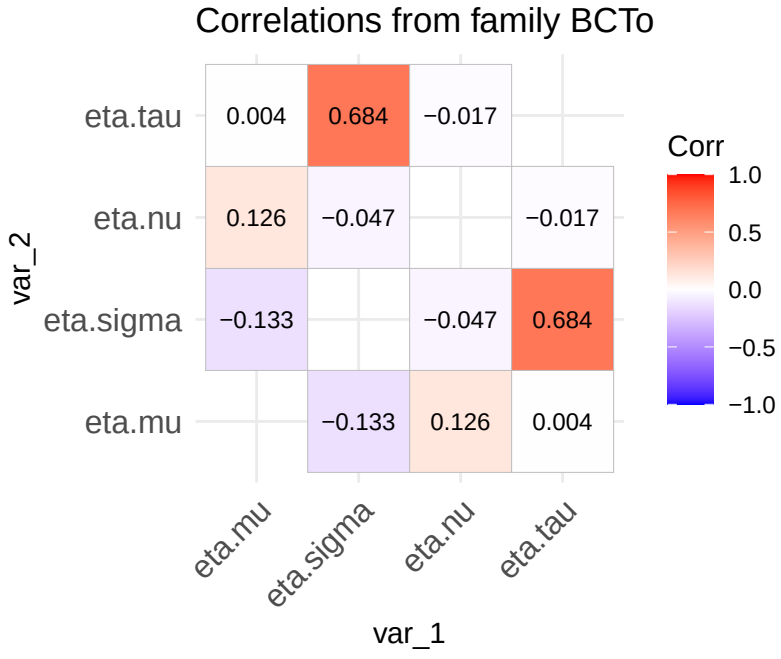
Practical 4

Mikis Stasinopoulos Bob Rigby Gillian Heller
Fernanda De Bastiani Niki Umlauf

Correlation of a distribution family

The code below shows how to create a correlation coefficient table for the parameters of the BCTo distribution. It generates 10000 obstbations from the BCTo distribution at given values for μ, σ, ν and τ them fits a model and plots the empirical correlation matrix of the parameters.

```
source("~/Dropbox/GAMLSS-development/ggplot/family_cor.R")
library(gamlss)
library(ggplot2)
family_cor("BCTo", mu=1, sigma=0.11, nu=1, tau=5, no.sim=10000)
```



1. Try different values for μ, σ, ν and τ for the BCTo distribution
2. Try different distribution i.e SHASH

Fitting different models

Fit the

- linear,
- smooth additive,
- regression tree,
- neural networks and
- random forest

described in the slides.

the data

```
library(gamlss2)
library(gamlss.ggplots)
library(gamlss.prepdata)
rent99 |> data_rm( c(2,9)) |>
 data_few2fac() -> da
```

```
************************************************************
    rent     area    yearc location     bath  kitchen cheating
    2723      132       68        3        2        2        2
************************************************************
4 vectors with fewer number of values than 10 were transformed to factors
************************************************************
************************************************************
```

the linear model

```
mlinear <- gamlss2(rent~area+poly(yearc,2)+location+bath+kitchen+
              cheating|area+yearc+location+bath+kitchen+cheating,
              family=BCTo, trace=FALSE, data=da)
summary(mlinear)
```

```
Call:
gamlss2(formula = rent ~ area + poly(yearc, 2) + location + bath +
    kitchen + cheating | area + yearc + location + bath + kitchen +
    cheating, data = da, family = BCTo, ... = pairlist(trace = FALSE))
---
Family: BCTo
Link function: mu = log, sigma = log, nu = identity, tau = log
*--------
Parameter: mu
---
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.0132864  0.0275445 182.007  < 2e-16 ***
area            0.0106293  0.0002329  45.640  < 2e-16 ***
location2       0.0878750  0.0104118   8.440  < 2e-16 ***
location3       0.1983303  0.0383017   5.178 2.39e-07 ***
bath1           0.0415152  0.0206509   2.010   0.0445 *
kitchen1        0.1129557  0.0236130   4.784 1.80e-06 ***
cheating1       0.3304758  0.0240475  13.743  < 2e-16 ***
poly(yearc, 2)1 5.0471663  0.3301520  15.287  < 2e-16 ***
poly(yearc, 2)2 3.3575299  0.2751069  12.204  < 2e-16 ***
*--------
Parameter: sigma
---
Coefficients:
              Estimate Std. Error   t value Pr(>|t|)
(Intercept)  1.047e+01  6.245e-02   167.650  < 2e-16 ***
area         1.218e-03  6.099e-04     1.996   0.0460 *
yearc       -5.972e-03  8.078e-07 -7392.764  < 2e-16 ***
location2    5.971e-02  2.911e-02     2.051   0.0403 *
location3    2.176e-01  9.224e-02     2.359   0.0184 *
bath1        5.069e-03  5.966e-02     0.085   0.9323
kitchen1     3.750e-02  7.106e-02     0.528   0.5977
cheating1   -2.431e-01  4.766e-02    -5.100  3.6e-07 ***
*--------
Parameter: nu
---
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.65342    0.05454   11.98   <2e-16 ***
*--------
Parameter: tau
---
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.18974    0.01758   181.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
*--------
n = 3082 df =  19 res.df =  3063
Deviance = 38254.617 Null Dev. Red. = 5.95%
AIC = 38292.617 elapsed =  0.32sec
```

the additive model

```
madditive <- gamlss2(rent~s(area)+s(yearc)+location+bath+kitchen+
                cheating|s(area)+s(yearc)+location+bath+kitchen+cheating,
              family=BCTo, data=da, trace=F)
GAIC(mlinear, madditive)
```

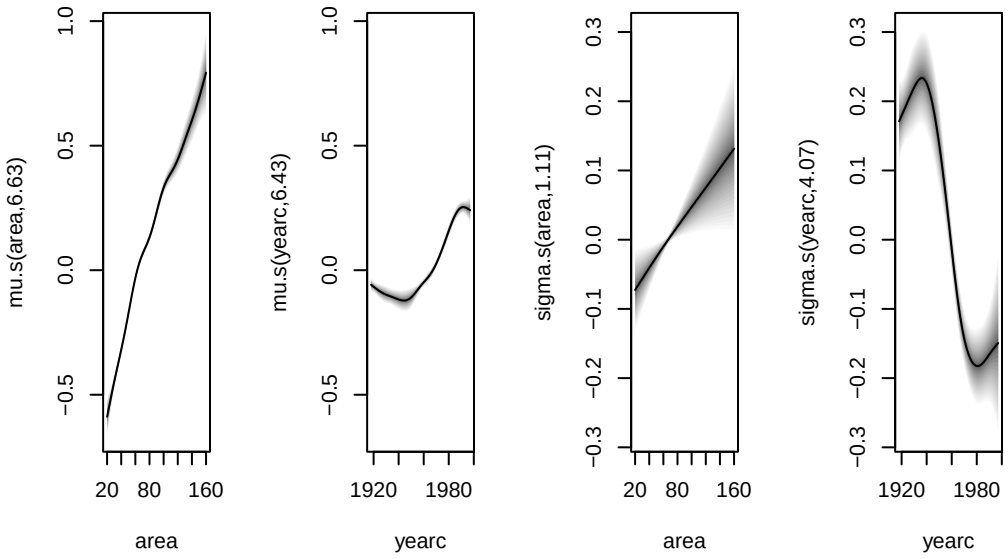
```
               AIC      df
madditive 38196.95 32.2364
mlinear   38292.62 19.0000
```

```
summary(madditive)
```

```
Call:
gamlss2(formula = rent ~ s(area) + s(yearc) + location + bath +
    kitchen + cheating | s(area) + s(yearc) + location + bath +
    kitchen + cheating, data = da, family = BCTo, ... = pairlist(trace = F))
---
Family: BCTo
Link function: mu = log, sigma = log, nu = identity, tau = log
*--------
Parameter: mu
---
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.71524    0.02082 274.541  < 2e-16 ***
location2    0.08808    0.01102   7.994 1.83e-15 ***
location3    0.20998    0.04965   4.229 2.41e-05 ***
bath1        0.05792    0.04280   1.353  0.17605
kitchen1     0.10859    0.03642   2.982  0.00289 **
```

```
cheating1    0.34594    0.02021  17.114  < 2e-16 ***
---
Smooth terms:
    s(area) s(yearc)
edf  6.6271   6.4275
*--------
Parameter: sigma
---
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.15569    0.03695 -31.277  < 2e-16 ***
location2    0.05293    0.02264   2.338   0.0195 *
location3    0.21218    0.08616   2.463   0.0138 *
bath1        0.03440    0.08318   0.414   0.6792
kitchen1     0.01861    0.07591   0.245   0.8063
cheating1   -0.23431    0.03754  -6.242 4.91e-10 ***
---
Smooth terms:
    s(area) s(yearc)
edf  1.1077   4.0742
*--------
Parameter: nu
---
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.69275    0.04548   15.23   <2e-16 ***
*--------
Parameter: tau
---
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.26960    0.07549   43.31   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
*--------
n = 3082 df =  32.24 res.df =  3049.76
Deviance = 38132.48 Null Dev. Red. = 6.25%
AIC = 38196.9528 elapsed =  0.47sec
```

```
plot(madditive)
```



Regression tree model

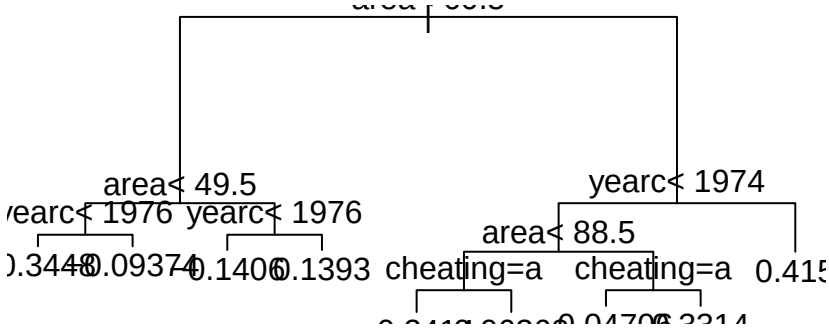
```
mregtree <- gamlss2(rent~tree(~area+yearc+location+bath+kitchen+cheating)|
                        tree(~area+yearc+location+bath+kitchen+cheating),
                    family=BCTo, data=da, trace=FALSE)
```

```
Loading required namespace: rpart
```

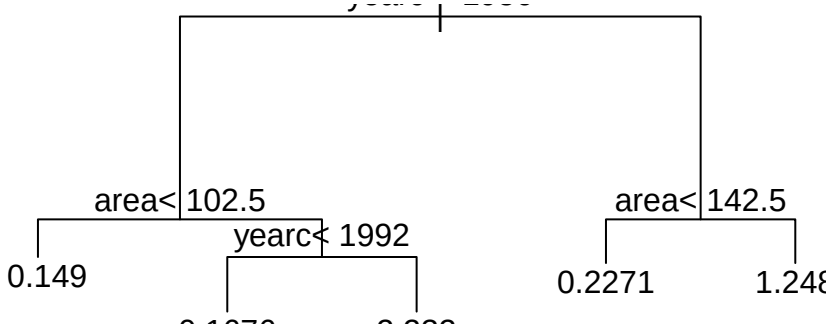
```
GAIC(mlinear, madditive, mregtree)
```

```
               AIC      df
madditive 38196.95 32.2364
mlinear   38292.62 19.0000
mregtree  38754.64 30.0000
```

```
pp<- specials(mregtree, model = "mu", elements = "model")
plot(pp)
text(pp)
```




```
pp<-specials(mregtree, model = "sigma", elements = "model")
plot(pp)
text(pp)
```



Neural network model

```
f <- rent ~ n(~area+yearc+location+bath+kitchen, size=10)|      n(~area+yearc+location+bath+
mneural <- gamlss2(f,family=BCTo, data=da)
```

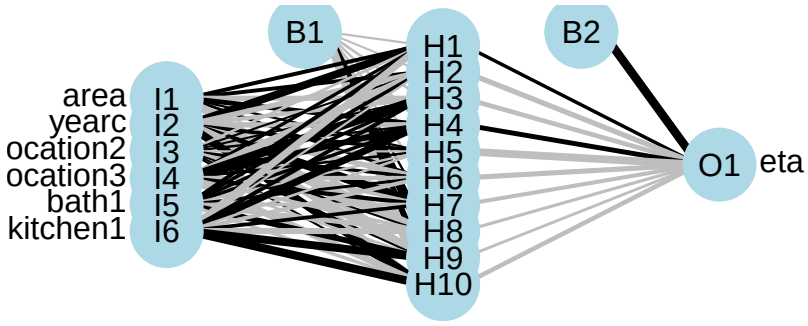
```
Loading required namespace: nnet

GAMLSS-RS iteration  1: Global Deviance = 38183.2379 eps = 0.290350
GAMLSS-RS iteration  2: Global Deviance = 38163.4158 eps = 0.000519
GAMLSS-RS iteration  3: Global Deviance = 38163.1172 eps = 0.000007
```

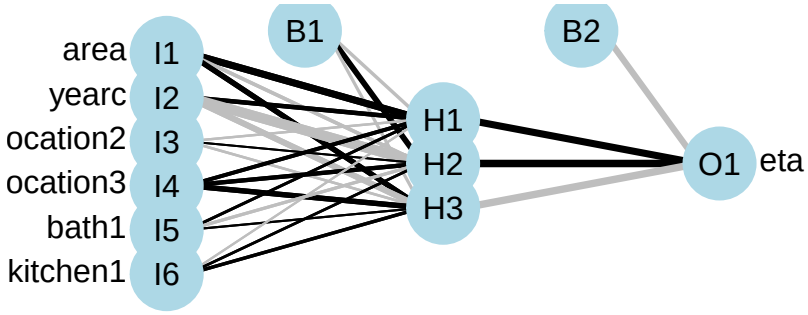
```
gamlss2::GAIC(mlinear, madditive, mneural, mregtree, k=2)
```

```
                AIC       df
madditive 38196.95  32.2364
mlinear   38292.62  19.0000
mneural   38383.12 110.0000
mregtree  38754.64  30.0000
```

```
source("~/Dropbox/GAMLSS-development/nnet/plot_NN.R")
plot(specials(mneural, model = "mu", elements = "model"))
```




```
plot(specials(mneural, model = "sigma", elements = "model"))
```



Random forest model

```
f <- rent ~ cf(~area+yearc+location+bath+kitchen)|    cf(~area+yearc+location+bath+kitchen)
mcf <- gamlss2(f,family=BCTo, data=da)
```

```
Loading required namespace: partykit
```

```
GAMLSS-RS iteration  1: Global Deviance = 38459.3856 eps = 0.285217
GAMLSS-RS iteration  2: Global Deviance = 38319.3501 eps = 0.003641
GAMLSS-RS iteration  3: Global Deviance = 38314.517 eps = 0.000126
GAMLSS-RS iteration  4: Global Deviance = 38313.124 eps = 0.000036
GAMLSS-RS iteration  5: Global Deviance = 38312.9247 eps = 0.000005
```

GAIC evaluation

```
gamlss2::GAIC(mlinear, madditive, mneural, mregtree, mcf, k=2)
```

```
               AIC       df
madditive 38196.95  32.2364
mlinear   38292.62  19.0000
mneural   38383.12 110.0000
mcf       38720.92 204.0000
mregtree  38754.64  30.0000
```

```
model_GAIC_lollipop(mlinear, madditive, mneural, mregtree, mcf)
```