Exploratory Hierarchical Factor Analysis

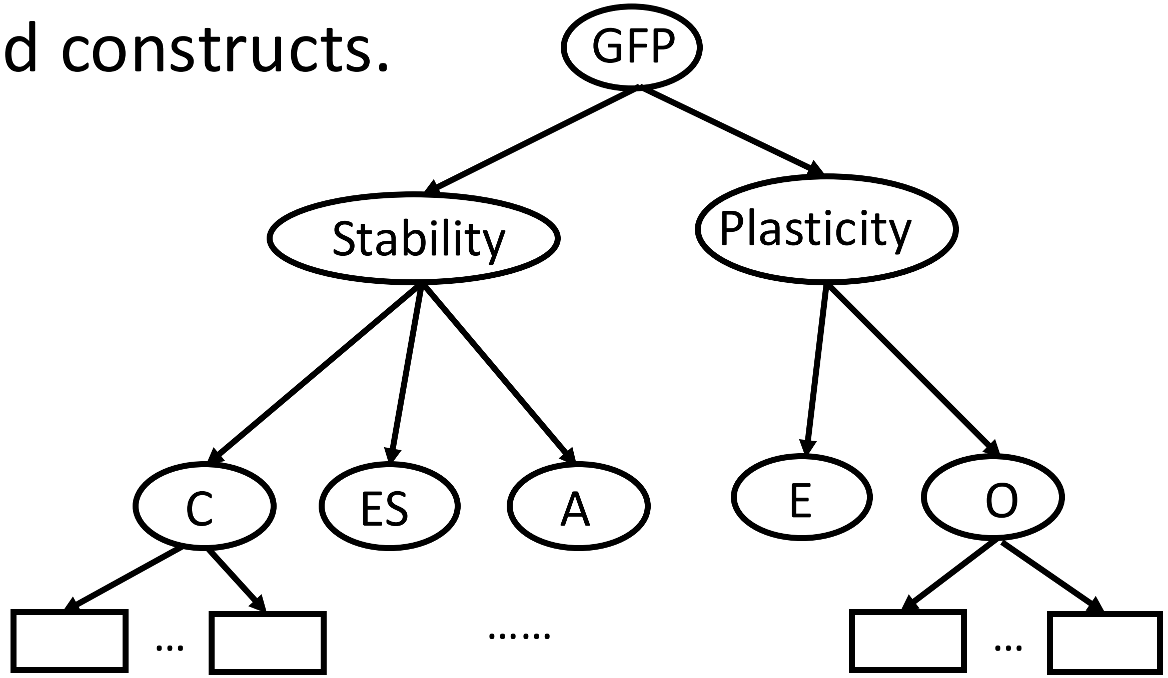
A Divide-and-Conquer Approach with Theoretical Guarantees

Yunxiao Chen, Department of Statistics, LSE

Joint work with Jiawei Qiao (Fudan U) and Zhiliang Ying (Columbia U)

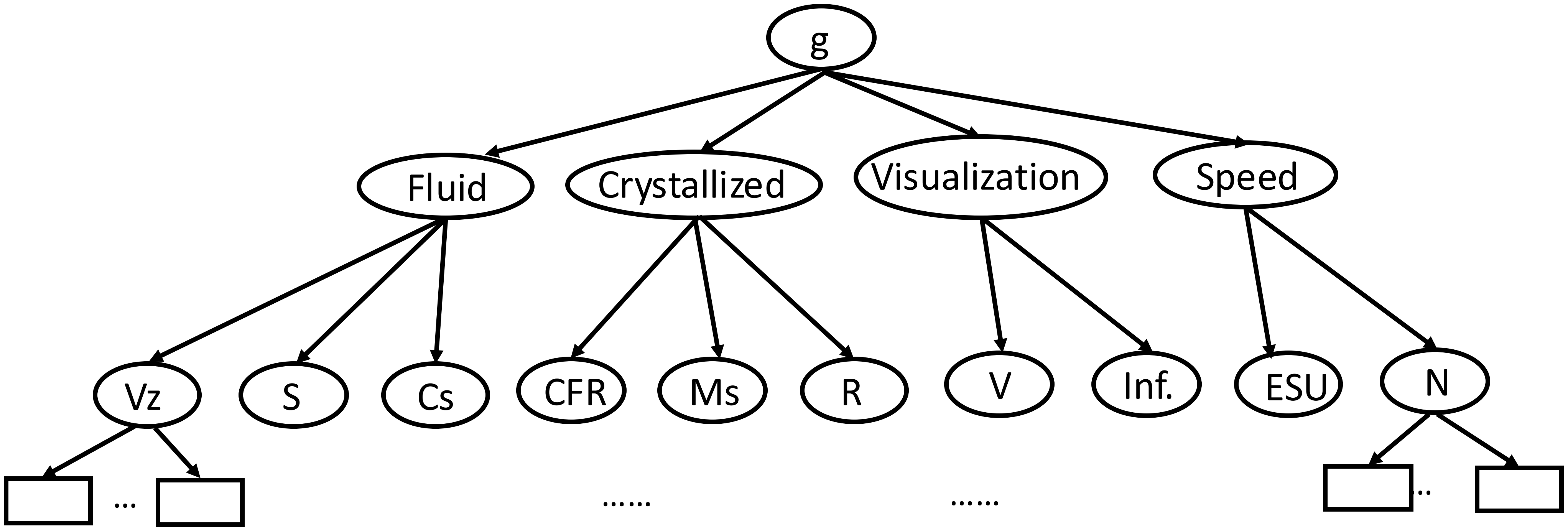
Background: Hierarchical Factors (con'd)

- Many social and behavioural theories assume hierarchically structured constructs.



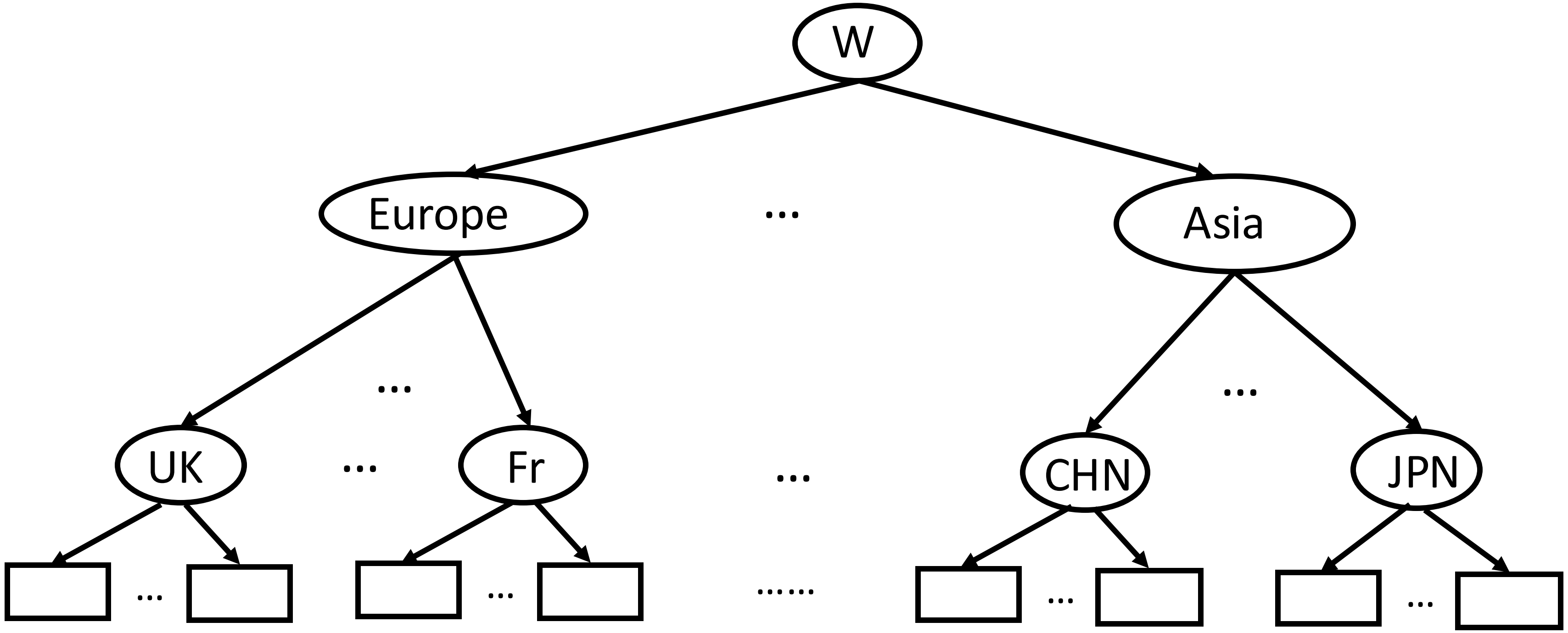
A hierarchical factor organization of human personality (Rushton and Irwing, 2008)

Background: Hierarchical Factors



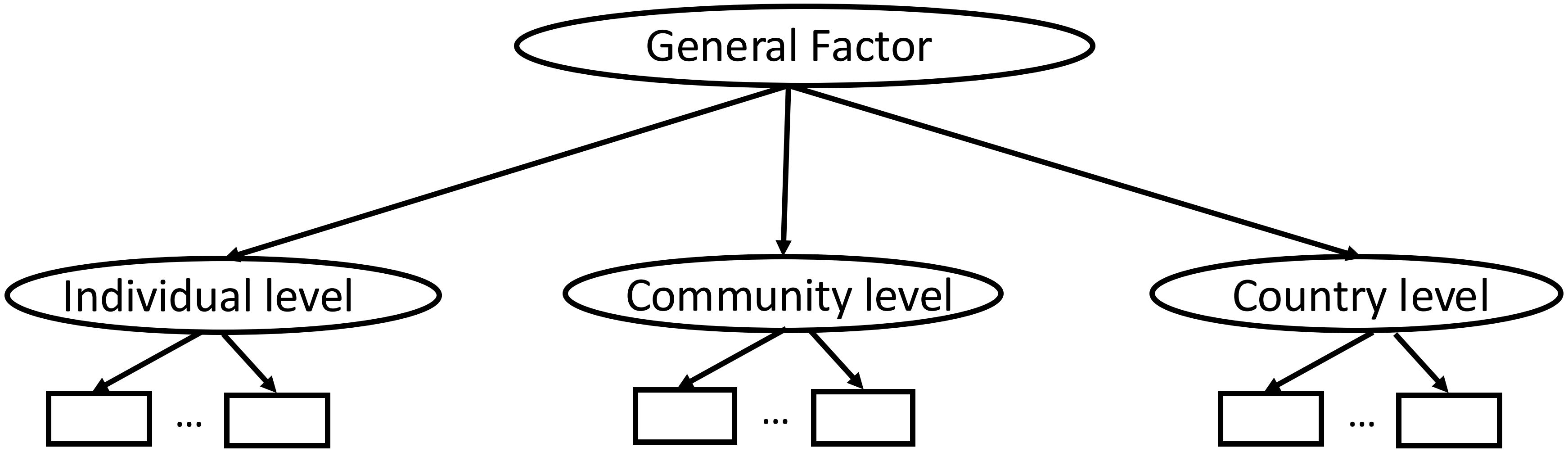
A (simplified) hierarchical factor organization of cognitive abilities (Undheim and Gustafasson, 1987)

Background: Hierarchical Factors



A hierarchical factor organization of international economic time series (Kose, Otrok, and Whiteman, 2003, 2008)

Background: Hierarchical Factors



A bi-factor factor organization (i.e., a two-layer hierarchical factor structure) of subjective well-being (Torres-Vallejo, et al., 2021)

Goals and Challenges

Goals:

- Learning the hierarchical factor structure from data without prior knowledge
- Provide theoretical guarantees to our learning algorithm

Challenge:

- Computational challenges brought by the combinatorial nature of the problem (similar to a hierarchical clustering problem).

Hierarchical Factor Model

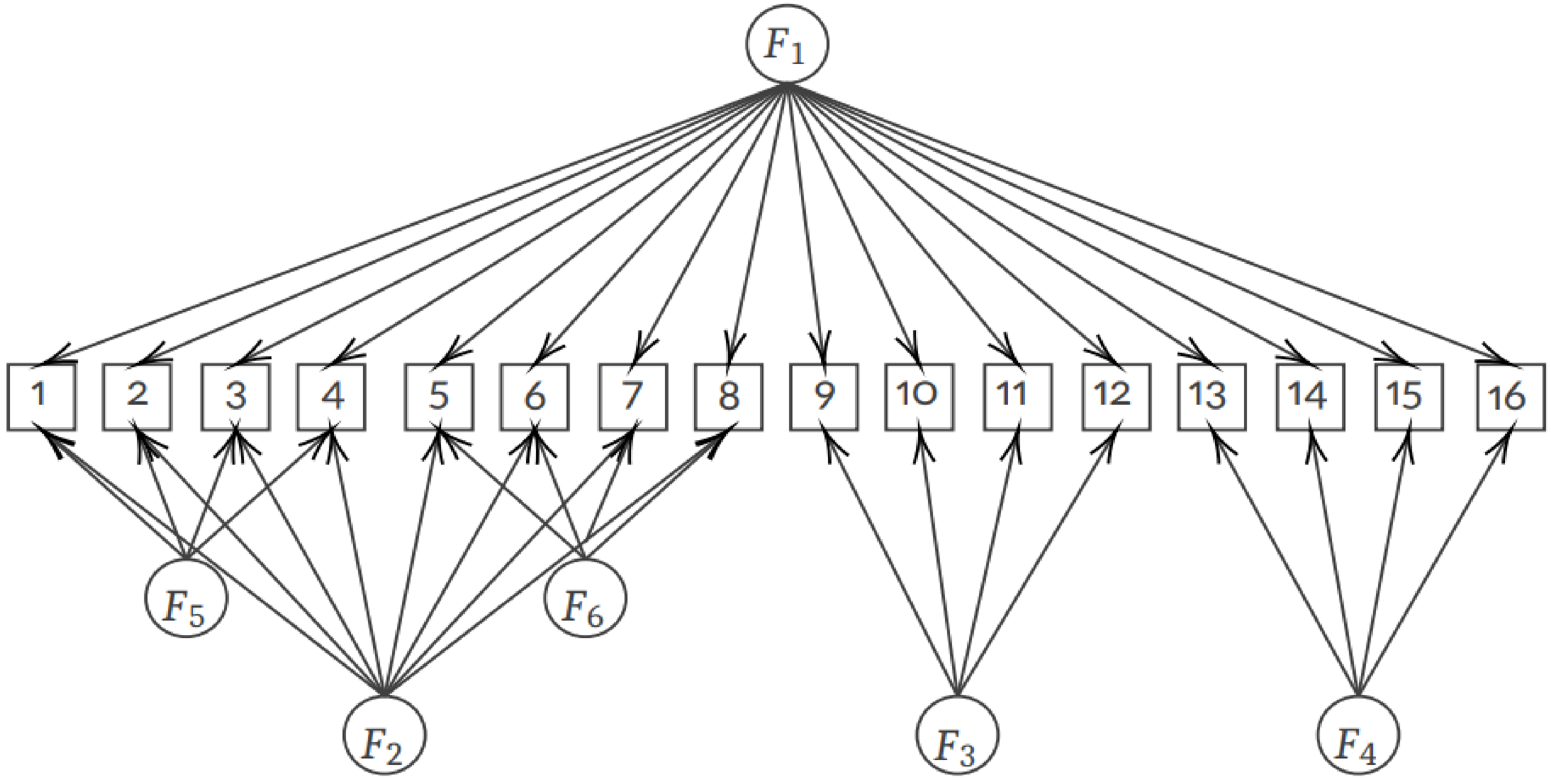
- A hierarchical factor model is a constrained linear factor model.
- Suppose we observe i.i.d. observations of $Y \in R^J$, with $Cov(Y) = \Sigma$.
- A linear factor model (with orthogonal factors) assumes that

$$\Sigma = \Lambda\Lambda^\top + \Psi,$$

where Λ is a low-dimensional constrained loading matrix and Ψ is a diagonal matrix capturing the unique variances.

Hierarchical Factor Model (con'd)

$$\Lambda = \begin{pmatrix}
\lambda_{11} & \lambda_{12} & 0 & 0 & \lambda_{15} & 0 \\
\lambda_{21} & \lambda_{22} & 0 & 0 & \lambda_{25} & 0 \\
\lambda_{31} & \lambda_{32} & 0 & 0 & \lambda_{35} & 0 \\
\lambda_{41} & \lambda_{42} & 0 & 0 & \lambda_{45} & 0 \\
\lambda_{51} & \lambda_{52} & 0 & 0 & 0 & \lambda_{56} \\
\lambda_{61} & \lambda_{62} & 0 & 0 & 0 & \lambda_{66} \\
\lambda_{71} & \lambda_{72} & 0 & 0 & 0 & \lambda_{76} \\
\lambda_{81} & \lambda_{82} & 0 & 0 & 0 & \lambda_{86} \\
\lambda_{91} & 0 & \lambda_{93} & 0 & 0 & 0 \\
\lambda_{10,1} & 0 & \lambda_{10,3} & 0 & 0 & 0 \\
\lambda_{11,1} & 0 & \lambda_{11,3} & 0 & 0 & 0 \\
\lambda_{12,1} & 0 & \lambda_{12,3} & 0 & 0 & 0 \\
\lambda_{13,1} & 0 & 0 & \lambda_{13,4} & 0 & 0 \\
\lambda_{14,1} & 0 & 0 & \lambda_{14,4} & 0 & 0 \\
\lambda_{15,1} & 0 & 0 & \lambda_{15,4} & 0 & 0 \\
\lambda_{16,1} & 0 & 0 & \lambda_{16,4} & 0 & 0
\end{pmatrix}$$



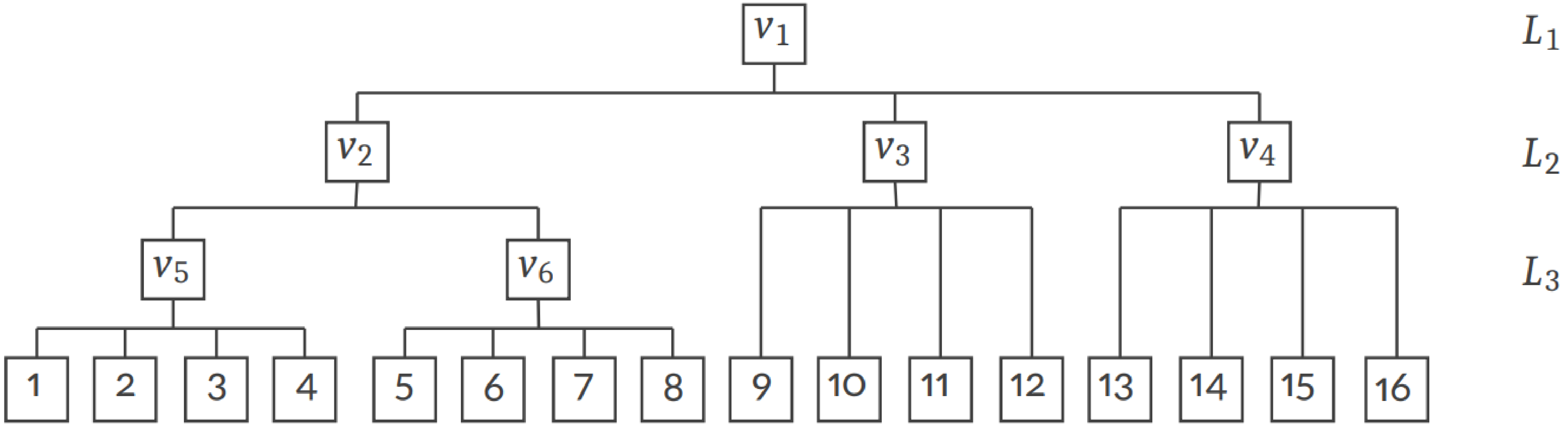
Hierarchical Factor Model (con’d)

- Let $v_k = \{j : \lambda_{jk} \neq 0\}$ be the variables loading on the kth factor. A hierarchical factor structure imposes following constraints on $v_k, k = 1, \ldots, K$:
  - C1. $v_1 = \{1, \ldots, J\}$ corresponds to a general factor that is loaded on by all the items.
  - C2. For any $k < l$, it holds that either $v_l \subset v_l$ or $v_l \cap v_l = \emptyset$. When $v_l \subset v_k$, we say factor l is a descendent factor of factor k. If further that there does not exist k’ such that $k < k’ < l$ and $v_l \subset v_{k'} \subset v_k$, we say factor l is a child factor of factor k, and factor k is a parent factor of factor l.

Hierarchical Factor Model (con'd)

- C3. For a given factor k, we denote all its child factors as Ch_k. Then its cardinality $|\mathrm{Ch}_k|$ satisfies that $|\mathrm{Ch}_k| = 0$ or $|\mathrm{Ch}_k| \geq 2$. When a factor k has two or more child factors, these child factors satisfy that $v_l \cap v_{l'} = \emptyset$, for any $l, l' \in \mathrm{Ch}_k$ and $\cup_{l \in \mathrm{Ch_k}} v_l = v_k$.

Hierarchical Factor Model (con'd)



Learning Factor Hierarchy: Existing Method

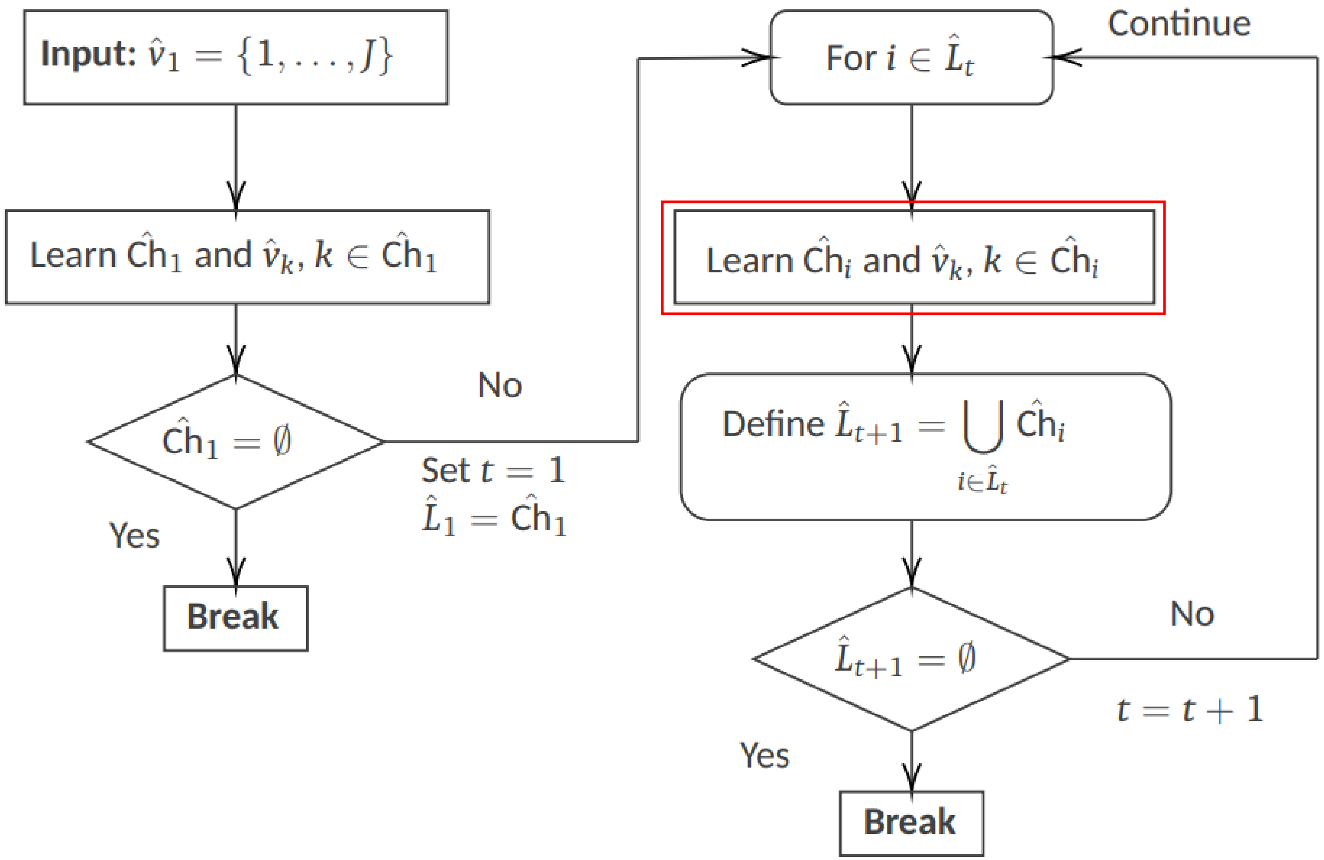
- Existing method for exploratory hierarchical factor analysis: Schmid-Leiman transformation (Schmid and Leiman, 1957).

  - Bottom-up construction by iteratively applying exploratory factor analysis
  - Leads to a rank-deficient loading matrix
  - Fail completely in some cases (Jennrich and Bentler, 2011)
  - No theoretical guarantee

Proposed Method

- Proposed divide-and-concur approach:
  - Top-down construction by iteratively solving equality constrained (nonconvex) optimization problem (using augmented Lagrangian method).
  - The number of factors in each layer is determined by BIC.
  - Assuming that global solutions are found for all the nonconvex optimization problems, together with some regularity conditions, as the sample size goes to infinity, we can consistently recover the hierarchical factor structure and its parameters.

Proposed Method (con'd)



Based on the submatrix of S, we answer the question:

1. How many child factors does factor i have?
2. What is the loading structure of these child factors.

A Special Case of Bifactor Model (Qiao et al., 2025+)

- Suppose that the factor hierarchy has two layers, and there are $K = 1 + G$ factors.
- We only need to solve the following problem to find the partition of the variables (associated with the group factors).

$$\begin{aligned} \min_{\Lambda,\Psi} \quad & l(\Lambda\Lambda^\top + \Psi; S) \\ s.t. \quad & \lambda_{jk}\lambda_{jk'} = 0, \text{ for all } k, k' = 2, ..., G+1, k \neq k', j = 1, ..., J, \end{aligned}$$

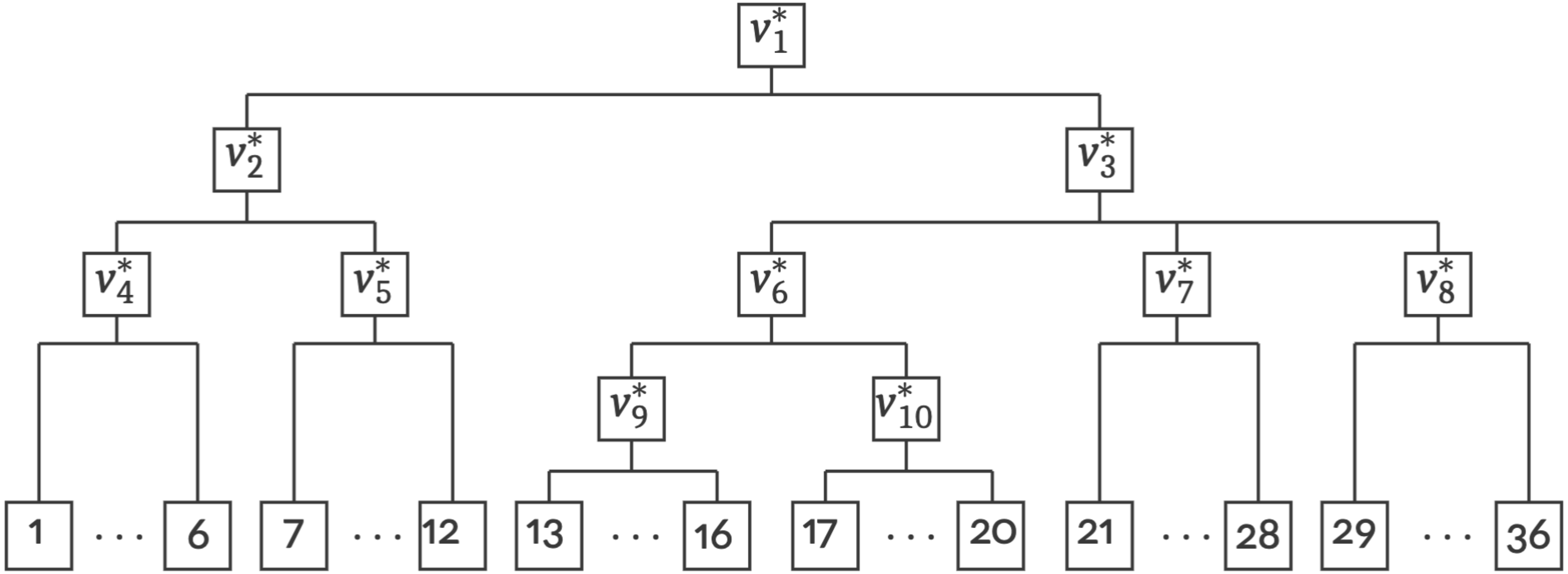
- This idea generalises to the hierarchical factor model (with a stepwise BIC-based procedure to adjust for the dimensions).

Theoretical Guarantee

Theorem: Assume that global solutions are found for all the nonconvex optimization problems. Under mild conditions, the proposed method is consistent. That is, as N goes to infinity, the probability of $\hat{T} = T$, $\hat{K} = K, \hat{L}_t = L_t, t = 1, \ldots, T$, and $\hat{v}_i = v_i^*, i = 1, \ldots, K$ goes to 1, and the corresponding parameters are root-N consistent.

Simulation Study

- True structure:



Simulation Study (con'd)

Table: The accurarcy of the overall estimates of hierarchical structure and parameters.

N	$\bar{K}$	$\bar{T}$	EMC	$\text{MSE}_{\hat{\Lambda}}$	$\text{MSE}_{\hat{\Psi}}$
500	10.10	4.00	0.90	3.22×10^{-3}	1.26×10^{-2}
2000	10.02	4.01	0.99	0.81×10^{-3}	3.01×10^{-3}

Table: The accurarcy of the estimated hierarchical structure on each layer.

N	$\lvert\hat{L}_2\rvert$	LMC_2	$\lvert\hat{L}_3\rvert$	LMC_3	$\lvert\hat{L}_4\rvert$	LMC_4
500	2.00	0.99	4.91	0.90	2.19	0.90
2000	2.00	0.99	4.97	0.99	2.03	0.99

Real Data Analysis

- Data: A sample of 1655 UK participants aged between 30 and 50 years who answered an Agreeableness scale of personality (Johnson, 2014).
- The scale consists of 24 items, designed to measure six facets of Agreeableness, including Trust (A1), Morality (A2), Altruism (A3), Cooperation (A4), Modesty (A5) and Sympathy (A6)

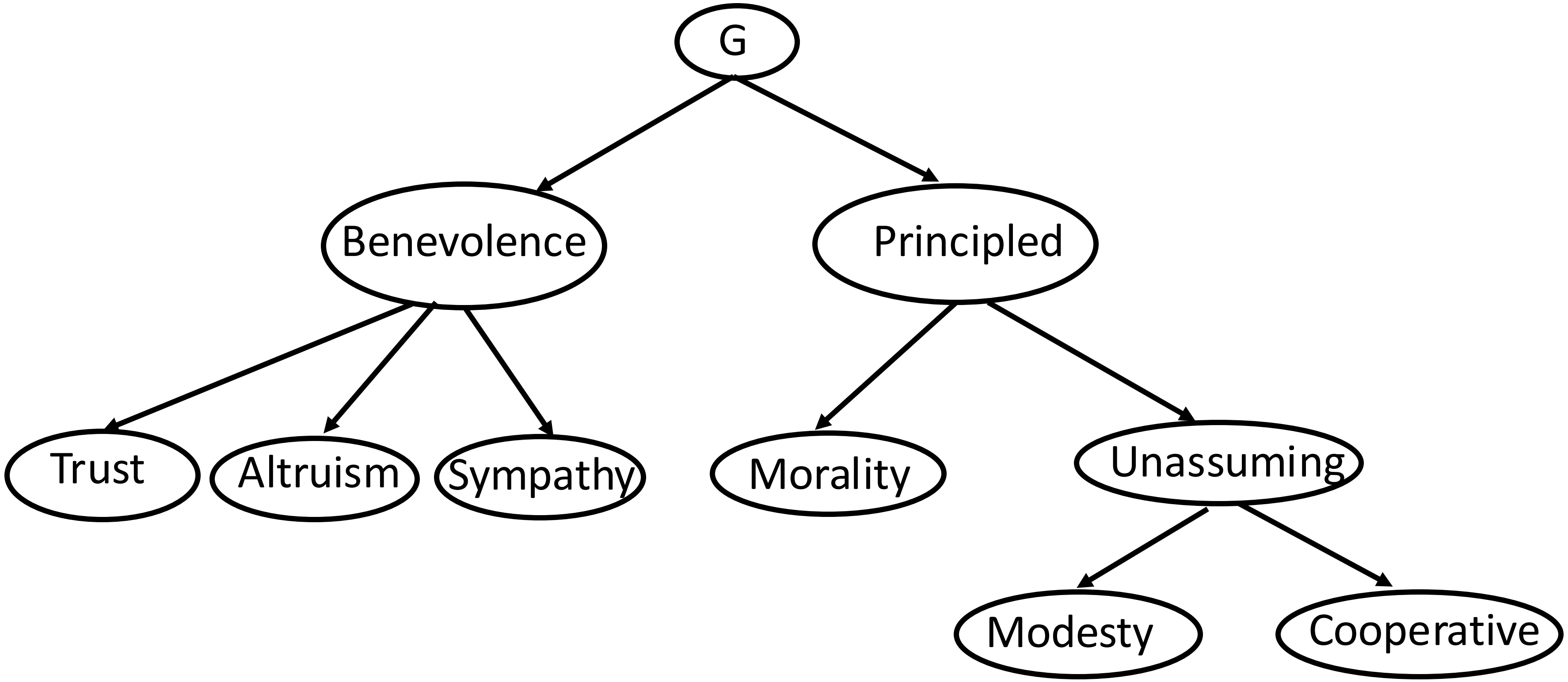
Real Data Analysis (con'd)

1	+	Trust(A1)	Trust others.
2	+	Trust(A1)	Believe that others have good intentions.
3	+	Trust(A1)	Trust what people say.
4	–	Trust(A1)	Distrust people.
5	–	Morality(A2)	Use others for my own ends.
6	–	Morality(A2)	Cheat to get ahead.
7	–	Morality(A2)	Take advantage of others.
8	–	Morality(A2)	Obstruct others' plans.
9	+	Altruism(A3)	Love to help others.
10	+	Altruism(A3)	Am concerned about others.
11	–	Altruism(A3)	Am indifferent to the feelings of others.
12	–	Altruism(A3)	Take no time for others.

Real Data Analysis (con'd)

13	–	Cooperation(A4)	Love a good fight.
14	–	Cooperation(A4)	Yell at people.
15	–	Cooperation(A4)	Insult people.
16	–	Cooperation(A4)	Get back at others.
17	–	Modesty(A5)	Believe that I am better than others.
18	–	Modesty(A5)	Think highly of myself.
19	–	Modesty(A5)	Have a high opinion of myself.
20	–	Modesty(A5)	Boast about my virtues.
21	+	Sympathy(A6)	Sympathize with the homeless.
22	+	Sympathy(A6)	Feel sympathy for those who are worse off than myself.
23	–	Sympathy(A6)	Am not interested in other people's problems.
24	–	Sympathy(A6)	Try not to think about the needy.

Real Data Analysis (con'd)



Item	Facet	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
1	A1	0.47	0.14	0	0.70	0	0	0	0	0	0
2	A1	0.36	0.23	0	0.59	0	0	0	0	0	0
3	A1	0.32	0.22	0	0.69	0	0	0	0	0	0
4	A1	0.59	0.11	0	0.64	0	0	0	0	0	0
5	A2	0.44	0	0.55	0	0	0	0.61	0	0	0
6	A2	0.46	0	0.27	0	0	0	0.34	0	0	0
7	A2	0.56	0	0.42	0	0	0	0.61	0	0	0
8	A2	0.45	0	0.21	0	0	0	0	-0.10	0.05	0
9	A3	0.26	0.37	0	0	0.48	0	0	0	0	0
10	A3	0.26	0.54	0	0	0.16	0	0	0	0	0
11	A3	0.46	0.51	0	0.11	0	0	0	0	0	0
12	A3	0.43	0.34	0	0	0.21	0	0	0	0	0
13	A4	0.21	0	0.48	0	0	0	0	-0.02	0	0.42
14	A4	0.46	0	0.14	0	0	0	0	-0.15	0	0.66
15	A4	0.63	0	0.21	0	0	0	0	0.00	0	0.48
16	A4	0.57	0	0.34	0	0	0	0	-0.21	0	0.20
17	A5	0.36	0	0.43	0	0	0	0	0.68	-0.06	0
18	A5	-0.09	0	0.46	0	0	0	0	0.70	0.48	0
19	A5	0.06	0	0.49	0	0	0	0	0.86	0.41	0
20	A5	0.29	0	0.43	0	0	0	0	0.15	0	0.13
21	A6	0.23	0.44	0	0	0	0.75	0	0	0	0
22	A6	0.22	0.56	0	0	0	0.41	0	0	0	0
23	A6	0.41	0.57	0	0.04	0	0	0	0	0	0
24	A6	0.34	0.40	0	0	0	0.38	0	0	0	0

Real Data Analysis (con'd)

- Comparison with other models

The BICs of the hierarchical factor model and the competing models

	HF	CFA	CBF	EBF
BIC	102,809.67	103,626.55	103,022.54	102,865.06

Thank you!