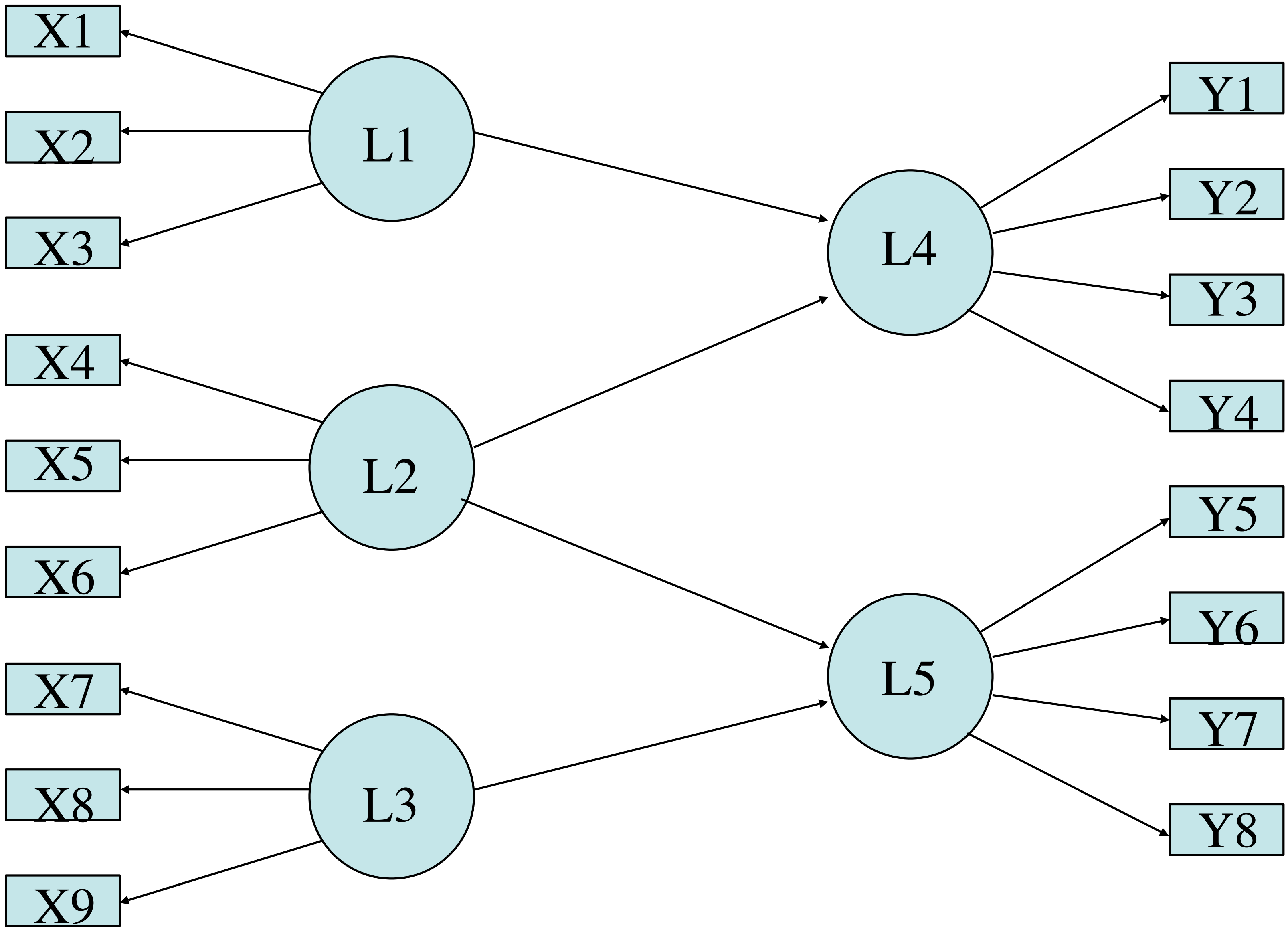
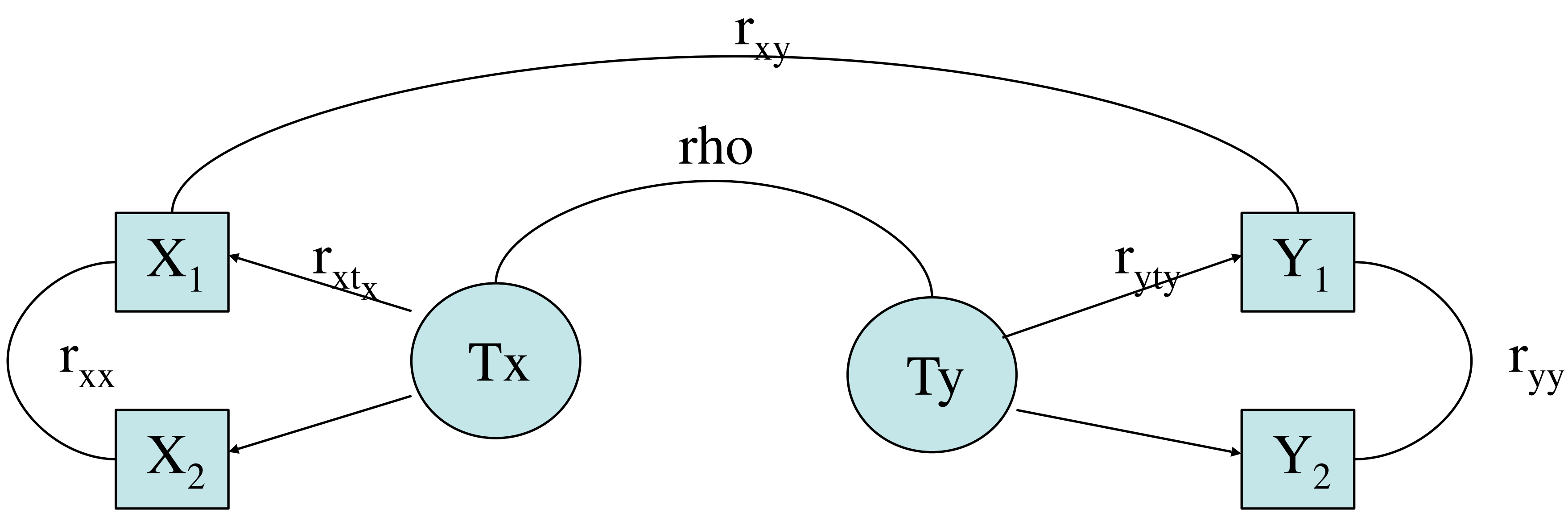
Psychometric Theory: A conceptual Syllabus



Reliability- Correction for attenuation

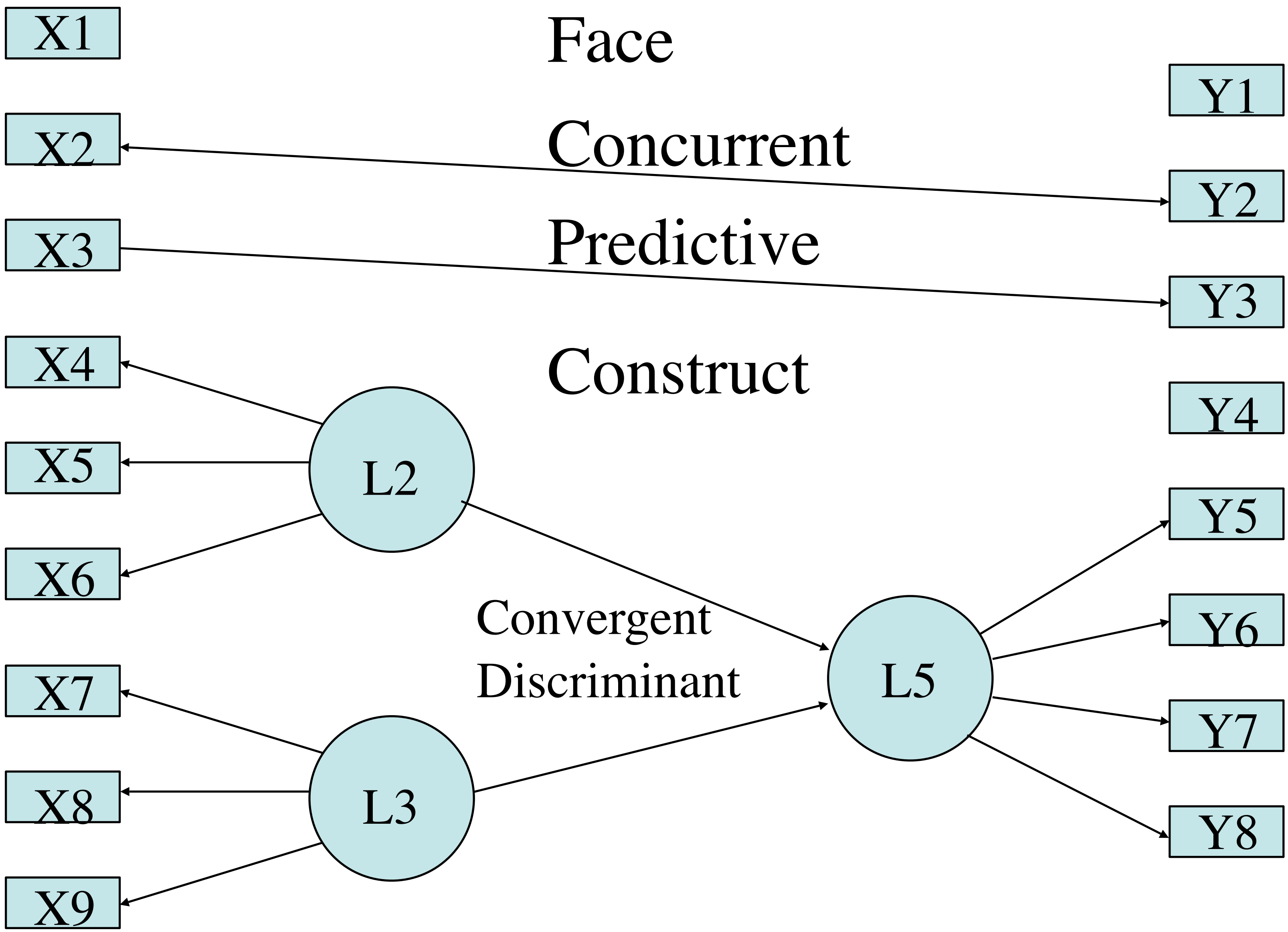


$r_{xt_x} = sqrt(r_{xx})$

$r_{yty} = sqrt(r_{yy})$

$Rho = rxy/sqrt(r_{xx} * r_{yy})$

Types of Validity: What are we measuring



Model Fitting

Structural Equation Models

Reliability + Validity

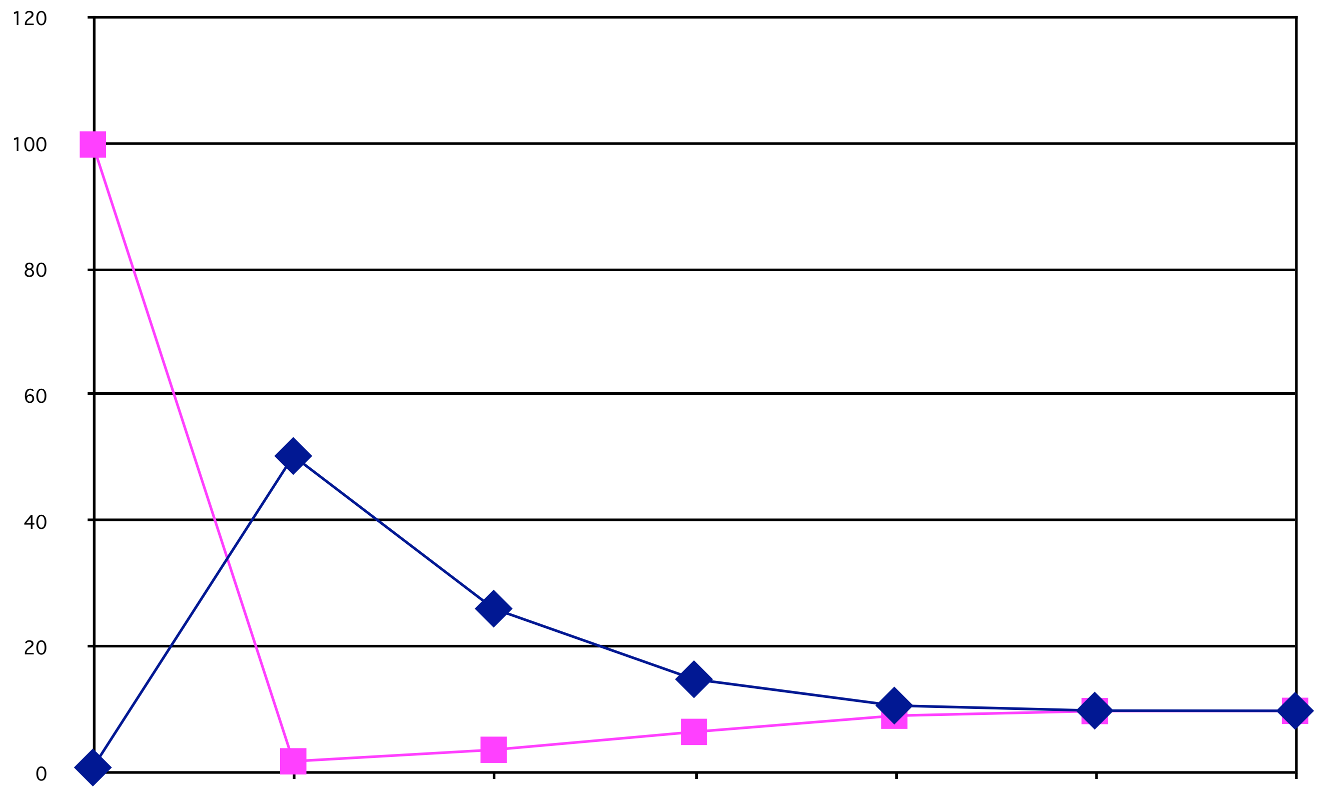
Basic concepts of iterative fit

- Most classical statistics (e.g. means, variances, regression slopes) may be found by algebraic solutions of closed form expressions
- More recent statistics are the results of iteratively fitting a model until some criterion is either minimized or maximized.

Simple example: the square root

Target	100		
Trial	**guess**	**fit**	**diff**
1	1.0	100.0	-99.0
2	50.5	2.0	48.5
3	26.2	3.8	22.4
4	15.0	6.7	8.4
5	10.8	9.2	1.6
6	10.0	10.0	0.1
7	10.0	10.0	0.0

Iteratively estimating the square root of 100



Applications: Factor Analysis

x1	1.00		
x2	0.70	1.00	
x3	0.60	0.58	1.00

Iterative Fit

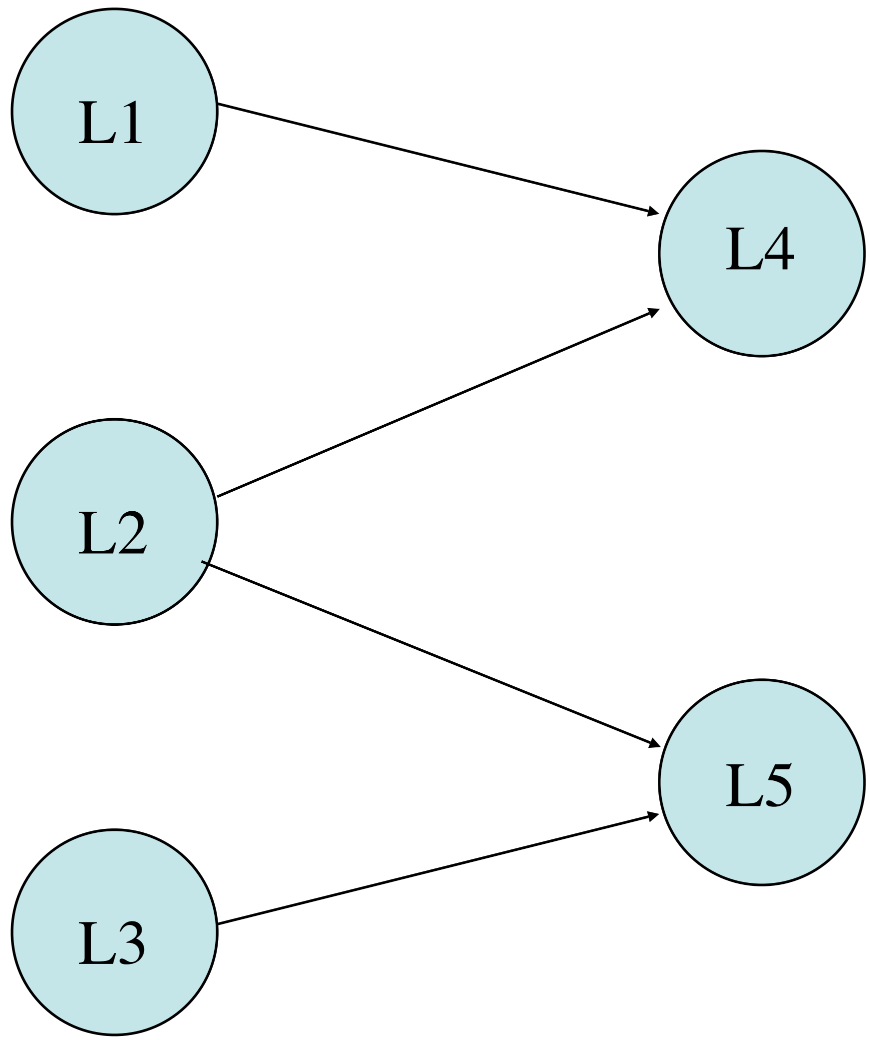
x1	x2	x3	fit
1.000	1.000	1.000	0.4264000
1.000	0.800	1.000	0.2180000
1.000	0.800	0.800	0.0536000
0.800	0.800	0.800	0.0088000
0.800	0.800	0.700	0.0056000
0.800	0.800	0.750	0.0040000
0.850	0.800	0.750	0.0022000
0.850	0.800	0.700	0.0008250
0.850	0.800	0.710	0.0005625
0.851	0.823	0.705	0.0000000

Fitted model

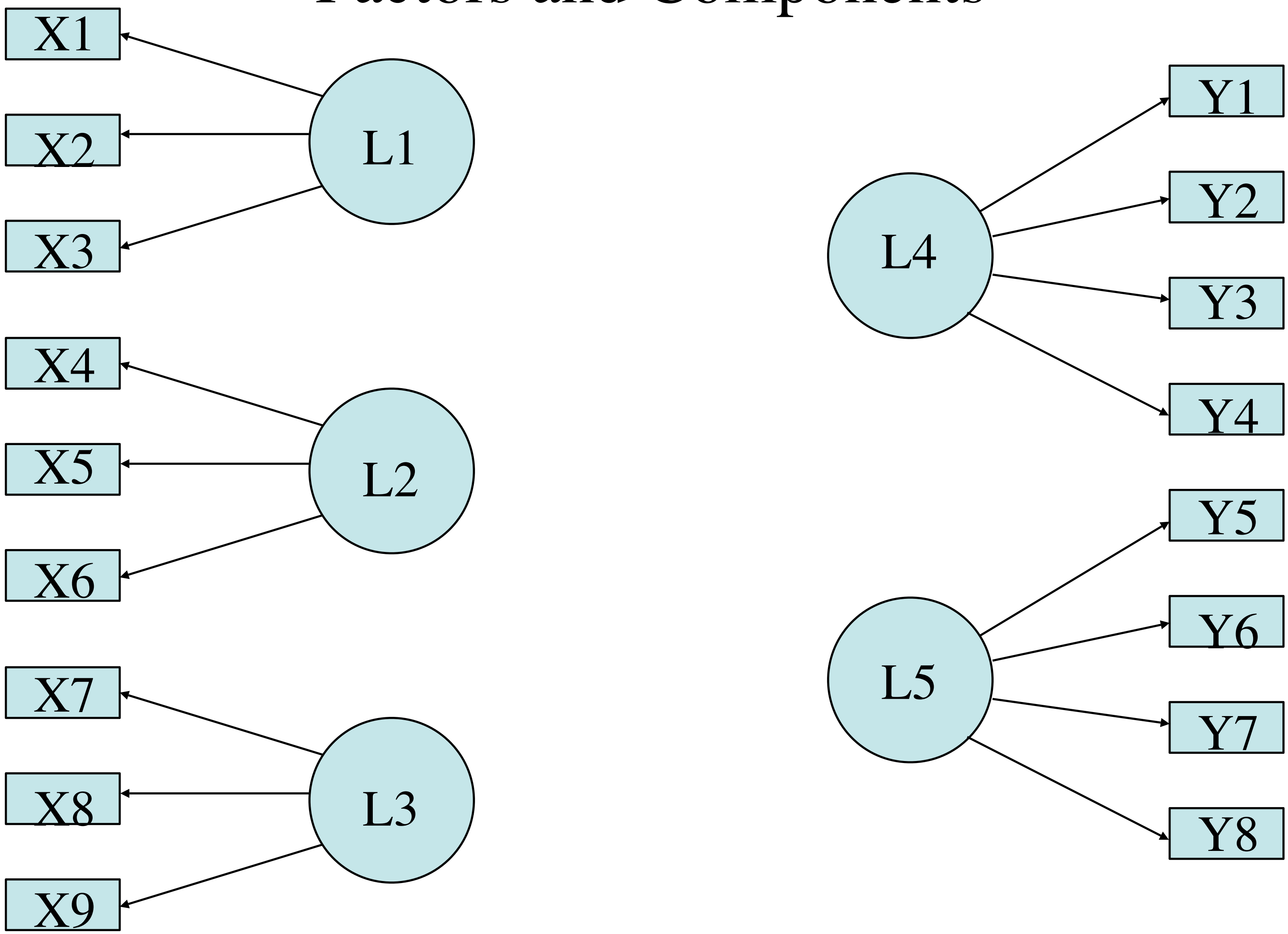
	F1
x1	0.851
x2	0.823
x3	0.705

0.72	0.70	0.60
0.70	0.68	0.58
0.60	0.58	0.50

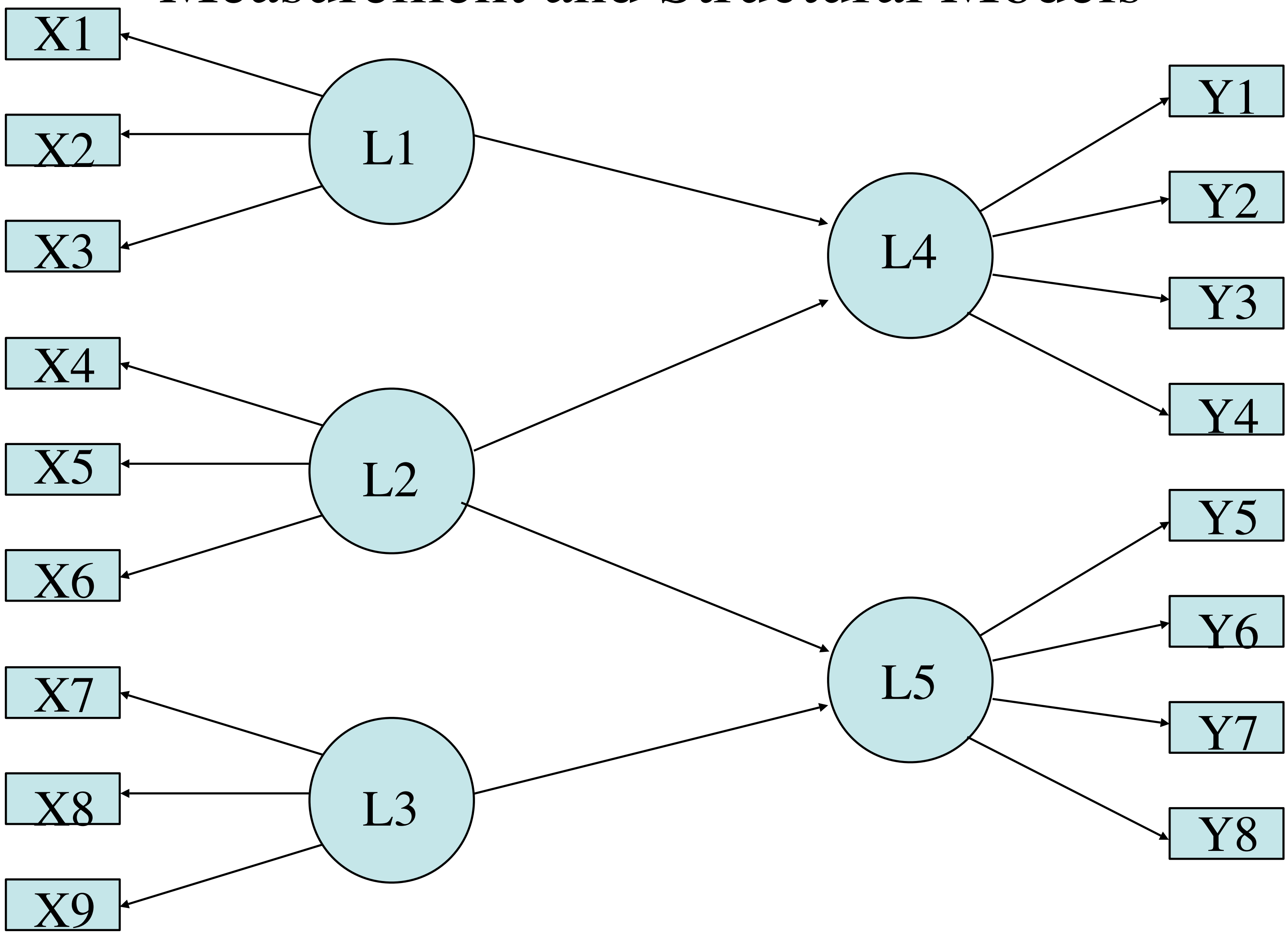
Theory as organization of constructs



Techniques of Data Reduction: Factors and Components



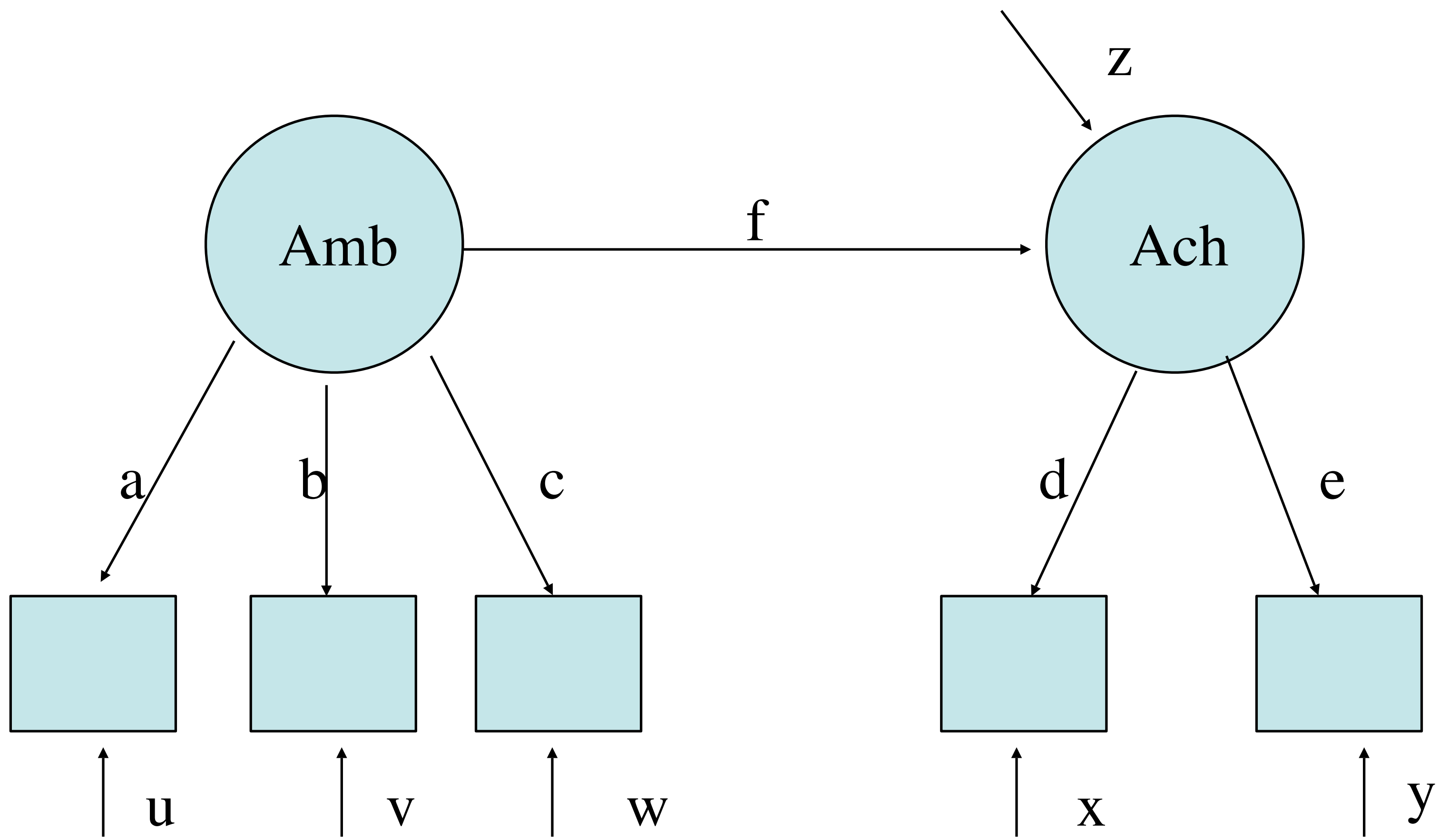
Structural Equation Modeling: Combining Measurement and Structural Models



SEM problem (Loehlin 2.5)

Ach1	Ach2	Amb1	Amb2	Amb3
1	0.6	0.3	0.2	0.2
0.6	1	0.2	0.3	0.1
0.3	0.2	1	0.7	0.6
0.2	0.3	0.7	1	0.5
0.2	0.1	0.6	0.5	1

Ambition and Achievement



R code for sem for Loehlin 2.5

First enter the correlation matrix

```
#Loehlin problem 2.5
 obs.var2.5 = c('Ach1',  'Ach2',  'Amb1',  'Amb2',  'Amb3')
 R.prob2.5 = matrix(c(
  1.00 ,  .60  , .30,  .20,   .20,
   .60,  1.00,   .20,   .30,   .10,
   .30,   .20,  1.00,   .70,   .60 ,
   .20,   .30,   .70,  1.00,   .50,
   .20,   .10,   .60,  .50,  1.00), ncol=5,byrow=TRUE)
```

R code for sem -Ram notation

```
model2.51=matrix(c(
        'Ambit ->  Amb1',       'a', NA,
        'Ambit -> Amb2' ,       'b', NA,
        'Ambit -> Amb3' ,       'c', NA,
        'Achieve -> Ach1',      'd', NA,
        'Achieve -> Ach2',      'e', NA,
        'Ambit -> Achieve',    'f', NA,
        'Amb1 <-> Amb1' ,       'u', NA,
        'Amb2 <-> Amb2' ,       'v', NA,
        'Amb3 <-> Amb3' ,       'w', NA,
        'Ach1 <-> Ach1' ,       'x', NA,
        'Ach2 <-> Ach2' ,       'y', NA,
        'Achieve <-> Achieve',  NA, 1,
        'Ambit <-> Ambit',      NA, 1),
        ncol=3, byrow=TRUE)
```

Run the R code and show results

```
sem2.5= sem(model2.5,R.prob2.5,60, obs.var2.5)
 summary(sem2.5,digits=3)

Model Chisquare =  9.74   Df =  4 Pr(>Chisq) = 0.0450
Goodness-of-fit index =  0.964
Adjusted goodness-of-fit index =  0.865
RMSEA index =  0.120   90 % CI: (0.0164, 0.219)
BIC =  -15.1

Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-5.77e-01 -3.78e-02 -2.04e-06  4.85e-03  3.87e-05  1.13e+00
```

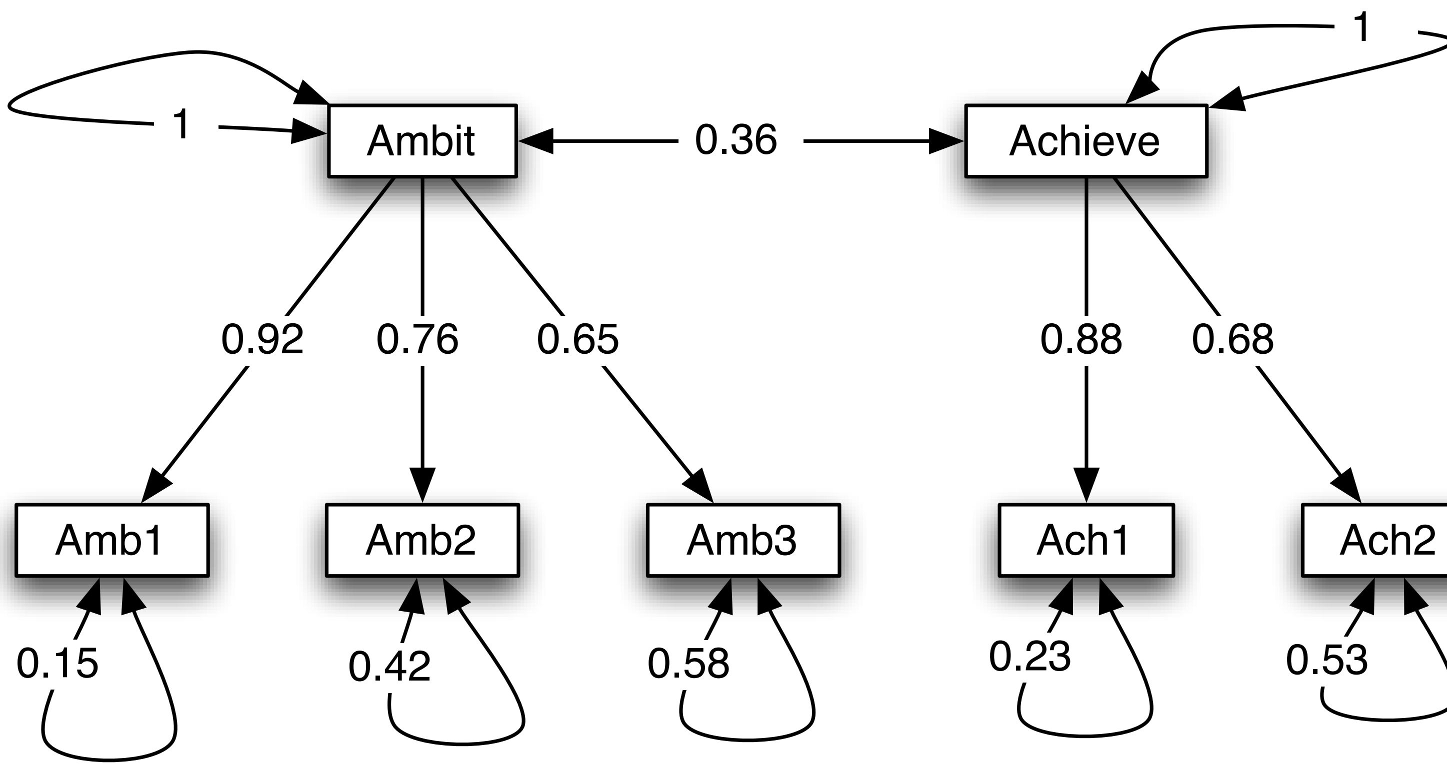
What are the parameters?

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
a	0.920	0.0924	9.966	0.00e+00	Amb1 <--- Ambit
b	0.761	0.0955	7.974	1.55e-15	Amb2 <--- Ambit
c	0.652	0.0965	6.753	1.45e-11	Amb3 <--- Ambit
d	0.879	0.1762	4.986	6.16e-07	Ach1 <--- Achieve
e	0.683	0.1509	4.525	6.03e-06	Ach2 <--- Achieve
f	0.356	0.1138	3.127	1.76e-03	Achieve <--> Ambit
u	0.153	0.0982	1.557	1.20e-01	Amb1 <--> Amb1
v	0.420	0.0898	4.679	2.88e-06	Amb2 <--> Amb2
w	0.575	0.0949	6.061	1.35e-09	Amb3 <--> Amb3
x	0.228	0.2791	0.816	4.15e-01	Ach1 <--> Ach1
y	0.534	0.1837	2.905	3.67e-03	Ach2 <--> Ach2

Iterations = 26

A simple sem



1
1
Ambit
0.36
Achieve
0.92
0.76
0.65
0.88
0.68
Amb1
Amb2
Amb3
Ach1
Ach2
0.15
0.42
0.58
0.23
0.53

Problems in interpretation

- Model fit does not imply best model
- Consider alternative models
  - Reverse the arrows of causality, nothing happens
  - Range of alternative models
  - Nested models can be compared
  - Non-nested alternative models might be better

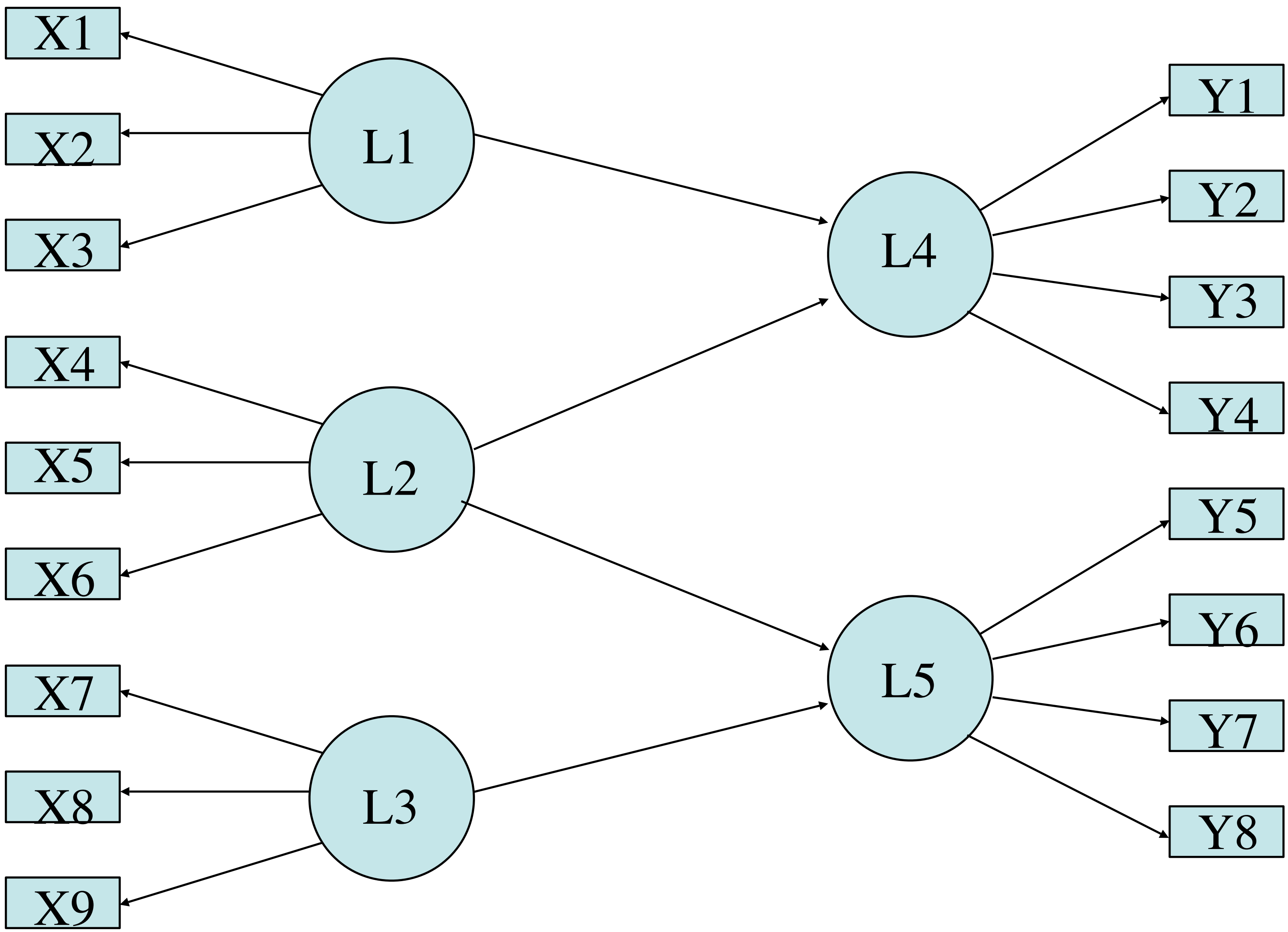
SEM points to consider

- Goodness of fit statistics
  - Statistical indices of size of residuals as compared to null model are sensitive to sample size
  - Comparisons of nested models
- Fits get better with more parameters -> development of df corrected fits
- Inspect residuals to see what is not being fit
- Avoid temptation to ‘fix’ model based upon results, or, at least be less confident in meaning of good fit

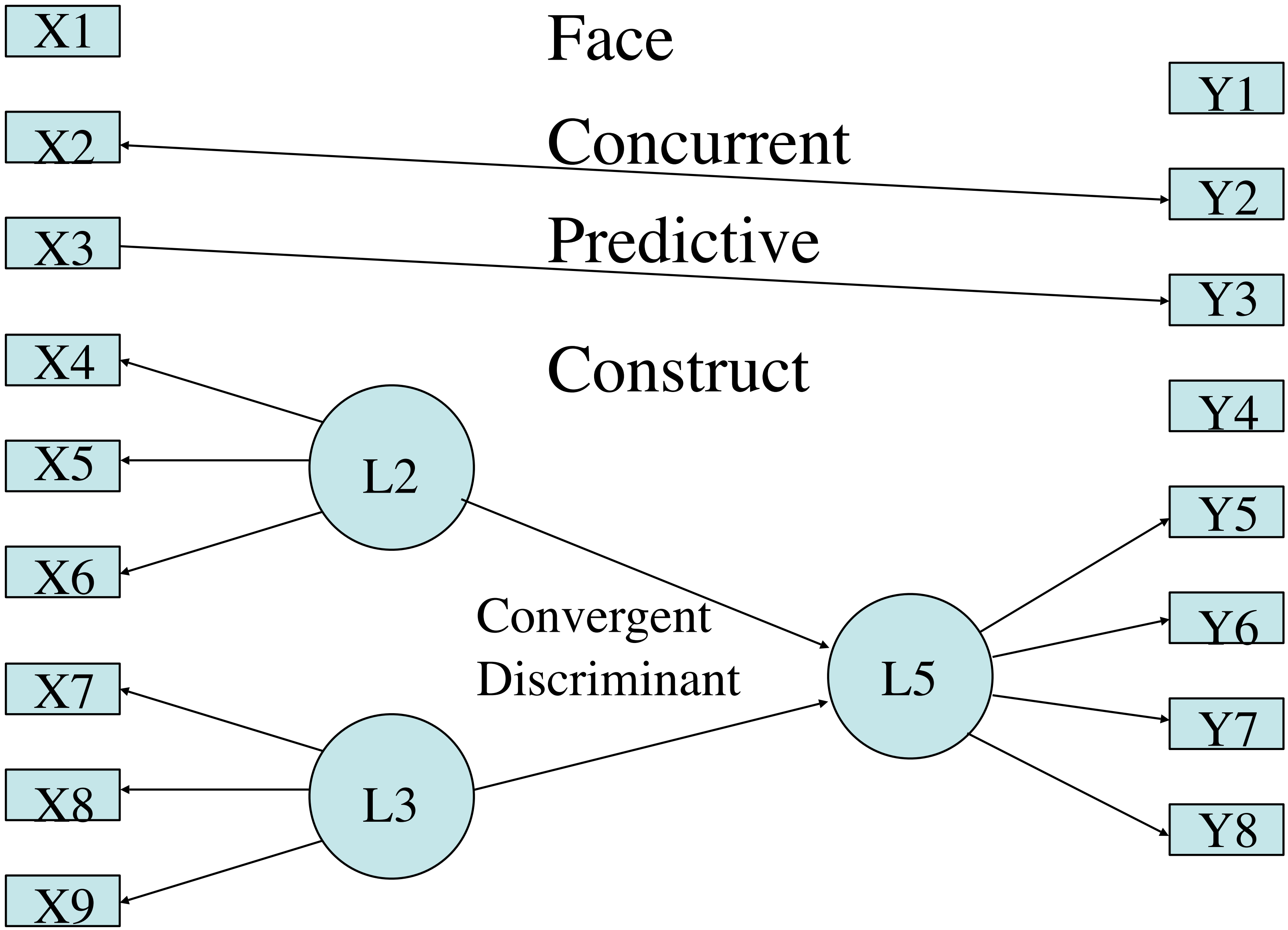
Applications of SEM techniques

- Confirmatory factor analysis
  - Does a particular structure fit the data
- Growth models (growth curve analysis)
- Multiple groups
  - Is the factor structure the same across groups
  - Is the factor structure the same across time

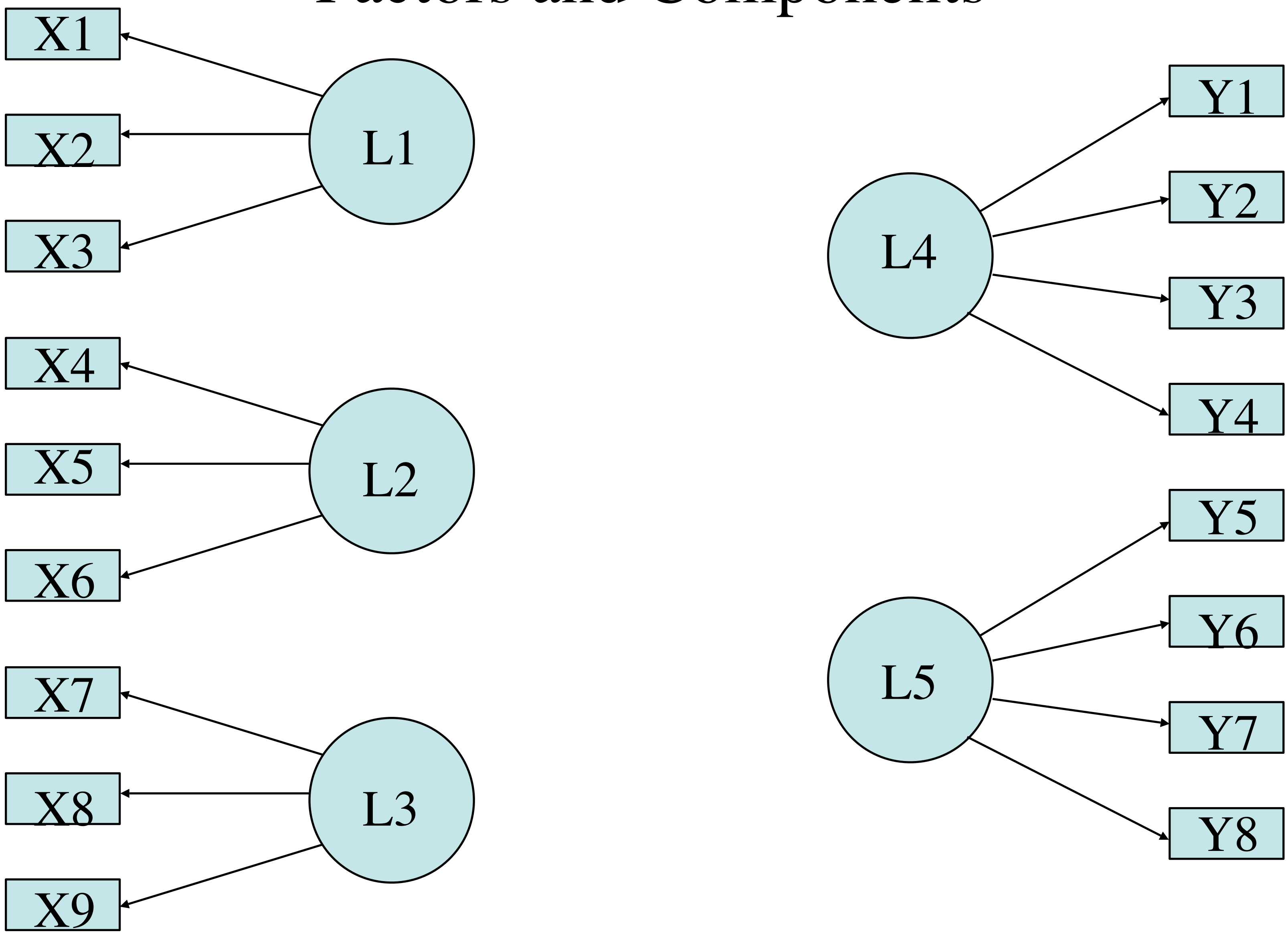
Psychometric Theory: A conceptual Syllabus



Types of Validity: What are we measuring



Techniques of Data Reduction: Factors and Components



Methods of Homogeneous Keying

Factor analysis, principal components analysis, and cluster analysis

Methods of Homogeneous Keying

- Factor Analysis
- Principal Components Analysis
- Cluster Analysis

Factor Analysis
Consider the following r matrix

	X_1	X_2	X_3	X_4	X_5
X_1	1.00				
X_2	0.72	1.00			
X_3	0.63	0.56	1.00		
X_4	0.54	0.48	0.42	1.00	
X_5	0.45	0.40	0.35	0.30	1.00

Factors as causes; variables as indicators of factors

- Can we recreate the correlation matrix R (of rank n) with a matrix F of rank 1 + a diagonal matrix of uniqueness U^2
- $R \approx FF' + U^2$
- Residual Matrix R* = R - (FF'+ U2)
- Try to minimize the residual

Create an example factor matrix

```
x <- c(.9,.8,.7,.6,.5,.4)
 r <- x %*% t(x)
 colnames(r) <- rownames(r) <- paste("V",1:6,sep="")
 diag(r) <- 1
 r
f1 <- factor.pa(r,1,TRUE,max.iter=1)
```

Observed Correlation matrix

	V1	V2	V3	V4	V5	V6
V1	1.00	0.72	0.63	0.54	0.45	0.36
V2	0.72	1.00	0.56	0.48	0.40	0.32
V3	0.63	0.56	1.00	0.42	0.35	0.28
V4	0.54	0.48	0.42	1.00	0.30	0.24
V5	0.45	0.40	0.35	0.30	1.00	0.20
V6	0.36	0.32	0.28	0.24	0.20	1.00

Find the characteristic roots

- Eigen vectors as characteristic roots of a matrix. (The vectors unchanged by the matrix.)
- Eigen values as a scaling of the roots
- Given a matrix R, the eigen vectors solve the equation
  - xR = vx   or RX = XD where X is a matrix of eigen vectors and D is a diagonal matrix of the eigen values
  - Thus since X X' = I, then R = XDX'
  - xR - vXI = 0 <=> x(R-vI) = 0 and we solve for x

Eigen vectors and eigen values

```
$values
                    ev1  ev2  ev3  ev4  ev5  ev6
                   3.16 0.82 0.72 0.59 0.44 0.26
$vectors
            ev1    ev2    ev3   ev4    ev5    ev6
     ev1 -0.50  0.061  0.092  0.14  0.238  0.816
     ev2 -0.47  0.074  0.121  0.21  0.657 -0.533
     ev3 -0.43  0.096  0.182  0.53 -0.675 -0.184
     ev4 -0.39  0.142  0.414 -0.78 -0.201 -0.104
     ev5 -0.34  0.299 -0.860 -0.20 -0.108 -0.067
     ev6 -0.28 -0.934 -0.178 -0.10 -0.067 -0.045
```

Principal components

I. Merely a transformation of original matrix to show the eigen vectors and eigen values

II. The vectors provide the basis space of the original matrix

III. The lengths of the vectors (the eigen values) reflect the importance of the vector

IV. If done on a rectangular (data) matrix, known as singular value decomposition, characteristic roots, characteristic vectors

Principal components as rescaled eigen vectors

- If V = the matrix of eigen vectors and v = vector of eigen values and D = diagonal with elements(v)
  - V <- eigen(r)$vectors
  - D <- diag(e$values)
- then R = V D V' R <- V %*% D %*% t(V)
- For symmetric matrices I = VV' = V'V
- if we let C = V * sqrt(D) C <- V %*% sqrt(D)
- R = CC' R <- C %*% t(C)

Principal Components

```
e <- eigen(r)
c <- e$vectors %*% diag(sqrt(e$values))
 rownames(c) <- paste('V',1:6,sep='')
 colnames(c) <- paste("C", seq(1:6),sep="")
 round(c,2)
```

	C1	C2	C3	C4	C5	C6
V1	-0.88	0.06	0.08	0.11	0.16	0.42
V2	-0.83	0.07	0.10	0.16	0.44	-0.27
V3	-0.77	0.09	0.15	0.41	-0.45	-0.09
V4	-0.69	0.13	0.35	-0.60	-0.13	-0.05
V5	-0.60	0.27	-0.73	-0.15	-0.07	-0.03
V6	-0.50	-0.85	-0.15	-0.08	-0.04	-0.02

R = CC'

round(c %*% t(c),2)

	V1	V2	V3	V4	V5	V6
V1	1.00	0.72	0.63	0.54	0.45	0.36
V2	0.72	1.00	0.56	0.48	0.40	0.32
V3	0.63	0.56	1.00	0.42	0.35	0.28
V4	0.54	0.48	0.42	1.00	0.30	0.24
V5	0.45	0.40	0.35	0.30	1.00	0.20
V6	0.36	0.32	0.28	0.24	0.20	1.00

Take just the first n principal components

I. R = CC' for all components

II. R ≈ CC' for first n components

III. Evaluate R* = R - CC'

IV. pc <- principal(r,1,residuals=TRUE)

V. notice that the sign of the column may be reversed

$loadings

	PC1
V1	0.88
V2	0.83
V3	0.77
V4	0.69
V5	0.60
V6	0.50

R for 1 component

```
c1 <- c[,1]
> round(c1 %*% t(c1),2)
```

	V1	V2	V3	V4	V5	V6
[1,]	0.78	0.74	0.68	0.61	0.53	0.44
[2,]	0.74	0.69	0.64	0.58	0.50	0.42
[3,]	0.68	0.64	0.59	0.53	0.47	0.39
[4,]	0.61	0.58	0.53	0.48	0.42	0.35
[5,]	0.53	0.50	0.47	0.42	0.37	0.30
[6,]	0.44	0.42	0.39	0.35	0.30	0.25

R*= R - CC' for 1 component

	V1	V2	V3	V4	V5	V6
V1	0.22	-0.02	-0.05	-0.07	-0.08	-0.08
V2	-0.02	0.31	-0.08	-0.10	-0.10	-0.10
V3	-0.05	-0.08	0.41	-0.11	-0.12	-0.11
V4	-0.07	-0.10	-0.11	0.52	-0.12	-0.11
V5	-0.08	-0.10	-0.12	-0.12	0.63	-0.10
V6	-0.08	-0.10	-0.11	-0.11	-0.10	0.75

Fit = $1 - R^{*2}/R^2$ = .85

R for 2 components

```
>   c2 <- c[,1:2]
> round(c2 %*% t(c2),2)
     V1   V2   V3   V4   V5   V6
V1 0.78 0.74 0.68 0.62 0.55 0.40
V2 0.74 0.70 0.65 0.59 0.52 0.36
V3 0.68 0.65 0.60 0.55 0.49 0.31
V4 0.62 0.59 0.55 0.50 0.45 0.24
V5 0.55 0.52 0.49 0.45 0.44 0.07
V6 0.40 0.36 0.31 0.24 0.07 0.97
```

R- CC' for 2 components

> round(r-c2 %*% t(c2),2)

```
        V1    V2    V3    V4    V5    V6
V1  0.22 -0.02 -0.05 -0.08 -0.10 -0.04
V2 -0.02  0.30 -0.09 -0.11 -0.12 -0.04
V3 -0.05 -0.09  0.40 -0.13 -0.14 -0.03
V4 -0.08 -0.11 -0.13  0.50 -0.15  0.00
V5 -0.10 -0.12 -0.14 -0.15  0.56  0.13
V6 -0.04 -0.04 -0.03  0.00  0.13  0.03
```

```
$fit
[1] 0.9
```

Factor Analysis as eigen value decomposition of reduced matrix

I. Estimate the R as FF' + U^2

A. U^2 is the amount not accounted for by the factor and is a diagonal matrix

II. Estimate $(R-U^2) = FF'$ using eigen value model

III. But what is the best value of U^2

IV. Solve for iteratively

The PC model - first pass

```
> round(pc1$loadings %*% t
(pc1$loadings),2)

      V1   V2   V3   V4   V5   V6
V1 0.77 0.73 0.68 0.61 0.53 0.44
V2 0.73 0.69 0.64 0.57 0.50 0.42
V3 0.68 0.64 0.59 0.53 0.46 0.38
V4 0.61 0.57 0.53 0.48 0.41 0.34
V5 0.53 0.50 0.46 0.41 0.36 0.30
V6 0.44 0.42 0.38 0.34 0.30 0.25
```

First iteration = PC

$loadings

	PA1
V1	0.88
V2	0.83
V3	0.77
V4	0.69
V5	0.60
V6	0.50

$residual

	V1	V2	V3	V4	V5	V6
V1	0.22	-0.02	-0.05	-0.07	-0.08	-0.08
V2	-0.02	0.31	-0.08	-0.10	-0.10	-0.10
V3	-0.05	-0.08	0.41	-0.11	-0.12	-0.11
V4	-0.07	-0.10	-0.11	0.52	-0.12	-0.11
V5	-0.08	-0.10	-0.12	-0.12	0.63	-0.10
V6	-0.08	-0.10	-0.11	-0.11	-0.10	0.75

$communality
[1] 3.16

$fit
[1] 0.85

$fitoff
[1] 0.95

2nd iteration

$loadings

	PA1
V1	0.88
V2	0.81
V3	0.72
V4	0.62
V5	0.52
V6	0.41

$residual

	V1	V2	V3	V4	V5	V6
V1	0.22	0.01	0.00	-0.01	0.00	0.00
V2	0.01	0.35	-0.02	-0.02	-0.02	-0.01
V3	0.00	-0.02	0.48	-0.03	-0.02	-0.01
V4	-0.01	-0.02	-0.03	0.61	-0.02	-0.01
V5	0.00	-0.02	-0.02	-0.02	0.73	-0.01
V6	0.00	-0.01	-0.01	-0.01	-0.01	0.83

$communality
[1] 3.16 2.77

$fit
[1] 0.83

$fitoff
[1] 1

3rd iteration

$loadings

	PA1
V1	0.89
V2	0.80
V3	0.71
V4	0.61
V5	0.50
V6	0.40

$residual

	V1	V2	V3	V4	V5	V6
V1	0.21	0.01	0.00	0.00	0.00	0.00
V2	0.01	0.36	-0.01	-0.01	0.00	0.00
V3	0.00	-0.01	0.50	-0.01	-0.01	0.00
V4	0.00	-0.01	-0.01	0.63	0.00	0.00
V5	0.00	0.00	-0.01	0.00	0.75	0.00
V6	0.00	0.00	0.00	0.00	0.00	0.84

$communality

[1] 3.16 2.77 2.71

$fit

[1] 0.82

$fitoff

[1] 1

4th iteration

$loadings

	PA1
V1	0.89
V2	0.80
V3	0.70
V4	0.60
V5	0.50
V6	0.40

```
$residual
    V1   V2   V3   V4   V5   V6
V1 0.2 0.00 0.00 0.00 0.00 0.00
V2 0.0 0.36 0.00 0.00 0.00 0.00
V3 0.0 0.00 0.51 0.00 0.00 0.00
V4 0.0 0.00 0.00 0.64 0.00 0.00
V5 0.0 0.00 0.00 0.00 0.75 0.00
V6 0.0 0.00 0.00 0.00 0.00 0.84
```

$communality
[1] 3.16 2.77 2.71 2.71

$fit
[1] 0.82

$fitoff
[1] 1

5th iteration

$residual

	V1	V2	V3	V4	V5	V6
V1	0.2	0.00	0.00	0.00	0.00	0.00
V2	0.0	0.36	0.00	0.00	0.00	0.00
V3	0.0	0.00	0.51	0.00	0.00	0.00
V4	0.0	0.00	0.00	0.64	0.00	0.00
V5	0.0	0.00	0.00	0.00	0.75	0.00
V6	0.0	0.00	0.00	0.00	0.00	0.84

$loadings

	PA1
V1	0.9
V2	0.8
V3	0.7
V4	0.6
V5	0.5
V6	0.4

$communality

[1] 3.16 2.77 2.71 2.71 2.71

$fit

[1] 0.82

$fitoff

[1] 1

Minimum Residual as Alternative to Principal Axis

- Find F such that if R* = R - FF', then
- 1 R*R*'1' - tr(R*) is a minimum. (minimize sum of squared off diagonal residuals)
- Done using the optim function
- Solutions are more similar to maximum likelihood than is principal axis.

Consider another matrix

	V1	V2	V3	V4	V5	V6
V1	1.00	0.72	0.63	0.00	0.00	0.00
V2	0.72	1.00	0.56	0.00	0.00	0.00
V3	0.63	0.56	1.00	0.00	0.00	0.00
V4	0.00	0.00	0.00	1.00	0.56	0.48
V5	0.00	0.00	0.00	0.56	1.00	0.42
V6	0.00	0.00	0.00	0.48	0.42	1.00

R = FF' (one factor)

round(pa1 %*% t(pa1),2)

$loadings

	PA1
V1	0.89
V2	0.81
V3	0.70
V4	0.00
V5	0.00
V6	0.00

	V1	V2	V3	V4	V5	V6
V1	0.79	0.72	0.62	0	0	0
V2	0.72	0.66	0.57	0	0	0
V3	0.62	0.57	0.49	0	0	0
V4	0.00	0.00	0.00	0	0	0
V5	0.00	0.00	0.00	0	0	0
V6	0.00	0.00	0.00	0	0	0

Factor with 1 factor

$loadings

	PA1
V1	0.89
V2	0.81
V3	0.70
V4	0.00
V5	0.00
V6	0.00

$residual

	V1	V2	V3	V4	V5	V6
V1	0.21	0.00	0.00	0.00	0.00	0.00
V2	0.00	0.35	-0.01	0.00	0.00	0.00
V3	0.00	-0.01	0.50	0.00	0.00	0.00
V4	0.00	0.00	0.00	1.00	0.56	0.48
V5	0.00	0.00	0.00	0.56	1.00	0.42
V6	0.00	0.00	0.00	0.48	0.42	1.00

$communality
[1] 2.28 2.04 1.96 1.94 1.94 1.94

$fit
[1] 0.51

$fitoff
[1] 0.63

R ≈ FF' (2 factors)

```
 $loadings
     PA1  PA2
V1 0.89 0.00
V2 0.80 0.00
V3 0.70 0.00
V4 0.00 0.79
V5 0.00 0.70
V6 0.00 0.60
```

```
r round(pa2 %*% t(pa2),2)
     V1   V2   V3   V4   V5   V6
V1 0.79 0.71 0.62 0.00 0.00 0.00
V2 0.71 0.64 0.56 0.00 0.00 0.00
V3 0.62 0.56 0.49 0.00 0.00 0.00
V4 0.00 0.00 0.00 0.62 0.55 0.47
V5 0.00 0.00 0.00 0.55 0.49 0.42
V6 0.00 0.00 0.00 0.47 0.42 0.36
```

Try 2 factors!

$loadings

	PA1	PA2
V1	0.89	0.00
V2	0.80	0.00
V3	0.70	0.00
V4	0.00	0.79
V5	0.00	0.70
V6	0.00	0.60

$residual

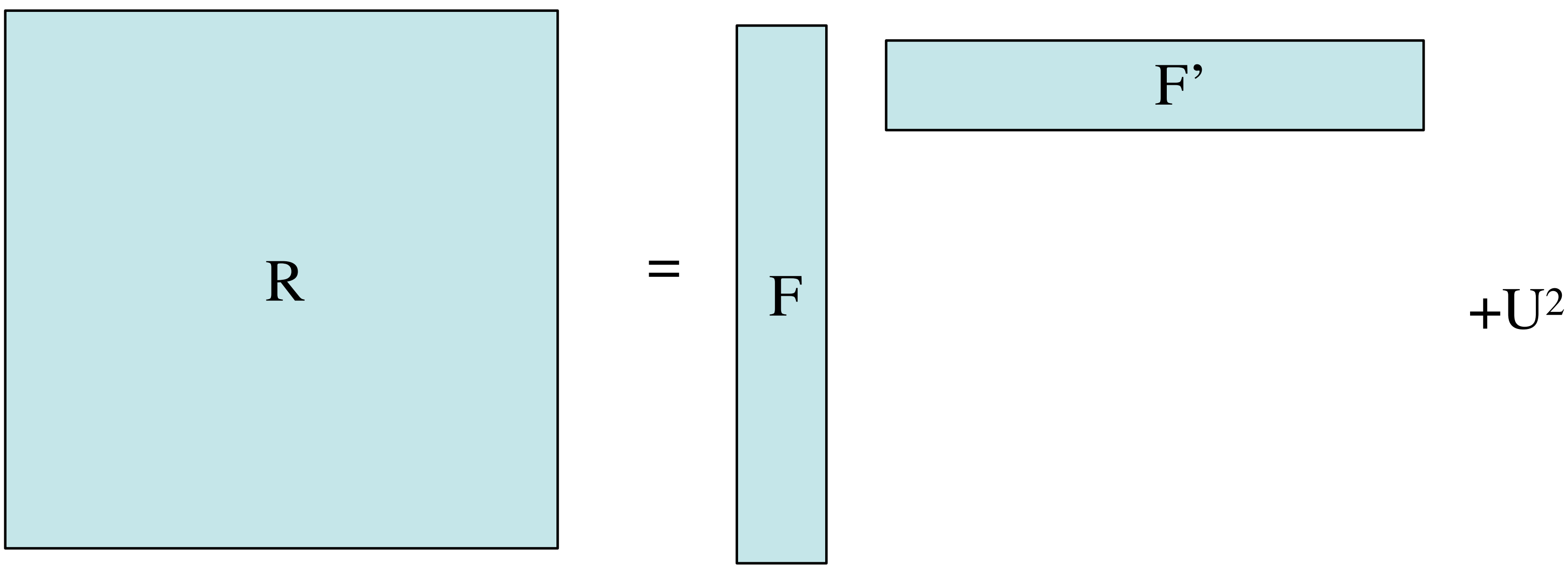
	V1	V2	V3	V4	V5	V6
V1	0.2	0.00	0.00	0.00	0.0	0.00
V2	0.0	0.35	-0.01	0.00	0.0	0.00
V3	0.0	-0.01	0.51	0.00	0.0	0.00
V4	0.0	0.00	0.00	0.37	0.0	0.00
V5	0.0	0.00	0.00	0.00	0.5	0.00
V6	0.0	0.00	0.00	0.00	0.0	0.64

```
        $communality                       $fit   $fitoff
4.25 3.68 3.50 3.44 3.43 3.42 3.42[1] 0.88  [1] 1
```

Factors as causes; variables as indicators of factors

- Can we recreate the correlation matrix R (of rank n) with a matrix F of rank 1 + a diagonal matrix of uniqueness U^2
- $R \approx FF' + U^2$
- Residual Matrix R* = R - (FF'+ U2)
- Try to minimize the residual

Factor analysis



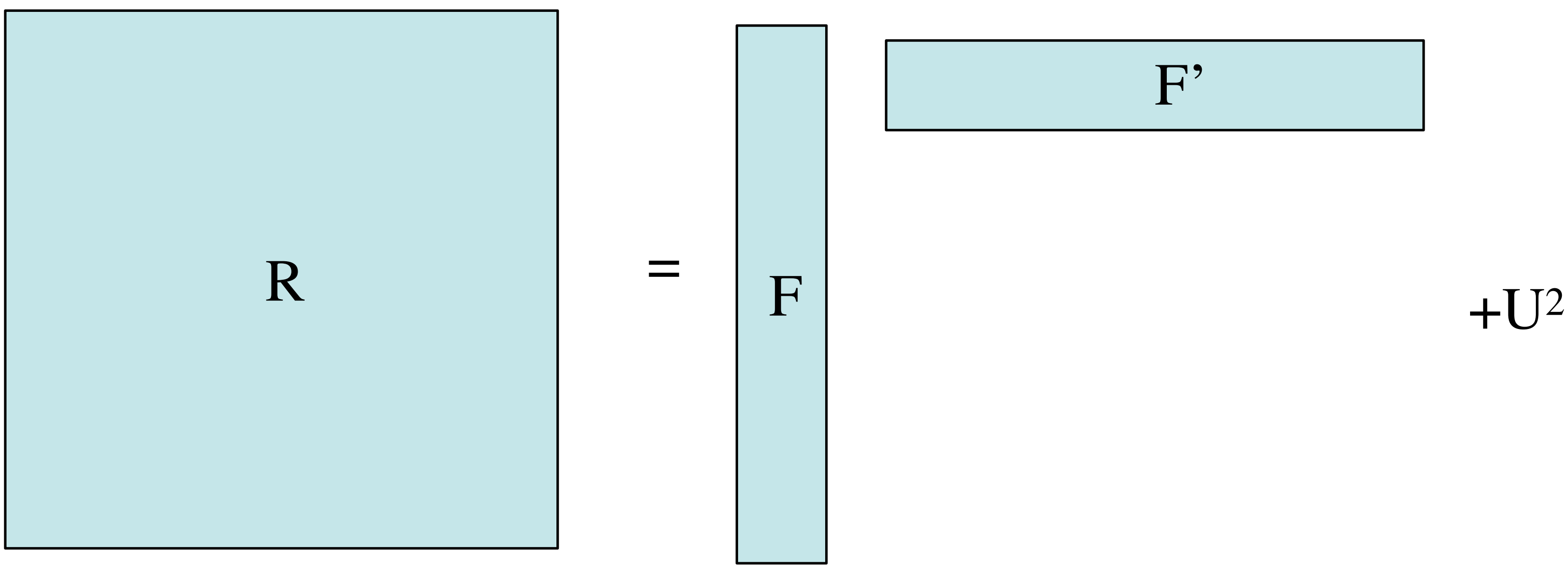
Factor Analysis: the model

- $R \approx FF' + U^2$
- Residual Matrix $R^* = R - (FF' + U2)$
- Try to minimize the residual
- Variables are linear composites of unknown (latent) factors.
- Covariance structures of observables in terms of covariance of unobservables

1 Factor solution $R \approx FF' + U^2$

	Correlation with (loading on) Factor 1
X_1	0.9
X_2	0.8
X_3	0.7
X_4	0.6
X_5	0.5

Factor analysis



Factor Analysis of more than 1 factor

	X_1	X_2	X_3	X_4	X_5
X_1	1.00				
X_2	0.72	1.00			
X_3	0.00	0.00	1.00		
X_4	0.00	0.00	0.42	1.00	
X_5	0.00	0.00	0.35	0.30	1.00

Factor Analysis: the model

I. $R \approx FF' + U^2$

II. Residual Matrix $R^* = R - (FF' + U2)$

III. Try to minimize the residual

IV. Variables are linear composites of unknown (latent) factors.

V. Covariance structures of observables in terms of covariance of unobservables

Factor analysis of mood

- What is the structure of affect?
- Limited number of variables taken from a larger set

Structure of mood - how not to display data

	AFRAID	AT_EASE	CALM	ENERGETI	HAPPY	LIVELY	SLEEPY	TENSE	TIRED
AFRAID	1.000								
AT_EASE	-0.209	1.000							
CALM	-0.157	0.586	1.000						
ENERGETI	0.019	0.230	0.056	1.000					
HAPPY	-0.070	0.452	0.294	0.595	1.000				
LIVELY	0.018	0.255	0.073	0.778	0.609	1.000			
SLEEPY	0.087	-0.112	0.031	-0.457	-0.264	-0.405	1.000		
TENSE	0.397	-0.337	-0.332	0.088	-0.103	0.084	0.044	1.000	
TIRED	0.082	-0.141	0.012	-0.484	-0.297	-0.439	0.808	0.044	1.000
UNHAPPY	0.350	-0.283	-0.187	-0.185	-0.314	-0.187	0.202	0.360	0.235

Structure of mood: "Alabama need not come first"

	AFRAID	AT_EASE	CALM	ENERGETI	HAPPY	LIVELY	SLEEPY	TENSE	TIRED
AFRAID	1.0								
AT_EASE	-0.2	1.0							
CALM	-0.2	0.6	1.0						
ENERGETI	0.0	0.2	0.1	1.0					
HAPPY	-0.1	0.5	0.3	0.6	1.0				
LIVELY	0.0	0.3	0.1	0.8	0.6	1.0			
SLEEPY	0.1	-0.1	0.0	-0.5	-0.3	-0.4	1.0		
TENSE	0.4	-0.3	-0.3	0.1	-0.1	0.1	0.0	1.0	
TIRED	0.1	-0.1	0.0	-0.5	-0.3	-0.4	0.8	0.0	1.0
UNHAPPY	0.3	-0.3	-0.2	-0.2	-0.3	-0.2	0.2	0.4	0.2

Structure of mood data

	ENERGETI	LIVELY	TIRED	SLEEPY	AFRAID	TENSE	AT_EASE	CALM	HAPPY
ENERGETI	1								
LIVELY	0.8	1							
TIRED	-0.5	-0.4	1						
SLEEPY	-0.5	-0.4	0.8	1					
AFRAID	0	0	0.1	0.1	1				
TENSE	0.1	0.1	0	0	0.4	1			
AT_EASE	0.2	0.3	-0.1	-0.1	-0.2	-0.3	1		
CALM	0.1	0.1	0	0	-0.2	-0.3	0.6	1	
HAPPY	0.6	0.6	-0.3	-0.3	-0.1	-0.1	0.5	0.3	1
UNHAPPY	-0.2	-0.2	0.2	0.2	0.3	0.4	-0.3	-0.2	-0.3

NUMBER OF OBSERVATIONS: 3748

Correlation of mood data possible structure

	ENERGETI	LIVELY	TIRED	SLEEPY	AFRAID	TENSE	AT_EASE	CALM	HAPPY
ENERGETI	1.0								
LIVELY	0.8	1.0							
TIRED	-0.5	-0.4	1.0						
SLEEPY	-0.5	-0.4	0.8	1.0					
AFRAID	0.0	0.0	0.1	0.1	1.0				
TENSE	0.1	0.1	0.0	0.0	0.4	1.0			
AT_EASE	0.2	0.3	-0.1	-0.1	-0.2	-0.3	1.0		
CALM	0.1	0.1	0.0	0.0	-0.2	-0.3	0.6	1.0	
HAPPY	0.6	0.6	-0.3	-0.3	-0.1	-0.1	0.5	0.3	1.0
UNHAPPY	-0.2	-0.2	0.2	0.2	0.3	0.4	-0.3	-0.2	-0.3

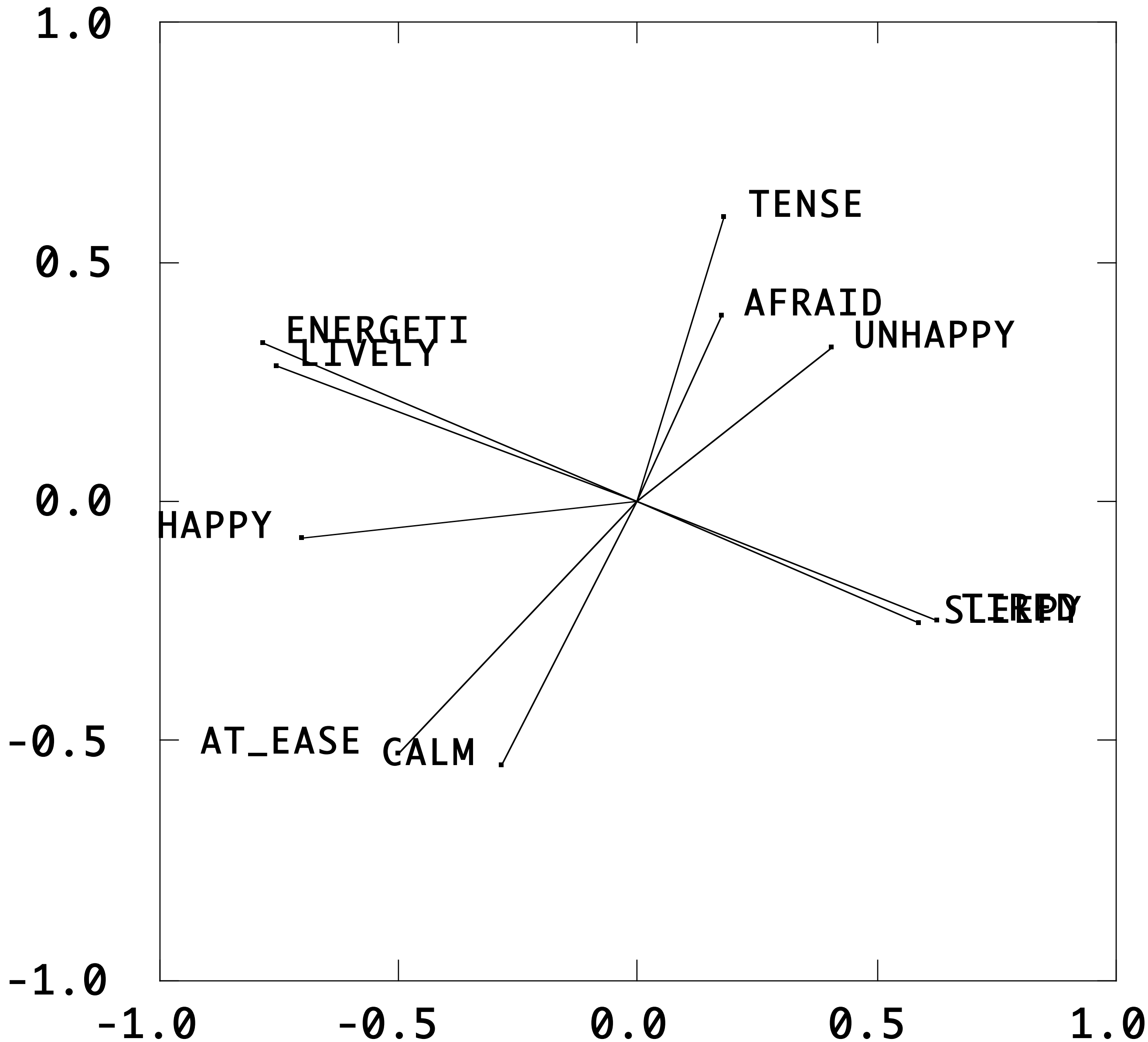
NUMBER OF OBSERVATIONS: 3748

Factor analysis 2 factor solution

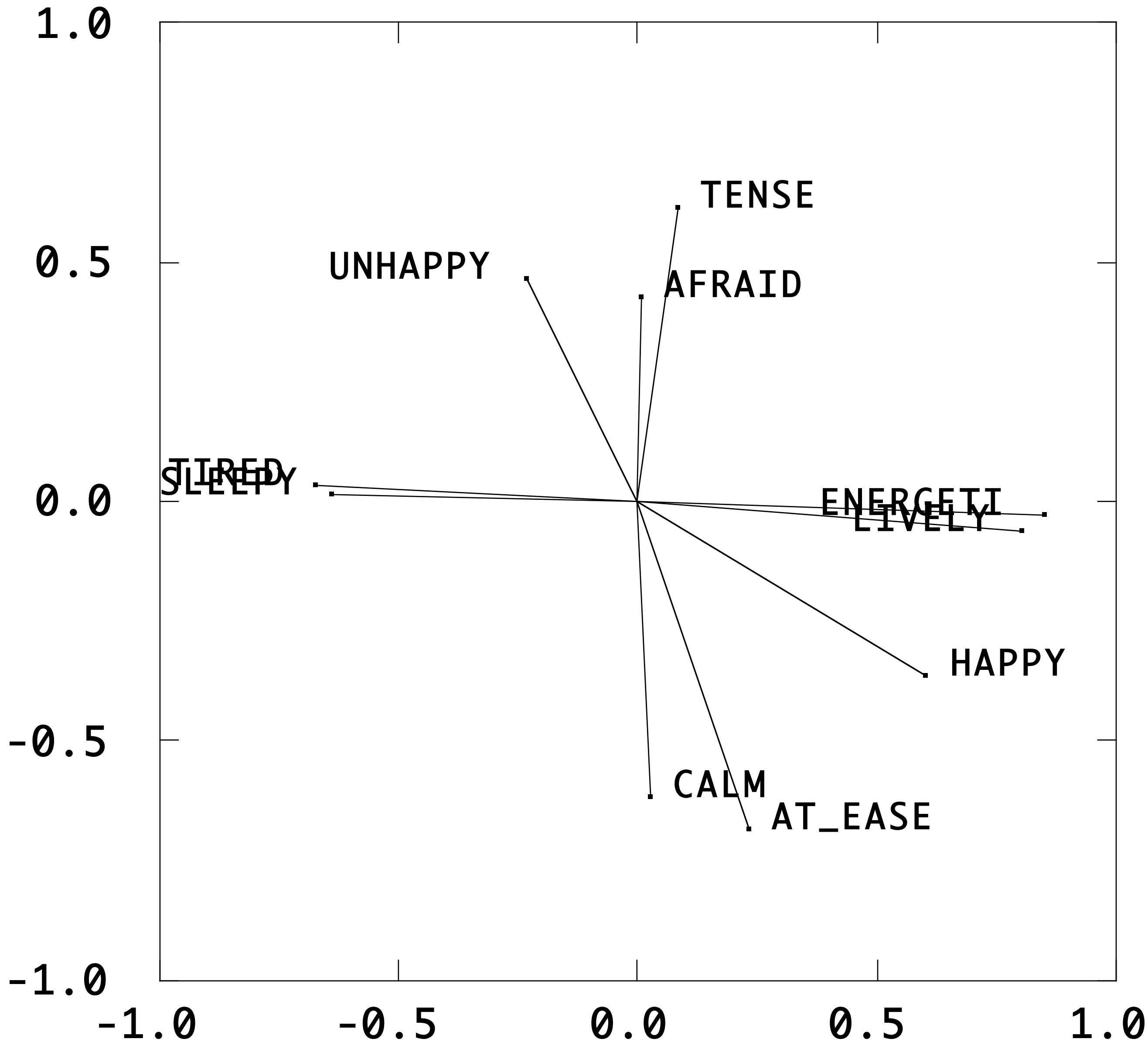
FACTOR PATTERN

	1	2	h2
ENERGETI	-0.8	0.3	0.73
LIVELY	-0.8	0.3	0.73
HAPPY	-0.7	-0.1	0.50
TIRED	0.6	-0.3	0.45
SLEEPY	0.6	-0.3	0.45
TENSE	0.2	0.6	0.40
CALM	-0.3	-0.6	0.45
AT_EASE	-0.5	-0.5	0.50
AFRAID	0.2	0.4	0.20
UNHAPPY	0.4	0.3	0.25
VARIANCE EXP	3.0	1.50	

2 factors of mood



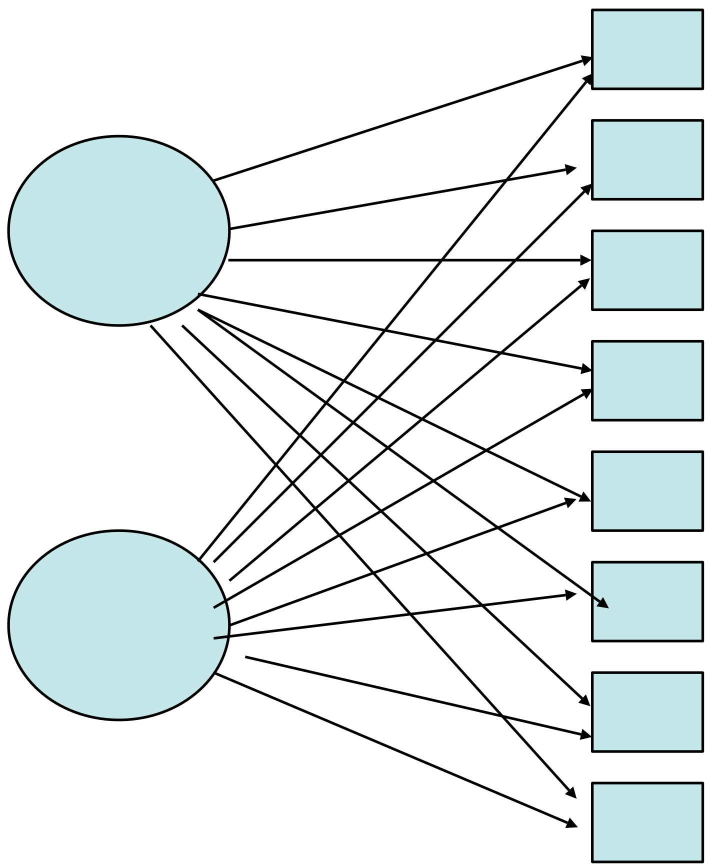
2 factors of mood (rotated)



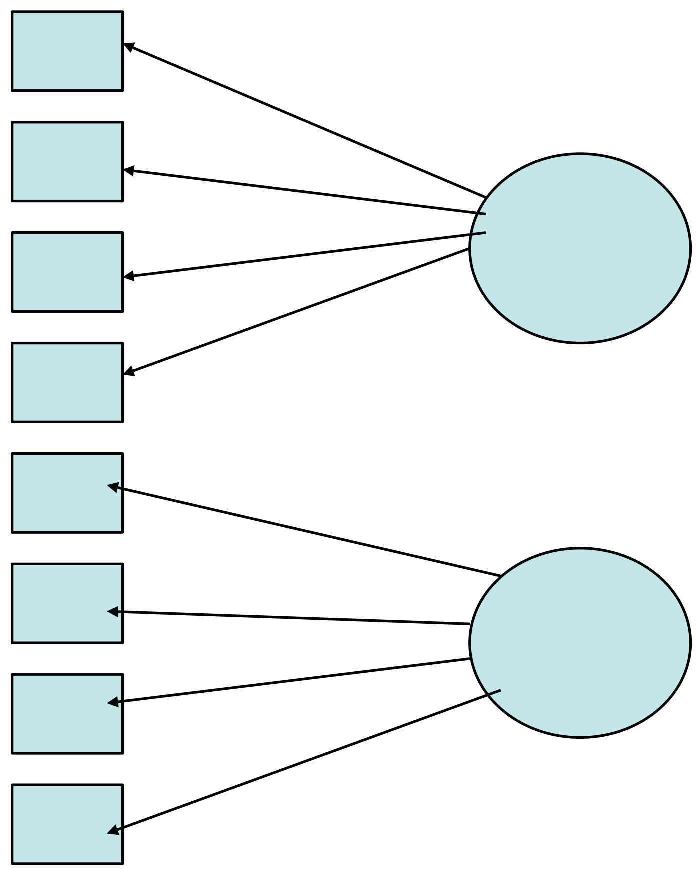
Rotation as orthogonal transformation

	F1	F2	F1'	F2'
ENERGETI	-0.8	0.3	0.8	0.0
LIVELY	-0.8	0.3	0.8	-0.1
HAPPY	-0.7	-0.1	0.6	-0.4
TIRED	0.6	-0.3	-0.7	0.0
SLEEPY	0.6	-0.3	-0.6	0.0
TENSE	0.2	0.6	0.1	0.6
CALM	-0.3	-0.6	0.0	-0.6
AT_EASE	-0.5	-0.5	0.2	-0.7
AFRAID	0.2	0.4	0.0	0.4
UNHAPPY	0.4	0.3	-0.2	0.5
eigen values	3	1.5	2.7	1.8

Rotation to simple structure
Original

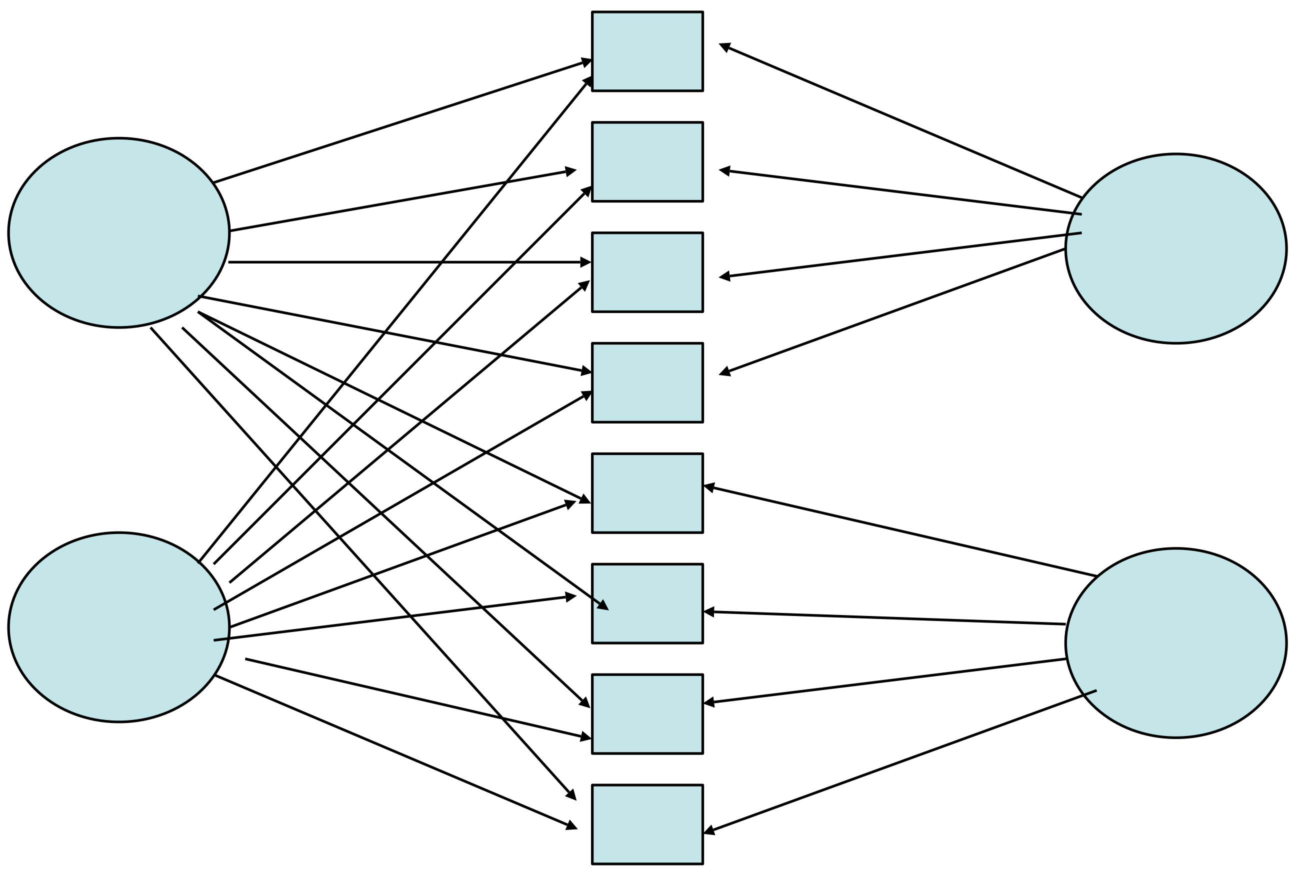


Rotation to simple structure
Simple Structure



Rotation to simple structure

Original Rotated



Principal components

I. R≈ CC'

II. Residual Matrix R* = R - (CC')

III. Try to minimize the residual

IV. Components are linear composites of known variables.

V. Covariance structures of observables in terms of covariance of observables

VI. Components account for observed variance

2 Principal Components of mood

	C1	C2
ENERGETI	-0.8	-0.4
LIVELY	-0.8	-0.3
HAPPY	-0.8	0
TIRED	0.7	0.3
SLEEPY	0.7	0.3
AT_EASE	-0.6	0.5
TENSE	0.2	-0.7
CALM	-0.3	0.6
AFRAID	0.2	-0.5
UNHAPPY	0.5	-0.4
eigen values	3.4	2.1

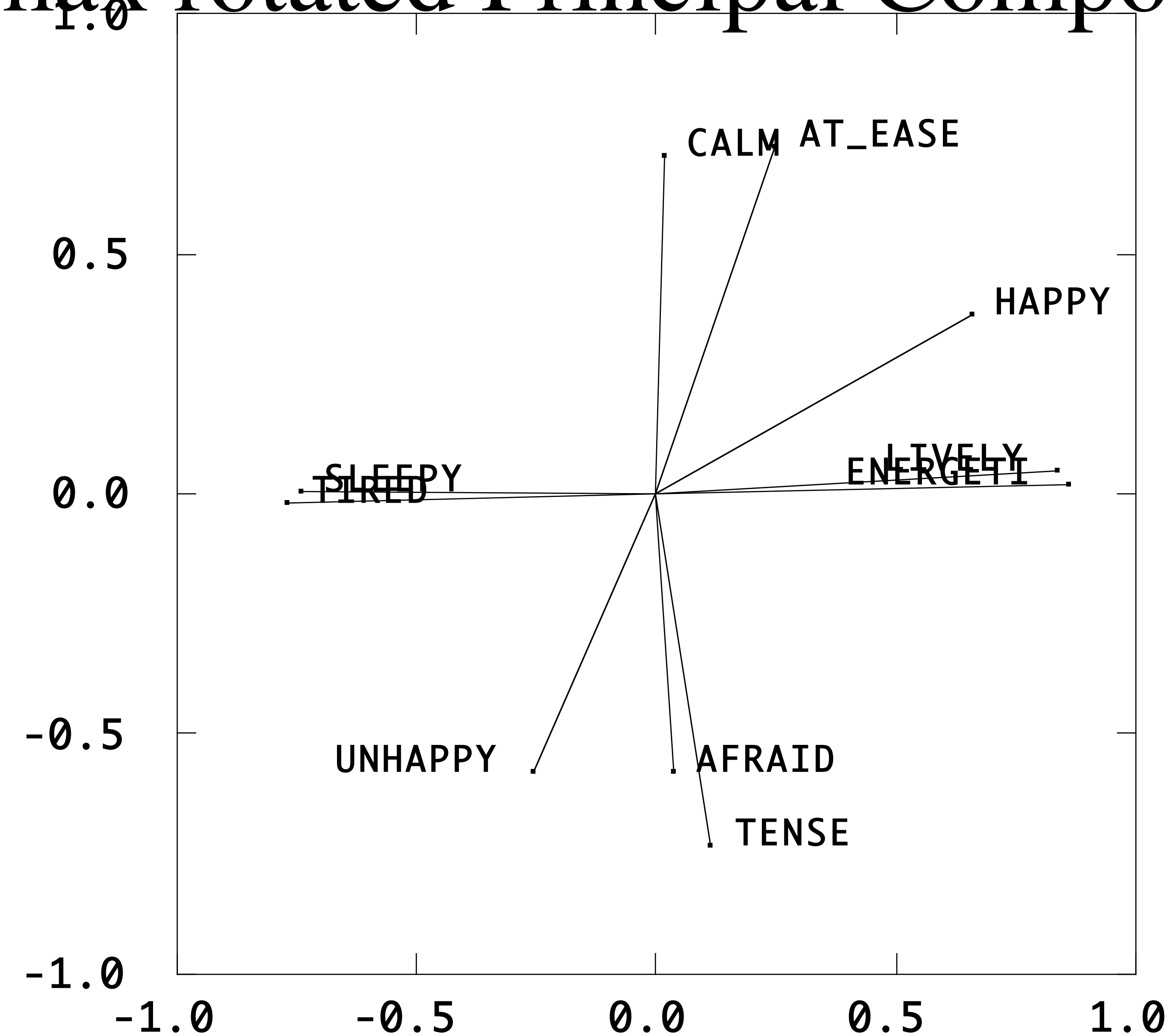
Unrotated and rotated PCs

	C1	C2	C1'	C2'
ENERGETI	-0.8	-0.4	0.9	0
LIVELY	-0.8	-0.3	0.8	0
HAPPY	-0.8	0	0.7	0.4
TIRED	0.7	0.3	-0.8	0
SLEEPY	0.7	0.3	-0.7	0
AT_EASE	-0.6	0.5	0.2	0.7
TENSE	0.2	-0.7	0.1	-0.7
CALM	-0.3	0.6	0	0.7
AFRAID	0.2	-0.5	0	-0.6
UNHAPPY	0.5	-0.4	-0.3	-0.6
eigen values	3.4	2.1	3.2	2.4

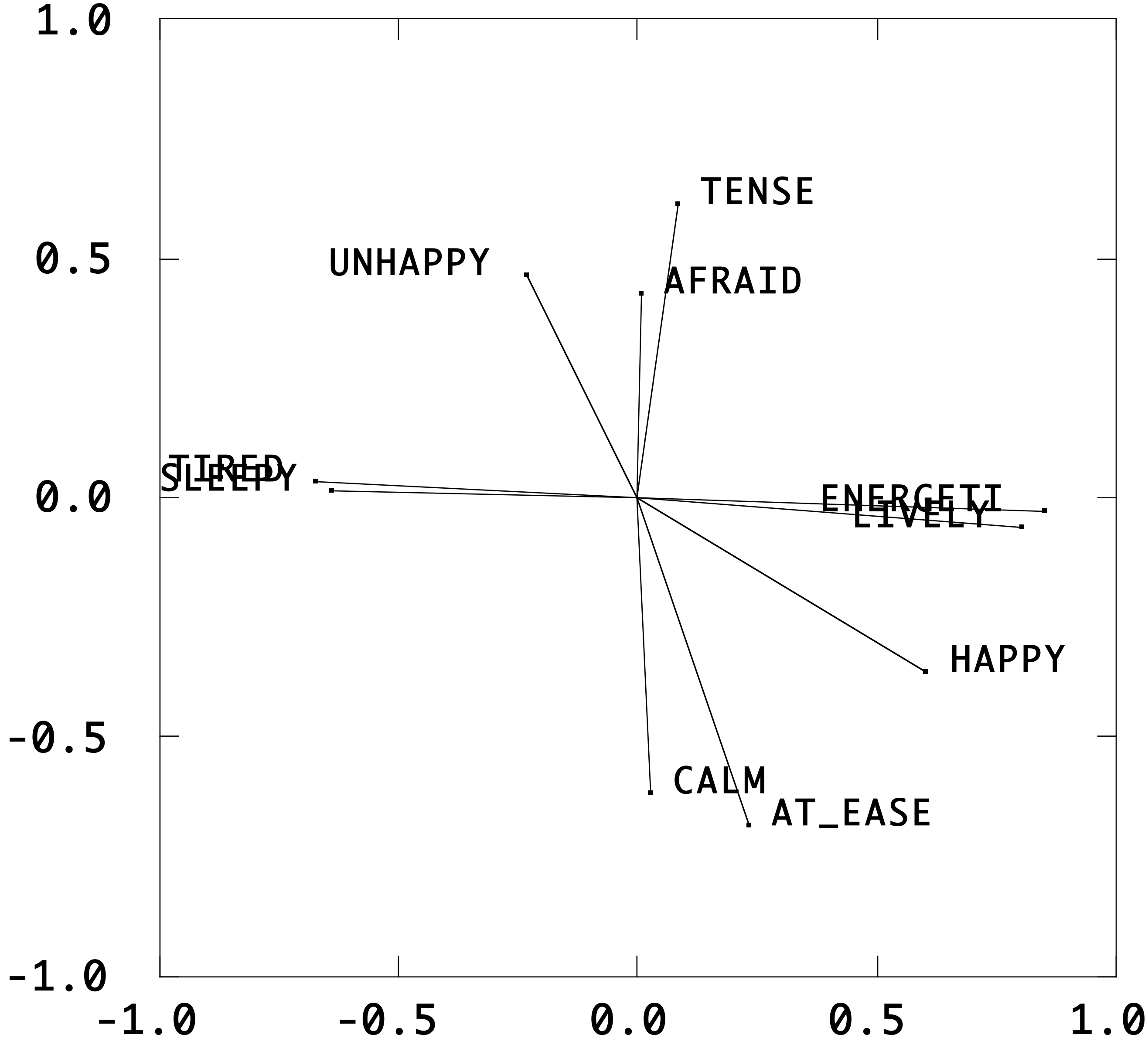
Varimax factors and components

	F1'	F2'	C1'	C2'
ENERGETI	0.8	0.0	0.9	0.0
LIVELY	0.8	-0.1	0.8	0.0
HAPPY	0.6	-0.4	0.7	0.4
TIRED	-0.7	0.0	-0.8	0.0
SLEEPY	-0.6	0.0	-0.7	0.0
AT_EASE	0.2	-0.7	0.2	0.7
TENSE	0.1	0.6	0.1	-0.7
CALM	0.0	-0.6	0.0	0.7
AFRAID	0.0	0.4	0.0	-0.6
UNHAPPY	-0.2	0.5	-0.3	-0.6
eigen values	2.7	1.8	3.2	2.4

Varimax rotated Principal Components



Varimax rotated factors of mood



Components vs. Factors

- Factors are theoretical constructs
- Components are observed linear sums
- Factors do not change when new variables are added
- Components change

Consider this matrix

```
> f <- matrix(c(.9,.8,.7,rep(0,6),.8,.7,.6),ncol=2) #the model
>  rownames(f) <- paste("V",seq(1:6),sep="") #add labels
> colnames(f) <- c("F1", "F2")
> R <- f %*% t(f) #create the correlation matrix
> diag(R) <- 1 #adjust the diagonal of the matrix
> R
```

	V1	V2	V3	V4	V5	V6
V1	1.00	0.72	0.63	0.00	0.00	0.00
V2	0.72	1.00	0.56	0.00	0.00	0.00
V3	0.63	0.56	1.00	0.00	0.00	0.00
V4	0.00	0.00	0.00	1.00	0.56	0.48
V5	0.00	0.00	0.00	0.56	1.00	0.42
V6	0.00	0.00	0.00	0.48	0.42	1.00

factanal(factors = 2, covmat = R)

V1	V2	V3	V4	V5	V6
0.19	0.36	0.51	0.36	0.51	0.64

Loadings:

	Factor1	Factor2
V1	0.9	
V2	0.8	
V3	0.7	
V4		0.8
V5		0.7
V6		0.6

	Factor1	Factor2
SS loadings	1.940	1.490
Proportion Var	0.323	0.248
Cumulative Var	0.323	0.572

pc <- principal(R,2)

	PC1	PC2
V1	0.90	
V2	0.88	
V3	0.83	
V4		0.85
V5		0.82
V6		0.77

	PC1	PC2
SS loadings	2.273	1.988
Proportion Var	0.379	0.331
Cumulative Var	0.379	0.710

Add two variables

	V1	V2	V3	V4	V5	V6	V7	V8
V1	1.00	0.72	0.63	0.00	0.00	0.00	0.63	0.00
V2	0.72	1.00	0.56	0.00	0.00	0.00	0.56	0.00
V3	0.63	0.56	1.00	0.00	0.00	0.00	0.49	0.00
V4	0.00	0.00	0.00	1.00	0.56	0.48	0.00	0.40
V5	0.00	0.00	0.00	0.56	1.00	0.42	0.00	0.35
V6	0.00	0.00	0.00	0.48	0.42	1.00	0.00	0.30
V7	0.63	0.56	0.49	0.00	0.00	0.00	1.00	0.00
V8	0.00	0.00	0.00	0.40	0.35	0.30	0.00	1.00

Factors don't change

	Factor1	Factor2
V1	0.9	
V2	0.8	
V3	0.7	
V4		0.8
V5		0.7
V6		0.6

	Factor1	Factor2
SS loadings	1.940	1.490
Proportion Var	0.323	0.248
Cumulative Var	0.323	0.572

	Factor1	Factor2
V1	0.9	
V2	0.8	
V3	0.7	
V4		0.8
V5		0.7
V6		0.6
V7	0.7	
V8		0.5

	Factor1	Factor2
SS loadings	2.430	1.740
Proportion Var	0.304	0.21
Cumulative Var	0.304	0.52

Components do

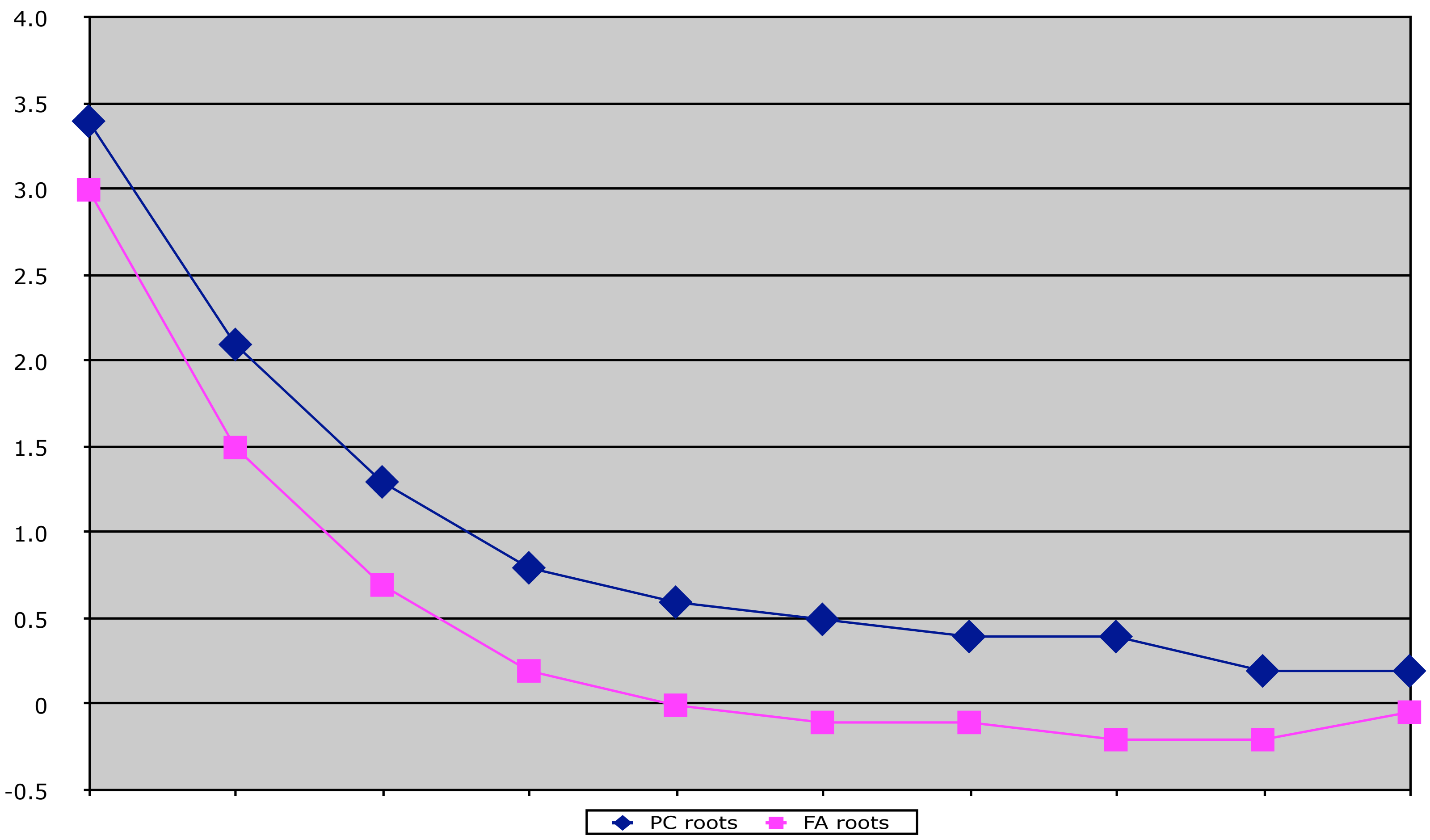
	PC1	PC2
V1	0.90	
V2	0.88	
V3	0.83	
V4		0.85
V5		0.82
V6		0.77

	PC1	PC2
SS loadings	2.273	1.988
Proportion Var	0.379	0.331
Cumulative Var	0.379	0.710

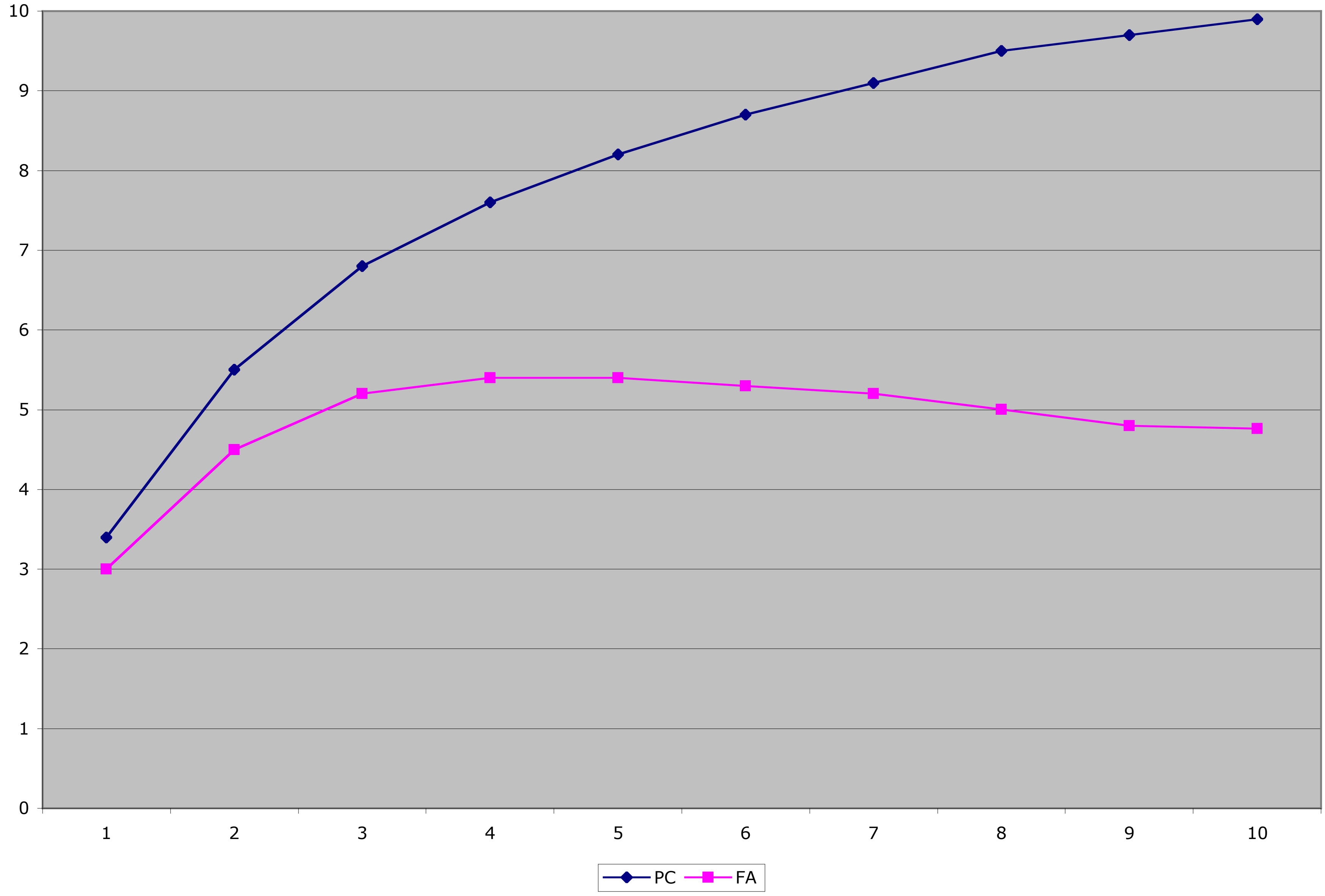
	PC1	PC2
V1	0.90	
V2	0.85	
V3	0.80	
V4		0.83
V5		0.79
V6		0.73
V7	0.80	
V8		0.65

	PC1	PC2
SS loadings	2.812	2.268
Proportion Var	0.352	0.28
Cumulative Var	0.352	0.63

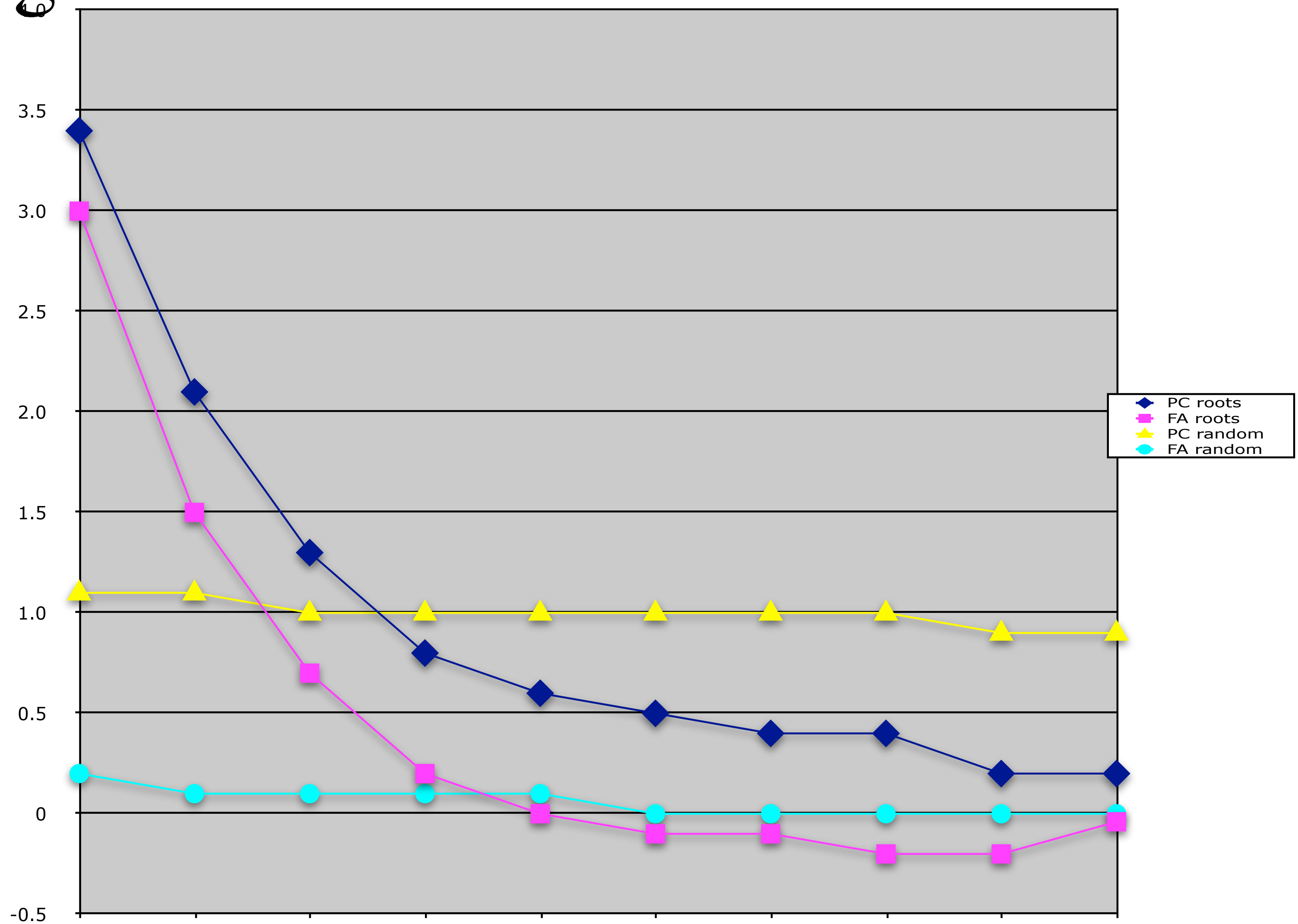
Eigenvalues and the scree test

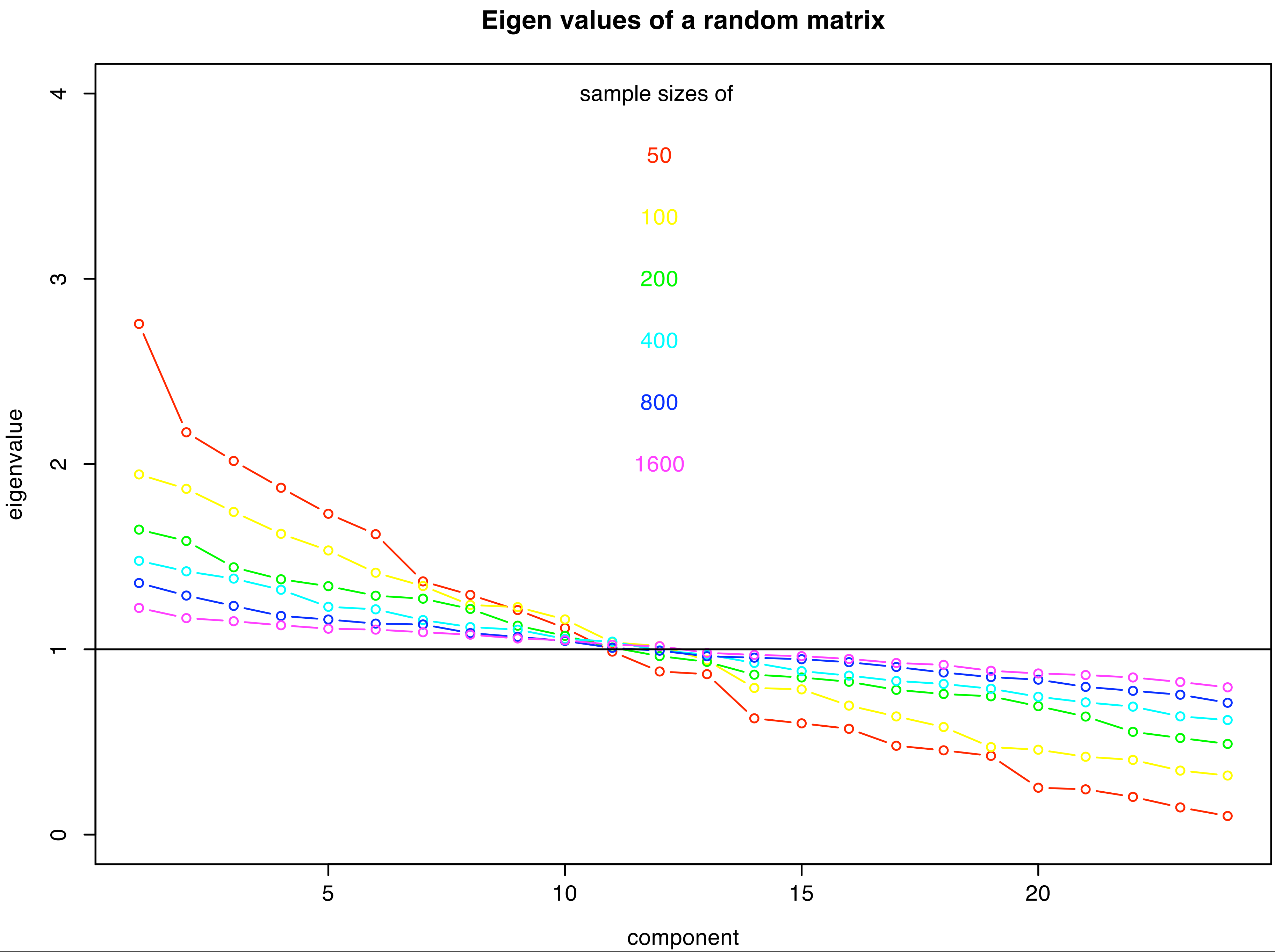


Cumulative variance explained



Eigen values for real vs. random data



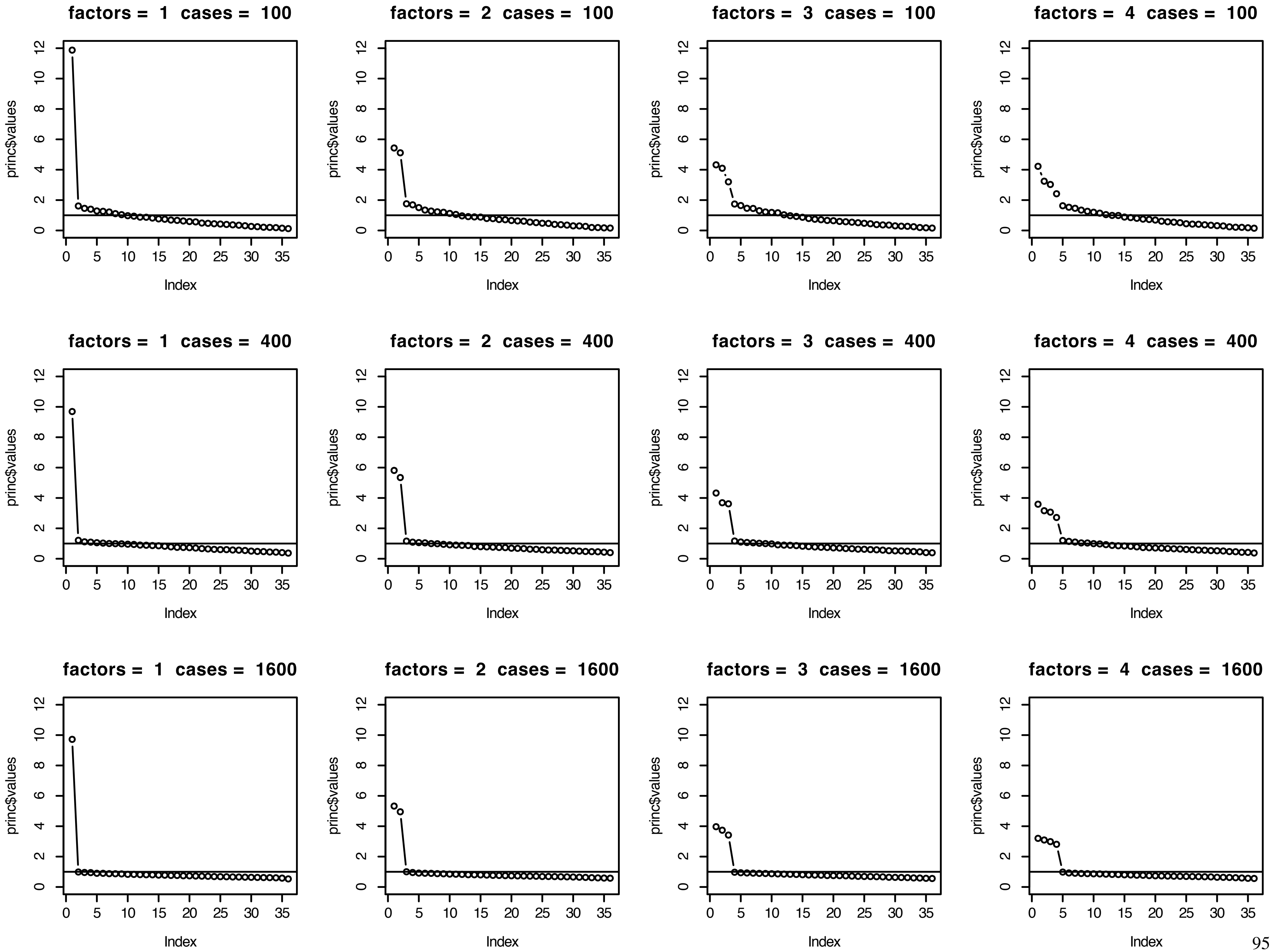

Eigen values of a random matrix
sample sizes of
50
100
200
400
800
1600
eigenvalue
component
0
1
2
3
4
5
10
15
20

Determining the number of factors/components

- Statistically: extract factors until residual matrix does not differ from 0
  - Sensitive to sample size (large N -> more factors)
- Theoretically: extract factors that make sense
  - Different theorists differ in their cognitive complexity
- Pragmatically: scree test
- Pragmatically: extract factors with eigen values greater than a random factor matrix
- Pragmatically: extract factors using Very Simple Structure Test or the minimal partial correlation
- Pragmatically: do not use eigen value>1 rule!

Scree vs. eigen value > 1

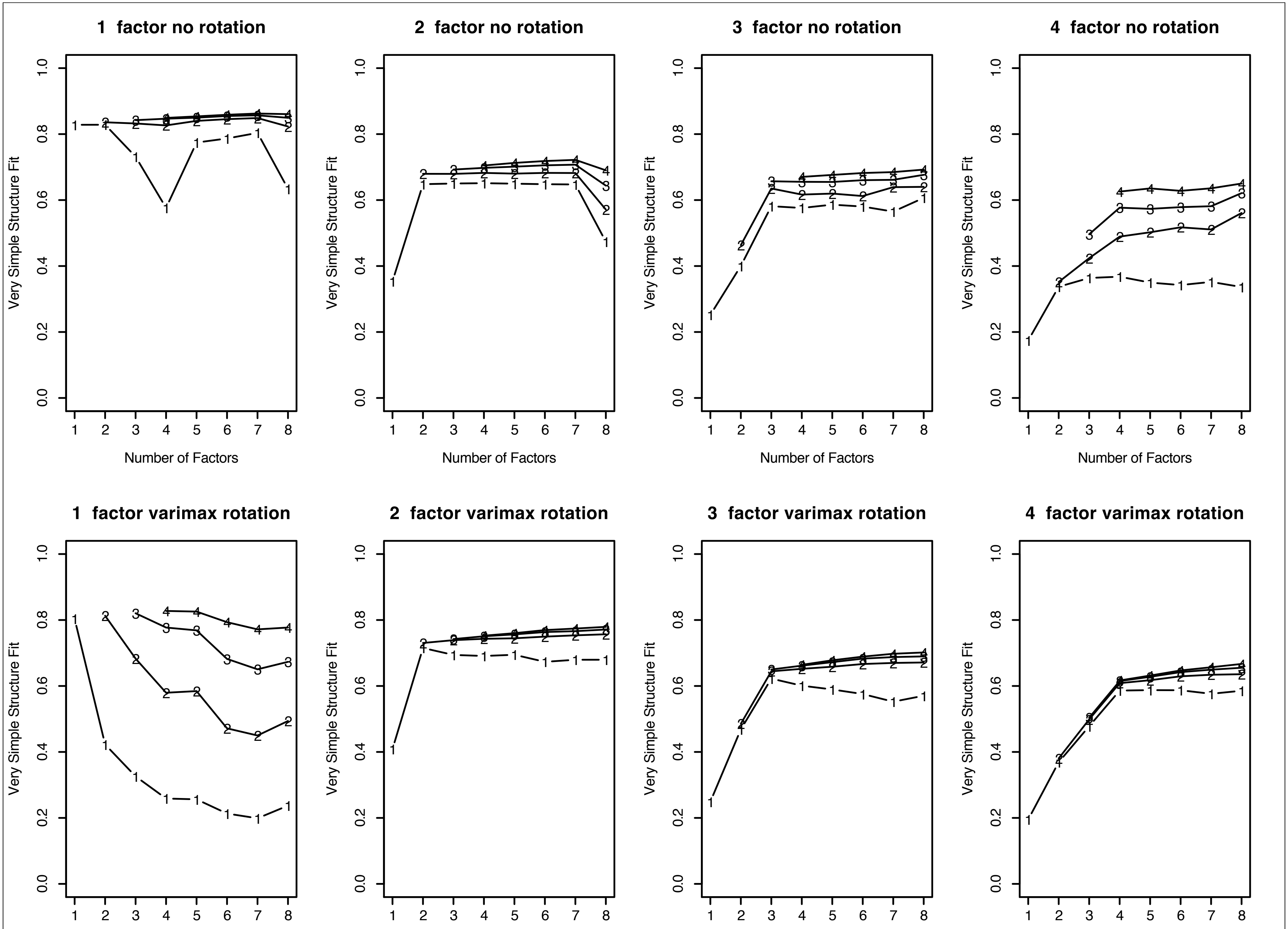
- Artificial data with number of factors ranging from 1 to 4
- Number of cases ranges from small (100) to normal (400) to large (1600)
- Item communality set to "typical" = .25
- Simulation using VSS.simulate in VSS "package"
  - `source("http://personality-project.org/r/vss.r")`
  - see http://personality-project.org/r/vss.html

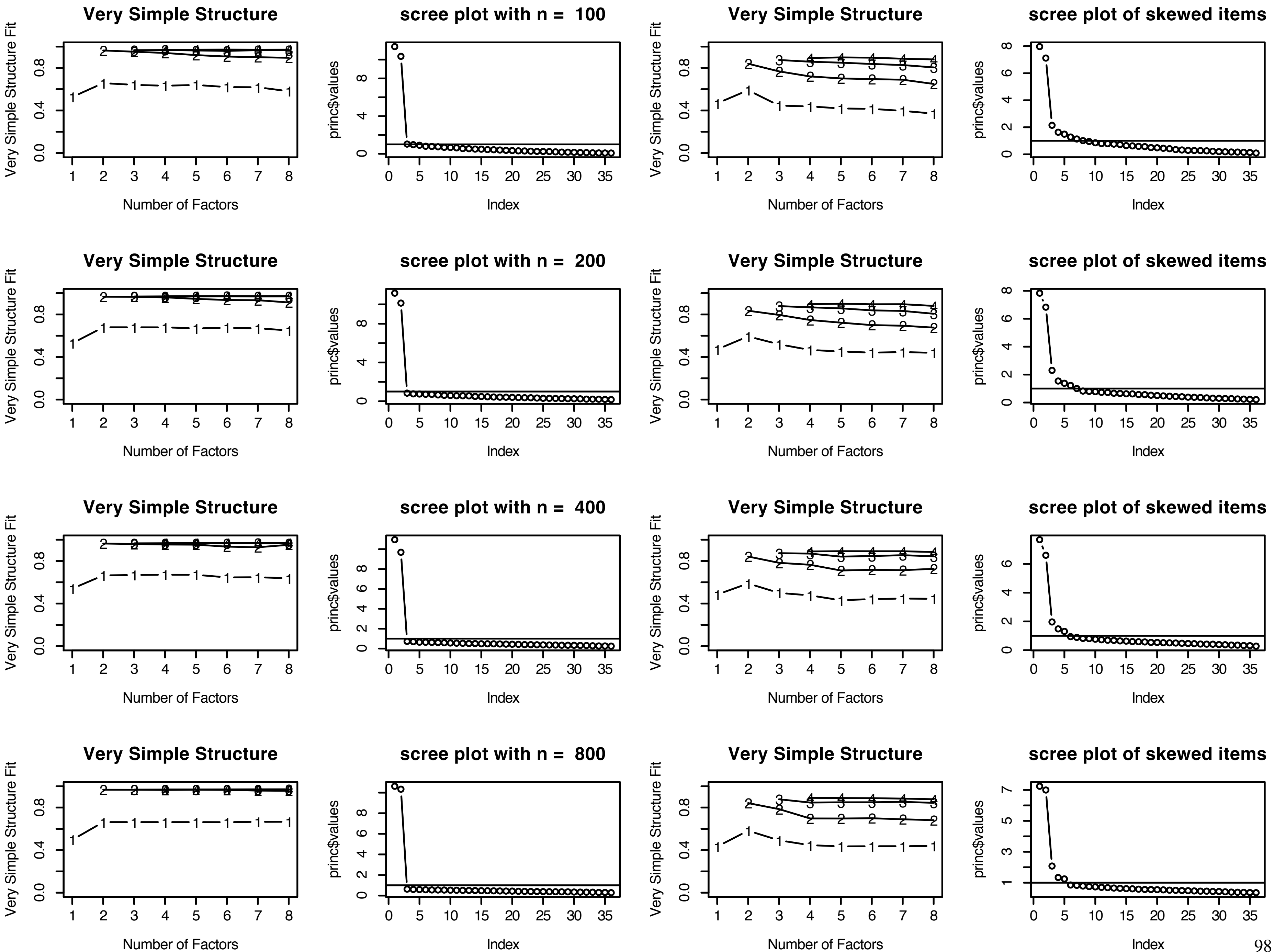

factors = 1 cases = 100
factors = 2 cases = 100
factors = 3 cases = 100
factors = 4 cases = 100
factors = 1 cases = 400
factors = 2 cases = 400
factors = 3 cases = 400
factors = 4 cases = 400
factors = 1 cases = 1600
factors = 2 cases = 1600
factors = 3 cases = 1600
factors = 4 cases = 1600
princ$values
Index

Very Simple Structure

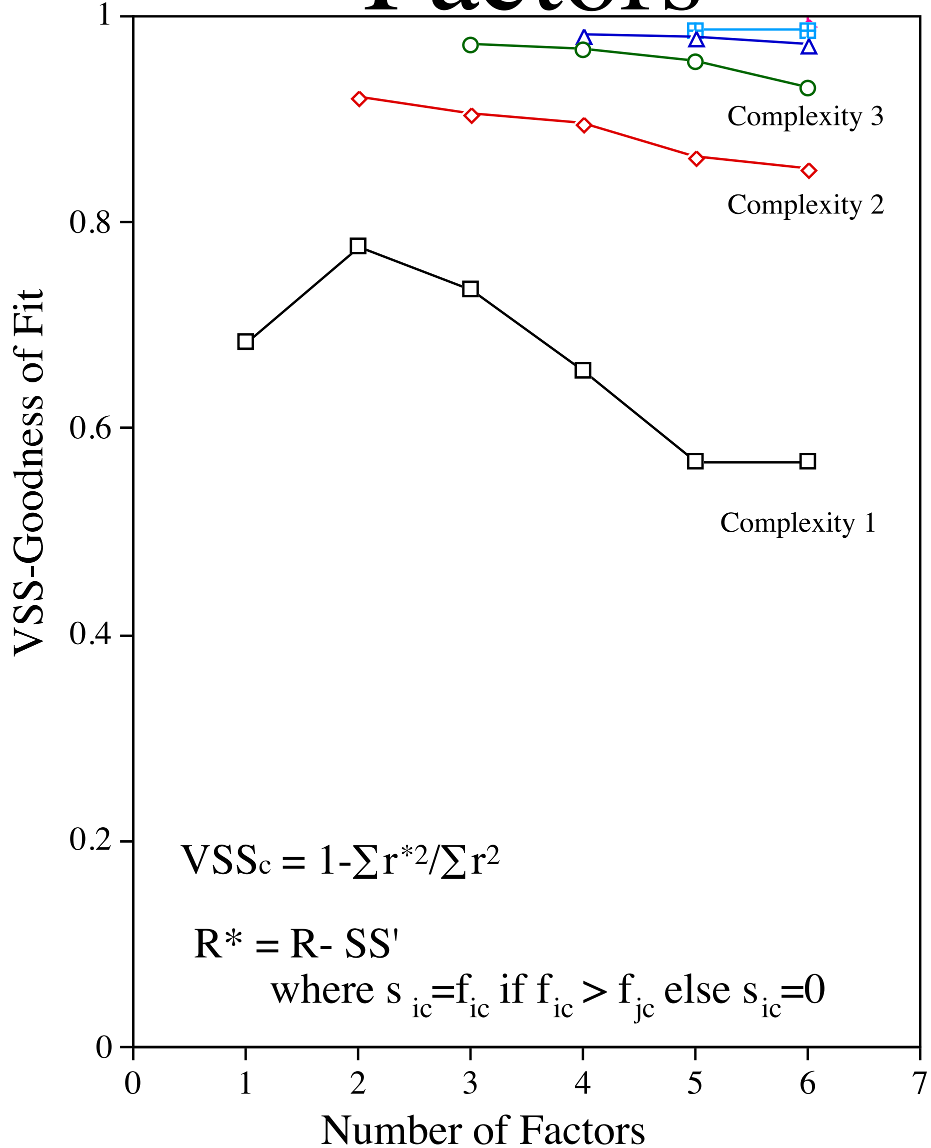
(Revelle and Rocklin)

- Consider the factors as interpreted by user
  - Small loadings are thought to be zero
- How well does this very simple interpretation of the actual structure reproduce the correlation matrix
- $R^*_{vss} = R - F_{vss} * F'_{vss}$
- Consider structures of various levels of simplicity (complexity) (1, 2, … loadings/item)
- Solution peaks at optimal number of factors

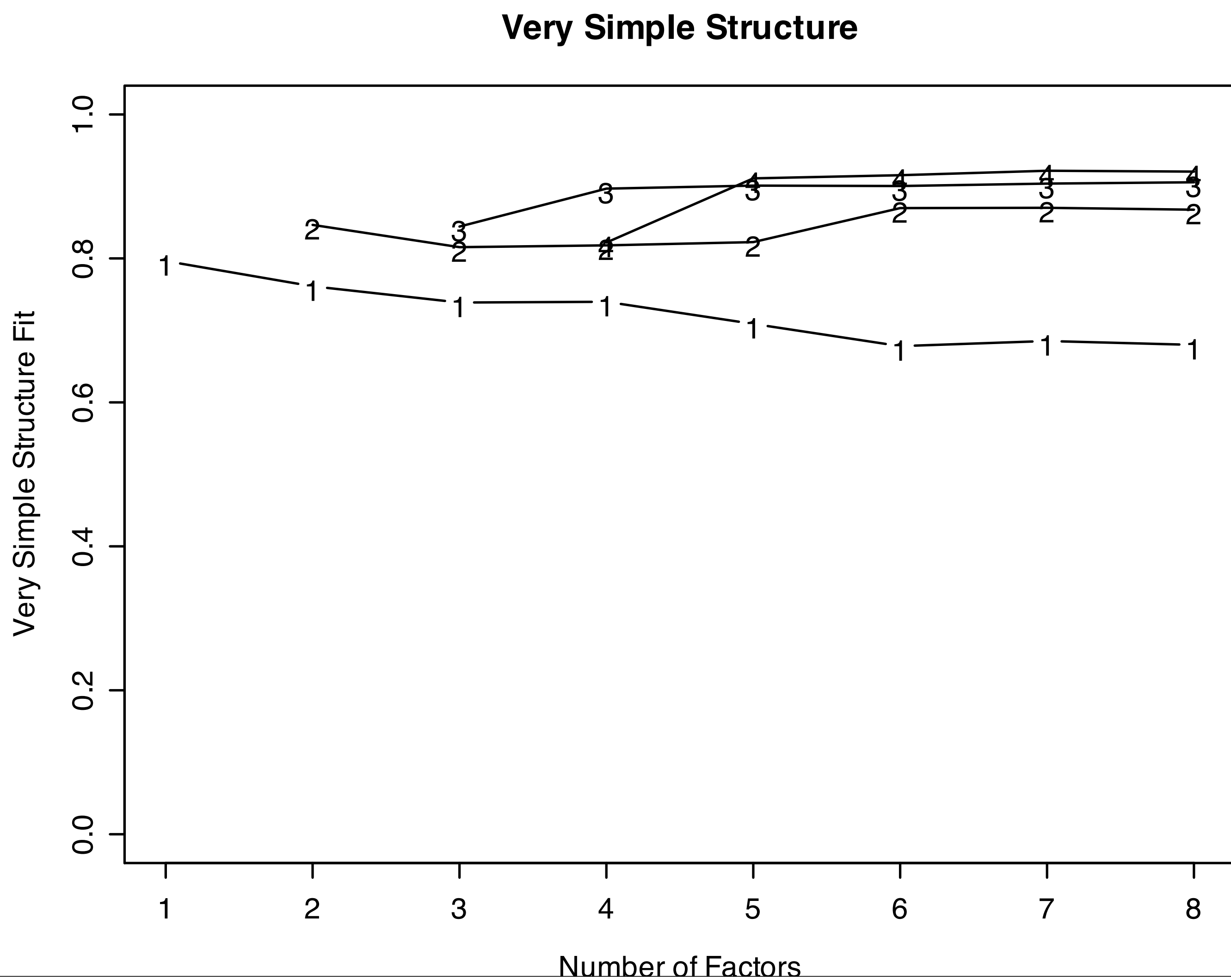

1 factor no rotation
2 factor no rotation
3 factor no rotation
4 factor no rotation
1 factor varimax rotation
2 factor varimax rotation
3 factor varimax rotation
4 factor varimax rotation
Very Simple Structure Fit
Number of Factors


Very Simple Structure
scree plot with n = 100
Very Simple Structure
scree plot of skewed items
Very Simple Structure
scree plot with n = 200
Very Simple Structure
scree plot of skewed items
Very Simple Structure
scree plot with n = 400
Very Simple Structure
scree plot of skewed items
Very Simple Structure
scree plot with n = 800
Very Simple Structure
scree plot of skewed items
Very Simple Structure Fit
Number of Factors
princ$values
Index

Very Simple Structure of affect => 2 Factors



VSS of Ability



Minimum Average Partial (Velicer)

R	C’
C	CC’

=

R	C’
C	I

Minimize partial R

- $R^* = R - CC'$ is the residual correlation
- $D = \text{diag}(R^*)$ is the variance of residuals
- $R^{\#} = D^{-.5} R^* D^{-.5}$ (matrix form)

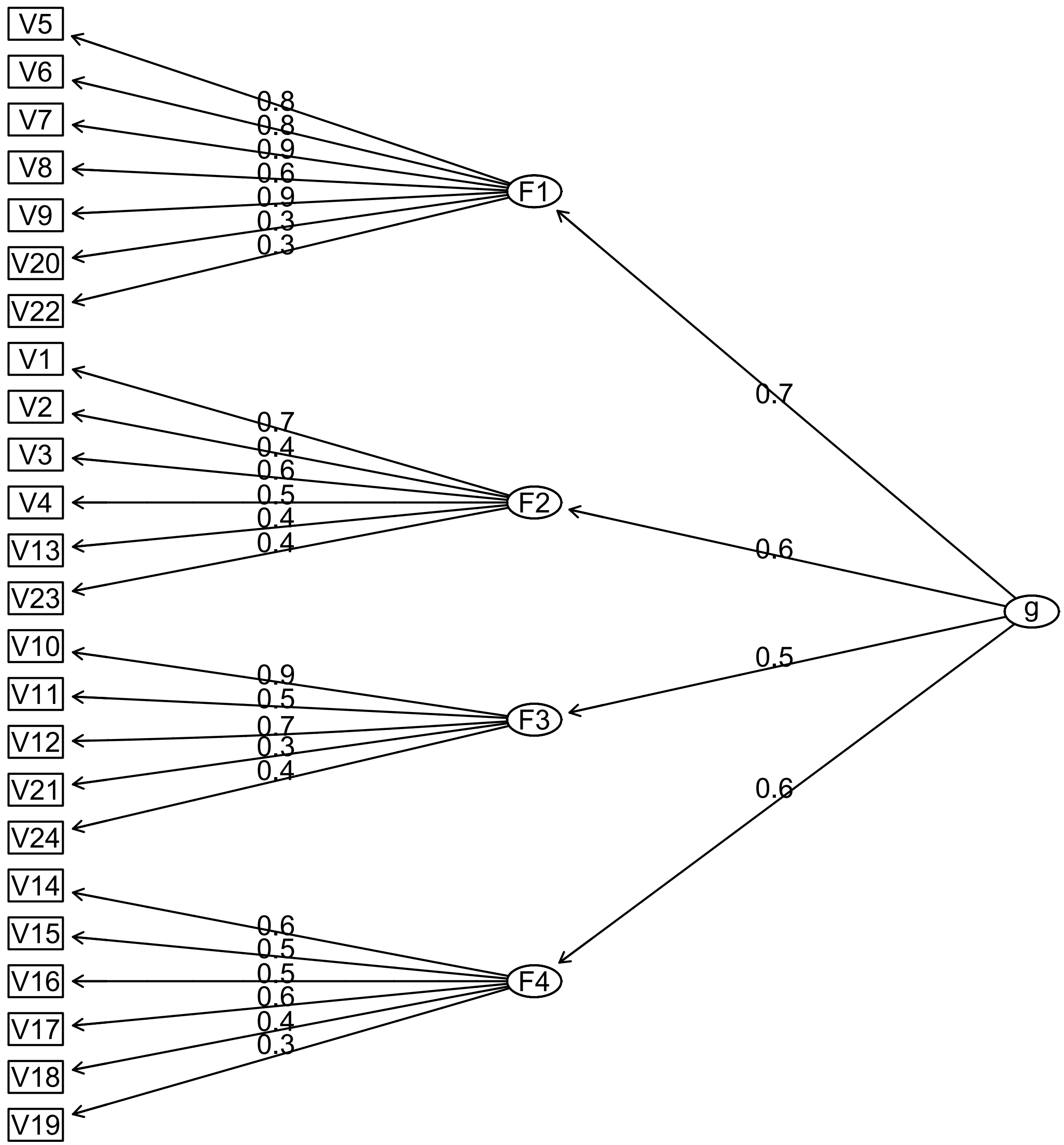
MAP

- $R^* = R - CC'$ is the residual correlation
- $D = diag(R^*)$ is the variance of residuals
- $R^\# = R^*/sqrt(D)$ (conceptually)
- $R^\# = D^{-.5} R^* D^{-.5}$ (matrix form)
- Now part of the VSS function

```
>round(harman.vss$map,4)
[1]0.0245 0.0216 0.0175 0.0174 0.0208 0.0237 0.0278 0.0301
>round(circ.vss$map,3)
[1]0.044 0.006 0.008 0.011 0.014 0.017 0.020 0.024
```

Hierarchical model

	g factor	PA1	PA2	PA3	PA4	h2	u2
VisualPerception	0.50	0.031	0.0417	0.5429	0.042	0.47	0.53
Cubes	0.31	0.037	0.0202	0.3593	0.015	0.21	0.79
PaperFormBoard	0.35	0.068	0.1275	0.4318	0.040	0.33	0.67
Flags	0.40	0.134	0.0327	0.4085	0.017	0.30	0.70
GeneralInformation	0.54	0.576	0.1055	0.0103	0.019	0.59	0.41
PargraphComprehension	0.55	0.605	0.0485	0.0034	0.072	0.65	0.35
SentenceCompletion	0.52	0.660	0.0410	0.0057	0.083	0.77	0.23
WordClassification	0.54	0.423	0.1000	0.1709	0.007	0.37	0.63
WordMeaning	0.56	0.652	0.0681	0.0134	0.059	0.75	0.25
Addition	0.38	0.057	0.7549	0.1274	0.028	0.76	0.24
Code	0.47	0.052	0.4380	0.0028	0.229	0.34	0.66
CountingDots	0.38	0.081	0.6169	0.1711	0.019	0.55	0.45
StraightCurvedCapitals	0.48	0.055	0.4090	0.3306	0.045	0.40	0.60
WordRecognition	0.39	0.107	0.0023	0.0890	0.447	0.36	0.64
NumberRecognition	0.36	0.027	0.0043	0.0044	0.420	0.29	0.71
FigureRecognition	0.46	0.041	0.0722	0.2656	0.409	0.40	0.60
ObjectNumber	0.41	0.030	0.1297	0.0674	0.457	0.38	0.62
NumberFigure	0.46	0.094	0.2185	0.1974	0.330	0.32	0.68
FigureWord	0.39	0.050	0.0799	0.1342	0.247	0.14	0.86
Deduction	0.51	0.248	0.0357	0.2476	0.160	0.25	0.75
NumericalPuzzles	0.49	0.053	0.2977	0.2886	0.082	0.26	0.74
ProblemReasoning	0.51	0.237	0.0052	0.2415	0.159	0.23	0.77
SeriesCompletion	0.57	0.225	0.0753	0.3514	0.079	0.30	0.70
ArithmeticProblems	0.53	0.213	0.3700	0.0600	0.134	0.29	0.71


V5
V6
V7
V8
V9
V20
V22
V1
V2
V3
V4
V13
V23
V10
V11
V12
V21
V24
V14
V15
V16
V17
V18
V19
0.8
0.8
0.9
0.6
0.9
0.3
0.3
F1
0.7
0.4
0.6
0.5
0.4
0.4
F2
0.9
0.5
0.7
0.3
0.4
F3
0.6
0.5
0.5
0.6
0.4
0.3
F4
0.7
0.6
0.5
0.6
g

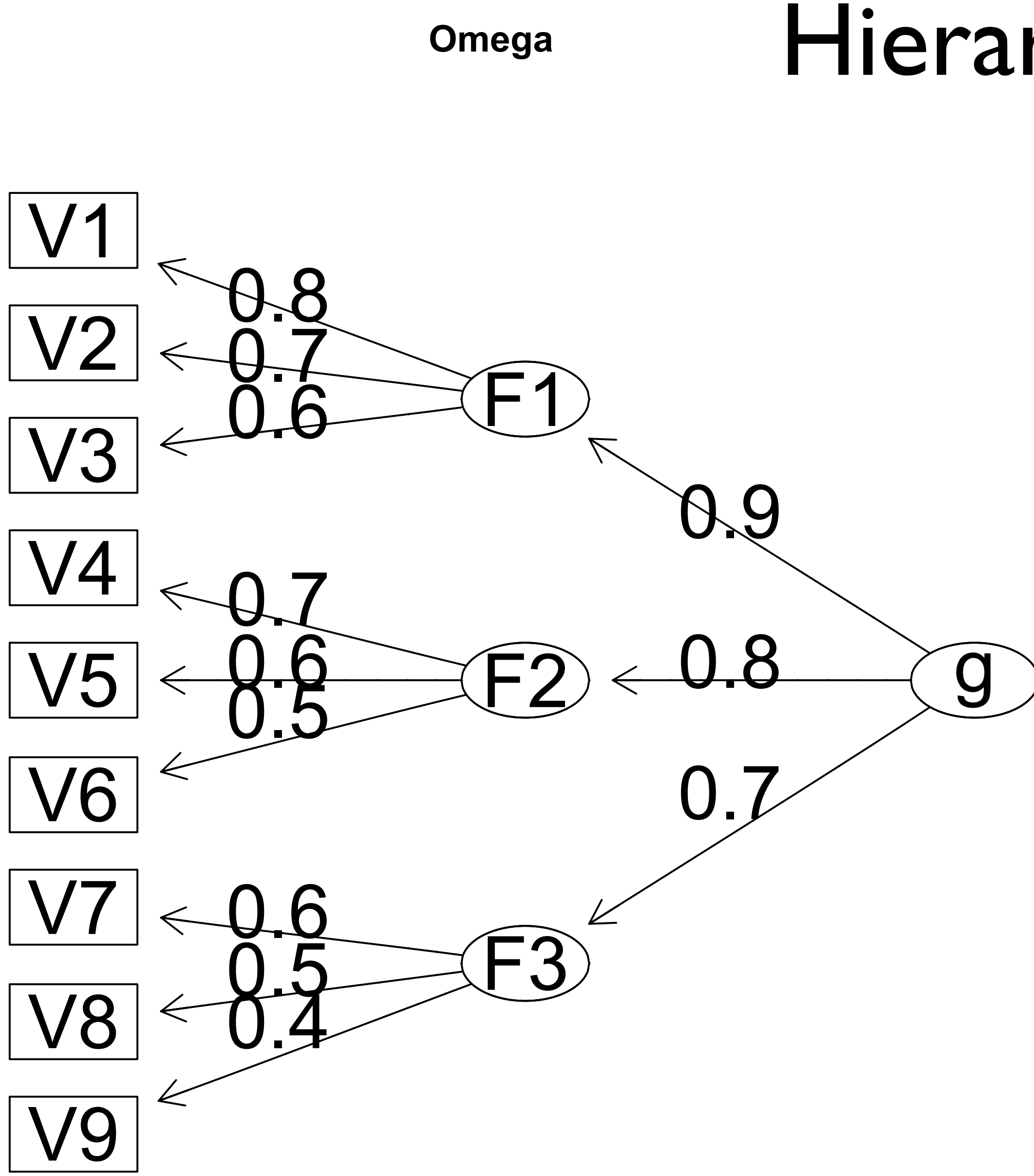
9 variable problem

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.56	0.48	0.40	0.35	0.29	0.30	0.25	0.20
V2	0.56	1.00	0.42	0.35	0.30	0.25	0.26	0.22	0.18
V3	0.48	0.42	1.00	0.30	0.26	0.22	0.23	0.19	0.15
V4	0.40	0.35	0.30	1.00	0.42	0.35	0.24	0.20	0.16
V5	0.35	0.30	0.26	0.42	1.00	0.30	0.20	0.17	0.13
V6	0.29	0.25	0.22	0.35	0.30	1.00	0.17	0.14	0.11
V7	0.30	0.26	0.23	0.24	0.20	0.17	1.00	0.30	0.24
V8	0.25	0.22	0.19	0.20	0.17	0.14	0.30	1.00	0.20
V9	0.20	0.18	0.15	0.16	0.13	0.11	0.24	0.20	1.00

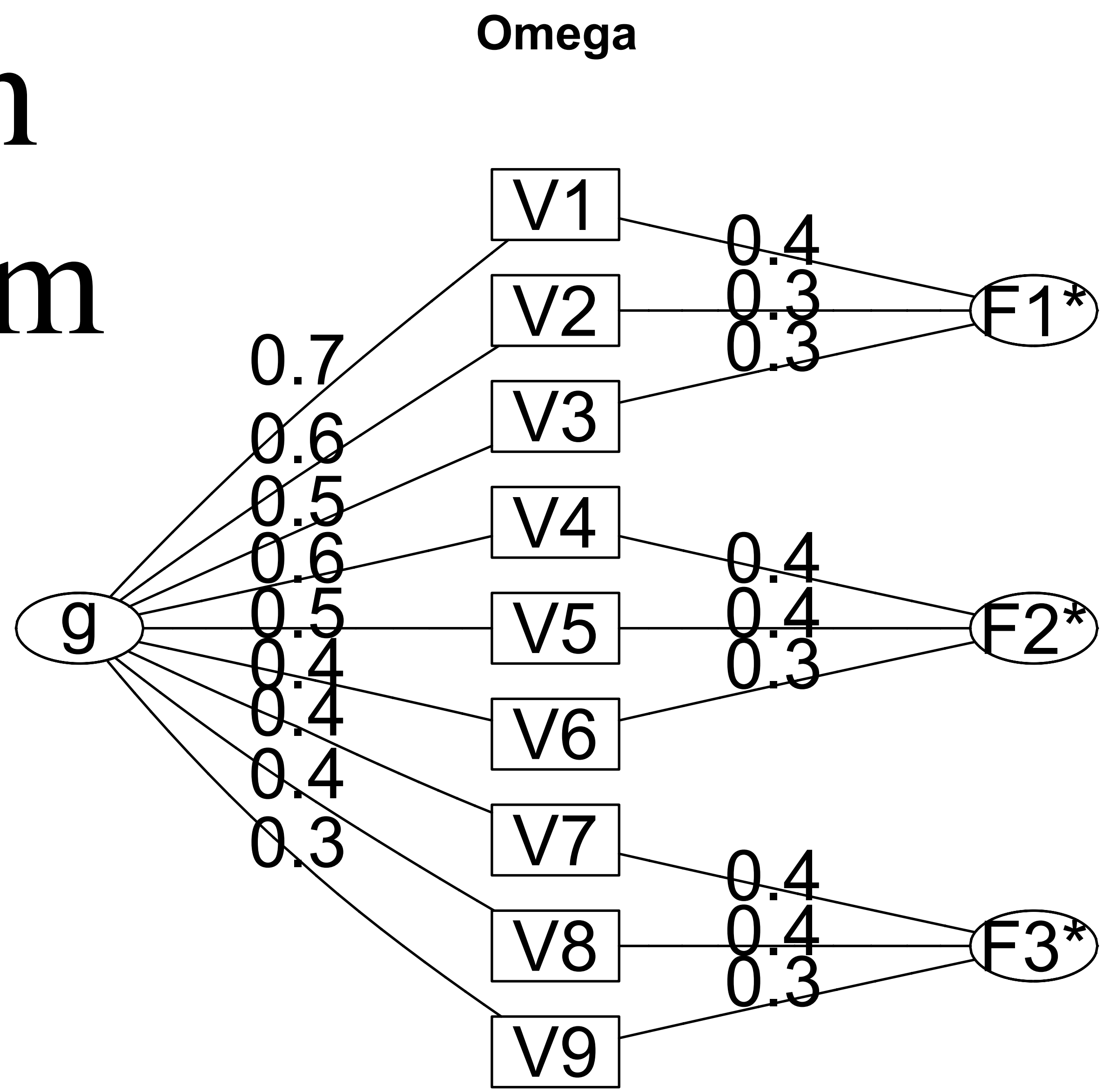
General + Group

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.56	0.48	0.40	0.35	0.29	0.30	0.25	0.20
V2	0.56	1.00	0.42	0.35	0.30	0.25	0.26	0.22	0.18
V3	0.48	0.42	1.00	0.30	0.26	0.22	0.23	0.19	0.15
V4	0.40	0.35	0.30	1.00	0.42	0.35	0.24	0.20	0.16
V5	0.35	0.30	0.26	0.42	1.00	0.30	0.20	0.17	0.13
V6	0.29	0.25	0.22	0.35	0.30	1.00	0.17	0.14	0.11
V7	0.30	0.26	0.23	0.24	0.20	0.17	1.00	0.30	0.24
V8	0.25	0.22	0.19	0.20	0.17	0.14	0.30	1.00	0.20
V9	0.20	0.18	0.15	0.16	0.13	0.11	0.24	0.20	1.00

Hierarchical

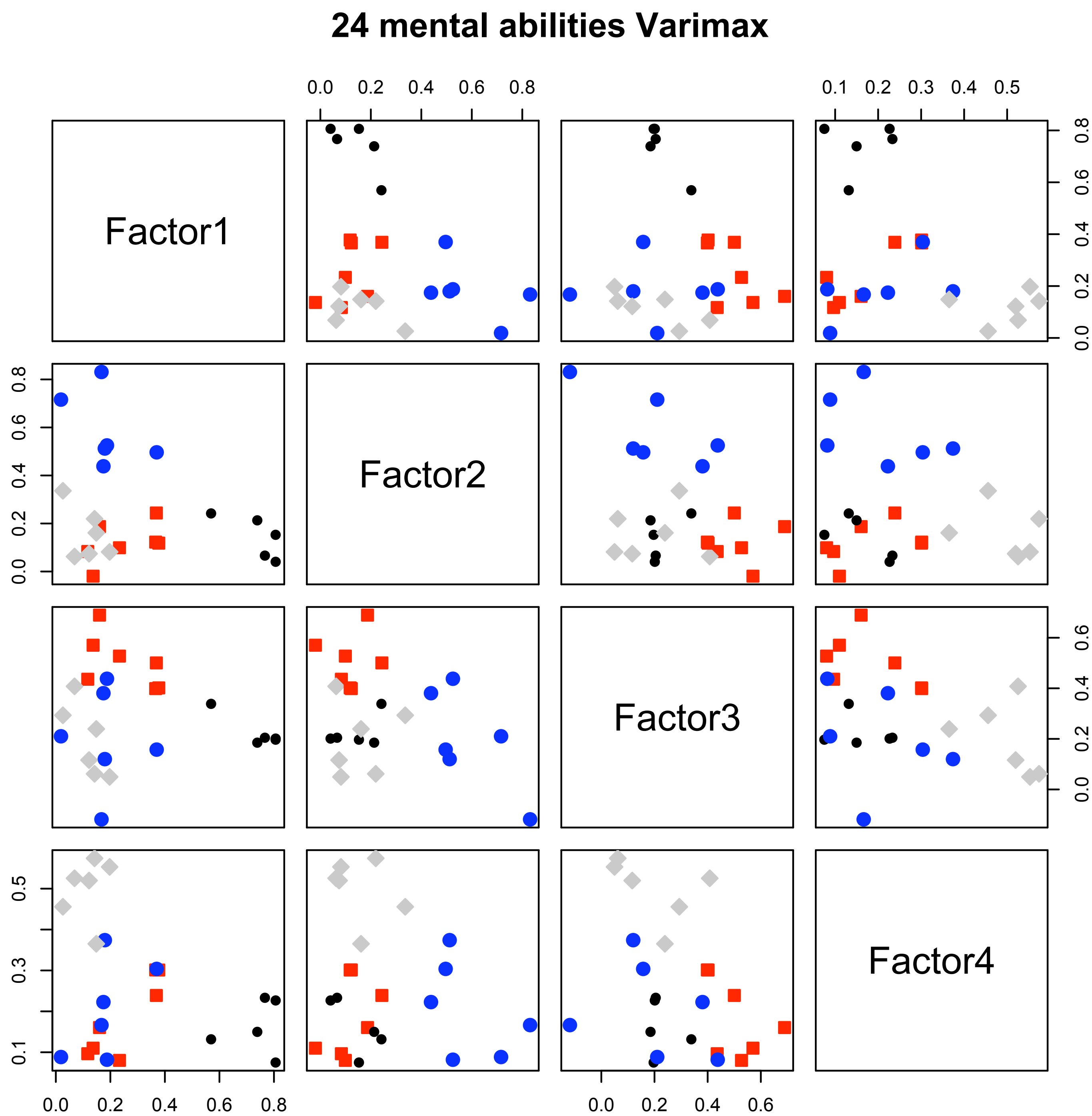


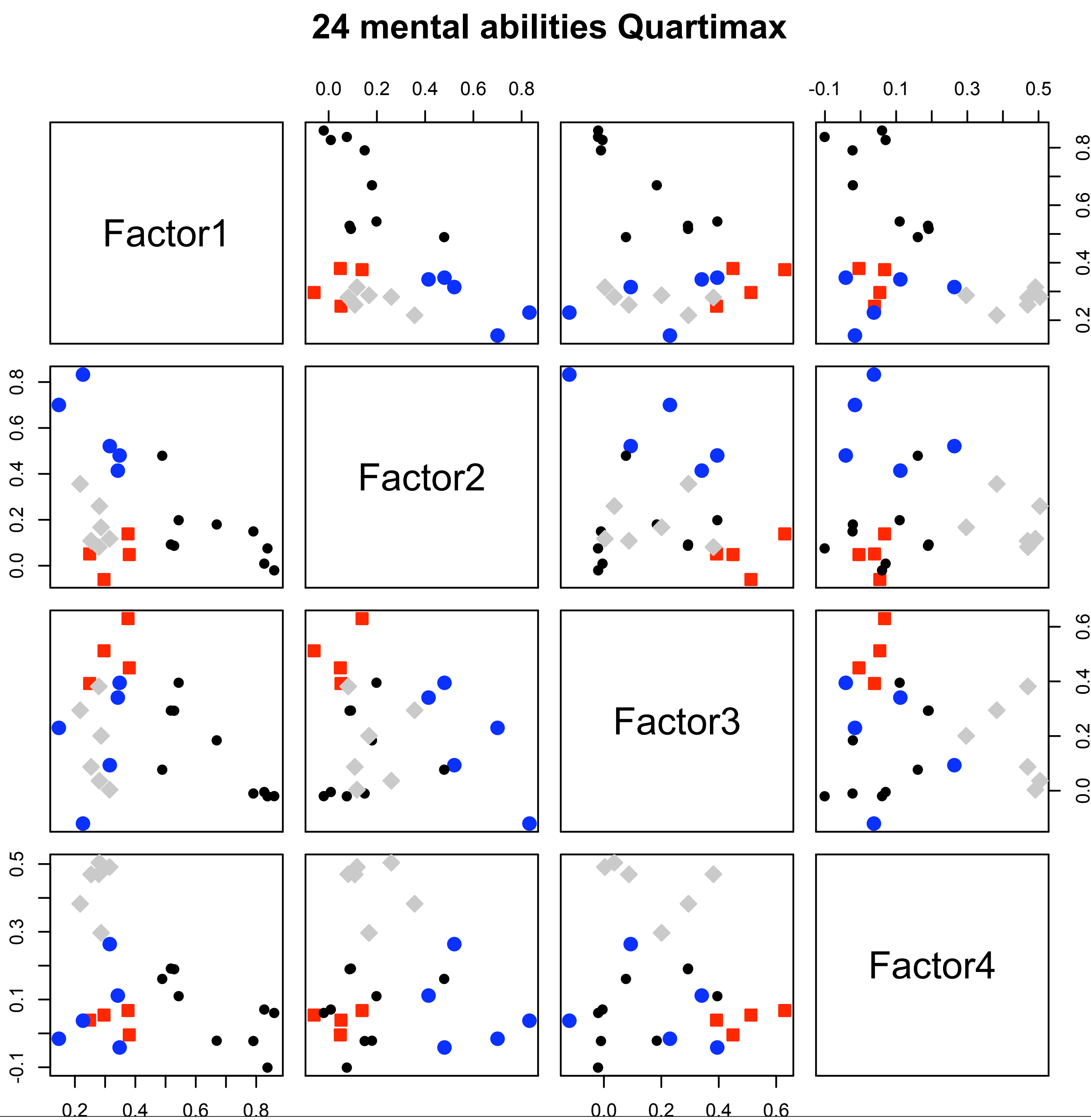
Schmid Leiman transform



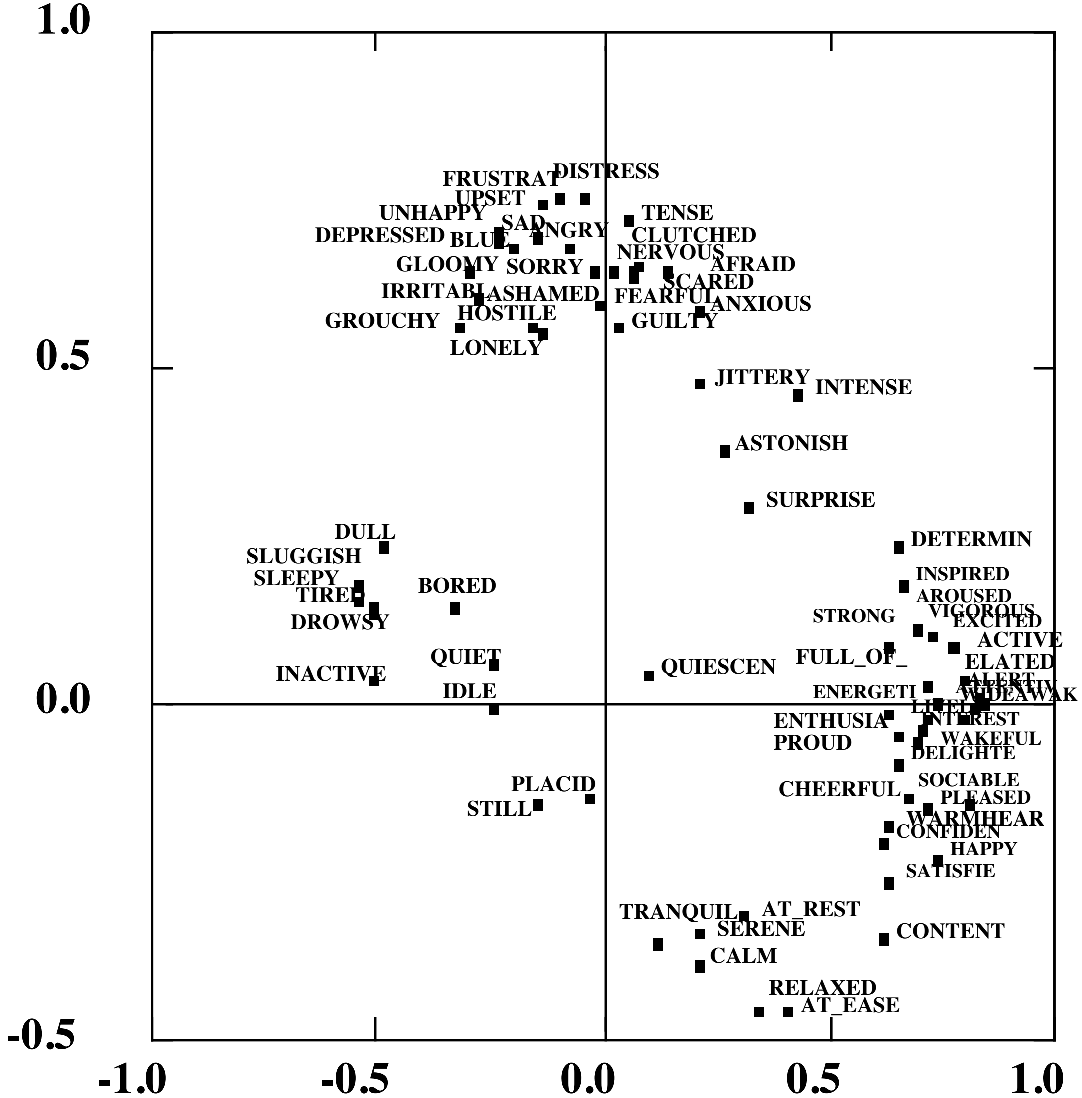
Schmid Leiman

g factor	Factor1	Factor2	Factor3	h2	u2
0.72	3.5e-01	7.3e-07	1.7e-07	0.64	0.36
0.63	3.1e-01	3.1e-07	5.4e-07	0.49	0.51
0.54	2.6e-01	8.3e-08	3.3e-07	0.36	0.64
0.56	3.2e-07	4.2e-01	1.5e-07	0.49	0.51
0.48	3.0e-07	3.6e-01	1.4e-07	0.36	0.64
0.40	3.6e-07	3.0e-01	1.7e-07	0.25	0.75
0.42	4.4e-06	6.3e-06	4.3e-01	0.36	0.64
0.35	2.3e-06	4.4e-06	3.6e-01	0.25	0.75
0.28	2.1e-06	3.6e-06	2.9e-01	0.16	0.84


24 mental abilities Varimax
Factor1
Factor2
Factor3
Factor4
0.0
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5

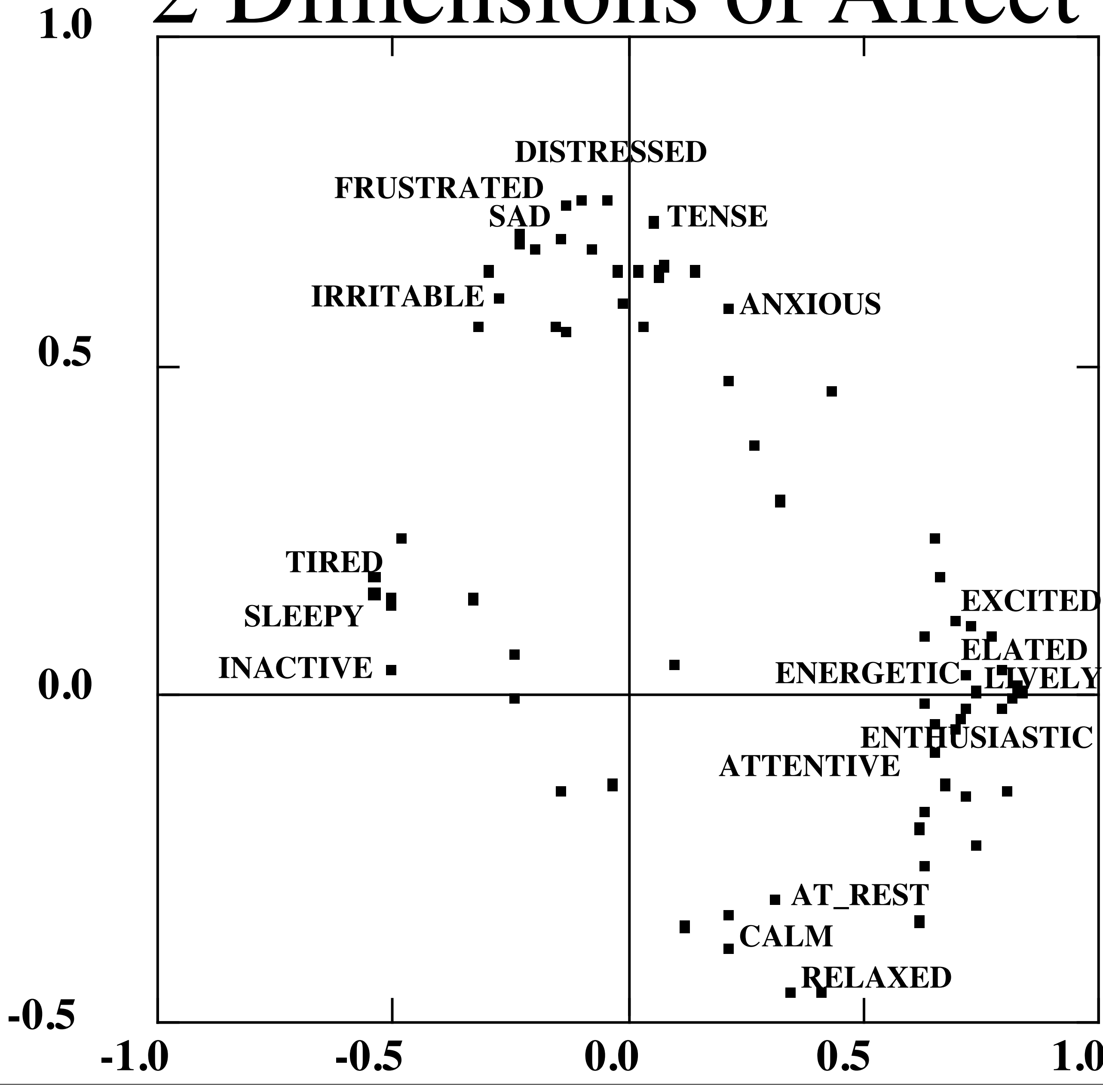

24 mental abilities Quartimax
Factor1
Factor2
Factor3
Factor4
0.0
0.2
0.4
0.6
0.8
-0.1
0.1
0.3
0.5

2 Dimensions of Affect



Energetic Arousal/Positive Affect

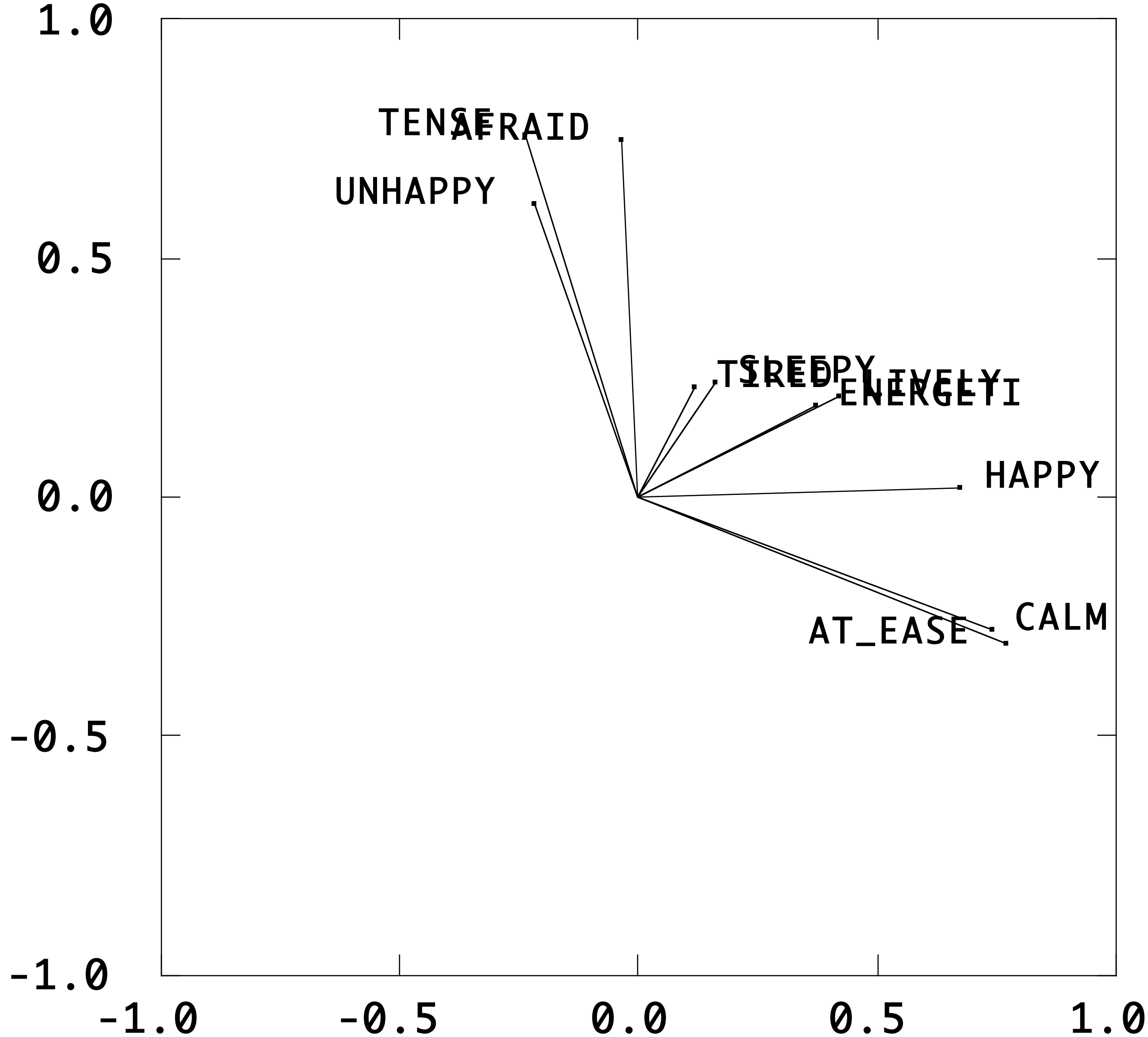
2 Dimensions of Affect



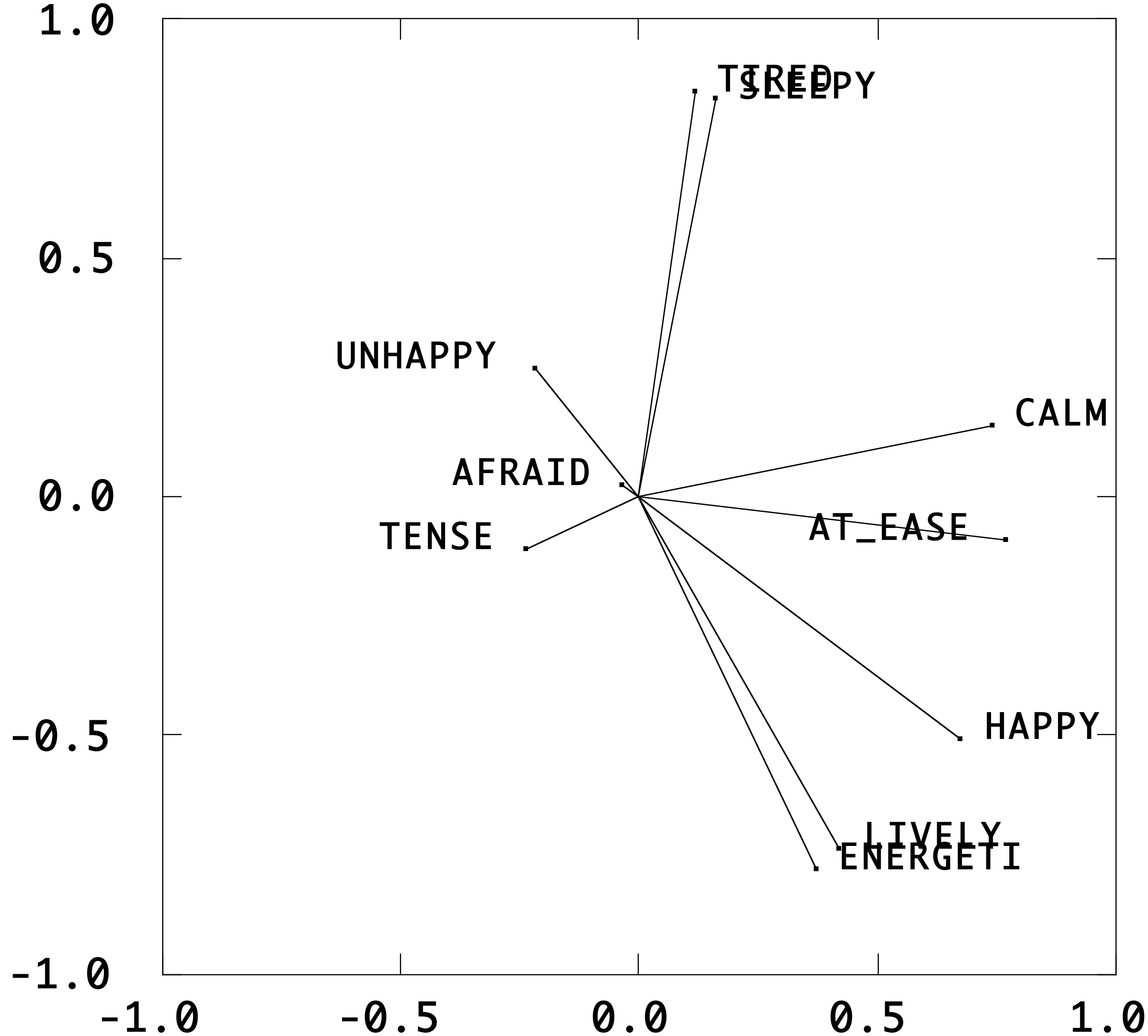
Representative MSQ items (arranged by angular location)

Item	EA-PA	TA-NA	Angle
energetic	**0.8**	0.0	1
elated	**0.7**	0.0	2
excited	**0.8**	0.1	6
anxious	0.2	**0.6**	70
tense	0.1	**0.7**	85
distressed	0.0	**0.8**	93
frustrated	-0.1	**0.8**	98
sad	-0.1	**0.7**	101
irritable	-0.3	**0.6**	114
sleepy	**-0.5**	0.1	164
tired	**-0.5**	0.2	164
inactive	**-0.5**	0.0	177
calm	0.2	**-0.4**	298
relaxed	0.4	**-0.5**	307
at ease	0.4	**-0.5**	312
attentive	**0.7**	0.0	357
enthusiastic	**0.8**	0.0	358
lively	**0.9**	0.0	360

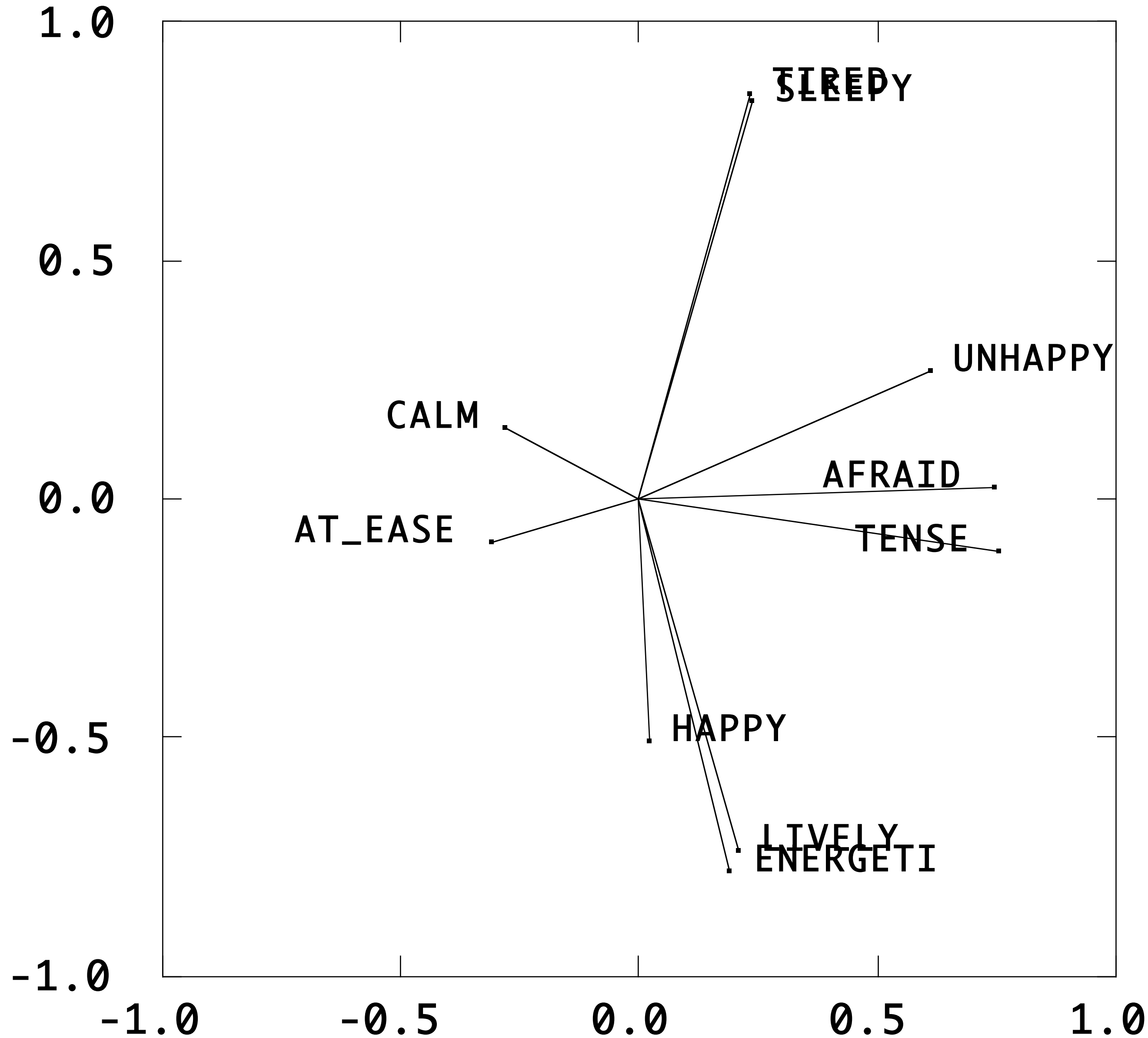
3 Principal Components (1 vs. 2)



3 Principal Components 1 vs. 3



3 Principal Components 2 vs 3



FA and PCA vocabulary

- Eigen values = $\sum(\text{loading}^2)$ across variables = amount of variance accounted for by factor
- Communality = $\sum(\text{loading}^2)$ across factors = amount of variance accounted for in a variable by all the factors
- Rotations versus Transformations
  - Rotations are orthogonal transformations
  - Oblique Transformations

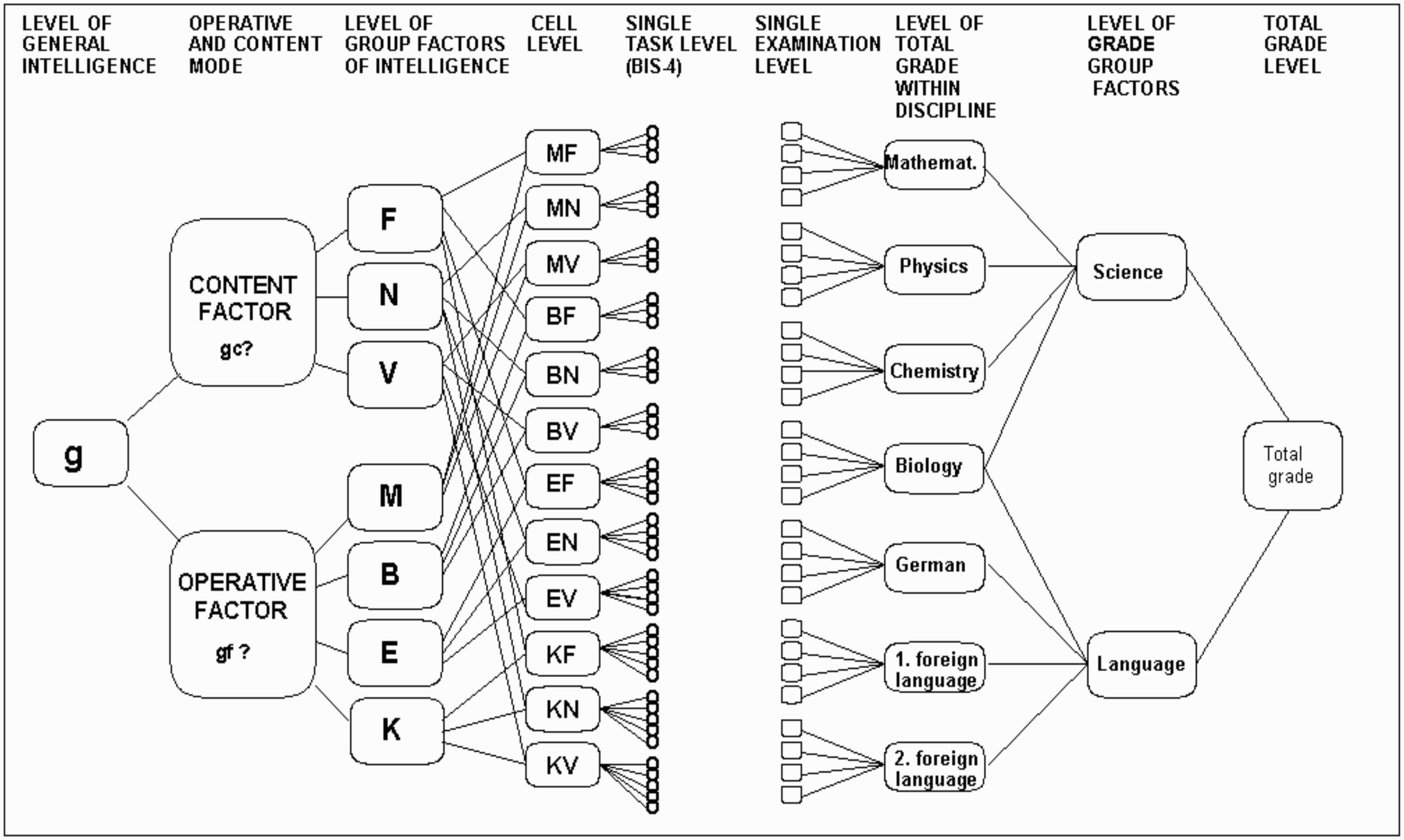
Rotations and transformations

- Simple structure as a criterion for rotation
- Simple structure in the eye of the beholder
- Simple factors (few high, many 0 loadings)
- Simple variables (few high, many 0 loadings)
- VARIMAX, Quartimax, Quartimin
- Procrustes

Rotations and transformations

- Orthogonal rotations
  - Factors are orthogonal, rotated to reduce (or maximize) particular definition of simple structure
- Oblique transformations and higher order factors
  - Allows factors to be correlated (and thus have higher order factors)
  - Many domains have hierarchical structure

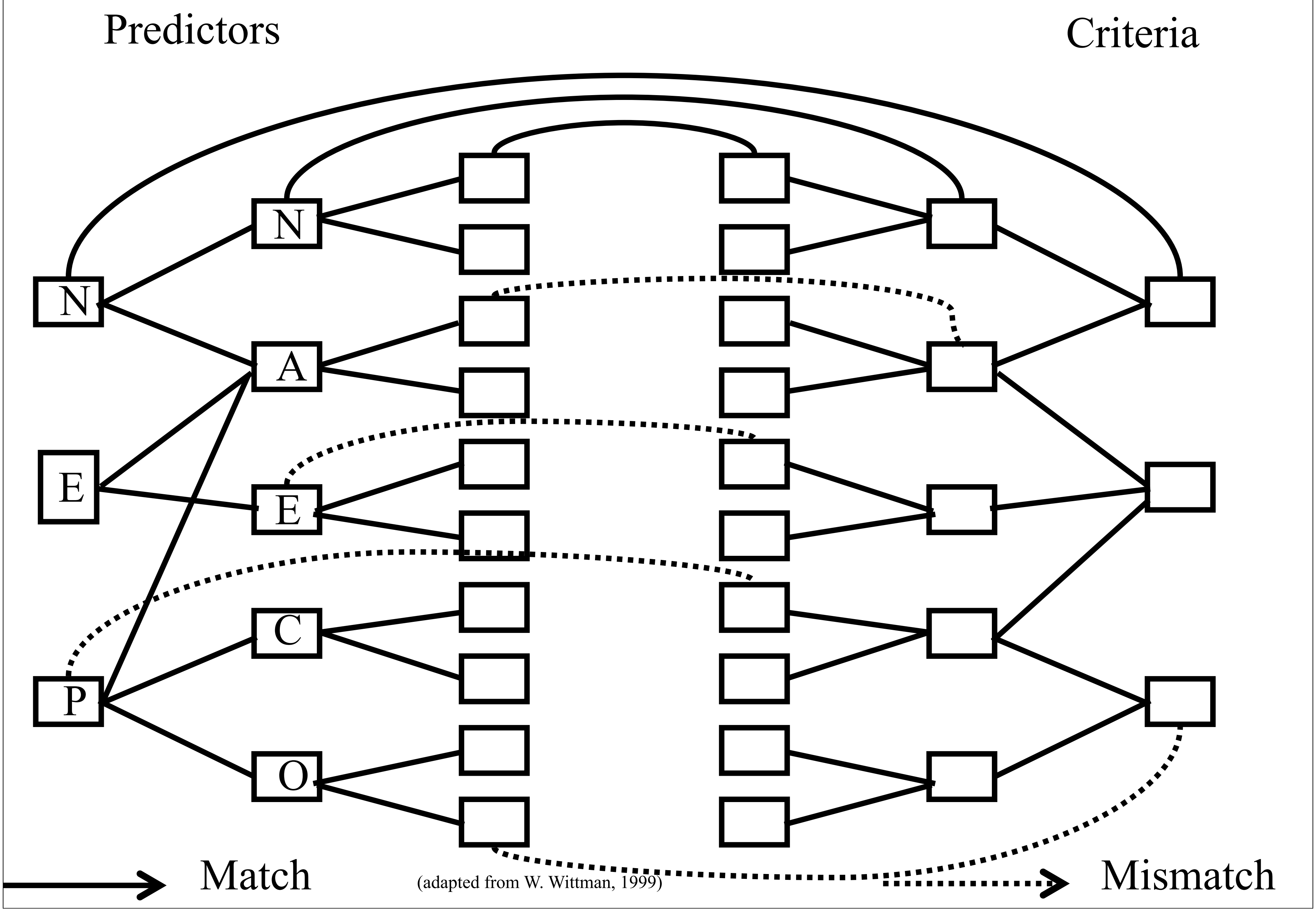
Fig. 9: **Hierarchical version of the Berlin model of intelligence and a grade hierarchy model**



K: Processing capacity for complex information, i.e. reasoning
E: Creativity
B: Speed on relatively simple tasks
M: Memory, i.e. storage capacity for information

F: figural Intelligence
N: numerical Intelligence
V: verbal intelligence

Specificity vs. generality --Matching predictors to outcomes



(adapted from W. Wittman, 1999)

Exploratory versus Confirmatory

- Exploratory:
  - How many factors and best rotation
  - Extraction
    - How many factors?
    - Algorithm for extraction
  - Rotation to simple structure -- what is best SS?
- Confirmatory: does a particular model fit?
  - Apply statistical test of fit
  - But larger N => less fit
  - Is the model the original one or has it been modified?

Components versus Factors

- Components are linear sums of variables and are thus defined at the data level
- Factors represent the covariances of variables and are only estimated (variables are linear sums of factors)
  - Model is undefined at data level, but is defined at structural level
  - Factor indeterminancy problem

Further adventures in hyperspace

I. Orthogonal PC, FA, MLE

II. Oblique FA

III. Cluster analysis

Harman 8 physical variables

	height	arm.span	forearm	lower.leg	weight	bitro.diameter	chest.girth	chest.width
height	1.00	0.85	0.81	0.86	0.47	0.40	0.30	0.38
arm.span	0.85	1.00	0.88	0.83	0.38	0.33	0.28	0.41
forearm	0.81	0.88	1.00	0.80	0.38	0.32	0.24	0.34
lower.leg	0.86	0.83	0.80	1.00	0.44	0.33	0.33	0.36
weight	0.47	0.38	0.38	0.44	1.00	0.76	0.73	0.63
bitro.diameter	0.40	0.33	0.32	0.33	0.76	1.00	0.58	0.58
chest.girth	0.30	0.28	0.24	0.33	0.73	0.58	1.00	0.54
chest.width	0.38	0.41	0.34	0.36	0.63	0.58	0.54	1.00

Principal Components

$values
[1] 4.67 1.77 0.48 0.42 0.23 0.19 0.14 0.10

$fit
[1] 0.98

$loadings

	PC1	PC2
height	0.86	-0.37
arm.span	0.84	-0.44
forearm	0.81	-0.46
lower.leg	0.84	-0.40
weight	0.76	0.52
bitro.diameter	0.67	0.53
chest.girth	0.62	0.58
chest.width	0.67	0.42

Rotated

	PC1	PC2
height	0.90	0.25
arm.span	0.93	0.19
forearm	0.92	0.16
lower.leg	0.90	0.22
weight	0.26	0.88
bitro.diameter	0.19	0.84
chest.girth	0.11	0.84
chest.width	0.26	0.75

Principal Factors

	PA1	PA2
height	0.86	-0.32
arm.span	0.85	-0.41
forearm	0.81	-0.41
lower.leg	0.83	-0.34
weight	0.75	0.57
bitro.diameter	0.63	0.49
chest.girth	0.57	0.51
chest.width	0.61	0.35

	PA1	PA2
height	0.87	0.28
arm.span	0.92	0.21
forearm	0.89	0.19
lower.leg	0.86	0.26
weight	0.23	0.91
bitro.diameter	0.18	0.78
chest.girth	0.12	0.76
chest.width	0.25	0.66

Maximum Likelihood

	Factor1	Factor2
height	0.865	0.287
arm.span	0.927	0.181
forearm	0.895	0.179
lower.leg	0.859	0.252
weight	0.233	0.925
bitro.diameter	0.194	0.774
chest.girth	0.134	0.752
chest.width	0.278	0.621

	Factor1	Factor2
SS loadings	3.335	2.617
Proportion Var	0.417	0.327
Cumulative Var	0.417	0.744

Factor congruence

congruence = sum(XY)/Sqrt(sum(X^2)*sum(Y^2))

	PC1	PC2
PA1	1.00	0.44
PA2	0.49	1.00

	Factor1	Factor2
PA1	1.0	0.47
PA2	0.5	1.00

Correlations of factor loadings ≠ factor congruence

	PC1	PC2
PA1	1.00	-0.99
PA2	-0.96	0.99

	Factor1	Factor2
PA1	1.00	-0.96
PA2	-0.97	1.00

Oblique factors

	PA1	PA2
height	0.88	0.06
arm.span	0.96	-0.03
forearm	0.93	-0.04
lower.leg	0.88	0.04
weight	0.02	0.93
bitro.diameter	0.00	0.80
chest.girth	-0.07	0.80
chest.width	0.10	0.65

Rotating matrix:

	[,1]	[,2]
[1,]	1.10	-0.284
[2,]	-0.26	1.094

Phi:

	[,1]	[,2]
[1,]	1.000	0.467
[2,]	0.467	1.000

Cluster Analysis: a poor person's factor analysis?

I. Clustering as a grouping procedure
   A. membership is -1, 0, 1
II. Clusters as partitions of variable space
III. Hierarchical clusters
   A. top down
   B. bottom up

Hierarchical Clustering

I. Form item proximity matrix (correlations?)

II. Search for most similar pair

III. Combine them (if criteria are reached)

IV. Re-estimate similarities

V. Repeat II-IV until stop criteria

ICLUST: hierarchical clustering of items

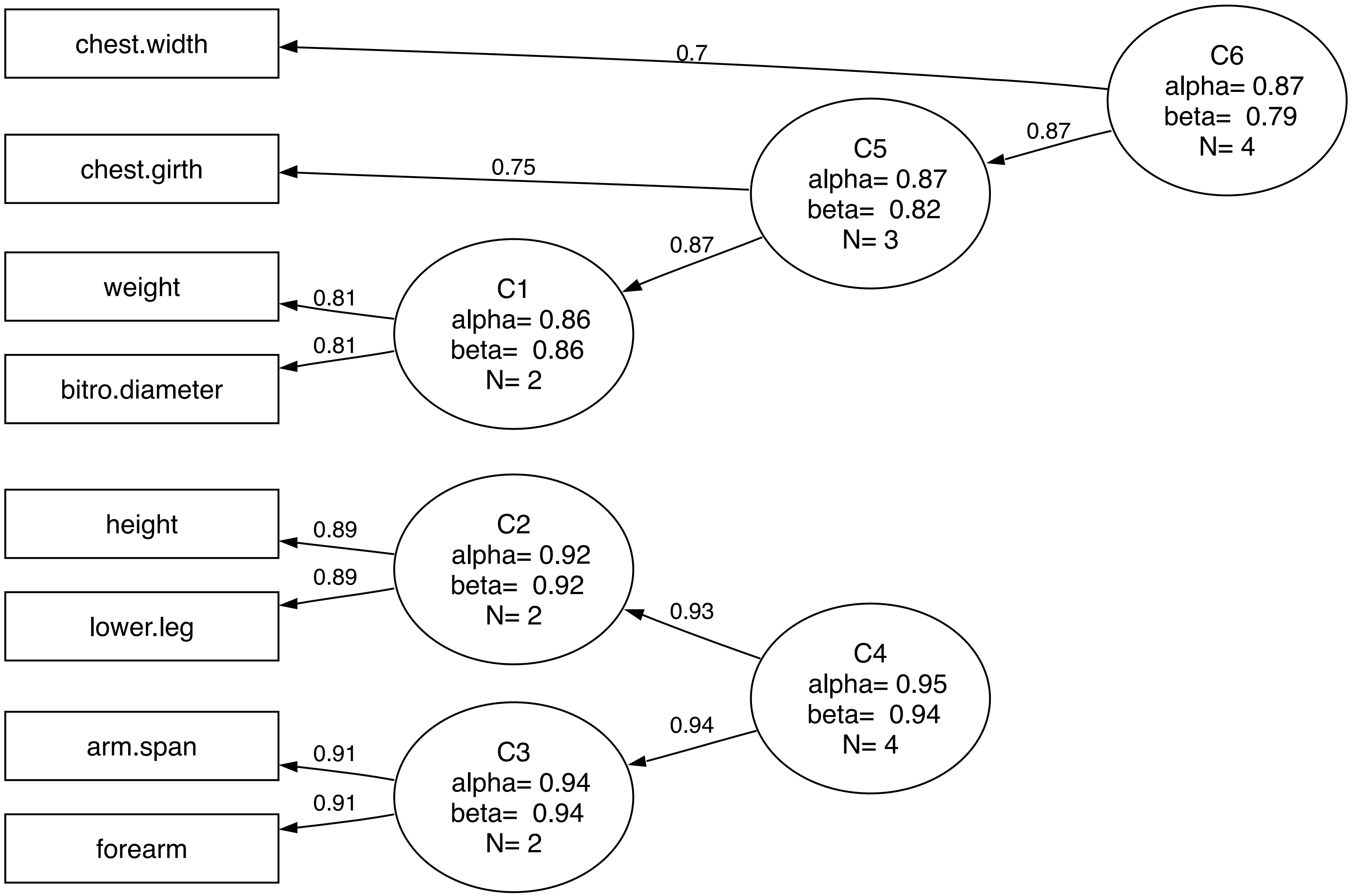
I. Proximities = correlations (corrected for reliability?)

II. Combination rule is matrix addition

III. Criteria are increase in alpha and beta

IV. Implemented in R: ICLUST

Harman - cluster



ICLUST

2 cluster solution

	item	content	cluster	C4	C6
arm.span	2	arm.span	1	0.91	0.41
height	1	height	1	0.89	0.46
forearm	3	forearm	1	0.89	0.38
lower.leg	4	lower.leg	1	0.88	0.43
weight	5	weight	2	0.44	0.83
bitro.diameter	6	bitro.diameter	2	0.37	0.75
chest.girth	7	chest.girth	2	0.30	0.72
chest.width	8	chest.width	2	0.40	0.68

Correlated clusters

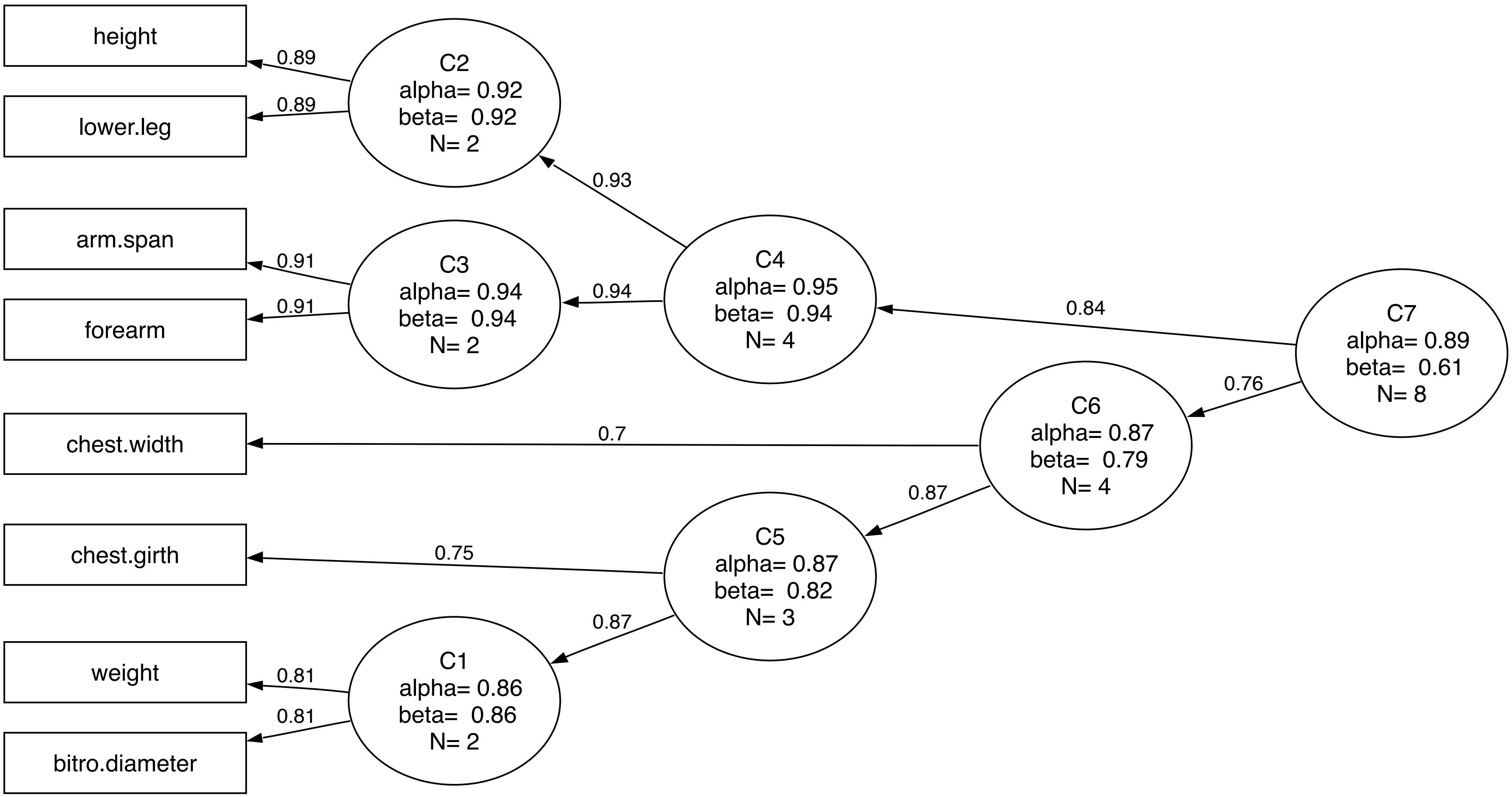
```
$purified$corrected
     C4   C6
C4 0.95 0.49
C6 0.44 0.88
```

```
$p.fit$clusterfit
[1] 0.79

$p.fit$factorfit
[1] 0.86
```

Equivalent to VSS complexity 1

One cluster solution

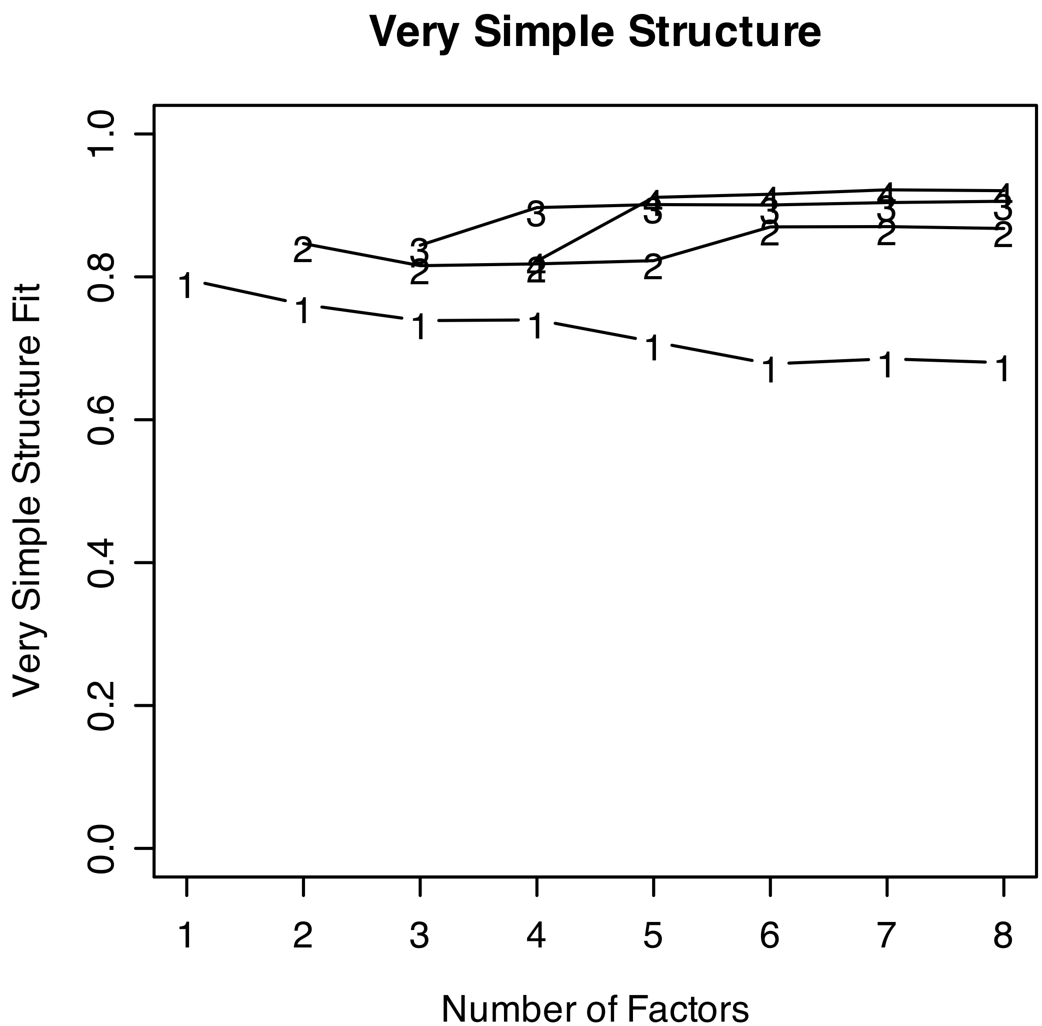


ICLUST

Yet another example

- A correlation matrix of 24 psychological tests given to 145 seventh and eight-grade children in a Chicago suburb by Holzinger and Swineford.
- Harman, H. H. (1976) *Modern Factor Analysis*, Third Edition Revised, University of Chicago Press, Table 7.4
- Factanal (maximum likelihood) suggests 5 factors are required to have non-significant residuals
- But is this overkill?

Holzinger Harmon 24



Oblique

	PA1	PA2	PA3	PA4
VisualPerception	0.041	0.0472	-0.6837	0.054
Cubes	0.049	-0.0229	-0.4526	0.020
PaperFormBoard	0.090	-0.1445	-0.5438	0.051
Flags	0.176	-0.0371	-0.5145	-0.022
GeneralInformation	0.759	0.1196	-0.0129	-0.024
rgraphComprehension	0.797	-0.0550	-0.0043	0.092
SentenceCompletion	0.869	0.0465	-0.0071	-0.107
WordClassification	0.557	0.1134	-0.2153	-0.009
WordMeaning	0.859	-0.0772	0.0168	0.076
Addition	0.076	0.8556	0.1605	0.036
Code	0.069	0.4964	-0.0035	0.295
CountingDots	-0.106	0.6992	-0.2155	-0.025
StraightCurvedCapitals	0.073	0.4635	-0.4163	-0.058
WordRecognition	0.141	-0.0026	0.1120	0.576
NumberRecognition	0.036	-0.0049	0.0056	0.541
FigureRecognition	-0.054	-0.0818	-0.3344	0.527
ObjectNumber	0.039	0.1470	0.0849	0.589
NumberFigure	-0.124	0.2477	-0.2486	0.426
FigureWord	0.066	0.0906	-0.1691	0.318
Deduction	0.327	-0.0404	-0.3119	0.207
NumericalPuzzles	0.070	0.3374	-0.3635	0.106
ProblemReasoning	0.313	-0.0059	-0.3041	0.205
SeriesCompletion	0.297	0.0853	-0.4425	0.102
ArithmeticProblems	0.281	0.4194	-0.0755	0.172

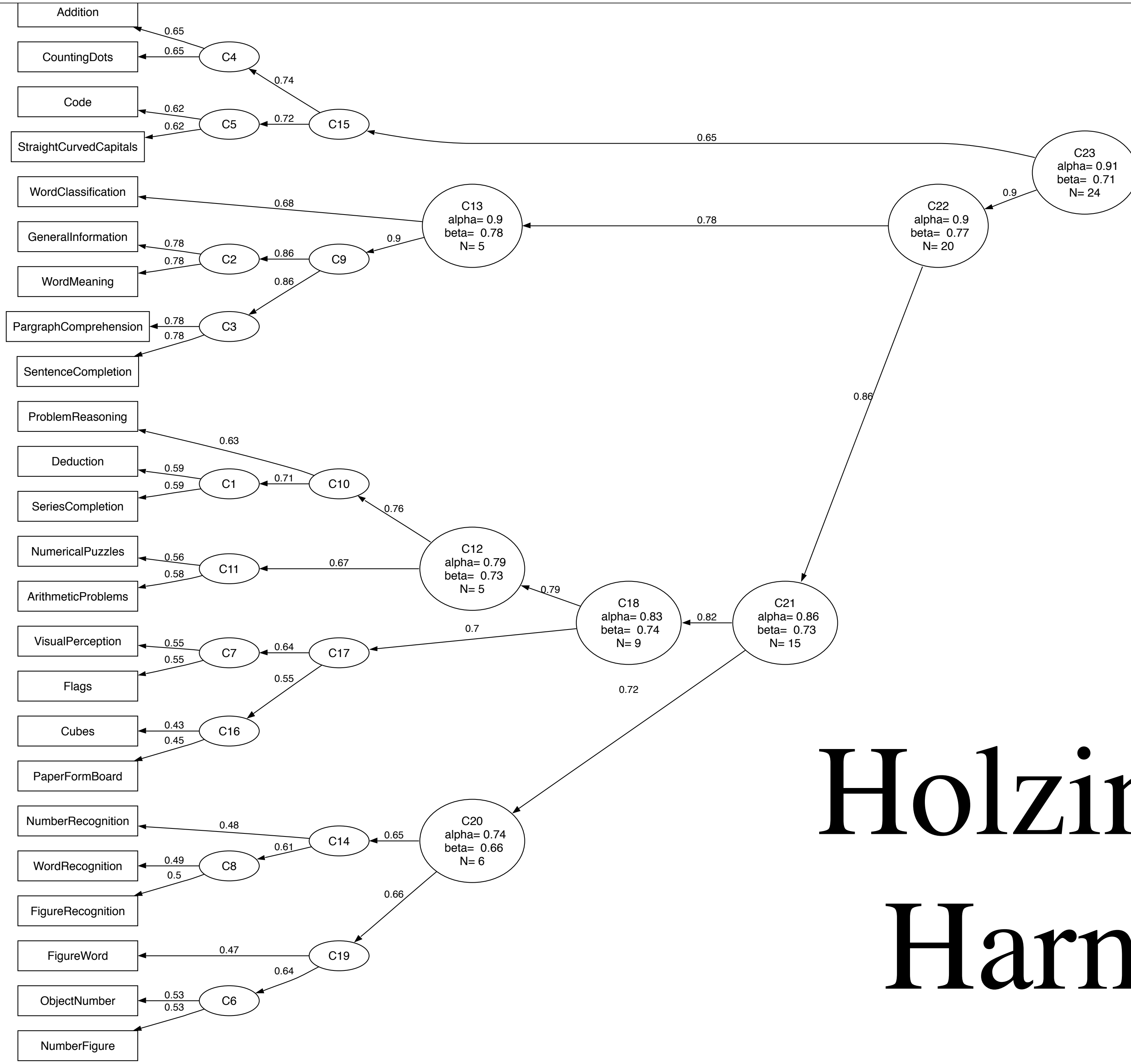
	[,1]	[,2]	[,3]	[,4]
[1,]	1.00	0.30	-0.41	0.40
[2,]	0.30	1.00	-0.27	0.32
[3,]	-0.41	-0.27	1.00	-0.38
[4,]	0.40	0.32	-0.38	1.00

Schmid

	g factor	PA1	PA2	PA3	PA4	h2	u2
VisualPerception	0.50	0.031	0.0417	0.5429	0.042	0.47	0.53
Cubes	0.31	0.037	0.0202	0.3593	0.015	0.21	0.79
PaperFormBoard	0.35	0.068	0.1275	0.4318	0.040	0.33	0.67
Flags	0.40	0.134	0.0327	0.4085	0.017	0.30	0.70
GeneralInformation	0.54	0.576	0.1055	0.0103	0.019	0.59	0.41
rgraphComprehension	0.55	0.605	0.0485	0.0034	0.072	0.65	0.35
SentenceCompletion	0.52	0.660	0.0410	0.0057	0.083	0.77	0.23
WordClassification	0.54	0.423	0.1000	0.1709	0.007	0.37	0.63
WordMeaning	0.56	0.652	0.0681	0.0134	0.059	0.75	0.25
Addition	0.38	0.057	0.7549	0.1274	0.028	0.76	0.24
Code	0.47	0.052	0.4380	0.0028	0.229	0.34	0.66
CountingDots	0.38	0.081	0.6169	0.1711	0.019	0.55	0.45
StraightCurvedCapitals	0.48	0.055	0.4090	0.3306	0.045	0.40	0.60
WordRecognition	0.39	0.107	0.0023	0.0890	0.447	0.36	0.64
NumberRecognition	0.36	0.027	0.0043	0.0044	0.420	0.29	0.71
FigureRecognition	0.46	0.041	0.0722	0.2656	0.409	0.40	0.60
ObjectNumber	0.41	0.030	0.1297	0.0674	0.457	0.38	0.62
NumberFigure	0.46	0.094	0.2185	0.1974	0.330	0.32	0.68
FigureWord	0.39	0.050	0.0799	0.1342	0.247	0.14	0.86
Deduction	0.51	0.248	0.0357	0.2476	0.160	0.25	0.75
NumericalPuzzles	0.49	0.053	0.2977	0.2886	0.082	0.26	0.74
ProblemReasoning	0.51	0.237	0.0052	0.2415	0.159	0.23	0.77
SeriesCompletion	0.57	0.225	0.0753	0.3514	0.079	0.30	0.70
ArithmeticProblems	0.53	0.213	0.3700	0.0600	0.134	0.29	0.71

$omega
[1] 0.64

$alpha
[1] 0.91

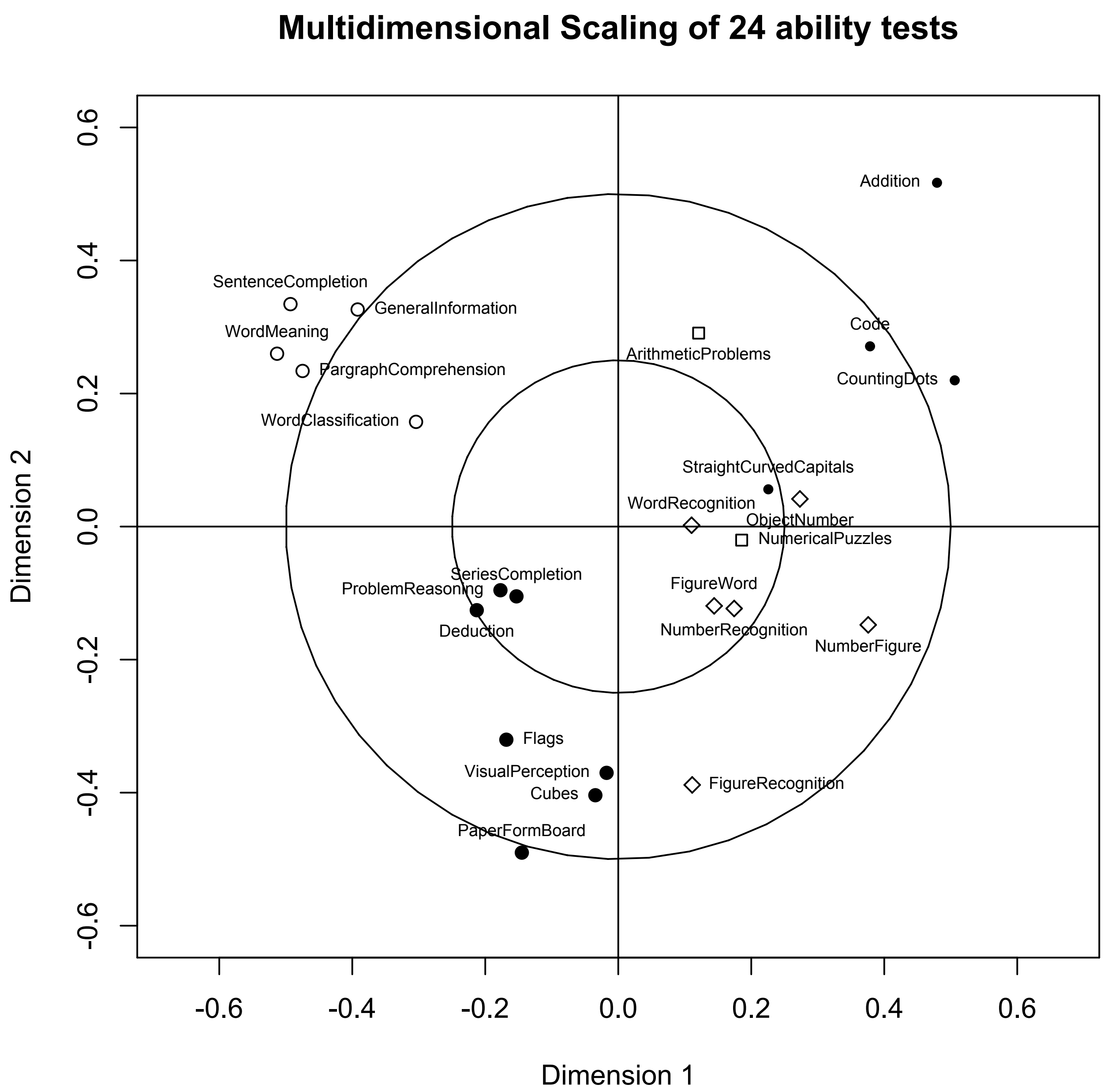


ICLUST of Harman's 24 mental variables

Holzinger-Harmon

Multidimensional scaling

- MDS will remove the first factor because it is grouping by similarity and ignores mean level of similarity
- Representations of distances in MDS can match those of a factor space, but distance from the center reflects the general factor loading
- distances from correlations
  - d = sqrt(2(1-r))


Multidimensional Scaling of 24 ability tests
Addition
SentenceCompletion
GeneralInformation
WordMeaning
PargraphComprehension
WordClassification
ArithmeticProblems
Code
CountingDots
StraightCurvedCapitals
WordRecognition
ObjectNumber
NumericalPuzzles
SeriesCompletion
ProblemReasoning
Deduction
FigureWord
NumberRecognition
NumberFigure
Flags
VisualPerception
Cubes
FigureRecognition
PaperFormBoard
Dimension 1
Dimension 2
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6

Threats to interpreting correlation and the benefits of structure

- Correlations can be attenuated due to differences in skew (and error)
- Bi polar versus unipolar scales (e.g., of affect)
- How happy do you feel?
  - Not at all A little Somewhat Very
- How sad do you feel?
  - Not at all A little Somewhat Very
- How do you feel?

very sad sad happy very happy

Unipolar scales allow data to speak for themselves

Simulated Example of unipolar scales and the problem of skew

- Consider X and Y as random normal variables
- Let X+ = X if X>0, 0 elsewhere
- Let X- = X if X<0, 0 elsewhere
- Reversed (X+) = - X+
- Similarly for Y
- Examine the correlational structure
- Note that although X and -X correlate -1.0, X+ and X- correlate only -.43 and that X+ correlates with X+Y+ .66

Determining Structure: zeros and skew

	X+	X-	Y+	Y-	X+Y+	X-Y-	X+Y-	X-Y+
X+	1.00							
X-	-0.47	1.00						
Y+	0.03	-0.01	1.00					
Y-	0.00	-0.03	-0.46	1.00				
X+Y+	0.65	-0.39	0.66	-0.39	1.00			
X-Y-	-0.40	0.63	-0.40	0.63	-0.46	1.00		
X+Y-	0.63	-0.40	-0.39	0.66	0.00	0.02	1.00	
X-Y+	-0.39	0.64	0.63	-0.40	0.00	0.00	-0.47	1.00

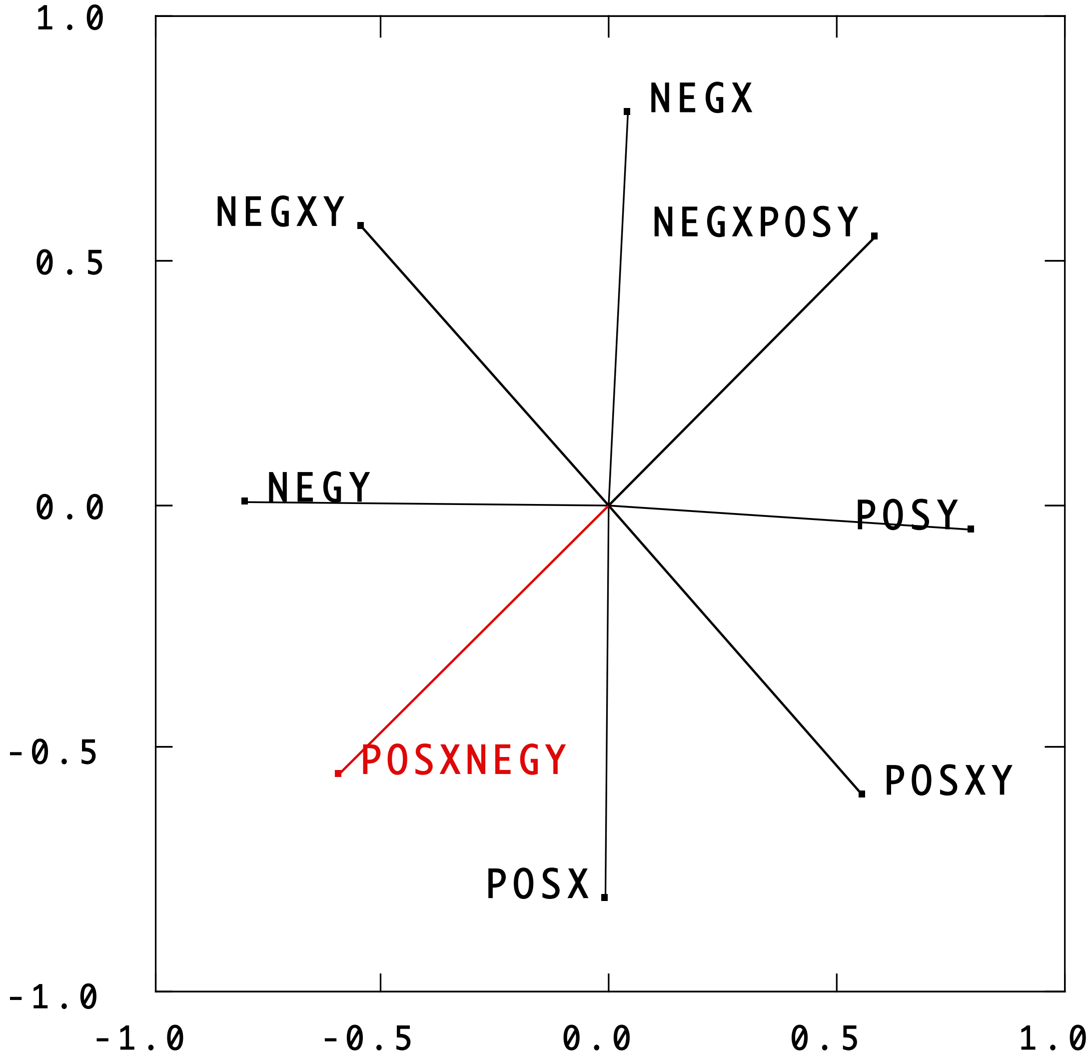
Skew and zeros: determining structure

	Y+	X+Y+	X+	X+Y-	r(Y-)	r(X-Y-)	r(X-)	X-Y+
Y	1.00							
X+Y+	0.66	1.00						
X+	0.03	0.65	1.00					
X+Y-	-0.39	0.00	0.63	1.00				
r(Y-)	0.46	0.39	0.00	-0.66	1.00			
r(X-Y-)	0.40	0.46	0.40	-0.02	0.63	1.00		
r(X-)	0.01	0.39	0.47	0.40	-0.03	0.63	1.00	
X-Y+	0.63	0.00	-0.39	-0.47	0.40	0.00	-0.64	1.00

Factor analysis shows structure

	1	2	
POSX	-0.01	-0.81	271
NEGX	0.04	0.80	87
POSY	0.80	-0.05	356
NEGY	-0.80	0.00	180
POSXY	0.56	-0.59	314
NEGXY	-0.55	0.57	136
POSXN	-0.59	-0.55	227
NEGXP	0.58	0.55	43

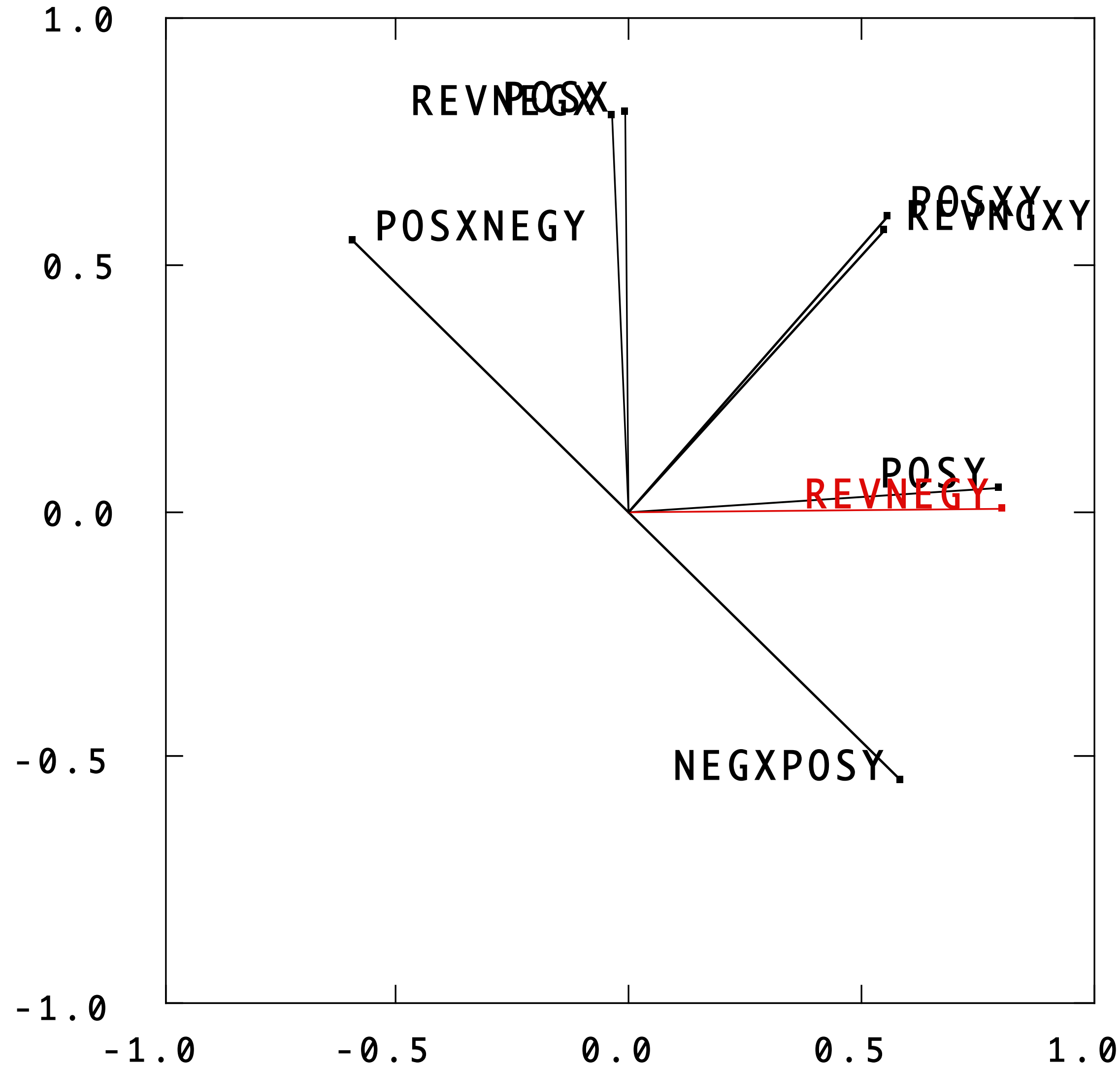
Structural representation



Factor analysis shows structure

	1	2	angle
POSX	-0.01	0.81	89
POSY	0.80	0.05	4
POSXY	0.56	0.59	46
POSXN	-0.59	0.55	137
NEGXP	0.58	-0.55	313
REVNE	-0.04	0.80	87
REVNE	0.80	0.00	90
REVNG	0.55	0.57	46

Factor analysis shows structure



Hyperplanes and Zeros: Defining variables by what they are not

- Tendency to interpret variables in terms of their patterns of high correlations
- But large correlations may be attenuated by skew or error
- Correlation of .7 is an angle of 45 degrees => lots of room for misinterpretation!
- Zero correlations reflect what a variable is not
  - Zeros provide definition by discriminant validity

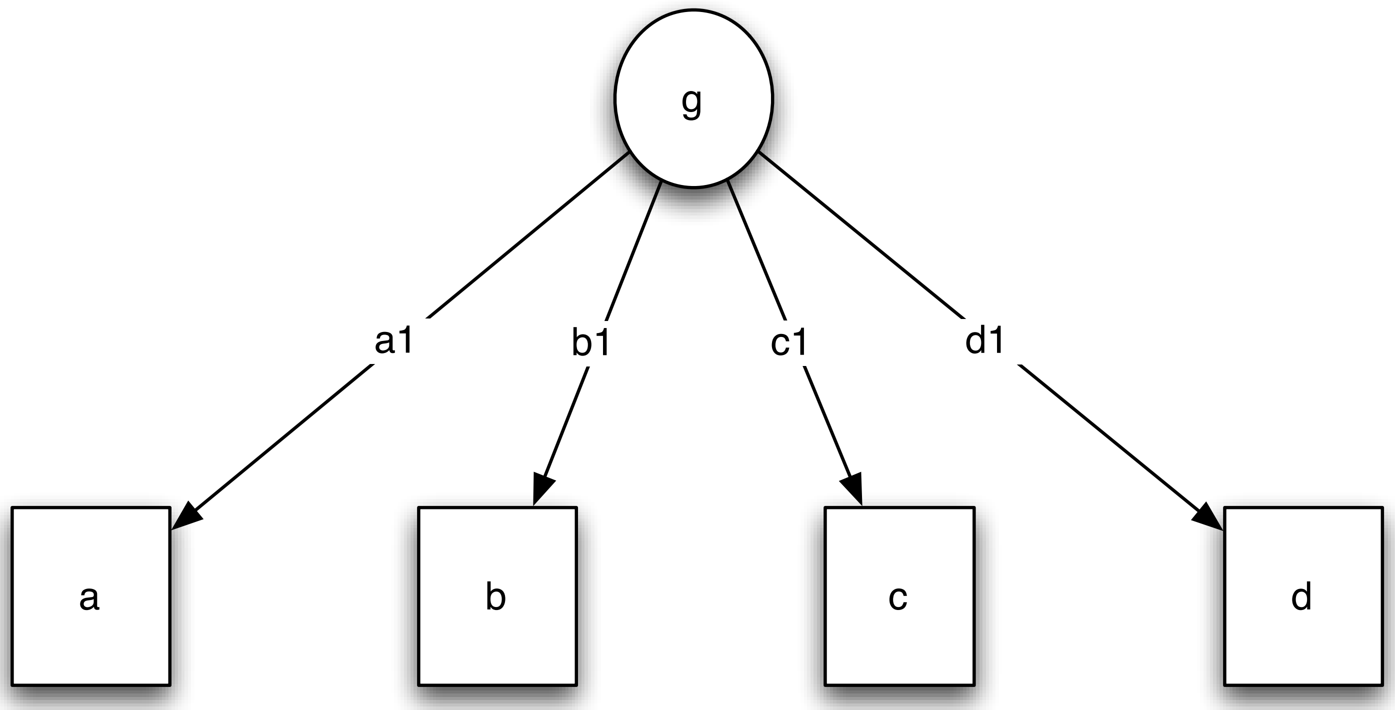
Structural Equation Models

- Estimating factor model of reliability (the measurement model
- Estimates of validity model corrected for attenuation (the structural model)
- Maximum likelihood solutions of covariance matrices

correlation matrix

	A	B	C	D
A	1.00	0.30	0.20	0.10
B	0.30	1.0	0.20	0.20
C	0.20	0.20	1.00	0.30
D	0.10	0.20	0.30	1.00

Hypothetical structure

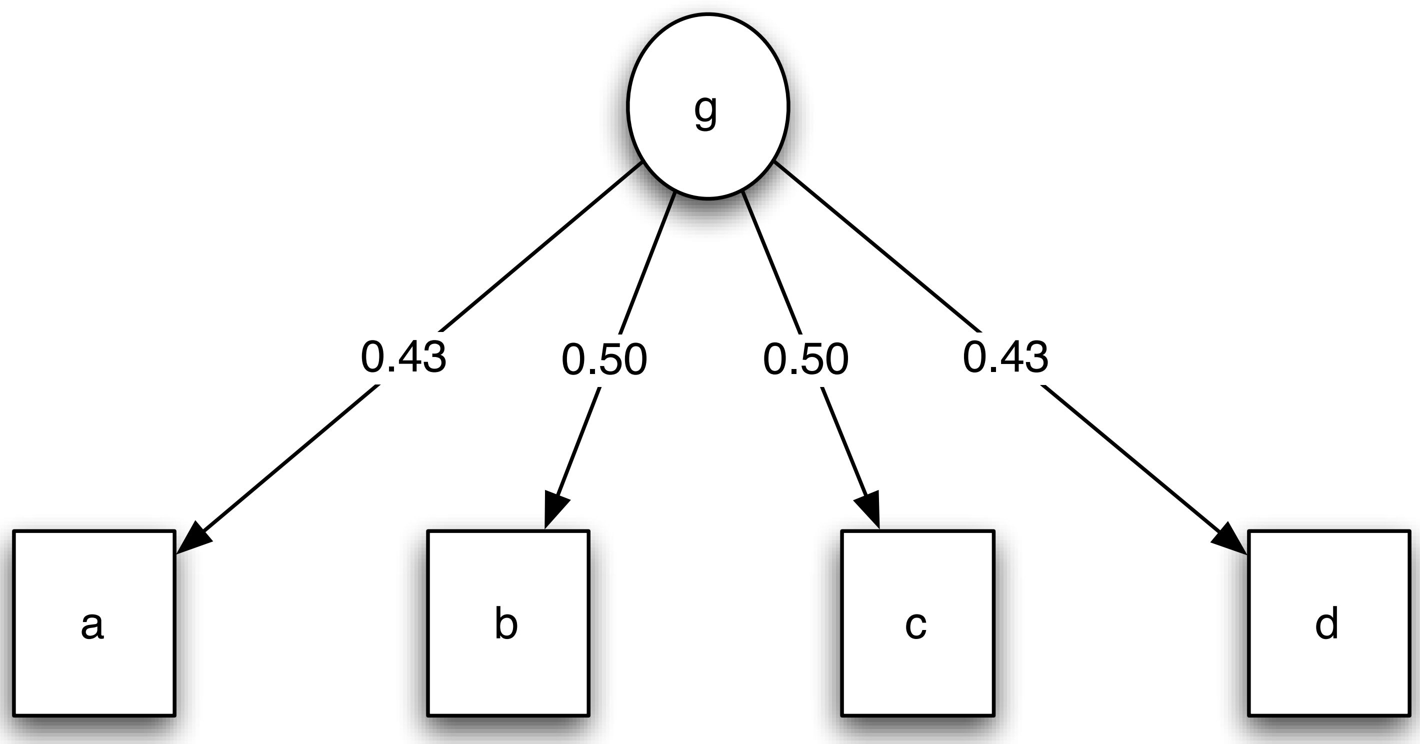


R for sem

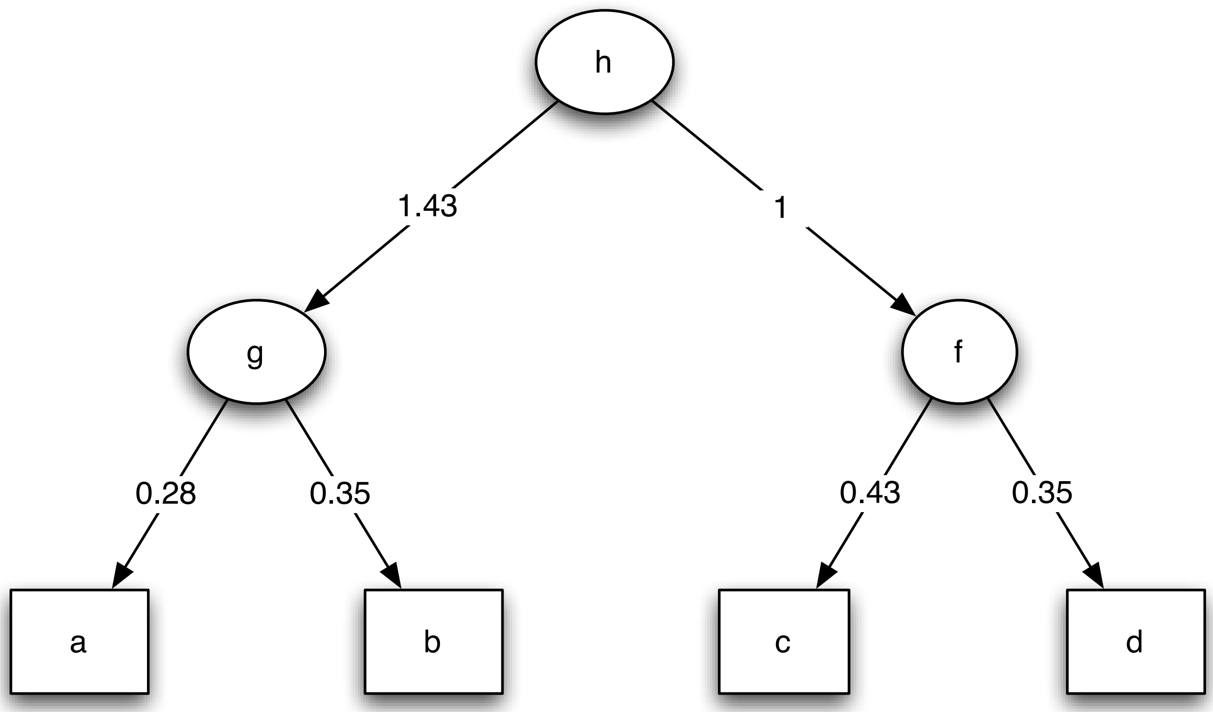
```
obs.var2.12 = c('a',  'b',  'c',  'd')
  R.prob2.12 = matrix(c(
  1.00 ,  .30,    .20,   .10,
   .30,  1.00,   .20,   .20,
   .20,   .20,  1.00,   .30,
   .10,   .20,   .30,  1.00),
   ncol=4,byrow=TRUE)

  model2.12a=matrix(c(
        'g ->  a',      'a1', NA,
        'g -> b' ,      'b1', NA,
        'g -> c' ,      'c1', NA,
        'g -> d',     'd1', NA,
        'a <-> a',      'e1', NA,
        'b <-> b',      'e2', NA,
        'c <-> c',      'e3', NA,
        'd <-> d',      'e4', NA,
        'g <-> g',       NA, 1),
        ncol=3, byrow=TRUE)
  sem2.12a= sem(model2.12a,R.prob2.12,120, obs.var2.12)
 summary(sem2.12a,digits=3)
```

SEM of Congeneric



an alternative model



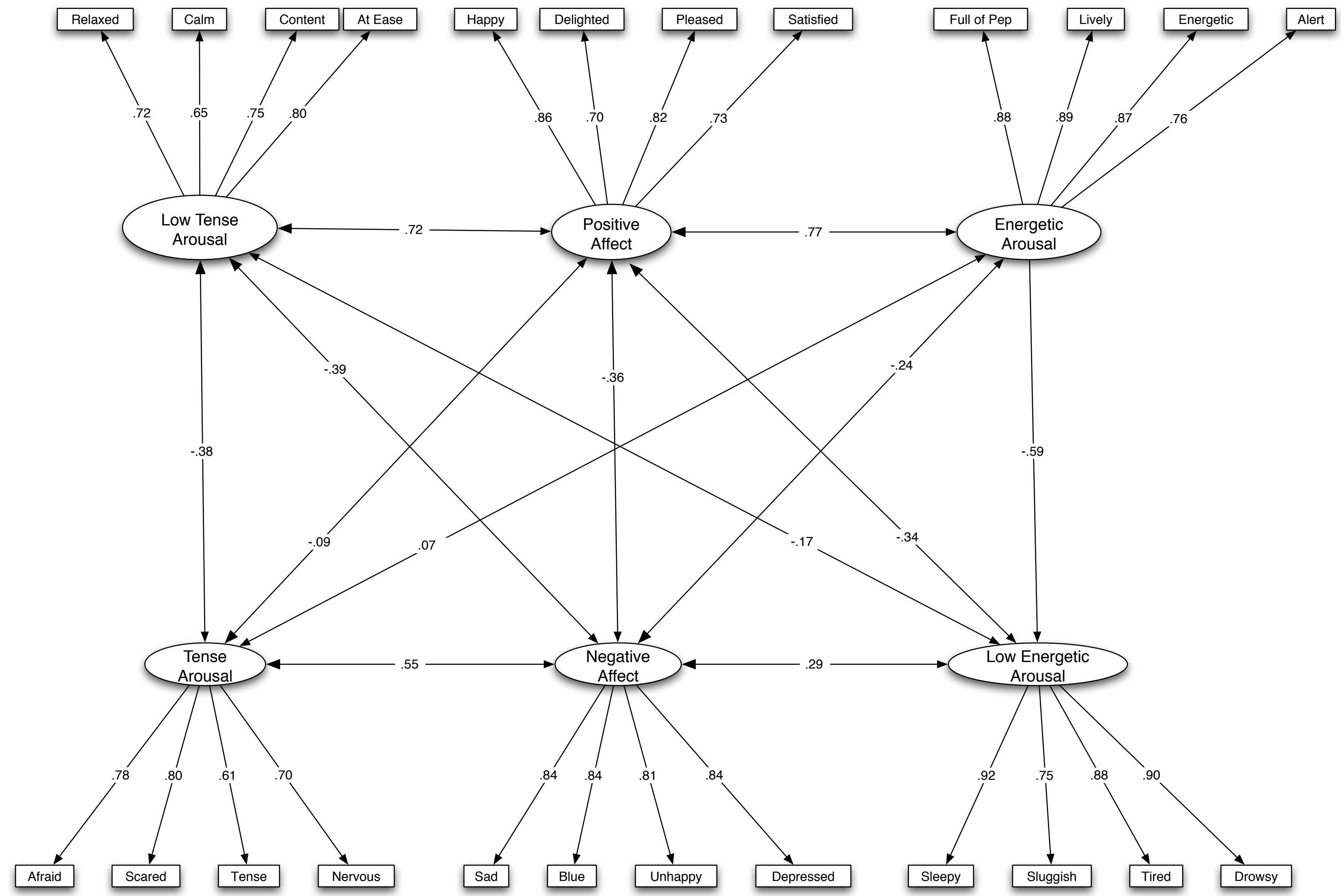
Structure of Affect

I. Is happy the opposite of sad?

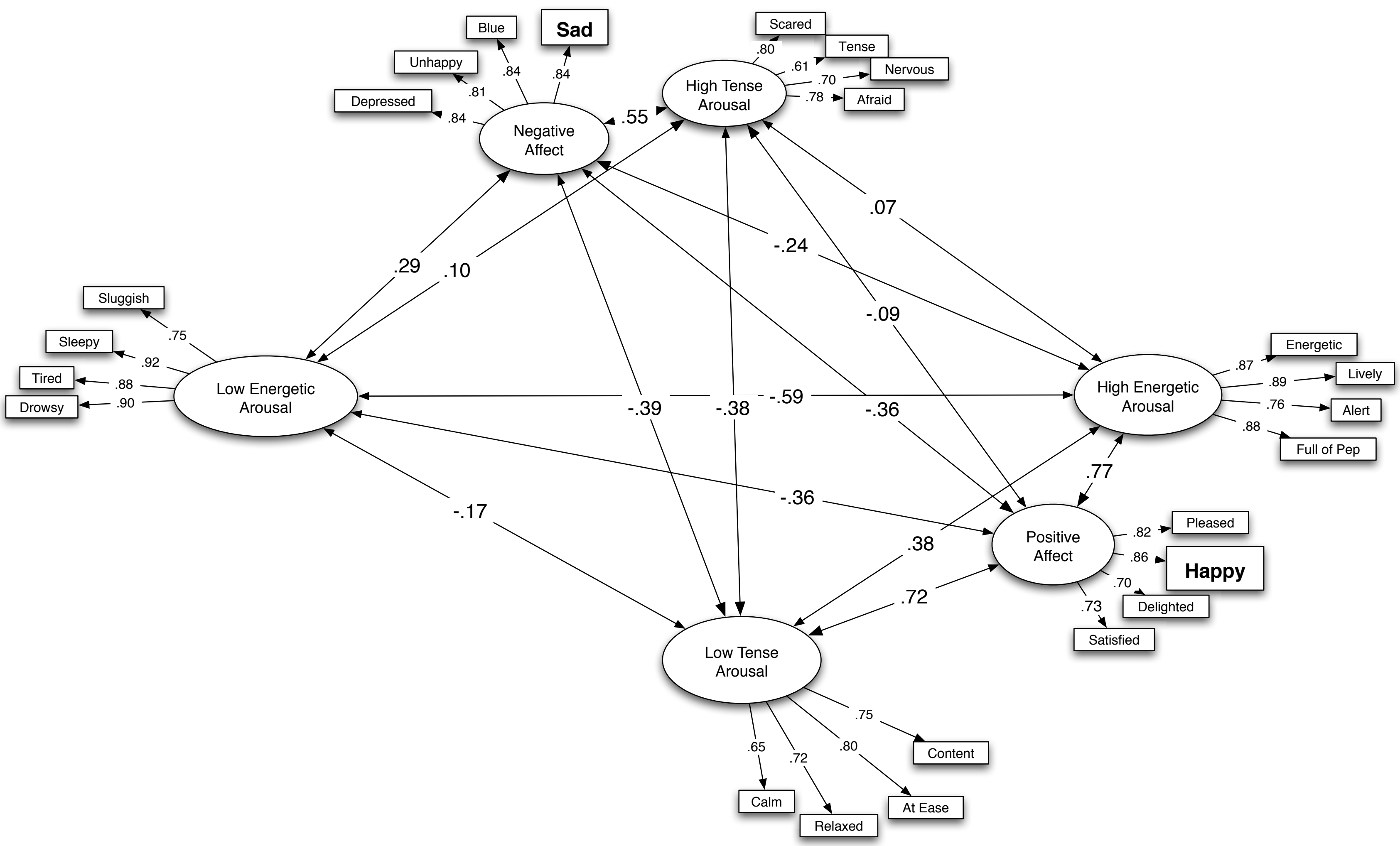
II. Is Positive Affect = - Negative Affect

III. What are the dimensions of Affect

IV. 75 affect words collected over multiple studies for > 3800 subjects


Relaxed
Calm
Content
At Ease
Happy
Delighted
Pleased
Satisfied
Full of Pep
Lively
Energetic
Alert
.72
.65
.75
.80
.86
.70
.82
.73
.88
.89
.87
.76
Low Tense Arousal
Positive Affect
Energetic Arousal
.72
.77
-.39
-.36
-.24
-.38
-.59
-.09
.07
-.17
-.34
Tense Arousal
Negative Affect
Low Energetic Arousal
.55
.29
.78
.80
.61
.70
.84
.84
.81
.84
.92
.75
.88
.90
Afraid
Scared
Tense
Nervous
Sad
Blue
Unhappy
Depressed
Sleepy
Sluggish
Tired
Drowsy

Structure of Affect



Comparing regression to latent modeling

I. Regression weights change as a function of number of correlated predictors

II. Path weights do not change

III. Consider the following 6 predictors of 1 criterion

6 predictors - 1 criterion

R

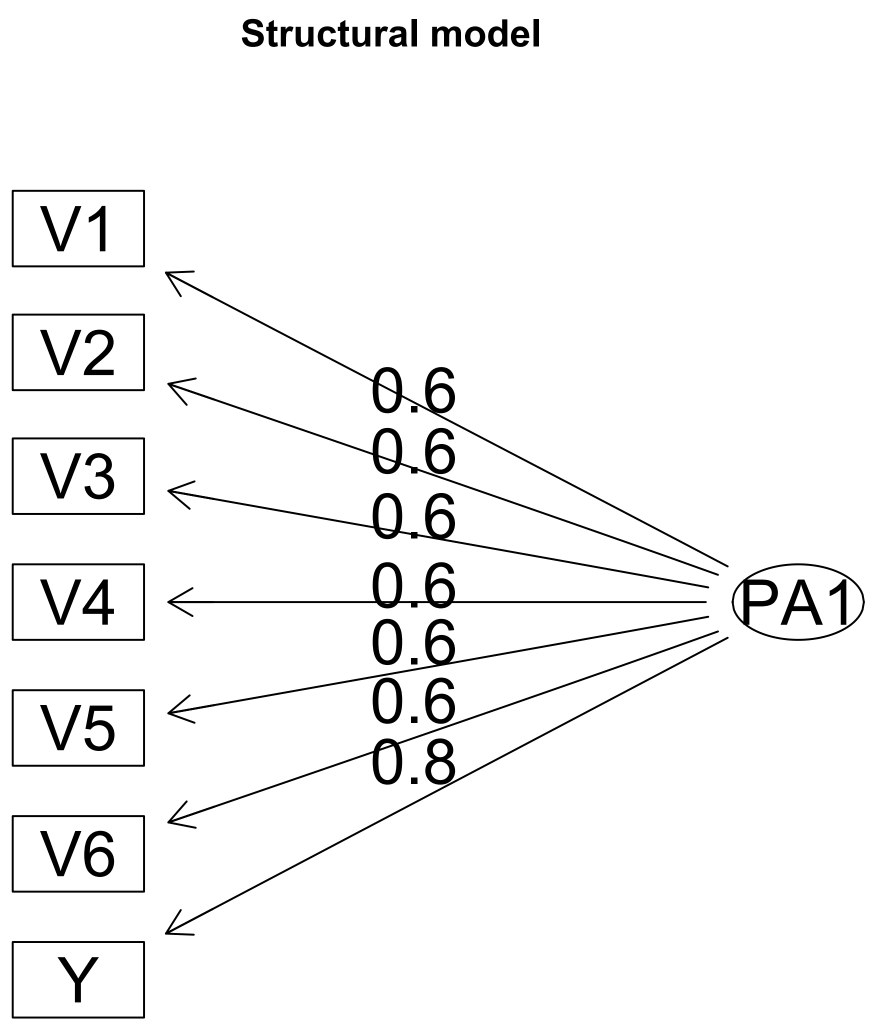
	V1	V2	V3	V4	V5	V6	Y
V1	1.00	0.36	0.36	0.36	0.36	0.36	0.48
V2	0.36	1.00	0.36	0.36	0.36	0.36	0.48
V3	0.36	0.36	1.00	0.36	0.36	0.36	0.48
V4	0.36	0.36	0.36	1.00	0.36	0.36	0.48
V5	0.36	0.36	0.36	0.36	1.00	0.36	0.48
V6	0.36	0.36	0.36	0.36	0.36	1.00	0.48
Y	0.48	0.48	0.48	0.48	0.48	0.48	1.00

A one factor model

```
> f
  V PA1
1 1 0.6
2 2 0.6
3 3 0.6
4 4 0.6
5 5 0.6
6 6 0.6
7 7 0.8
               PA1
SS loadings     2.8
Proportion Var 0.4
Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the model is 14  and the fit was  0
```

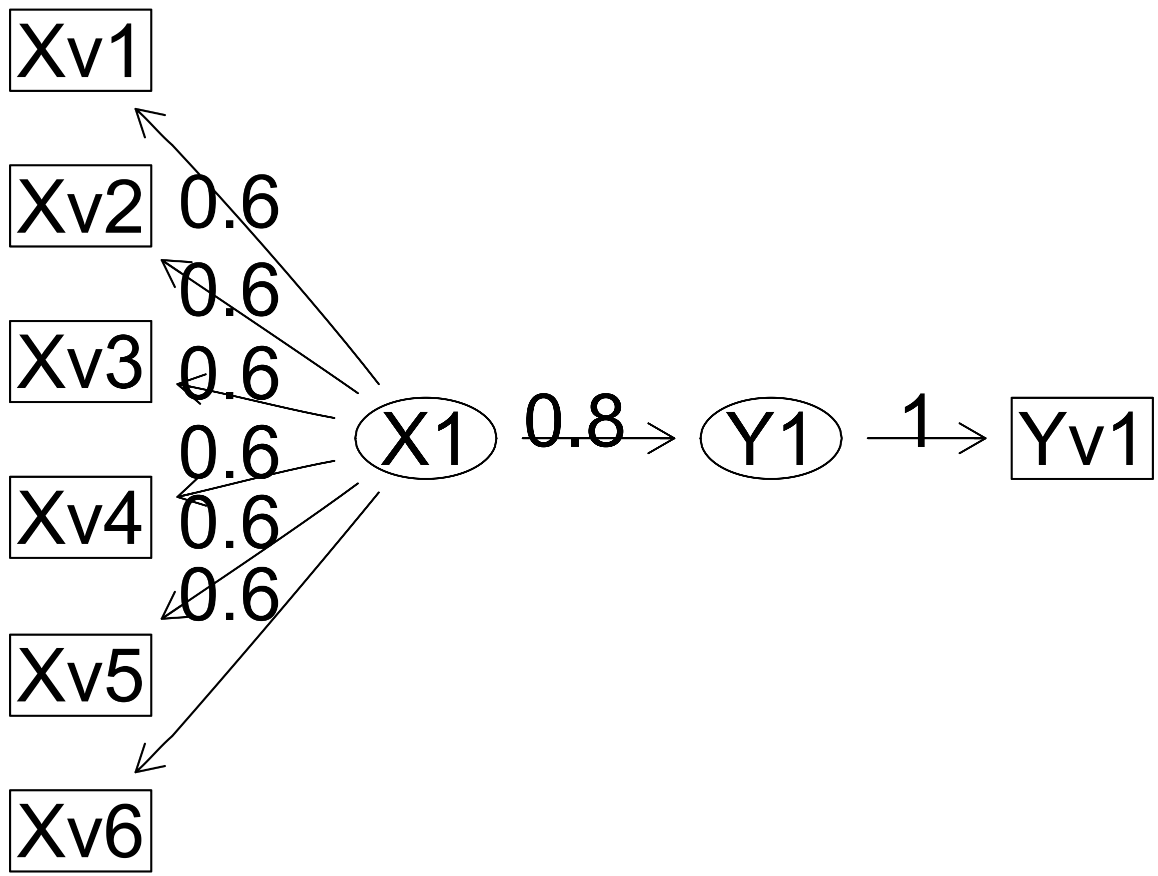
Congeneric factor model

Congeneric factor model accounts for all the observed correlations



Structural model

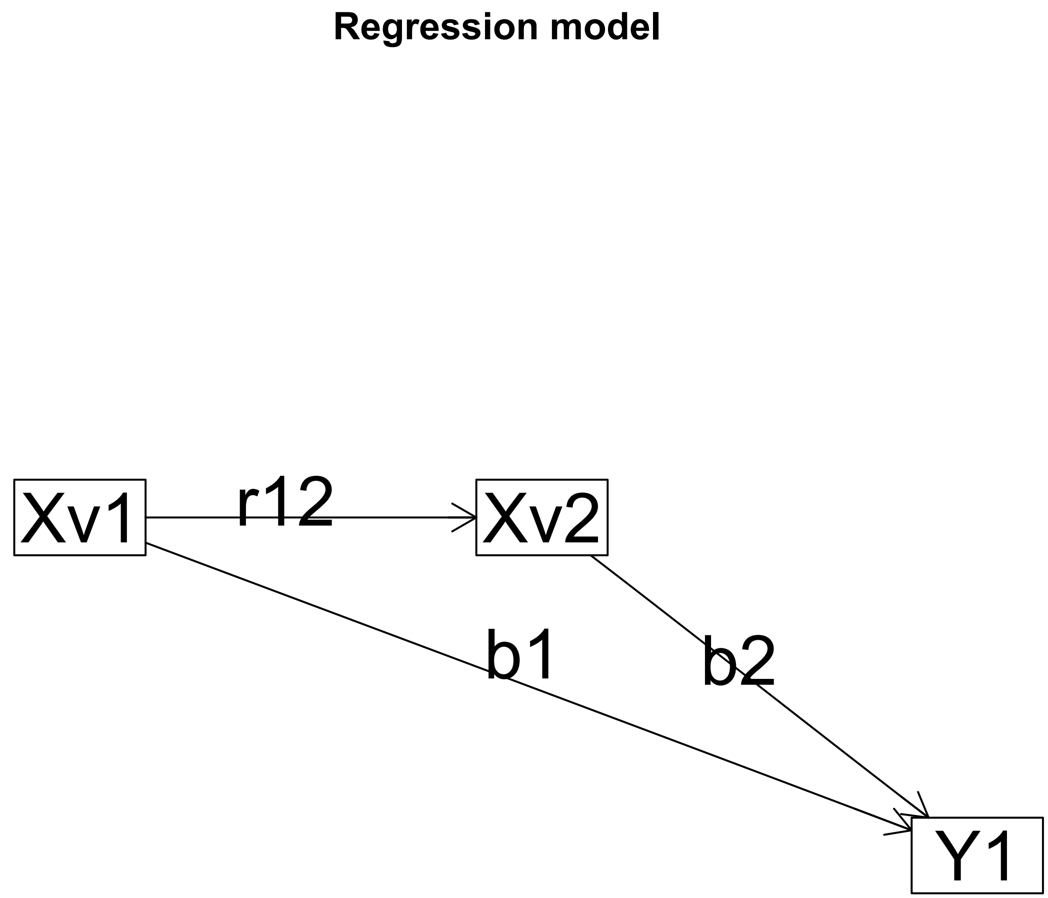
All correlations are modeled



Regression model has direct and indirect paths

$r_{x1y} = b_1 + r_{12}b_2$

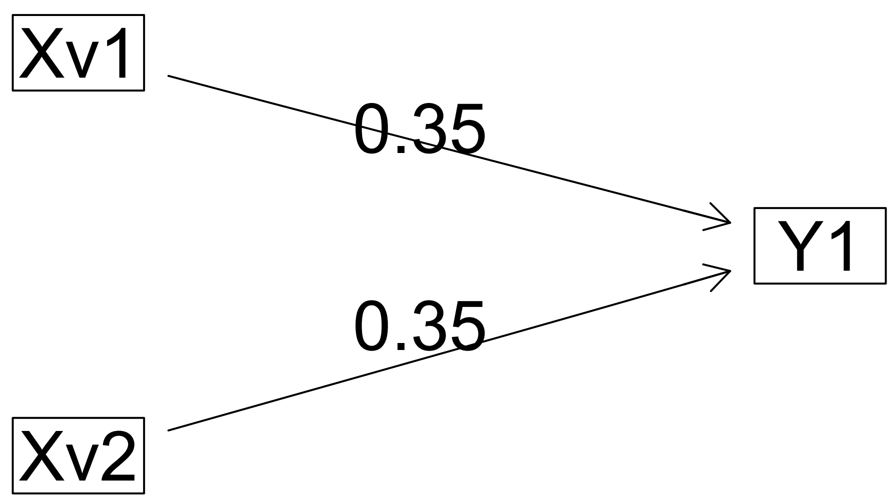
$r_{x2y} = b_2 + r_{12}b_1$



But, consider regression R = .58

Regression model

Correlations of .36 between predictors are not shown

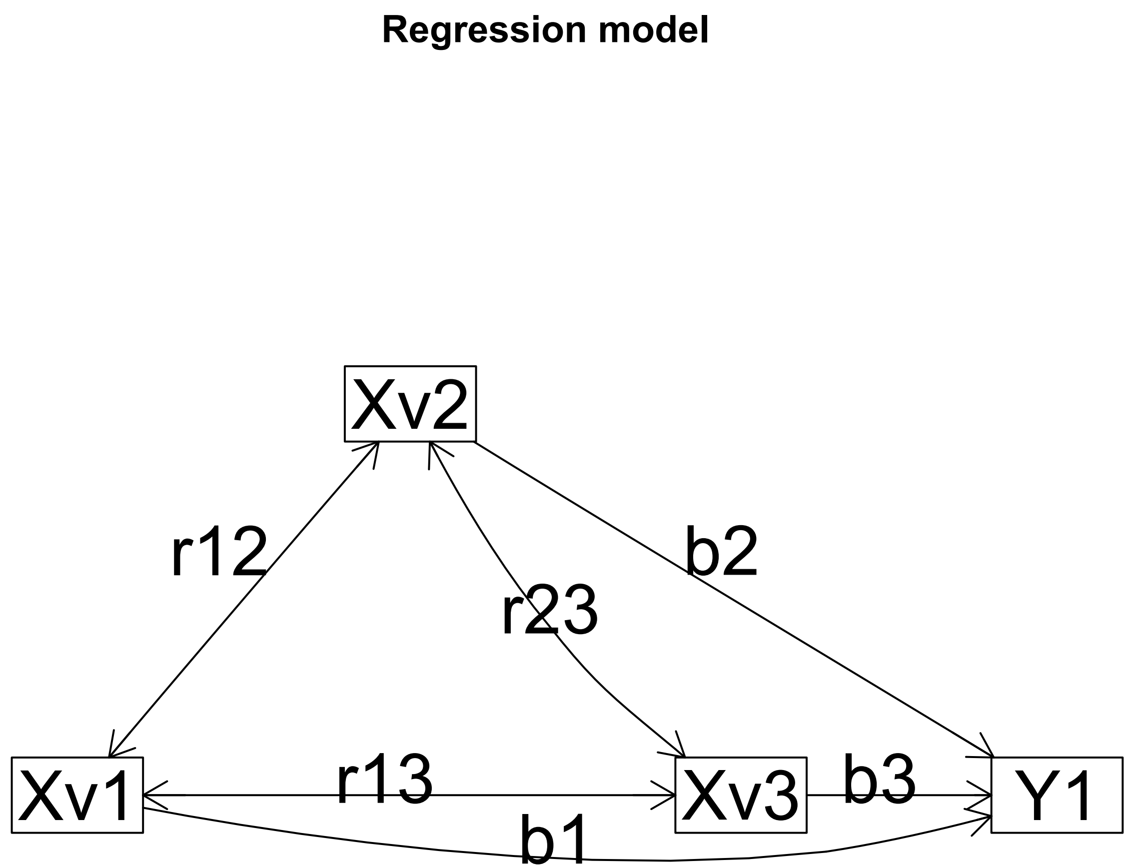


Regression weights

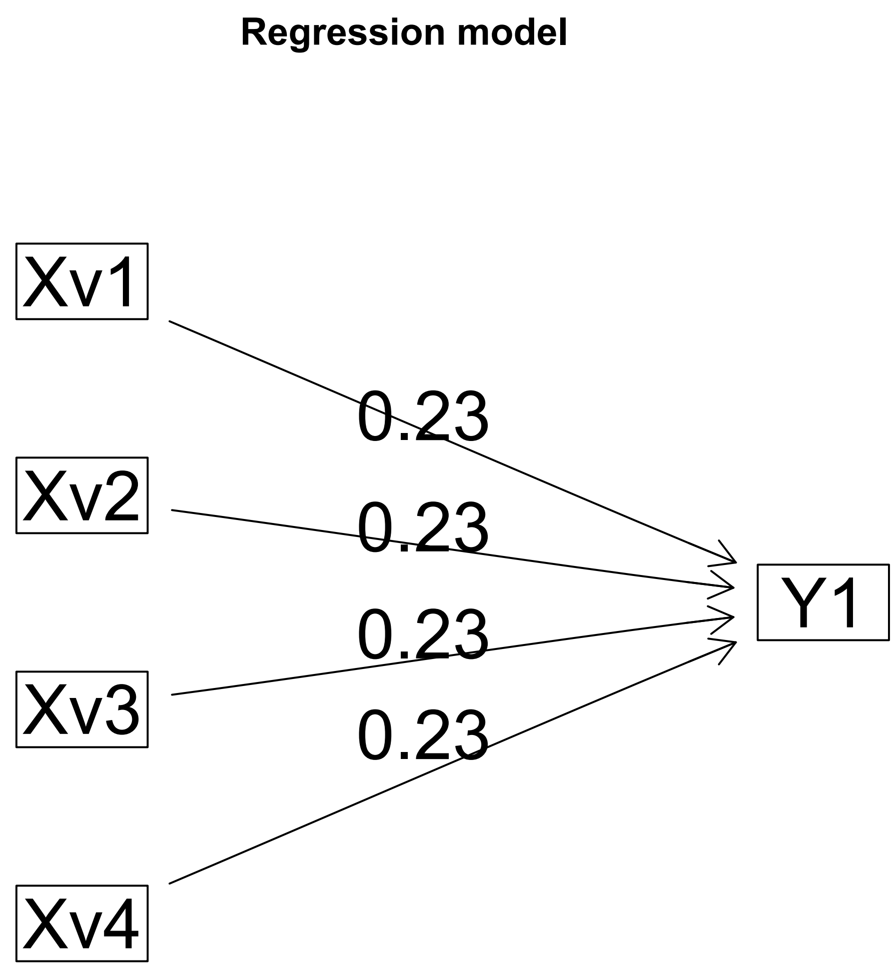
$r_{x1y} = b_1 + r_{12}b_2 + r_{13}b_3$

$r_{x2y} = b_2 + r_{12}b_1 + r_{23}b_3$

$r_{x3y} = b_3 + r_{13}b_1 + r_{23}b_2$



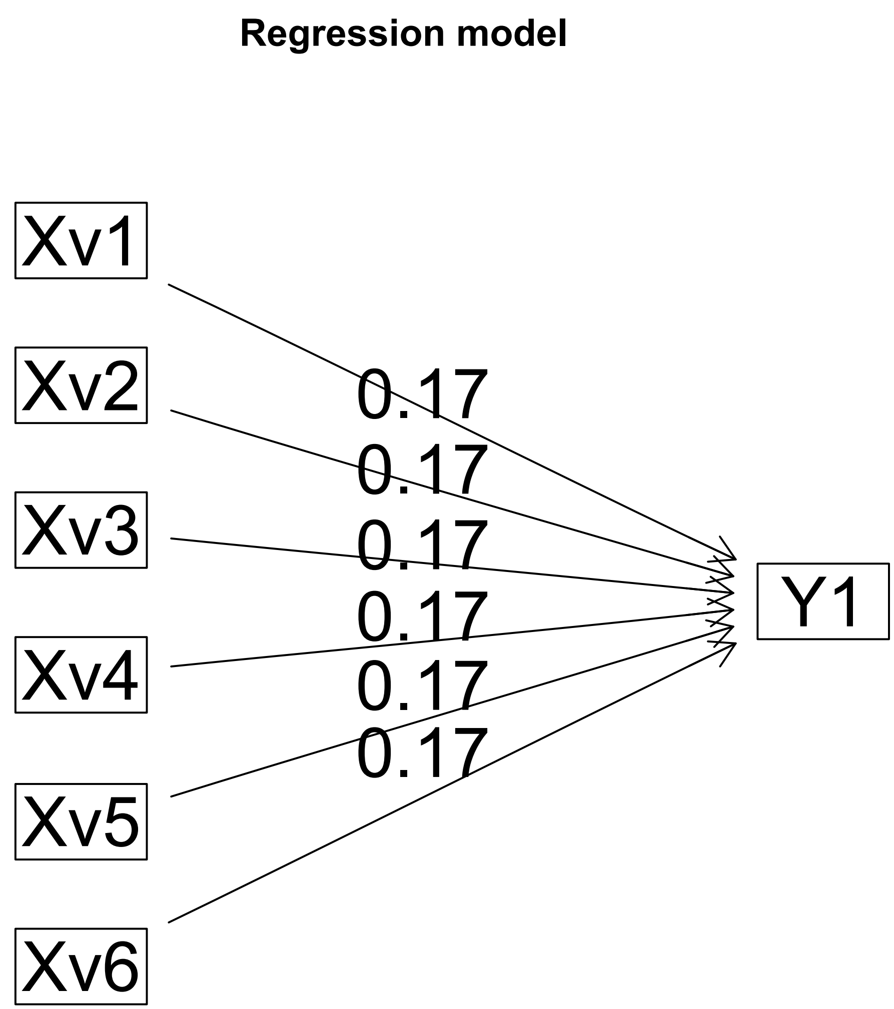
Four predictors R = .67



Correlations of .36 between predictors are not shown

6 predictors: R = .7

Correlations of .36 between predictors are not shown



Regression vs. structural model

- Regression weights decrease as number of predictors increase
- Structural weights do not change
- If N predictors are equally correlated with each other (a_{ij}) and with criterion (r_{iy}), then $r_{iy} = b_{iy} + (N-1)a_{ij}\, b_{iy}$ which implies
  - $b_{iy} = r_{ij}/(1+(N-1)a_{ij}$ and $R^2 = \text{sum}(\ b_{iy}\ r_{iy}) = N\ b_{iy}\ r_{iy}$
  - $R^2 = N\ r_{ij}^2/(1+(N-1)a_{ij}$ which at limit =>
  - $R = r_{ij}/\text{sqrt}(a_{ij})$ which is just the disattenuated r

Methods of Scale Construction

- Empirical
  - MMPI, Strong
- Rational
  - CPI
- Theoretical
  - NAch
- Homogeneous

EPI, 16PF, NEO

Empirical Keying

- Ask items that discriminate known groups
  - People in general versus specific group
  - Choose items that are maximally independent and that have highest validities
- Example:
  - MMPI
  - Strong-Campbell
- Problem:
  - What is the meaning of the scale?
  - Need to develop new scale for every new group

Rational Keying

- Ask items with direct content relevance
- Example: California Psychological Inventory
- Problems
  - Not all items predict in obvious way
  - Need evidence for validity
  - Easy to fake

Theoretical Keying

- Ask items with theoretical relevance
- Example: Jackson Personality Research Form
- Problems:
  - Theoretical circularity
  - Need evidence for validity

Homogeneous Keying

- Select items to represent single domain
  - Exclude items based upon internal consistency
- Examples:
  - 16PF, EPI/EPQ, NEO
- Problems
  - Garbage In, Garbarge Out
  - Need evidence for validity

Methods of Homogeneous Keying

- Factor Analysis
- Principal Components Analysis
- Cluster Analysis