Bayesian Lasso Factor Analysis Models with Ordered Categorical Data

報告人：張瀝今
中山大學心理學系
潘俊豪 中山大學心理學系
王雨吟 中山大學心理學系

第十三屆海峽兩岸心理與教育測驗學術研討會, 臺灣

CFA

- Confirmatory factor analysis (CFA) is a commonly used technique for studying theory-driven hypotheses regarding observed and unobserved variables.

  - The number of factors is assumed to be known a priori and is thus treated as a fixed number in the estimation procedure.

  - The observed variables have zero loadings on all factors except the appropriate ones, which are a priori specified in the model.

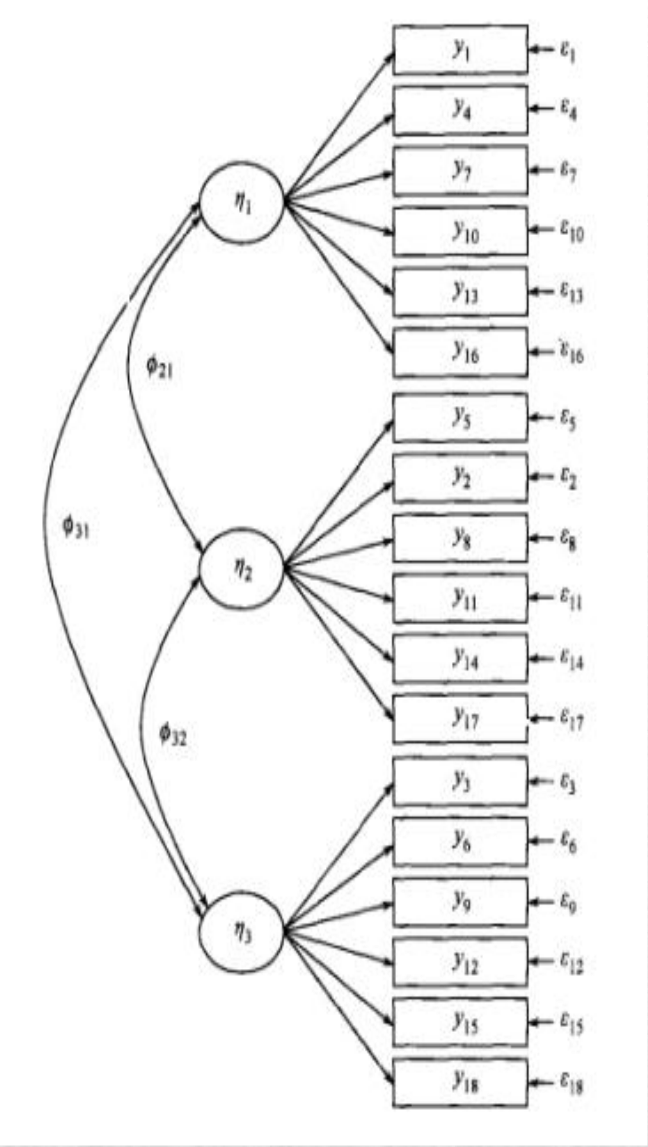
CFA

Suppose $y_1, y_2, \cdots, y_n$ are independent random observations, and each y_i = $(y_{i1}, y_{i2}, \cdots, y_{ip})^T$ satisfies the following factor analysis model:

$$y_i = \mu + \Lambda\omega_i + \varepsilon_i \quad i=1,2,\cdots,n,$$

where:

- $\boldsymbol{\mu}$: intercepts.
- **Λ**: factor loading matrix.
- $\boldsymbol{\omega_i}$: latent factors, ~N[0,Φ].
- $\boldsymbol{\varepsilon_i}$: measurement errors , ~N[0, Ψ].



Bayesian Approach

- The residual covariance structure in the error terms across all variables is not necessarily diagonal, but only a few of the off-diagonal elements are bounded away from zero.

  - In particular, these covariance terms are treated as nuisance parameters : deviations from the theory-driven factor model that need to be accounted for but otherwise of little substantive interest(Ip et al., 2004).

Bayesian Lasso

- The Bayesian Lasso (Least Absolute Shrinkage and Selection Operator) approach is implemented.
- The entire residual covariance matrix for all observed measures are modeled as a sparse structure that contains only a few covariance entries bounded away from zero (Pan, Ip and Dubé, 2017).

Bayesian Lasso

- Provides a model that fits the researcher's prior beliefs better than ML.
- Avoids the post hoc guessing work regarding when to stop making changes, especially when the MIs identify a large number of potential changes to the CFA.
- Achieves model parsimony and maintains the positive definiteness property of the covariance matrix.

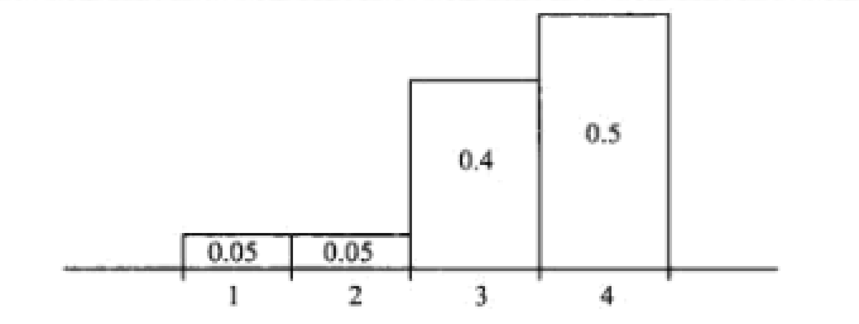
- This methods can only deal with the continuous data with normal distribution.

Ordered Categorical Data

- Most analyses of the model have been carried out under the framework of confirmatory factor analysis with the assumption that the observed variables are continuous and have normal distribution.
- To satisfy the assumption, most subjects are required to select intermediate options from all options.
- However, in practical applications, the histogram of most variable is biased.

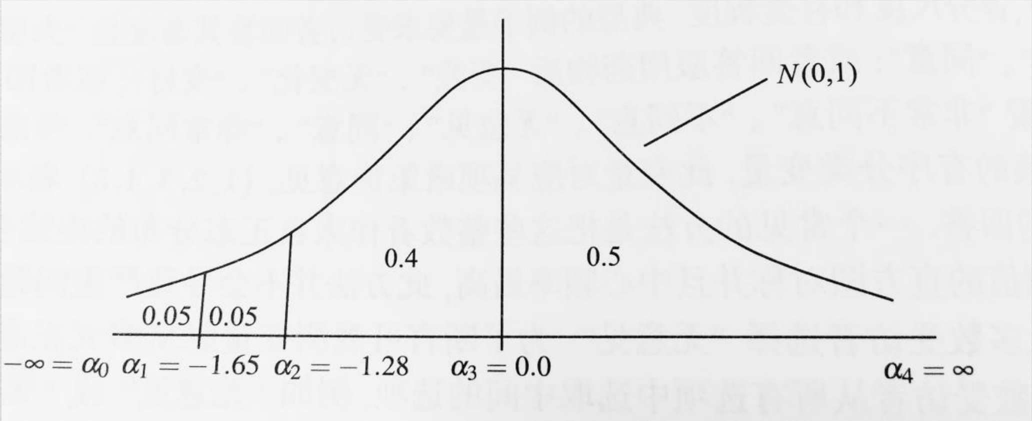
Implicit Normal Distribution

- Assume a data set has such biased histogram:



- The corresponding continuous measurements y ($y \sim N[0,1]$) are unobservable. The relationship between y and the observable variable: For k =1, 2, 3, 4, $\alpha_{k-1} < y < \alpha_k$,

$$-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 = +\infty$$



Model Description

- The corresponding continuous measurements $y_1, y_2, \cdots, y_n$ are unobservable, and their information is given by ordered categorical variables ***Z***.

- The relationship between Y and ***Z*** is defined by a set of unknown thresholds α_k as follows:

$$\mathbf{Z} = \begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix} \text{ if } \begin{cases} \alpha_{1,z_1} < y_1 \leq \alpha_{1,z_1+1}, \\ \vdots \\ \alpha_{s,z_s} < y_s \leq \alpha_{s,z_s+1}, \end{cases}$$

- where z_k is an integral value in {0, 1, . . . , b_k} and α_k= {$\alpha_{k,1}$, . . ., α_{k,b_k}

In general, we set $\alpha_{k,0} = -\infty$, $\alpha_{k,b_k+1} = +\infty$.

Model Description

- Each y_i = $(y_{i1}, y_{i2}, \cdots, y_{ik})^T$ satisfies the following factor analysis model:

$$y_i = \mu + \Lambda\omega_i + \varepsilon_i \quad i=1,2, \cdots, n, \quad y_k \sim N[\mu', \sigma^2]$$

- Because μ', σ^2, α_k are uncertain, models with ordered categorical variables are not identified without imposing identification conditions.

  To solve this problem, the $\alpha_{k,1}$ and α_{k,b_k} should be fixed.

Prior Distributions

For the structural parameters involved in $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Phi}$, the following conjugate prior distributions are assigned (Lee, 2007). For $k = 1, \cdots, p$,

$$\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \mathbf{H}_{\mu 0}), \quad \boldsymbol{\Lambda}_k \sim N(\boldsymbol{\Lambda}_{0k}, \mathbf{H}_{0k}), \quad \boldsymbol{\Phi}^{-1} \sim \mathrm{Wishart}(\mathbf{R}_0, \rho_0),$$

where $\boldsymbol{\Lambda}_k^T$ is the kth row of $\boldsymbol{\Lambda}$; $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_{0k}$, ρ_0, and positive definite matrices $\mathbf{H}_{\mu 0}$, $\mathbf{H}_{0k}$, and $\mathbf{R}_0$ are hyperparameters whose values are assumed to be given.

Fix $\alpha_{k,1} = \Phi^{*-1}(f_{k,1}^*)$, $\alpha_{k,b_k} = \Phi^{*-1}(f_{k,b_k}^*)$, the following non-informative prior distribution is used for other threshold:

$p(\alpha_k) = p(\alpha_{k,2}, \ldots, \alpha_{k,b_k-1}) \propto c$, where c is a constant.

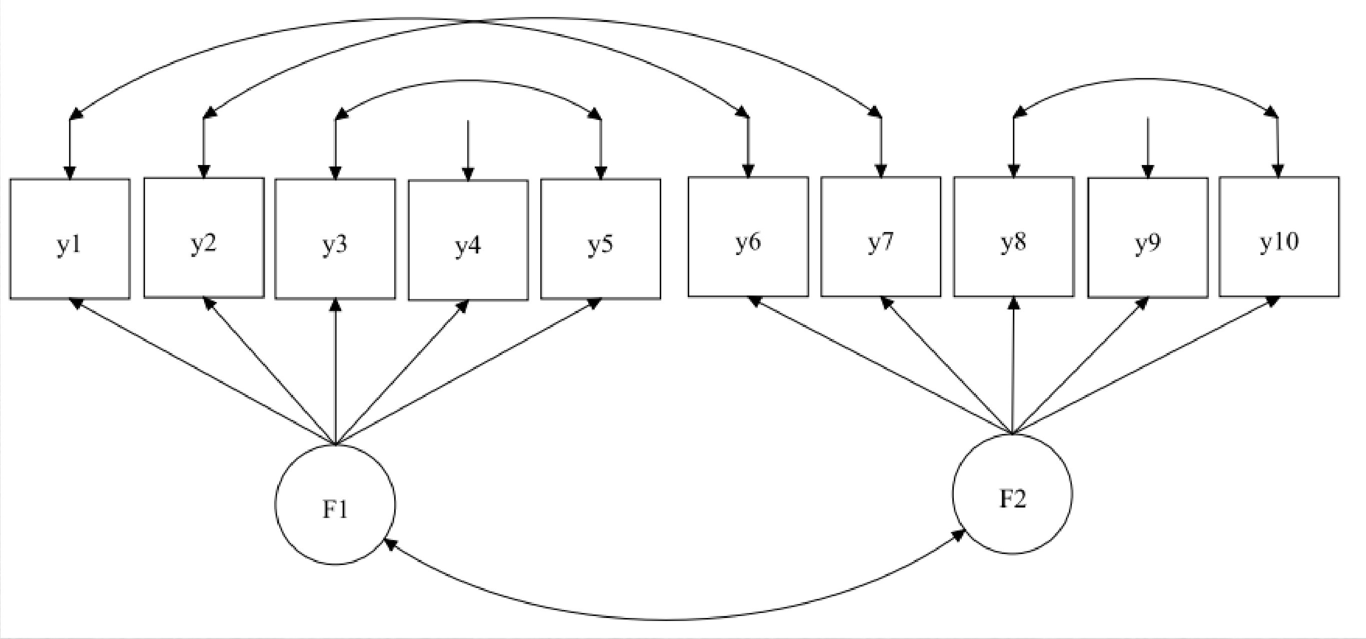
Bayesian Inference

- The following Gibbs sampler algorithm and the Metropolis Hastings (MH) algorithm for generating observations is proposed:

1. Generate Ω from $p(\Omega|\alpha, \mathrm{Y}, \mathrm{Z}, \mu, \Lambda, \Phi, \Psi)$;
2. Generate Λ from $p(\Lambda|\alpha, \mathrm{Y}, \mathrm{Z}, \Omega, \mu, \Phi, \Psi)$;
3. Generate Φ from $p(\Phi|\alpha, \mathrm{Y}, \mathrm{Z}, \Omega, \mu, \Lambda, \Psi)$;
4. Generate μ from $p(\mu|\alpha, \mathrm{Y}, \mathrm{Z}, \Omega, \Phi, \Lambda, \Psi)$;
5. Generate ∑ from $p(\sum|\alpha, \mathrm{Y}, \mathrm{Z}, \Omega, \mu, \Phi, \Lambda, \tau, \lambda)$ and compute $\Psi = \sum^{-1}$;
   - A block Gibbs sampler (Pan, Ip and Dubé, 2017).
6. Generate (α, Y) from $p(\alpha, \mathrm{Y}|\mathrm{Z}, \Omega, \mu, \Phi, \Lambda, \Psi)$.
   - Metropolis Hastings (MH) algorithm (Lee, 2007).

Simulation Studies

- The main purpose of these studies is to evaluate the parameter recovery of the proposed CFA model.
- A data set was simulated based on the model specified in equation ($y_i = \mu + \Lambda\omega_i + \varepsilon_i$). The number of observed variables set at *p* = 10 and the number of factors set at *q* = 2.



Simulation Studies

- The true values of the structural parameters were set as follows:

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots \mu_{10})^T = (0.5, 0.5, \cdots 0.5)^T$$

- The continuous measurements y_k are transformed to ordered categorical observations via with the following thresholds:

$$\alpha_k = (-0.5, 0.2, 0.8, 1.5). \quad k = 1, 2, \dots, 10.$$

- It was assumed that equation contained the following structure:

$$\boldsymbol{\Lambda}^T = \begin{pmatrix} 1.0 & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & \lambda_{72} & \lambda_{82} & \lambda_{92} & \lambda_{10,2} \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

Simulation Studies

- It was assumed that equation contained the following structure:
- Two levels of sample sizes were used, N = 200 and N = 400.

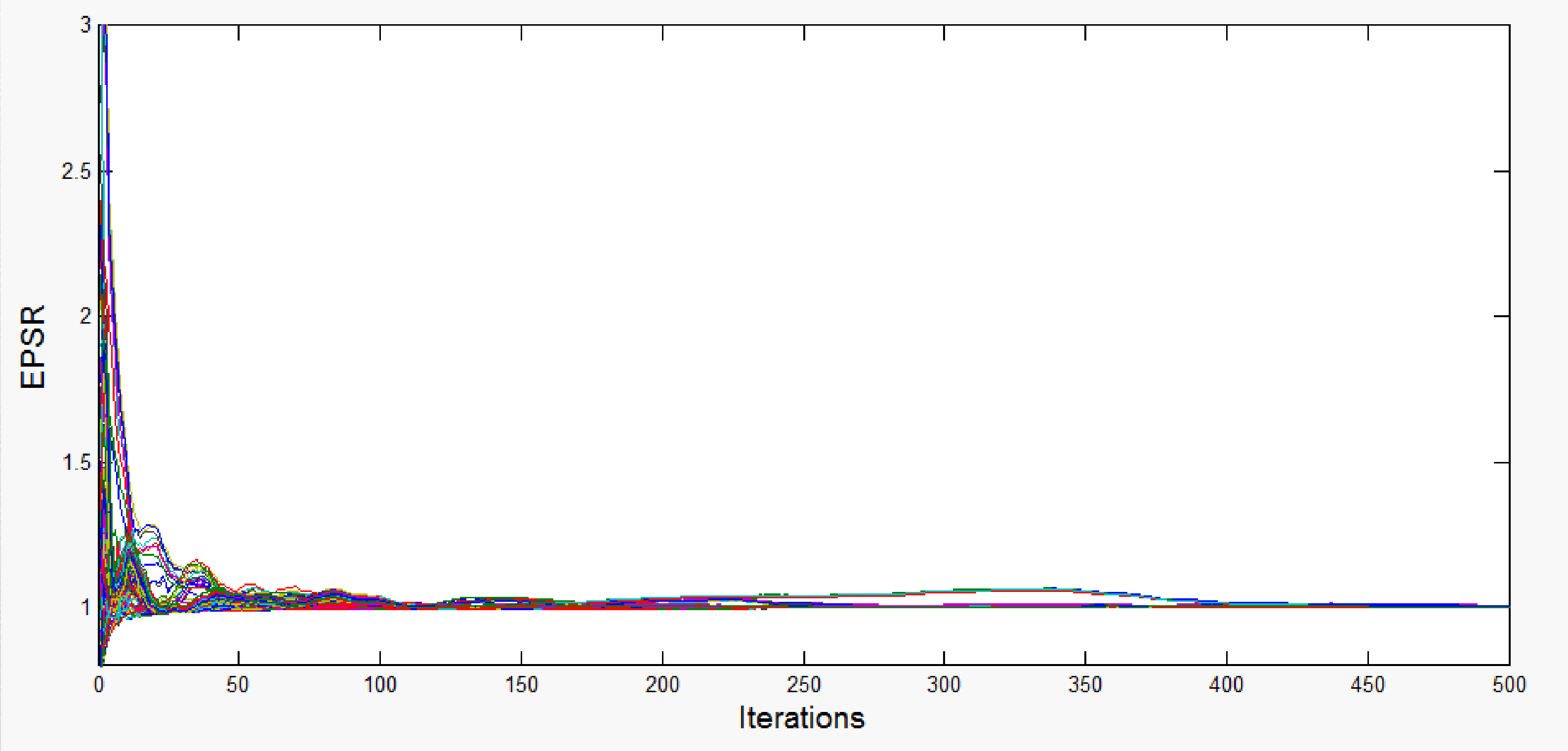
$$\mathbf{\Psi}=\begin{pmatrix}
\psi_{11} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{16} & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & \psi_{22} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{27} & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & \psi_{33} & 0.0 & \psi_{35} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & \psi_{44} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & \psi_{53} & 0.0 & \psi_{55} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
\psi_{61} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{66} & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & \psi_{72} & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{77} & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{88} & 0.0 & \psi_{8,10} \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{99} & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \psi_{10,8} & 0.0 & \psi_{10,10}
\end{pmatrix}.$$

Simulation Studies

- Additionally, a sensitivity analysis regarding inputs in the prior distributions was conducted by perturbing the prior input as follows:
  - Perturbed Input I: The elements in $\boldsymbol{\mu}_0$ and $\boldsymbol{\Lambda}_{0k}$ were taken as 0.0, and $\mathbf{H}_{\mu 0}$ and $\mathbf{H}_{0k}$ were taken as diagonal matrices with all diagonal elements equal to 4.0 in appropriate order. $\rho_0 = 6$, $\mathbf{R}_0 = 6\mathbf{I}_2$, where $\mathbf{I}_2$ is an identity matrix, $\alpha_{\lambda 0} = 1$, and $\beta_{\lambda 0} = 0.01$.
  - Perturbed Input II: The elements in $\boldsymbol{\mu}_0$ and $\boldsymbol{\Lambda}_{0k}$ were taken as true value, and $\mathbf{H}_{\mu 0}$ and $\mathbf{H}_{0k}$ were taken as identity matrices in appropriate order. $\rho_0 = 3$, $\mathbf{R}_0 = 6\mathbf{I}_2$, where $\mathbf{I}_2$ is an identity matrix, $\alpha_{\lambda 0} = 1$, and $\beta_{\lambda 0} = 0.005$.

Simulation Studies

- It was observed that the algorithm converged in less than 1,000 iterations, as indicated by the EPSR values being less than 1.2.
- In the Bayesian approach, 5000 posterior samples of the model parameters were recorded after 5000 burn-in iterations.



Simulation Studies

- Based on 100 replications, the BIAS, SE, and RMS between the estimates and the true values were computed.
- Bayesian estimates of the unknown parameters are generally accurate.
- The analysis results are not sensitive to how prior inputs are specified.

		Perturbed Input I			Perturbed Input II		
Par	True	BIAS	SE	RMS	BIAS	SE	RMS
μ_1	0.5	0.010	0.095	0.105	0.012	0.092	0.104
μ_2	0.5	0.008	0.084	0.080	0.010	0.081	0.081
μ_3	0.5	0.007	0.084	0.084	0.008	0.081	0.085
μ_4	0.5	0.006	0.081	0.094	0.008	0.080	0.093
μ_5	0.5	0.006	0.080	0.081	0.005	0.079	0.082
μ_6	0.5	-0.005	0.094	0.086	0.001	0.090	0.089
μ_7	0.5	-0.001	0.083	0.084	0.004	0.080	0.081
μ_8	0.5	-0.013	0.080	0.086	-0.007	0.080	0.083
μ_9	0.5	0.004	0.079	0.086	0.007	0.080	0.089
μ_{10}	0.5	-0.011	0.078	0.096	-0.009	0.080	0.094
λ_{21}	0.8	0.014	0.074	0.061	0.030	0.076	0.068
λ_{31}	0.8	0.026	0.091	0.084	0.043	0.090	0.092
λ_{41}	0.8	0.023	0.084	0.084	0.040	0.090	0.091
λ_{51}	0.8	0.020	0.087	0.080	0.038	0.090	0.086
λ_{72}	0.8	0.014	0.074	0.069	0.031	0.076	0.077
λ_{82}	0.8	0.030	0.087	0.088	0.050	0.091	0.099
λ_{92}	0.8	0.015	0.083	0.091	0.032	0.089	0.097
$\lambda_{10,2}$	0.8	0.023	0.089	0.084	0.043	0.091	0.093

Simulation Studies

- Bayesian estimates of the unknown parameters are generally accurate.
- The analysis results are not sensitive to how prior inputs are specified.

φ_{11}	1.0	0.012	0.177	0.143	-0.041	0.165	0.141
φ_{22}	1.0	0.045	0.174	0.162	-0.013	0.170	0.154
φ_{12}	0.3	0.004	0.098	0.088	-0.011	0.092	0.084
ψ_{11}	0.36	0.046	0.087	0.076	0.048	0.085	0.078
ψ_{22}	0.36	0.016	0.071	0.062	0.016	0.068	0.062
ψ_{33}	0.36	0.003	0.074	0.070	0.003	0.070	0.070
ψ_{44}	0.36	0.010	0.067	0.062	0.010	0.067	0.064
ψ_{55}	0.36	0.010	0.070	0.077	0.010	0.070	0.076
ψ_{66}	0.36	0.044	0.087	0.082	0.045	0.085	0.084
ψ_{77}	0.36	0.028	0.071	0.073	0.028	0.070	0.073
ψ_{88}	0.36	0.008	0.071	0.075	0.007	0.071	0.075
ψ_{99}	0.36	0.016	0.071	0.063	0.014	0.068	0.062
$\psi_{10,10}$	0.36	0.013	0.072	0.077	0.012	0.072	0.077
ψ_{16}	0.3	0.001	0.070	0.052	0.001	0.067	0.052
ψ_{27}	0.3	-0.009	0.057	0.057	-0.009	0.055	0.057
ψ_{35}	0.3	-0.009	0.065	0.066	-0.009	0.063	0.066
$\psi_{8,10}$	0.3	-0.007	0.064	0.067	-0.008	0.064	0.067

Simulation Studies

- The result for N = 400 is similar to that for N = 200. For most unknown parameters, the BIAS, SE and RMS reduced when sample size was increased to N = 400.

	Perturbed Input Ⅰ			
Par	True	BIAS	SE	RMS
μ_1	0.5	0.010	0.095	0.105
μ_2	0.5	0.008	0.084	0.080
μ_3	0.5	0.007	0.084	0.084
μ_4	0.5	0.006	0.081	0.094
μ_5	0.5	0.006	0.080	0.081
μ_6	0.5	-0.005	0.094	0.086
μ_7	0.5	-0.001	0.083	0.084
μ_8	0.5	-0.013	0.080	0.086
μ_9	0.5	0.004	0.079	0.086
μ_{10}	0.5	-0.011	0.078	0.096
λ_{21}	0.8	0.014	0.074	0.061
λ_{31}	0.8	0.026	0.091	0.084
λ_{41}	0.8	0.023	0.084	0.084
λ_{51}	0.8	0.020	0.087	0.080
λ_{72}	0.8	0.014	0.074	0.069
λ_{82}	0.8	0.030	0.087	0.088
λ_{92}	0.8	0.015	0.083	0.091
$\lambda_{10,2}$	0.8	0.023	0.089	0.084

	Perturbed Input Ⅰ			
Par	True	BIAS	SE	RMS
μ_1	0.5	-0.002	0.066	0.068
μ_2	0.5	-0.002	0.058	0.055
μ_3	0.5	-0.004	0.057	0.062
μ_4	0.5	0.001	0.055	0.066
μ_5	0.5	-0.002	0.056	0.072
μ_6	0.5	-0.008	0.064	0.066
μ_7	0.5	-0.005	0.055	0.058
μ_8	0.5	-0.008	0.055	0.060
μ_9	0.5	0.003	0.056	0.060
μ_{10}	0.5	-0.004	0.055	0.057
λ_{21}	0.8	0.006	0.050	0.042
λ_{31}	0.8	0.009	0.060	0.058
λ_{41}	0.8	0.001	0.060	0.056
λ_{51}	0.8	0.014	0.050	0.065
λ_{72}	0.8	0.003	0.060	0.046
λ_{82}	0.8	0.010	0.060	0.061
λ_{92}	0.8	0.012	0.060	0.065
$\lambda_{10,2}$	0.8	0.012	0.060	0.063

Real Data Analysis

- We analyzed a self-compassion measure with the data collected from 363 participants. A 6-factor 26-item CFA model was assumed.
- self-kindness (5 items), self-judgement(5 items), common humanity(4 items), feelings of isolation(4 items), mindfulness(4 items) and identification with thoughts(4 items).

$$\Lambda^T = \begin{pmatrix} 1 & \lambda_{1,2} & \lambda_{1,3} & \lambda_{1,4} & \lambda_{1,5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{2,7} & \lambda_{2,8} & \lambda_{2,9} & \lambda_{2,10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{3,12} & \lambda_{3,13} & \lambda_{3,14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{4,16} & \lambda_{4,17} & \lambda_{4,18} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{5,20} & \lambda_{5,21} & \lambda_{5,22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{6,24} & \lambda_{6,25} & \lambda_{6,26} \end{pmatrix}$$

Real Data Analysis

- It can be seen that all factor loading estimates are statistically significant and substantial in magnitude, and the correlations between the factors were all significant.

Parameters	est	se
$\lambda_{1,2}$	1.009	0.087
$\lambda_{1,3}$	0.957	0.096
$\lambda_{1,4}$	0.794	0.118
$\lambda_{1,5}$	0.752	0.110
$\lambda_{2,7}$	0.530	0.088
$\lambda_{2,8}$	0.647	0.089
$\lambda_{2,9}$	0.679	0.087
$\lambda_{2,10}$	0.695	0.093
$\lambda_{3,12}$	0.778	0.124
$\lambda_{3,13}$	0.783	0.134
$\lambda_{3,14}$	1.021	0.126
$\lambda_{4,16}$	0.898	0.091
$\lambda_{4,17}$	0.755	0.092
$\lambda_{4,18}$	0.881	0.083
$\lambda_{5,20}$	0.880	0.115
$\lambda_{5,21}$	0.782	0.108
$\lambda_{5,22}$	0.965	0.113
$\lambda_{6,24}$	0.811	0.113
$\lambda_{6,25}$	0.893	0.117
$\lambda_{6,26}$	0.882	0.103

$$\Phi = \begin{pmatrix} 0.682 & & & & & \\ 0.312 & 0.816 & & & & \\ 0.406 & 0.165 & 0.541 & & & \\ 0.265 & 0.603 & 0.203 & 0.737 & & \\ 0.483 & 0.272 & 0.372 & 0.268 & 0.584 & \\ 0.235 & 0.557 & 0.144 & 0.516 & 0.271 & 0.633 \end{pmatrix}$$

Real Data Analysis

- 21 or 6.5% significant item-pair errors were detected out of 325 (=C(26,2)) simultaneously by our proposed method.

- Iteam 7: 當我感到不走運的時候，我經常勸勉自己其實很多人也正經歷黴運。Item 10: 當我感到自己在某些方面不足時，我儘量提醒自己大部分人和我一樣，都不完美。

- This method also circumvents the problem of having to handle correlated error terms sequentially in traditional post hoc modification approach.

Discussion

- By treating the ordered categorical data as observations that are coming from a hidden continuous normal distribution with a threshold specification, the proposed method extends the Bayesian Lasso CFA in Pan, Ip and Dubé (2017).
- The simulation studies and real data analysis show that Bayesian estimates of the unknown parameters are generally accurate. The analysis results are not sensitive to how prior inputs are specified.

Discussion

- The proposed method here can also be extended to handle Bayesian Lasso CFA with mixed continuous and ordered categorical variables.

- Due to the similarity of CFA with ordered categorical data and Item Response Theory (IRT), the proposed methodology has potential to deal the local dependence problem in IRT analysis.

Thank You!

中山大學 心理學系
DEPARTMENT OF PSYCHOLOGY
SUN YAT-SEN UNIVERSITY