ddsPLS Exploration

Harpeth Lee

2/4/2022

```r
library(ddsPLS2)
```

```r
ddsPLS2::ddsPLS2_App()
```

This code chunk opens an applet that can be used to build models using ddsPLS. Note that it requires the X and Y variables as separate csv files.

Code copied from the simulation_ssdpls2 repository created by Hadrien Lorenzo.

```r
# Creates a toy data set for the ddsPLS function
toy_ex <- get_toy_example()
```

```r
# Creates model from the toy data
toy_mod <- ddsPLS(toy_ex$X, toy_ex$Y)

toy_results <- toy_mod$results
```

Recreate Toy Example

This is a recreation of the toy example created by Hadrien Lorenzo, the original example can be found here.

```r
# Creates toy data set to be used
simu_toy <- get_toy_example(n=50,sqrt_1_minus_sig2 = 0.9025,p = 1000)

# Creates vector of lambda values to be used
lambdas <- seq(0,1,length.out = 30)

# Sets number of bootstrap samples to run
n_B <- 100

# Creates model using ddsPLS algorithm
model_toy <- ddsPLS(simu_toy$X,simu_toy$Y,
                    doBoot = FALSE,
                    lambdas = lambdas,
                    n_B = n_B,
                    verbose = T # whether trace during process
                    )
```

```r
model_toy_2 <- ddsPLS(simu_toy$X,simu_toy$Y,
                      doBoot = FALSE,
                      criterion = "Q2",
                      lambdas = lambdas,
                      n_B = n_B,
                      verbose = T # whether trace during process
                      )
```

Design 1

```
simu_1 <- get_design_1(n=50,sqrt_1_minus_sig2 = 0.99,p = 1000,q = 3)
```
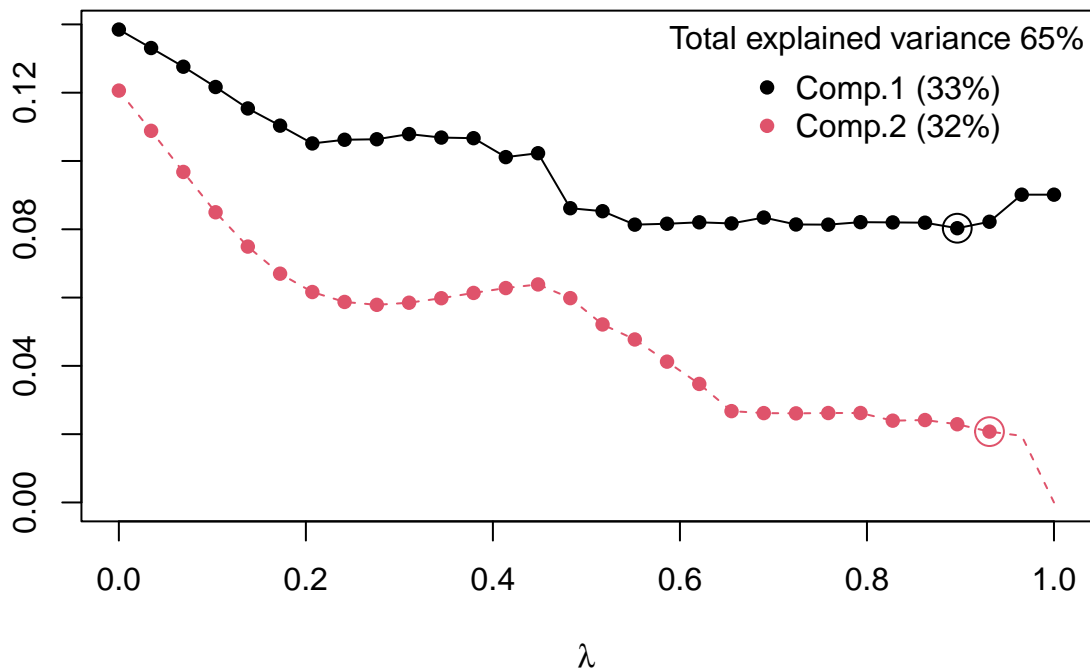
What does the `NCORES` argument do? Setting it to integers greater than 1 gives an error.

Is there a way to include more components in the model?

```
model_1 <- ddsPLS(simu_1$X,simu_1$Y,
                  lambdas = lambdas,
                  n_B=n_B,
                  verbose=T)
```

```
##                         ---------------
##                        |    ddsPLS     |
## =====================---------------=====================
## Should we build component 1 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##         0.9 0.36 0.36 0.27 0.27     33%         33%
##                                     ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.93 0.41 0.1 0.39 0.26     32%         65%
##                                     ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                              ...component 3 not built!
## =====================                 =====================
##                        ================
```

$$\overline{R}^2_B - \overline{Q}^2_B$$


```
model_1_Q2 <- ddsPLS(simu_1$X,simu_1$Y,
                     criterion = "Q2",
```

```
                    lambdas = lambdas,
                    n_B=n_B,
                    verbose=T)
```

```
##                   --------------
##                  |    ddsPLS    |
## ====================--------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda   R2   R2h   Q2   Q2h  VarExpl  VarExpl.Tot
##        0.24 0.38  0.38 0.29  0.29      33%          33%
##                                  ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2   R2h   Q2  Q2h  VarExpl  VarExpl.Tot
##        0.21 0.65  0.28 0.57  0.4      32%          65%
##                                  ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                              ...component 3 not built!
## ====================             ====================
##                   ================
```

$$\overline{Q}^2_B$$
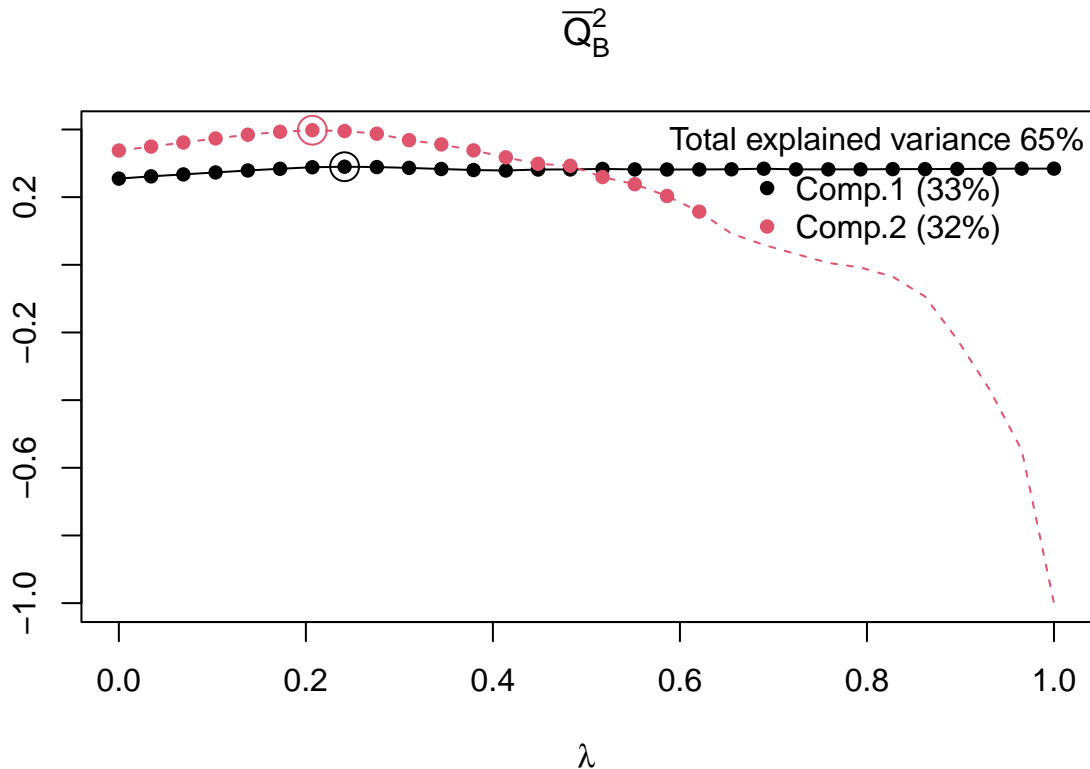

```
model_1_lambda <- ddsPLS(simu_1$X, simu_1$Y,
                         criterion = "Q2",
                         lambdas = seq(0,1,length.out = 100),
                         n_B = n_B,
                         verbose = T)
```
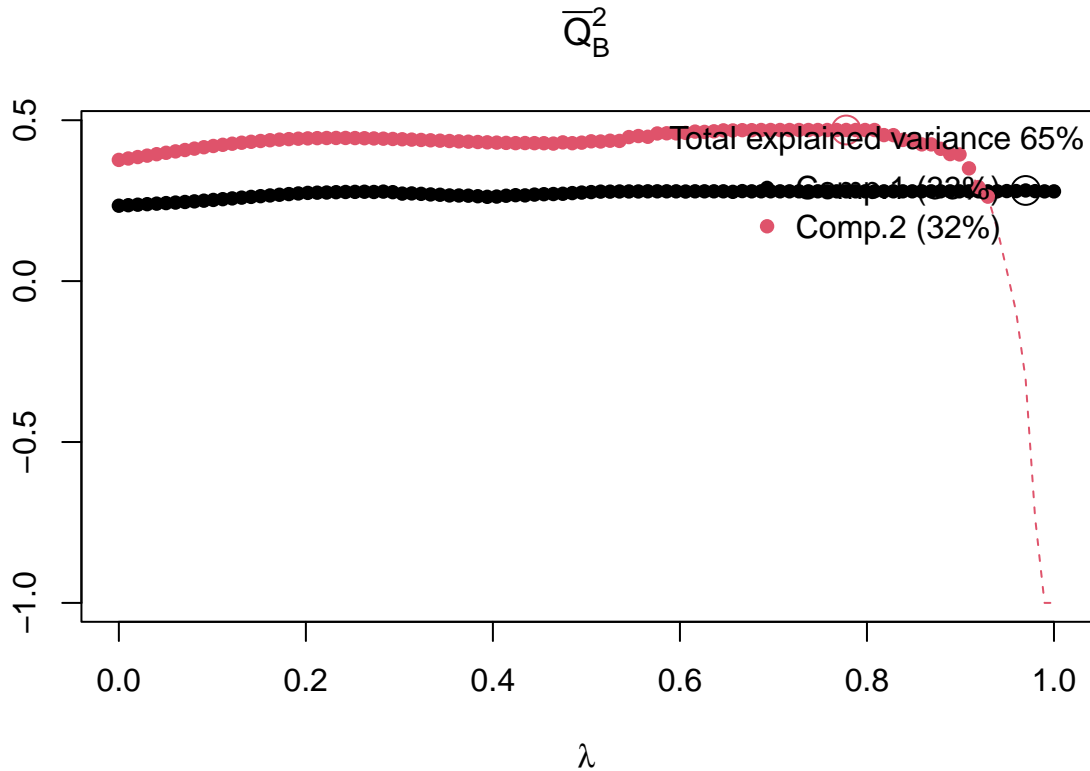
```
##                   --------------
##                  |    ddsPLS    |
## ====================--------------====================
## Should we build component 1 ? Bootstrap pending...
```

```
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.97 0.36 0.36 0.28 0.28     33%          33%
##                                    ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.78 0.62 0.27 0.61 0.47     32%          65%
##                                    ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                                    ...component 3 not built!
## ====================          ====================
##                   ================
```
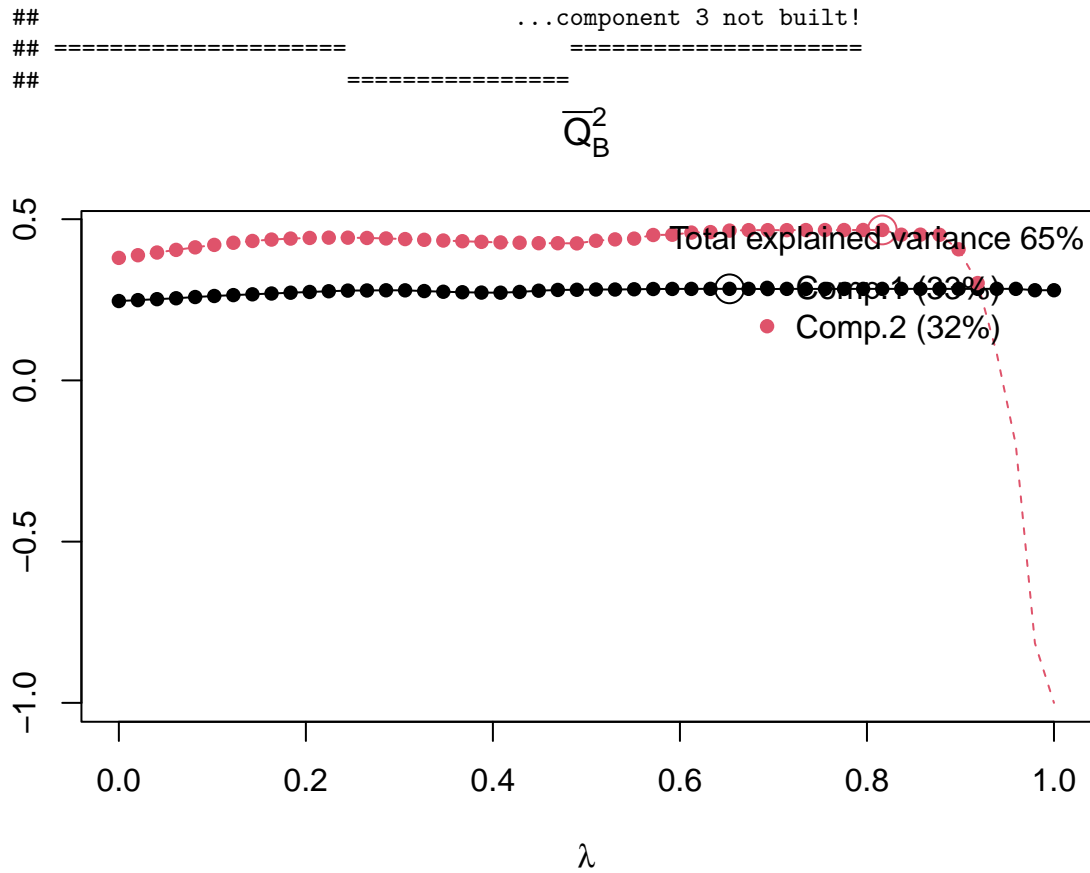
$$\overline{Q}_B^2$$


```
model_1_lambda_2 <- ddsPLS(simu_1$X, simu_1$Y,
                    criterion = "Q2",
                    lambdas = seq(0,1,length.out = 50),
                    n_B = n_B,
                    verbose = T)
```

```
##                   ---------------
##                   |    ddsPLS    |
## =====================---------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.65 0.36 0.36 0.28 0.28     33%          33%
##                                    ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.82 0.62 0.27 0.62 0.47     32%          65%
##                                    ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
```

4

```
##                              ...component 3 not built!
## ====================        =======================
##                     ================
```

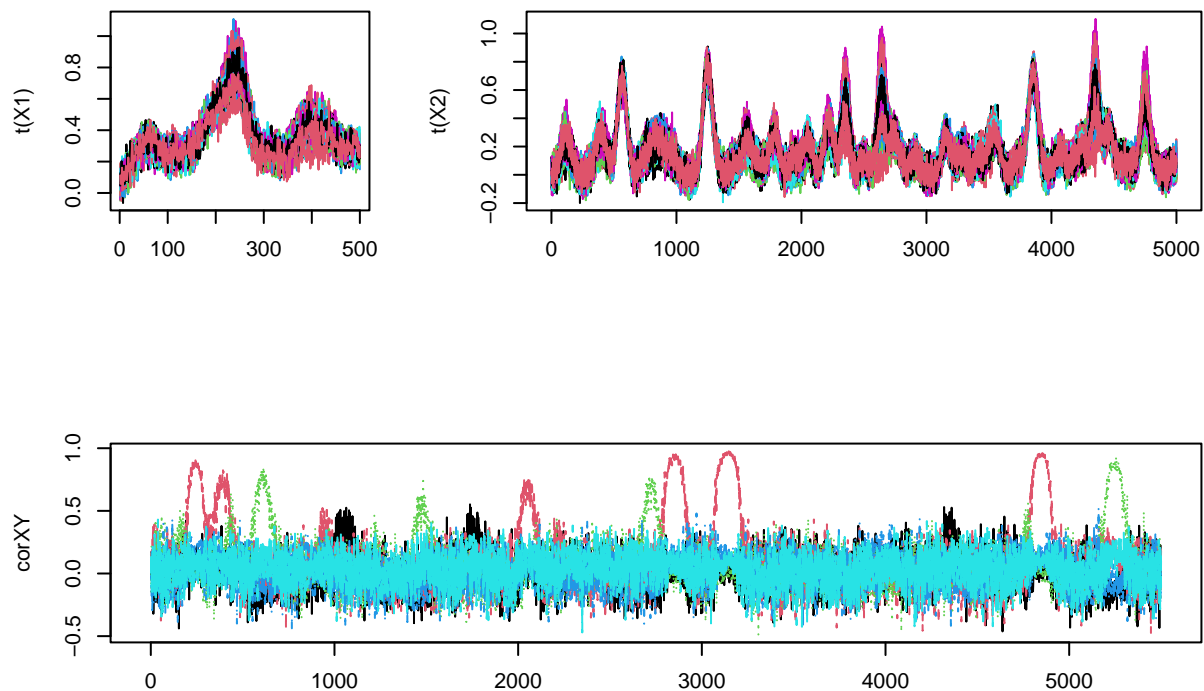$$\overline{Q}^2_B$$

Different Simulations of Design 1 Data

There is a problem with `get_design_1`, q cannot take values other than 5.

```r
data_1 <- get_design_1(n = 100, p = 1000, q = 5)

ddsPLS(data_1$X,data_1$Y,
                criterion = "Q2",
                lambdas = lambdas,
                n_B=n_B,
                verbose=T)
```
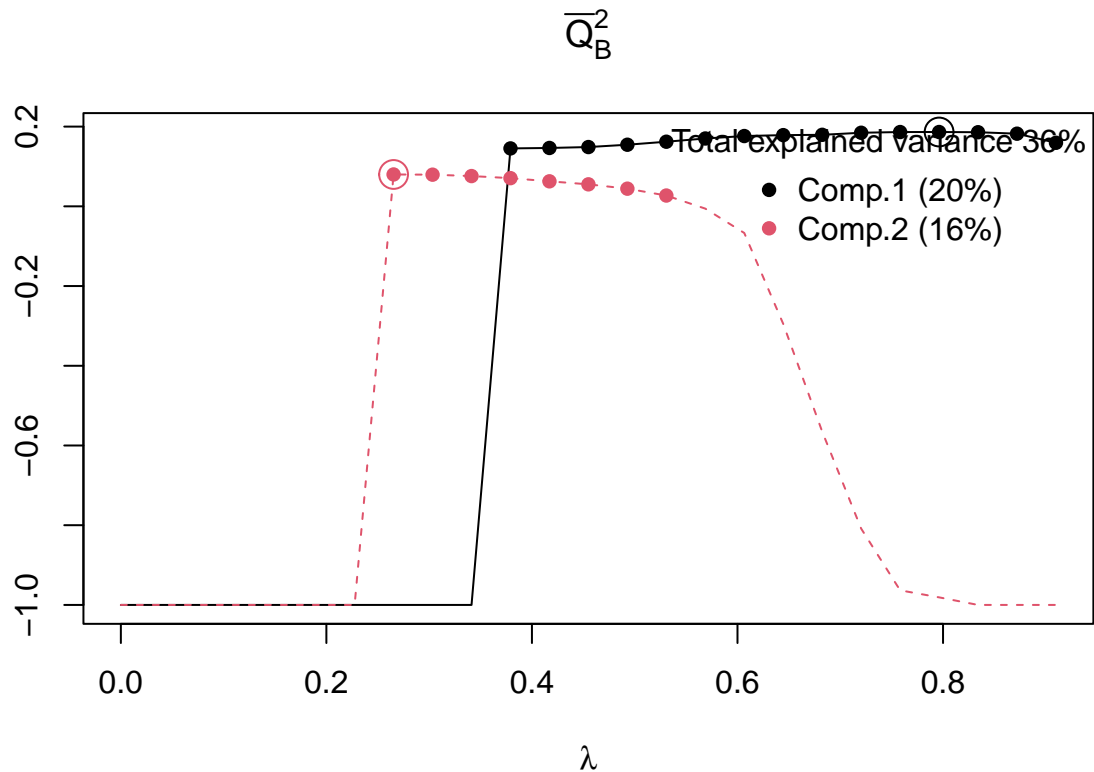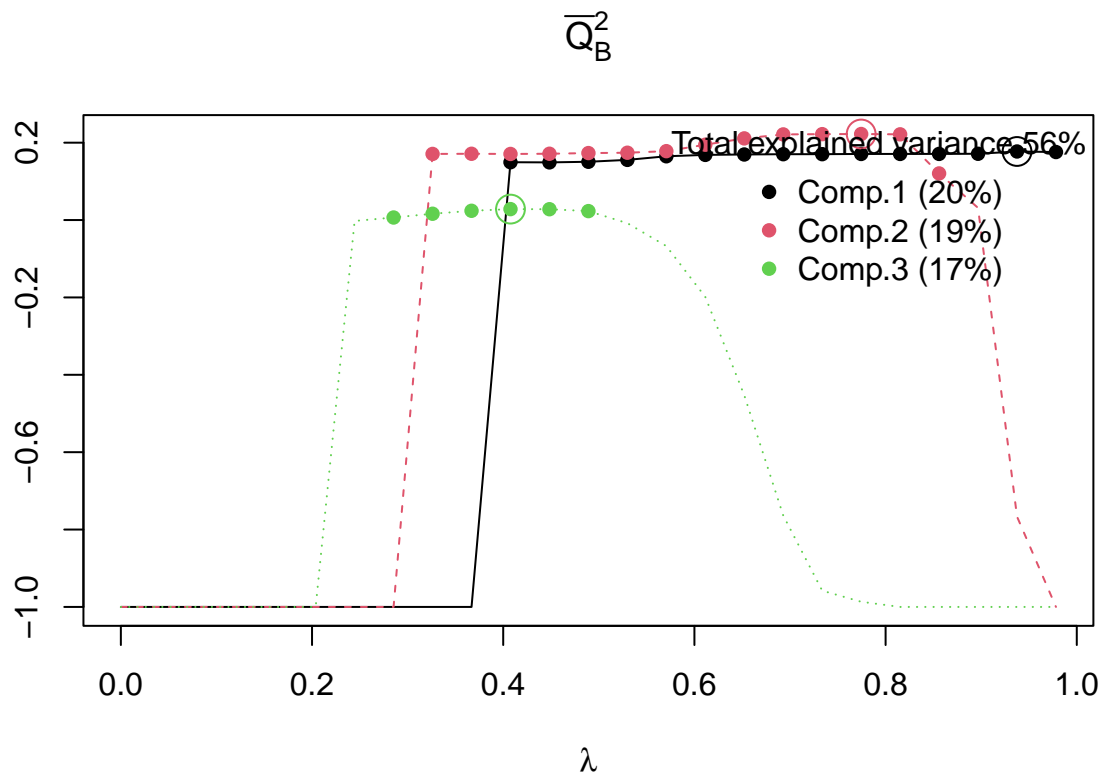
Design 2

```r
simu_2.1 <- get_design_2(plot = T)
```

```r
model_2.1 <- ddsPLS(simu_2.1$X$X1, simu_2.1$Y,
                    criterion = "Q2",
                    n_lambdas = 25,
                    verbose = TRUE)
```

```
##                     --------------
##                    |    ddsPLS    |
## ====================--------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda  R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##         0.8 0.2 0.2 0.19 0.19     20%         20%
##                                     ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.27 0.39 0.2 0.26 0.08     16%         36%
##                                     ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                                     ...component 3 not built!
## ====================                ====================
##                     ================
```
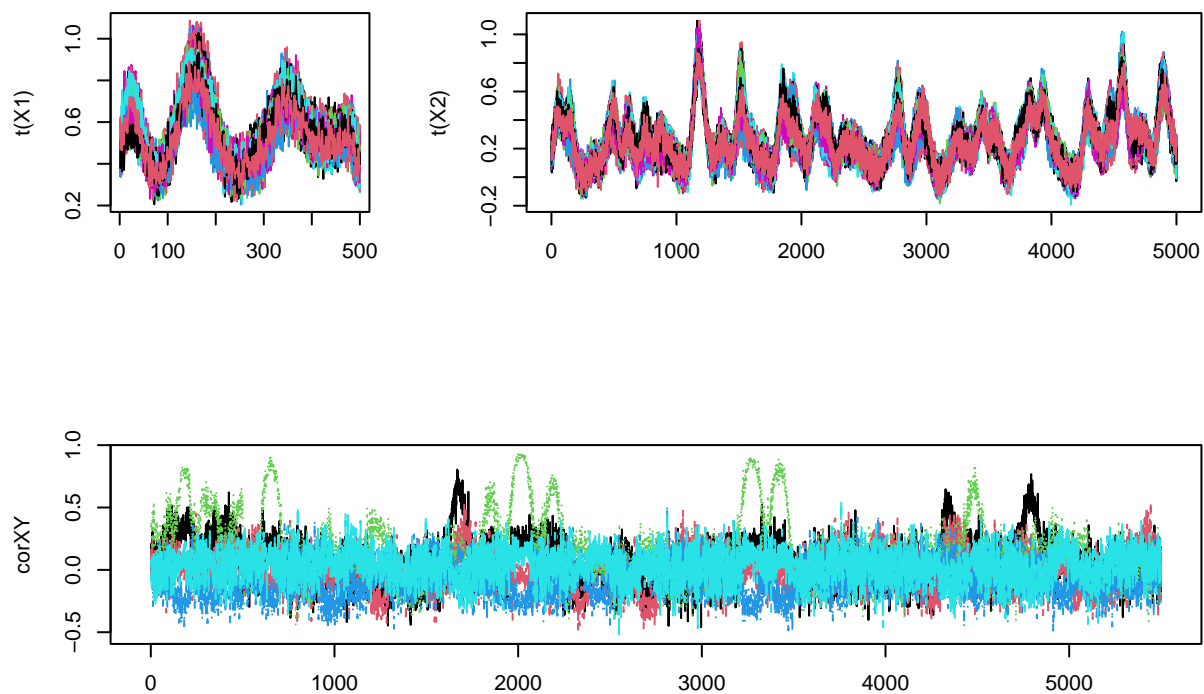
$$\overline{Q_B^2}$$

Total explained variance 36%
- Comp.1 (20%)
- Comp.2 (16%)

λ

```r
model_2.2 <- ddsPLS(simu_2.1$X$X2, simu_2.1$Y,
                    criterion = "Q2",
                    n_lambdas = 25,
                    verbose = TRUE)
```

```
##                     --------------
##                    |    ddsPLS    |
## ====================--------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.94 0.2 0.2 0.18 0.18     20%         20%
##                                ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.77 0.38 0.18 0.36 0.22     19%         39%
##                                ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.41 0.57 0.18 0.39 0.03     17%         56%
##                                ...component 3 built!
## Should we build component 4 ? Bootstrap pending...
##                                ...component 4 not built!
## ====================              ====================
##                     ==============
```

$$\overline{Q}_B^2$$

Total explained variance 56%
- Comp.1 (20%)
- Comp.2 (19%)
- Comp.3 (17%)

λ

```r
simu_2.2 <- get_design_2(seed = 2, ncpX = 20, plot = T)
```
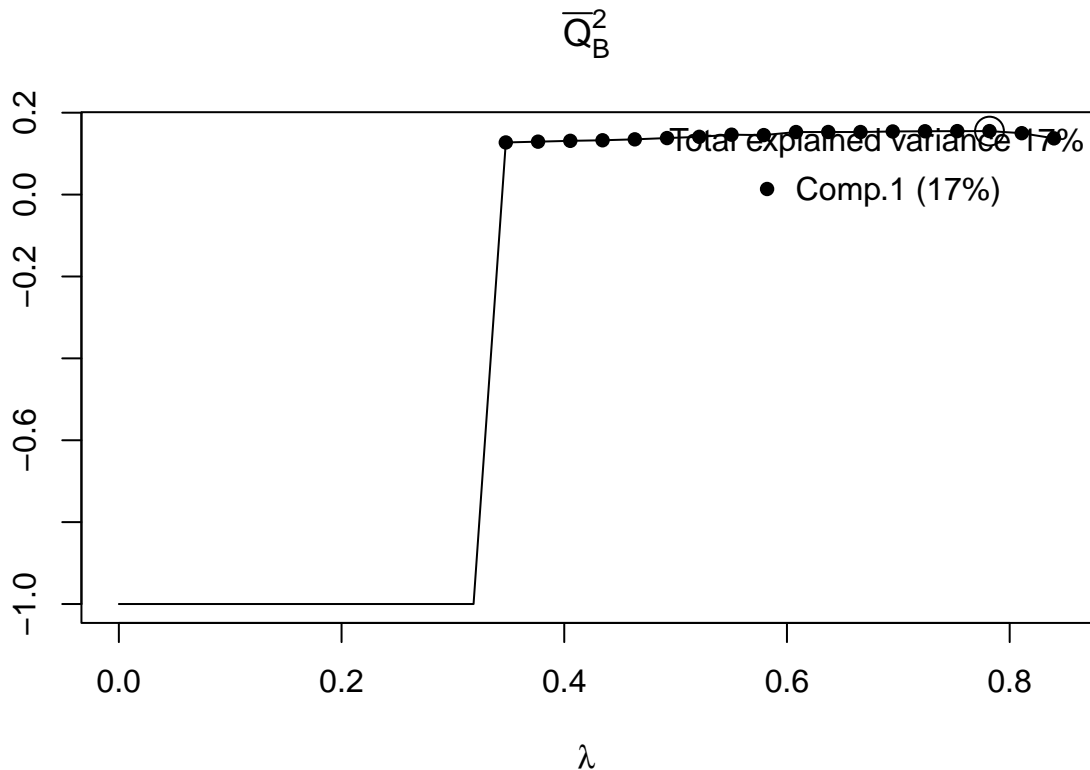


```r
model_2.3 <- ddsPLS(simu_2.2$Xs$X1, simu_2.2$Y,
                    criterion = "Q2",
                    n_lambdas = 30,
                    verbose = TRUE)
```
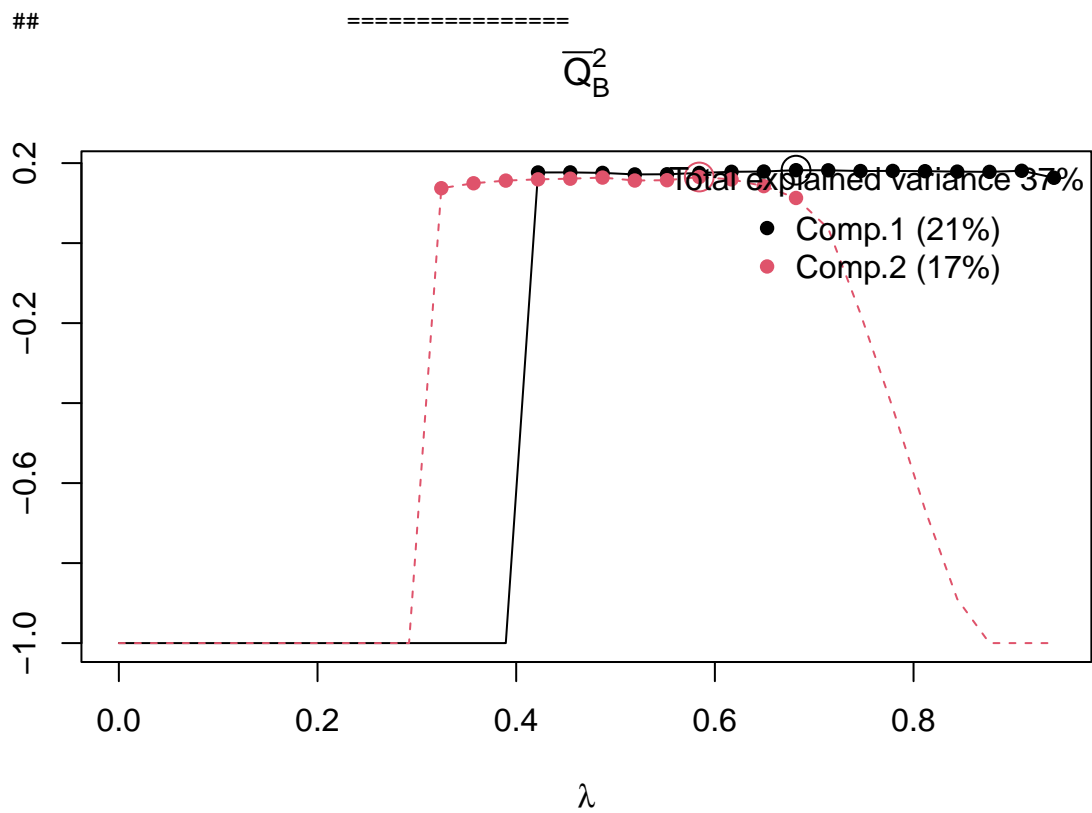
```
##                    --------------
```

```
##                          |    ddsPLS   |
## ====================---------------=====================
## Should we build component 1 ? Bootstrap pending...
##       lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##         0.78 0.18 0.18 0.16 0.16     17%          17%
##                                      ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##                                      ...component 2 not built!
## ====================             =====================
##                  ================
```

$$\overline{Q}^2_B$$


```
model_2.4 <- ddsPLS(simu_2.2$Xs$X2, simu_2.2$Y,
                    criterion = "Q2",
                    n_lambdas = 30,
                    verbose = TRUE)
```

```
##                      --------------
##                          |    ddsPLS   |
## ====================---------------=====================
## Should we build component 1 ? Bootstrap pending...
##       lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##         0.68 0.21 0.21 0.18 0.18     21%          21%
##                                      ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##       lambda  R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##         0.58 0.4 0.19 0.31 0.17     17%          37%
##                                      ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                                      ...component 3 not built!
## ====================             =====================
```

$$\overline{Q}^2_B$$

Model results can also be plotted using the `plot` function.

```
plot(model_2.1,type="Q2",legend.position = "bottomleft")
```

$$\overline{Q}^2_B$$
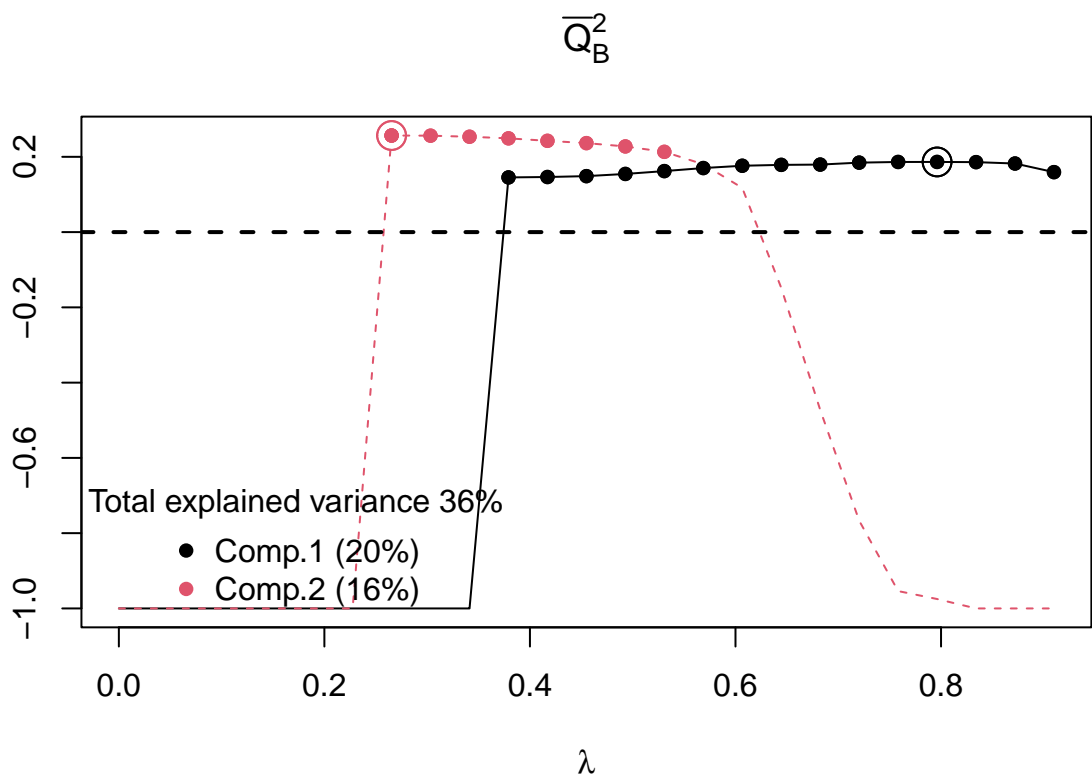
Get Data Simulation

The following `get_data` function is from the vignette for the `ddsPLS` package

```
data_3.1 <- get_data()

model_3.1 <- ddsPLS(data_3.1$X, data_3.1$Y,
                    criterion = "Q2",
                    n_lambdas = 30,
                    verbose = TRUE)
```

```
##                       --------------
##                      |    ddsPLS     |
## ====================---------------====================
## Should we build component 1 ? Bootstrap pending...
##       lambda   R2   R2h   Q2   Q2h VarExpl VarExpl.Tot
##         0.35 0.44 0.44 0.41 0.41     44%          44%
##                                     ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##                                 ...component 2 not built!
## ====================                 ====================
##                    ================
```

$$\overline{Q}^2_B$$
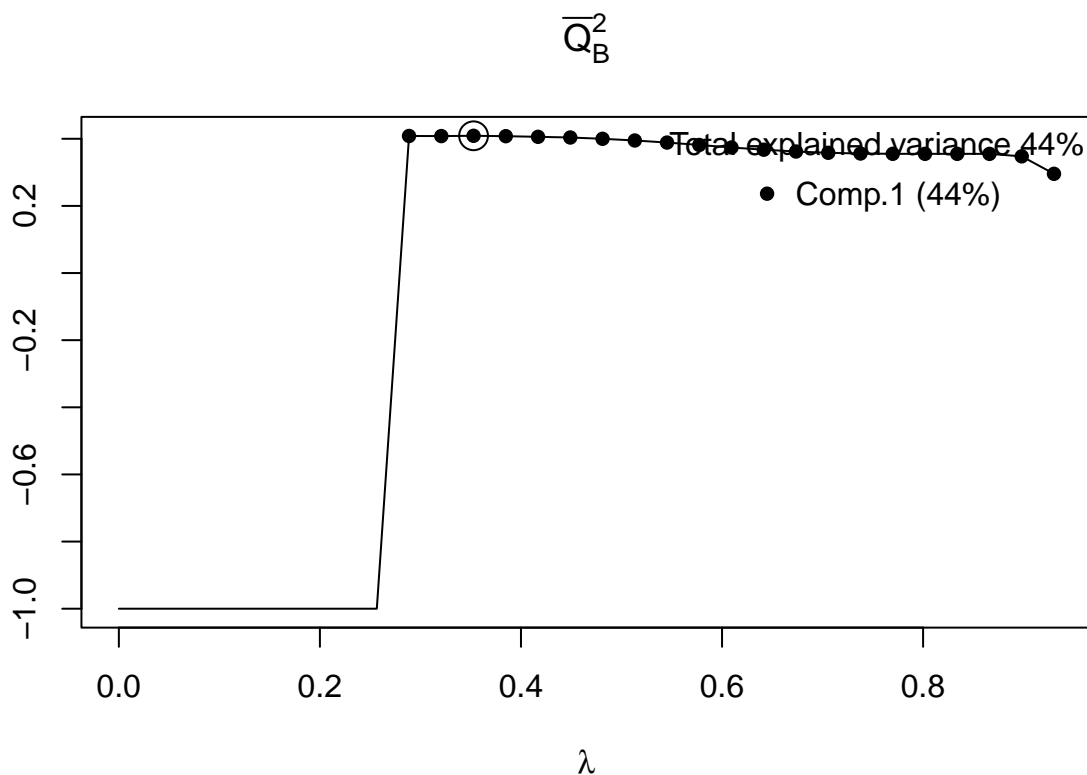

```
data_3.2 <- get_data(p1 = 50, p2 = 50, p3 = 50, p = 250)

model_3.2 <- ddsPLS(data_3.2$X, data_3.2$Y,
                    criterion = "Q2",
                    n_lambdas = 30,
                    verbose = TRUE)
```

```
##                       --------------
##                      |    ddsPLS     |
```

```
## ====================---------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda R2 R2h   Q2 Q2h VarExpl VarExpl.Tot
##        0.26 0.4 0.4 0.39 0.39     41%          41%
##                                  ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.16 0.61 0.2 0.56 0.28     21%          62%
##                                  ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                                  ...component 3 not built!
## ====================         ====================
##                     ================
```
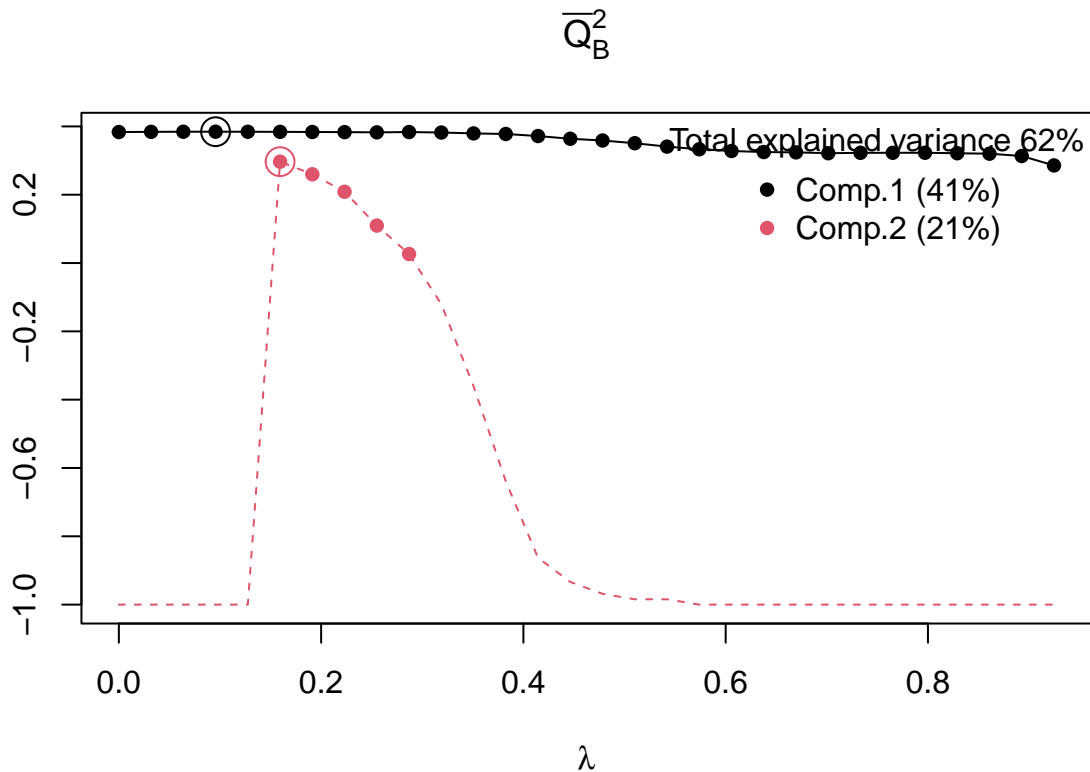
$$\overline{Q}^2_B$$
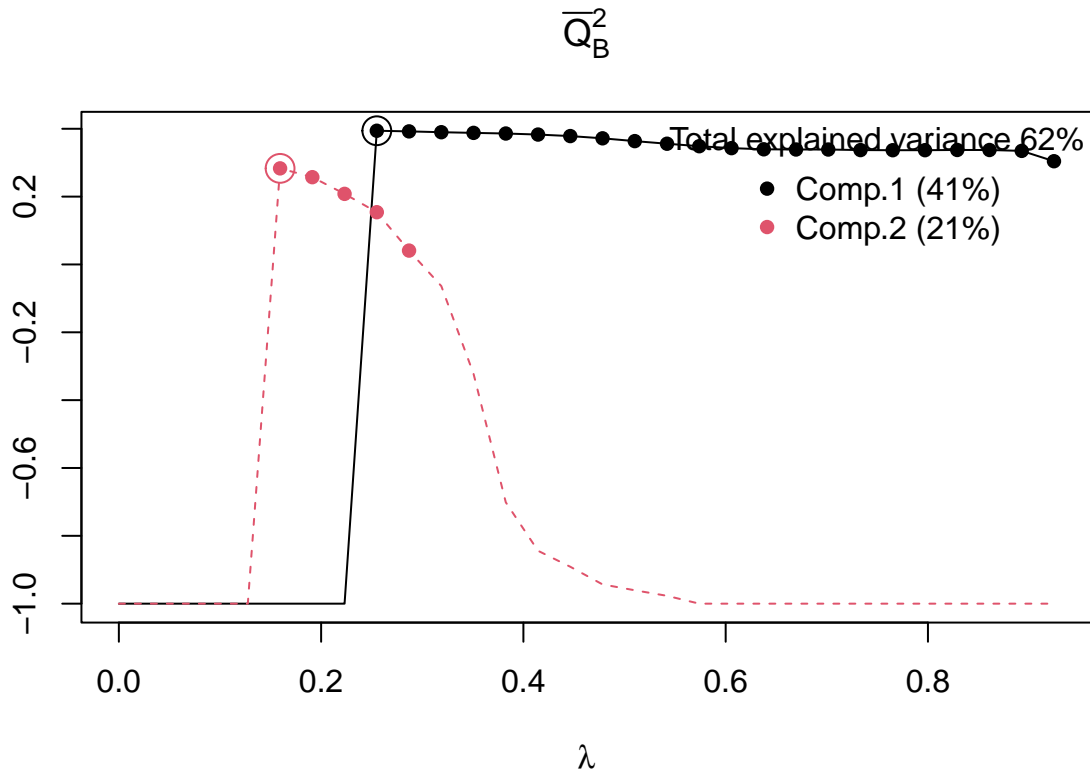

```
model_3.3 <- ddsPLS(data_3.2$X, data_3.2$Y,
                    criterion = "Q2",
                    n_lambdas = 30,
                    verbose = TRUE,
                    LD = TRUE)
```

```
##                     --------------
##                    |   ddsPLS     |
## ====================---------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##         0.1 0.4 0.4 0.38 0.38     41%          41%
##                                  ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2  R2h   Q2 Q2h VarExpl VarExpl.Tot
##        0.16 0.62 0.21 0.57 0.3     21%          62%
```

12

```
##                                       ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                                       ...component 3 not built!
## ====================             ====================
##                    ================
```

$$\overline{Q}_B^2$$


```
model_3.4 <- ddsPLS(data_3.2$X, data_3.2$Y,
                    criterion = "diffR2Q2",
                    n_lambdas = 30,
                    verbose = TRUE,
                    LD = TRUE)
```

```
##                   --------------
##                  |    ddsPLS    |
## ===================--------------====================
## Should we build component 1 ? Bootstrap pending...
##      lambda  R2 R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.38 0.4 0.4 0.38 0.38     40%         40%
##                                       ...component 1 built!
## Should we build component 2 ? Bootstrap pending...
##      lambda   R2  R2h   Q2  Q2h VarExpl VarExpl.Tot
##        0.16 0.61 0.21 0.56 0.28     21%         61%
##                                       ...component 2 built!
## Should we build component 3 ? Bootstrap pending...
##                                       ...component 3 not built!
## ====================             ====================
##                    ================
```

$$\overline{R}^2_B - \overline{Q}^2_B$$

Total explained variance 61%
- ● Comp.1 (40%)
- ● Comp.2 (21%)

λ