Dimensions of Well-Being and Their Statistical Measurements:

Statistical Composite Indicators to convey consistent policy messages

Maurizio Vichi

Department of Statistical Sciences
Sapienza University of Rome

em: maurizio.vichi@uniroma1.it

2016 Annual meeting of Community of Practice on Composite Indicators and Scoreboards, 29 September 2016

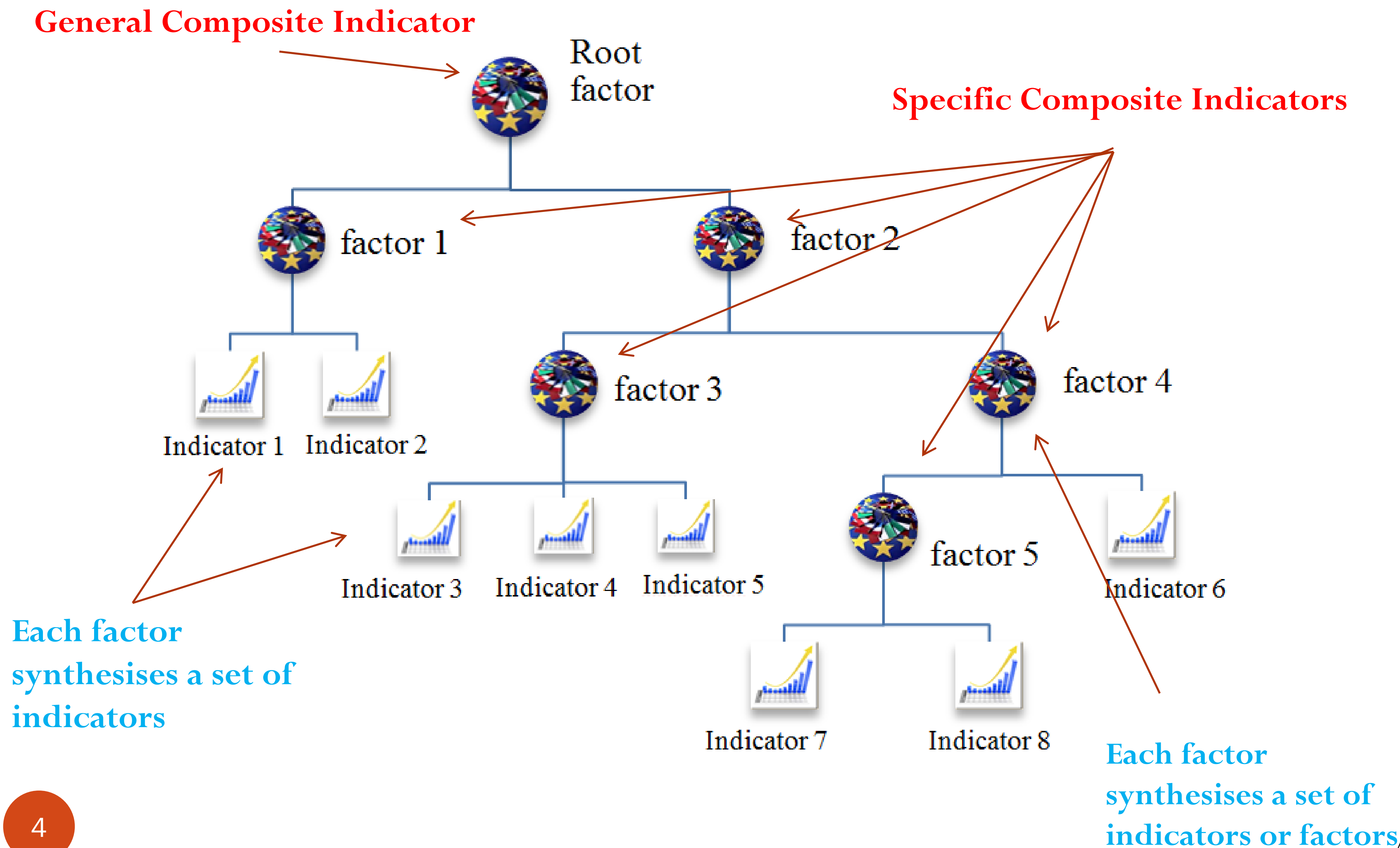
Outline of the Presentation

- A model-based Composite Indicator (CI) which is the result of the joint dimensional reduction of the observed multivariate data.
  The methodology has two aims:
  - Indicator reduction: find a hierarchical simple structure model to identify a CI
  - Units reduction: obtain the largest number of clusters with CI statistically different

  Methodology: Clustering & Hierarchical Disjoint Non-negative Factor analysis
  Properties of the CHDNFA
  CHDNFA detects a General and some Specific Composite Indicators that best (MLE) reconstruct the observed indicators (via a reflective model : data=CI model + error);
  CHDNFA is scale equivariant, thus normalization of observed indicators does not effect the final composite indicator;
  CHDNFA identifies unique (latent) composite indicators which interpretation cannot be improved by any orthogonal transformation;
  CHDNFA produces reliable composite indicators by the best non-negative loadings;
  CHDNFA defines Unidimensional Composite Indicators;
  CHDNFA detects Composite Indicators with a robust ranking of individuals by means of clustering.

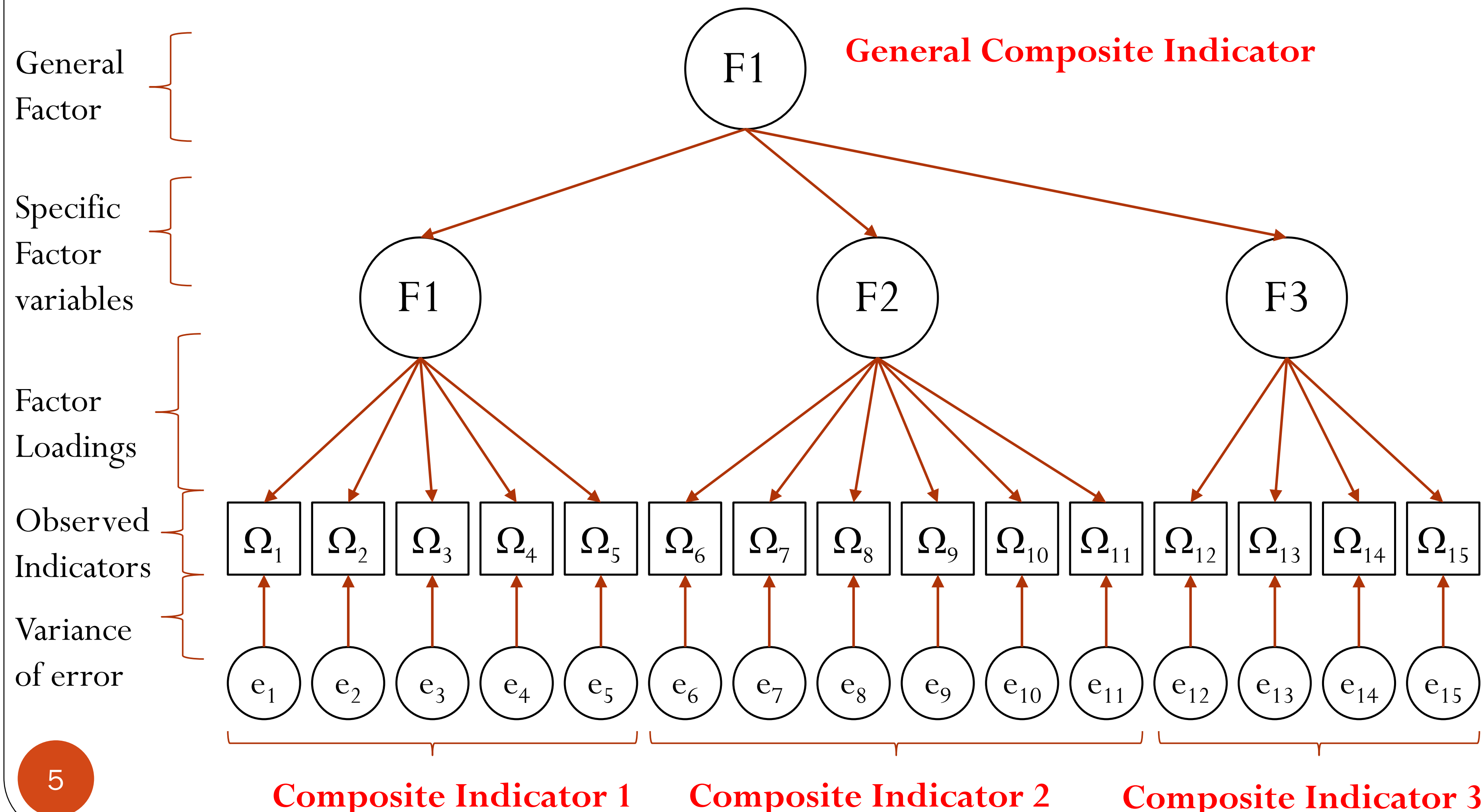
- **Hierarchical simple structure model to construct a Composite Indicator**

Hierarchical simple structure model



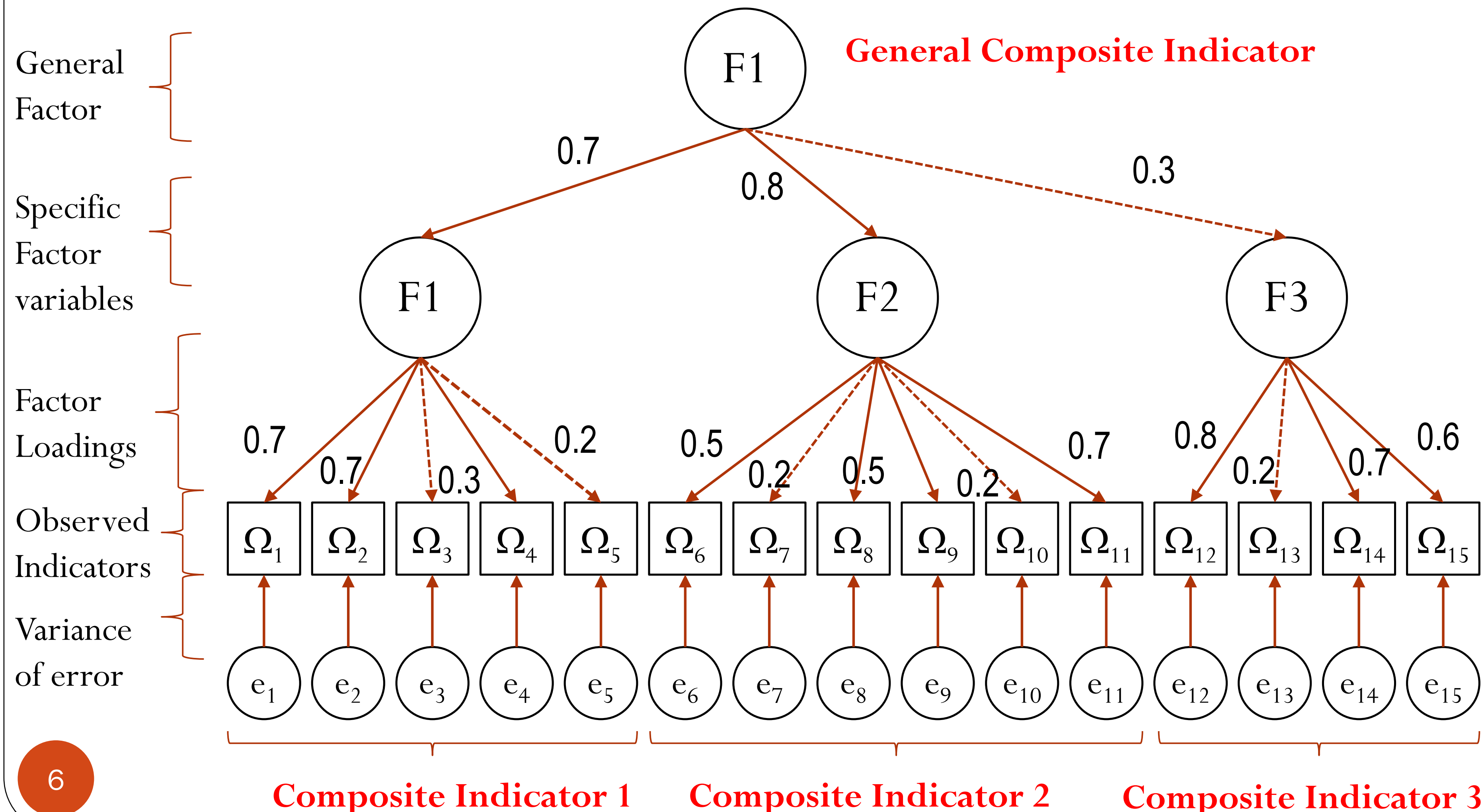
Two level Hierarchical Simple Structure Model

Confirmatory approach: associations the number of specific factors, and association between observed indicators and specific composite indicators (represented by arrows) are supposed known, the level of correlation has to be estimated.



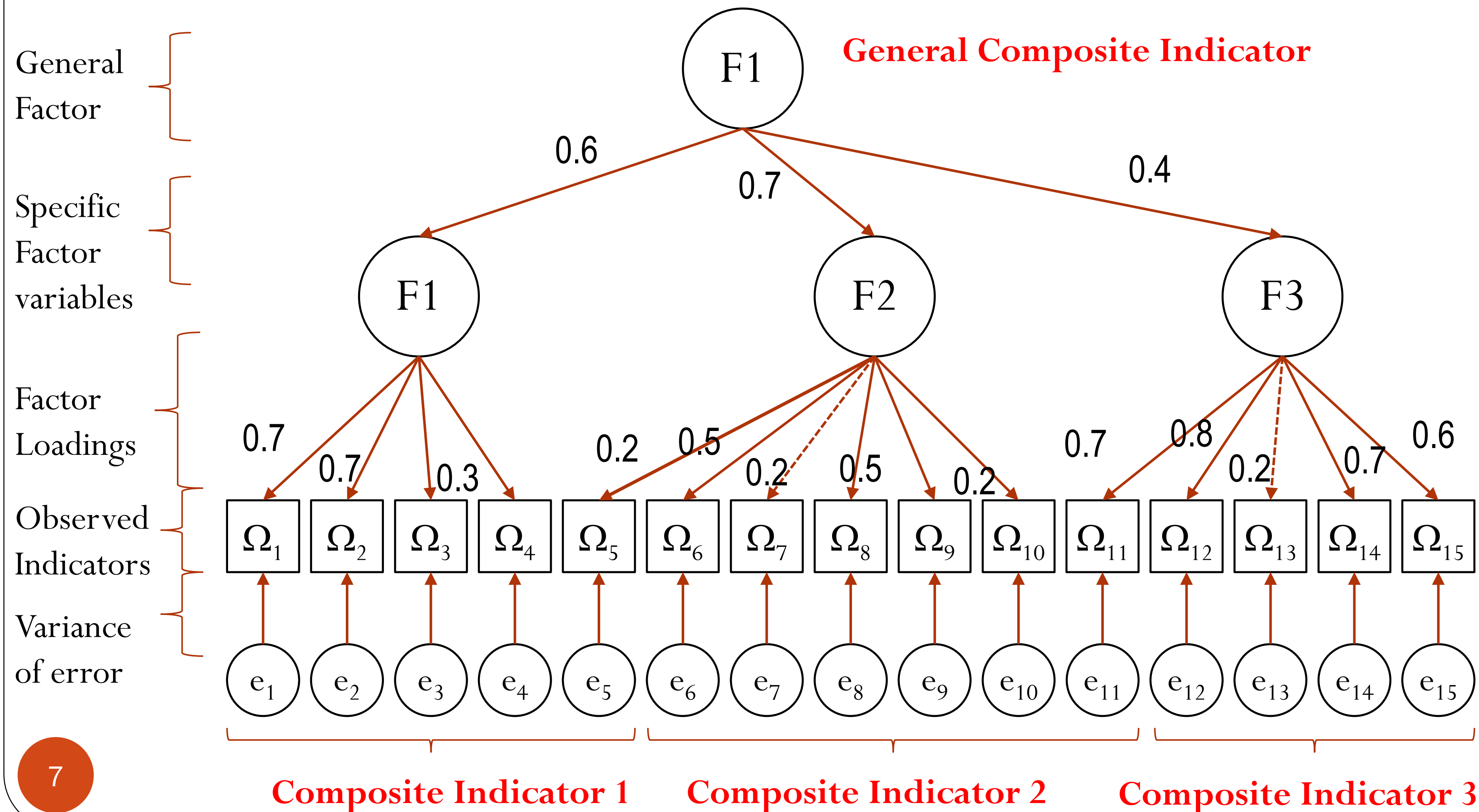
Two level Hierarchical Simple Structure Model

ESTIMATION: correlation between variables and factors, between general and specific factors some associations may be not statistically significant (correlations are substantially null) and FIT POOR



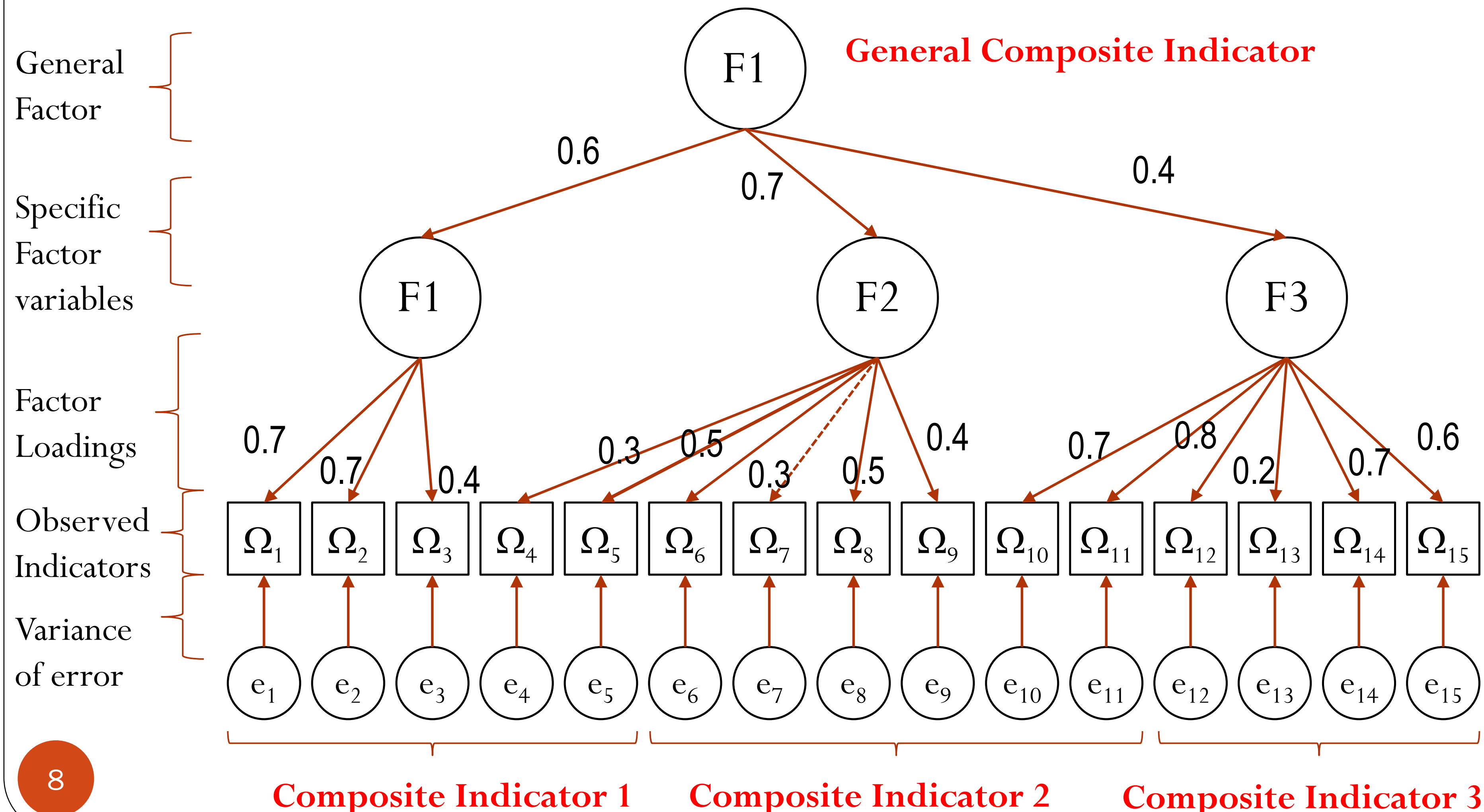
Two level Hierarchical Simple Structure Model

At this point the researcher start to play with different models hypothesizing some changes that do not have a theory behind.



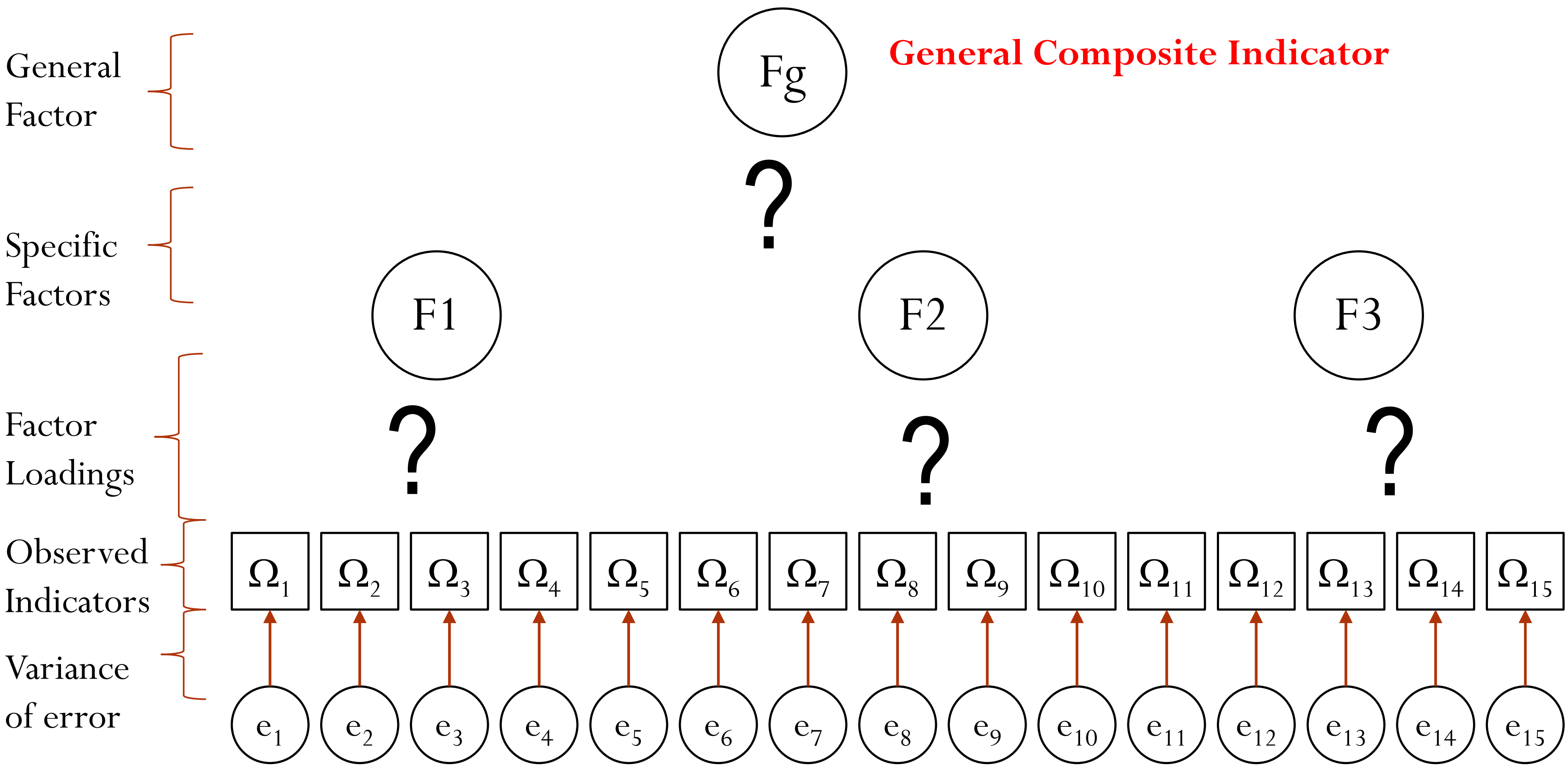
Two level Hierarchical Simple Structure Model

The final model is one obtained by the researcher only "partially" sustained by a theory. The modification is not 'optimised' and thus us the model selection becomes an "artisanal skill" of the researcher



Two level Hierarchical Simple Structure Model

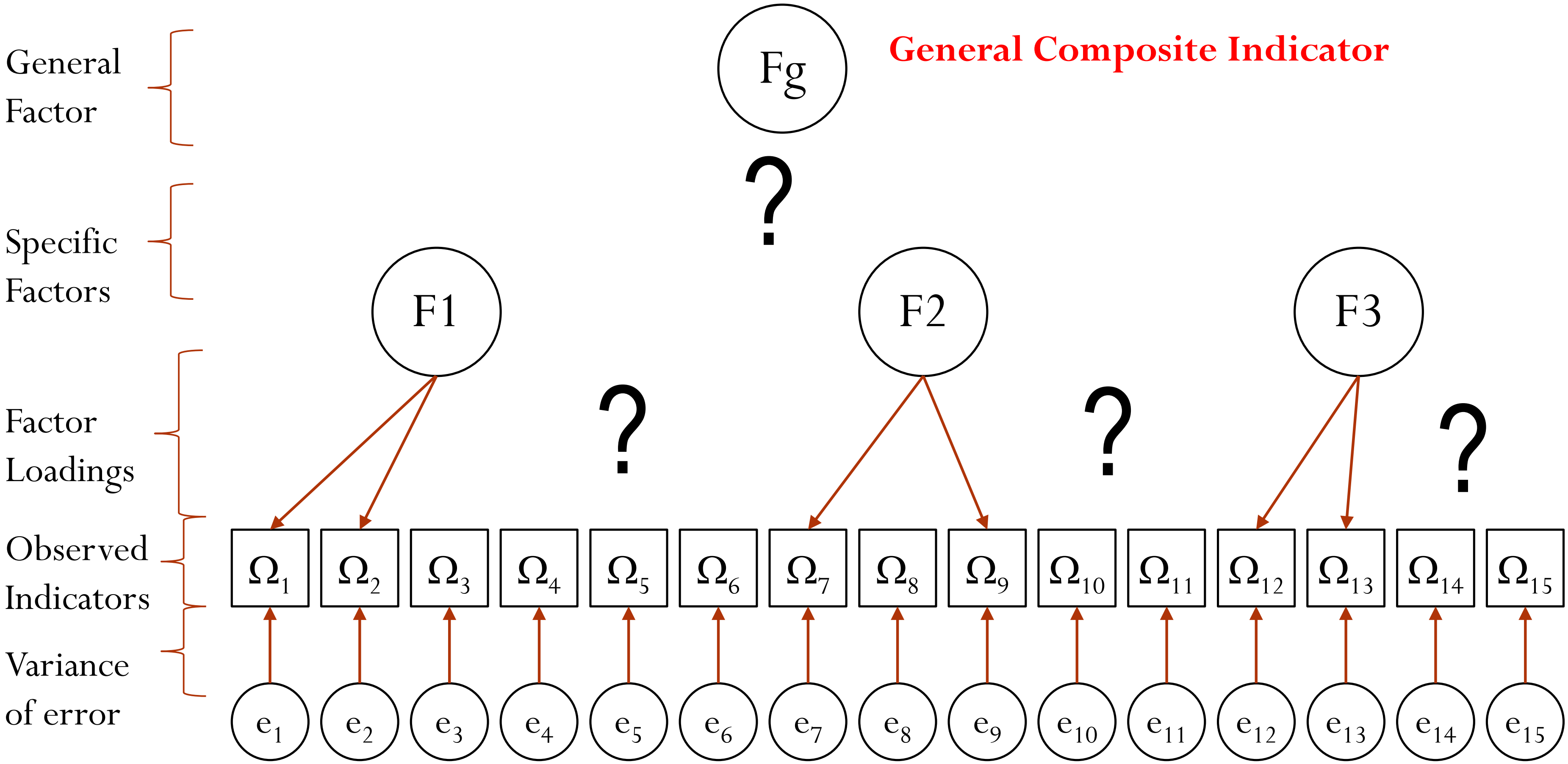
OUR PROPOSAL (1/3): Only the # of specific CIs is known



Association between CIs and between CIS and Indicators **are unknown**

Two level Hierarchical Simple Structure Model

OUR PROPOSAL (2/3) SOME FLEXIBILITY: also part of associations are known, because these are sustained by a theory.

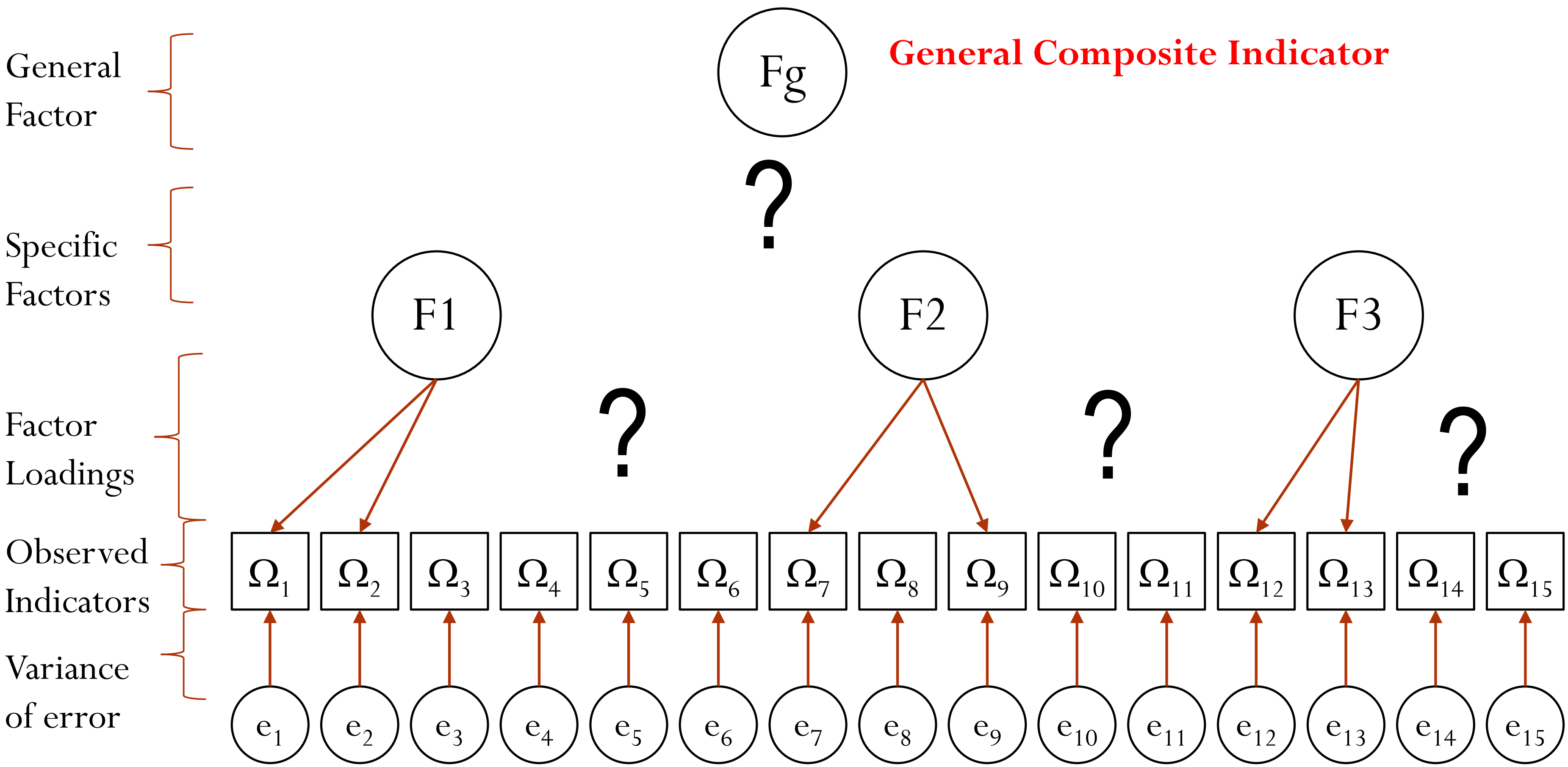


Association between CIs and observed Indicators **are partially known**

Two level Hierarchical Simple Structure Model

OUR PROPOSAL (3/3) Statistical Coherence of correlations.

Weights must be non-negative, unidimensionality and reliability of specific CIs



NON-NEGATIVE CORRELATIONS

Clustering &
Hierarchical Disjoint Factor Analysis

a model to identify the latent Hierarchical Composite Indicator and set of specific Composite Indicators that best reconstruct the observed data and specify a ranking of clusters

SOME METODOLOGICAL CONSIDERATIONS

Clustering &
Hierarchical Disjoint Factor Analysis

$$\mathbf{x} - \boldsymbol{\mu}_\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{e}_\mathbf{x}, \quad (\mathbf{y} \text{ Specific factors}) \quad (1)$$
$$\mathbf{y} = \mathbf{c}g + \mathbf{e}_\mathbf{y}, \quad (g \text{ General factor}) \quad (2)$$

Let include model (2) into model (1) the loading matrix **A** is restricted to the product **A**=**BV**
Including (4) in HFA the **HDFA model** is defined

$$\mathbf{x} - \boldsymbol{\mu}_\mathbf{x} = \mathbf{BV}(\mathbf{c}g + \mathbf{e}_\mathbf{y}) + \mathbf{e}_\mathbf{x} = \mathbf{BVc}g + \mathbf{BVe}_\mathbf{y} + \mathbf{e}_\mathbf{x}. \quad (3)$$

Let rewrite the model in matrix form

$$\mathbf{X} = \mathbf{gc}'\mathbf{V}'\mathbf{B} + \mathbf{E}_\mathbf{x}. \quad (4)$$

Additionally the general factor scores **g** is partitioned into K disjoint clusters, where matrix **U** is the membership matrix and $\bar{\mathbf{g}}$ is the centroid vector

$$\mathbf{X} = \mathbf{U}\bar{\mathbf{g}}\mathbf{c}'\mathbf{V}'\mathbf{B} + \mathbf{E}_\mathbf{x}, \quad (5)$$

with

$$\boldsymbol{\Sigma}_\mathbf{x} = \mathbf{BVc}\frac{1}{n}(\bar{\mathbf{g}}'\mathbf{U}'\mathbf{U}\bar{\mathbf{g}})\mathbf{c}'\mathbf{V}'\mathbf{B} + \boldsymbol{\Psi}_\mathbf{x}, \quad (6)$$

where $\boldsymbol{\Sigma}_\mathbf{y} = \mathbf{c}\frac{1}{n}(\bar{\mathbf{g}}'\mathbf{U}'\mathbf{U}\bar{\mathbf{g}})\mathbf{c}' + \boldsymbol{\Psi}_\mathbf{y}.$ (7)

such that

$\mathbf{V} = [\, v_{jh} : \forall\ v_{jh} \in \{0,1\}]$	(binary)	(8)
$\mathbf{V}\mathbf{1}_H = \mathbf{1}_J$	(row stochastic)	(9)
$\mathbf{U} = [\, u_{jk} : \forall\ u_{ik} \in \{0,1\}]$	(binary)	(8)
$\mathbf{U}\mathbf{1}_K = \mathbf{1}_n$	(row stochastic)	(9)
$\mathbf{B} = diag(b_1, \ldots, b_J)$ with $b_j^2 > 0$	(diagonal, non-null)	(10)
$\mathbf{V}'\mathbf{BBV} = diag(b_{.1}^2, \ldots, b_{.H}^2)$, with $b_{.h}^2 = \sum_{j=1}^{J} b_{jh}^2 > 0$	(orthogonal, non-empty)	(11)

Estimation of Clustering & HDFA

Minimization of the **discrepancy functions** w.r.t. **B, V, U, $\bar{\mathbf{Y}}$** and **Ψ**

Least-Squares Estimation

$$LSE(\mathbf{B},\mathbf{V}, \mathbf{\Psi}, \mathbf{U}, \bar{\mathbf{Y}}) = \|\mathbf{S} - \mathbf{BV}\frac{1}{n}(\bar{\mathbf{g}}'\mathbf{U}'\mathbf{U}\bar{\mathbf{g}}))\mathbf{V}'\mathbf{B} - \mathbf{\Psi}_{\mathbf{x}}\|^2 \rightarrow \min_{\mathbf{B},\mathbf{V}, \mathbf{\Psi}, \mathbf{U}, \bar{\mathbf{Y}}} \quad (12)$$

Maximum likelihood Estimation

$$MLE(\mathbf{B},\mathbf{V}, \mathbf{\Psi}, \mathbf{U}, \bar{\mathbf{Y}}) = ln\left|\mathbf{BV}\frac{1}{n}(\bar{\mathbf{g}}\mathbf{U}'\mathbf{U}\bar{\mathbf{g}})\mathbf{V}'\mathbf{B}+\mathbf{\Psi}\right| - ln\,|\mathbf{S}| + tr\left(\left(\mathbf{BV}\frac{1}{n}(\bar{\mathbf{g}}\mathbf{U}'\mathbf{U}\bar{\mathbf{g}})\mathbf{V}'\mathbf{B}+\mathbf{\Psi}\right)^{-1}\mathbf{S}\right) - J \rightarrow \min_{\mathbf{B},\mathbf{V}, \mathbf{\Psi}, \mathbf{U}, \bar{\mathbf{Y}}} \quad (12)$$

Generalised Least-Squares Estimation

$$GLSE(\mathbf{B},\mathbf{V}, \mathbf{\Psi}, \mathbf{U}, \bar{\mathbf{Y}}) = \|(\mathbf{S} - \mathbf{BV}\frac{1}{n}(\bar{\mathbf{g}}'\mathbf{U}'\mathbf{U}\bar{\mathbf{g}})\mathbf{V}'\mathbf{B} - \mathbf{\Psi}_{\mathbf{x}})\mathbf{S}^{-1/2}\|^2 \rightarrow \min_{\mathbf{B},\mathbf{V}, \mathbf{\Psi}, \mathbf{U}, \bar{\mathbf{Y}}} \quad (12)$$

such that

$\mathbf{V} = [\, v_{jh} : \forall\ v_{jh} \in \{0,1\}]$	(binary)	(13)
$\mathbf{V1}_H = \mathbf{1}_J$	(row stochastic)	(14)
$\mathbf{U} = [\, u_{ik} : \forall\ v_{jk} \in \{0,1\}]$	(binary)	(15)
$\mathbf{U1}_K = \mathbf{1}_n$	(row stochastic)	(16)
$\mathbf{B} = diag(b_1, \ldots, b_J)$ with $b_j^2 > 0$	(diagonal, non-null)	(17)
$\mathbf{V}'\mathbf{BBV} = diag(b_{.1}^2, \ldots, b_{.H}^2)$, with $b_{.h}^2 = \sum_{j=1}^{J} b_{jh}^2 > 0$	(orthogonal, non-empty)	(18)

A coordinated descendent algorithm has been developed this problem.

NOTE: **This is a discrete and continuous problem that cannot be solved by a quasi-Newton type algorithm**

APPLICATION on WELL-BEING

OECD defines a Well-Being Index called **Better Life index**, considering 34 Countries and 24 indicators essential to identify well-being.

OECD, following Stiglitz et al. Committee (2009) identifies two dimensions

- **Material Living Conditions (MLC)**
- **Quality of Life (QL)**

OECD assumes that 8 observed variables define MLC and
the remain 16 variables the QL.

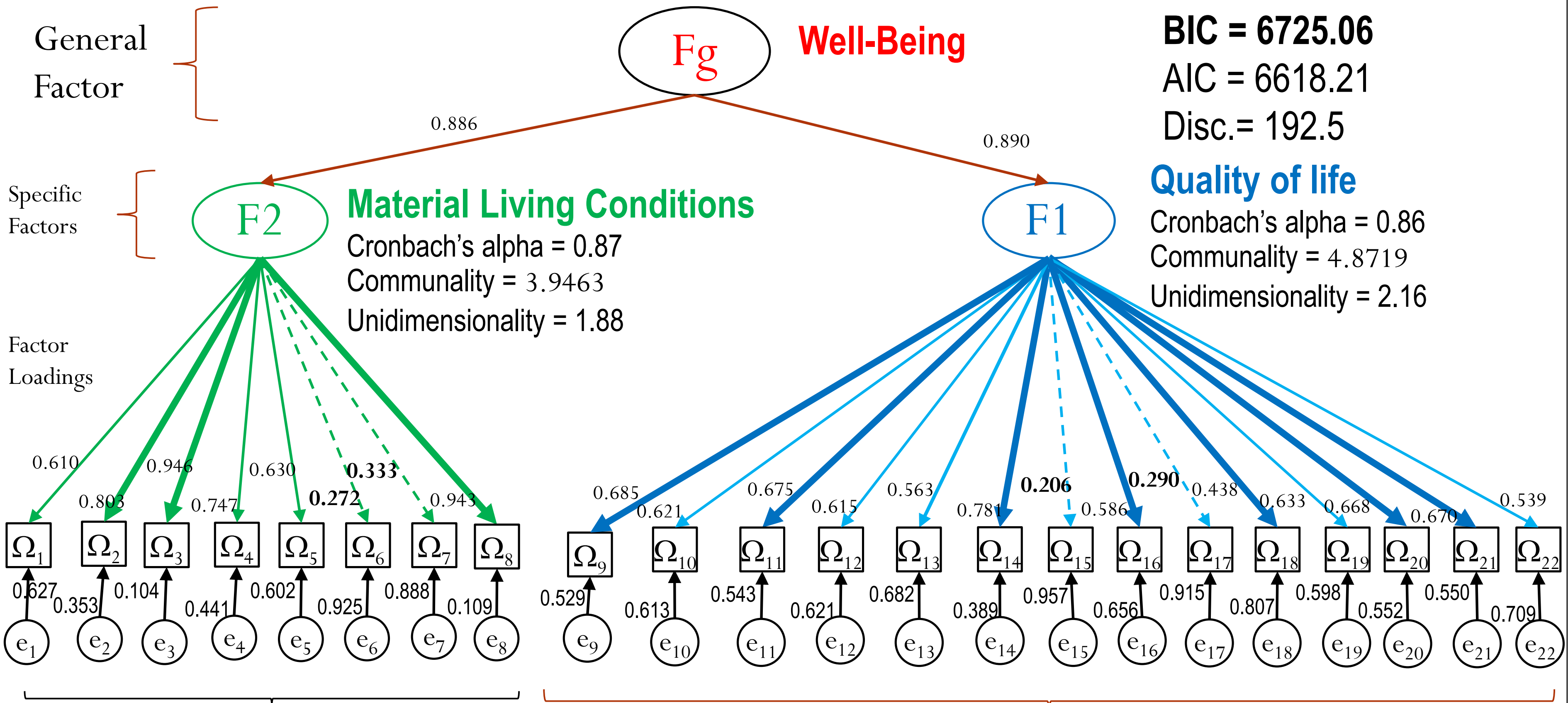
- Variables have been standardized. Min-max normalization can be used giving the same results because the scale equivariance property.

OECD observes that some manifest variables are imperfect proxies of the concepts that one would like to measure (MLC and QL) (Dolan, P., Peasgood, T. and White, M. (2008))
OECD leaves individual users to give subjective weighs to the variables

First a two-level hierarchy completely confirmatory model has been applied, requiring weights to be non-negative. 8 indicators have weight zero (ex. Dwellings without basic facilities). This means that they measure a negative component of the WB and for coherence must be reversed. Then the model has been reapplied. *Housing expenditure* and *Consultation on rule-making* have not significant correlation and therefore have been discarded. Then the model has been reapplied with 22 indicators.

Application: Better Life Index, OECD 2015

34 Countries, 22 indicators reflecting the pillars essential to identify well-being



Material Living Conditions

*(**Housing**: (1) Dwellings without basic facilities; (2) Rooms per person); (**Income**: (3) Household net adjusted disposable income; (4) Household net financial wealth); (**Jobs**: (5) Employment rate; (6) **Job security**; (7) **Long-term unemployment rate**; (8) Personal earnings);*

***Quality of Life** (QL)*

*(**Community**: (9) Quality of social support network); (**Education**: (10) Educational attainment; (11) Student skills; (12) Years in education); (**Environment**: (13) Air pollution; (14) Water quality); (**Civic engagement**: (15) **Voter turnout**); (**Health**: (16) Life expectancy; (17) **Self-reported health**); (Life **Satisfaction**: (18) Life satisfaction); (**Safety**: (19) Assault rate; (20) Homicide rate); (**Work-Life Balance**: (21) Employees working very long hours; (22) Time devoted to leisure and personal care)*

Can we obtain a better result?

- find a composite indicator that best reconstruct the 22 initial variables by using two dimensions?

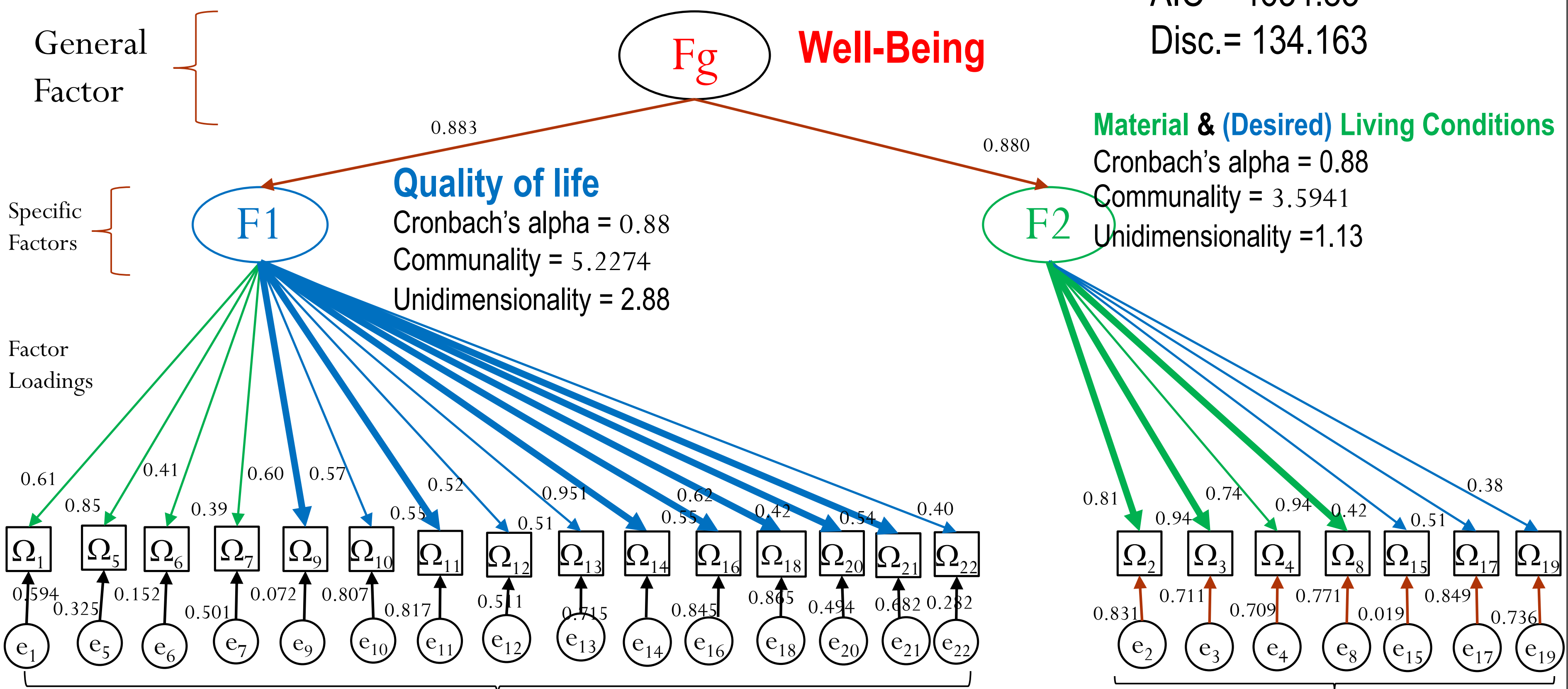
- Can we found more than two dimensions?

Additional information: (11 dimensions),
3 define MLC; and 8 define QL

Application: Better Life Index, OECD 2015

34 Countries, 22 indicators

BIC = 4741.67
AIC = 4634.83
Disc.= 134.163



Quality of Life *(QL))*

(Housing: (1) Dwellings without basic facilities; (Jobs: (5) Employment rate; (6) Job insecurity; (7) Long-term unemployment rate; (Community: (9) Quality of social support network); (Education: (10) Educational attainment; (11) Student skills; (12) Years in education); (Environment: (13) Air pollution; (14) Water quality); (Health: (16) Life expectancy; (Life Satisfaction: (18) Life satisfaction; (Civic engagement: (20) Homicide rate); (Work-Life Balance: (21) Employees working very long hours; (22) Time devoted to leisure and personal care)

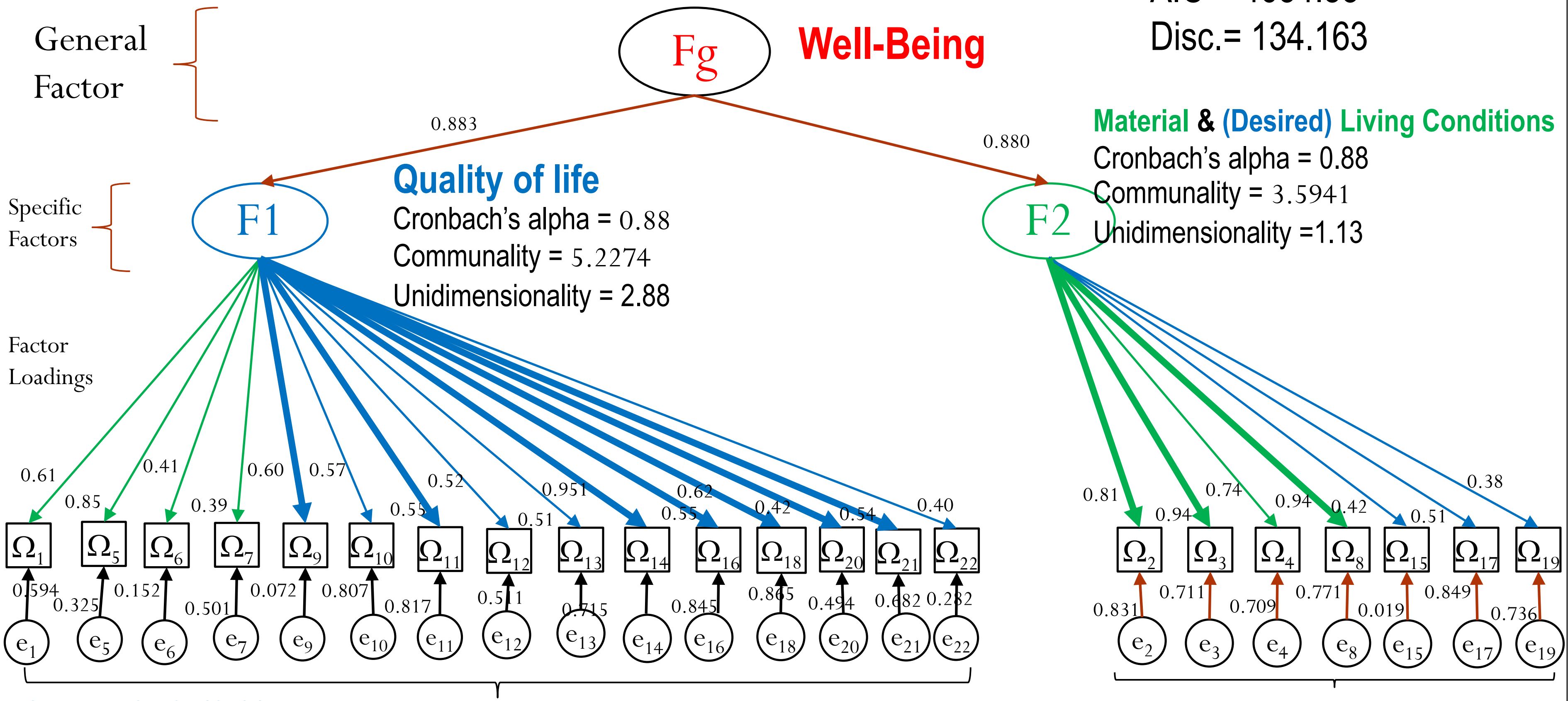
Material & Desired Living Conditions

(Housing: (2) Rooms per person); (Income: (3) Household net adjusted disposable income; (4) Household net financial wealth (Job:(8) Personal earnings); (15) Voter turnout); (17) Self-reported health); (Safety: (19) Assault rate;

Application: Better Life Index, OECD 2015

34 Countries, 22 indicators

BIC = 4741.67
AIC = 4634.83
Disc.= 134.163



Quality of Life *(QL))*

(Housing: (1) Dwellings without basic facilities; (Jobs: (5) Employment rate; (6) Job insecurity; (7) Long-term unemployment rate; (Community: (9) Quality of social support network); (Education: (10) Educational attainment; (11) Student skills; (12) Years in education); (Environment: (13) Air pollution; (14) Water quality); (Health: (16) Life expectancy; (Life Satisfaction: (18) Life satisfaction; (Civic engagement: (20) Homicide rate); (Work-Life Balance: (21) Employees working very long hours; (22) Time devoted to leisure and personal care)

Material & Desired Living Conditions

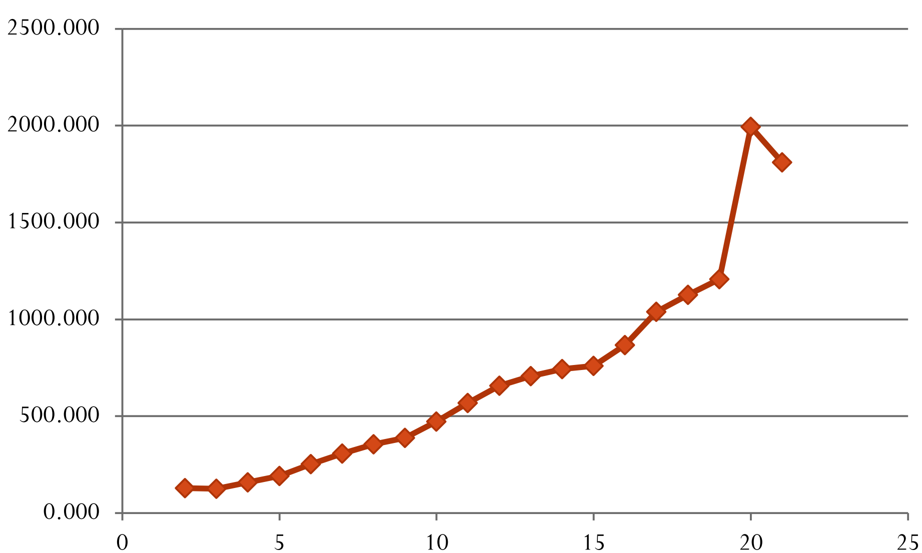
(Housing: (2) Rooms per person); (Income: (3) Household net adjusted disposable income; (4) Household net financial wealth (Job:(8) Personal earnings); (15) Voter turnout); (17) Self-reported health); (Safety: (19) Assault rate;

Ranking of clusters

K=20 according to pF

		Fg
1	United States	1.99
2	Norway	1.71
3	Australia	1.57
	Switzerland	1.56
	Belgium	1.53
4	Germany	1.25
5	Denmark	1.08
	New Zealand	1.04
6	Ireland	0.93
	Japan	0.83
7	Finland	0.76
	Netherlands	0.73
8	Luxembourg	0.61
	Sweden	0.57
	Austria	0.56
	Italy	0.56
9	Iceland	0.42
	Canada	0.41
10	United Kingdom	0.26
	France	0.12
11	Greece	-0.05
12	Korea	-0.49
	Slovenia	-0.52
13	Spain	-0.84
14	Poland	-1.05
15	Israel	-1.27
	Mexico	-1.3
16	Chile	-1.49
	Czech Republic	-1.49
17	Slovak Republic	-1.61
18	Hungary	-1.66
19	Turkey	-2.2
	Portugal	-2.21
20	Estonia	-2.34

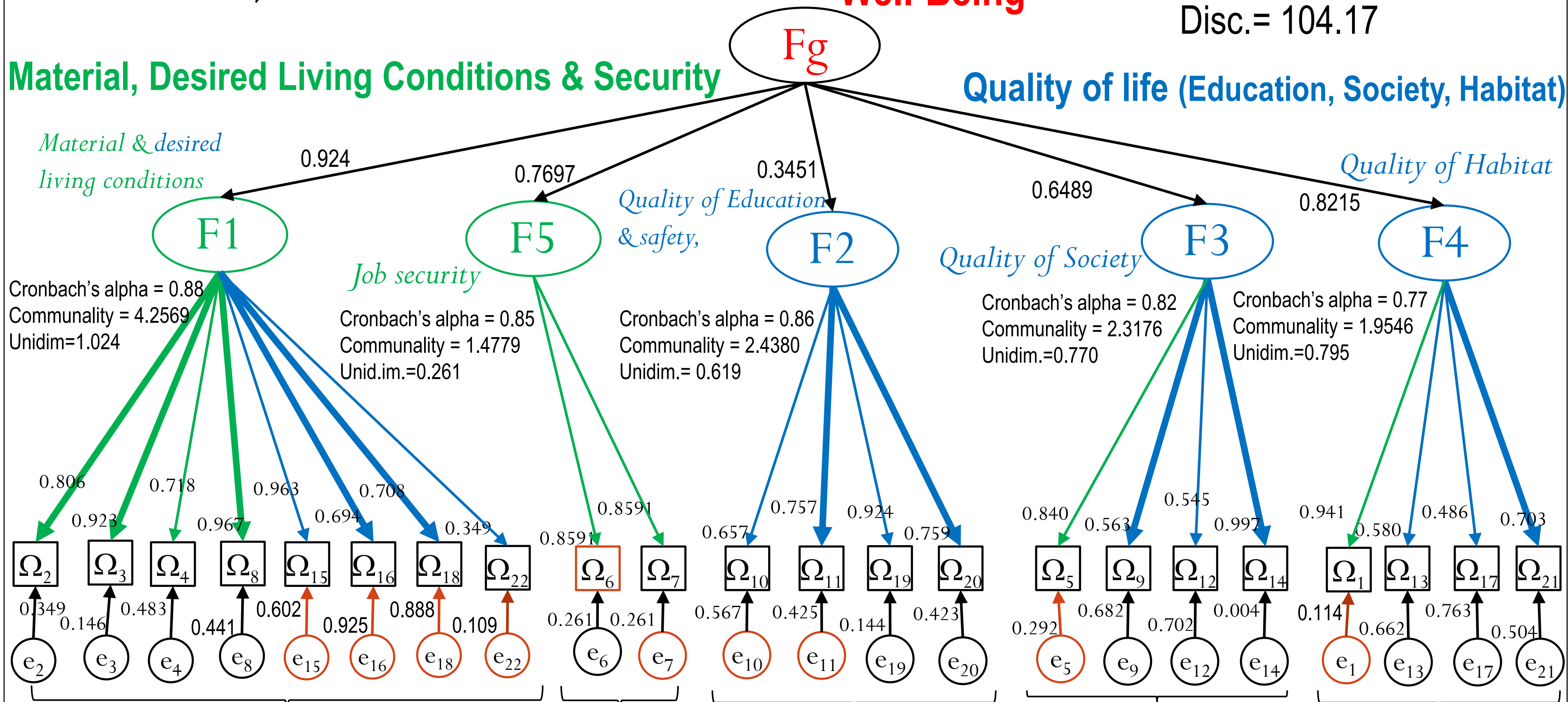
pseudo F statistics



Application: Better Life Index, OECD 2015

34 Countries, 22 indicators

BIC = 4007.78
AIC = 3791.04
Disc.= 104.17



Material, Desired Living Conditions & Security

Material ,desired conditions (2) *Rooms per person;* (3) *Household net adjusted disposable income;* (4) *Household net financial wealth;* (8) *Personal earnings);*

(15) *Voter turnout);* (16) (18) *Life satisfaction);* (22) *Time devoted to leisure and personal care*

Job security (6) *Job insecurity;* (7) *Long-term unemployment rate*

Quality of Life (QL) ***(Education, Society, Habitat)***

Quality of Safety and Education: (10) *Educational attainment;* (11) *Student skills;* (19) *Assault rate;* (20) *Homicide rate;*

Quality of Society (9) *Quality of social support network);* (14) *Water quality);* (5) *Employment rate;* (12) *Years in education;*

Quality of the habitat (1) *Dwellings without basic facilities* (13) *Air pollution;* (17) *Self-reported health;* (21) *Employees working very long hours;*

CONCLUSION

- **CI to convey consistent policy messages**

- **Properties of the CHDNFA**
- CHDNFA detects a *General* and some *Specific Composite Indicators* that *best* (MLE) reconstruct the observed indicators (via a reflective model);

- CHDNFA is *scale equivariant,* thus allowing any scaling of indicators necessary to normalise the observed indicators with different units of measurements;
- CHDNFA identifies *unique* (latent) composite indicators which interpretation cannot be improved by any orthogonal transformation;
- CHDNFA produces *reliable* composite indicators by the best non-negative loadings;
- CHDNFA, with the correct model selection, defines Unidimensional Composite Indicators;
- CHDNFA detects Composite Indicators with a robust ranking of individuals by means of clustering.

Thank you for your attention!

SIMULATION STUDY

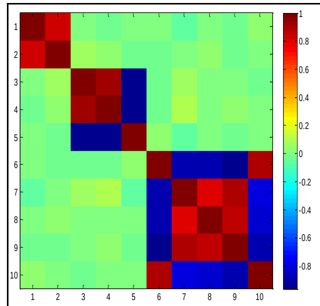
n=100, J=10, H=3
Three levels of error low, medium high

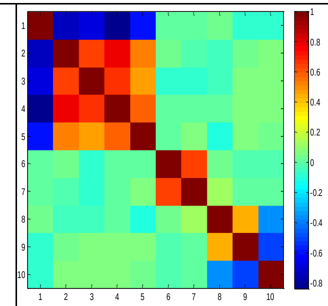
n=500, J=50 H=3
Three levels of error low, medium high

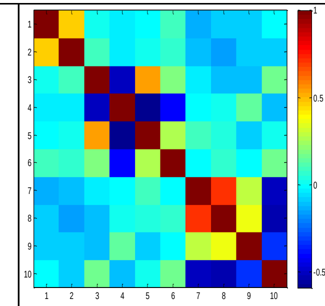
$\mathbf{y} \sim N_H(\mathbf{0}, \mathbf{I})$ and
$\mathbf{e} \sim N_J(\mathbf{0}, d\mathbf{\Psi})$, ($d$ = 0.1, 1, 2),
with $\psi_j \sim$U(0, 1),

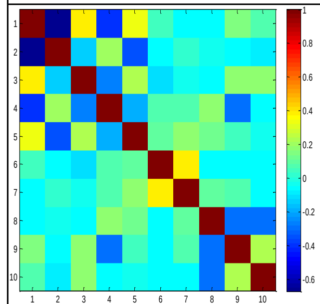
$\mathbf{B}=diag(b_1,\ldots,b_J)$
with $b_j = 0.7\text{sign}(a) + 0.1a$, with $a \sim N(0, 1)$,

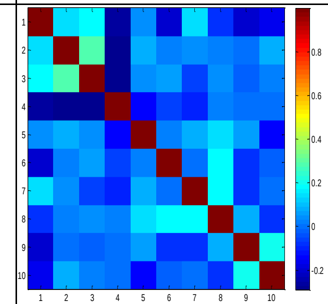
$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_J]'$
with $\mathbf{v}_J \sim Multinomial(H: p_h=1/H, h=1,\ldots,H)$,

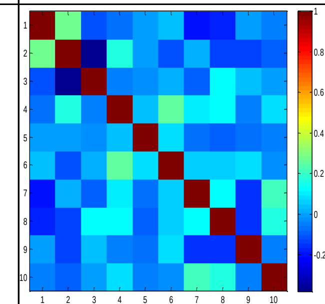

(a) n=100,J=10, d=0.33
Low error – blocks are well visible

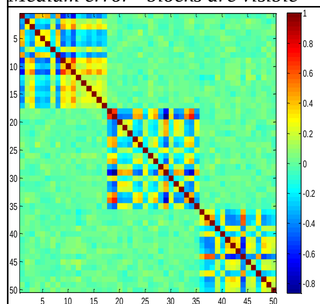

(b) n=100,J=10, d=1.00
Low error –blocks are well visible

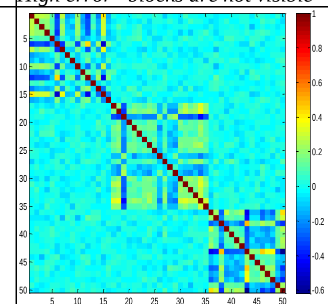

(c) n=100, J=10, d=1.33
Medium error –blocks are visible

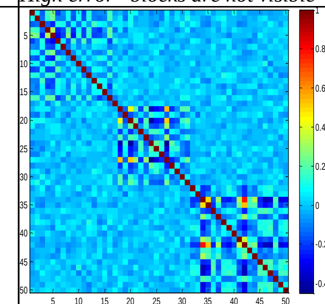

(e) n=100, J=10, d=1.66
Medium error –blocks are visible

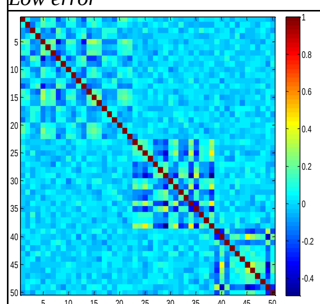

(f) n=100, J=10, d=2.00
High error –blocks are not visible

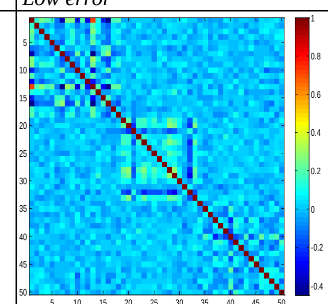

(g) n=100,J=10, d=2.33
High error –blocks are not visible


(h) n=500, J=50, d=1.66
Low error


(i) n=500, J=50, d=2.00
Low error


(l) n=500, J=50, d=2.66
Medium error


(m) n=500,J=50, d=3.00
Medium error


(p) n=500,J=50, d=3.66
High error-blocks are not visible

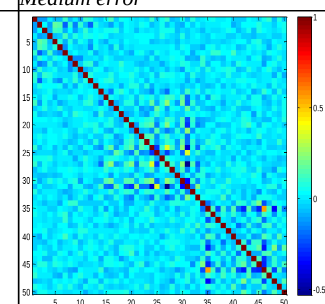

(q) n=500,J=50, d=4.66
High error- blocks are not visible

Table 1: *simulated data sets with n*=100, *J*=10, *H*=3 and different level of error.

	Error low			Error medium		Error high	
	d=0.33	*d*=0.66	*d*=1	*d*=1.33	*d*=1.66	*d*=2.00	*d*=2.33
ARI	**1.000**	**1.000**	**0.997**	**0.973**	**0.934**	**0.809**	**0.640**
GFI	0.923	0.919	0.930	0.921	0.921	0.923	0.924
AGFI	0.864	0.857	0.876	0.860	0.861	0.863	0.865
RMSEA	0.156	0.162	0.148	0.161	0.161	0.159	0.156
RMSR	0.0012	0.0034	0.0046	0.0055	0.0058	0.0060	0.006
BIC	-78.00	706.23	984.12	1149.68	1223.60	1268.50	1287.20
AIC	-140.52	643.71	940.18	1087.15	1161.07	1205.98	1224.67
BIC_{H1}/BIC_{H0}	157	144	120	81	76	73	72
AIC_{H1}/AIC_{H0}	161	140	116	85	80	77	75

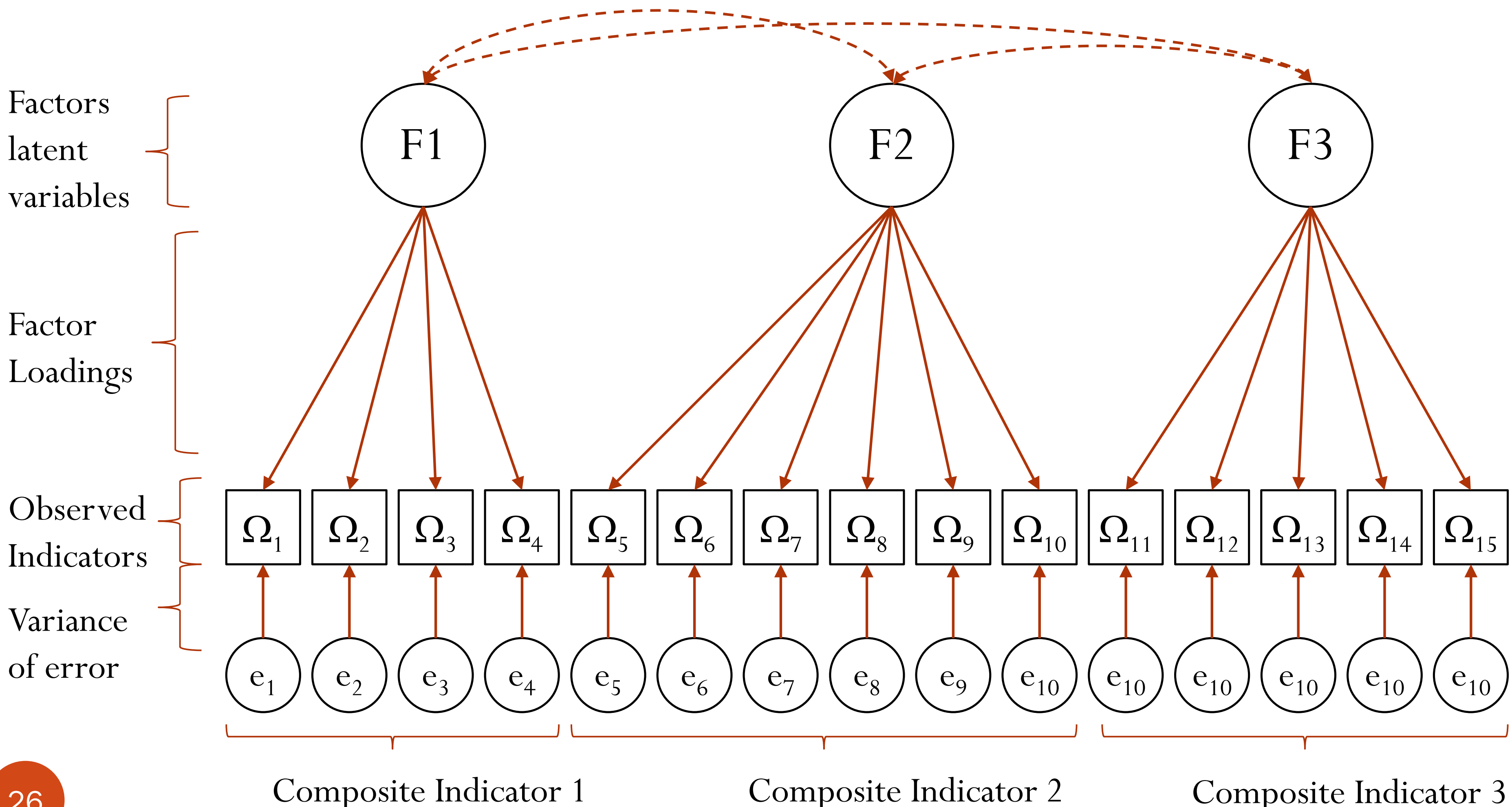
Table 2: *simulated data sets with n*=500, *J*=50, *H*=3 and different level of error.

	Error low			Error medium		Error high	
	d=1.66	*d*=2.00	*d*=2.66	*d*=3.00	*d*=3.66	*d*=4.00	*d*=4.66
ARI	**1.000**	**1.000**	**0.999**	**0.994**	**0.978**	**0.958**	**0.878**
GFI	0.897	0.897	0.896	0.896	0.895	0.897	0.897
AGFI	0.883	0.884	0.883	0.882	0.882	0.884	0.884
RMSEA	0.069	0.069	0.069	0.069	0.069	0.069	0.069
RMSR	0.001	0.002	0.002	0.002	0.002	0.002	0.002
BIC	11480.53	21096.27	23385.81	26073.98	26745.66	27751.42	27957.30
AIC	10873.63	20489.36	22778.91	25467.08	26138.76	27144.52	27350.40
BIC_{H1}/BIC_{H0}	7.52	3.32	2.99	2.67	2.60	2.51	2.49
AIC_{H1}/AIC_{H0}	6.78	3.39	3.04	2.71	2.64	2.54	2.52

Cross-Loadings

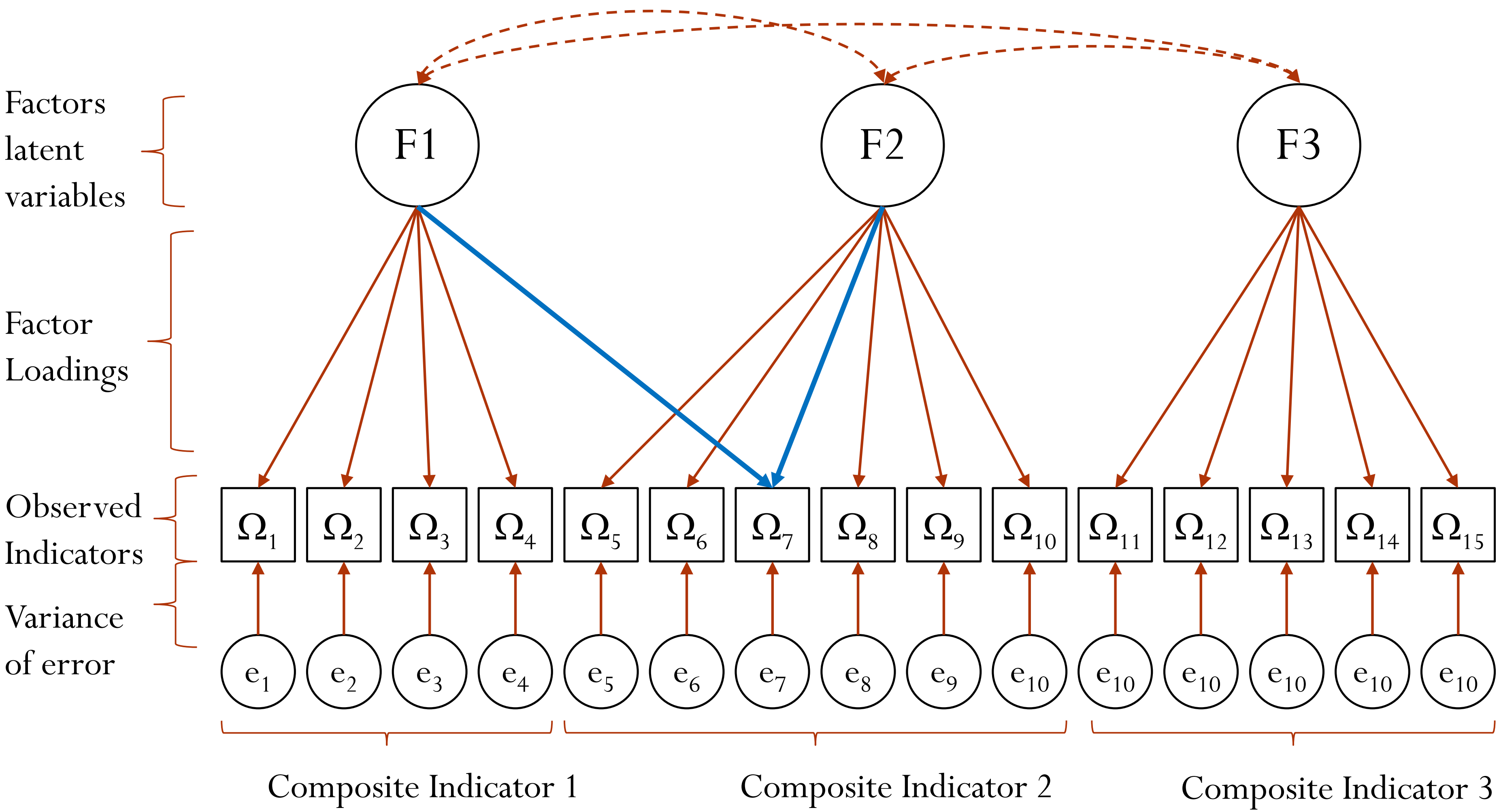
The fit of the SSM may be poor: for uncorrect choice of the number of factors
for the presence of cross-loadings

PROCEDURE: FIRST ESTIMATE the best SSM



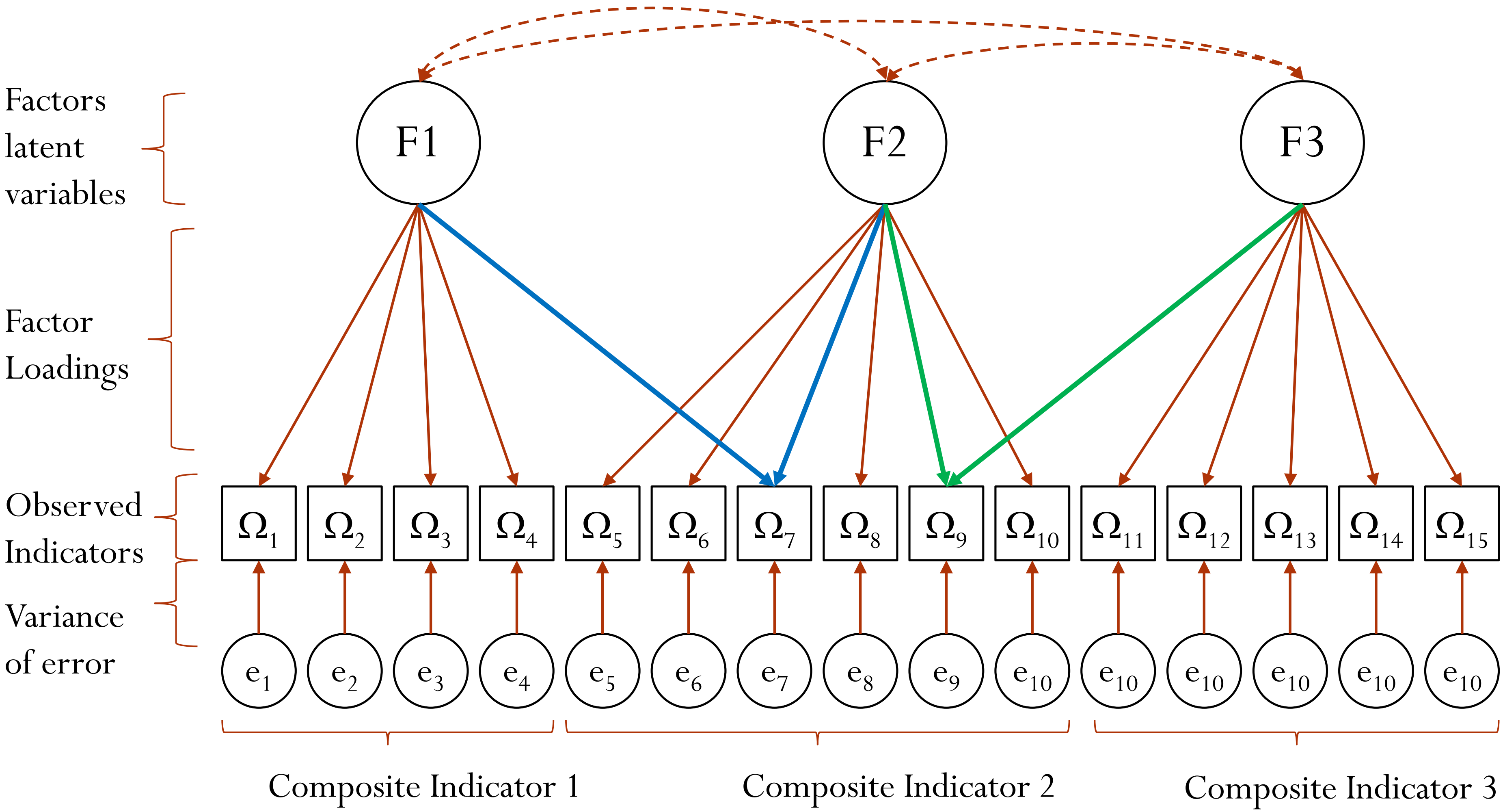
Cross-Loadings

- IDENTIFY the Cross-Loanding that most reduce BIC



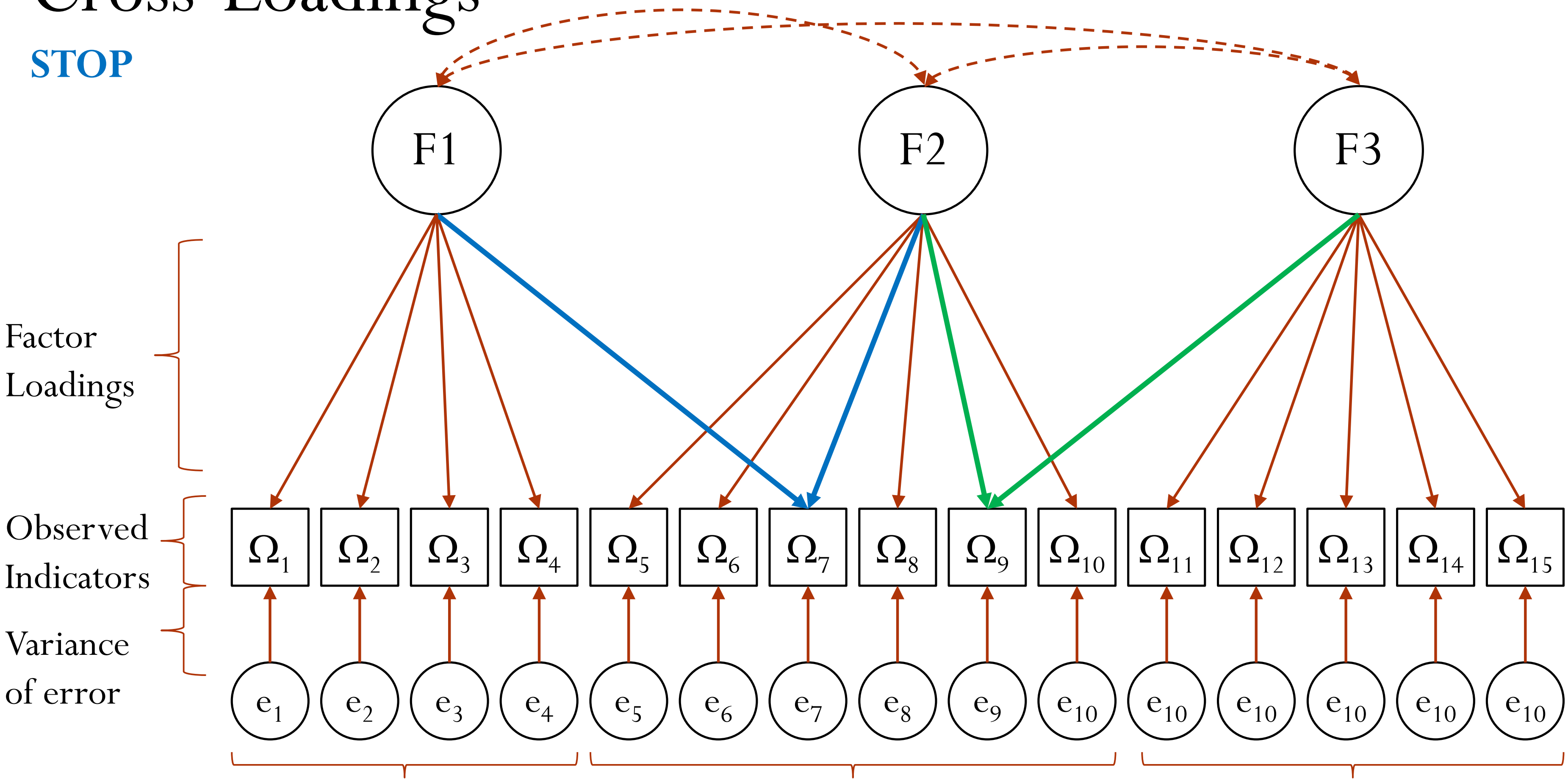
Cross-Loadings

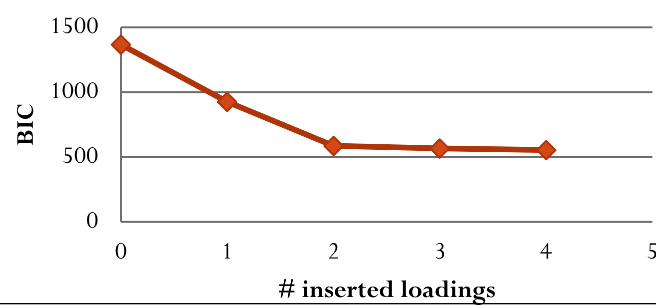
Continue to IDENTIFY Cross-Loadings that reduce BIC



Cross-Loadings

STOP

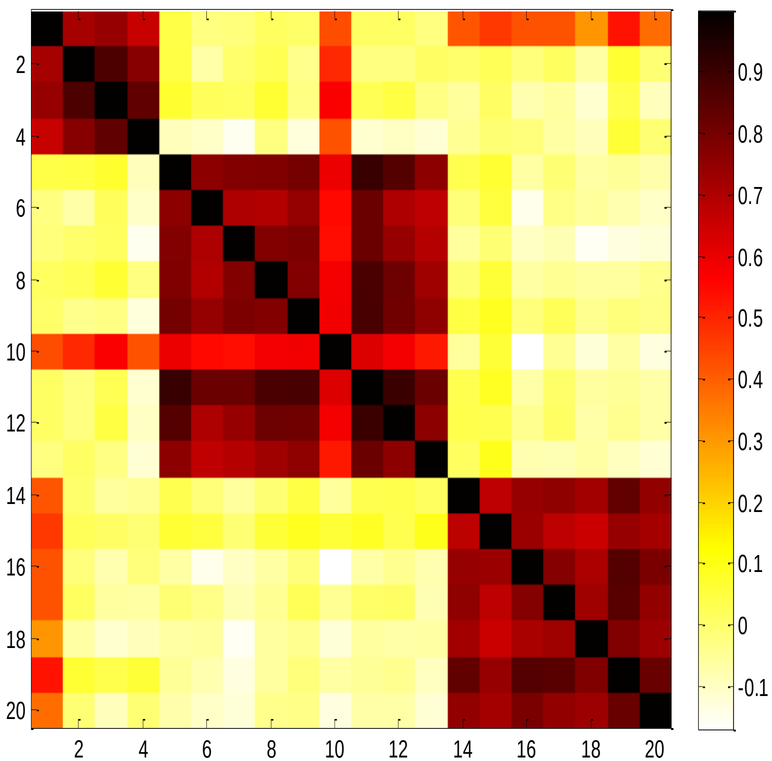




Cross-Loadings

Example: 20 x 20 correlation matrix

3 blocks generated according DFA + cross-loadings $a_{10,1}$, $a_{1,3}$



Estimation:

first estimate the best Simple Structure Model [3 blocks (V1-V4), (V5-13), (V14-V20)]

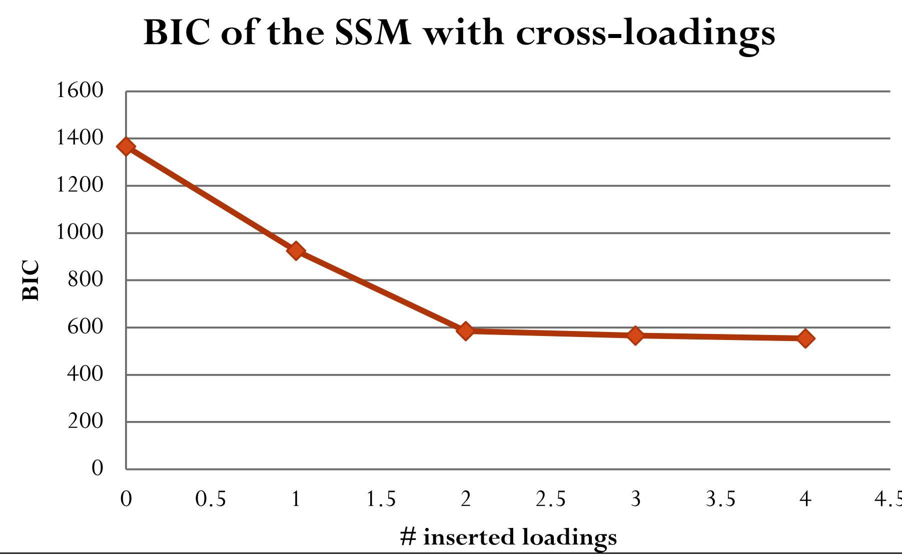
BIC = 1366, AIC = 785

First cross-loading estimated $a_{10,1}$ =0.57;

BIC = 925, AIC = 785

Second cross-loading estimated $a_{1,3}$ =0.52;

BIC = 585, AIC = 445



Hierarchical Disjoint Non-Negative Factor Analysis

The **General factor frequently corresponds** to a **Composite indicator** where **each subset of variable is consistent and reliable, thus the loadings must be positive**

Recall that the discrepancy function $D(\mathbf{B}, \hat{\mathbf{V}}, \hat{\mathbf{\Psi}})$ is minimized with respect to $\mathbf{B}_h = diag(\mathbf{b}_h)$ by

$$\widehat{\boldsymbol{b}_h} = \hat{\mathbf{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} \mathbf{u}_{1h} (\lambda_{1h} - 1)^{\frac{1}{2}} \quad h=1,\dots,H.$$

where λ_{1h} is the largest eigenvalue and $\mathbf{u}_{1h}$ is the corresponding eigenvector of the variance-covariance matrix $\hat{\mathbf{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} \mathbf{S}_h \hat{\mathbf{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}}$ corresponding to variables identified by $\mathbf{v}_{\cdot h}$, that corresponds to h-th column of $\mathbf{V}$. Values λ_{1h} and $\mathbf{u}_{1h}$ minimize $\|\hat{\mathbf{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} \mathbf{S}_h \hat{\mathbf{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} - \lambda_{1h} \mathbf{u}_{1h} \mathbf{u}_{1h}'\|^2$, or equivalently

$$\| \mathbf{X}_h \hat{\mathbf{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} - \sqrt{\lambda_{1h}} \mathbf{y}_h \mathbf{u}_{1h}' \|^2, \qquad (52)$$

where $\mathbf{X}_h$ is the centered data matrix formed by variables identified by $\mathbf{v}_{\cdot h}$ and $\mathbf{y}_h$ is the factor score vector.

ariables.

$$\| \mathbf{X}_h \widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} - \sqrt{\lambda_{1h}}\mathbf{y}_h \mathbf{u}_{1h}' \|^2, \quad (52)$$

SOLUTION OF a REGRESSION PROBLEM

can be solved by an **ALS algorithm that alternates two regression problems**.

Given $\hat{\mathbf{u}}_{1h}$ compute $\mathbf{y}_h$ by

$$\mathbf{y}_h = \mathbf{X}_h \widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} \hat{\mathbf{u}}_{1h} (\hat{\mathbf{u}}_h' \hat{\mathbf{u}}_{1h})^{-1}. \quad (53)$$

Given $\hat{\mathbf{y}}_h$ compute $\mathbf{u}_{1h}$ by

$$\mathbf{u}_{1h} = \widehat{\boldsymbol{\Psi}}_h^{-\frac{1}{2}} \mathbf{X}_h' \hat{\mathbf{y}}_h (\hat{\mathbf{y}}_h' \hat{\mathbf{y}}_h)^{-1}. \quad (54)$$

At each reiteration of the two steps (53) and (54), the loss function (52) decreases or at least does not increase. The algorithm stops when function (52) decreases less than a positive arbitrary constant.

Now **vector $\mathbf{u}_{1h}$ must be non-negative**.

The **solution** can be **found** by the **Non-Negative LS Algorithm** (Lawson and Hanson, 1974).
This is an **active set algorithm**, where the H inequality constraints are active if $\mathbf{u}_{1h}'$ are negative (or zero) when estimated unconstrained, otherwise constraints are passive.
The non-negative **solution** of (52) with respect to $\mathbf{u}_{1h}$ is the **unconstrained least squares solution using only the variables of the passive set**, setting the regression coefficients of the active set to zero. Therefore

$$\mathbf{u}_{1h} = \begin{cases} \widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} \mathbf{X}_{h+}' \hat{\mathbf{y}}_h (\hat{\mathbf{y}}_h' \hat{\mathbf{y}}_h)^{-1} \\ 0 \quad \text{otherwise} \end{cases}. \quad (55)$$

where $\mathbf{X}_{h+}$ is the set of passive variables.