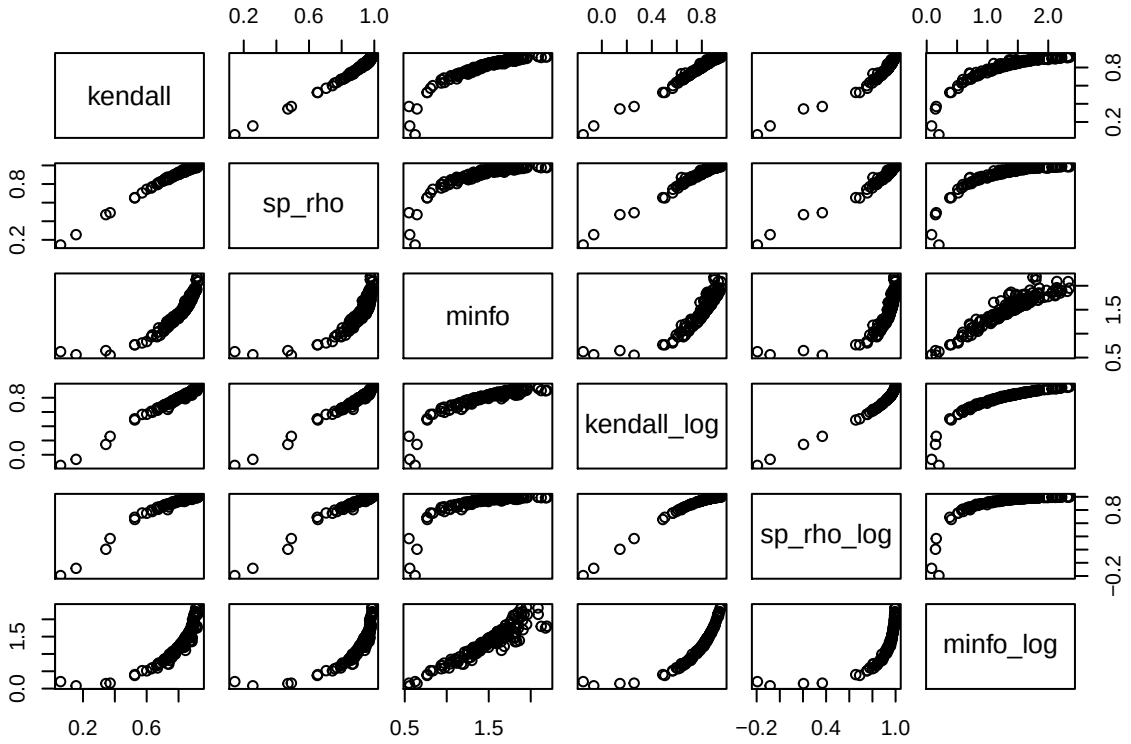

Kendall's tau
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho
Frequency
0
20
60
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC
Frequency
0
40
80
120
0.0
0.2
0.4
0.6
0.8
1.0
ICC
Kendall's tau (ratio)
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho (ratio)
Frequency
0
40
80
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC (ratio)
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
ICC

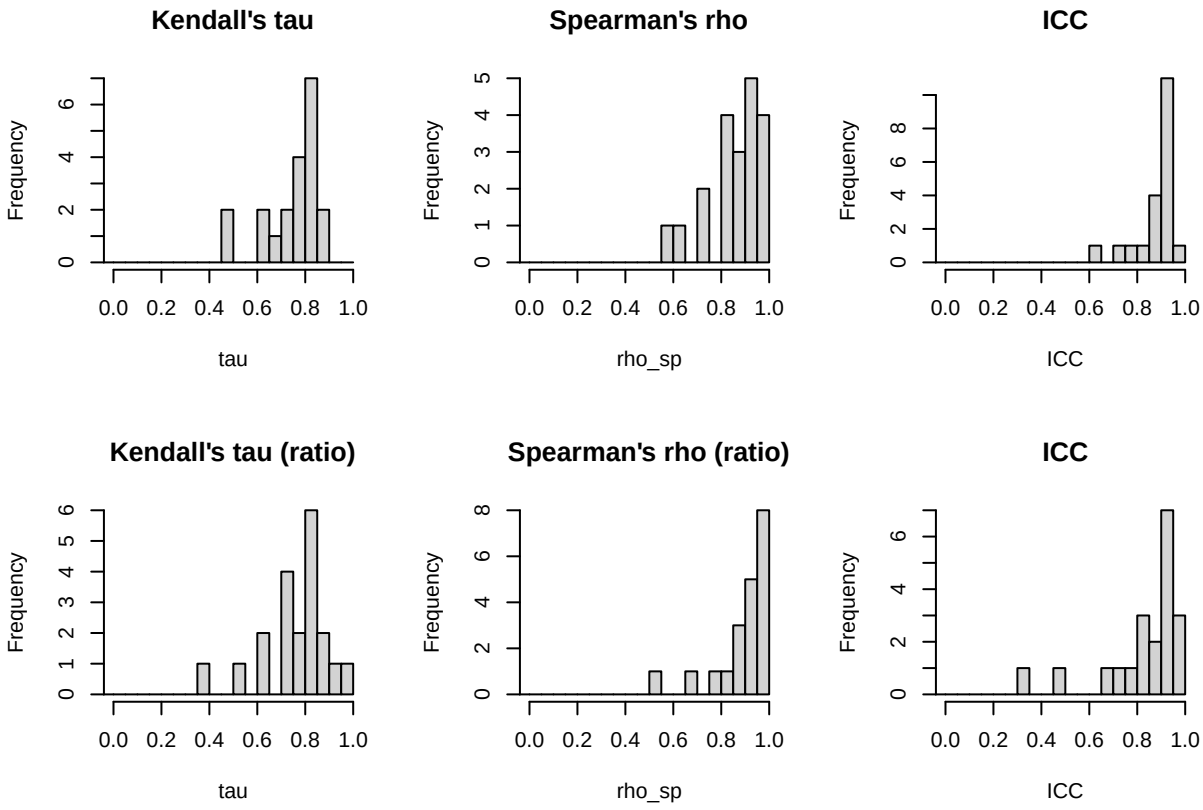


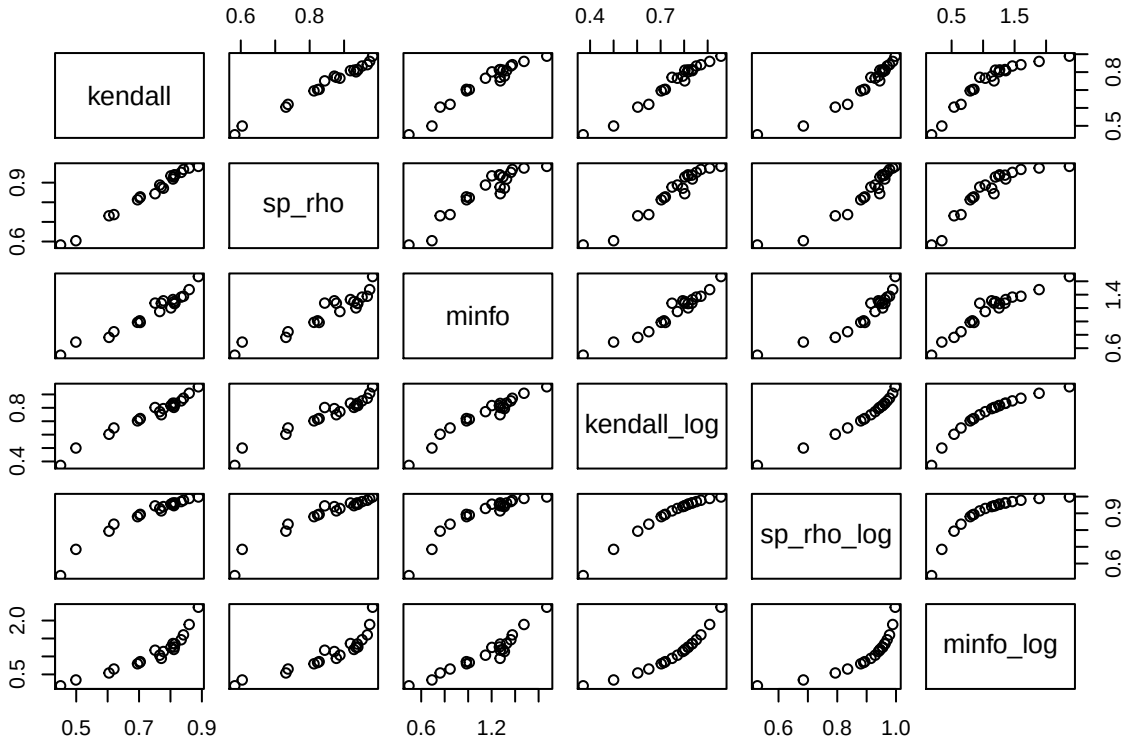
Gaussian Copula Model

The sensitivity analysis is repeated for the gaussian copula model because this is the model that was originally proposed. Moreover, this is a somewhat simpler model that can be seen independent of the vine copula formulation.

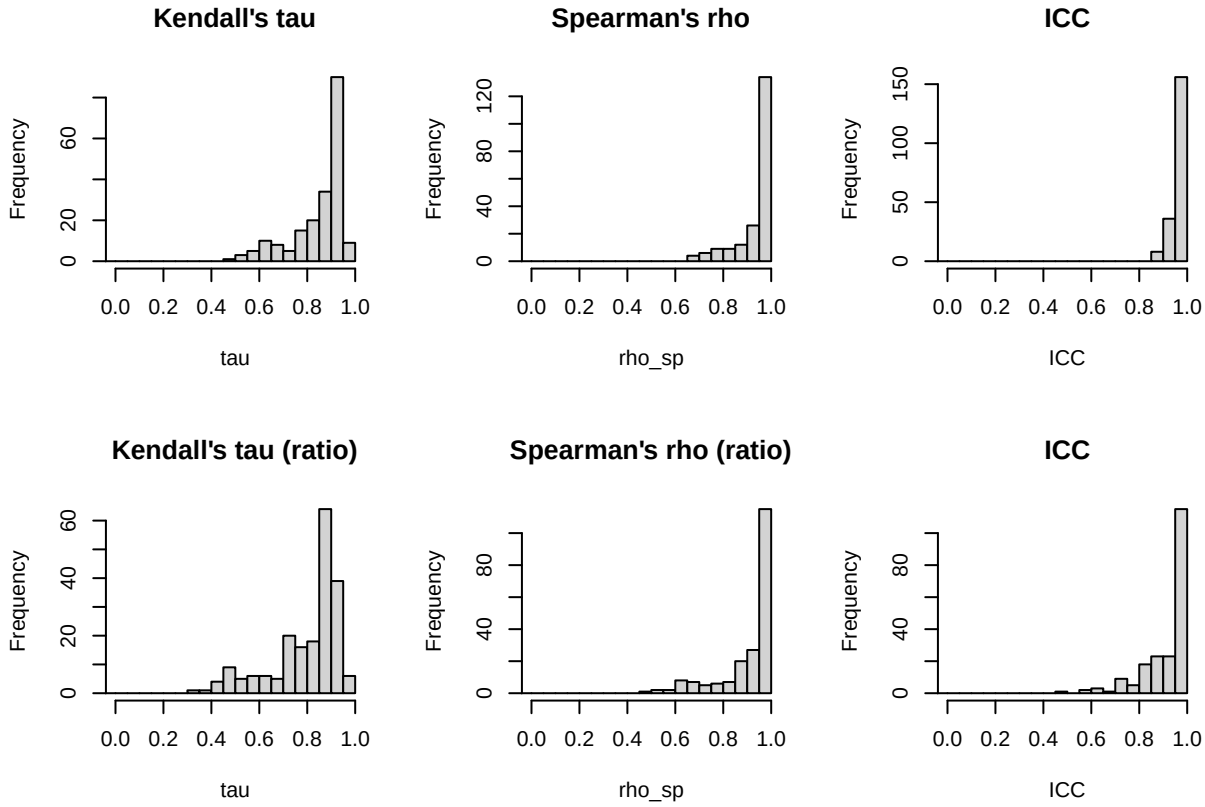

Kendall's tau
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho
Frequency
0
20
60
100
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC
Frequency
0
40
80
120
0.0
0.2
0.4
0.6
0.8
1.0
ICC
Kendall's tau (ratio)
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho (ratio)
Frequency
0
40
80
120
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC (ratio)
Frequency
0
20
40
60
80
0.0
0.2
0.4
0.6
0.8
1.0
ICC

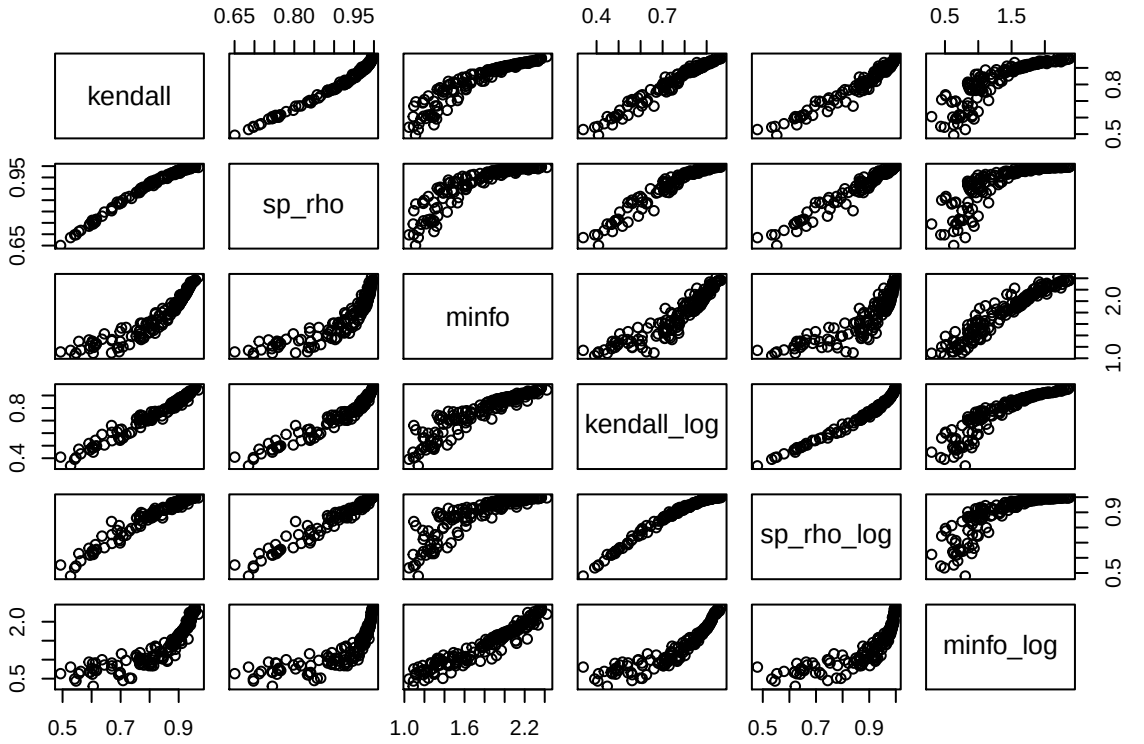

kendall
sp_rho
minfo
kendall_log
sp_rho_log
minfo_log
0.2
0.6
1.0
0.0
0.4
0.8
0.0
1.0
2.0
0.2
0.8
0.2
0.8
0.5
1.5
0.0
0.8
0.8
−0.2
0.0
1.5
0.2
0.6
0.5
1.5
−0.2
0.4
1.0

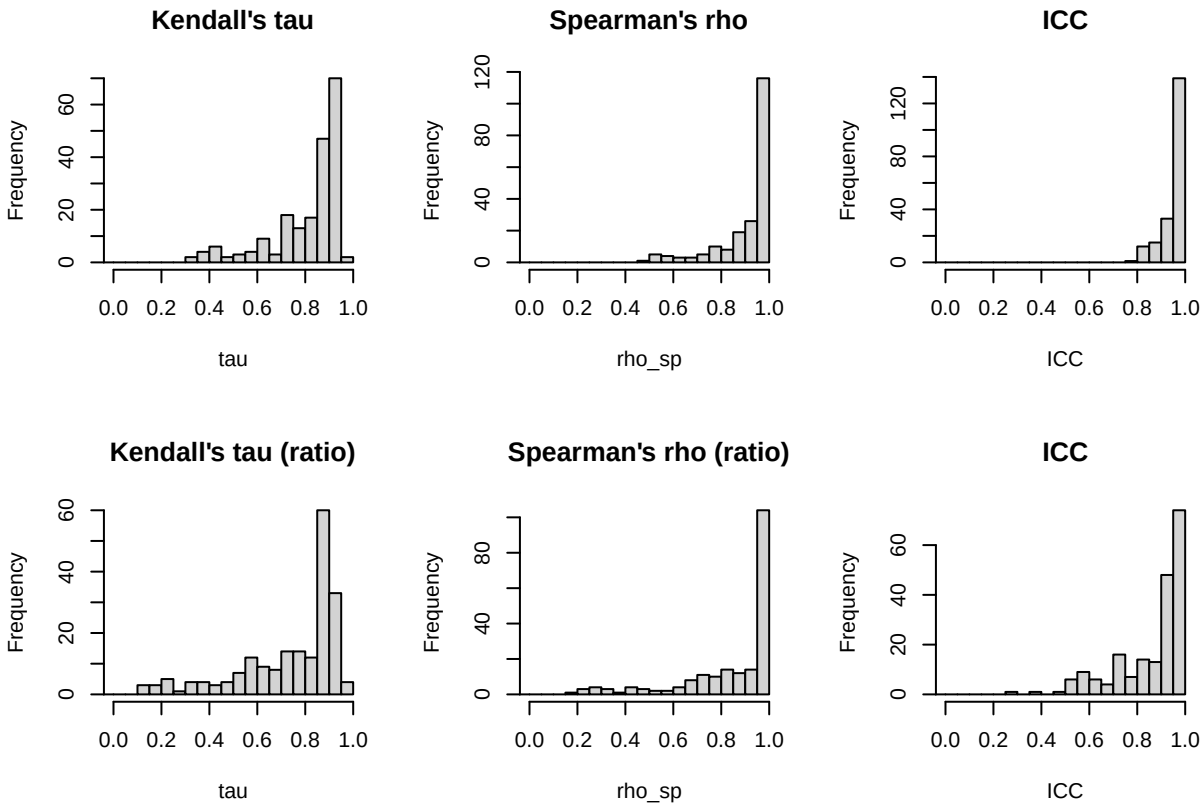

Kendall's tau
Frequency
0 2 4 6
0.0 0.2 0.4 0.6 0.8 1.0
tau
Spearman's rho
Frequency
0 1 2 3 4 5
0.0 0.2 0.4 0.6 0.8 1.0
rho_sp
ICC
Frequency
0 2 4 6 8
0.0 0.2 0.4 0.6 0.8 1.0
ICC
Kendall's tau (ratio)
Frequency
0 1 2 3 4 5 6
0.0 0.2 0.4 0.6 0.8 1.0
tau
Spearman's rho (ratio)
Frequency
0 2 4 6 8
0.0 0.2 0.4 0.6 0.8 1.0
rho_sp
ICC
Frequency
0 2 4 6
0.0 0.2 0.4 0.6 0.8 1.0
ICC

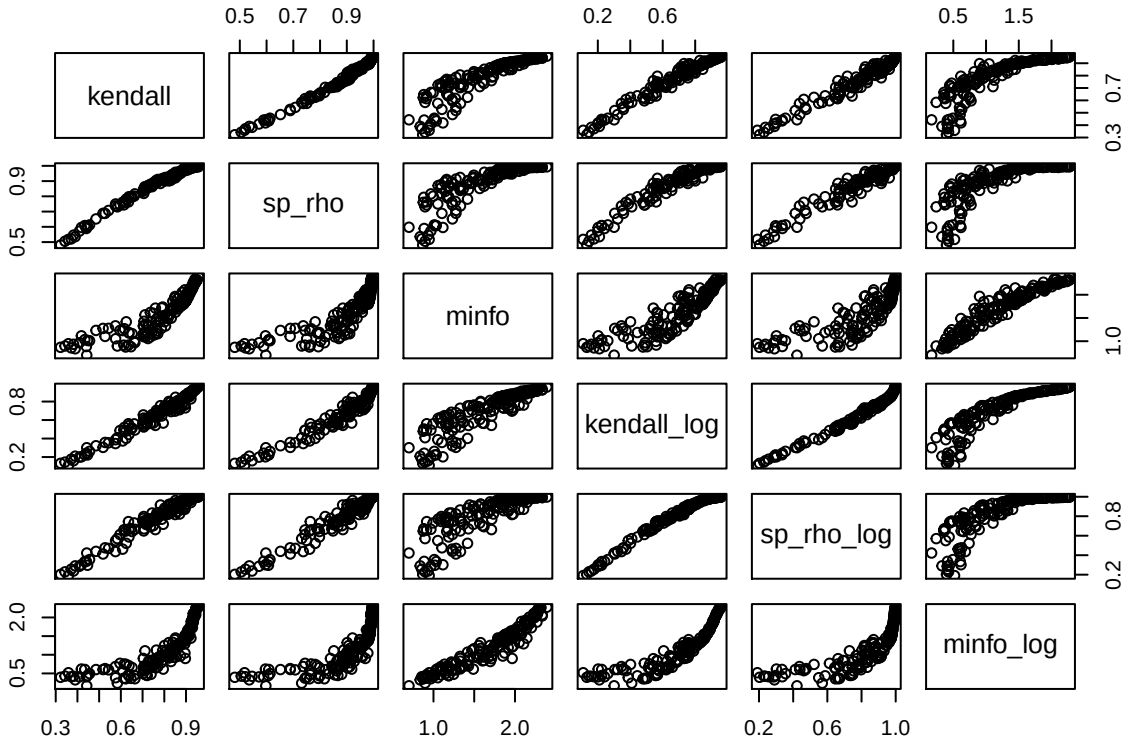

kendall
sp_rho
minfo
kendall_log
sp_rho_log
minfo_log
0.6
0.8
0.4
0.7
0.5
1.5
0.9
0.6
0.8
0.5
1.4
0.6
0.8
0.4
0.9
0.6
2.0
0.5
0.5
0.7
0.9
0.6
1.2
0.6
0.8
1.0

Clayton Copula Model

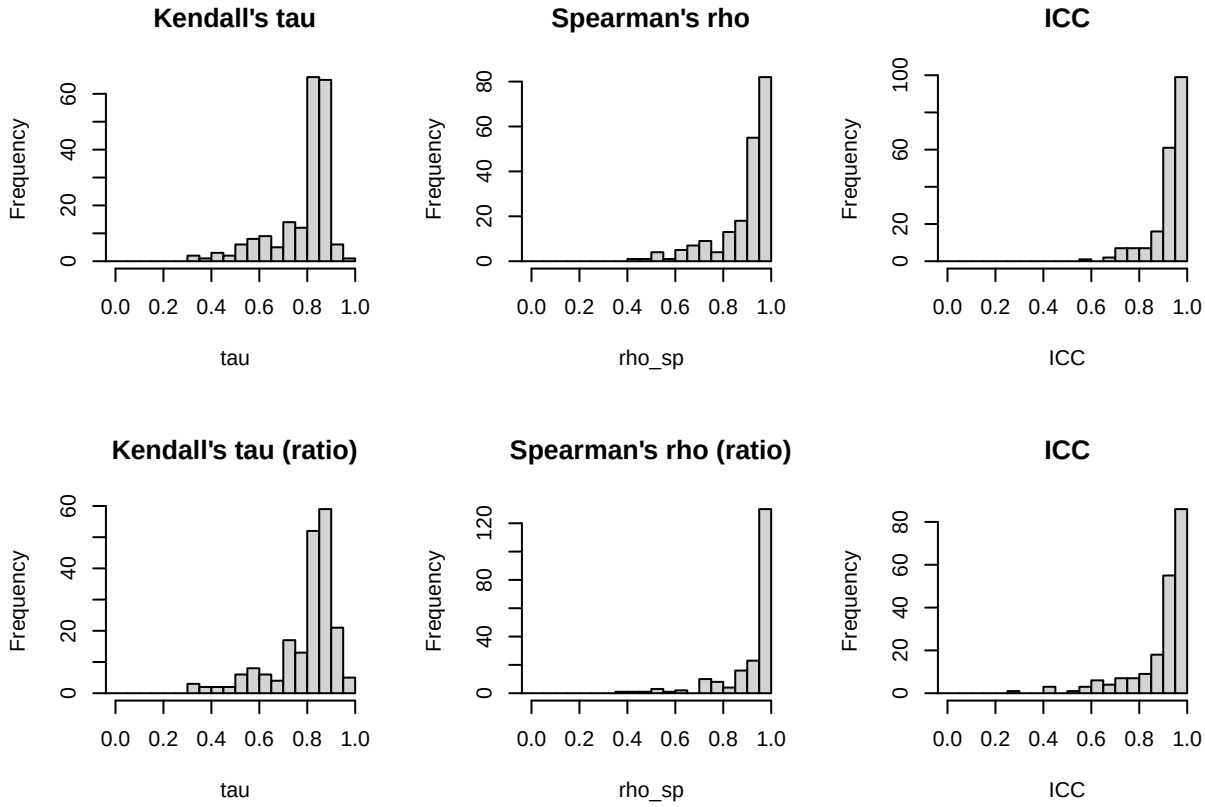


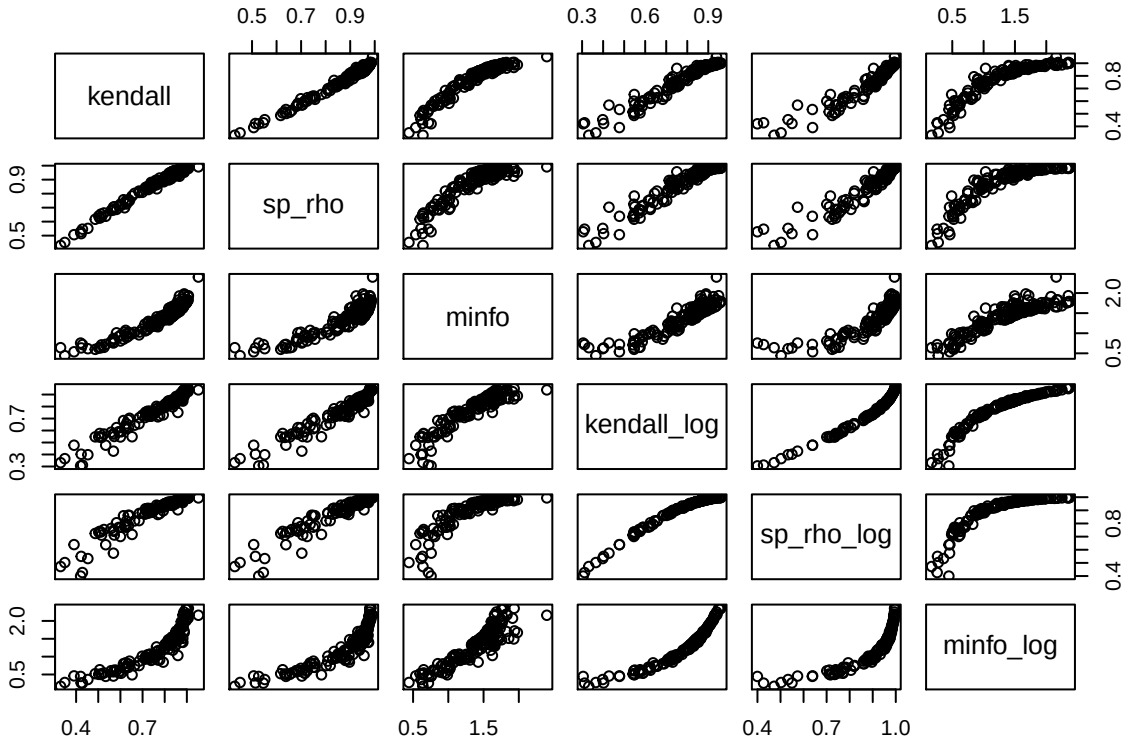


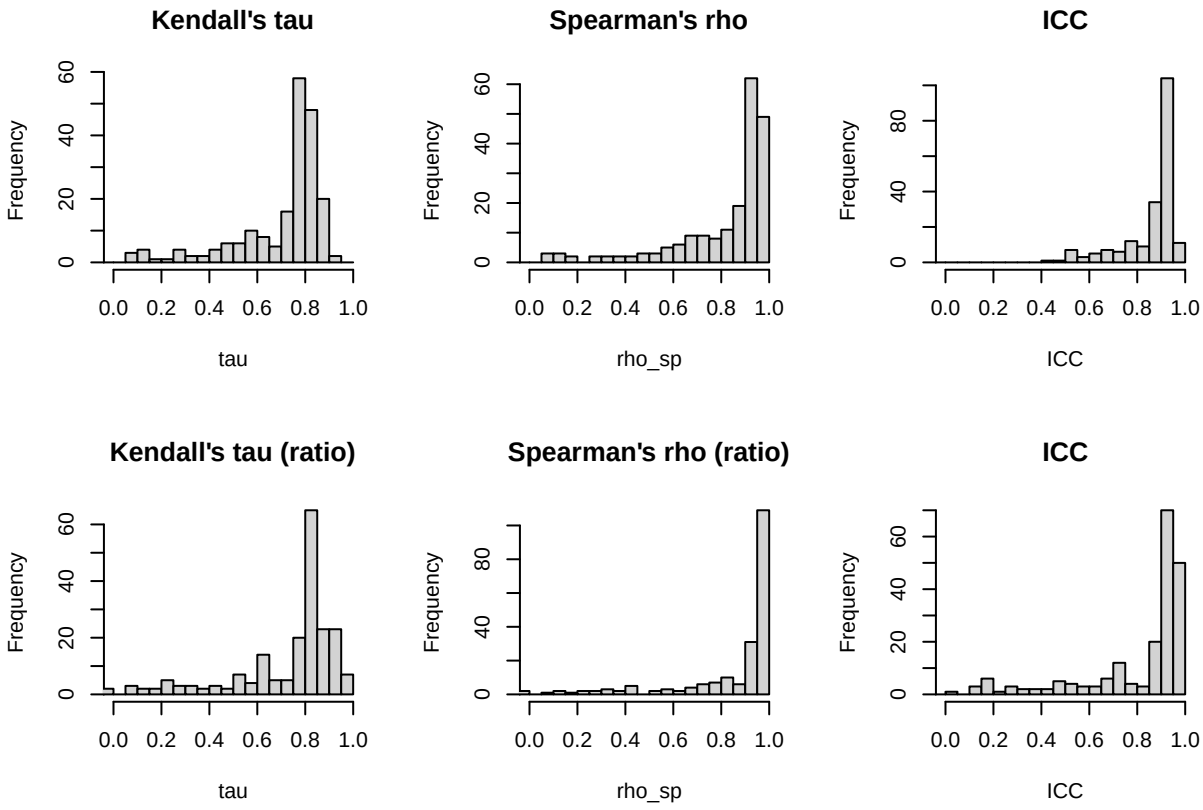

Kendall's tau
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho
Frequency
0
40
80
120
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC
Frequency
0
40
80
120
0.0
0.2
0.4
0.6
0.8
1.0
ICC
Kendall's tau (ratio)
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho (ratio)
Frequency
0
40
80
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
ICC

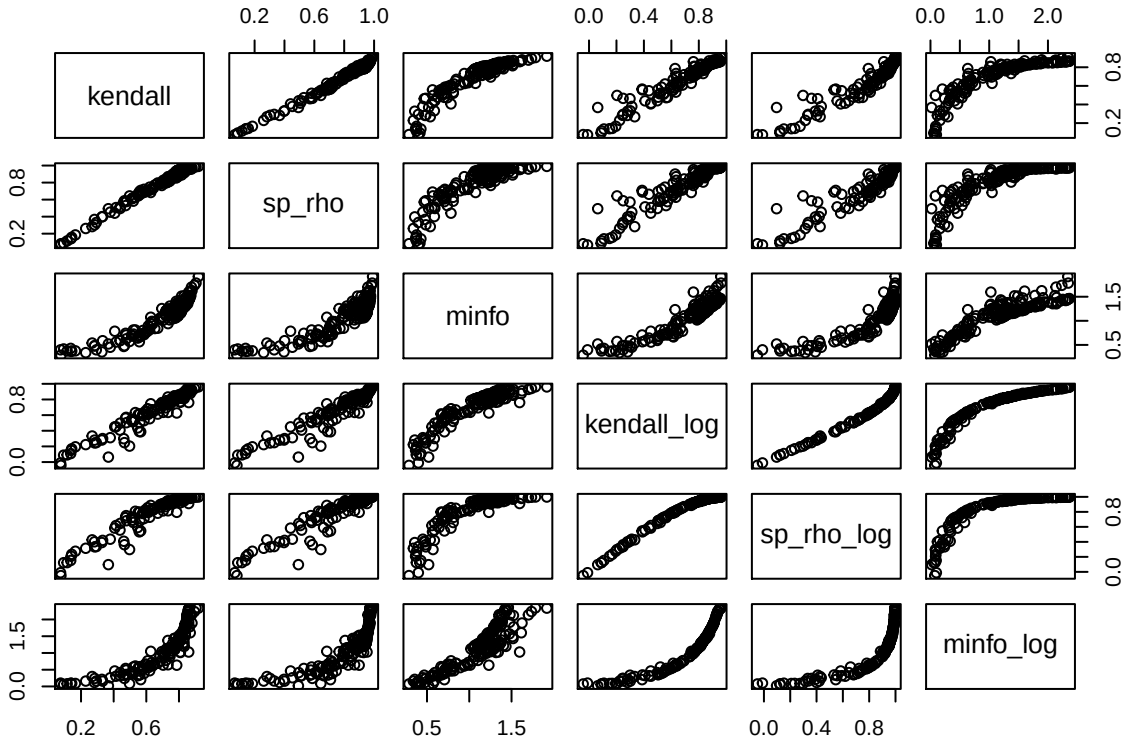

kendall
sp_rho
minfo
kendall_log
sp_rho_log
minfo_log
0.5 0.7 0.9
0.2 0.6
0.5 1.5
0.3 0.7
0.5 0.9
1.0
0.2 0.8
0.2 0.8
0.5 2.0
0.3 0.6 0.9
1.0 2.0
0.2 0.6 1.0

Gumbel Copula Model






Kendall's tau
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC
Frequency
0
40
80
0.0
0.2
0.4
0.6
0.8
1.0
ICC
Kendall's tau (ratio)
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
tau
Spearman's rho (ratio)
Frequency
0
40
80
0.0
0.2
0.4
0.6
0.8
1.0
rho_sp
ICC
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
ICC



Information Theoretic (unfinished)

The density of ΔT can be computed by integration (sampling) as before. The histogram is obtained by sampling, the red line is the computed density. They correspond closely, indicating that the computed density is correct.

The density for ΔS is less straightforward because of the atomicity. Still, it can be computed in a similar way is for ΔT.

IT other approach