Constrained versus unconstrained exploratory approach to Bayesian Item Factor Analysis*

S. Fontanella[1], L. Fontanella[2], P. Valentini[3], N. Trendafilov[1]

[1] *Department of Mathematics and Statistics, The Open University- Milton Keynes, UK*
[2] *Department of Legal and Social Sciences, University of Chieti-Pescara - Pescara, Italy*
[3] *Department of Economics, University of Chieti-Pescara - Pescara, Italy*

1 Introduction

Within the social and behavioural sciences, item-level data are often categorical in nature. Dichotomous or polytomous item response formats may fail to maintain the scale and distributional properties assumed by models such as common linear factor analysis. In such a case, item factor analysis (IFA) offers an appropriate alternative. An item factor analytic model is often viewed as consisting of two component models. First, a classical factor model represents some continuous latent response variables as linear functions of the multiple latent abilities. Then, as second component, a threshold model represents the nonlinear relationship between the latent response variable and the probability of a given categorical response to an item (Wirth and Edwards, 2007). Various IFA models can be found in both the Structural Equation Models (SEM) and Item Response Theory (IRT) literatures. In particular, in the latter context, great attention has been recently paid to the multidimensional formulation of IRT models, which account for distinct underlying latent traits, involved in producing the manifest responses to the selected items. IRT can be applied in a way that is analogous to exploratory and confirmatory factor analysis for continuous variables (Reckase, 1997). In an exploratory perspective, identication problems has to be considered. This would require fixing a minimal number of constraints for the model parameters. In particular, to solve the rotational indeterminacy, a commom approach is to impose a positive lower triangular (PLT) structure on the factor loadings matrix. In Bayesian EFA, this is achieved with the definition of appropriate prior distributions for the factor loadings (Geweke and Zhou, 1996; Lopes and West, 2004). The PLT approach guarantees a unique global mode of the likelihood underlying the posterior distribution. It does not, however, preclude the existence of local modes. The constraints influence the shape of the likelihood and thus the shape of the posterior distribution. This is problematic since local modes can negatively affect the convergence behavior of Markov Chain Monte Carlo (MCMC) sampling schemes used for estimation purposes. As the constraints are imposed on particular elements of the loadings matrix, inference results may depend on the ordering of the variables (Aßmann et al., 2016).

Since the PLT results in the *ordering problem* (Lopes and West, 2004), Aßmann et al. (2016) argue that an adequate solution would be to design a sampler which works without such constraints. To this end, the authors developed an unconstrained Gibbs sampler for EFA where the rotational invariance is addressed in a post-processing procedure based on the Weighted Orthogonal Procrustes (WOP) approach. Here, we consider only the Orthogonal Procrustes (OP) problem, but the weighted version can be easily adapted (for more details see Pape (2015)).

In this paper, following a Bayesian approach, we propose two alternative exploratory MIRT models based on the PLT constraints and the unconstrained Gibbs sampler with OP post-processing procedur, respectively.

2 The multidimensional normal ogive model

Given a test consisting of K ordered categorical and assuming M latent traits, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)'$, a factorial measurement model can be used to represent some continuous underlying response variables, $\mathbf{Z} = (Z_1, \ldots, Z_K)'$, as linear functions of the multiple latent traits

$$\mathbf{Z} = \boldsymbol{A}\boldsymbol{\Theta} + \boldsymbol{\epsilon}. \tag{1}$$

The factorial structure of the model is represented by the $(K \times M)$ matrix $\boldsymbol{A}$ containing the discrimination parameters, or factor loadings, $\{\alpha_{k,m}\}$, and $\boldsymbol{\epsilon}$ is a vector of disturbances.
In addition, each categorical response X_{ik}, for subject $i = 1, \ldots, n$ and item $k = 1, \ldots, K$, is linked to the latent continuous response Z_{ik} through the following threshold model

$$X_{ik} = c \quad if\ \gamma_{k,c-1} \le Z_{ik} \le \gamma_{k,c},\ c = 1, \ldots, C;\ \gamma_{k,0} = -\infty, \gamma_{k,C} = \infty \tag{2}$$

*This work is supported by a grant RPG-2013-211 from The Leverhulme Trust, UK.

where C it the number of response categories and $\boldsymbol{\gamma}_k = (\gamma_{k,1} \dots \gamma_{k,C-1})'$ are the threshold parameters.
In the IRT literature, different unidimensional and multidimensional IFA models have been proposed. In our study, we adopt the two-parameter normal ogive (2PNO) formulation which derives from the assumption of normally distributed error, i.e. $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
If we consider polytomous items, the probability of responding a certain category c to a given item k can be expressed as (Béguin and Glas, 2001)

$$P(X_{ik} = c|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k) = \Phi\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i - \gamma_{k,c-1}\right) - \Phi\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i - \gamma_{k,c}\right) \tag{3}$$

where X_{ik} is the observed response of person i ($i = 1, \dots, N$) to item k ($k = 1, \dots, K$); c denotes the category of the ordered response scale ($c = 1, ..., C$) and Φ is the cumulative distribution function for the standard normal distribution. The probability of responding a certain category c, to a given item k, depends on the M-dimensional vector $\boldsymbol{\theta}_i$ of the unobserved latent scores for subject i, on the M-dimensional vector $\boldsymbol{\alpha}_k$ of item discrimination parameters and on the $(C-1)$-dimensional vector $\boldsymbol{\gamma}_k$ of category thresholds.

In a confirmatory framework, considering a multi-unidimensional IRT model, we have that each item loads only onto a construct. The assumption that each of the M correlated constructs is being measured by its own set Ω_m containing K_m items, can be written as $\alpha_{k,m} \neq 0$ if $k \in \Omega_m$, $\alpha_{k,m} = 0$ if $k \notin \Omega_m$ and the matrix of discrimination parameter has a block structure.

In an exploratory framework, since each item might potentially load onto all the M constructs, the discrimination parameter matrix has not a block structure and identifiability issues have to be considered.

3 Identification issues

In the Bayesian approach to MIRT models, identification restrictions required by the over-parameterisation can be imposed by choosing appropriate prior hyperparameters.

For location and scale indeterminacy, we consider the constraint $\sum_{k=1}^{K} \gamma_{k,c} = 0$ for a given category c and assume that each latent component has unit variance.

With regards to the prior specification, on the person side of the model, we assume that all person parameters $\boldsymbol{\theta}_i$ are independent and identically distributed samples from a multivariate normal distribution, that is $\boldsymbol{\theta}_i \sim \mathcal{N}_M(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ (Béguin and Glas, 2001). The prior for $\boldsymbol{\mu}_\theta$ is normal with mean $\boldsymbol{\mu}_0 = \mathbf{0}$ and covariance matrix $\sigma_\mu^2\mathbf{I}$, where $\sigma_\mu^2 = 100$. The prior for the inverse covariance matrix $\boldsymbol{\Sigma}_\theta$ is a Wishart with scale matrix $0.1\mathbf{I}$ and degree of freedom $M+1$. On the item side, a uniform prior is assigned to the ordered thresholds $\gamma_{k,c} \sim uniform,\ c = 1, \dots, C-1,\ \gamma_{k,1} \leq \dots \leq \gamma_{k,C-1},\ \forall k$.

The remaining issue to address is the joint rotational indeterminacy of the vector of item discriminations, $\boldsymbol{\alpha}_k$, and the vector of abilities, $\boldsymbol{\theta}_i$. Indeed for any orthogonal matrix $\boldsymbol{D}$, we can transform the model parameter as $\boldsymbol{\alpha}_k^* = \boldsymbol{D}\boldsymbol{\alpha}_k$ and $\boldsymbol{\theta}_i^* = \boldsymbol{D}\boldsymbol{\theta}_i$ so that the probability of a certain response given ability values is not altered. Thus, an infinite number of pairwise parameter estimates for $\{\boldsymbol{\alpha}_k^*, \boldsymbol{\theta}_i^*\}$ retain the same probability $P(X_{ik} = c|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k)$.

In this paper we addressed the rotational indeterminacy issue in the Sections 3.1 and 3.2 by considering the PLT constraints and the unconstrained Gibbs sampler with OP post-processing procedure, respectively.

3.1 The constrained MIRT model

In the constrained Gibbs sampler for MIRT models, we incorporate the PLT constraint by appropriately choosing fully informative prior distributions for the discrimination parameters.

Here, for each free $\alpha_{k,m}$, we assume independent priors such that $\alpha_{k,m} \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$ when $k \neq m$ and $\alpha_{k,m} \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)I(\alpha_{k,m} > 0)$ if $k = m$. Here, $I(\cdot)$ represents the indicator function and μ_α and σ_α^2 are fixed at 1 and 100, respectively. The upper-diagonal elements are then set to 0 using a Dirac delta distribution $\delta_0(\alpha_{k,m})$ for $k \neq m$ and $k \leq M$.

In order to draw samples for the posterior distribution of the parameters, it is convenient to use data augmentation technique. Coherently with the IFA specification in equation (1), we assume that a continuous variable Z_k underlies the observed ordinal measure X_k, and that there is a linear relationship between item and person parameters and the underlying variable such that $Z_{i,k} = \boldsymbol{\alpha}_k'\boldsymbol{\theta}_i + \epsilon_{i,k}$, with $\epsilon_{i,k} \sim \mathcal{N}(0, 1)$. The relation between the observed items and the underlying variables is given by the threshold model in equation (2).

The full conditional of most parameters can be specified in closed form, which allows for a Gibbs sampler although, for the polytomous model, a Metropolis-Hastings step is required to sample the ordered threshold parameters. Details on the MCM scheme are given in the Appendix.

3.2 The Unconstrained MIRT model with OP post-processing procedure

The PLT constrain has been extensively used in Bayesian EFA literature to solve the rotational indeterminacy. However, as pointed out by Lopes and West (2004), this configuration gives rise to the ordering problem. Indeed, for a $K \times K$ permutation matrix $\boldsymbol{O}$ that is premultiplied to the IFA model in equation (1), we have

$$\boldsymbol{O}\mathbf{Z} = \boldsymbol{O}\boldsymbol{A}\boldsymbol{\Theta} + \boldsymbol{O}\boldsymbol{\epsilon}$$

This equation implies that the reordered data have the same latent traits, with new parameters $\boldsymbol{A}^* = \boldsymbol{O}\boldsymbol{A}$. It is clear then that the choice to constrain $\boldsymbol{A}^*$ instead of $\boldsymbol{A}$ may have a substantial impact on the shape of the posterior distribution and the estimates derived therefrom (Aßmann et al., 2016).

To circumvent this issue, Aßmann et al. (2016) propose an ex-post approach to solve rotation indeterminacy in EFA. In particular, starting from the theory of mixture models, the authors show that the ordering problem strongly resembles the well-known *label switching* issue: finite mixture models are typically not identified, as labels of the mixture components can be changed by permutation. The suggested approach does not constrain the parameter space a priori, but fixes the rotation problem ex-post via minimizing posterior expected loss. To achieve the minimization in the MCMC context, the output of the unconstrained Gibbs sampler is transformed using a sequence of orthogonal matrices obtained solving a orthogonal Procrustes problem (Gower and Dijksterhuis, 2004). We adapt this method to the MIRT framework.

When no constraints are imposed on the parameter models, the MCMC sampler is said to be orthogonally mixed(Aßmann et al., 2016), meaning that at each iteration of the sampler, the sample space of the discrimination matrix $\boldsymbol{A}$ may be subject to an orthogonal transformation, expressed by the matrix $\boldsymbol{D}$. In order to solve this issue, the OP algorithm is applied (Pape, 2015). The aim is that to solve the following optimization problem for a sequence of R draws from an orthogonally mixing posterior distribution of $\boldsymbol{A}$, which is denoted as $\{\boldsymbol{A}^r\}_{r=1}^R$:

$$\left\{\{\boldsymbol{D}^r\}_{r=1}^R, \boldsymbol{A}^*\right\} = \arg\min \sum_{r=1}^{R} L_D\left(\boldsymbol{A}^*, \boldsymbol{A}^{(r)}\boldsymbol{D}^{(r)}\right) \quad s.t. \;\; \boldsymbol{D}^{(r)'}\boldsymbol{D}^{(r)} = \boldsymbol{I}, \;\; r = 1, \ldots, R. \tag{4}$$

L_D denotes the quadratic loss function

$$L_D\left(\boldsymbol{A}^*, \boldsymbol{A}^{(r)}\boldsymbol{D}^{(r)}\right) = tr\left[(\boldsymbol{A}^{(r)}\boldsymbol{D}^{(r)} - \boldsymbol{A}^*)'(\boldsymbol{A}^{(r)}\boldsymbol{D}^{(r)} - \boldsymbol{A}^*)\right]$$

This optimization problem can be easily solved by the OP algorithm that prescribes a 2-step procedure:

1. For a given $\boldsymbol{A}^*$ minimize equation (4) for each s$= 1, \ldots, S$. This is solved by:
   - Define $\boldsymbol{S}_r = \boldsymbol{A}^{(r)'}\boldsymbol{A}^*$.
   - Compute the SVD of $\boldsymbol{S}_r = \boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r$
   - Obtain the orthogonal transformation matrix $\boldsymbol{D}^{(r)} = \boldsymbol{U}_r\boldsymbol{V}_r'$.
2. Choose $\boldsymbol{A}^*$ as

$$\boldsymbol{A}^* = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{A}^{(r)}\boldsymbol{D}^{(r)}$$

The algorithm converges when

$$||\boldsymbol{A}^{*(r)} - \boldsymbol{A}^{*(r-1)}||^2 < \omega$$

with ω set close to 0. Note that the algorith requires to set an initial value for $\boldsymbol{A}^*$. For convenience, Pape (2015) suggests to use the last draw of the unconstrained MCMC algorithm.

Appendix : MCMC for the constrained MIRT model

(1) Sample $Z_{i,k}|\boldsymbol{\alpha}_k, \boldsymbol{\theta}_i, \boldsymbol{\gamma}_k, \mathbf{X}$

The underlying variable scores are drawn from doubly truncated normal distributions implied by the introduction of the thresholds.

$$Z_{i,k}|\boldsymbol{\alpha}_k, \boldsymbol{\theta}_i, \boldsymbol{\gamma}_k, \mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i, 1\right) I(\gamma_{k,c-1} < Z_{i,k} < \gamma_{k,c-1}) \quad if \;\; X_{i,k} = c$$

(2) Sample $\boldsymbol{\theta}_i|\boldsymbol{Z}, \{\boldsymbol{\alpha}_k\}, \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta$

Considering the prior $\boldsymbol{\theta}_i \sim \mathcal{N}_M(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$, the full conditional is normal

$$\boldsymbol{\theta}_i|\boldsymbol{Z}, \{\boldsymbol{\alpha}_k\}, \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta, \sim \mathcal{N}_M\left((\mathbf{A}'\mathbf{A} + \boldsymbol{\Sigma}_\theta^{-1})^{-1}(\mathbf{A}'\boldsymbol{Z}_i + \boldsymbol{\Sigma}_\theta^{-1}\boldsymbol{\mu}_i), (\mathbf{A}'\mathbf{A} + \boldsymbol{\Sigma}_\theta^{-1})^{-1}\right)$$

(3) Sample $\boldsymbol{\alpha}_k|\boldsymbol{Z}, \{\boldsymbol{\theta}_i\}, \mu_\alpha, \sigma_\alpha^2$

Due to the PLT constraint we have full conditionals as follows:

- for $k = 1, \ldots, M$

$$\boldsymbol{\alpha}_k|\boldsymbol{Z}, \{\boldsymbol{\theta}_i\}, \mu_\alpha, \sigma_\alpha^2 \sim \mathcal{N}_M \left((\boldsymbol{\Theta}_k'\boldsymbol{\Theta}_k + \sigma_\alpha^{-2}\boldsymbol{I}_k)^{-1}(\sigma_\alpha^{-2}\mu_\alpha \mathbf{1}_k + \boldsymbol{\Theta}_k'\boldsymbol{z}_k), (\boldsymbol{\Theta}_k'\boldsymbol{\Theta}_k + \sigma_\alpha^{-2}\boldsymbol{I}_k)^{-1} \right)$$

  where $\boldsymbol{\Theta}_k$ is the $N \times k$ matrix containing the first k columns of $\boldsymbol{\Theta}$ and $\boldsymbol{z}_k$ is the k-th column of $\boldsymbol{Z}$.

- for $k = M + 1, \ldots, K$

$$\boldsymbol{\alpha}_k|\boldsymbol{Z}, \{\boldsymbol{\theta}_i\}, \mu_\alpha, \sigma_\alpha^2 \sim \mathcal{N}_M \left((\boldsymbol{\Theta}'\boldsymbol{\Theta} + \sigma_\alpha^{-2}\boldsymbol{I}_K)^{-1}(\sigma_\alpha^{-2}\mu_\alpha \mathbf{1}_K + \boldsymbol{\Theta}'\boldsymbol{z}_k), (\boldsymbol{\Theta}'\boldsymbol{\Theta} + \sigma_\alpha^{-2}\boldsymbol{I}_K)^{-1} \right).$$

(4) Sample $\boldsymbol{\gamma}_k|\{\boldsymbol{\theta}_i\}, \{\boldsymbol{\alpha}_k\}$
To draw the threshold parameters for a given item, we consider a Metropolis-Hastings step based on Cowles' algorithm (Cowles, 1996), and draw the candidate from

$$\gamma_{k,c}^* \sim \mathcal{N}(\gamma_{k,c}, \sigma_{MH}^2) I(\gamma_{k,c-1}^* < \gamma_{k,c} < \gamma_{k,c+1}) \; for \; c = 1, \ldots, C-1$$

The Metropolis-Hastings acceptance probability is then given by

$$min \left[\prod_{i=1}^{N} \frac{Pr(X_{i,k} = c|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k^*) f(\boldsymbol{\gamma}_k|\boldsymbol{\gamma}_k^*, \sigma_{MH}^2)}{Pr(X_{i,k} = c|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k^*|\boldsymbol{\gamma}_k, \sigma_{MH}^2)}, 1 \right]$$

(5) Sample $\boldsymbol{\mu}_\theta|\{\boldsymbol{\theta}_i\}, \boldsymbol{\Sigma}_\theta$
Given the prior distribution for the M-dimensional vector $\boldsymbol{\mu}_\theta$, $\mathcal{N}_M(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_{\theta,0})$, and considering the $(N \times M)$ matrix $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_N]'$, the posterior is normal

$$\boldsymbol{\mu}_\theta|\{\boldsymbol{\theta}_i\}, \boldsymbol{\Sigma}_\theta \sim \mathcal{N}_M \left((N\boldsymbol{\Sigma}_\theta^{-1} + \boldsymbol{\Sigma}_{\theta,0}^{-1})^{-1}(\boldsymbol{\Sigma}_\theta^{-1} \sum_{i=1}^{N} \boldsymbol{\theta}_i + \boldsymbol{\Sigma}_{\theta,0}^{-1}\boldsymbol{\mu}_0), (N\boldsymbol{\Sigma}_\theta^{-1} + \boldsymbol{\Sigma}_{\theta,0}^{-1})^{-1} \right)$$

We consider $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\theta,0} = 100\mathbf{I}_M$
(6) Sample $\boldsymbol{\Sigma}_\theta|\{\boldsymbol{\theta}_i\}, \boldsymbol{\mu}_\theta, \boldsymbol{S}_0, N_0$
The prior is $\boldsymbol{\Sigma}^{-1} \sim Wish(N_0, \boldsymbol{S}_0)$ con $N_0 = M + 1$ and $\boldsymbol{S}_0 = 0.1\boldsymbol{I}$. Therefore, the posterior is

$$\boldsymbol{\Sigma} \sim Inv - Wish \left(N_0 + N, \boldsymbol{S}_0 + \sum_{i=1}^{N} (\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta)(\boldsymbol{\theta}_i - \boldsymbol{\mu}_\theta)' \right)$$

.

References

C. Aßmann, J. Boysen-Hogrefe, and M. Pape. Bayesian analysis of static and dynamic factor models: An ex-post approach towards the rotation problem. *Journal of Econometrics*, 192(1):190–206, 2016.

A. A. Béguin and C. A. W. Glas. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66(4):541–561, 2001.

M.K. Cowles. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Stat. Comput.*, 6(2):101–111, 1996.

J. Geweke and G. Zhou. Measuring the pricing error of the arbitrage pricing theory. *Rev. Financ. Stud*, 9(2): 557–587, 1996.

J. C. Gower and G. B. Dijksterhuis. *Procrustes problems.* Oxford Statistical Science Series. Oxford University Press, Oxford, UK, 2004.

H. F. Lopes and M. West. Bayesian model assessment in factor analysis. *Statistica Sinica*, 14:41–67, 2004.

M. Pape. *Model Identification in Bayesian Analysis of Static and Dynamic Factor Models.* PhD thesis, Christian-Albrechts-Universität zu Kiel, 2015.

M. D. Reckase. The Past and Future of Multidimensional Item Response Theory. *Appl. Psych. Meas.*, 21(1): 25–36, 1997.

R. J. Wirth and M. C. Edwards. Item factor analysis: Current approaches and future directions. *Psychological methods*, 12(1):58–79, 2007.