STAT 5020
Chapter 3 Bayesian Methods for Estimating Structural Equation Models

Department of Statistics
2021/2022 Term 2

Section 1: Estimating with “lavaan”

lavaan

lavaan (LAtent VAriable ANalysis) is an R package designed for general structural equation modeling. Information and documentation about it can be found on the package's web page: http://www.lavaan.org.

```
> library("lavaan")
This is lavaan 0.6-8
lavaan is FREE software! Please report any bugs.
Warning message:
package 'lavaan' was built under R version 4.0.4
```

lavaan

To compute a model in lavaan requires two steps:

1. **specify the path model**
2. analyze the model.

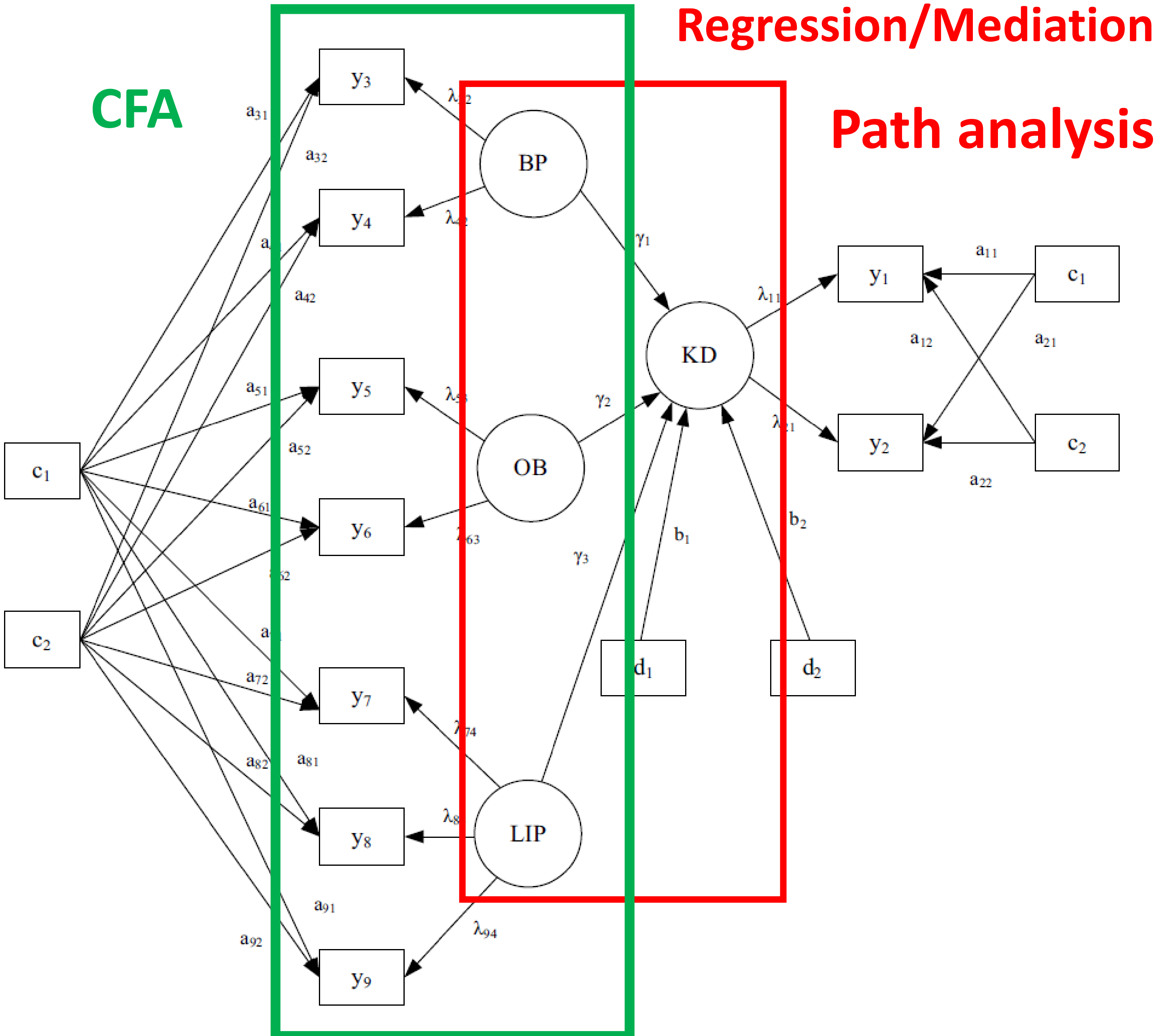
Specification commands

Syntax	Command	Example
`~`	Regress onto	Regress B onto A: `B ~ A`
`~~`	(Co)variance	Variance of A: `A ~~ A` Covariance of A and B: `A ~~ B`

`~1`	Constant/mean/intercept	Regress B onto A, and and include the intercept in the model: `B ~ 1 + A` or
		`B ~ A`
		`B ~ 1`

`=~`	Define reflective latent variable	Define Factor 1 by A-D: `F1 =~A+B+C+D`
`<~`	Define formative latent variable	Define Factor 1 by A-D: `F1 <~ 1*A+B+C+D`
`:=`	Define non-model parameter	Define parameter u2 to be twice the square of u: `u2 := 2*(u^2)`
`*`	Label parameters (the label has to be pre-multiplied)	Label the regression of Z onto X as b: `Z ~b*X`
`|`	Define the number of thresholds (for categorical endogenous variables)	Variable u has three thresholds: `u | t1 + t2 + t3`

lavaan



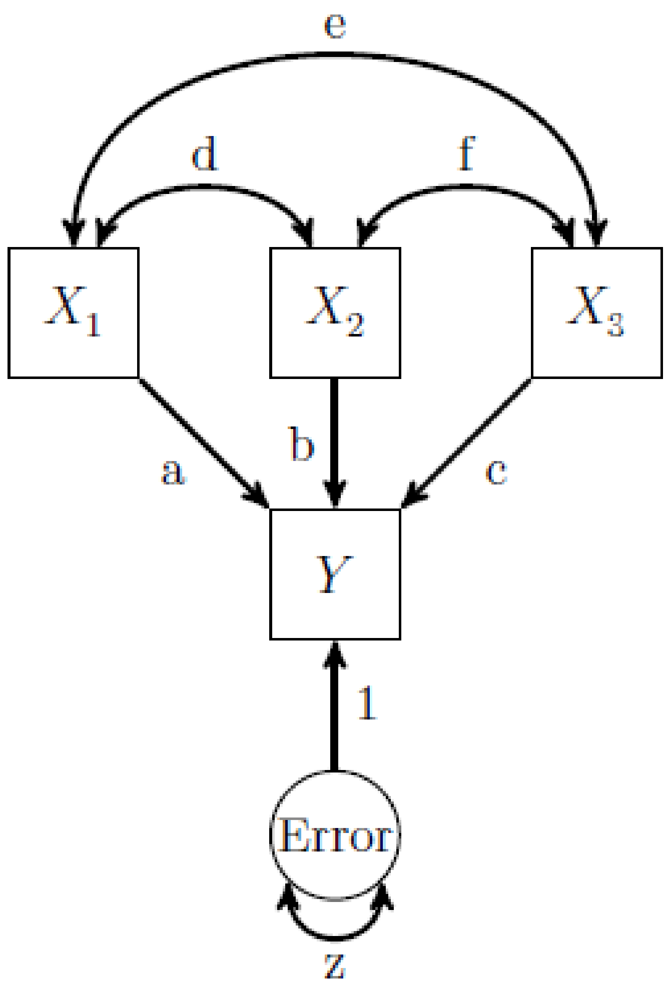
To compute a model in lavaan requires two steps:

1. **specify the path model**
2. analyze the model.

Path analysis

Multiple Regression

```
regression.model<-'
# structural model for Y
Y ~ a*X1 + b*X2 + c*X3
# label the residual variance of Y
Y ~~ z*Y
'
```



Path analysis

Multiple Regression

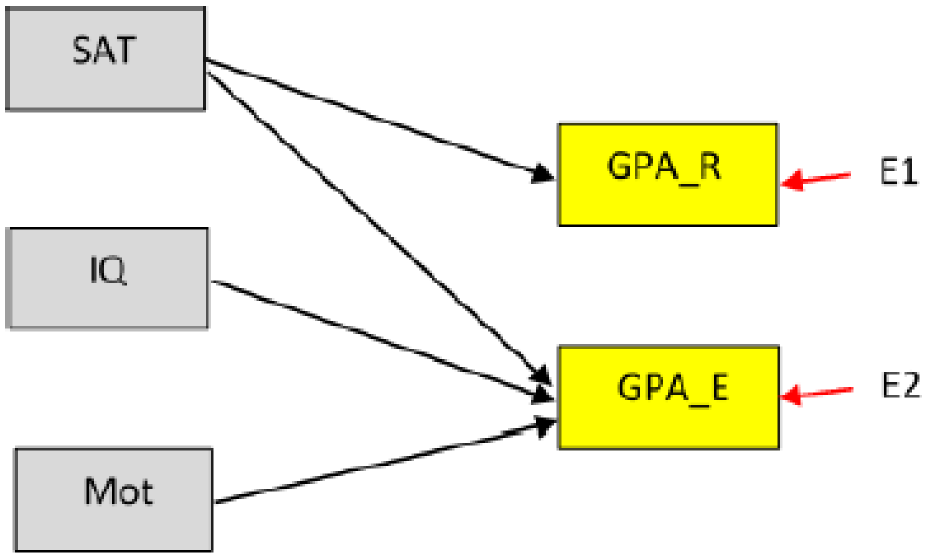
```
regression.model<-'
# structural model for Y
Y ~ a*X1 + b*X2 + c*X3
# label the residual variance of Y
Y ~~ z*Y
'
```

```
> summary(regression.fit, rsquare = TRUE)
lavaan 0.6-8 ended normally after 25 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         4

  Number of observations                          1000

Model Test User Model:

  Test statistic                                 0.000
  Degrees of freedom                                 0

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  Y ~
    X1         (a)    0.571    0.008   74.539    0.000
    X2         (b)    0.700    0.008   89.724    0.000
    X3         (c)   -0.047    0.008   -5.980    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .Y          (z)    0.054    0.002   22.361    0.000

R-Square:
                   Estimate
    Y                 0.946
```

Path analysis

Multiple Regression [Extension]

College Academic Performance (Raykov, 2006)

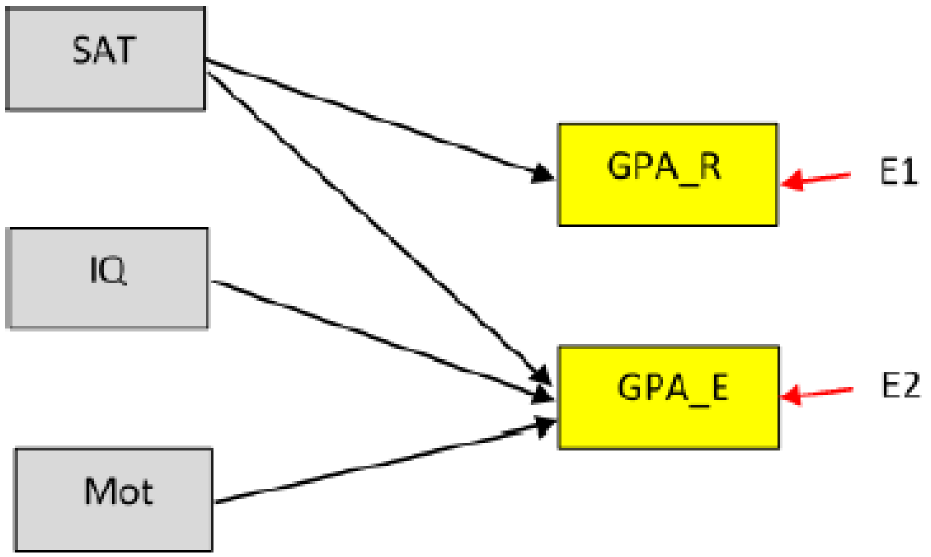


	GPA_R	GPA_E	SAT	IQ	Motiv
GPA_R	.594				
GPA_E	.483	.754			
SAT	3.993	3.626	47.457		
IQ	.426	1.757	4.100	10.267	
Motiv	.500	.722	6.394	.525	2.675

GPA_R=GPA in required courses, GPA_E=GPA in elective courses, SAT=Scholastic aptitude test, IQ=intelligence score, Motiv=motivation

Path analysis

Multiple Regression [Extension]




```
college.model <- "
# structural model for Y
GPA_R ~ a*SAT
GPA_E ~ b*SAT + c*IQ + d*Motiv
# error Variance and Covariance (psi)
GPA_R ~~ e*GPA_R
GPA_E ~~ f*GPA_E
GPA_R ~~ g*GPA_E
"
```

Path analysis

Multiple Regression [Extension]

```
> summary(fit1)
lavaan 0.6-8 ended normally after 32 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         7

  Number of observations                           150
```

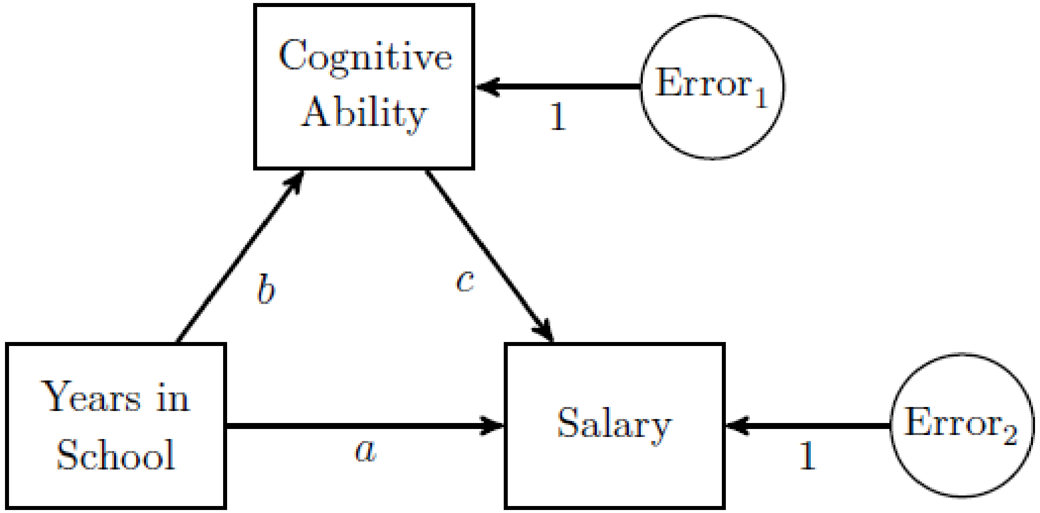
```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  GPA_R ~
    SAT       (a)     0.084    0.006   13.975    0.000
  GPA_E ~
    SAT       (b)     0.045    0.006    6.910    0.000
    IQ        (c)     0.141    0.009   15.550    0.000
    Motiv     (d)     0.145    0.021    6.851    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
 .GPA_R ~~
   .GPA_E     (g)     0.171    0.024    7.001    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .GPA_R     (e)     0.256    0.030    8.660    0.000
   .GPA_E     (f)     0.235    0.027    8.660    0.000
```

Path analysis

Mediation: Indirect Effect

	Salary	Years in School	Cognitive Ability
Salary	648.07	30.05	140.17
Years in School	30.05	8.64	25.57
Cognitive Ability	140.17	25.57	233.20

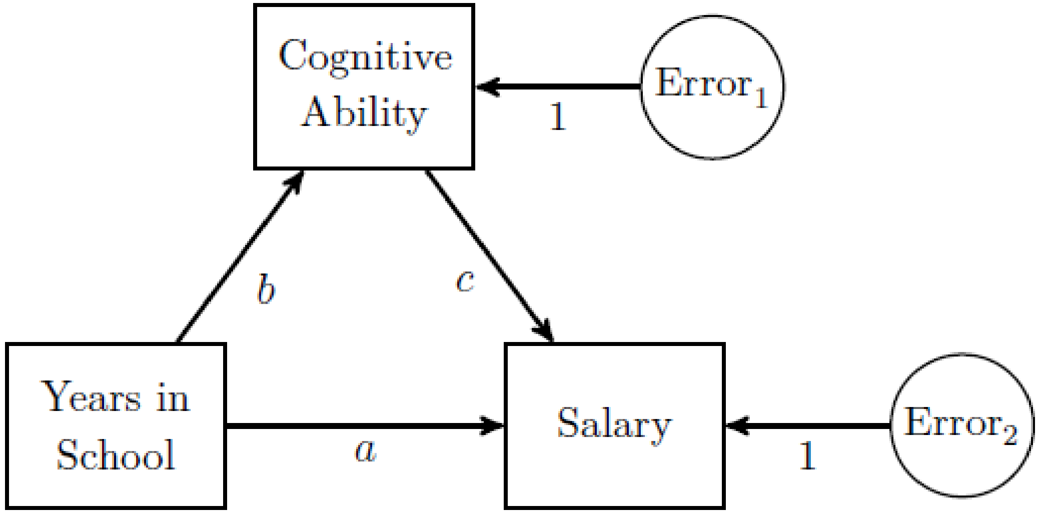
Data taken from Beaujean (2008), and are available at `http://jwosborne.com/bp_ch28.html`



Path analysis

Mediation: Indirect Effect

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  salary ~
    school    (a)     2.515    0.300    8.371    0.000
    iq        (c)     0.325    0.058    5.625    0.000
  iq ~
    school    (b)     2.959    0.135   21.917    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .salary    (d)   526.358   23.539   22.361    0.000
   .iq        (e)   157.378    7.038   22.361    0.000

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ind               0.963    0.177    5.448    0.000
```



Path analysis

Mediation: Indirect Effect

```
Regressions:
                     Estimate  Std.Err  z-value  P(>|z|)
  salary ~
    school     (a)      0.290    0.035    8.371    0.000
    iq         (c)      0.195    0.035    5.625    0.000
  iq ~
    school     (b)      0.570    0.026   21.917    0.000

Variances:
                     Estimate  Std.Err  z-value  P(>|z|)
     .salary   (d)      0.812    0.036   22.361    0.000
     .iq       (e)      0.675    0.030   22.361    0.000

R-Square:
                     Estimate
    salary              0.187
    iq                  0.324

Defined Parameters:
                     Estimate  Std.Err  z-value  P(>|z|)
    ind                 0.111    0.020    5.448    0.000
```

With out direct effect

```
Regressions:
                     Estimate  Std.Err  z-value  P(>|z|)
  salary ~
    iq         (c)      0.361    0.029   12.225    0.000
  iq ~
    school     (b)      0.570    0.026   21.917    0.000

Variances:
                     Estimate  Std.Err  z-value  P(>|z|)
     .salary   (d)      0.869    0.039   22.361    0.000
     .iq       (e)      0.675    0.030   22.361    0.000

R-Square:
                     Estimate
    salary              0.130
    iq                  0.324

Defined Parameters:
                     Estimate  Std.Err  z-value  P(>|z|)
    ind                 0.205    0.019   10.676    0.000
```

Path analysis

Mediation: Indirect Effect

Only direct effect

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  salary ~
    school     (a)    0.402    0.029   13.866    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .salary     (d)    0.838    0.037   22.361    0.000

R-Square:
                   Estimate
    salary            0.161
```

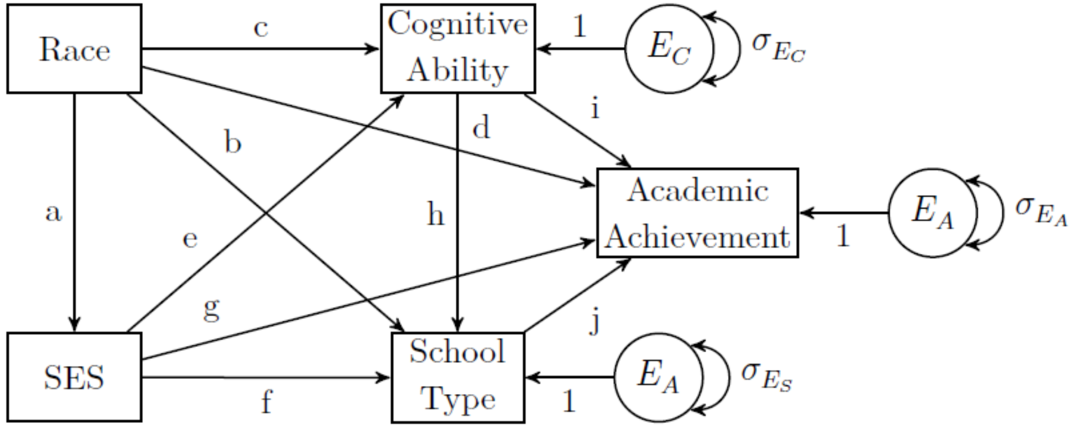
With out direct effect

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  salary ~
    iq         (c)    0.361    0.029   12.225    0.000
  iq ~
    school     (b)    0.570    0.026   21.917    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .salary     (d)    0.869    0.039   22.361    0.000
   .iq         (e)    0.675    0.030   22.361    0.000

R-Square:
                   Estimate
    salary            0.130
    iq                0.324

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ind               0.205    0.019   10.676    0.000
```

Path analysis

Extension

Page and Keith (1981) used respondents from the 1980 High School and Beyond data to investigate the following question: What is the relationship between school type and academic achievement?

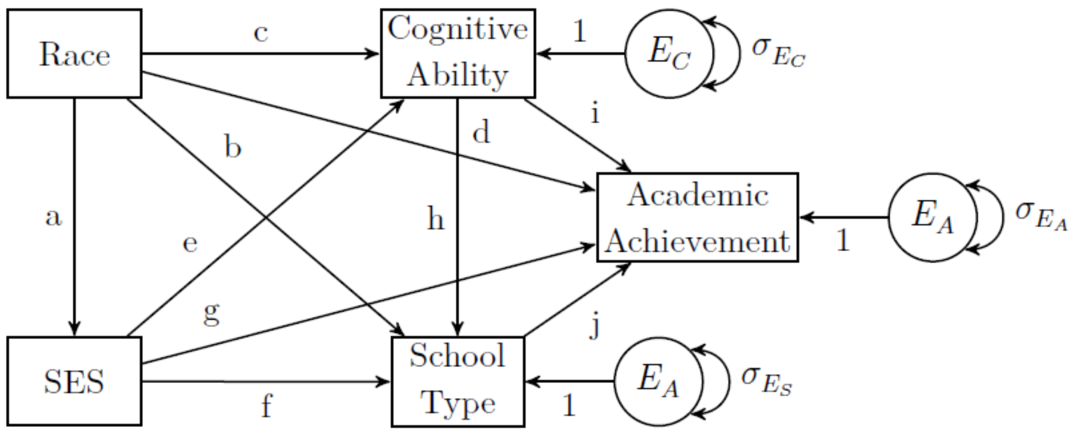


	Race	SES	CogAbil	SchTyp	AcadAch
Race	1.000	0.178	0.230	0.106	0.195
SES	0.178	1.000	0.327	0.245	0.356
Cognitive Ability	0.230	0.327	1.000	0.183	0.721
School Type	0.106	0.245	0.183	1.000	0.178
Academic Achievement	0.195	0.356	0.721	0.178	1.000

Path analysis

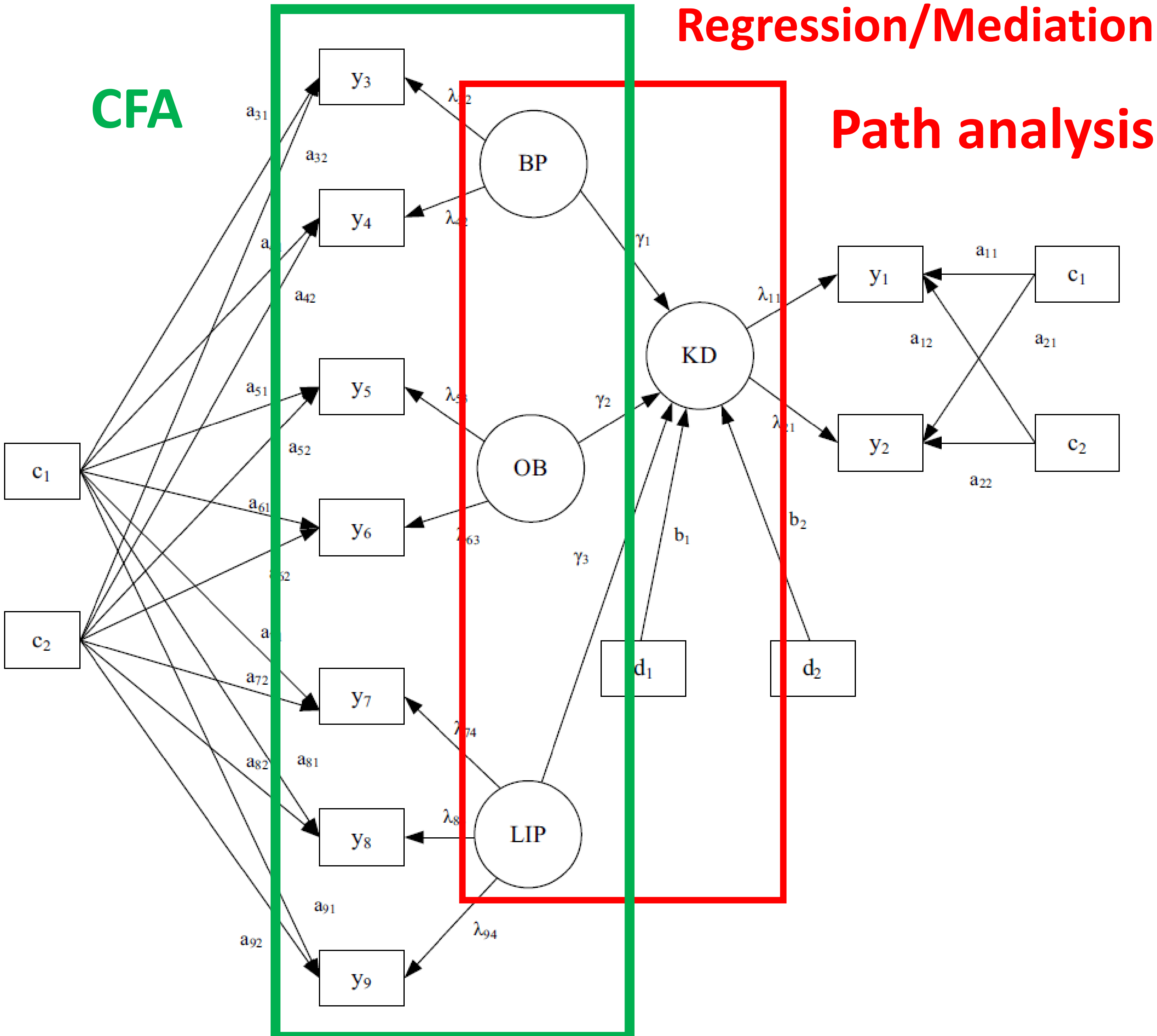
Extension

Page and Keith (1981) used respondents from the 1980 High School and Beyond data to investigate the following question: What is the relationship between school type and academic achievement?




```
full.model <- '
# structural model
AcadAch ~ j*SchTyp + g*SES + d*Race + i*CogAbil
SchTyp ~ f*SES + b*Race + h*CogAbil
CogAbil ~ e*SES + c*Race
SES ~ a*Race
# error Variance and Covariance (psi)
AcadAch ~~  AcadAch
SchTyp ~~  SchTyp
CogAbil ~~  CogAbil
SES ~~ SES
'
```

lavaan

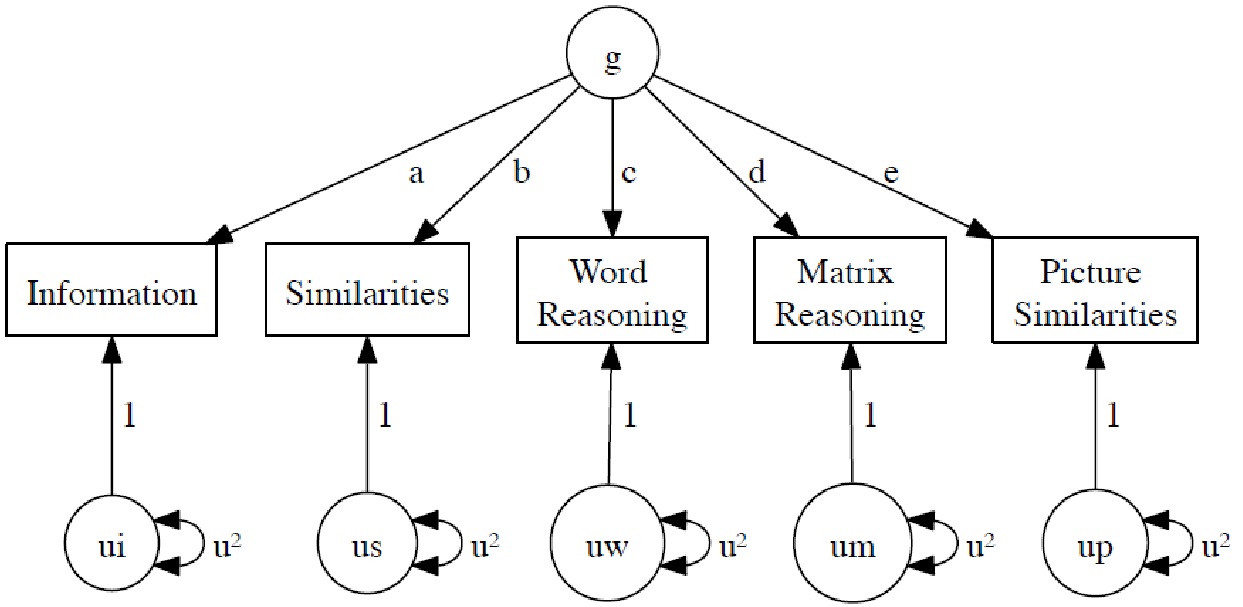


To compute a model in lavaan requires two steps:

1. <u>specify the path model</u>
2. analyze the model.

Factor analysis

Confirmatory Factor Analysis

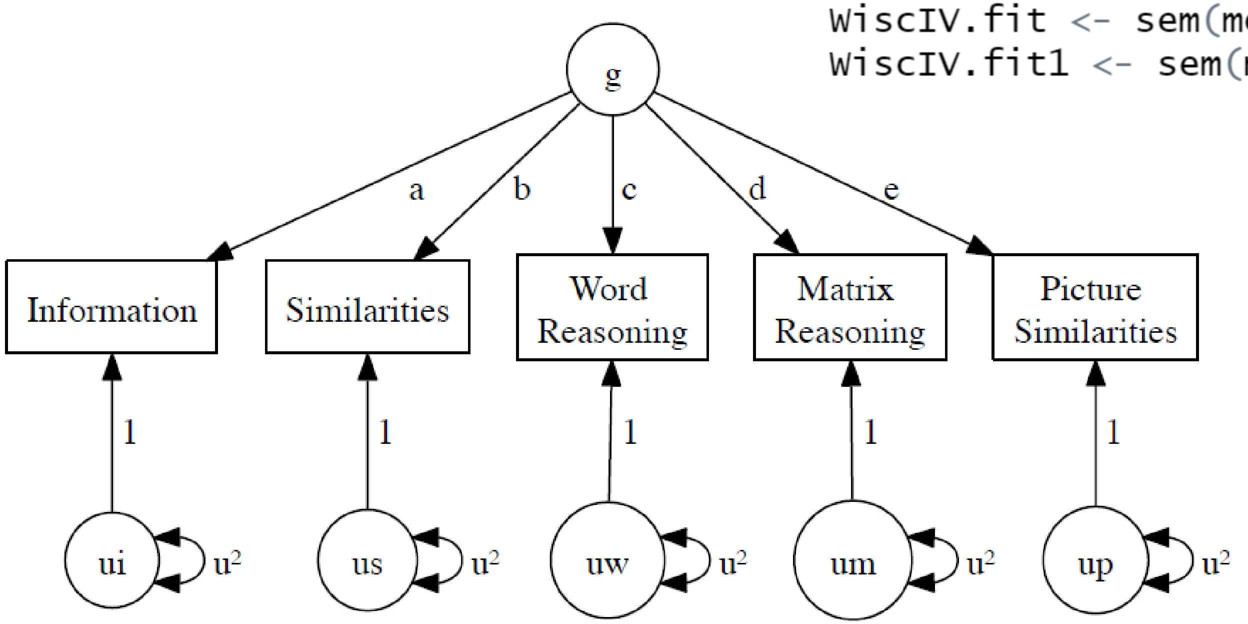


Correlations for the WISC-IV data

	Info	Sim	Word Reas	Matrix Reas	Picture Sim
inss	1.00	0.72	0.64	0.51	0.37
siss	0.72	1.00	0.63	0.48	0.38
wrss	0.64	0.63	1.00	0.37	0.38
mrss	0.51	0.48	0.37	1.00	0.38
psss	0.37	0.38	0.38	0.38	1.00

Factor analysis

Confirmatory Factor Analysis

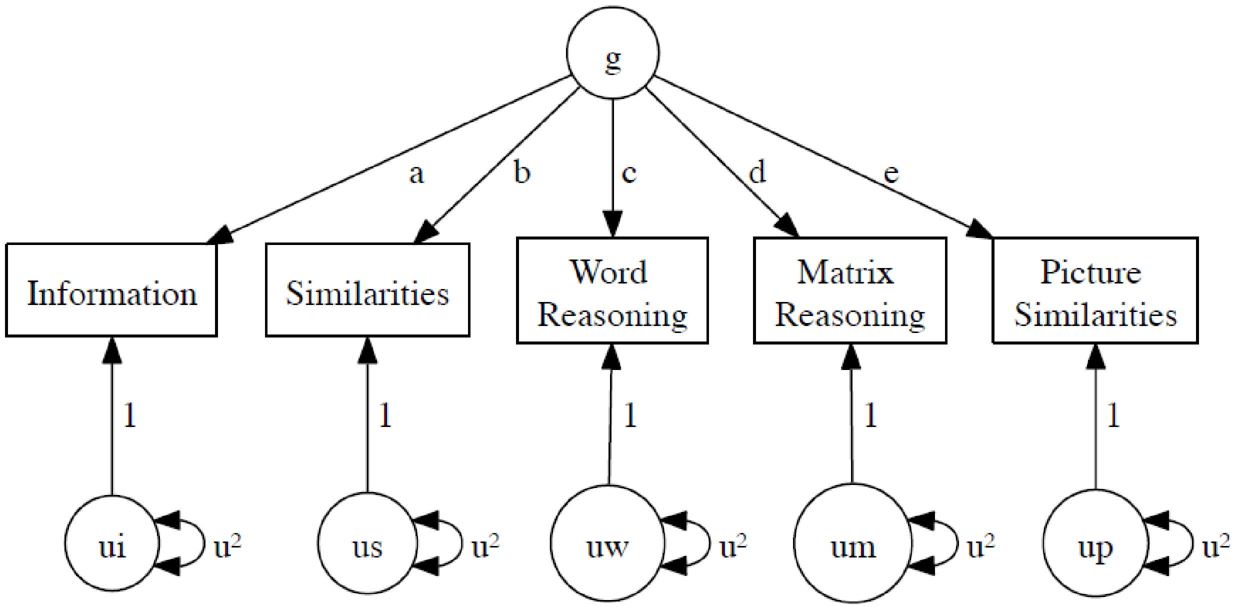



```
model <- '
g=~ a*Info + b*Sim + c*Word + d*Matrix + e*Pict
'
WiscIV.fit <- sem(model,sample.cov= WiscIV.cor,sample.nobs =550)
WiscIV.fit1 <- sem(model,sample.cov= WiscIV.cor,sample.nobs =550,std.lv=T)
```

```
Latent Variables:
                    Estimate  Std.Err
  g =~
    Info        (a)    1.000
    Sim         (b)    0.979    0.045
    Word        (c)    0.866    0.046
    Matrix      (d)    0.674    0.049
    Pict        (e)    0.548    0.050

Variances:
                    Estimate  Std.Err
   .Info               0.264    0.028
   .Sim                0.295    0.028
   .Word               0.448    0.033
   .Matrix             0.665    0.043
   .Pict               0.778    0.049
    g                  0.734    0.062
```

Factor analysis

Confirmatory Factor Analysis

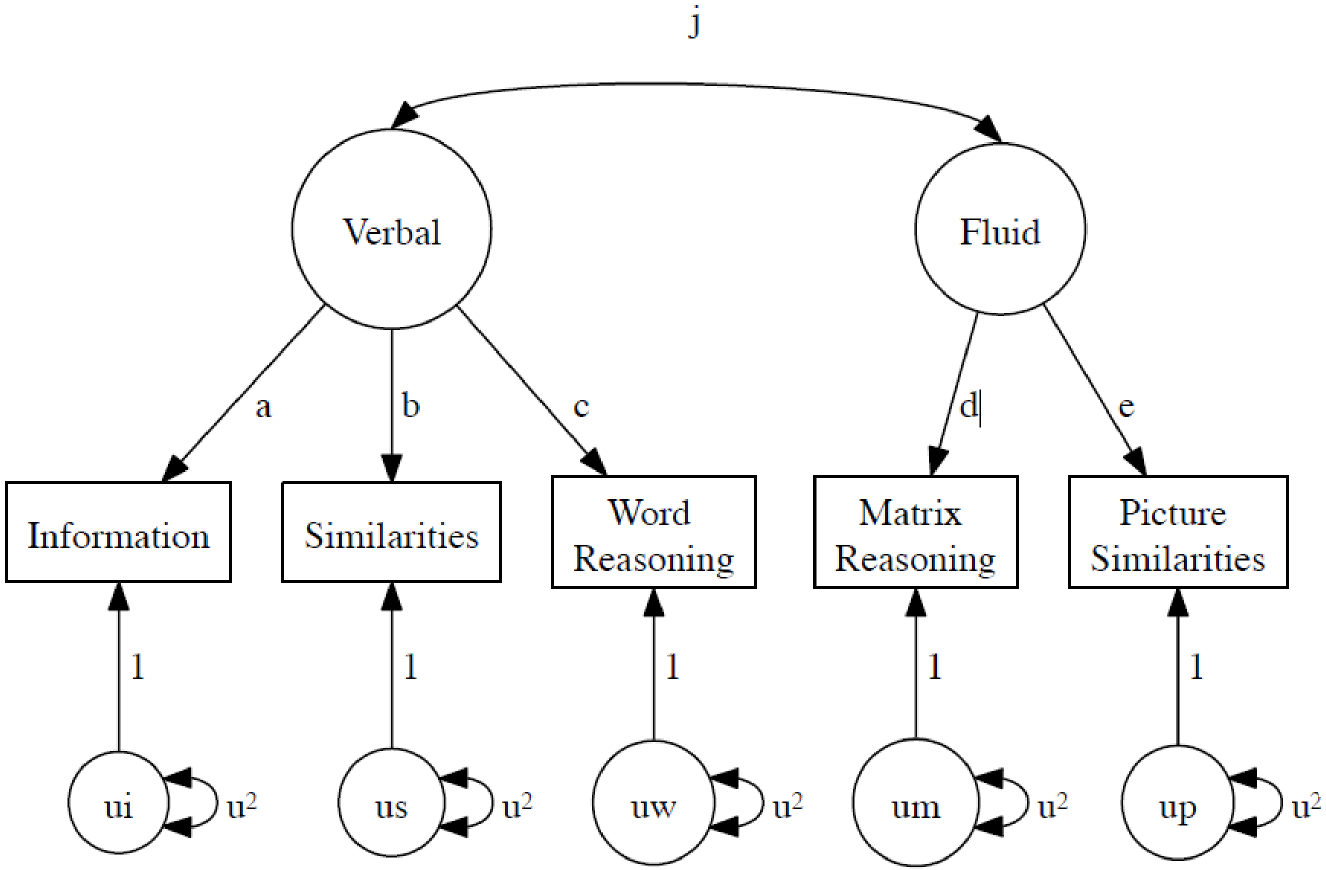



```
model <- '
g=~ NA*Info + b*Sim + c*Word + d*Matrix + e*Pict
g ~~ 1*g
'
```

```
Latent Variables:
                    Estimate
  g =~
    Info       (a)     0.857
    Sim        (b)     0.839
    Word       (c)     0.741
    Matrix     (d)     0.578
    Pict       (e)     0.469

Variances:
                    Estimate
   .Info               0.264
   .Sim                0.295
   .Word               0.448
   .Matrix             0.665
   .Pict               0.778
    g                  1.000
```

Factor analysis

Confirmatory Factor Analysis




```
model3 <- '
# Measurement model
Verbal=~ a*Info + b*Sim + c*Word
Fluid =~ d*Matrix + e*Pict
# error Variance and Covariance (psi)
Verbal ~~ Fluid
'
```

```
model31 <- '
# Measurement model
Verbal=~ NA*Info + b*Sim + c*Word
Fluid =~ NA*Matrix + e*Pict
# error Variance and Covariance (psi)
Verbal ~~ 1*Verbal
Fluid ~~ 1*Fluid
Verbal ~~ Fluid
'
```

Factor analysis

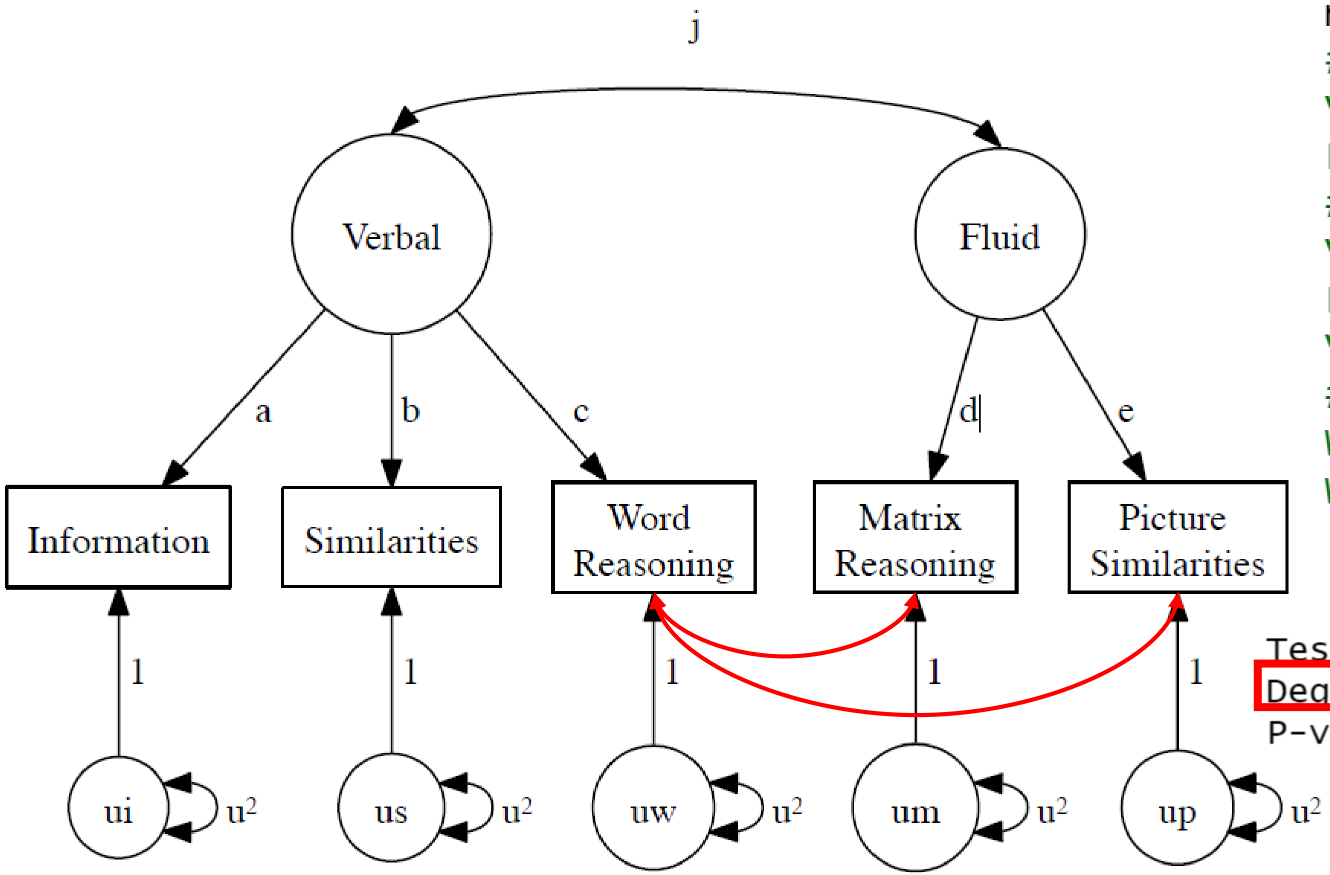
Confirmatory Factor Analysis

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  Verbal =~
    Info              0.859    0.036   23.613    0.000
    Sim        (b)    0.840    0.037   22.867    0.000
    Word       (c)    0.742    0.038   19.303    0.000
  Fluid =~
    Matrix            0.688    0.049   13.920    0.000
    Pict       (e)    0.551    0.047   11.699    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  Verbal ~~
    Fluid             0.823    0.043   19.280    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    Verbal            1.000
    Fluid             1.000
   .Info              0.260    0.028    9.295    0.000
   .Sim               0.292    0.028   10.282    0.000
   .Word              0.447    0.033   13.555    0.000
   .Matrix            0.524    0.055    9.557    0.000
   .Pict              0.695    0.051   13.673    0.000
```

```
Model Test User Model:

  Test statistic                                12.687
  Degrees of freedom                                 4
  P-value (Chi-square)                           0.013

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  Verbal =~
    Info       (a)    1.000
    Sim        (b)    0.978    0.045   21.625    0.000
    Word       (c)    0.864    0.046   18.958    0.000
  Fluid =~
    Matrix     (d)    1.000
    Pict       (e)    0.801    0.082    9.747    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  Verbal ~~
    Fluid             0.487    0.046   10.604    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .Info              0.260    0.028    9.295    0.000
   .Sim               0.292    0.028   10.282    0.000
   .Word              0.447    0.033   13.555    0.000
   .Matrix            0.524    0.055    9.557    0.000
   .Pict              0.695    0.051   13.673    0.000
    Verbal            0.739    0.063   11.807    0.000
    Fluid             0.474    0.068    6.960    0.000
```

Factor analysis

Confirmatory Factor Analysis

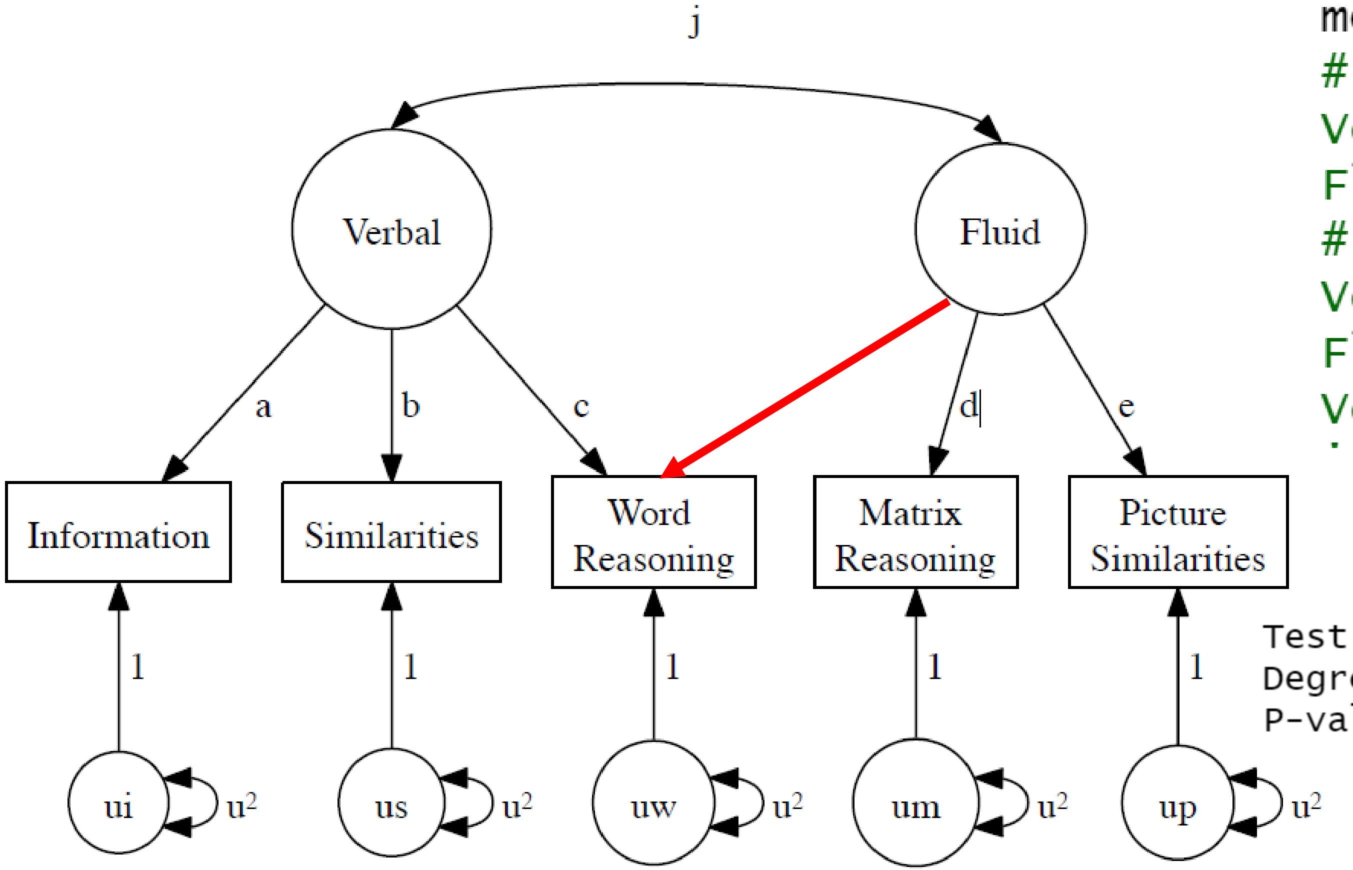



```
model32 <- '
# Measurement model
Verbal=~ NA*Info + b*Sim + c*Word
Fluid =~ NA*Matrix + e*Pict
# error Variance and Covariance (psi)
Verbal ~~ 1*Verbal
Fluid ~~ 1*Fluid
Verbal ~~ Fluid
# error covariances
Word ~~ Matrix
Word ~~ Pict
'
```

```
Test statistic                                 1.117
Degrees of freedom                                 2
P-value (Chi-square)                           0.572
```

Factor analysis

Confirmatory Factor Analysis




```
model34 <- '
# Measurement model
Verbal=~ NA*Info + b*Sim + c*Word
Fluid =~ NA*Matrix + e*Pict + f*Word
# error Variance and Covariance (psi)
Verbal ~~ 1*Verbal
Fluid ~~ 1*Fluid
Verbal ~~ Fluid
'
```

```
Test statistic                     11.612
Degrees of freedom                      3
P-value (Chi-square)                0.009
```

Factor analysis

Goodness of Fit Assessment

- ✓ **Chi-square goodness of fit**
- ✓ **Goodness of fit indices**
  1. ***Absolute fit indexes***
  2. ***Incremental fit indexes***
  3. ***Parsimony indexes***

```
> fitMeasures(WiscIV.fit34)
          npar           fmin              chisq                df
        12.000          0.011             11.612             3.000
        pvalue baseline.chisq        baseline.df   baseline.pvalue
         0.009       1073.427             10.000             0.000
           cfi            tli               nnfi               rfi
         0.992          0.973              0.973             0.964
           nfi           pnfi                ifi               rni
         0.989          0.297              0.992             0.992
          logl unrestricted.logl             aic               bic
     -3368.671      -3362.865           6761.342          6813.061
        ntotal           bic2              rmsea    rmsea.ci.lower
       550.000       6774.968              0.072             0.032
rmsea.ci.upper   rmsea.pvalue                rmr        rmr_nomean
         0.118          0.160              0.019             0.019
          srmr    srmr_bentler srmr_bentler_nomean            crmr
         0.019          0.019              0.019             0.024
   crmr_nomean     srmr_mplus  srmr_mplus_nomean             cn_05
         0.024          0.019              0.019           371.155
         cn_01            gfi               agfi              pgfi
       538.364          0.992              0.958             0.198
           mfi           ecvi
         0.992          0.065
```

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
           npar            fmin           chisq              df
         12.000           0.011          11.612           3.000
         pvalue  baseline.chisq     baseline.df baseline.pvalue
          0.009        1073.427          10.000           0.000
```

✓ **Chi-square goodness of fit**

$$T_{\mathrm{ML}} = (N-1)F_{\mathrm{ML}}(S, \Sigma(\widehat{\theta}_{\mathrm{ML}}))$$

$$T_{\mathrm{GLS}} = (N-1)F_{\mathrm{GLS}}(S, \Sigma(\widehat{\theta}_{\mathrm{GLS}}))$$

$$F_{\mathrm{ML}}(S,\Sigma(\theta)) = \mathrm{tr}(S\Sigma(\theta)^{-1}) + \ln|\Sigma(\theta)| - \ln|S| - p$$

$$F_{\mathrm{GLS}}(S,\Sigma(\theta)) = \mathrm{tr}[(S-\Sigma(\theta))S^{-1}]^2$$

In practice, it is very **difficult** to get a nonsignificant result

- Under H_o: $\Sigma = \Sigma(\theta)$, both T_{ML}, T_{GLS}, and T_{ADF} are asymptotically distributed as a χ^2 variate with $df = p^* - q$
- Reject H_o at α level of significance if $T > \chi^2_\alpha(df = p^* - q)$

Strongly influenced by **sample size**: the larger the sample size the more likely the *chisq* value will be statistically "significant".

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
           npar            fmin           chisq              df 
         12.000           0.011          11.612           3.000 
         pvalue  baseline.chisq     baseline.df baseline.pvalue 
          0.009        1073.427          10.000           0.000 
```

✓ **Chi-square goodness of fit**

Likelihood-ratio (LR) or chi-square difference test

H_o: no difference between M_0 and M_1 in terms of goodness of fit

$$\Delta\chi^2 = T_0 - T_1 \sim \chi^2(df_0 - df_1)$$

Reject H_o if $\Delta\chi^2$ is significant at $\alpha = 0.05$

- LR test evaluates the **decrease in the goodness-of-fit** when a less restricted model M_1 is further restricted
- The **greater** the reduction, the **less** likely that the restricted model M_0 is true

```
> anova(WiscIV.fit31,WiscIV.fit32)
Chi-Squared Difference Test

             Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)   
WiscIV.fit32  2 6752.8 6808.9  1.117                                 
WiscIV.fit31  4 6760.4 6807.8 12.687      11.57       2   0.003074 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Factor analysis

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
          npar           fmin          chisq             df
        12.000          0.011         11.612          3.000
        pvalue baseline.chisq    baseline.df baseline.pvalue
         0.009       1073.427         10.000           0.000
```

- ✓ **Chi-square goodness of fit**

Likelihood-ratio (LR) or chi-square difference test

H_o: no difference between M_0 and M_1 in terms of goodness of fit

$$\Delta\chi^2 = T_0 - T_1 \sim \chi^2(df_0 - df_1)$$

Reject H_o if $\Delta\chi^2$ is significant at $\alpha = 0.05$

```
> anova(WiscIV.fit31,WiscIV.fit34)
Chi-Squared Difference Test

             Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
WiscIV.fit34  3 6761.3 6813.1 11.612
WiscIV.fit31  4 6760.4 6807.8 12.687     1.0751       1     0.2998
```

- LR test evaluates the **decrease in the goodness-of-fit** when a less restricted model M_1 is further restricted
- The **greater** the reduction, the **less** likely that the restricted model M_0 is true

```
> modificationIndices(WiscIV.fit31,sort.=TRUE)
     lhs op    rhs    mi    epc sepc.lv sepc.all sepc.nox
26  Word ~~ Matrix 8.421 -0.078  -0.078   -0.162   -0.162
27  Word ~~   Pict 5.574  0.067   0.067    0.120    0.120
21  Info ~~ Matrix 4.640  0.056   0.056    0.153    0.153
22  Info ~~   Pict 3.143 -0.047  -0.047   -0.111   -0.111
18 Fluid =~   Word 1.009 -0.111  -0.111   -0.111   -0.111
19  Info ~~    Sim 1.009 -0.042  -0.042   -0.154   -0.154
16 Fluid =~   Info 0.526  0.082   0.082    0.082    0.082
```

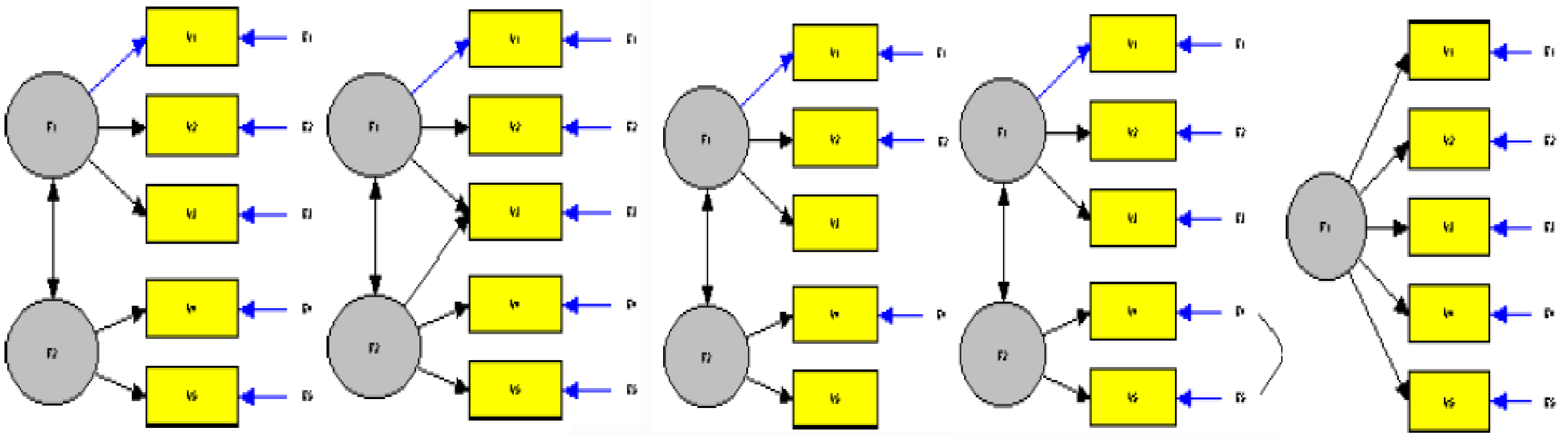
Factor analysis

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
          npar           fmin          chisq             df
        12.000          0.011         11.612          3.000
        pvalue baseline.chisq    baseline.df baseline.pvalue
         0.009       1073.427         10.000          0.000
```

✓ **Chi-square goodness of fit**

Likelihood-ratio (LR) or chi-square difference test



- **Nested models :** any model which requires that some function of its free parameters equals another free parameter or equals a constant is nested in the identical model that has no such restriction

Factor analysis

Goodness of Fit Assessment

- ✓ **Goodness of fit indices**
  1. ***Absolute fit indexes***
  2. ***Incremental fit indexes***
  3. ***Parsimony indexes***

```
> fitMeasures(WiscIV.fit34)
               npar                fmin               chisq                  df 
             12.000               0.011              11.612               3.000 
             pvalue      baseline.chisq         baseline.df     baseline.pvalue 
              0.009            1073.427              10.000               0.000 
                cfi                 tli                nnfi                 rfi 
              0.992               0.973               0.973               0.964 
                nfi                pnfi                 ifi                 rni 
              0.989               0.297               0.992               0.992 
               logl   unrestricted.logl                 aic                 bic 
          -3368.671           -3362.865            6761.342            6813.061 
             ntotal                bic2               rmsea      rmsea.ci.lower 
            550.000            6774.968               0.072               0.032 
     rmsea.ci.upper        rmsea.pvalue                 rmr          rmr_nomean 
              0.118               0.160               0.019               0.019 
               srmr        srmr_bentler srmr_bentler_nomean                crmr 
              0.019               0.019               0.019               0.024 
        crmr_nomean          srmr_mplus   srmr_mplus_nomean               cn_05 
              0.024               0.019               0.019             371.155 
              cn_01                 gfi                agfi                pgfi 
            538.364               0.992               0.958               0.198 
                mfi                ecvi 
              0.992               0.065 
```

Factor analysis

Goodness of Fit Assessment

- ✓ **Goodness of fit indices**
  1. ***Absolute fit indexes***

- Measure how well a model reproduces the sample covariance matrix

<u>**Closer to 0.0 indicate better fit.**</u>

```
> fitMeasures(WiscIV.fit34)
```

Cov or *Cor*

(Standardized) Root Mean Square Residual (SRMR/RMR)

(a) produce the residual correlation matrix;
(b) remove the redundant values;
(c) square the remaining values;
(d) sum all the squared values;
(e) divide the sum by the number of non-redundant elements in the matrix;
(f) take the square root of the resulting number.

```
> residuals(WiscIV.fit34)$cov
        Info   Sim    Word   Matrix Pict
Info     0.000
Sim      0.000  0.000
Word    -0.002  0.002  0.000
Matrix   0.015 -0.004 -0.036  0.000
Pict    -0.019  0.000  0.062  0.000  0.000
```

```
                                                  rmr         rmr_nomean
                                                0.019              0.019
         srmr        srmr_bentler srmr_bentler_nomean               crmr
        0.019               0.019               0.019              0.024
  crmr_nomean         srmr_mplus   srmr_mplus_nomean
        0.024               0.019               0.019
```

Factor analysis

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
```

- ✓ **Goodness of fit indices**
  1. ***Absolute fit indexes***

- Measure how well a model reproduces the sample covariance matrix

Closer to 1.0 indicate better fit.

Goodness of Fit Index

$$GFI = 1 - \frac{\mathrm{tr}[(\hat{\Sigma}^{-1}S - I)^2]}{\mathrm{tr}[(\hat{\Sigma}^{-1}S)^2]}$$

GFI measures the relative amount of the variance and covariance in S that are predicted by $\hat{\Sigma}$ (similar to in regression)

Adjusted Goodness of Fit Index

$$AGFI = 1 - \frac{p(p+1)}{2df}(1 - GFI)$$

$AGFI$ adjusts for the df of a model, it rewards simpler model with fewer parameters

$$\text{When } S = \hat{\Sigma},\ GFI = AGFI = 1.00$$

```
  gfi    agfi    pgfi
0.992   0.958   0.198
```

Factor analysis

Goodness of Fit Assessment

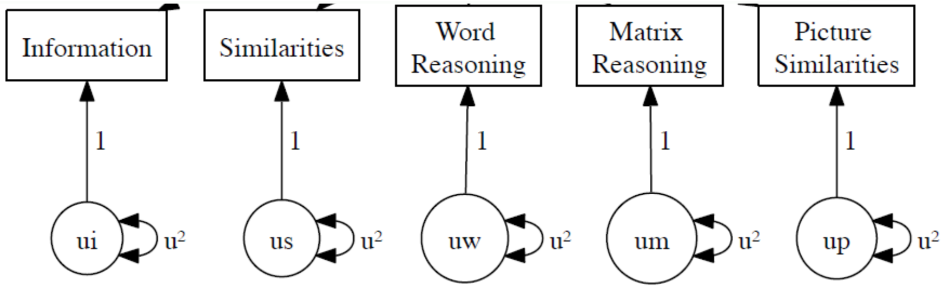
```
> fitMeasures(WiscIV.fit34)
           npar            fmin           chisq              df
         12.000           0.011          11.612           3.000
         pvalue  baseline.chisq     baseline.df baseline.pvalue
          0.009        1073.427          10.000           0.000
            cfi             tli            nnfi             rfi
          0.992           0.973           0.973           0.964
            nfi            pnfi             ifi             rni
          0.989           0.297           0.992           0.992
```

✓ **Goodness of fit indices**

2. Incremental fit indexes

- Measure the relative improvement in fit by comparing a target model with a baseline (independence) model

Baseline (independence) model




```
modelbaseline <- '
# Measurement model
Info ~~ Info
Sim ~~ Sim
Word ~~ Word
Matrix ~~ Matrix
Pict ~~ Pict
'
```

Factor analysis

Goodness of Fit Assessment

✓ **Goodness of fit indices**

2. *Incremental fit indexes*

- Measure the relative improvement in fit by comparing a target model with **a baseline (independence) model**

```
> fitMeasures(WiscIV.fit34)
           npar            fmin           chisq              df
         12.000           0.011          11.612           3.000
         pvalue  baseline.chisq     baseline.df baseline.pvalue
          0.009        1073.427          10.000           0.000
            cfi             tli            nnfi             rfi
          0.992           0.973           0.973           0.964
            nfi            pnfi             ifi             rni
          0.989           0.297           0.992           0.992
```

Normed Fit Index

Measure the proportionate reduction in the chi-square values when moving from baseline to hypothesized model

$$NFI = \frac{T_B - T_T}{T_B}$$

***NFI* values closer to 1.0 indicate better fit**

[No control for degrees of freedom]

Factor analysis

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
```

✓ **Goodness of fit indices**

3. Parsimony indexes

- Model complexity is taken into account when assessing model fit.
- <u>**Models with more parameters (fewer degrees of freedom) are penalized**</u>

```
          logl unrestricted.logl               aic               bic
     -3368.671         -3362.865          6761.342          6813.061
        ntotal              bic2             rmsea    rmsea.ci.lower
       550.000          6774.968             0.072             0.032
rmsea.ci.upper     rmsea.pvalue
         0.118             0.160

           mfi              ecvi
         0.992             0.065
```

Factor analysis

Goodness of Fit Assessment

```
> fitMeasures(WiscIV.fit34)
```

Information-Theoretic Criterion

Information-theoretic methods emphasize minimizing the amount of parameters needed in a model to explain the data. Thus, these indexes attempt to select models that are the most parsimonious/efficient representations of the observed data.

✓ **Goodness of fit indices**

```
     logl   unrestricted.logl                 aic                 bic
-3368.671           -3362.865            6761.342            6813.061
   ntotal                bic2
  550.000            6774.968
```

smaller values indicating better fit

3. Parsimony indexes

- Model complexity is taken into account when assessing model fit.
- **Models with more parameters (fewer degrees of freedom) are penalized**

1. Akaike's information criterion (AIC)
2. Schwarz's Bayesian information criterion (BIC)
3. Sample-Size Adjusted Bayesian information criterion (SABIC)

$$AIC^* = -2\ln(L) + 2q$$

$$BIC^* = -2\ln(L) + q\ln(n)$$

$$SABIC^* = -2\ln(L) + q \times \ln((n+2)/24)$$

L is the likelihood function's value.

Factor analysis

Confirmatory Factor Analysis

The correlations from McIver, Carmines, and Zeller's (1980) study of attitudes toward police. They are based on telephone interviews with a total of some 11,000 respondents in 60 neighborhoods in three U.S. metropolitan areas..

General dimension of attitude toward police

Likelihood of burglary, vandalism, robbery in the neighborhood

	1	2	3	4	5	6	7	8	9
1. Police service	1.00	.50	.41	.33	.28	.30	−.24	−.23	−.20
2. Responsiveness		1.00	.35	.29	.26	.27	−.19	−.19	−.18
3. Response time			1.00	.30	.27	.29	−.17	−.16	−.14
4. Honesty				1.00	.52	.48	−.13	−.11	−.15
5. Courtesy					1.00	.44	−.11	−.09	−.10
6. Equal treatment						1.00	−.15	−.13	−.13
7. Burglary							1.00	.58	.47
8. Vandalism								1.00	.42
9. Robbery									1.00

Factor analysis

Confirmatory Factor Analysis

The correlations from McIver, Carmines, and Zeller's (1980) study of attitudes toward police. They are based on telephone interviews with a total of some 11,000 respondents in 60 neighborhoods in three U.S. metropolitan areas..

Reflecting attitudes toward the quality of police services

Likelihood of burglary, vandalism, robbery in the neighborhood

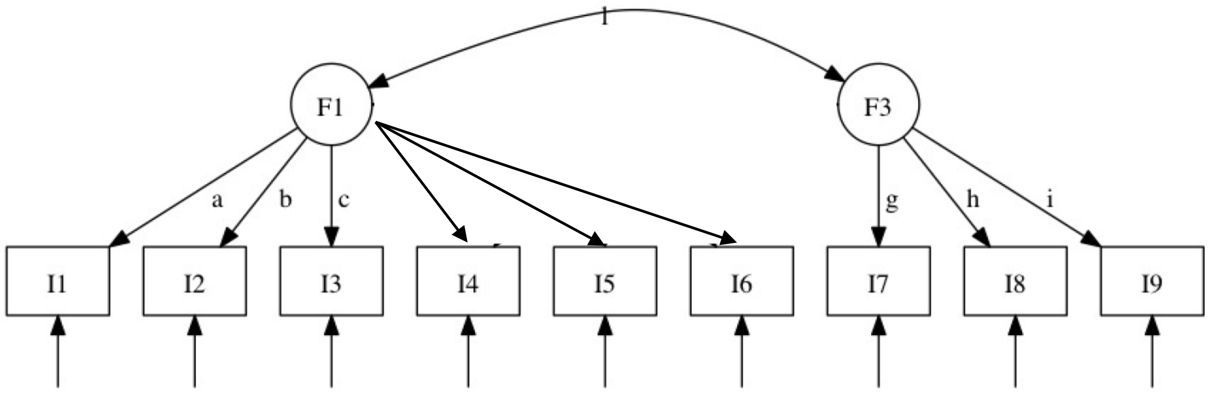
Personal qualities of the police

	1	2	3	4	5	6	7	8	9
1. Police service	1.00	.50	.41	.33	.28	.30	−.24	−.23	−.20
2. Responsiveness		1.00	.35	.29	.26	.27	−.19	−.19	−.18
3. Response time			1.00	.30	.27	.29	−.17	−.16	−.14
4. Honesty				1.00	.52	.48	−.13	−.11	−.15
5. Courtesy					1.00	.44	−.11	−.09	−.10
6. Equal treatment						1.00	−.15	−.13	−.13
7. Burglary							1.00	.58	.47
8. Vandalism								1.00	.42
9. Robbery									1.00

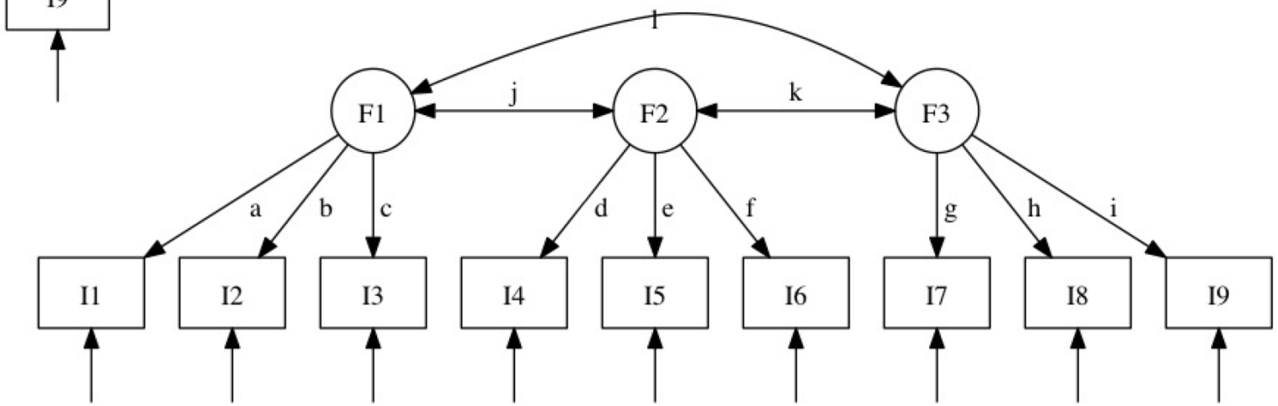
Factor analysis

Confirmatory Factor Analysis

The correlations from McIver, Carmines, and Zeller’s (1980) study of attitudes toward police. They are based on telephone interviews with a total of some 11,000 respondents in 60 neighborhoods in three U.S. metropolitan areas..



vs



Factor analysis

Confirmatory Factor Analysis

```
> residuals(p1.1.fit)
$type
[1] "raw"

$cov
       PS     RE     RT     HO     CO     ET     BU     VA     RO
PS  0.000                                                        
RE  0.173  0.000                                                 
RT  0.094  0.062  0.000                                          
HO -0.069 -0.074 -0.052  0.000                                   
CO -0.086 -0.074 -0.052  0.112  0.000                            
ET -0.068 -0.066 -0.034  0.070  0.064  0.000                     
BU -0.072 -0.037 -0.022  0.057  0.061  0.022  0.000              
VA -0.079 -0.052 -0.027  0.059  0.065  0.025  0.003  0.000       
RO -0.076 -0.067 -0.031 -0.012  0.026 -0.003 -0.002 -0.006  0.000
```

```
> residuals(p1.fit)
$type
[1] "raw"

$cov
       PS     RE     RT     HO     CO     ET     BU     VA     RO
PS  0.000                                                        
RE  0.016  0.000                                                 
RT -0.009 -0.019  0.000                                          
HO -0.015 -0.013  0.038  0.000                                   
CO -0.033 -0.015  0.032  0.009  0.000                            
ET  0.001  0.007  0.063 -0.008 -0.003  0.000                     
BU  0.000  0.021  0.013  0.013  0.019 -0.026  0.000              
VA -0.011  0.002  0.006  0.020  0.028 -0.017  0.003  0.000       
RO -0.022 -0.023 -0.005 -0.044 -0.004 -0.038  0.000 -0.007  0.000
```

Factor analysis

Confirmatory Factor Analysis

Item	Factor pattern F_1	F_2	F_3	h^2
1. Police service	.74	.00	.00	.55
2. Responsiveness	.65	.00	.00	.43
3. Response time	.56	.00	.00	.32
4. Honesty	.00	.75	.00	.56
5. Courtesy	.00	.68	.00	.46
6. Equal treatment	.00	.65	.00	.42
7. Burglary	.00	.00	.80	.63
8. Vandalism	.00	.00	.72	.52
9. Robbery	.00	.00	.59	.35

Factor correlations

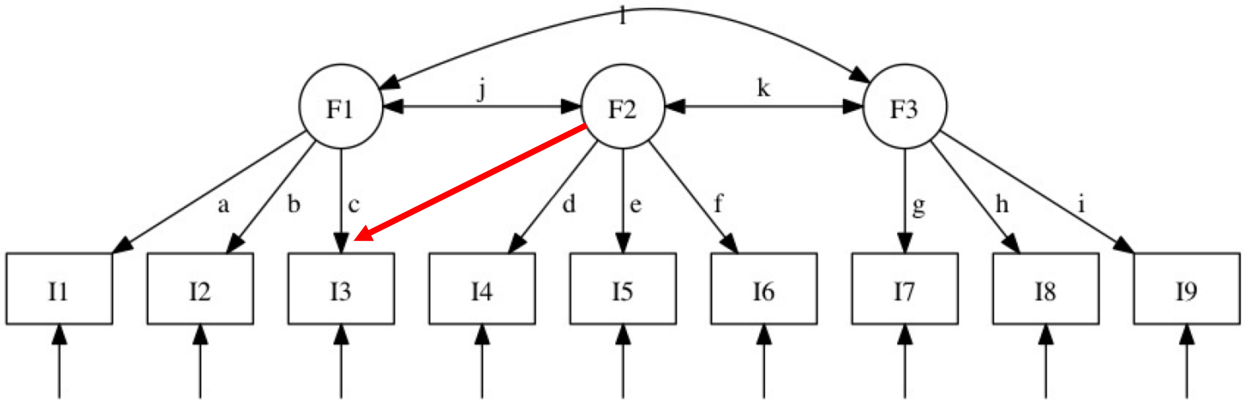
	F_1	F_2	F_3
F_1	1.00	.62	−.41
F_2		1.00	−.24
F_3			1.00

```
> modificationindices(p1.fit,sort=T)
   lhs op rhs      mi    epc sepc.lv sepc.all sepc.nox
33  F2 =~  RT 100.693  0.160   0.160    0.160    0.160
43  PS ~~  RE  69.015  0.102   0.102    0.201    0.201
31  F2 =~  PS  43.204 -0.122  -0.122   -0.122   -0.122
36  F2 =~  RO  31.732 -0.057  -0.057   -0.057   -0.057
68  HO ~~  RO  30.842 -0.036  -0.036   -0.068   -0.068
60  RT ~~  ET  29.191  0.038   0.038    0.061    0.061
```

```
> residuals(p1.fit)
$type
[1] "raw"

$cov
   PS     RE     RT     HO     CO     ET     BU     VA     RO
PS  0.000
RE  0.016  0.000
RT -0.009 -0.019  0.000
HO -0.015 -0.013  0.038  0.000
CO -0.033 -0.015  0.032  0.009  0.000
ET  0.001  0.007  0.063 -0.008 -0.003  0.000
BU  0.000  0.021  0.013  0.013  0.019 -0.026  0.000
VA -0.011  0.002  0.006  0.020  0.028 -0.017  0.003  0.000
RO -0.022 -0.023 -0.005 -0.044 -0.004 -0.038  0.000 -0.007  0.000
```

Factor analysis

Confirmatory Factor Analysis



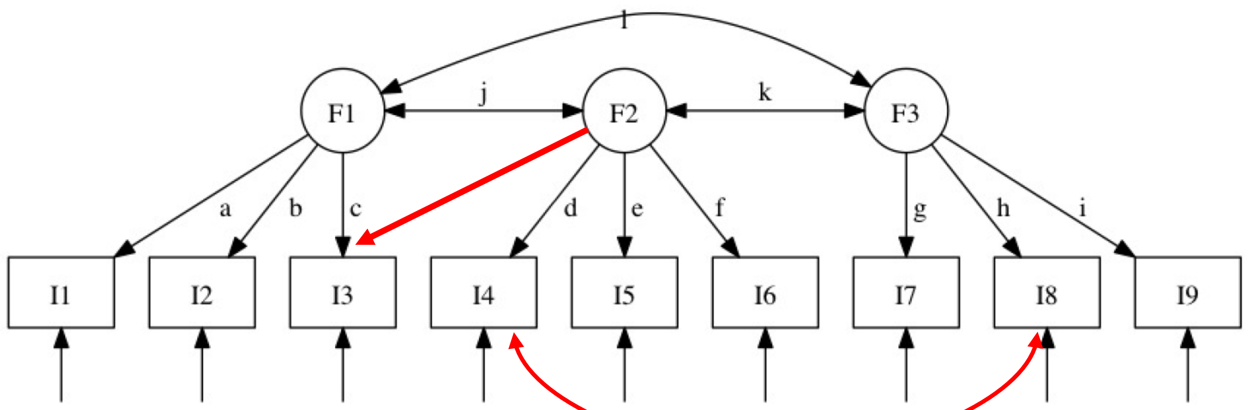

```
modelp2 <- '
# Measurement model
F1 =~ a*PS + b*RE + c*RT
F2 =~ d*HO + e*CO + f*ET + RT
F3 =~ g*BU + h*VA + i*RO
# error Variance and Covariance (psi)
F1 ~~ j*F2
F1 ~~ k*F3
F2 ~~ l*F3
'
p2.fit=sem(modelp2,sample.cov= police.cor,sample.nobs =11000,std.lv=T)
```

```
Latent Variables:
                   Estimate
  F1 =~
    PS         (a)    0.760
    RE         (b)    0.657
    RT         (c)    0.451
  F2 =~
    HO         (d)    0.749
    CO         (e)    0.681
    ET         (f)    0.651
    RT                0.148
  F3 =~
    BU         (g)    0.796
    VA         (h)    0.725
    RO         (i)    0.590

Covariances:
                   Estimate
  F1 ~~
    F2         (j)    0.582
    F3         (k)   -0.404
  F2 ~~
    F3         (l)   -0.239
```

```
> anova(p1.fit,p2.fit)
Chi-Squared Difference Test

       Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
p2.fit 23 256978 257139 127.32
p1.fit 24 257075 257228 226.23     98.915       1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Factor analysis

Confirmatory Factor Analysis




```
> modificationindices(p2.fit,sort=T)
   lhs op rhs     mi    epc sepc.lv sepc.all sepc.nox
68  HO ~~  RO 31.322 -0.036  -0.036   -0.068   -0.068
36  F2 =~  RO 31.013 -0.056  -0.056   -0.057   -0.057
64  HO ~~  CO 30.481  0.067   0.067    0.138    0.138
42  F3 =~  ET 28.044 -0.053  -0.053   -0.053   -0.053
28  F1 =~  ET 24.836  0.074   0.074    0.074    0.074
76  BU ~~  VA 19.161  0.101   0.101    0.241    0.241
```

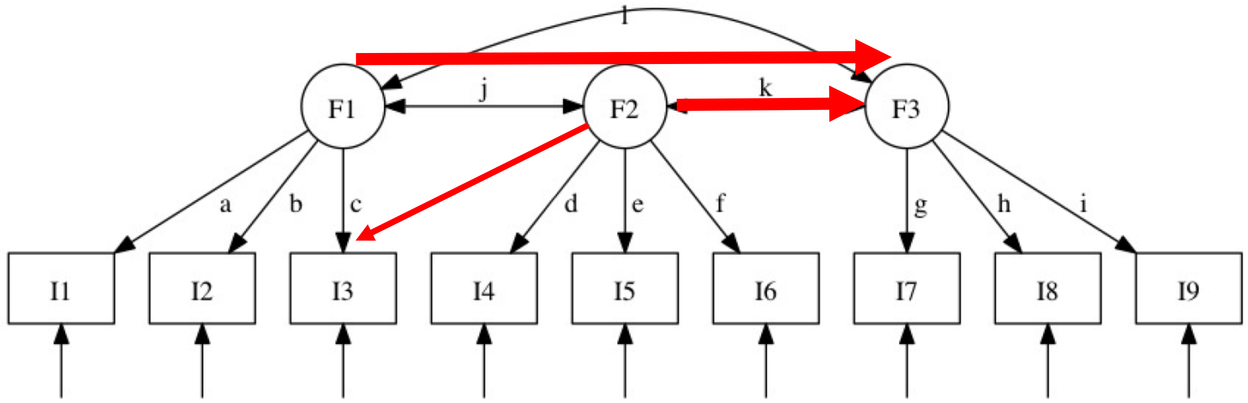
```
modelp2.1 <- '
# Measurement model
F1 =~ a*PS + b*RE + c*RT
F2 =~ d*HO + e*CO + f*ET + RT
F3 =~ g*BU + h*VA + i*RO
# error Variance and Covariance (psi)
F1 ~~ j*F2
F1 ~~ k*F3
F2 ~~ l*F3
#covariance
HO ~~ RO
'
```

```
Covariances:
                    Estimate
  F1 ~~
    F2         (j)     0.582
    F3         (k)    -0.405
  F2 ~~
    F3         (l)    -0.234
 .HO ~~
   .RO                -0.036
```

Factor analysis

Structural Equation Model

$CFA + Regression$




```
modelp3 <- '
# Measurement model
F1 =~ a*PS + b*RE + c*RT
F2 =~ d*HO + e*CO + f*ET + RT
F3 =~ g*BU + h*VA + i*RO
# error Variance and Covariance (psi)
F1 ~~ j*F2
#Structural model
F3 ~ k*F1+l*F2
# error Variance
F3 ~~ m*F3
'
```

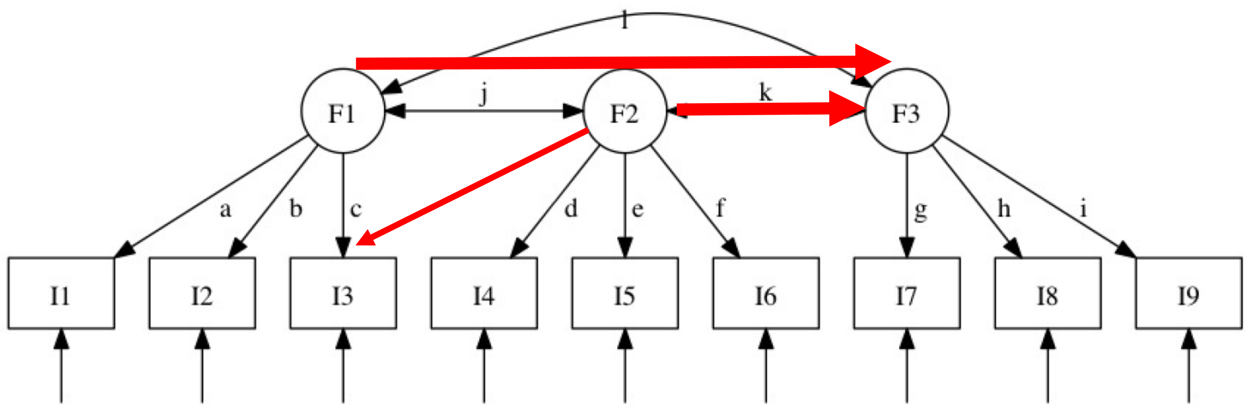
Reflecting attitudes toward the quality of police services

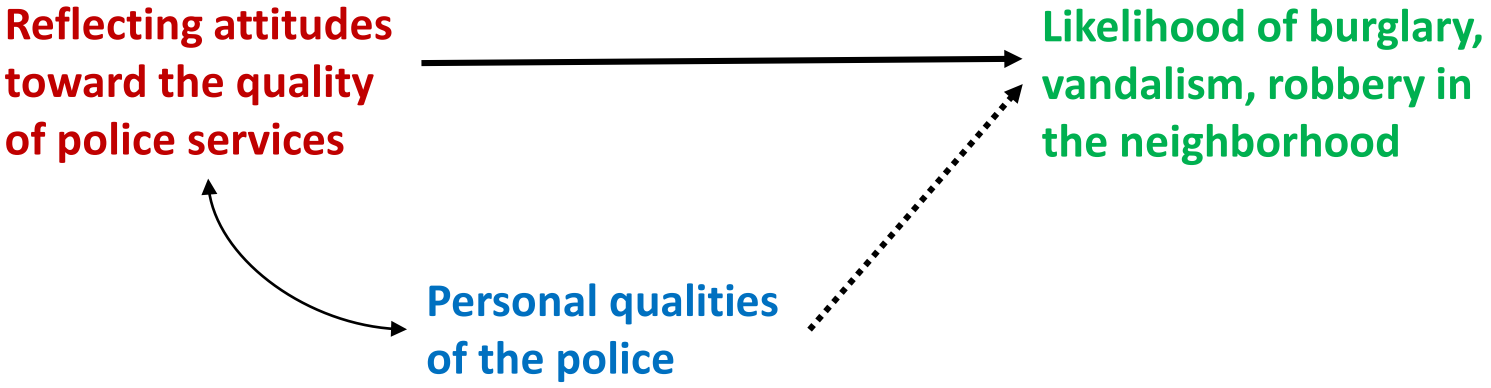
Likelihood of burglary, vandalism, robbery in the neighborhood

Personal qualities of the police

Factor analysis

Structural Equation Model





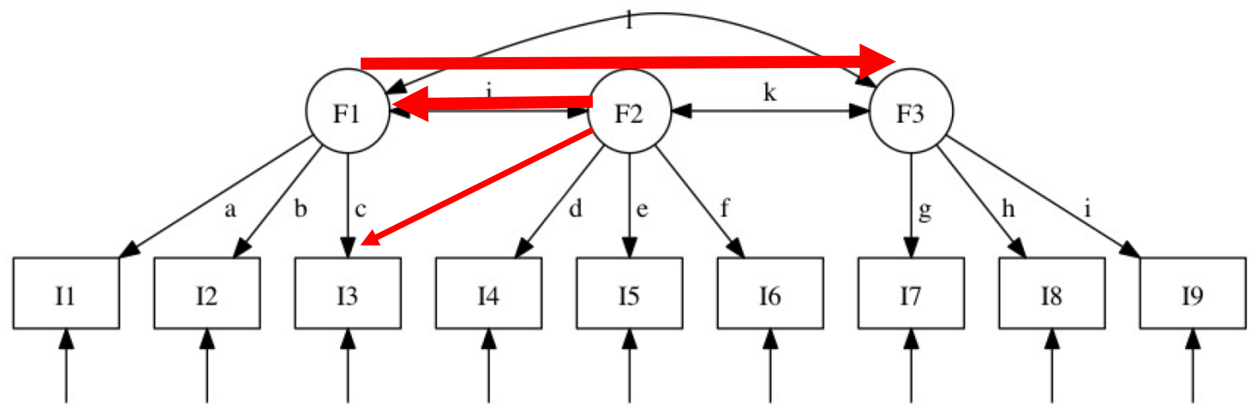
$CFA + Regression$

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  F1 =~
    PS        (a)     1.000
    RE        (b)     0.864    0.018   49.141    0.000
    RT        (c)     0.593    0.021   28.017    0.000
  F2 =~
    HO        (d)     1.000
    CO        (e)     0.908    0.017   55.034    0.000
    ET        (f)     0.869    0.016   53.810    0.000
    RT                0.197    0.019   10.325    0.000
  F3 =~
    BU        (g)     1.000
    VA        (h)     0.910    0.017   54.714    0.000
    RO        (i)     0.741    0.015   50.410    0.000

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  F3 ~
    F1        (k)    -0.420    0.020  -21.243    0.000
    F2        (l)    -0.006    0.018   -0.304    0.761

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  F1 ~~
    F2        (j)     0.332    0.009   36.520    0.000
```

Factor analysis

Structural Equation Model

$CFA + Mediantion$




```
modelp3.1 <- '
# Measurement model
F1 =~ a*PS + b*RE + c*RT
F2 =~ d*HO + e*CO + f*ET + RT
F3 =~ g*BU + h*VA + i*RO
#Structural model
F1 ~ j*F2
F3 ~ k*F1+l*F2
#define a new parameter
ind := j*k
# error Variance
F3 ~~ m*F3
F1 ~~ n*F1
'
```

Reflecting attitudes toward the quality of police services

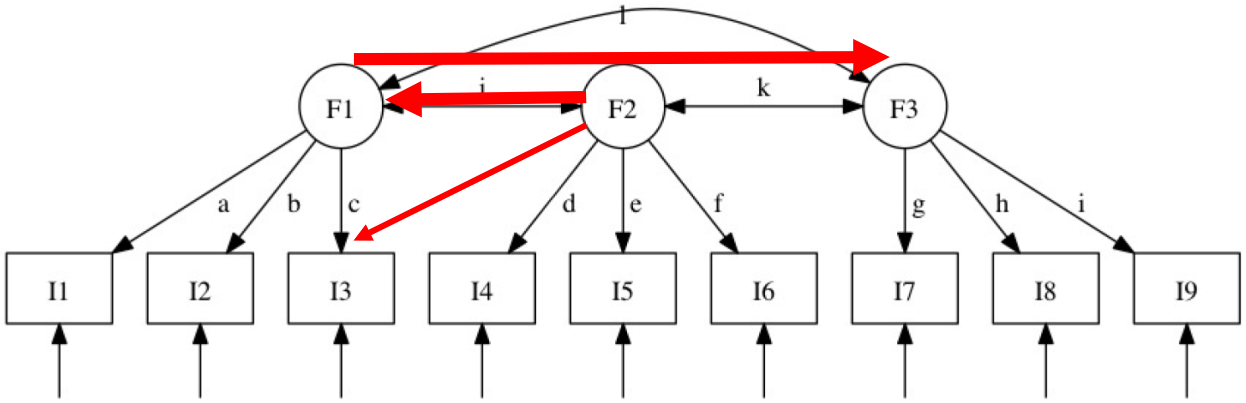
Likelihood of burglary, vandalism, robbery in the neighborhood

Personal qualities of the police

Factor analysis

Structural Equation Model

$CFA + Mediantion$




```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  F1 ~
    F2        (j)     0.591    0.015   40.155    0.000
  F3 ~
    F1        (k)    -0.420    0.020  -21.243    0.000
    F2        (l)    -0.006    0.018   -0.304    0.761

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ind              -0.248    0.013  -19.197    0.000
```

Reflecting attitudes toward the quality of police services

Likelihood of burglary, vandalism, robbery in the neighborhood

Personal qualities of the police

Factor analysis

Structural Equation Model

```
modelp3.1 <- '
# Measurement model
F1 =~ a*PS + b*RE + c*RT
F2 =~ d*HO + e*CO + f*ET + RT
F3 =~ g*BU + h*VA + i*RO
#Structural model
F1 ~ j*F2
F3 ~ k*F1+l*F2
#define a new parameter
ind := j*k
# error Variance
F3 ~~ m*F3
F1 ~~ n*F1
'
```

```
modelp3.2 <- '
# Measurement model
F1 =~ a*PS + b*RE + c*RT
F2 =~ d*HO + e*CO + f*ET +
F3 =~ g*BU + h*VA + i*RO
#Structural model
F1 ~ j*F2
F3 ~ k*F1
# error Variance
F3 ~~ m*F3
F1 ~~ n*F1
'
```

```
> anova(p3.1.fit,p3.2.fit)
Chi-Squared Difference Test

          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
p3.1.fit 23 256978 257139 127.32
p3.2.fit 24 256976 257129 127.41   0.091983       1     0.7617
```

```
Latent Variables:
                         Estimate  Std.Err  z
  F1 =~
    PS          (a)        1.000
    RE          (b)        0.865    0.018
    RT          (c)        0.594    0.021
  F2 =~
    HO          (d)        1.000
    CO          (e)        0.908    0.017
    ET          (f)        0.869    0.016
    RT                     0.197    0.019
  F3 =~
    BU          (g)        1.000
    VA          (h)        0.910    0.017
    RO          (i)        0.741    0.015

Regressions:
                         Estimate  Std.Err  z
  F1 ~
    F2          (j)        0.591    0.015
  F3 ~
    F1          (k)       -0.425    0.014   -
```

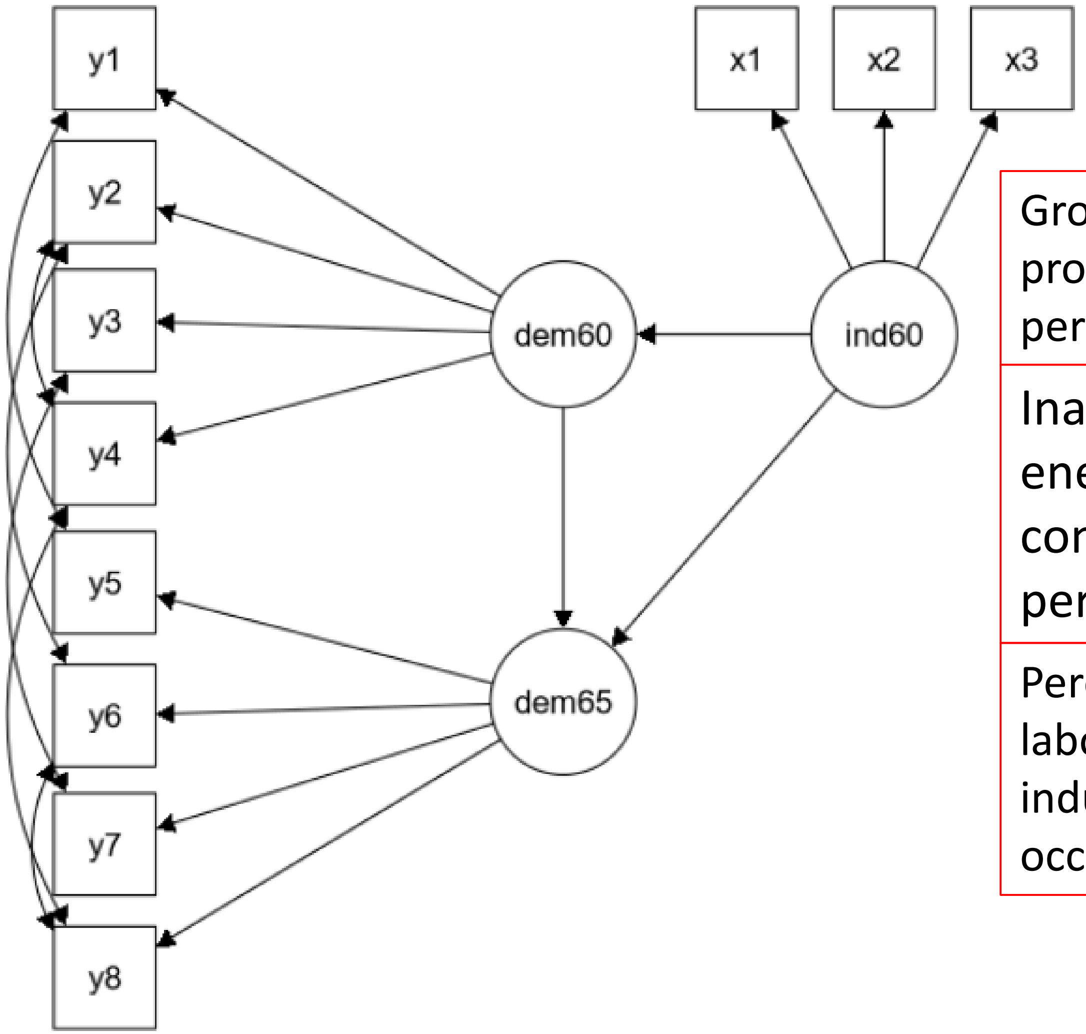
Factor analysis

Structural Equation Model

Political democracy to industrialization in developing countries

Three latent variables
Political democracy in 1965
Political democracy in 1960
Industrialization in 1960

Assume that political democracy in 1965 is **a function of** 1960 political democracy and industrialization.
The 1960 industrialization level **also affects** the 1960 political democracy level.



Indicators of industrialization in 1960

Gross national product (GNP) per capita (X_1)
Inanimate energy consumption per capita (X_2)
Percent of the labor force in industrial occupations (X_3)

Factor analysis

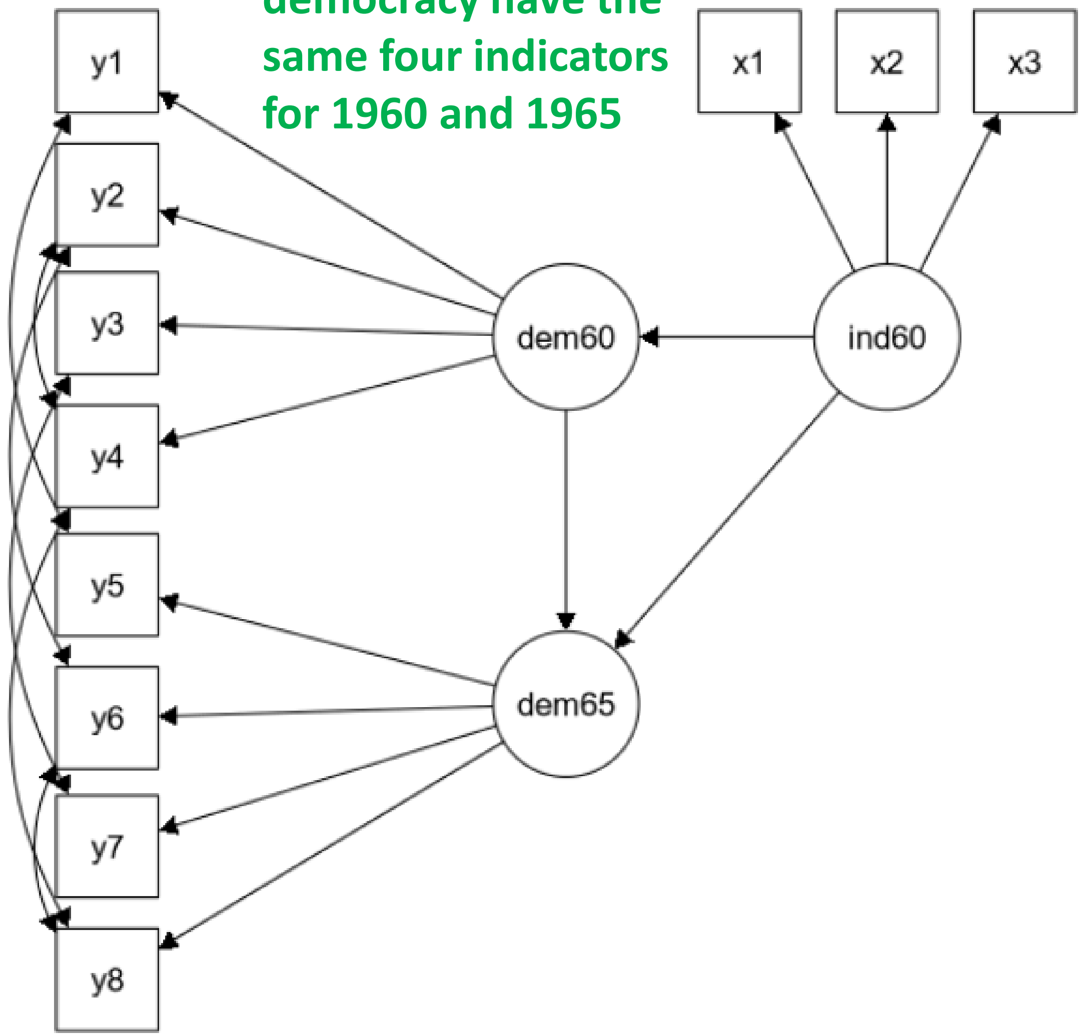
Structural Equation Model

Political democracy to industrialization in developing countries

Three latent variables
Political democracy in 1965
Political democracy in 1960
Industrialization in 1960

Expert ratings of the freedom of the press (Y_1 in 1960, Y_5 in 1965)	Fairness of elections (Y_3 and Y_7)
Freedom of political opposition (Y_2 and Y_6)	Effectiveness of legislative body (Y_4 and Y_8)

For political democracy have the same four indicators for 1960 and 1965



Factor analysis

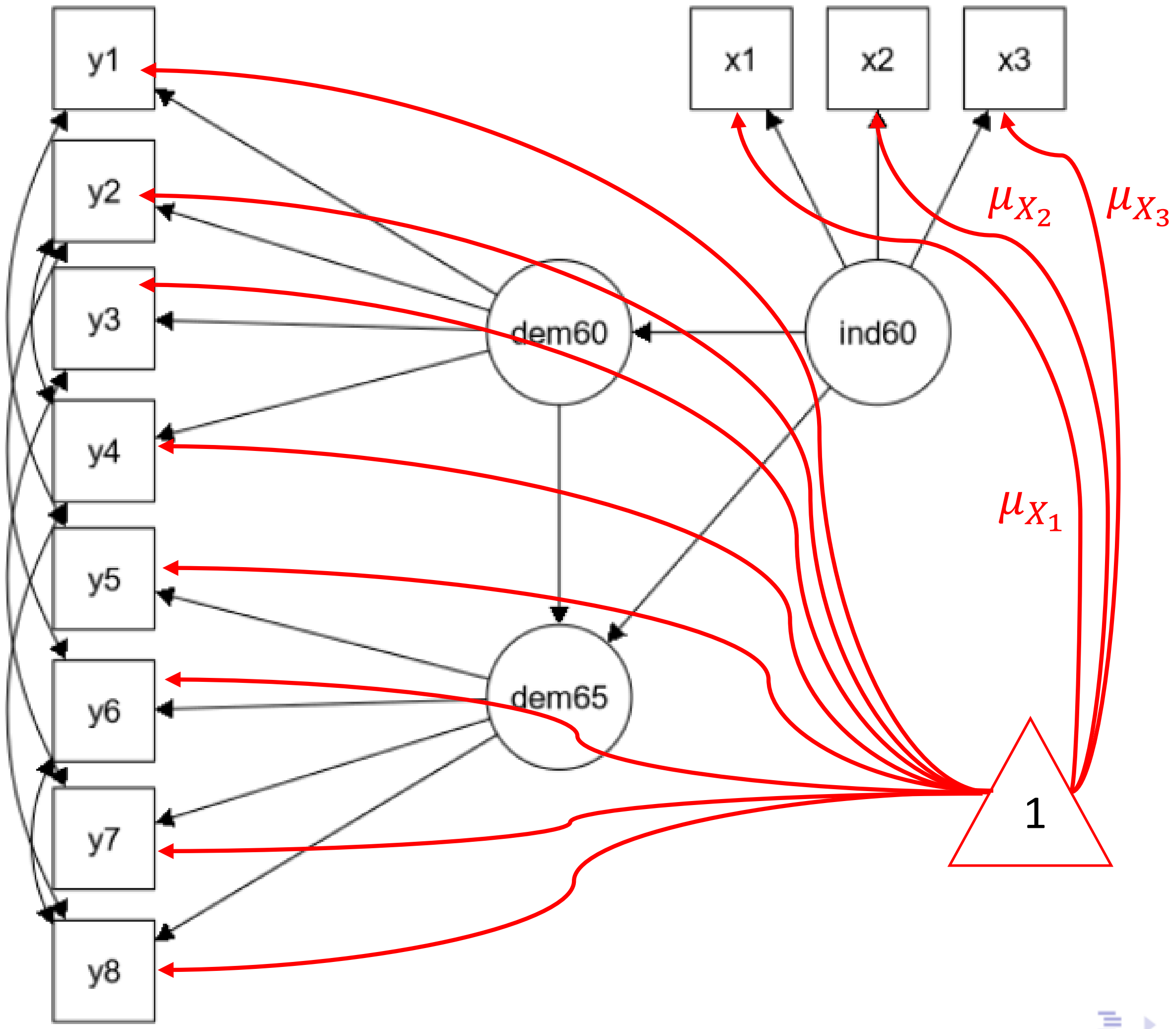
Structural Equation Model

To extend traditional SEM models by decomposing the means of the measured variables into more basic model parameters.

[mean and covariance structure]

$$\mathbf{y} = \boxed{\mu} + \Lambda\omega + \epsilon$$

$$\eta = \Gamma\xi + \delta$$



Means and intercepts are measures of a variable's location.

Factor analysis

Structural Equation Model

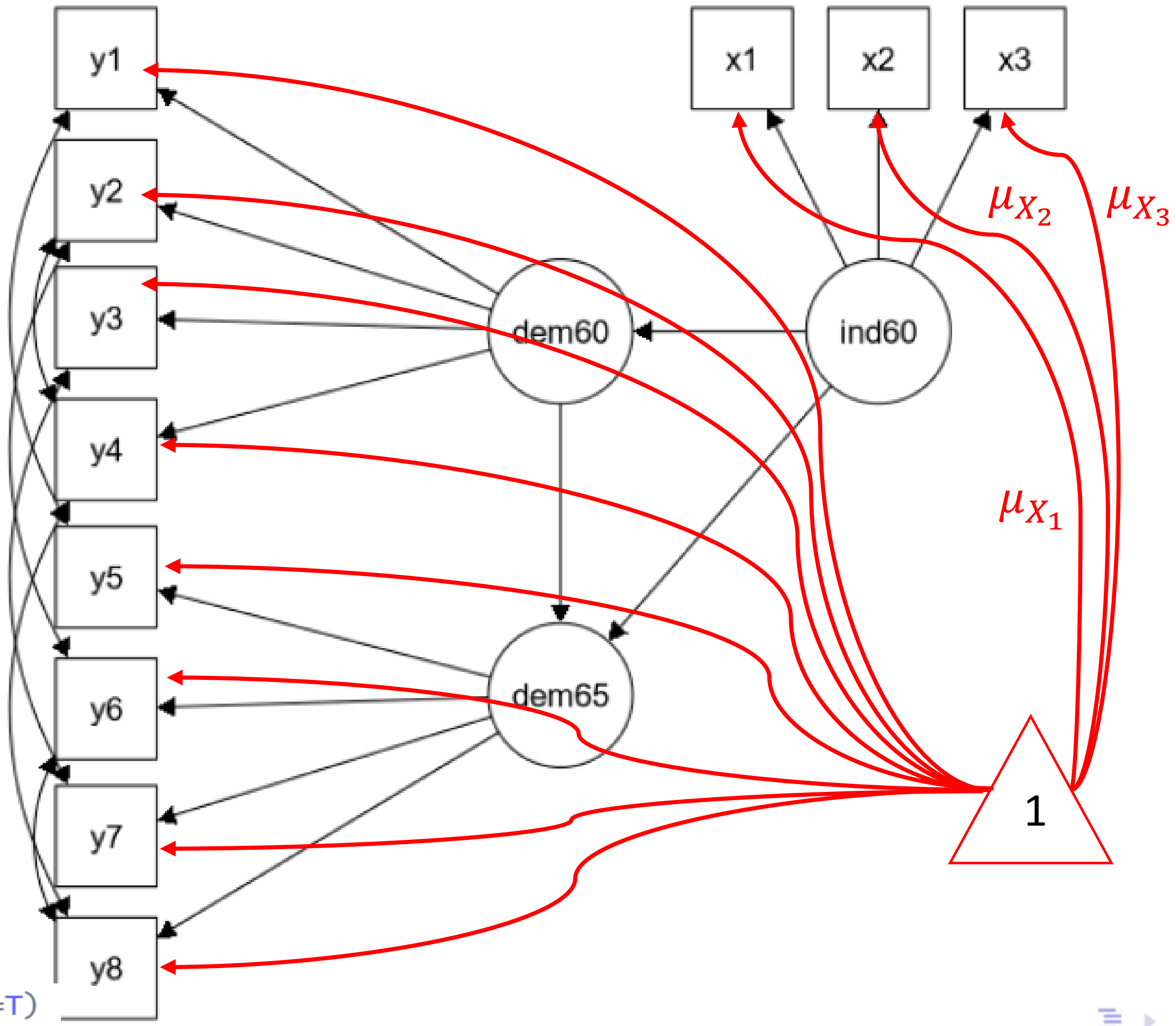
```
modelint <- '
  # Measurement Models
    ind60 =~ x1 + x2 + x3
    dem60 =~ y1 + y2 + y3 + y4
    dem65 =~ y5 + y6 + y7 + y8
  # Structural Models
    dem60 ~ ind60
    dem65 ~ ind60 + dem60
  # error Variance and Covariance
    dem60 ~~ dem60
    dem65 ~~ dem65
    y1 ~~ y5
    y2 ~~ y4 + y6
    y3 ~~ y7
    y4 ~~ y8
    y6 ~~ y8
    #intercept
    x1 ~ 1
    x2 ~ 1
    x3 ~ 1
    y1 ~ 1
    y2 ~ 1
    y3 ~ 1
    y4 ~ 1
    y5 ~ 1
    y6 ~ 1
    y7 ~ 1
    y8 ~ 1
'
fitint1 <- sem(model, data = PoliticalDemocracy,meanstructure=T)
```

In practice, there are two reasons why a user would add intercept formulas in the model syntax

1. Some constraints must be specified on them.

   `x1 + x2 + x3 + x4 ~ 0.5*1`

2. Exam group differences in multiple group data.

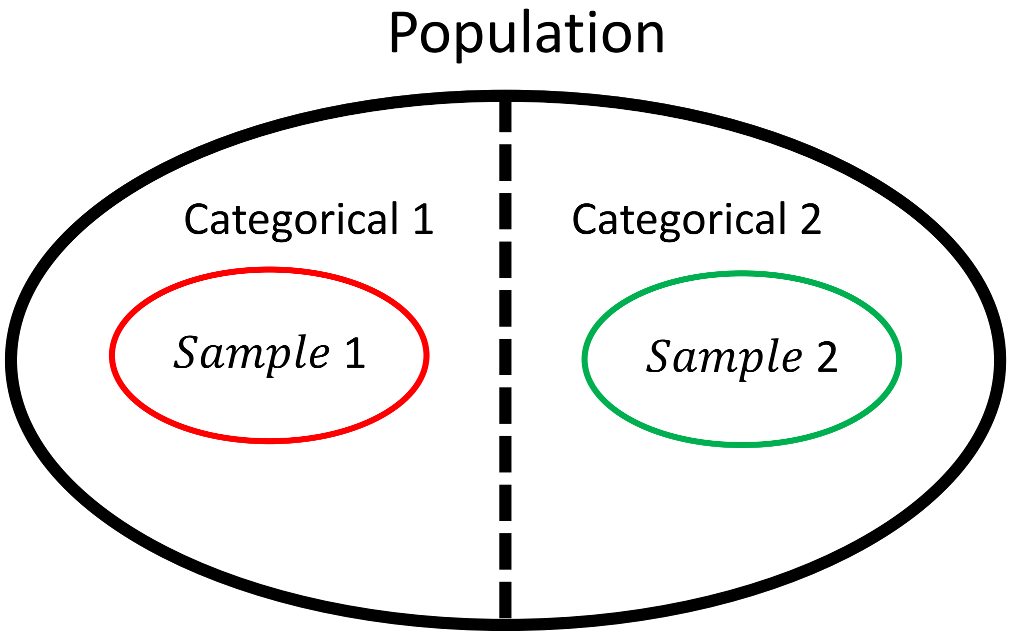


Structural Equation Model

Multiple Sample extension

When we have data form different groups (e.g., gender, countries, organizations), an interesting question to ask is whether a structural model would work the same for these groups.

Universality vs. Specificity



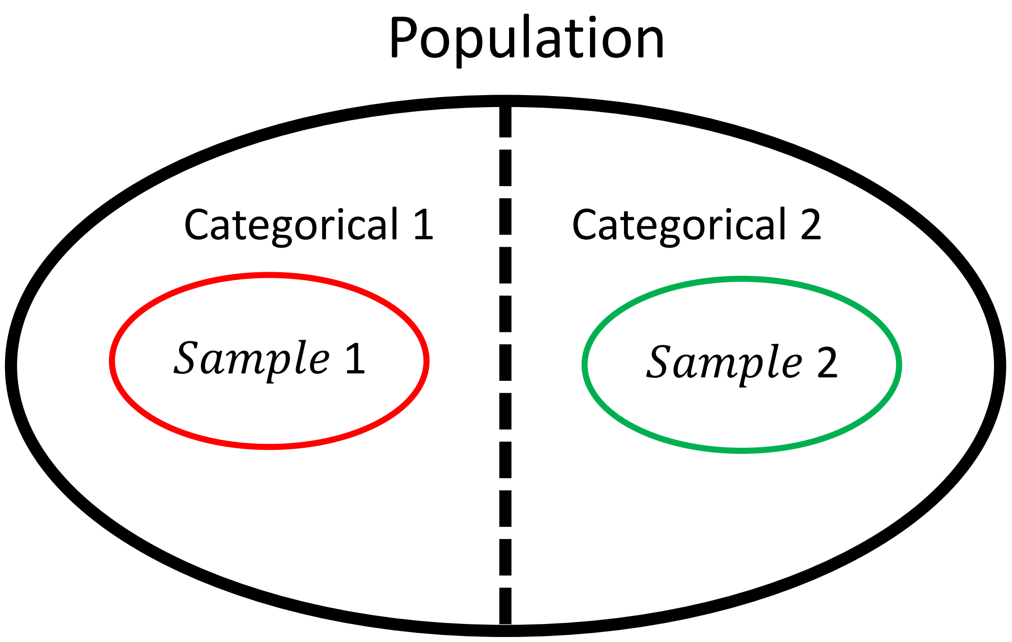
Invariance across the groups

This kind of analysis is useful for investigating the behaviors of different groups of employees, different cultures, different treatment groups, etc..

Structural Equation Model

Multiple Sample extension

In many substantive researches, one may encounter the following types of data:



1. Multisample data: come from a number of distinct groups (populations), where the number of groups is known, and the group membership of each observation can be specified exactly.

2. Hierarchical data: come from a number of different groups (clusters) with a known hierarchical structure, the number of groups is large.

3. Mixture data: come from one of the K populations with different distributions, and **no information** is available on which of the K populations an individual observation belongs to

Structural Equation Model

Multiple Sample extension

In many substantive researches, one may encounter the following types of data:

1. Multisample data: come from a number of distinct groups (populations), where the number of groups is known, and the group membership of each observation can be specified exactly.

2. Hierarchical data: come from a number of different groups (clusters) with a known hierarchical structure, the number of groups is large.

(1) the number of observations within each group is usually large; but is relatively small for hierarchical data.
(2) observations within each group are assumed independent. But dependent for hierarchical data because individuals within a group share certain common influential factors.

Examples:
investigate the behaviors of different groups of employees,
compare different cultures,
examine the drug effect in different treatment groups.

Examples:
patients from within random samples of clinics or hospitals;
individuals from within random samples of families;
students from within random samples of schools

Structural Equation Model

Multiple Sample extension

An important issue is to test about the invariances among the models in different groups.

Group 1

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad \epsilon_i$ are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

Group 2

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad \epsilon_i$ are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

	Type of Invariance	Constraints	Between-Groups Comparisons Allowed
1	Configural	Same model. No parameter constraints.	None
2	Weak	1 + all loadings constrained to be equal between groups (but can vary within a group). Latent (co)variances allowed to vary between groups.	Latent (co)variances [weak evidence]
3	Strong	2 + all intercepts are constrained to be equal between groups (but can vary within a group).	Latent means, latent (co)variances [strong evidence]
4	Strict	3 + error variances are constrained to be the same between groups (but can vary within a group).	

Structural Equation Model

Multiple Sample extension

Want to know if the groups share a common structure

Group 1

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

Group 2

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

	Type of Invariance	Constraints	Between-Groups Comparisons Allowed
1	Configural	Same model. No parameter constraints.	None
2	Weak	1 + all loadings constrained to be equal between groups (but can vary within a group). Latent (co)variances allowed to vary between groups.	Latent (co)variances [weak evidence]
3	Strong	2 + all intercepts are constrained to be equal between groups (but can vary within a group).	Latent means, latent (co)variances [strong evidence]
4	Strict	3 + error variances are constrained to be the same between groups (but can vary within a group).	

Structural Equation Model

Multiple Sample extension

Want to know if the factor loadings are equal across the groups

Group 1

$$\mathbf{y} = \boldsymbol{\mu} + \boxed{\boldsymbol{\Lambda}}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

Group 2

$$\mathbf{y} = \boldsymbol{\mu} + \boxed{\boldsymbol{\Lambda}}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

	Type of Invariance	Constraints	Between-Groups Comparisons Allowed
1	Configural	Same model. No parameter constraints.	None
2	Weak	1 + all loadings constrained to be equal between groups (but can vary within a group). Latent (co)variances allowed to vary between groups.	Latent (co)variances [weak evidence]
3	Strong	2 + all intercepts are constrained to be equal between groups (but can vary within a group).	Latent means, latent (co)variances [strong evidence]
4	Strict	3 + error variances are constrained to be the same between groups (but can vary within a group).	

Structural Equation Model

Multiple Sample extension

Want to know if the Intercepts are equal

Group 1

$$\mathbf{y} = \boxed{\boldsymbol{\mu}} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

Group 2

$$\mathbf{y} = \boxed{\boldsymbol{\mu}} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

	Type of Invariance	Constraints	Between-Groups Comparisons Allowed
1	Configural	Same model. No parameter constraints.	None
2	Weak	1 + all loadings constrained to be equal between groups (but can vary within a group). Latent (co)variances allowed to vary between groups.	Latent (co)variances [weak evidence]
3	Strong	2 + all intercepts are constrained to be equal between groups (but can vary within a group)	Latent means, latent (co)variances [strong evidence]
4	Strict	3 + error variances are constrained to be the same between groups (but can vary within a group).	

Structural Equation Model

Multiple Sample extension

Want to know if the factor Variance/Covariance or structural coefficients are equal

Group 1

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boxed{\boldsymbol{\Gamma}}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boxed{\boldsymbol{\Phi}}]$

ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$

δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

Group 2

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\eta} = \boxed{\boldsymbol{\Gamma}}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ϵ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$

ξ_i are i.i.d. $N[\mathbf{0}, \boxed{\boldsymbol{\Phi}}]$

δ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

	Type of Invariance	Constraints	Between-Groups Comparisons Allowed
1	Configural	Same model. No parameter constraints.	None
2	Weak	1 + all loadings constrained to be equal between groups (but can vary within a group). Latent (co)variances allowed to vary between groups.	Latent (co)variances [weak evidence]
3	Strong	2 + all intercepts are constrained to be equal between groups (but can vary within a group).	Latent means, latent (co)variances [strong evidence]
4	Strict	3 + error variances are constrained to be the same between groups (but can vary within a group).	

Structural Equation Model

Multiple Sample extension

Want to know if the error Variance/Covariance are equal

All of the parameters in the common model are equal across groups

Group 1

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boxed{\boldsymbol{\Psi}_\epsilon}]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boxed{\boldsymbol{\Psi}_\delta}]$

Group 2

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega} + \boldsymbol{\epsilon}$$

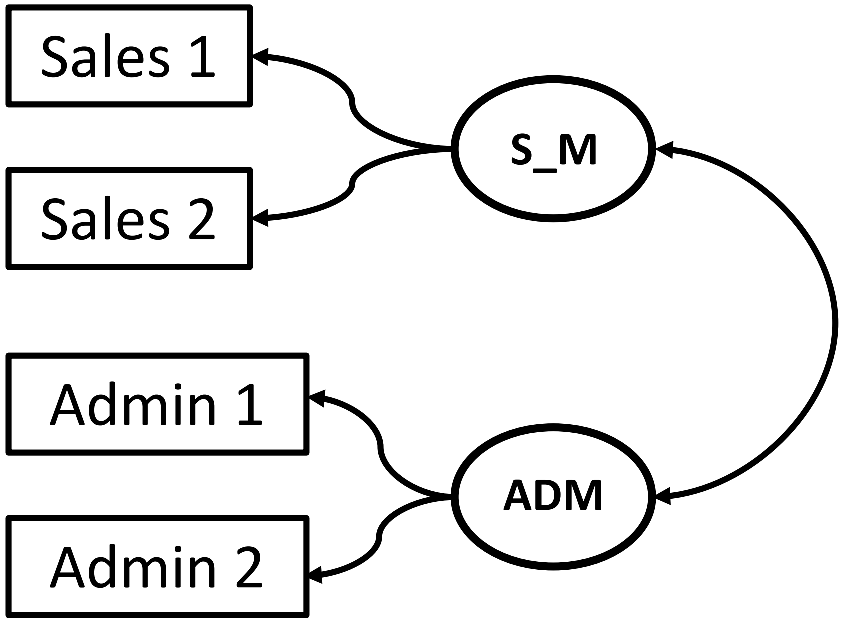
$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

ξ_i are i.i.d. $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\quad$ ϵ_i are i.i.d. $N[\mathbf{0}, \boxed{\boldsymbol{\Psi}_\epsilon}]$ $\quad$ δ_i are i.i.d. $N[\mathbf{0}, \boxed{\boldsymbol{\Psi}_\delta}]$

	Type of Invariance	Constraints	Between-Groups Comparisons Allowed
1	Configural	Same model. No parameter constraints.	None
2	Weak	1 + all loadings constrained to be equal between groups (but can vary within a group). Latent (co)variances allowed to vary between groups.	Latent (co)variances [weak evidence]
3	Strong	2 + all intercepts are constrained to be equal between groups (but can vary within a group).	Latent means, latent (co)variances [strong evidence]
4	Strict	3 + error variances are constrained to be the same between groups (but can vary within a group).	

Structural Equation Model

Multiple Sample extension



Sales/Marketing and Administration (SMA) Ability

- Group 1: Males Officers ($N = 265$, filename = *sma_male.cov*)

sales1	63			
sales2	70	110		
admin1	41	52	60	
admin2	30	37	36	32

- Group 2: Females Officers ($N = 300$, filename = *sma_female.cov*)

sales1	67			
sales2	72	107		
admin1	40	55	63	
admin2	28	38	39	35

Structural Equation Model

Multiple Sample extension

Want to know if the groups share a common structure

```
model1 <- "
# measurement model
S_M =~ NA*sales1 + sales2
ADM =~ NA*admin1 + admin2
# factor Variances and covariance
S_M ~~ 1*S_M
ADM ~~ 1*ADM
S_M ~~ ADM
"
```

```
model1 <- "
# measurement model
S_M =~ sales1 + sales2
ADM =~ admin1 + admin2
# factor Variances and covariance
S_M ~~ S_M
ADM ~~ ADM
S_M ~~ ADM
"
```

```
fit1 <-sem(model1, sample.cov=list(Group1=(sma_male.cov), Group2=(sma_female.cov)), sample.nobs=c(265, 300))
```

Structural Equation Model

Multiple Sample extension

Want to know if the factor loadings are equal across the groups

```
model2 <- "
# measurement model
S_M =~ c(eq1,eq1)*sales1 + c(eq2,eq2)*sales2
ADM =~ c(eq3,eq3)*admin1 + c(eq4,eq4)*admin2
# factor Variances and covariance
S_M ~~ c(1,NA)*S_M
ADM ~~ c(1,NA)*ADM
S_M ~~ ADM
"
```

```
> fit2 <-sem(model2, sample.cov=list(Group1=(sma_male.
cov), Group2=(sma_female.cov)), sample.nobs=c(265, 30
0),std.lv=T)
```

```
model2 <- "
# measurement model
S_M =~ c(eq1,eq1)*sales1 + c(eq2,eq2)*sales2
ADM =~ c(eq3,eq3)*admin1 + c(eq4,eq4)*admin2
# factor Variances and covariance
S_M ~~ S_M
ADM ~~ ADM
S_M ~~ ADM
"
```

Structural Equation Model

Multiple Sample extension

Want to know if the factor Variance/Covariance or structural coefficients are equal

```
model3 <- "
# measurement model
S_M =~ c(eq1,eq1)*sales1 + c(eq2,eq2)*sales2
ADM =~ c(eq3,eq3)*admin1 + c(eq4,eq4)*admin2
# factor Variances and covariance
S_M ~~ c(1,1)*S_M
ADM ~~ c(1,1)*ADM
S_M ~~ ADM
"
```

```
model4 <- "
# measurement model
S_M =~ c(eq1,eq1)*sales1 + c(eq2,eq2)*sales2
ADM =~ c(eq3,eq3)*admin1 + c(eq4,eq4)*admin2
# factor Variances and covariance
S_M ~~ c(1,1)*S_M
ADM ~~ c(1,1)*ADM
S_M ~~ c(eq5,eq5)*ADM
"
```

Structural Equation Model

Multiple Sample extension

Want to know if the error Variance/Covariance are equal

```
model5 <- "
# measurement model
S_M =~ c(eq1,eq1)*sales1 + c(eq2,eq2)*sales2
ADM =~ c(eq3,eq3)*admin1 + c(eq4,eq4)*admin2
# factor Variances and covariance
S_M ~~ c(1,1)*S_M
ADM ~~ c(1,1)*ADM
S_M ~~ c(eq5,eq5)*ADM
# error Variances
sales1 ~~ c(eq6,eq6)*sales1
sales2 ~~ c(eq7,eq7)*sales2
admin1 ~~ c(eq8,eq8)*admin1
admin2 ~~ c(eq9,eq9)*admin2
"
```

```
total=(sma_male.cov*265+sma_female.cov*300)/(265+300)
model1 <- "
# measurement model
S_M =~ NA*sales1 + sales2
ADM =~ NA*admin1 + admin2
# factor Variances and covariance
S_M ~~ 1*S_M
ADM ~~ 1*ADM
S_M ~~ ADM
"
fit1t <-sem(model1, sample.cov=total, sample.nobs=sum(265, 300))
summary(fit1t,standardize=T)
```

Structural Equation Model

Multiple Sample extension

```
> lavTestLRT(fit1, fit2, fit3, fit4, fit5)
Chi-Squared Difference Test

     Df   AIC   BIC   Chisq Chisq diff Df diff Pr(>Chisq)
fit1  2 14023 14101  0.9504
fit2  4 14021 14090  2.9205     1.9701       2     0.3734
fit3  6 14017 14078  3.3459     0.4253       2     0.8084
fit4  7 14017 14073  4.9327     1.5869       1     0.2078
fit5 11 14017 14056 12.6678     7.7351       4     0.1018
```

Do not reject (non-significant)

Imply that all of the parameters in the common model are equal across groups

Structural Equation Model

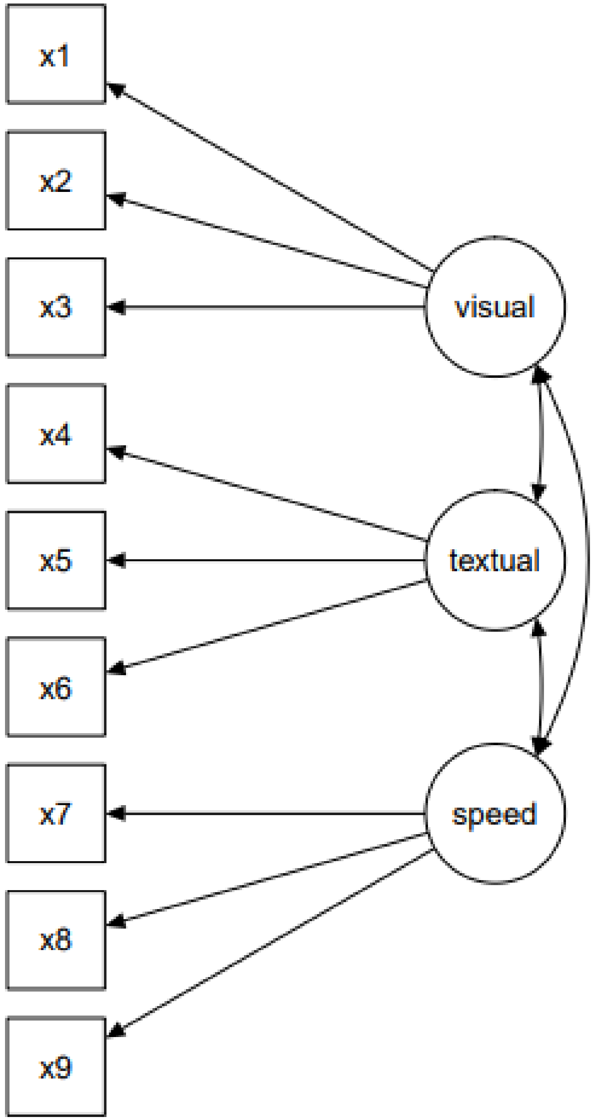
Multiple Sample extension

The lavaan package contains a built-in dataset called **HolzingerSwineford1939.** The classic Holzinger and Swineford (1939) dataset consists of mental ability test scores of seventh- and eighth-grade children from **two different schools (Pasteur and Grant-White)**.

X_1: Visual perception
X_2: Cubes
X_3: Lozenges

X_4: Paragraph comprehension
X_5: Sentence completion
X_6: Word meaning

X_7: Speeded addition
X_8: Speeded counting of dots
X_9: Speeded discrimination straight and curved capitals



Structural Equation Model

Multiple Sample extension

Want to know if the groups share a common structure

```
HSmodel <- "
# measurement model
visual  =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed   =~ x7 + x8 + x9
# factor Variances and covariance
visual ~~ visual
textual ~~ textual
speed ~~ speed
visual ~~ textual
visual ~~ speed
textual ~~ speed
"
# configural invariance
fit1 <- cfa(HSmodel, data = HolzingerSwineford1939, group = "school")
summary(fit1)
```

Structural Equation Model

Multiple Sample extension

Want to know if the factor loadings are equal across the groups

```
# weak invariance
fit2 <- sem(HSmodel, data = HolzingerSwineford1939, group = "school",
            group.equal = "loadings")
```

- intercepts: the intercepts of the observed variables
- means: the intercepts/means of the latent variables
- residuals: the residual variances of the observed variables
- residual.covariances: the residual covariances of the observed variables
- lv.variances: the variances of the latent variables
- lv.covariances: the covariances of the latent variables
- regressions: all regression coefficients in the model

Structural Equation Model

Multiple Sample extension

Want to know if the Intercepts are equal

```
# strong invariance
fit3 <- sem(HSmodel, data = HolzingerSwineford1939, group = "school",
            group.equal = c("loadings","intercepts"))
```

```
Intercepts:
                   Estimate
   .x1       (.25.)   5.001
   .x2       (.26.)   6.151
   .x3       (.27.)   2.271
   .x4       (.28.)   2.778
   .x5       (.29.)   4.035
   .x6       (.30.)   1.926
   .x7       (.31.)   4.242
   .x8       (.32.)   5.630
   .x9       (.33.)   5.465
    visual            0.000
    textual           0.000
    speed             0.000
```

```
Intercepts:
                   Estimate
   .x1       (.25.)   5.001
   .x2       (.26.)   6.151
   .x3       (.27.)   2.271
   .x4       (.28.)   2.778
   .x5       (.29.)   4.035
   .x6       (.30.)   1.926
   .x7       (.31.)   4.242
   .x8       (.32.)   5.630
   .x9       (.33.)   5.465
    visual           -0.148
    textual           0.576
    speed            -0.177
```

- intercepts: the intercepts of the observed variables
- means: the intercepts/means of the latent variables
- residuals: the residual variances of the observed variables
- residual.covariances: the residual covariances of the observed variables
- lv.variances: the variances of the latent variables
- lv.covariances: the covariances of the latent variables
- regressions: all regression coefficients in the model

Structural Equation Model

Multiple Sample extension

```
> lavTestLRT(fit1, fit2, fit3)
Chi-Squared Difference Test

     Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit1 48 7484.4 7706.8 115.85
fit2 54 7480.6 7680.8 124.04      8.192       6     0.2244
fit3 60 7508.6 7686.6 164.10     40.059       6  4.435e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Because the first p-value is non-significant, we may conclude that weak invariance (equal factor loadings) is supported in this dataset. However, because the second p-value is significant, strong invariance is not. Therefore, it is unwise to directly compare the values of the latent means across the two groups.

Structural Equation Model

Multiple Sample extension

In many substantive researches, one may encounter the following types of data:

$Two\ level$ SEM

With-in Group

Measurement Model

$$\mathbf{u}_{gi} = \mathbf{v}_g + \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi},$$

$$g = 1, \cdots, G, \quad i = 1, \cdots, N_g,$$

Between Group

$$\mathbf{v}_g = \boldsymbol{\mu} + \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \quad g = 1, \cdots, G, \Longrightarrow \mathbf{u}_{gi} = \boldsymbol{\mu} + \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g} + \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}$$

1. Multisample data: come from a number of distinct groups (populations), where the number of groups is known, and the group membership of each observation can be specified exactly.

2. Hierarchical data: come from a number of different groups (clusters) with a known hierarchical structure, the number of groups is large.

Structural Model

$$\boldsymbol{\eta}_{1gi} = \boldsymbol{\Pi}_{1g}\boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_{1g}\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi},$$

$$\boldsymbol{\eta}_{2g} = \boldsymbol{\Pi}_2\boldsymbol{\eta}_{2g} + \boldsymbol{\Gamma}_2\mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g},$$

Structural Equation Model

Ordered/Categorical data extension

$$\mathbf{y} = \boxed{\mathbf{Ac}} + \Lambda\omega + \epsilon$$

$$\eta = \boxed{\mathbf{Bd}} + \xi + \delta$$

Exogenous categorical variables (covariates)

Dummy (0/1) variable

$K > 2$ levels, you need to replace it by a set of $K - 1$ dummy variables. (Similar in classic regression model)

Structural Equation Model

Ordered/Categorical data extension

$$\boxed{\mathbf{y}} = \mathbf{A}\mathbf{c} + \Lambda\omega + \epsilon$$

$$\eta = \mathbf{B}\mathbf{d} + \xi + \delta$$

Underlying variable approach assumes the categorical outcomes are realizations of continuous underlying response variables that are incompletely observed

For each categorical outcome, x_i , there is an incompletely observed continuous variable x_i^* and $x_i^* \sim N(\mu_i, \sigma_i^2)$ nothing is lost if μ_i, σ_i^2 are constrained to certain values (0,1).

$$x_i = \begin{cases} 1 & \text{if } x_i^* \geq \boxed{\tau_i} \\ 0 & \text{if } x_i^* < \boxed{\tau_i} \end{cases}$$

threshold

$$x_i^* = v_i + \lambda_i \omega + \epsilon_i$$

Structural Equation Model

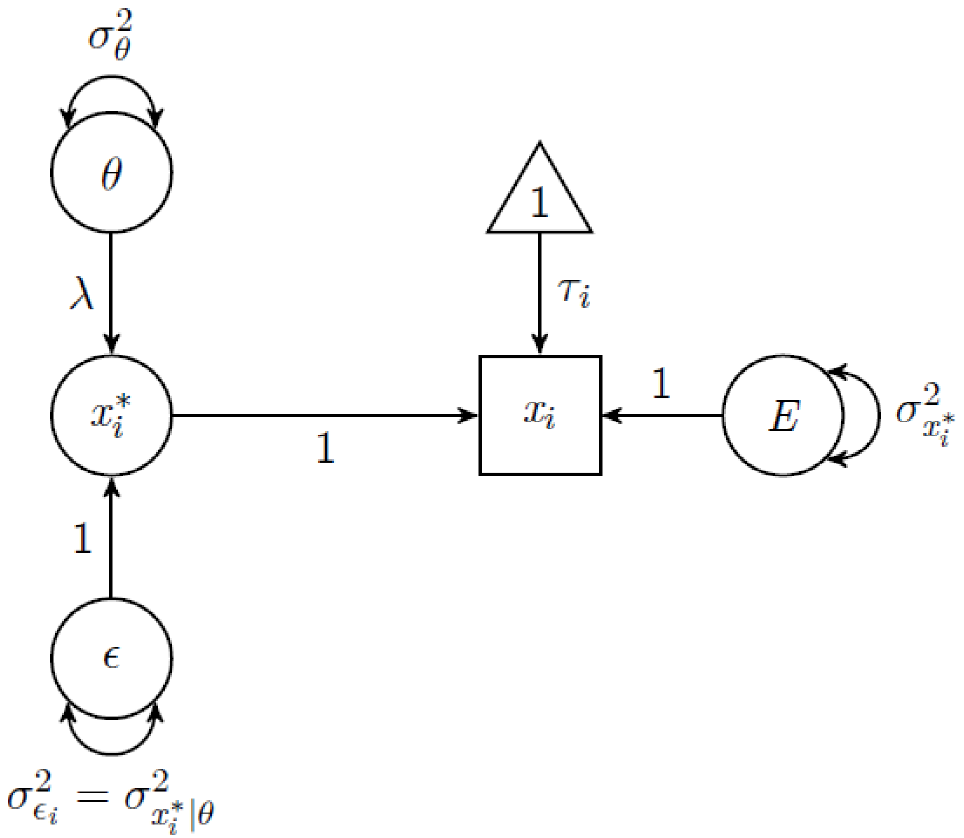
Ordered/Categorical data extension

```r
fit <- cfa(myModel, data = myData,
           ordered = c("item1","item2",
                       "item3","item4"))
```

"Lavaan" can only deal with 0 vs 1

$$x_i = \begin{cases} 1 & \text{if } x_i^* \geq \boxed{\tau_i} \\ 0 & \text{if } x_i^* < \boxed{\tau_i} \end{cases}$$

threshold



$$x_i^* = v_i + \lambda_i \omega + \epsilon_i$$

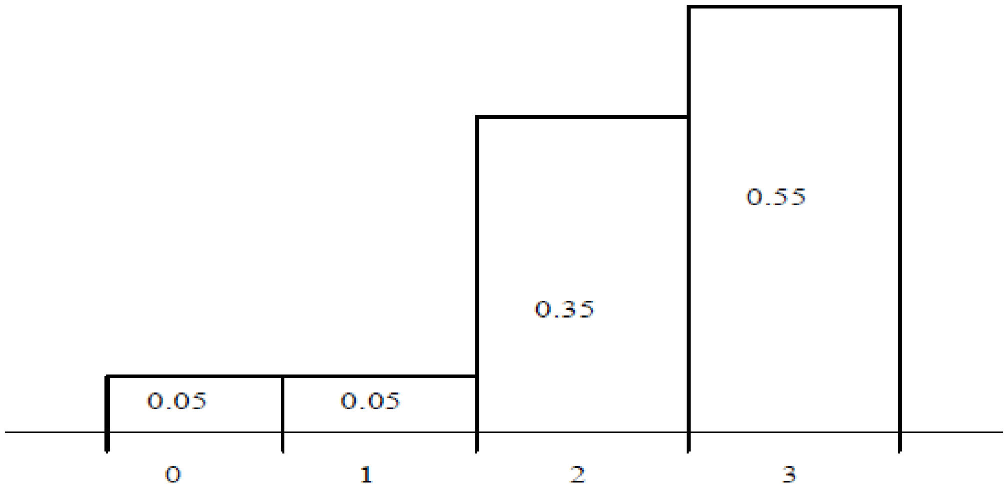
Structural Equation Model

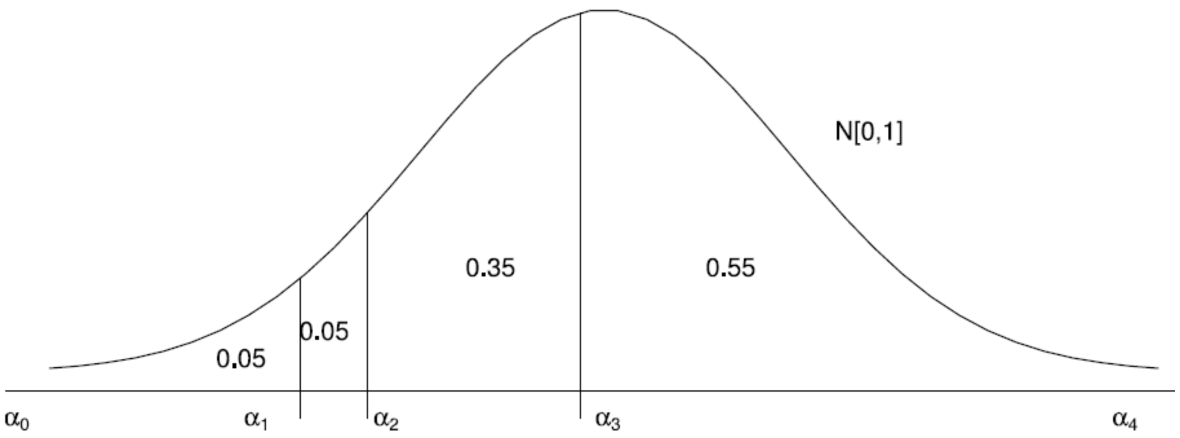
Ordered/Categorical data extension

Suppose that for a given data set, the proportions of 0, 1, 2, 3 are 0.05, 0.05, 0.35, and 0.55, respectively.

threshold	$a_1 < a_2 < a_3$

$$x_i = \begin{cases} 0, & x_i^* < a_1 \\ 1, & a_1 < x_i^* < a_2 \\ 2, & a_2 < x_i^* < a_3 \\ 3, & a_3 < x_i^* \end{cases}$$





Section 2: Estimating with Bayesian Method

Parameter Estimation

$$\Sigma = \Lambda\Psi\Lambda^T + \Theta$$

Want to find $\widehat{\Lambda}$, $\widehat{\Psi}$, and $\widehat{\Theta}$ (estimates of Λ, Ψ, Θ) such that $\widehat{\Sigma}$ and S are ***as close as*** possible

$$S \simeq \widehat{\Lambda}\widehat{\Psi}\widehat{\Lambda}' + \widehat{\Theta} = \widehat{\Sigma} = \Sigma(\widehat{\theta})$$

$\widehat{\Sigma} = \Sigma(\widehat{\theta})$ is called the ***implied, reproduced, or fitted*** covariance matrix

Assume that the data follow a multivariate Normal distribution [Normal-theory methods]

Maximum Likelihood Estimation

$$F_{\mathrm{ML}}(S,\Sigma(\theta)) = \mathrm{tr}(S\Sigma(\theta)^{-1}) + \ln|\Sigma(\theta)| - \ln|S| - p$$

Generalized least squares estimation

over admissible choices of θ

$$F_{\mathrm{GLS}}(S,\Sigma(\theta)) = \mathrm{tr}[(S-\Sigma(\theta))S^{-1}]^2$$

When sample size is large, the ML or GLS estimators are consistent, efficient and jointly normal

Both the ML, and GLS estimates are found by numerically searching methods with **repeated iterations**. To begin the search, starting values are needed

The basic objective of this section is to introduce a Bayesian approach for analyzing not only the standard SEMs but also their useful generalizations which have been developed in recent years. In contrast to the existing covariance structure analysis approach, we focus on the use of the raw observations rather than the sample covariance matrix.

Introduction

The basic attractive feature of a Bayesian approach is its flexibility to utilize useful prior information for achieving better results. In many practical problems, statisticians may have good **prior information** from some sources, for example *the knowledge of experts and analyses of similar data and/or past data.*

For example, in research relating to organization and management that involves latent variables about job performance and job satisfaction, we may have some prior information about the correlation of these latent variables, say a relatively large value is one that is larger than 0.4; we also may have some prior information on the values of the factor loadings, say a relatively large loading that corresponds to ‘salary’ and ‘job satisfaction’.

For situations **without accurate prior information**, some type of non-informative prior distributions can be used in a Bayesian approach. In these cases, the accuracy of the **Bayesian estimates is close to that of the ML estimates**.

Introduction

Before the 20th century, the Bayesian approach received little attention in SEM. Contributions are only limited to factor analysis (see, for example, Martin and McDonald, 1975; Lee, 1981; Bartholomew, 1981). More recently, the idea of data augmentation (Tanner and Wong, 1987) and the powerful tools in statistical computing for simulating observations from posterior distributions have greatly enhanced the applicability of the general Bayesian approach

A number of generalizations of the standard SEM have been separately developed by this approach. These include the developments of models with fixed covariates, nonlinear models, multilevel models, multisample models, mixture models, models with mixed continuous, dichotomous and/or ordered categorical variables, models with missing data, and models with data that are coming from an exponential family of distributions.

Introduction

Notations and concepts related to the Bayesian approach:

In a non-Bayesian approach, for example in an ML approach, $\boldsymbol{\theta}$ is not considered as random.
In a Bayesian approach, $\boldsymbol{\theta}$ is considered to be random with a distribution.

Bayesian inference is based on the observed data $\boldsymbol{Y}$ and the prior distribution of $\boldsymbol{\theta}$.

M — an arbitrary SEM with a vector of unknown parameters $\boldsymbol{\theta}$.

$\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ — the observed data set

$\boxed{p(\boldsymbol{\theta}|M) \text{ (or } p(\theta))}$ — the prior density function of $\boldsymbol{\theta}$.

$p(\mathbf{Y}, \boldsymbol{\theta}|M)$ — the probability density function of the joint distribution of $\mathbf{Y}$ and $\boldsymbol{\theta}$ under M.

$p(\boldsymbol{\theta}|\mathbf{Y}, M)$ — the density function of the posterior distribution of $\boldsymbol{\theta}$ under M. This function fully describes the behavior of $\boldsymbol{\theta}$ under the given data $\mathbf{Y}$.

Most Important

$p(\mathbf{Y}|\boldsymbol{\theta}, M)$ — the likelihood function under M.

$$\boxed{p(\mathbf{Y}, \boldsymbol{\theta}|M) = p(\mathbf{Y}|\boldsymbol{\theta}, M)p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{Y}, M)p(\mathbf{Y}|M)}$$

Prior Distribution and Posterior Analysis

Notations and concepts related to the Bayesian approach:

It is the probability density of $\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_n}$ conditional on the parameter vector $\boldsymbol{\theta}$.
[Sample Information]

Likelihood function / Log likelihood function

$$p(\theta|\mathbf{Y}, M) \propto \boxed{p(\mathbf{Y}|\theta, M)}\boxed{p(\theta)} \qquad \log p(\theta|\mathbf{Y}, M) = \boxed{\log p(\mathbf{Y}|\theta, M)} + \boxed{\log p(\theta)} + \text{constant}$$

Prior density function
[Prior Information]

As $p(\mathbf{Y}|M)$ does not depend on θ, and can be regarded as a constant with fixed $\mathbf{Y}$

$$p(\mathbf{Y}, \theta|M) = \boxed{p(\mathbf{Y}|\theta, M)p(\theta) = p(\theta|\mathbf{Y}, M)p(\mathbf{Y}|M)}$$

Prior Distribution and Posterior Analysis

Notations and concepts related to the Bayesian approach:

$p(\boldsymbol{Y}|\boldsymbol{\theta},M)$ depends on sample size n, whereas $p(\boldsymbol{\theta})$ does not.

Likelihood function / Log likelihood function

It is the probability density of $\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_n}$ conditional on the parameter vector $\boldsymbol{\theta}$.

[Sample Information]

$$p(\theta|\mathbf{Y}, M) \propto \boxed{p(\mathbf{Y}|\theta, M)}\boxed{p(\theta)}$$

$$\log p(\theta|\mathbf{Y}, M) = \boxed{\log p(\mathbf{Y}|\theta, M)} + \boxed{\log p(\theta)} + \text{constant}$$

Prior density function

[Prior Information]

The selection of the prior density is an important issue in Bayesian analysis.

When the **sample size n becomes arbitrarily large**, $\log p(\boldsymbol{Y}|\boldsymbol{\theta},M)$ could dominate $\log p(\boldsymbol{\theta})$. In this situation, $p(\boldsymbol{\theta})$ plays a less important role, and $\log p(\boldsymbol{\theta}|\boldsymbol{Y},M)$ is close to $\log p(\boldsymbol{Y}|\boldsymbol{\theta},M)$. Hence, asymptotically Bayesian and ML approaches are equivalent, and the Bayesian estimates have the same optimal properties as the ML estimates.

When **the sample sizes n are small or moderate**, the prior distribution of $\boldsymbol{\theta}$ plays a more substantial role in Bayesian estimation. Hence, in the problems with small or moderate sample sizes, $p(\boldsymbol{\theta})$ is useful for achieving better results

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

Noninformative prior distribution: we have little prior information about $\boldsymbol{\theta}$, and hence $p(\boldsymbol{\theta})$ plays a minimal role in $p(\boldsymbol{\theta}|\boldsymbol{Y},M)$. The associated prior density is chosen to be vague, diffuse, flat, or noninformative, for example a density that is proportional to a constant or has a huge variance. In this case, the Bayesian estimation is unaffected by information external to the observed data.

Informative prior distributions: we may have useful prior knowledge about $\boldsymbol{\theta}$, either from closely related data or from subjective knowledge of experts. Usually, an informative prior distribution has its own parameters, which are called **hyperparameters**.

The selection of the prior density is an important issue in Bayesian analysis.

When the **sample size n becomes arbitrarily large**, $\log p(\boldsymbol{Y}|\boldsymbol{\theta},M)$ could dominate $\log p(\boldsymbol{\theta})$. In this situation, $p(\boldsymbol{\theta})$ plays a less important role, and $\log p(\boldsymbol{\theta}|\boldsymbol{Y},M)$ is close to $\log p(\boldsymbol{Y}|\boldsymbol{\theta},M)$. Hence, asymptotically Bayesian and ML approaches are equivalent, and the Bayesian estimates have the same optimal properties as the ML estimates.

When **the sample sizes n are small or moderate**, the prior distribution of $\boldsymbol{\theta}$ plays a more substantial role in Bayesian estimation. Hence, in the problems with small or moderate sample sizes, $p(\boldsymbol{\theta})$ is useful for achieving better results

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

A commonly used informative prior distribution in the general Bayesian approach to statistical problems is the **conjugate prior distribution**.

Let us consider an example with the univariate binomial model. Considered as a function of $\boldsymbol{\theta}$, the likelihood of an observation $\boldsymbol{y}$ is of the form

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

If the prior density of $\boldsymbol{\theta}$ is of the same form ($Beta$ Distribution with hyperparameters α and β)

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

posterior density will also be of this form (Same parametric form)

$$\begin{aligned} f(x;\alpha,\beta) &= \text{constant} \cdot x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \end{aligned}$$

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \end{aligned}$$

$Beta$ Distribution with parameters $y+\alpha$ and $n-y+\beta$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

A commonly used informative prior distribution in the general Bayesian approach to statistical problems is the **conjugate prior distribution**.

A sample of independently and identically distributed (iid) observations $y_1, y_2, \cdots, y_n$ from $N[\theta, \sigma^2]$,where σ^2 is **known**.

$$p(\mathbf{Y}|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right]$$

A conjugate prior distribution of $\boldsymbol{\theta}$ can be parameterized as **Normal distribution** $N(\mu_0, \tau_0^2)$

$$p(\theta) \propto \exp\left[-\frac{1}{2\boxed{\tau_0^2}}(\theta - \boxed{\mu_0})^2\right]$$

hyperparameters

$$p(\theta|\mathbf{Y}) \propto p(\theta)p(\mathbf{Y}|\theta) \quad \propto \exp\left[-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right]p(\mathbf{Y}|\theta)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\tau_0^2}(\theta - \mu_0)^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right]\right\}$$

Normal distribution

$N(\tilde{\mu}, \tilde{\sigma}^2)$

$$\tilde{\mu} = \left(\frac{\sum_{i=1}^{n} y_i}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)\tilde{\sigma}^2$$

$$\tilde{\sigma}^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

A commonly used informative prior distribution in the general Bayesian approach to statistical problems is the **conjugate prior distribution**.

A sample of independently and identically distributed (iid) observations $y_1, y_2, \cdots, y_n$ from $N[\theta, \sigma^2]$,where θ **is known**, σ^2 is **unknown**.

$$p(\mathbf{Y}|\sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right]$$

$$\propto (\sigma^2)^{-n/2}\prod_{i=1}^{n}\exp\left\{\frac{1}{2\sigma^2}(y_i-\theta)^2\right\}$$

A conjugate prior distribution of $\boldsymbol{\theta}$ can be parameterized as **Inverted Gamma distribution**

$$X \sim \Gamma(\alpha,\beta) \equiv \mathrm{Gamma}(\alpha,\beta) \qquad f(x;\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)} \quad \text{for } x>0 \quad \alpha,\beta>0$$

$$Y = g(X) = \frac{1}{X} \qquad f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right| = \frac{\beta^\alpha}{\Gamma(\alpha)}(y)^{-\alpha-1}\exp\left(\frac{-\beta}{y}\right)$$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

A commonly used informative prior distribution in the general Bayesian approach to statistical problems is the <u>**conjugate prior distribution**</u>.

A sample of independently and identically distributed (iid) observations $y_1, y_2, \cdots, y_n$ from $N[\theta, \sigma^2]$,where θ <u>**is known**</u>, σ^2 is <u>**unknown**</u>.

$$p(\mathbf{Y}|\sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right]$$

$$\propto (\sigma^2)^{-n/2}\prod_{i=1}^{n}\exp\left\{\frac{1}{2\sigma^2}(y_i-\theta)^2\right\}$$

A conjugate prior distribution of σ^2 can be parameterized as <u>**Inverted Gamma distribution**</u>

$$p(\sigma^2|\mathbf{Y}) \propto p(\sigma^2)p(\mathbf{Y}|\sigma^2) \longleftarrow p(\sigma^2) \propto (\sigma^2)^{-(\alpha_0+1)}\exp(-\beta_0/\sigma^2)$$

$$\propto (\sigma^2)^{\boxed{-[n/2+(\alpha_0+1)]}}\exp\left[-\frac{1}{2\sigma^2}\boxed{(nv+2\beta_0)}\right]$$

$$v = n^{-1}\Sigma_{i=1}^{n}(y_i-\theta)^2$$

$$\frac{\beta^\alpha}{\Gamma(\alpha)}(y)^{-\alpha-1}\exp\left(\frac{-\beta}{y}\right)$$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

A commonly used informative prior distribution in the general Bayesian approach to statistical problems is the **conjugate prior distribution**.

A sample of independently and identically distributed (iid) observations $y_1, y_2, \cdots, y_n$ from $N[\mu, \sigma^2]$,where μ **is unknown**, σ^2 is **unknown**.

$$p(\mathbf{Y}|\mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} \ \sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

A conjugate prior distribution of μ and σ^2 $\xrightarrow{\text{Same Form}}$ $\propto \ (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right\}$

$$p(\mu, \sigma^2) = p(\sigma^2) p(\mu|\sigma^2)$$

$$p(\mu, \sigma^2|\mathbf{Y}) \propto p(\mathbf{Y}|\mu, \sigma^2) p(\mu, \sigma^2)$$

inverted-gamma distribution

$$p(\sigma^2) \propto (\sigma^2)^{-(\alpha_0+1)} \exp(-\beta_0/\sigma^2)$$

normal distribution $N[\mu_0, \sigma^2\tau_0^2]$

$$p(\mu|\sigma^2) \ \propto \ (\sigma^2)^{-1/2} \ \exp[-(\mu-\mu_0)^2/(2\sigma^2\tau_0^2)]$$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

A commonly used informative prior distribution in the general Bayesian approach to statistical problems is the **conjugate prior distribution**.

A sample of independently and identically distributed (iid) observations $y_1, y_2, \cdots, y_n$ from $N[\mu, \sigma^2]$,where μ **is unknown**, σ^2 is **unknown**.

$$p(\mathbf{Y}|\mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

A conjugate prior distribution of μ and σ^2 $\xrightarrow{\text{Same Form}}$ $\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right\}$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu|\sigma^2)$$

$$p(\mu, \sigma^2|\mathbf{Y}) \propto p(\mathbf{Y}|\mu, \sigma^2)p(\mu, \sigma^2)$$

normal-inverted gamma distribution

$$\begin{aligned} p(\mu, \sigma^2) &\propto (\sigma^2)^{-(\alpha_0+1)}(\sigma^2)^{-1/2} \exp(-\beta_0/\sigma^2)\exp[-(\mu-\mu_0)^2/(2\sigma^2\tau_0^2)], \\ &\propto (\sigma^2)^{-(\alpha_0+1)}(\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(2\beta_0 + \frac{(\mu-\mu_0)^2}{\tau_0^2}\right)\right], \end{aligned}$$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

The above discussion on scalar parameters motivates the selection of conjugate type prior distributions for the parameters in Bayesian analyses of SEMs

$$\mathbf{y}_i = \boxed{\boldsymbol{\Lambda}}\omega_i + \boldsymbol{\epsilon}_i$$

$\boldsymbol{\Lambda}_k^T$ be the k^{th} row of $\boldsymbol{\Lambda}$

$\boldsymbol{\omega}_i(q \times 1)$ is distributed as $N[\mathbf{0}, \boxed{\boldsymbol{\Phi}}]$

$\boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \boxed{\boldsymbol{\Psi}_\epsilon}]$

Independent

diagonal elements ψ_{ϵ_k}

$$[\boldsymbol{\Lambda}_k|\psi_{\epsilon k}] \overset{D}{=} N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}]$$

related to the mean vectors

related to the co/variance

$$\psi_{\epsilon k} \overset{D}{=} IG[\alpha_{0\epsilon k}, \beta_{0\epsilon k}] \text{ or } \psi_{\epsilon k}^{-1} \overset{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$$

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

The above discussion on scalar parameters motivates the selection of conjugate type prior distributions for the parameters in Bayesian analyses of SEMs

$$\mathbf{y}_i = \boxed{\mathbf{\Lambda}}\omega_i + \boldsymbol{\epsilon}_i$$

$\mathbf{\Lambda}_k^T$ be the k^{th} row of $\mathbf{\Lambda}$

$\boldsymbol{\omega}_i(q \times 1)$ is distributed as $N[0, \boxed{\mathbf{\Phi}}]$

$\boldsymbol{\epsilon}_i$ is distributed as $N[0, \boxed{\mathbf{\Psi}_\epsilon}]$

Independent

diagonal elements ψ_{ϵ_k}

$$\mathbf{\Phi}^{-1} \stackrel{D}{=} W_q[\mathbf{R}_0, \rho_0], \text{ or equivalently } \mathbf{\Phi} \stackrel{D}{=} IW_q(\mathbf{R}_0^*, \rho_0)$$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{|\mathbf{x}|^{(n-p-1)/2} e^{-\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x})/2}}{2^{\frac{np}{2}} |\mathbf{V}|^{n/2} \Gamma_p(\frac{n}{2})}$$

related to the co/variance

A multivariate extension of Gamma distribution

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

Noninformative prior distribution

Hyperparameters

Take a LARGE variance in the corresponding prior distribution

Informative prior distributions

Select the prior distribution with a SMALL variance

$$[\boldsymbol{\Lambda}_k|\psi_{\epsilon k}] \overset{D}{=} N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}]$$

related to the mean vectors

related to the co/variance

$$\psi_{\epsilon k} \overset{D}{=} IG[\alpha_{0\epsilon k}, \beta_{0\epsilon k}] \text{ or } \psi_{\epsilon k}^{-1} \overset{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$$

$$\boldsymbol{\Phi}^{-1} \overset{D}{=} W_q[\mathbf{R}_0, \rho_0], \text{ or equivalently } \boldsymbol{\Phi} \overset{D}{=} IW_q(\mathbf{R}_0^*, \rho_0)$$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{|\mathbf{x}|^{(n-p-1)/2} e^{-\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x})/2}}{2^{\frac{np}{2}} |\mathbf{V}|^{n/2} \Gamma_p(\frac{n}{2})}$$

related to the co/variance

A multivariate extension of Gamma distribution

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

Noninformative prior distribution

Hyperparameters

Take a LARGE variance in the corresponding prior distribution

Informative prior distributions

Select the prior distribution with a SMALL variance

$$[\boldsymbol{\Lambda}_k | \psi_{\epsilon k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}]$$ **related to the mean vectors**

related to the co/variance $\psi_{\epsilon k} \stackrel{D}{=} IG[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$ or $\psi_{\epsilon k}^{-1} \stackrel{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$

If we have confidence that the true $\boldsymbol{\Lambda}_k$ is not too far away from the preassigned hyperparameter value $\boldsymbol{\Lambda}_{0k}$, then $\boldsymbol{H}_{0yk}$ should be taken as a matrix with small variances (such as $0.5\boldsymbol{I}$).

If we think that the variation of ϵ_k is small (that is $\boldsymbol{\Lambda}_k^T\boldsymbol{\omega}_i$ is a good predictor of y_{ik}), then the prior distribution of $\psi_{\epsilon k}$ should have a small mean value as well as a small variance. Otherwise, the prior distribution of $\psi_{\epsilon k}$ should have a large mean value and/or a large variance.

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

Noninformative prior distribution

Hyperparameters

Take a LARGE variance in the corresponding prior distribution

Informative prior distributions

Select the prior distribution with a SMALL variance

$$[\boldsymbol{\Lambda}_k|\psi_{\epsilon k}] \overset{D}{=} N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}]$$ **related to the mean vectors**

related to the co/variance $$\psi_{\epsilon k} \overset{D}{=} IG[\alpha_{0\epsilon k}, \beta_{0\epsilon k}] \text{ or } \psi_{\epsilon k}^{-1} \overset{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$$

$$E(\psi_{\epsilon k}) = \beta_{0\epsilon k}/(\alpha_{0\epsilon k} - 1),$$
$$Var(\psi_{\epsilon k}) = \beta_{0\epsilon k}^2/\{(\alpha_{0\epsilon k} - 1)^2(\alpha_{0\epsilon k} - 2)\}$$

if $\alpha_{0\epsilon k} = 9$ and $\beta_{0\epsilon k} = 4$, then

$E(\psi_{\epsilon k}) = 4/8 = 0.5,\quad Var(\psi_{\epsilon k}) = 4^2/\{(9-1)^2(9-2)\} = 1/28$

if $\alpha_{0k} = 6$ and $\beta_{0\epsilon k} = 10$, then $E(\psi_{\epsilon k}) = 2.0$ $Var(\psi_{\epsilon k}) = 1.0$

If we think that the variation of ϵ_k is small (that is $\boldsymbol{\Lambda}_k^T\boldsymbol{\omega}_i$ is a good predictor of y_{ik}), then the prior distribution of $\psi_{\epsilon k}$ should have a small mean value as well as a small variance. Otherwise, the prior distribution of $\psi_{\epsilon k}$ should have a large mean value and/or a large variance.

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

Noninformative prior distribution

Hyperparameters

Take a LARGE variance in the corresponding prior distribution

Informative prior distributions

Select the prior distribution with a SMALL variance

mean of $\boldsymbol{\Phi}$ is $\mathbf{R}_0^{-1}/\{\rho_0 - q - 1\}$

If we have confidence that is not too far away from $\boldsymbol{\Phi}_0$, we can choose $\mathbf{R}_0^{-1}$ and ρ_0 such that $\mathbf{R}_0^{-1} = (\rho_0 - q - 1)\boldsymbol{\Phi}_0$

Other values of $\mathbf{R}_0^{-1}$ and ρ_0 may be considered for situations without good prior information. (Large value of ρ_0)

$$\boldsymbol{\Phi}^{-1} \overset{D}{=} W_q[\mathbf{R}_0, \rho_0], \text{ or equivalently } \boldsymbol{\Phi} \overset{D}{=} IW_q(\mathbf{R}_0^*, \rho_0)$$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{|\mathbf{x}|^{(n-p-1)/2} e^{-\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x})/2}}{2^{\frac{np}{2}} |\mathbf{V}|^{n/2} \Gamma_p(\frac{n}{2})}$$

related to the co/variance

A multivariate extension of Gamma distribution

Prior Distribution and Posterior Analysis

Prior Distribution $\log p(\boldsymbol{\theta})$ / $p(\boldsymbol{\theta})$

Noninformative prior distribution

Hyperparameters

Take a LARGE variance in the corresponding prior distribution

Informative prior distributions

Select the prior distribution with a SMALL variance

If the sample size is large, one possible method to get $\boldsymbol{\Lambda}_{0k}$ and $\boldsymbol{\Phi}_0$ is to use a portion of the data, say one-third or less, to conduct an auxiliary Bayesian estimation with noninformative priors to get initial Bayesian estimates. Then, use the remaining data to do the actual Bayesian analysis, by utilizing the initial Bayesian estimates as hyperparameter values.

For situations with moderate sample sizes, Bayesian analysis may be done by applying data dependent prior inputs that are obtained from an initial estimation with the whole data set.

In general, a **sensitivity analysis** should be conducted to see whether the results are robust to prior inputs. This can be done by perturbing the given hyperparameter values or considering some ad hoc prior inputs.

Prior Distribution and Posterior Analysis

Posterior analysis using Markov chain Monte Carlo methods

MCMC methods

Bayesian estimate of θ is usually defined as the mean of the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y}, M)$.

Theoretically, it could be obtained via **integration**. For most situations, the integration does not have a closed form. However, if we can simulate a suffciently large number of observations from $p(\boldsymbol{\theta}|\mathbf{Y}, M)$, we can approximate the mean and other useful statistics through the simulated observations.

$$p(\boldsymbol{\theta}|\mathbf{Y}, M) \quad + \quad p(\boldsymbol{\theta}, \boldsymbol{\Omega}|\mathbf{Y}, M) \quad \longrightarrow \quad p(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{Y}, M) \quad p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \mathbf{Y}, M)$$

Intractable posterior density

Ω is the set latent variable

Gibbs Sampler

The **presence of latent variables** causes the difficulties in analyzing the model. The feature that makes SEMs different from the common regression model and the simultaneous equation model is the existence of latent random variables. However, if the latent random variables are observed, the SEM will become the familiar regression model or simultaneous equation model that can be handled without much difficulty.

Prior Distribution and Posterior Analysis

Posterior analysis using Markov chain Monte Carlo methods

MCMC methods

Gibbs Sampler

The Gibbs sampler is a Markov chain algorithm which performs an alternating conditional sampling at each of its iterations.

At the j^{th} iteration with current values

$$
\begin{array}{lll}
\boldsymbol{\theta}_1^{(j+1)} & \text{from} & p\left(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(j)}, \ldots, \boldsymbol{\theta}_a^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{Y}\right), \\
\boldsymbol{\theta}_2^{(j+1)} & \text{from} & p\left(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(j+1)}, \ldots, \boldsymbol{\theta}_a^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{Y}\right), \\
\vdots & & \vdots \\
\boldsymbol{\theta}_a^{(j+1)} & \text{from} & p\left(\boldsymbol{\theta}_a | \boldsymbol{\theta}_1^{(j+1)}, \ldots, \boldsymbol{\theta}_{a-1}^{(j+1)}, \boldsymbol{\Omega}^{(j)}, \mathbf{Y}\right), \\
\boldsymbol{\Omega}_1^{(j+1)} & \text{from} & p\left(\boldsymbol{\Omega}_1 | \boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}_2^{(j)}, \cdots, \boldsymbol{\Omega}_b^{(j)}, \mathbf{Y}\right), \\
\boldsymbol{\Omega}_2^{(j+1)} & \text{from} & p\left(\boldsymbol{\Omega}_2 | \boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}_1^{(j+1)}, \ldots, \boldsymbol{\Omega}_b^{(j)}, \mathbf{Y}\right), \\
\vdots & & \vdots \\
\boldsymbol{\Omega}_b^{(j+1)} & \text{from} & p\left(\boldsymbol{\Omega}_b | \boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}_1^{(j+1)}, \ldots, \boldsymbol{\Omega}_{b-1}^{(j+1)}, \mathbf{Y}\right).
\end{array}
$$

Usually normal, gamma, or inverted Wishart.

It cycles through the components of $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$, drawing each component conditional on the others.

Prior Distribution and Posterior Analysis

Posterior analysis using Markov chain Monte Carlo methods

The Metropolis–Hastings algorithm

~~Usually normal, gamma, or inverted Wishart.~~

Suppose we wish to simulate observations, say $\{X_j, j = 1, 2, \ldots\}$, from a conditional distribution with target density $p(\cdot)$.

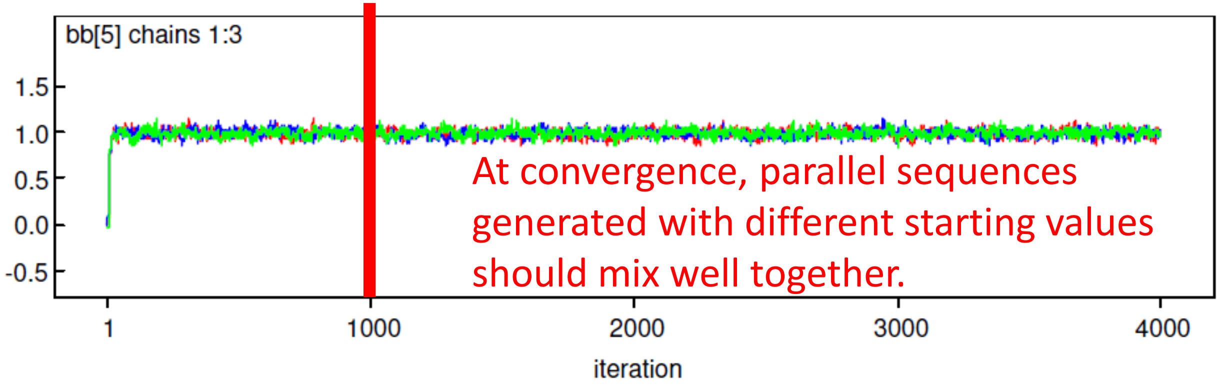
Step 1 At the j^{th} iteration of the Metropolis–Hastings algorithm with a current value X_j, the next value X_{j+1} is chosen by first sampling a candidate point Y from a proposal distribution $q(\cdot\,|X_j)$ which is easy to sample. (Normal distribution with Mean X_j and a fixed covariance matrix)

Step 2 This candidate point Y is accepted as X_{j+1} with probability $\min\left(1, \frac{p(Y)q(X_j|Y)}{p(X_j)q(Y|X_j)}\right)$

If the candidate point Y is rejected, then $X_{j+1} = X_j$ and the chain does not move.

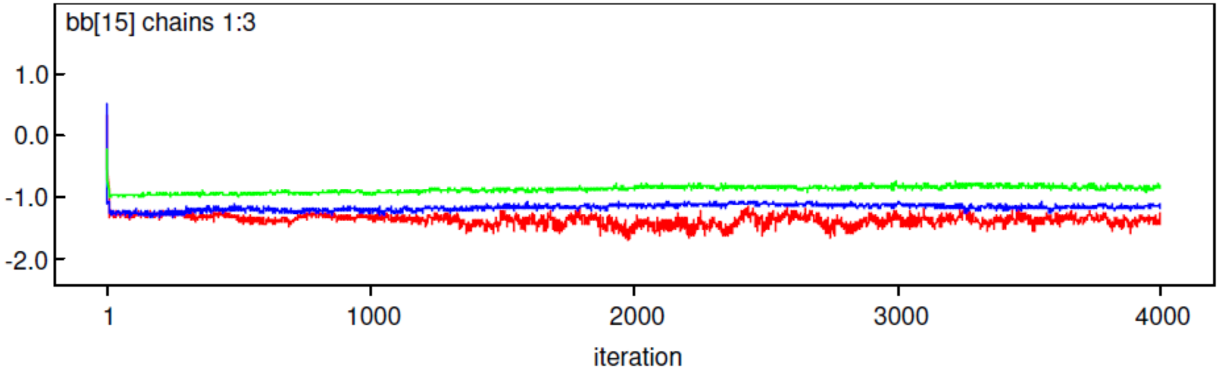
Posterior analysis using Markov chain Monte Carlo methods

It has been shown that under mild regularity conditions, the joint distribution of $(\boldsymbol{\theta}^{(j)}, \boldsymbol{\Omega}^{(j)})$ converges to the desired posterior distribution $\boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{\Omega}|\mathbf{Y}, M)$ after a sufficiently large number of iterations, say J.



a

The required number of iterations for achieving convergence of the Gibbs sampler, that is the burn-in iterations J, can be determined by plots of the simulated sequences of the individual parameters.



b

Prior Distribution and Posterior Analysis

Posterior analysis using Markov chain Monte Carlo methods

This gives a direct Bayesian estimate which is not expressed in terms of the structural parameter estimates

$$\hat{\boldsymbol{\theta}} = T^{*-1} \sum_{t=1}^{T^*} \boldsymbol{\theta}^{(t)},$$

tends to $E(\theta|Y)$ as T^* tends to infinity.

$$\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\mathbf{Y}) = (T^* - 1)^{-1} \sum_{t=1}^{T^*} (\boldsymbol{\theta}^{(t)} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(t)} - \hat{\boldsymbol{\theta}})^T$$

Bayesian credible interval

Estimate of the standard deviation

A Bayesian estimate $\widehat{\boldsymbol{\omega}}_i$ can be obtained through

$$\hat{\boldsymbol{\omega}}_i = T^{*-1} \sum_{t=1}^{T^*} \boldsymbol{\omega}_i^{(t)}$$

tends to $E(\omega_i|\boldsymbol{y}_i)$ as T^* tends to infinity.

used for outlier and residual analyses

Prior Distribution and Posterior Analysis

Posterior analysis using Markov chain Monte Carlo methods

The Bayesian approach has several advantages.

- It has potential to be applied to more complex situations.
- It produces a direct estimation of latent variables, which cannot be obtained with classical methods.
- In addition to the information that is available in the observed data, the Bayesian approach allows the use of genuine prior information for producing better results.
- It can give more reliable results for small samples.

Bayesian Estimation

CFA Model

$\mathbf{y}_i$ is a $p \times 1$ observed random vector

For $i = 1, \ldots, n$ $\qquad$ $\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$

$\boldsymbol{\omega}_i (q \times 1)$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\qquad$ $\boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$

Once $\boldsymbol{\Omega}$ is given rather than random, it becomes the familiar regression model.

STEP A: Generate a random variate $\boldsymbol{\Omega}^{(j+1)}$ from the conditional distribution $[\boldsymbol{\Omega}|\mathbf{Y}, \boldsymbol{\theta}^{(j)}]$.

STEP B: Generate a random variate $\boldsymbol{\theta}^{(j+1)}$ from the conditional distribution $[\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Omega}^{(j+1)}]$, and return to 'Step a' if necessary.

Bayesian Estimation

Conditional Distributions

$$p(\Omega|\mathrm{Y},\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\mathbf{y}_i,\boldsymbol{\theta}) \propto \prod_{i=1}^{n} \boxed{p(\boldsymbol{\omega}_i|\boldsymbol{\theta})}\ \boxed{p(\mathbf{y}_i|\boldsymbol{\omega}_i,\boldsymbol{\theta})}$$

$$\boxed{N(0,\boldsymbol{\Phi})} \quad \boxed{N(\boldsymbol{\Lambda}\boldsymbol{\omega}_i,\boldsymbol{\Psi}_\epsilon)}$$

Still Normal Distribution

$$[\boldsymbol{\omega}_i|\mathbf{y}_i,\boldsymbol{\theta}] \stackrel{D}{=} N\left[(\boldsymbol{\Phi}^{-1}+\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i,(\boldsymbol{\Phi}^{-1}+\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda})^{-1}\right]$$

Bayesian Estimation

Conditional Distributions

Normal distribution

$$p(\Lambda, \Psi_\epsilon, \Phi | Y, \Omega) = p(\theta | Y, \Omega) \propto p(Y, \Omega | \theta) p(\theta) = \boxed{p(Y|\theta, \Omega)}\boxed{p(\Omega|\theta)}\boxed{p(\theta)}$$

Prior distribution of $\boldsymbol{\theta}$

Parameters in regression model (Given Ω)

Φ is only involved in the distribution of Ω

$$p(\theta) = p(\Lambda, \Phi, \Psi_\epsilon) = \boxed{p(\Lambda, \Psi_\epsilon)}\boxed{p(\Phi)}.$$

Independent

$$= p(Y|\theta, \Omega) p(\Omega|\theta) p(\Lambda, \Psi_\epsilon) p(\Phi)$$

$$= \boxed{[p(\Lambda, \Psi_\epsilon) p(Y|\Lambda, \Psi_\epsilon, \Omega)]}\boxed{[p(\Omega|\Phi) p(\Phi)]}$$

Treat Separately

Generate $\Psi_\epsilon^{(j+1)}$ from $p(\Psi_\epsilon | \Omega^{(j+1)}, \Lambda^{(j)}, \Phi^{(j)}, Y)$.

Generate $\Lambda^{(j+1)}$ from $p(\Lambda | \Omega^{(j+1)}, \Psi_\epsilon^{(j+1)}, \Phi^{(j)}, Y)$.

Generate $\Phi^{(j+1)}$ from $p(\Phi | \Omega^{(j+1)}, \Psi_\epsilon^{(j+1)}, \Lambda^{(j+1)}, Y)$

Bayesian Estimation

Conditional Distributions

$$p(\boldsymbol{\Phi}|\mathbf{Y},\boldsymbol{\Omega}) \propto \boxed{[p(\boldsymbol{\Omega}|\boldsymbol{\Phi})p(\boldsymbol{\Phi})]} \propto \boxed{p(\boldsymbol{\Phi})}\boxed{\prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\boldsymbol{\Phi})}$$

$$\boxed{IW_q[\mathbf{R}_0^{-1}, \rho_0]} \qquad \boxed{N(\mathbf{0}, \boldsymbol{\Phi})}$$

$$p(\boldsymbol{\Phi}|\mathbf{Y},\boldsymbol{\Omega}) \propto \left[|\boldsymbol{\Phi}|^{-(\rho_0+q+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}[\mathbf{R}_0^{-1}\boldsymbol{\Phi}^{-1}]\right)\right]$$

$$\times \left\{|\boldsymbol{\Phi}|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} \boldsymbol{\omega}_i^T\boldsymbol{\Phi}^{-1}\boldsymbol{\omega}_i\right]\right\}$$

$$\boxed{IW_q[(\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \mathbf{R}_0^{-1}),\ n+\rho_0]}$$

$$= |\boldsymbol{\Phi}|^{-(n+\rho_0+q+1)/2} \exp\left\{-\frac{1}{2}\ \mathrm{tr}[\boldsymbol{\Phi}^{-1}(\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \mathbf{R}_0^{-1})]\right\}$$

Bayesian Estimation

Conditional Distributions

$\boldsymbol{\Lambda}_k^T$ be the k^{th} row of $\boldsymbol{\Lambda}$
ψ_{ϵ_k} be the k^{th} diagonal elements of $\boldsymbol{\Psi}_\epsilon$ $k = 1,2,\cdots,p$

$$p(\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon | \mathbf{Y}, \boldsymbol{\Omega}) \propto \boxed{[p(\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon) p(\mathbf{Y}|\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\Omega})]}$$

$$[\boldsymbol{\Lambda}_k|\psi_{\epsilon k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}]$$ **related to the mean vectors**

related to the co/variance $\psi_{\epsilon k} \stackrel{D}{=} IG[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$ or $\psi_{\epsilon k}^{-1} \stackrel{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$

$$[\psi_{\epsilon k}^{-1}|\mathbf{Y}, \boldsymbol{\Omega}] \stackrel{D}{=} Gamma\,[n/2 + \alpha_{0\epsilon k},\ \beta_{\epsilon k}]$$

$$[\boldsymbol{\Lambda}_k|\mathbf{Y}, \boldsymbol{\Omega}, \psi_{\epsilon k}^{-1}] \stackrel{D}{=} N\,[\mathbf{a}_k, \psi_{\epsilon k}\mathbf{A}_k]$$

$$\beta_{\epsilon k} = \beta_{0\epsilon k} + 2^{-1}(\mathbf{Y}_k^T\mathbf{Y}_k - \mathbf{a}_k^T\mathbf{A}_k^{-1}\mathbf{a}_k + \boldsymbol{\Lambda}_{0k}^T\mathbf{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k})$$

$$\mathbf{A}_k = (\mathbf{H}_{0yk}^{-1} + \boldsymbol{\Omega}\boldsymbol{\Omega}^T)^{-1}$$

$$\mathbf{a}_k = \mathbf{A}_k(\mathbf{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k} + \boldsymbol{\Omega}\mathbf{Y}_k)$$

Normal – (Inverted) gamma distribution

$\mathbf{Y}_k^T$ be the kth row of $\mathbf{Y}$

Bayesian Estimation

Conditional Distributions

$$p(\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon | \mathbf{Y}, \mathbf{\Omega}) \propto \boxed{[p(\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon) p(\mathbf{Y} | \mathbf{\Lambda}, \mathbf{\Psi}_\epsilon, \mathbf{\Omega})]}$$

$\mathbf{\Lambda}_k^T$ be the k^{th} row of $\mathbf{\Lambda}$
ψ_{ϵ_k} be the k^{th} diagonal elements of $\mathbf{\Psi}_\epsilon$ $\quad k = 1,2,\cdots,p$

Extended to handle the general situation with fixed known elements in $\mathbf{\Lambda}$

Let $\boldsymbol{c}_k$ be the corresponding $1 \times q$ row vector

$c_{kj} = 0$ if λ_{kj} is a **fixed parameter**

$c_{kj} = 1$ if λ_{kj} is an **unknown parameter**

$$r_k = c_{k1} + \cdots + c_{kq}$$

Number of unknown parameter in k

$\mathbf{\Lambda}_k^{*T}$ be the $1 \times r_k$ vector contains the **unknown parameters** in of $\mathbf{\Lambda}_k$

$\mathbf{\Omega}_k^*$ be the $r_k \times n$ vector submatrix of $\mathbf{\Omega}$
(Delete the rows corresponding $c_{kj} = 0$)

$\mathbf{Y}_k^{*T}$ be the $1 \times n$ vector $(y_{1k}^*, \cdots, y_{nk}^*)$

$$y_{ik}^* = y_{ik} - \sum_{j=1}^{q} \lambda_{kj} \omega_{ij} (1 - c_{kj})$$

Bayesian Estimation

Conditional Distributions

$$p(\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon | \mathbf{Y}, \mathbf{\Omega}) \propto \boxed{[p(\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon) p(\mathbf{Y} | \mathbf{\Lambda}, \mathbf{\Psi}_\epsilon, \mathbf{\Omega})]}$$

$\mathbf{\Lambda}_k^T$ be the k^{th} row of $\mathbf{\Lambda}$
ψ_{ϵ_k} be the k^{th} diagonal elements of $\mathbf{\Psi}_\epsilon$

Extended to handle the general situation with fixed known elements in $\mathbf{\Lambda}$

$$[\psi_{\epsilon k}^{-1} | \mathbf{Y}, \mathbf{\Omega}] \overset{D}{=} Gamma\,[n/2 + \alpha_{0\epsilon k},\ \beta_{\epsilon k}]$$

$$[\mathbf{\Lambda}_k^* | \mathbf{Y}, \mathbf{\Omega}, \psi_{\epsilon k}^{-1}] \overset{D}{=} N\,[\mathbf{a}_k^*, \psi_{\epsilon k} \mathbf{A}_k^*]$$

Normal – (Inverted) gamma distribution

$$\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(\mathbf{Y}_k^{*T} \mathbf{Y}_k^* - \mathbf{a}_k^{*T} \mathbf{A}_k^{*-1} \mathbf{a}_k^* + \mathbf{\Lambda}_{0k}^{*T} \mathbf{H}_{0yk}^{*-1} \mathbf{\Lambda}_{0k}^*)$$

$$\mathbf{A}_k^* = (\mathbf{H}_{0yk}^{*-1} + \mathbf{\Omega}_k^* \mathbf{\Omega}_k^{*T})^{-1}$$

$$\mathbf{a}_k^* = \mathbf{A}_k^* (\mathbf{H}_{0yk}^{*-1} \mathbf{\Lambda}_{0k}^* + \mathbf{\Omega}_k^* \mathbf{Y}_k^*)$$

Bayesian Estimation

CFA Model <u>With Mean structure and Covariates</u>

$\mathbf{y}_i$ is a $p \times 1$ observed random vector

For $i = 1, \ldots, n$ $\qquad \mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i \longrightarrow \boldsymbol{y_i} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{c_i} + \boldsymbol{\Lambda}\boldsymbol{\omega_i} + \boldsymbol{\epsilon_i}$

$\boldsymbol{\omega}_i(q \times 1)$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}]$ $\qquad \boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$

Once $\boldsymbol{\Omega}$ is given rather than random, it becomes the familiar regression model.

STEP A: Generate a random variate $\Omega^{(j+1)}$ from the conditional distribution $[\Omega|\mathbf{Y}, \theta^{(j)}]$.

STEP B: Generate a random variate $\theta^{(j+1)}$ from the conditional distribution $[\theta|\mathbf{Y}, \Omega^{(j+1)}]$, and return to 'Step a' if necessary.

Bayesian Estimation

Conditional Distributions

$$p\left(\Omega|\mathrm{Y},\boldsymbol{\theta}\right)=\prod_{i=1}^{n}p\left(\boldsymbol{\omega}_i|\mathrm{y}_i,\boldsymbol{\theta}\right)\propto\prod_{i=1}^{n}\boxed{p\left(\boldsymbol{\omega}_i|\boldsymbol{\theta}\right)}\,\boxed{p\left(\mathrm{y}_i|\boldsymbol{\omega}_i,\boldsymbol{\theta}\right)}$$

$$\boxed{N(0,\,\boldsymbol{\Phi})}\quad\boxed{N(\boldsymbol{\Lambda}\boldsymbol{\omega}_i,\,\boldsymbol{\Psi}_\epsilon)}$$

Still Normal Distribution

$$[\boldsymbol{\omega}_i|\mathrm{y}_i,\boldsymbol{\theta}]\overset{D}{=}N\left[(\boldsymbol{\Phi}^{-1}+\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boxed{\mathrm{y}_i},\,(\boldsymbol{\Phi}^{-1}+\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda})^{-1}\right]$$

$$\boxed{\boldsymbol{y_i-\mu-Ac_i}}$$

Bayesian Estimation

Conditional Distributions

$$\boldsymbol{y_i} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{c_i} + \boldsymbol{\Lambda}\boldsymbol{\omega_i} + \boldsymbol{\epsilon_i}$$

$$p(\Lambda, \Psi_\epsilon | \mathbf{Y}, \boldsymbol{\Omega}) \propto \boxed{[p(\Lambda, \Psi_\epsilon) p(\mathbf{Y}|\Lambda, \Psi_\epsilon, \boldsymbol{\Omega})]}$$

$$\boldsymbol{u_i} = (\mathbf{1}^T, \boldsymbol{c_i}^T, \boldsymbol{\omega_i}^T)^T \quad \boldsymbol{\Lambda_y} = (\boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{\Lambda})$$

Extended to handle the general situation with fixed known elements in $\boldsymbol{\Lambda}$

$$\boldsymbol{y_i} = \boldsymbol{\Lambda_y}\boldsymbol{u_i} + \boldsymbol{\epsilon_i}$$

$$[\psi_{\epsilon k}^{-1} | \mathbf{Y}, \boldsymbol{\Omega}] \stackrel{D}{=} Gamma[n/2 + \alpha_{0\epsilon k}, \ \beta_{\epsilon k}]$$

$$[\boldsymbol{\Lambda}_k^* | \mathbf{Y}, \boldsymbol{\Omega}, \psi_{\epsilon k}^{-1}] \stackrel{D}{=} N[\mathbf{a}_k^*, \psi_{\epsilon k}\mathbf{A}_k^*]$$

$$\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(\mathbf{Y}_k^{*T}\mathbf{Y}_k^* - \mathbf{a}_k^{*T}\mathbf{A}_k^{*-1}\mathbf{a}_k^* + \boldsymbol{\Lambda}_{0k}^{*T}\mathbf{H}_{0yk}^{*-1}\boldsymbol{\Lambda}_{0k}^*)$$

$$\mathbf{A}_k^* = (\mathbf{H}_{0yk}^{*-1} + \boldsymbol{\Omega}_k^*\boldsymbol{\Omega}_k^{*T})^{-1}$$

$$\mathbf{a}_k^* = \mathbf{A}_k^*(\mathbf{H}_{0yk}^{*-1}\boldsymbol{\Lambda}_{0k}^* + \boldsymbol{\Omega}_k^*\mathbf{Y}_k^*)$$

Normal – (Inverted) gamma distribution

Bayesian Estimation

SEM Model

$\mathbf{y}_i$ is a $p \times 1$ observed random vector

For $i = 1, \ldots, n$

$$\boldsymbol{y_i} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{c_i} + \boldsymbol{\Lambda}\boldsymbol{\omega_i} + \boldsymbol{\epsilon_i}$$

$$\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i$$

$\boldsymbol{\xi}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}]$

$\boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$

$\boldsymbol{\delta}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

$$\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$$

STEP A: Generate a random variate $\Omega^{(j+1)}$ from the conditional distribution $[\Omega|\mathbf{Y}, \theta^{(j)}]$.

STEP B: Generate a random variate $\theta^{(j+1)}$ from the conditional distribution $[\theta|\mathbf{Y}, \Omega^{(j+1)}]$, and return to 'Step a' if necessary.

Bayesian Estimation

SEM Model

$\mathbf{y}_i$ is a $p \times 1$ observed random vector

For $i = 1, \ldots, n$

$$\boldsymbol{y_i} = \boldsymbol{\mu} + \boldsymbol{A c_i} + \boldsymbol{\Lambda \omega_i} + \boldsymbol{\epsilon_i}$$

$$\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i$$

$\boldsymbol{\xi}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}]$

$\boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$

$\boldsymbol{\delta}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$

$$\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$$

$$[\boldsymbol{\omega}_i|\boldsymbol{\theta}] \stackrel{D}{=} N\left[\begin{pmatrix} \boldsymbol{\Pi}_0^{-1}\mathbf{B}\mathbf{d}_i \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\Sigma}_\omega\right] \quad \boldsymbol{\Sigma}_\omega = \begin{bmatrix} \boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma\Phi\Gamma}^T + \boldsymbol{\Psi}_\delta)\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma\Phi} \\ \boldsymbol{\Phi\Gamma}^T\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Phi} \end{bmatrix}$$

$$\boldsymbol{\Pi}_0 = \mathbf{I} - \boldsymbol{\Pi}$$

Bayesian Estimation

Conditional Distributions

$$[\boldsymbol{\omega}_i|\mathbf{y}_i,\boldsymbol{\theta}] \stackrel{D}{=} N\left[\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}-\boldsymbol{A}\boldsymbol{c}_i)+\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_\omega^{-1}\begin{pmatrix}\boldsymbol{\Pi}_0^{-1}\mathbf{B}\mathbf{d}_i\\ \mathbf{0}\end{pmatrix}, \boldsymbol{\Sigma}^{*-1}\right]$$

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}$$

$$\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$$

$$[\boldsymbol{\omega}_i|\boldsymbol{\theta}] \stackrel{D}{=} N\left[\begin{pmatrix}\boldsymbol{\Pi}_0^{-1}\mathbf{B}\mathbf{d}_i\\ \mathbf{0}\end{pmatrix}, \boldsymbol{\Sigma}_\omega\right] \quad \boldsymbol{\Sigma}_\omega = \begin{bmatrix}\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^T+\boldsymbol{\Psi}_\delta)\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\\ \boldsymbol{\Phi}\boldsymbol{\Gamma}^T\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Phi}\end{bmatrix}$$

$$\boldsymbol{\Pi}_0 = \mathbf{I} - \boldsymbol{\Pi}$$

Bayesian Estimation

Conditional Distributions

$$p(\boldsymbol{\Phi}|\mathbf{Y},\boldsymbol{\Omega}) \propto \boxed{[p(\boldsymbol{\Omega}|\boldsymbol{\Phi})p(\boldsymbol{\Phi})]} \propto \underbrace{\boxed{p(\boldsymbol{\Phi})}}_{IW_q[\mathbf{R}_0^{-1},\,\rho_0]} \underbrace{\boxed{\prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\boldsymbol{\Phi})}}_{N(\mathbf{0},\,\boldsymbol{\Phi})}$$

$$p(\boldsymbol{\Phi}|\mathbf{Y},\boldsymbol{\Omega}) \propto \left[|\boldsymbol{\Phi}|^{-(\rho_0+q+1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}[\mathbf{R}_0^{-1}\boldsymbol{\Phi}^{-1}]\right)\right]$$

$$\times\left\{|\boldsymbol{\Phi}|^{-n/2}\ \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\boxed{\boldsymbol{\xi}_i^T}\boldsymbol{\Phi}^{-1}\boxed{\boldsymbol{\xi}_i}\right]\right\}$$

$$\boxed{IW_q[(\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T+\mathbf{R}_0^{-1}),\ n+\rho_0]}$$

$$=|\boldsymbol{\Phi}|^{-(n+\rho_0+q+1)/2}\exp\left\{-\frac{1}{2}\ \mathrm{tr}[\boldsymbol{\Phi}^{-1}(\boxed{\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T}+\mathbf{R}_0^{-1})]\right\}$$

$$\boxed{\boldsymbol{\Omega}_2=(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n)}$$

Bayesian Estimation

Conditional Distributions

$$y_i = \mu + Ac_i + \Lambda\omega_i + \epsilon_i$$

$$u_i = (1^T, c_i^T, \omega_i^T)^T \quad \Lambda_y = (\mu, A, \Lambda)$$

$$y_i = \Lambda_y u_i + \epsilon_i$$

$$p(\Lambda, \Psi_\epsilon | Y, \Omega) \propto \boxed{[p(\Lambda, \Psi_\epsilon) p(Y | \Lambda, \Psi_\epsilon, \Omega)]}$$

Extended to handle the general situation with fixed known elements in $\boldsymbol{\Lambda}$

$$\boxed{\begin{aligned} [\psi_{\epsilon k}^{-1} | \mathbf{Y}, \boldsymbol{\Omega}] &\stackrel{D}{=} Gamma\,[n/2 + \alpha_{0\epsilon k},\ \beta_{\epsilon k}] \\ [\boldsymbol{\Lambda}_k^* | \mathbf{Y}, \boldsymbol{\Omega}, \psi_{\epsilon k}^{-1}] &\stackrel{D}{=} N\,[\mathbf{a}_k^*, \psi_{\epsilon k} \mathbf{A}_k^*] \end{aligned}}$$

Normal – (Inverted) gamma distribution

$$\beta_{\epsilon k} = \beta_{0\epsilon k} + 2^{-1}(\mathbf{Y}_k^T \mathbf{Y}_k - \mathbf{a}_k^T \mathbf{A}_k^{-1} \mathbf{a}_k + \boldsymbol{\Lambda}_{0k}^T \mathbf{H}_{0yk}^{-1} \boldsymbol{\Lambda}_{0k})$$

$$\mathbf{A}_k^* = (\mathbf{H}_{0yk}^{*-1} + \boldsymbol{\Omega}_k^* \boldsymbol{\Omega}_k^{*T})^{-1}$$

$$\mathbf{a}_k^* = \mathbf{A}_k^* (\mathbf{H}_{0yk}^{*-1} \boldsymbol{\Lambda}_{0k}^* + \boldsymbol{\Omega}_k^* \mathbf{Y}_k^*)$$

Bayesian Estimation

Conditional Distributions

$$\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i$$

$$\boldsymbol{v}_i = (\boldsymbol{d}_i^T, \boldsymbol{\eta}_i^T, \boldsymbol{\omega}_i^T)^T \quad \boldsymbol{\Lambda}_\omega = (\boldsymbol{B}, \boldsymbol{\Pi}, \boldsymbol{\Gamma})$$

$$\boldsymbol{\eta}_i = \boldsymbol{\Lambda}_\omega \boldsymbol{v}_i + \boldsymbol{\delta}_i$$

$$[\boldsymbol{\Lambda}_{\omega k} | \psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0\omega k}, \psi_{\delta k}\mathbf{H}_{0\omega k}]$$

$$\psi_{\delta k}^{-1} \stackrel{D}{=} \text{Gamma}[\alpha_{0\delta k}, \beta_{0\delta k}]$$

$$\boldsymbol{\Xi}_k^T = (\eta_{1k}^*, \ldots, \eta_{nk}^*) \qquad \eta_{ik}^* = \eta_{ik} - \sum_{j=1}^{r_2+q} \lambda_{\omega kj} v_{ij}(1 - l_{\omega kj})$$

V_k be the submatrix of $\boldsymbol{V}$ such that all the rows corresponding to $l_{\omega kj} = 0$ are deleted

$$[\psi_{\delta k}^{-1} | \boldsymbol{\Omega}] \stackrel{D}{=} \text{Gamma}[n/2 + \alpha_{0\delta k}, \beta_{\delta k}]$$

$$[\boldsymbol{\Lambda}_{\omega k} | \boldsymbol{\Omega}, \psi_{\delta k}^{-1}] \stackrel{D}{=} N[\mathbf{a}_{\omega k}, \psi_{\delta k}\mathbf{A}_{\omega k}]$$

$$\beta_{\delta k} = \beta_{0\delta k} + \frac{1}{2}(\boldsymbol{\Xi}_k^T\boldsymbol{\Xi}_k - \mathbf{a}_{\omega k}^T\mathbf{A}_{\omega k}^{-1}\mathbf{a}_{\omega k} + \boldsymbol{\Lambda}_{0\omega k}^T\mathbf{H}_{0\omega k}^{-1}\boldsymbol{\Lambda}_{0\omega k})$$

$$\mathbf{A}_{\omega k} = (\mathbf{H}_{0\omega k}^{-1} + \mathbf{V}_k\mathbf{V}_k^T)^{-1}$$

Normal – (Inverted) gamma distribution

$$\mathbf{a}_{\omega k} = \mathbf{A}_{\omega k}(\mathbf{H}_{0\omega k}^{-1}\boldsymbol{\Lambda}_{0\omega k} + \mathbf{V}_k\boldsymbol{\Xi}_k)$$

End of Chapter 3