Structural Equation Models 2019 / WEEK 4

Mikko Patronen

13 February, 2019

Exercise 4.1

Let us import the data file in R, prepare the data and build the model according to the input file. Then let us draw the graph of the initial, full structural equation model.

```
wk4 <- read.table(file = "allsecondary.DAT")

colnames(wk4) <- c("ROLEA1", "ROLEA2", "ROLEC1", "ROLEC2", "WORK1", "WORK2",
                   "CCLIM1", "CCLIM2", "CCLIM3", "CCLIM4", "DEC1", "DEC2",
                   "SSUP1", "SSUP2", "PSUP1", "PSUP2", "SELF1", "SELF2",
                   "SELF3", "ELC1", "ELC2", "ELC3", "ELC4", "ELC5",
                   "EE1", "EE2", "EE3", "DP1", "DP2", "PA1", "PA2", "PA3")

library(lavaan)
```

```
## This is lavaan 0.6-3
```

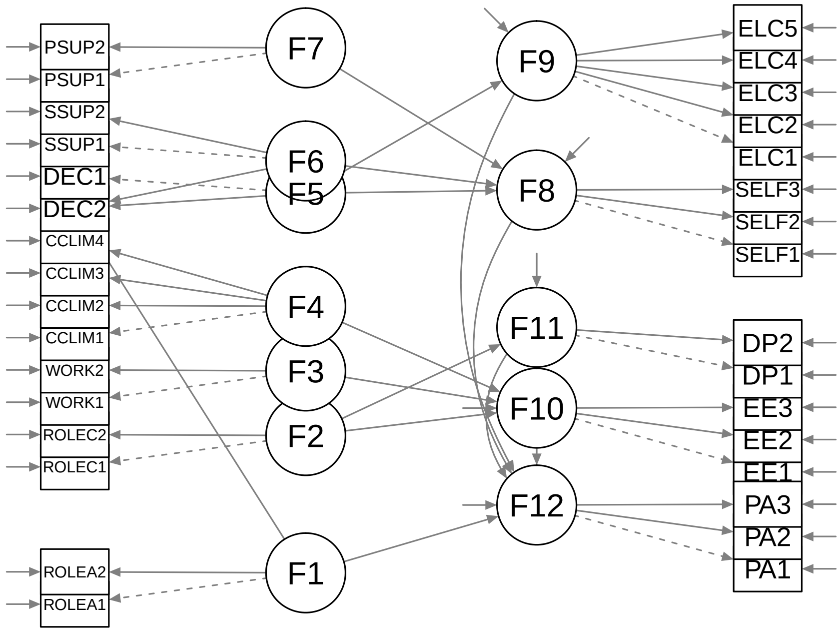
```
## lavaan is BETA software! Please report any bugs.
```

```
model <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6 + F7
F9 ~ F5
F10 ~ F2 + F3 + F4
F11 ~ F2 + F10
F12 ~ F1 + F8 + F9 + F10 + F11
'

fit <- sem(model, data = wk4, estimator="MLM")
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated lv
## variances are negative
```

```
library(semPlot)
```

```r
semPaths(fit, what = "path", whatLabels = "hide", style = "lisrel",
         layout = "tree2", residuals = TRUE, nCharNodes = 0, rotation = 2,
         sizeMan = 6, sizeMan2= 4, sizeLat = 7, esize = 1, label.cex = 1.2,
         optimizeLatRes = TRUE, exoCov = FALSE)
```



Then let us estimate the initial model using the robust MLM estimator and present a brief summary of the model fit:

```r
# summary(fit, fit.measures = T)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
msr <- c("chisq.scaled", "df.scaled", "pvalue.scaled",
         "chisq.scaling.factor", "cfi.robust", "tli.robust",
         "rmsea.robust")

M1 <- c(fitMeasures(fit)[msr]) %>% round(3)
print(M1)
```

```
##         chisq.scaled            df.scaled        pvalue.scaled
##             1541.844              427.000                0.000
## chisq.scaling.factor           cfi.robust           tli.robust
##                1.127                0.945                0.936
##         rmsea.robust
##                0.045
```

In the output Chi square statistic 1541.844 (df = 427, p = 0.000) suggests against a good fit. However the fit indices robust CFI (0.945), robust TLI (0.936) and robust RMSEA (0.045) indicate good fit.

Exercise 4.2

Then let us print the modification indices (MI) for the model:

```r
modindices(fit, minimum.value = 60, sort = TRUE, op="~")
```

```
##      lhs op rhs      mi    epc sepc.lv sepc.all sepc.nox
## 1038 F11  ~  F4 112.597 -0.974  -0.339   -0.339   -0.339
## 1104  F4  ~ F11  90.818 -0.141  -0.404   -0.404   -0.404
## 1037 F11  ~  F3  85.322 -4.799  -4.467   -4.467   -4.467
```

Model 2

F11 ~ F4 has the highest MI value. Following the lecture material we specify another model by adding F11 ~ F4:

```r
model2 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6 + F7
F9 ~ F5
F10 ~ F2 + F3 + F4
F11 ~ F2 + F10 + F4
F12 ~ F1 + F8 + F9 + F10 + F11
'

fit2 <- sem(model2, data = wk4, estimator = "MLM")
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated lv
## variances are negative
```

```r
M2 <- c(fitMeasures(fit2)[msr]) %>% round(3)
print(M2)
```

```
##         chisq.scaled            df.scaled        pvalue.scaled
##             1440.864              426.000                0.000
## chisq.scaling.factor           cfi.robust           tli.robust
##                1.125                0.950                0.942
##         rmsea.robust
```

```
##                 0.043
```

Let us list the indices from two first models so they can be compared:

```
tbl1 <- matrix(c(M1, M2), nrow=2, byrow = T)
nms <- c("ChiSquare", "DF", "p", "ChiSq.sc.factor", "CFI", "TLI", "RMSEA")
colnames(tbl1) <- nms
rownames(tbl1) <- c("M1", "M2")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
```

The model 2 seems to have better fit than the original model: Chi square statistic is still significant 1440.864 (df = 426, p = 0.000) and suggests against a good fit. However the fit indices robust CFI (0.95), robust TLI (0.942) and robust RMSEA (0.043) all indicate slightly better fit than original model.

Then let us calculate the Chi-Square Difference Test:

```
cd <- ((427*1.127)-(426*1.125))/(427-426)
TRd <- (1541.844*1.127 - 1440.864*1.125)/cd
TRd
```

```
## [1] 58.9622
```

```
anova(fit2, fit)
```

```
## Scaled Chi Square Difference Test (method = "satorra.bentler.2001")
##
##       Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit2 426 94568 95105 1620.4
## fit  427 94682 95214 1737.1      58.82       1  1.728e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both tests show almost exactly the same results and indicate significant difference between models.

Model 3

Next, we specify a third model by adding F12 ~ F5:

```
model3 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6 + F7
```

```r
F9 ~ F5
F10 ~ F2 + F3 + F4
F11 ~ F2 + F10 + F4
F12 ~ F1 + F8 + F9 + F10 + F11 + F5
'

fit3 <- sem(model3, data = wk4, estimator = "MLM")
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated lv
## variances are negative
```

```r
M3 <- c(fitMeasures(fit3)[msr]) %>% round(3)
print(M3)
```

```
##         chisq.scaled            df.scaled        pvalue.scaled
##             1396.527              425.000                0.000
## chisq.scaling.factor           cfi.robust           tli.robust
##                1.125                0.952                0.944
##         rmsea.robust
##                0.042
```

```r
tbl1 <- rbind(tbl1, M3)
rownames(tbl1) <- c("M1", "M2", "M3")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
## M3  1396.527 425 0           1.125 0.952 0.944 0.042
```

The third model seems to improve the fit from the 2nd model slightly in the same manner as the second model improved the original model.

```r
cd2 = (426*1.125 - 425*1.125)/(426 - 425)
TRd2 = (1440.864*1.125 - 1396.527*1.125)/cd2
TRd2
```

```
## [1] 44.337
```

The Chi-Square Difference Test indicates that the difference between models is significant.

Model 4

Next, we specify a fourth model by adding F9 ~ F2:

```r
model4 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
```

```
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6 + F7
F9 ~ F5 + F2
F10 ~ F2 + F3 + F4
F11 ~ F2 + F10 + F4
F12 ~ F1 + F8 + F9 + F10 + F11 + F5
'

fit4 <- sem(model4, data = wk4, estimator = "MLM")
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated lv
## variances are negative
```

```
M4 <- c(fitMeasures(fit4)[msr]) %>% round(3)
print(M4)
```

```
##         chisq.scaled            df.scaled        pvalue.scaled
##             1356.810              424.000                0.000
## chisq.scaling.factor           cfi.robust           tli.robust
##                1.125                0.954                0.946
##         rmsea.robust
##                0.042
```

```
tbl1 <- rbind(tbl1, M4)
rownames(tbl1) <- c("M1", "M2", "M3", "M4")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
## M3  1396.527 425 0           1.125 0.952 0.944 0.042
## M4  1356.810 424 0           1.125 0.954 0.946 0.042
```

The fourth model seems to improve the fit from the third model slightly in the same manner as the third model improved the second model.

Model 5

Next, we specify a fifth model by adding F9 ~ F8:

```
model5 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3
```

```
# regressions
F8 ~ F5 + F6 + F7
F9 ~ F5 + F2 + F8
F10 ~ F2 + F3 + F4
F11 ~ F2 + F10 + F4
F12 ~ F1 + F8 + F9 + F10 + F11 + F5
'

fit5 <- sem(model5, data = wk4, estimator = "MLM")
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated lv
## variances are negative
```

```
M5 <- c(fitMeasures(fit5)[msr]) %>% round(3)
```

```
tbl1 <- rbind(tbl1, M5)
rownames(tbl1) <- c("M1", "M2", "M3", "M4", "M5")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
## M3  1396.527 425 0           1.125 0.952 0.944 0.042
## M4  1356.810 424 0           1.125 0.954 0.946 0.042
## M5  1323.061 423 0           1.124 0.956 0.948 0.041
```

The fifth model seems to improve the fit from the fourth model slightly in the same manner as the fourth model improved the third model.

Model 6

Next, we specify a sixth model by adding F10 ~ F8:

```
model6 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6 + F7
F9 ~ F5 + F2 + F8
F10 ~ F2 + F3 + F4 + F8
F11 ~ F2 + F10 + F4
F12 ~ F1 + F8 + F9 + F10 + F11 + F5
'
```

```r
fit6 <- sem(model6, data = wk4, estimator = "MLM")
M6 <- c(fitMeasures(fit6)[msr]) %>% round(3)
```

```r
tbl1 <- rbind(tbl1, M6)
rownames(tbl1) <- c("M1", "M2", "M3", "M4", "M5", "M6")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
## M3  1396.527 425 0           1.125 0.952 0.944 0.042
## M4  1356.810 424 0           1.125 0.954 0.946 0.042
## M5  1323.061 423 0           1.124 0.956 0.948 0.041
## M6  1288.246 422 0           1.124 0.958 0.950 0.040
```

The sixth model seems to improve the fit from the fifth model slightly in the same manner as the fifth model improved the fourth model.

Model 7

Next, we specify a seventh model by adding F12 ~ F2:

```r
model7 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6 + F7
F9 ~ F5 + F2 + F8
F10 ~ F2 + F3 + F4 + F8
F11 ~ F2 + F10 + F4
F12 ~ F1 + F8 + F9 + F10 + F11 + F5 + F2
'

fit7 <- sem(model7, data = wk4, estimator = "MLM")
M7 <- c(fitMeasures(fit7)[msr]) %>% round(3)
```

```r
tbl1 <- rbind(tbl1, M7)
rownames(tbl1) <- c("M1", "M2", "M3", "M4", "M5", "M6", "M7")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
```

```
## M3  1396.527 425 0           1.125 0.952 0.944 0.042
## M4  1356.810 424 0           1.125 0.954 0.946 0.042
## M5  1323.061 423 0           1.124 0.956 0.948 0.041
## M6  1288.246 422 0           1.124 0.958 0.950 0.040
## M7  1267.581 421 0           1.124 0.958 0.951 0.040
```

The seventh model seems to improve the fit from the sixth model slightly in the same manner as the sixth model improved the fifth model.

Model 8

The final model 8 is specified by removing F8 ~ F7 and F12 ~ F1:

```
model8 <- '
# latent variables
F1  =~ ROLEA1 + ROLEA2 + DEC2
F2  =~ ROLEC1 + ROLEC2
F3  =~ WORK1 + WORK2
F4  =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
F5  =~ DEC1 + DEC2
F6  =~ SSUP1 + SSUP2 + DEC2
F7  =~ PSUP1 + PSUP2
F8  =~ SELF1 + SELF2 + SELF3
F9  =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
F10  =~ EE1 + EE2 + EE3
F11  =~ DP1 + DP2
F12  =~ PA1 + PA2 + PA3

# regressions
F8 ~ F5 + F6
F9 ~ F5 + F2 + F8
F10 ~ F2 + F3 + F4 + F8
F11 ~ F2 + F10 + F4
F12 ~ F8 + F9 + F10 + F11 + F5 + F2
'

fit8 <- sem(model8, data = wk4, estimator = "MLM")
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: covariance matrix of latent variables
##                 is not positive definite;
##                 use lavInspect(fit, "cov.lv") to investigate.
```

```
M8 <- c(fitMeasures(fit8)[msr]) %>% round(3)
```

```
tbl1 <- rbind(tbl1, M8)
rownames(tbl1) <- c("M1", "M2", "M3", "M4", "M5", "M6", "M7", "M8")
tbl1
```

```
##    ChiSquare  DF p ChiSq.sc.factor   CFI   TLI RMSEA
## M1  1541.844 427 0           1.127 0.945 0.936 0.045
## M2  1440.864 426 0           1.125 0.950 0.942 0.043
## M3  1396.527 425 0           1.125 0.952 0.944 0.042
## M4  1356.810 424 0           1.125 0.954 0.946 0.042
## M5  1323.061 423 0           1.124 0.956 0.948 0.041
## M6  1288.246 422 0           1.124 0.958 0.950 0.040
## M7  1267.581 421 0           1.124 0.958 0.951 0.040
```

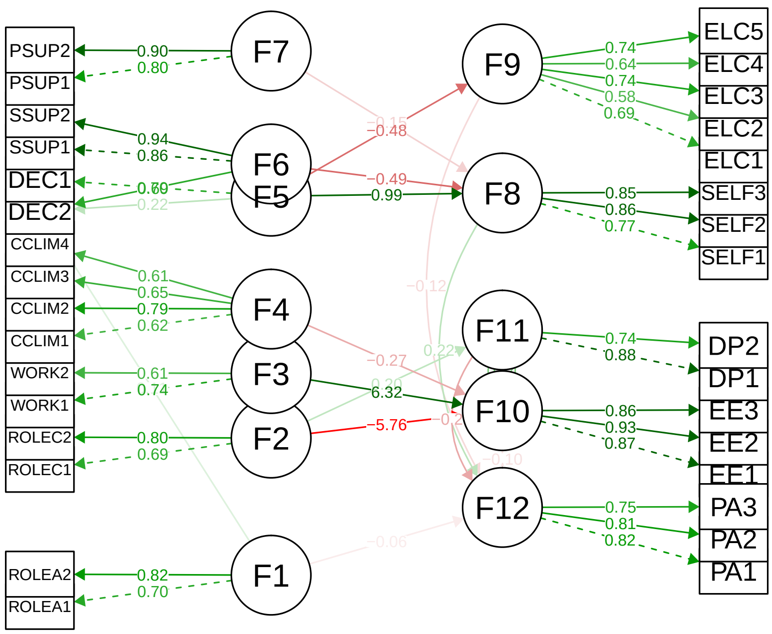
```
## M8  1275.427 423 0           1.123 0.958 0.951 0.040
```

The final model seems to not improve the fit from the seventh model. All in all all the steps have improved the fit nicely: the Chi Square statistic decreased from 1541.844 to 1275.427, CFI increased from 0.945 to 0.958, TLI increased from 0.936 to 0.951 and RMSEA decreased from 0.045 to 0.040.

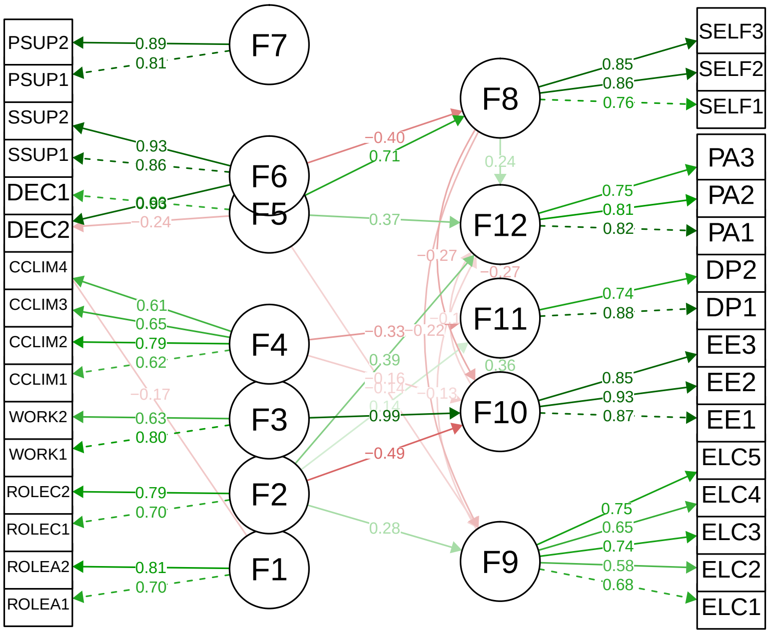
Exercise 4.3

Let us draw the graphs of the original and the final models:

```r
semPaths(fit, what = "std", style = "lisrel", optimizeLatRes = TRUE,
         layout = "tree2", residuals = FALSE, nCharNodes = 0, rotation = 2,
         sizeMan = 6, sizeMan2= 4, sizeLat = 7, esize = 1, label.cex = 1.2,
         exoCov = FALSE)
```




```r
semPaths(fit8, what = "std", style = "lisrel", optimizeLatRes = TRUE,
         layout = "tree2", residuals = FALSE, nCharNodes = 0, rotation = 2,
         sizeMan = 6, sizeMan2= 4, sizeLat = 7, esize = 1, label.cex = 1.2,
         exoCov = FALSE)
```



PSUP2
PSUP1
SSUP2
SSUP1
DEC1
DEC2
CCLIM4
CCLIM3
CCLIM2
CCLIM1
WORK2
WORK1
ROLEC2
ROLEC1
ROLEA2
ROLEA1
F7
F6
F5
F4
F3
F2
F1
F8
F12
F11
F10
F9
SELF3
SELF2
SELF1
PA3
PA2
PA1
DP2
DP1
EE3
EE2
EE1
ELC5
ELC4
ELC3
ELC2
ELC1
0.89
0.81
0.93
0.86
−0.24
0.61
0.65
0.79
0.62
−0.17
0.63
0.80
0.79
0.70
0.81
0.70
−0.40
0.71
0.37
−0.33
0.39
0.99
−0.49
0.28
0.24
−0.27
−0.27
0.36
0.85
0.86
0.76
0.75
0.81
0.82
0.74
0.88
0.85
0.93
0.87
0.75
0.65
0.74
0.58
0.68