LIMITED INFORMATION & PAIRWISE LIKELIHOOD ESTIMATION OF MIXED LATENT VARIABLE MODELS

Tasouli Foteini†

†Athens University of Economics and Business

Introduction

Latent Variable Models (LVMs) enable modelling relationships between observed variables (items) and latent variables (factors). We will focus on the estimation of LVMs with mixed continuous and ordinal items with a limited information method & pairwise maximum likelihood (PML).

Modelling Approach

Let p_1 continuous items w_j

Underlying Variable Approach

Let p_2 ordinal items y_i with m_i categories & y_i^* underlying *standard normal* variables such that:

$$y_i = c_i \quad \Leftrightarrow \quad \tau^{(y_i)}_{c_{i-1}} < y_i^* \le \tau^{(y_i)}_{c_i}$$

where $\tau^{(y_i)}$ *thresholds*:
$-\infty = \tau_0^{(y_i)} < \ldots < \tau_{m_i}^{(y_i)} = +\infty$

Let the p-dimensional ($p = p_1 + p_2$) vector $\boldsymbol{x} = (\boldsymbol{w}, \boldsymbol{y}^*)$, then the model is of the form:

$$\boldsymbol{x} = \Lambda \boldsymbol{z} + \boldsymbol{\epsilon}$$

where factors $\boldsymbol{z} \sim N_q(\mathbf{0}, \boldsymbol{\Phi})$, errors $\epsilon \sim N_p(\mathbf{0}, \boldsymbol{\Psi})$, $\boldsymbol{\epsilon}$,$\boldsymbol{z}$ independent & Λ a pxq matrix of *factor loadings*

Maximization of log-likelihood requires the calculation of p_2-dimensional integrals, therefore...

Limited Information Method

Unweighted Least Squares (ULS)

Estimation in three stages [1]:

- Thresholds $\tau^{(y_i)}$
- Polychoric, polyserial & Pearson correlations
- Minimization of a fit function F:
  $$F(\boldsymbol{\theta}) = (S - \Sigma(\boldsymbol{\theta}))' W^{-1} (S - \Sigma(\boldsymbol{\theta}))$$
  where S the vector of sample statistics, $\Sigma(\boldsymbol{\theta})$ the vector of their model-implied counterparts & $W = I$ the weight matrix

Implementation: Mplus software & R package lavaan

Pairwise Maximum Likelihood

Maximization of pairwise likelihood was proposed in the context of ordinal variables [2] as an alternative estimation method. PL is defined as the sum of all the bivariate log-likelihoods. For *one* observation the bivariate log-likelihoods of the ***pairs*** are of the form:

Both continuous: Bivariate normal density

Both ordinal: $\log L(y_i, y_k) = \sum_{c_i=1}^{m_i} \sum_{c_k=1}^{m_k} I(y_i{=}c_i, y_k{=}c_k) \log \pi_{c_i c_k}^{(y_i y_k)}$

where $\pi_{c_i c_k}^{(y_i y_k)} = P(y_i{=}c_i, y_k{=}c_k) =$

$$\int_{\tau_{c_{i-1}}^{(y_i)}}^{\tau_{c_i}^{(y_i)}} \int_{\tau_{c_{k-1}}^{(y_k)}}^{\tau_{c_k}^{(y_k)}} f_2(y_i^*, y_k^*) dy_i^* dy_k^*$$

& f_2 bivariate normal density

Ordinal & continuous: $\log L(y_i, w_j) = \sum_{c_i=1}^{m_i} I(y_i{=}c_i | w_j) \log P(y_i{=}c_i | w_j) + \log g_1(w_j)$, where

$$P(y_i{=}c_i | w_j) = \int_{\tau_{c_{i-1}}^{(y_i)}}^{\tau_{c_i}^{(y_i)}} f_1(y_i^* | w_j) dy_i^*$$

& g_1,f_1 univariate normal densities

Implementation PML is available for mixed items in R package lavaan

Simulation

Evaluation of ULS & PML: One- & three-factor orthogonal models, with continuous, ordinal (5 categories) & binary items. Sample size n=1000, $\boldsymbol{\Lambda}$ generated from Uniform $(0.4, 0.9)$, error variances of continuous items 1.0, B=100 replications.

Graphical representation of the simulated three-factor model with 30 items

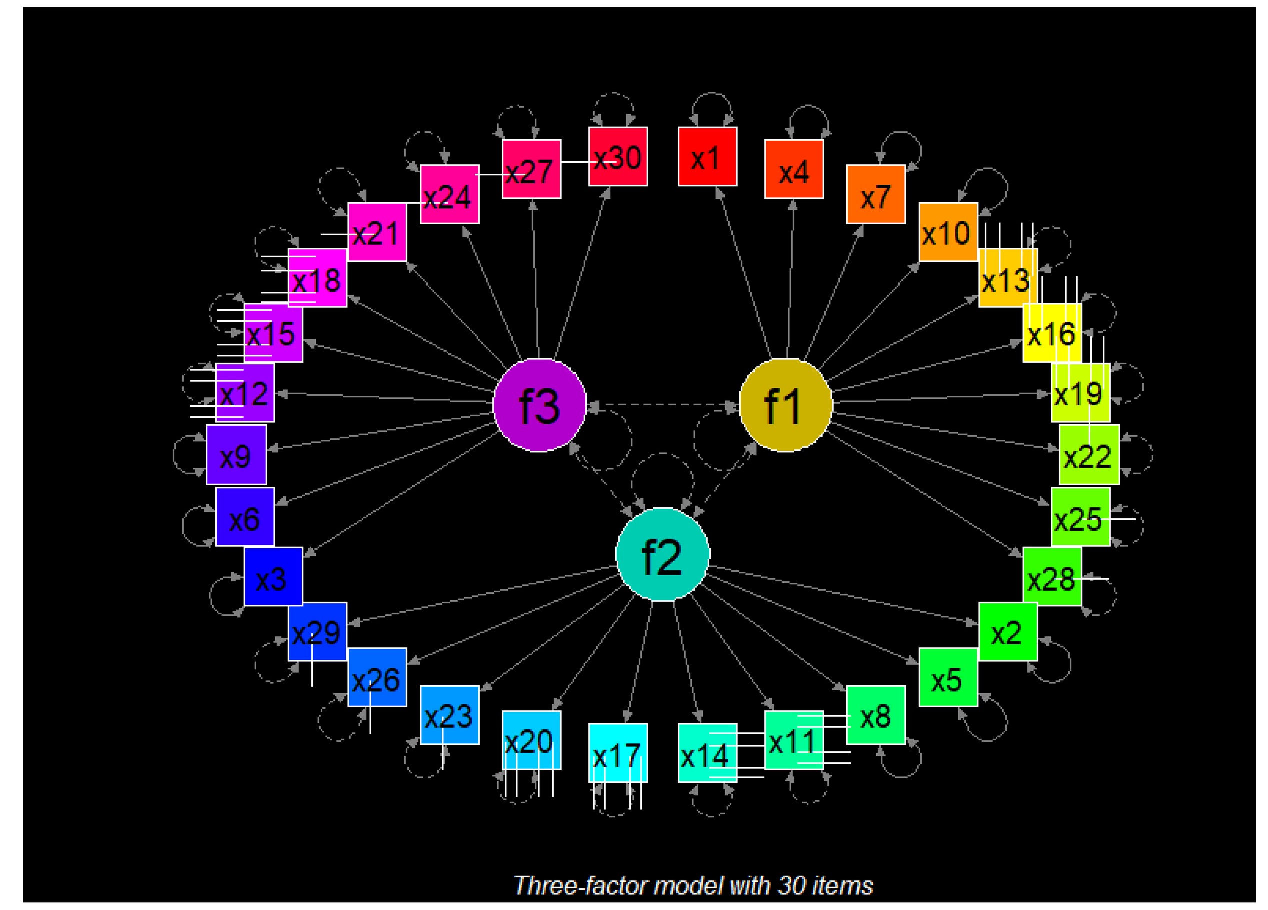

Fig. 1: Three-factor model with 30 items

Average absolute relative bias of loadings across items (%)

Model	ULS	PML
One factor & p=9	0.412	0.401
One factor & p=15	0.637	0.649
Three factors & p=21	0.482	0.442
Three factors & p=30	0.584	0.592

Table. 1: Average absolute relative bias of loadings across items (%)

References

[1] K. G. Jöreskog. On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, 59(3):381–389, 1994.

[2] M. Katsikatsou, I. Moustaki, F. Yang-Wallentin, and K. G. Jöreskog. Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*, 56(12):4243–4258, 2012.