



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Applied Psychological Measurement (2014) 38.7: 549-562

DOI: <http://dx.doi.org/10.1177/0146621614536272>

Copyright: © 2014 by Sage Publications

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

Multidimensional Item Response Model for Nominal Variables

Javier Revuelta

Autonoma University of Madrid

Correspondence should be sent to:

Javier Revuelta

Department of Social Psychology and Methodology

Autonoma University of Madrid

28049 Madrid

SPAIN

javier.revuelta@uam.es

Published in: Revuelta, J. (2014). Multidimensional ítem response model for nominal variables. *Applied Psychological Measurement*, 38(7), 549–562.

Abstract

This paper describes a multidimensional generalization of the nominal categories model that serves to estimate factorial models from nominal and ordinal observed responses, and includes a structural model for latent variables that distinguishes between endogenous and exogenous factors. The model includes a scale parameter for each response category in each factor. Item parameters relate the logit between categories to the vector of latent variables. The inferential framework is marginal maximum likelihood, implemented via static and adaptive Gauss–Hermite quadrature and Monte Carlo EM. The properties of estimators are investigated in a simulation study. An example with real data illustrates the utility of the model in analyzing local dependencies in testlets composed of multiple-choice items that are clustered in several groups around a common theme.

Keywords: multidimensional nominal categories model; item factor analysis; marginal maximum likelihood; adaptive GH quadrature; local independence; multidimensional item response theory.

1. Introduction

Multidimensional item response models have provided a means to analyze the dimensionality of a set of items since the pioneering work by Bock, Gibbons, and Muraki (1988). The item factor model typically applies to dichotomous and ordinal data (Reckase, 2009), whereas the factorial analysis of nominal responses is more recent (Bock & Gibbons, 2010). For example, Thissen, Cai, and Bock (2010) proposed a flexible multidimensional model that can accommodate a variety of models between ordinal and nominal scales. This paper continues this line of research and presents a generalization of the nominal categories model (Bock, 1972) to the multidimensional case. Moreover, a structural model can be imposed on the latent variables to distinguish between exogenous and endogenous factors (Jöreskog & Sörbom, 1979).

The nominal categories model may be regarded as the most popular model for nominal responses because of its unique combination of flexibility and mathematical simplicity. For these reasons, it is taken as the basis to develop a multidimensional nominal categories model (MNCM). The MNCM, in its full generality, imposes no constraints on item parameters. However, the paper describes a method of linear constraints that provides particular cases suitable for the analysis of ordinal data or for testing a hypothesis about parameter structure (Revuelta, 2009, 2010).

The MNCM has a factor loading for each category in each factor, instead of having one parameter for each variable in each factor, as in the factor analysis of quantitative and ordinal data. There are a number of new applications in which the factor loadings may differ between categories. For example, consider one multiple-choice item measuring mathematics in which one of the categories has an involved wording. This distractor may

have a higher loading on a verbal factor than other categories written in more simple terms. More generally, two-dimensional item response models have been proposed in the psychometric literature to distinguish between a factor of substantive interest and a factor that explains individual differences in solution strategy or response style. For example, Bolt and Johnson (2009) and Johnson and Bolt (2010) analyzed ordinal responses from a self-report of smoking dependence. Their second factor explains response tendencies that are unrelated to the construct measured by the items. A similar idea was translated to maximum performance testing by Bolt and Wollack (2012), who analyzed data from multiple-choice items in which there was a primary factor related to the correct option and a second factor that influences selection among distractors.

The present paper includes an example with real data that illustrates the application of the MNCM to investigating local dependencies originated by testlets. In this example, several items are clustered around a common theme, which leads to violation of the local independence assumption in a one-dimensional model. The paper shows how a polytomous multidimensional model can be used to analyze associations between categories from one item to another.

The paper also includes a number of technical developments, such as a marginal likelihood/EM estimation algorithm (Schilling & Bock, 2005), the computation of the information matrix, a structural latent model, and a Taylor series expansion to approximate the correlation between response categories. Finally, the recovery of theoretical parameters is investigated in a simulation study. Most technical material has been moved to appendices and the main body of the paper focuses on the description of the model, results of its application and the conclusions.

2. Multidimensional nominal categories model

Suppose that each individual is described by a vector of D factors, referred to as $\boldsymbol{\theta}$.

According to the MNMCM, the probability of category r of item i is given by:

$$P_{ir}(\boldsymbol{\theta}) = \frac{\exp(z_{ir})}{\sum_{k=1}^{K_i} \exp(z_{ik})}, \quad (1)$$

where K_i is the number of categories, and $z_{ir} = \tau_{ir} + \boldsymbol{\lambda}_{ir}' \boldsymbol{\theta} = \tau_{ir} + \lambda_{ir1} \theta_1 + \cdots + \lambda_{iK_i D} \theta_D$ is the latent propensity of category r , where the prime denotes the vector transpose, τ_{ir} is a scalar intercept parameter, and $\boldsymbol{\lambda}_{ir}$ is a vector of scale parameters that includes a parameter for each factor. In matrix form, the vector of latent propensities for item i is:

$$\mathbf{z}_i = \boldsymbol{\tau}_i + \boldsymbol{\Lambda}_i \boldsymbol{\theta}, \quad (2)$$

where $\boldsymbol{\tau}_i$ is a column vector of K_i elements and $\boldsymbol{\Lambda}_i$ is a matrix of $K_i \times D$ scalar parameters. The parameters of the MNMCM describe the logit between pairs of alternatives. The logit between categories r and K_i is given by

$$\log \frac{P_{ir}(\boldsymbol{\theta})}{P_{iK_i}(\boldsymbol{\theta})} = (\tau_{ir} - \tau_{iK_i}) + (\lambda_{ir1} - \lambda_{iK_i1}) \theta_1 + \cdots + (\lambda_{irD} - \lambda_{iK_iD}) \theta_D. \quad (3)$$

The parameters of equation (3) are unidentifiable because the logit depends on differences between parameters and does not vary if a constant value is added to all the parameters. The identifiability problem is more easily solved when equation (3) is written in matrix form (Revuelta, 2008). Let \mathbf{P}_i be the vector of response probabilities for item i and \mathbf{S}_i be an identity matrix supplemented with a column vector of -1 ; that is, $\mathbf{S}_i = (\mathbf{I}; -\mathbf{1})$ and the size of \mathbf{S}_i is $(K_i - 1) \times K_i$. The following equation describes in matrix form the logit of categories 1 to $K_i - 1$ against category K_i :

$$\mathbf{S}_i \log \mathbf{P}_i = \mathbf{S}_i (\boldsymbol{\tau}_i + \boldsymbol{\Lambda}_i \boldsymbol{\theta}) = \mathbf{S}_i (\boldsymbol{\tau}_i + \boldsymbol{\lambda}_i^{(1)} \theta_1 + \cdots + \boldsymbol{\lambda}_i^{(D)} \theta_D), \quad (4)$$

where $\lambda_i^{(d)}$ is the d column of Λ_i . Parameter indeterminacy is removed by computing τ_i and Λ_i from a smaller number of parameters using the linear equation:

$$\tau_i + \lambda_i^{(1)}\theta_1 + \dots + \lambda_i^{(D)}\theta_D = \mathbf{V}_i\boldsymbol{\iota} + \mathbf{W}_{1i}\boldsymbol{\varsigma}^{(1)}\theta_1 + \dots + \mathbf{W}_{Di}\boldsymbol{\varsigma}^{(D)}\theta_D, \quad (5)$$

where $\boldsymbol{\iota}$, $\boldsymbol{\varsigma}^{(1)}$, ..., $\boldsymbol{\varsigma}^{(D)}$ are vectors of parameters, and \mathbf{V}_i , \mathbf{W}_{1i} , ..., \mathbf{W}_{Di} are matrices of fixed coefficients that impose a constraint on the parameters of equation (2). The intercepts and the columns of Λ_i are given by $\tau_i = \mathbf{V}_i\boldsymbol{\iota}$, and $\lambda_i^{(d)} = \mathbf{W}_{di}\boldsymbol{\varsigma}^{(d)}$. The matrices of coefficients have K_i rows and their number of columns must be smaller than K_i , although the exact number of columns depends on the number of parameters in vectors $\boldsymbol{\iota}$, $\boldsymbol{\varsigma}^{(1)}$, ..., $\boldsymbol{\varsigma}^{(D)}$. Item parameters are identifiable when $\mathbf{S}_i\mathbf{V}_i$ and $\mathbf{S}_i\mathbf{W}_{di}$ (for $d = 1, \dots, D$) have full column rank (Rothenberg, 1971).

Two particular cases of equation (5) are considered in this paper. First, the same constraints may apply to the scale parameters of all dimensions. In that case, equation (5) reduces to

$$\tau_i + \lambda_i^{(1)}\theta_1 + \dots + \lambda_i^{(D)}\theta_D = \mathbf{V}_i\boldsymbol{\iota} + \mathbf{W}_i(\boldsymbol{\varsigma}^{(1)}\theta_1 + \dots + \boldsymbol{\varsigma}^{(D)}\theta_D) = \mathbf{V}_i\boldsymbol{\iota} + \mathbf{W}_i\boldsymbol{\varsigma}\boldsymbol{\theta}, \quad (6)$$

where $\boldsymbol{\varsigma}$ is the matrix with column vectors $\boldsymbol{\varsigma}^{(1)}$, ..., $\boldsymbol{\varsigma}^{(D)}$; thus $\tau_i + \Lambda_i\boldsymbol{\theta} = \mathbf{V}_i\boldsymbol{\iota} + \mathbf{W}_i\boldsymbol{\varsigma}\boldsymbol{\theta}$.

The second particular case imposes the same constraint to all intercept and scale parameters. In that case, equation (5) becomes

$$\tau_i + \Lambda_i\boldsymbol{\theta} = \mathbf{W}_i(\boldsymbol{\iota} + \boldsymbol{\varsigma}\boldsymbol{\theta}). \quad (7)$$

In the context of linearly constrained item response models, identifiability is traditionally assessed by verifying that the augmented matrix $\mathbf{W}_i^+ = (\mathbf{W}_i; \mathbf{1})$ has full column rank (Fischer, 1983). The identifiability conditions based on \mathbf{W}_i^* and \mathbf{W}_i^+ can be shown to be equivalent (see Appendix A).

3. Examples of linear constraints

3.1. Nominal models

Models for nominal responses, in their most general form, impose no constraints apart from those that ensure identifiability. This section shows one example of the many forms that matrix \mathbf{W}_i may take. Multiple-choice items have a correct category, say, category K_i , whose parameters τ_{iK_i} and λ_{iK_i} are set to 0. The incorrect categories $1, \dots, K_i - 1$ are unconstrained and describe the logit against the correct category:

$$\log \frac{P_{ir}(\boldsymbol{\theta})}{P_{iK_i}(\boldsymbol{\theta})} = \tau_{ir} + \lambda_{ir1}\theta_1 + \dots + \lambda_{irD}\theta_D. \quad (8)$$

Equation (7) for an item with three categories and two factors reads as follows:

$$\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \tau_{i3} \end{pmatrix} + \begin{pmatrix} \lambda_{i11} & \lambda_{i12} \\ \lambda_{i21} & \lambda_{i22} \\ \lambda_{i31} & \lambda_{i32} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \left(\begin{pmatrix} \iota_{i1} \\ \iota_{i2} \end{pmatrix} + \begin{pmatrix} \varsigma_{i11} & \varsigma_{i12} \\ \varsigma_{i21} & \varsigma_{i22} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \right). \quad (9)$$

In this example, \mathbf{W}_i satisfies the identifiability condition because the matrix

$$\mathbf{W}_i^* = \mathbf{S}_i \mathbf{W}_i = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (10)$$

has full column rank. The linear restriction reduces the number of effective parameters from the nine elements contained in $\boldsymbol{\tau}$ and $\boldsymbol{\Lambda}$ to the six elements of $\boldsymbol{\iota}$ and $\boldsymbol{\varsigma}$. One example in which constraints are different for each dimension is as follows. Suppose that equation (5) reads:

$$\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \tau_{i3} \end{pmatrix} + \begin{pmatrix} \lambda_{i11} & \lambda_{i12} \\ \lambda_{i21} & \lambda_{i22} \\ \lambda_{i31} & \lambda_{i32} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \iota_{i1} \\ \iota_{i2} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \varsigma_{i11} \\ \varsigma_{i21} \end{pmatrix} \theta_1 + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \varsigma_{i12} \\ \varsigma_{i22} \end{pmatrix} \theta_2. \quad (11)$$

The matrices \mathbf{V}_i and \mathbf{W}_{1i} in equation (11) are the same as \mathbf{W}_i in equation (10), and thus identifiability is ensured. Regarding \mathbf{W}_{2i} , the scale parameters in the second dimension

are identifiable because the matrix

$$\mathbf{S}_i \mathbf{W}_{2i} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (12)$$

has full column rank.

3.2. Ordinal models

Ordinal model use different constraints for intercept and scale parameters, as in equation (6). The category scale parameters are proportional to an item scale parameter, $\lambda_{ird} = (r-1)\varsigma_{id}$, and \mathbf{W}_i is a column vector with integer values between 0 and $K_i - 1$.

For an item with three categories and two factors, equation (6) reads:

$$\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \tau_{i3} \end{pmatrix} + \begin{pmatrix} \lambda_{i11} & \lambda_{i12} \\ \lambda_{i21} & \lambda_{i22} \\ \lambda_{i31} & \lambda_{i32} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} t_{i1} \\ t_{i2} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} (\varsigma_{i1} \quad \varsigma_{i2}) \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}. \quad (13)$$

The matrix \mathbf{W}_i has a smaller rank for an ordinal than for a nominal model because the ordinal model does not have separate scale parameters for each category.

Thissen, Cai, and Bock (2010) developed a multidimensional model, the TCB, that is based on a nonlinear constraint on scale parameters and uses a Fourier basis to specify the matrix of coefficients. The TCB is more flexible than the model in equation (13) and provides useful and interesting particular cases for the analysis of ordinal data by relaxing the assumption that the response categories are equally spaced. However, the TCB does not reach the generality of the full rank model in equation (9) because of the effect of the nonlinear constraint.

4. Inferential algorithms

The inferential algorithm used in this paper this section assumes that θ is a random variable whose density function, $f(\theta)$, is MNormal $(\mathbf{0}, \Sigma)$. Moreover, the measurement model in equations (1) and (2) is supplemented with a structural model to analyze relations of dependence between factors. The basic assumption is that factors are interrelated according to the linear equation

$$\theta = \Delta\theta + \mathcal{G}, \quad (14)$$

where Δ is a parameter matrix, \mathcal{G} is a vector of random disturbances, and the distribution of \mathcal{G} is MNormal $(\mathbf{0}, \Omega)$. Equation (14) implies $(\mathbf{I} - \Delta)\theta = \mathcal{G}$, where \mathbf{I} is an identity matrix of order $D \times D$. Defining $\Upsilon = (\mathbf{I} - \Delta)^{-1}$, it follows that $\theta = \Upsilon\mathcal{G}$ and $\Sigma = \Upsilon\Omega\Upsilon'$. In summary, two matrices have to be defined to set up a structural model following the formalism in equation (14): Δ and Ω . The model in equation (14) can accommodate the same relations between factors as a LISREL model (Jöreskog & Sörbom, 1979).

The sample consists of a vector of frequencies $(n_1, \dots, n_t, \dots, n_T)$, where the element n_t is the number of individuals whose response pattern is $\mathbf{r}_t = (r_{t1}, \dots, r_{ti}, \dots, r_{tM})$, θ is regarded as a random effect, and the estimation proceeds by maximizing the multinomial marginal distribution of the response patterns. The log likelihood is given by $l(\epsilon) = \sum_{t=1}^T n_t \log P_t$, where ϵ is a vector that contains all the parameters to be estimated, and P_t is the marginal probability of response pattern t :

$$P_t = \int_{R^D} f(\mathbf{r}_t, \theta) d\theta. \quad (15)$$

One important difficulty in estimation is the evaluation of the D -dimensional integral in equation (15) and in its derivatives. Three numerical integration methods have been

compared in this investigation: static Gauss–Hermite quadrature, adaptive Gauss–Hermite, and Monte Carlo integration via importance sampling (MCEM), these methods are described in Schilling and Bock (2005). The details of the estimation algorithms appear in Appendix B, and are based on methods for matrix differential calculus described in Magnus and Neudecker (1999) and Harville (2008).

5. Simulation study

A simulation study was conducted to investigate the performance of the estimation algorithm in recovering true parameter values. The performance of the three integration methods for samples of different size was compared in the simulations.

The description of the simulated conditions and the results has been moved to Appendix C to preserve journal space, and this section contains only the main conclusions. Taken together, simulations show that adaptive–GH and MCEM improve static–GH when the number of items per factor is large and the structural model is simple. Static–GH selects quadrature points to optimize precision in the evaluation of the integral of a function over the population distribution of the factor: $f(\boldsymbol{\theta})$. Adaptive–GH and MCEM are based on a normal approximation the posterior distribution of the factor for each response pattern: $f(\boldsymbol{\theta}|\mathbf{r})$. The normal approximation to $f(\boldsymbol{\theta}|\mathbf{r})$ is based on the MAP estimate of $\boldsymbol{\theta}$ and its dispersion matrix, and its quality improves as the test length increases. The advantage of adaptive–GH and MCEM may be more prominent in long tests, whereas in short tests the imprecision in the MAP estimate may present a challenge to these two methods. Moreover, when the structural model is complicated, static–GH compares well with the other methods. One tentative explanation is that static–GH provides a stable basis for

computing the integral over the latent space, whereas the instability inherent to adaptive–GH and MCEM may compromise the estimation of involved structural models for latent variables.

6. Empirical example

The use of the MNCM to account for measurement problems is demonstrated with a real data set of an examination about inferential statistics. A sample of $n = 353$ second-course undergraduate students of psychology responded to 20 multiple-choice items as part of the final examination of the subject of data analysis, which ranged from one-sample t test to ANOVA, multiple comparisons, and linear regressions. The participants ranged from 19 to 30 years old and had 60 minutes to complete the test. Because data were collected for a real evaluation, the identity of the participants and any other information apart from the item responses were removed before the data matrix was submitted to investigation. The test was structured in six testlets. Each testlet consisted of two or more items related to a common theme, and each item had three response categories. Identification constraints were as in Sec. 3.1. Table 1 illustrates two of the testlets.

(Table 1)

Six models were applied to these data. All the models are variants of a bifactor model, depicted in Figure 1, that has a general factor and six group factors that explain the associations within each testlet. The six models are as follows:

- M1. Bifactor oblique. The correlations between the seven factors are free parameters. That is, the latent propensity is $z_{ir} = \tau_{ir} + \lambda_{ir1}\theta_1 + \lambda_{ir2}\theta_2$ and Σ contains 1s in the main diagonal and correlations between factors are estimated.
- M2. Bifactor partly orthogonal. The correlations between the general factor and

each of the group factors are free parameters. The correlations between group factors are set to zero. That is, z_{ir} is as in M1, Σ contains 1s in the main diagonal, correlations involving Factor 1 are free parameters and the other correlations are set to 0.

- M3. Bifactor orthogonal. The seven factors are uncorrelated; z_{ir} is as in M1 and Σ is an identity matrix.
- M4. Testlet model. This is a polytomous generalization of the dichotomous testlet model by Bradlow, Wainer, and Wang, X. (1999). In that model, the scale parameters of the general and the group factor have the same value to obtain a correlation between latent propensities with a convenient structure; that is, the latent propensity is $z_{ir} = \tau_{ir} + \lambda_{ir}(\theta_1 + \theta_2)$. The diagonal elements of Σ (factor variances) are free parameters and off-diagonal elements are zero.
- M5. Testlet model, equal variances. That is, z_{ir} is as in M4 and Σ is an identity matrix.
- M6. The one-dimensional nominal categories model; $z_{ir} = \tau_{ir} + \lambda_{ir}\theta$ and $\text{Var}(\theta)=1$.

(Figure 1)

These models are obtained as particular cases of the MNCM by using the method for linear constraints described in Appendix B. The model was estimated with two quadrature points for each dimension, as recommended by Schilling and Bock (2005) for a model with seven dimensions. However, note that this recommendation is based on the amount of computational resources that would be demanded by a larger number of points and not because two points is an optimal figure in a statistical sense. A number of statistics were used to test the model fit: the likelihood ratio for the hypothesis of exact fit,

the likelihood ratio against the most general model, M1, and the statistics AIC, BIC, and AICc, which consider the model fit and parsimony (Burnham & Anderson, 2004). Table 2 shows the results. Model M3 was selected for interpretation because of the values G^2 , AIC, and AICc. Notwithstanding, BIC differs from the other statistics and points to M5 as the preferred model. Table 2 contains also the mean estimation time in minutes from five different runs of the estimation algorithm using different random starting values. Estimation time depends largely on the number of estimated parameters in the structural model in equation (14).

(Table 2)

Table 3 presents the parameter estimates for M3. The item responses have been coded such that the correct response is the third one, and categories one and two are the distractors. The scale parameters in the six group factors have been placed in the same column, labeled *Group factor* in the table, although they refer to a different factor for each testlet. All but two of the scale parameters in the general factor were negative, indicating negative correlations between the distractors and the general factor. Although estimates were unbounded in the simulation study in Sect. 5, estimates were bounded between -10 and 10 in the real data analysis because the sample contains insufficient information to estimate some parameters. Table 3 shows that eleven intercept parameters reached the boundary during estimation. This is due to insufficient sample size. For example, the two distractors of Item 6, whose parameters τ are in the boundary, receive 6 and 8 responses. If the estimation is repeated without the bound of the parameters, these estimates converge at very extreme values, -18.4 and -19.8 for Item 6 with first order derivatives of 0.01 and 0.20 at convergence. Table 3 contains the results with the boundary because their actual estimated value is not important, the real problem is the lack of statistical information and the boundary helps to identify easily the weakly

identifiable parameters.

(Table 3)

Testlet 3 presents a number of interesting features. First, distractor B of item 10 (in the following, distractor 10-B) has a positive loading on the general factor ($\lambda_{10,2,1} = 0.37$). That is, the propensity to commit the misconception reflected in category B increases with θ_1 , which indicates that this topic was not well clarified during the teaching of the subject. Second, the categories 9-A and 10-B have remarkable loadings on the group factor, with reversed signs. The group factor θ_4 seems to be bipolar in the sense that individuals high in θ_4 feel attracted by the error reflected in 10-B, and individuals low in θ_4 tend toward the irreconcilable error contained in 9-A. The association between 9-A and 10-B is quite predictable in relation to item content and is explored more deeply in the following paragraphs.

In the model M3, the principle of independence between alternatives conditional on θ_i breaks down. To illustrate, the top row of Figure 2 shows the marginal response functions for items 10 and 17, with the group factor marginalized out. The response functions vary depending on the response given to the other item in the same testlet (items 9 and 16, respectively), as shown in the second, third, and fourth rows of Figure 2. The first item of the two testlets entails making a statistical decision about a null hypothesis, and the second item is the conclusion that follows from that decision. The interplay between the decisions and conclusions explains the associations. For example, the conclusion that two population means do not differ (category 10-B) has a smaller probability for those individuals that reject H_0 (category 9-A) than for those who do not reject H_0 (category 9-B).

(Figure 2)

Clustered items present an extra covariance beyond that explained by a one-dimensional latent trait. For the one-dimensional model, the covariance between items conditional on θ is zero. Regarding M3, if the group factor is integrated out, the result is a marginal distribution of responses that depends only on θ_I and has a non-null correlation between items. Ip (2010, Corollary 1) showed that the correlation between two dichotomous items, i and j , depends on the product of the scale parameters of the group factor. Appendix D extends that result to the polytomous case and shows how to compute item correlations in the marginal distribution that depends only on θ_I . The correlation depends largely on the product of the scale parameters but is not proportional to it. The key element that explains correlations in the polytomous case is the difference between a parameter and the mean of the scale parameters of the item. In particular, the correlation between categories r and s of items i and j is proportional to:

$$P_{ir\theta_1} \left(\lambda_{ir2} - \sum_k \lambda_{ik2} P_{ik\theta_1} \right) P_{js\theta_1} \left(\lambda_{js2} - \sum_k \lambda_{jk2} P_{jk\theta_1} \right), \quad (16)$$

where $P_{ik\theta_1}$ is the probability of the category when $\theta_2=0$. When data are dichotomous, the scale parameter of the first category is set to zero, and equation (16) reduces to $\lambda_{i22} \lambda_{j22} P_{i2\theta_1} (1 - P_{i2\theta_1}) P_{j2\theta_1} (1 - P_{j2\theta_1})$, which is the result by Ip (2010).

Figure 3 presents the correlation between pairs of categories conditional on θ_I for the two testlets in Table 1. As expected, categories 9-B (the correct one) and 10-B (an incorrect one) have a positive relation, which singles out a frequent misconception about the conclusion of a significance test when H_0 is maintained. Moreover, the correlation between the two correct categories, 9-B and 10-C, is negative because the distractor 10-B is more attractive for those who pass item 9 than for those who fail it, which is also predictable from item content. Regarding the second testlet, categories 16-A and 16-C have similar content and a comparable association pattern with 17-C, which is their

logical consequence. Categories 17-A and 17-B have a subtle difference and obtain similar patterns of association with item 16.

(Figure 3)

7. Final remarks

The purpose of the MNCM is to measure factors by using nominal response data. The key aspect of the model is that there is one slope parameter for each response category on each factor, instead of one slope for each item on each factor, as in factorial models for quantitative and ordinal data. In its more general form, it has minimal parameter constraints to ensure identifiability, although the paper shows how to impose more stringent restrictions by using linear constraints.

From a statistical point of view, the combination of multidimensionality and a heavy parameterization presents important challenges for estimation. The empirical example has shown that insufficient sample size may lead to weakly identifiable parameters. Exploratory analyses, in which the number of factors is increased until the model fits and item parameters are unconstrained, demand samples with thousands of examinees and time-consuming computational resources. Problems of weakly identified parameters may arise in exploratory analyses with smaller samples. Weakly identifiable parameters are identifiable in a mathematical sense, but the data provide scarce information about them, originating problems of convergence and large standard errors. Confirmatory analysis seems more tenable because fixed parameters and a smaller number of dimensions per item reduce the dimensionality of the optimization problem involved in estimation and increase statistical information about the remaining parameters.

The model has been developed in a traditional item response theory framework, with

strong distributional assumptions for latent factors and a marginal likelihood estimation algorithm. Three integration methods have been compared: adaptive GH, static GH, and MCEM. The simulation results revealed that adaptive GH provides the greatest precision, especially with a large number of items per factor. However, the estimation of a structured variance-covariance matrix for the factors could be more precise with static GH because of the increased stability of this method. The numerical integration problem limits the dimensionality of the latent space to about eight factors, and clearly, an area of further research is the development of more effective estimation methods for higher dimensionality with a reasonable demand for computational resources.

Ideally, an improved algorithm should allow faster estimation and increased dimensionality. The development of better algorithms, within the maximum likelihood framework, may be based on the Metropolis-Hastings Robins-Monro integration algorithm by Cai (2010-b). Estimation of the bifactor model can also be accelerated by applying the dimension reduction method (Cai, Yang, & Hansen, 2011), which reduces the D -dimensional integration problem to two-dimensional integration. Fully Bayesian simulation techniques are another promising inferential framework (Edwards, 2010). The inclusion of diffuse priors may speed up the estimation algorithm by avoiding cases in which weakly identified parameters use a long sequence of iterations to converge. Nevertheless, Bayesian simulation methods would also benefit from the use of the output of the EM maximization to construct a multivariate normal approximation of the posterior distribution of the parameters, which can be used to obtain samples of parameters. The assumption of multivariate normality of the latent variables has been made to avoid further complications although its relaxation is another topic for future research.

References

- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bock, R. D. & Gibbons, R. (2010). Factor analysis of categorical item responses. In: Nering, M. L. & Ostini, R. (Eds.). *Handbook of polytomous item response theory models*. New York. Routledge.
- Bock, R. D., Gibbons, R. & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.
- Bolt, D. M. & Johnson, T. R. (2009). Addressing score bias and differential Item functioning due to individual differences in response style. *Applied Psychological Measurement*, 33, 335-352.
- Bolt, D. M. & Wollack, J. A. (2012). Application of a multidimensional nested logit model to multiple-choice test items. *Psychometrika*, 77, 339-357.
- Bradlow, E. T., Wainer, H. & Wang, X. (1999). A Bayesian random effects model for testlet. *Psychometrika*, 64, 153-168.
- Burnham, K. P. & Anderson, P. R. (2004). Multimodel inference. Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33, 261-304.
- Cai, L. (2010-b). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, 75, 33-57.
- Cai, L., Yang, J. S. & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16, 221-248.
- Edwards, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika*, 75, 474-497.
- Fischer, G. H. (1983). Logistic latent trait model with linear constraints. *Psychometrika*, 48, 3-26.

- Harville, D. A. (2008). *Matrix algebra from a statistician's perspective*. New York. Springer.
- Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *British Journal of Mathematical and Statistical Psychology*, 63, 395-416.
- Johnson, T. R. & Bolt, D. M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style. *Journal of Educational and Behavioral Statistics*, 35, 92–114.
- Jöreskog, K. & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA. Abt Books.
- Magnus, J.R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. 2nd edition. New York: Wiley.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York. Springer.
- Revuelta, J. (2008). The generalized logit linear item response model for binary designed items. *Psychometrika*, 73, 385-405.
- Revuelta, J. (2009). Identifiability and equivalence of GLLIRM models. *Psychometrika*, 74, 257-272.
- Revuelta, J. (2010). Estimating difficulty from polytomous categorical data. *Psychometrika*, 75, 331-350.
- Rothenberg, T. (1971). Identification in parametric models. *Econometrica*, 39, 577-591.
- Schilling, S. & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70, 533-555.
- Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories model. In Nering, M. L. and Ostini, R. (Eds.). *Handbook of polytomous item response theory models*. New York, Routledge.

Acknowledgements

I would like to thank the editor and three anonymous reviewers for their comments that contributed to improve the manuscript.

This article has been supported by the grant PSI2012-31958 from the Spanish Ministry of Economy and Competitiveness

Appendix A. Equivalence of identifiability conditions

For any matrix \mathbf{W} of order $K \times L$, write \mathbf{W} , $\mathbf{W}^+ = (\mathbf{W}; \mathbf{1})$, and $\mathbf{W}^* = \mathbf{S}\mathbf{W}$ as:

$$\mathbf{W} = \begin{pmatrix} \mathbf{A} \\ \mathbf{b} \end{pmatrix}, \quad \mathbf{W}^+ = \begin{pmatrix} \mathbf{A} & \mathbf{1} \\ \mathbf{b} & 1 \end{pmatrix}, \text{ and } \mathbf{W}^* = (\mathbf{A} - \mathbf{1}\mathbf{b}), \quad (\text{A1})$$

where \mathbf{A} is a matrix of order $(K-1) \times L$, $\mathbf{1}$ is a vector of ones, and \mathbf{b} is a row vector. If vector $(\mathbf{b}; 1)$ is subtracted from the first $K-1$ rows of \mathbf{W}^+ , one gets:

$$\mathbf{G} = \begin{pmatrix} \mathbf{A} - \mathbf{1}\mathbf{b} & \mathbf{0} \\ \mathbf{b} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{W}^* & \mathbf{0} \\ \mathbf{b} & 1 \end{pmatrix}. \quad (\text{A2})$$

The demonstration is based on two facts. First, the rank of \mathbf{G} is the same as the rank of \mathbf{W}^+ because \mathbf{G} is obtained from \mathbf{W}^+ by the application of linear transformations that preserve the rank. Second, the last row of \mathbf{G} is linearly independent of the other rows and thus the rank of \mathbf{G} and \mathbf{W}^+ is the same as the rank of \mathbf{W}^* plus one.

Because \mathbf{W}^+ has order $K \times (L+1)$ and \mathbf{W}^* has order $(K-1) \times L$, a full column rank of \mathbf{W}^* implies a full column rank of \mathbf{W}^+ and vice versa, and the two identifiability conditions are equivalent.

The matrix \mathbf{W}^* has been used in the analysis of identifiability for linear transformation of the canonical parameters of the multinomial distribution in its exponential-family form, whereas \mathbf{W}^+ is more frequent in the item response theory literature. This appendix shows that both traditions lead to a similar result.

Appendix B. Inferential algorithms

B.1. Estimating factor scores

Let $\mathbf{r} = (r_1, \dots, r_M)$ be a vector of responses to M items. The Bayesian modal a posteriori estimate of $\boldsymbol{\theta}$ is the value that maximizes the joint distribution of \mathbf{r} and $\boldsymbol{\theta}$,

$f(\mathbf{r}, \boldsymbol{\theta}) = f(\boldsymbol{\theta}) \prod_{i=1}^M P_{ir_i}$, where P_{ir_i} denotes $P_{ir_i}(\boldsymbol{\theta})$. The estimation equation

$\partial \log f(\mathbf{r}, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}' = \mathbf{0}$ is solved using an iterative algorithm, such as Newton–Raphson.

The derivative is:

$$\frac{\partial}{\partial \boldsymbol{\theta}'} \log f(\mathbf{r}, \boldsymbol{\theta}) = \sum_{i=1}^M \left(\boldsymbol{\lambda}_{ir_i}' - \sum_{k=1}^{K_i} \boldsymbol{\lambda}_{ik}' P_{ik} \right) - \boldsymbol{\theta}' \boldsymbol{\Sigma}^{-1}. \quad (\text{B1})$$

The asymptotic variance–covariance matrix of the estimate of $\boldsymbol{\theta}$ is computed by inverting the observed information matrix: $I(\boldsymbol{\theta}) = -\partial^2 \log f(\mathbf{r}, \boldsymbol{\theta}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}')$, where:

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log f(\mathbf{r}, \boldsymbol{\theta}) = \sum_{i=1}^M \left(-\sum_{k=1}^{K_i} \boldsymbol{\lambda}_{ik} \boldsymbol{\lambda}_{ik}' P_{ik} + \left(\sum_{k=1}^{K_i} \boldsymbol{\lambda}_{ik} P_{ik} \right) \left(\sum_{k=1}^{K_i} \boldsymbol{\lambda}_{ik}' P_{ik} \right) \right) - \boldsymbol{\Sigma}^{-1}. \quad (\text{B2})$$

The last term on the right-hand side of equations (B1) and (B2) comes from $f(\boldsymbol{\theta})$ and is modified if this distribution is changed to another parametric family or dropped to obtain the maximum likelihood estimate instead of the Bayesian one.

B.2. Linear constraints on item parameters

In this paper, parameter constraints are imposed by linear equations, which are easily handled and reduce the dimensionality of the optimization problem involved in estimation by searching for the estimate in an unconstrained parameter space.

Constraints are imposed by the linear equations $\boldsymbol{\tau}_i = \mathbf{c}_{1i} + \mathbf{U}_{1i} \boldsymbol{\varepsilon}_{1i}$, $\boldsymbol{\lambda}_i = \mathbf{c}_{2i} + \mathbf{U}_{2i} \boldsymbol{\varepsilon}_{2i}$,

$\omega = \mathbf{c}_3 + \mathbf{U}_3 \boldsymbol{\varepsilon}_3$, and $\boldsymbol{\delta} = \mathbf{c}_4 + \mathbf{U}_4 \boldsymbol{\varepsilon}_4$; where \mathbf{c}_{1i} , \mathbf{c}_{2i} , \mathbf{c}_3 and \mathbf{c}_4 , are vector of constants that are used when parameter are fixed to nonzero values, \mathbf{U}_{1i} , \mathbf{U}_{2i} , \mathbf{U}_3 and \mathbf{U}_4 are matrices of coefficients that impose constraints, $\boldsymbol{\varepsilon}_{1i}$, $\boldsymbol{\varepsilon}_{2i}$, $\boldsymbol{\varepsilon}_3$ and $\boldsymbol{\varepsilon}_4$ are vectors of parameters, and $\boldsymbol{\lambda}_i$, ω and $\boldsymbol{\delta}$ are vectors that contain the non-duplicated elements of matrices $\boldsymbol{\Lambda}_i$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Delta}$, respectively. The vector that contains all estimated parameters ($\boldsymbol{\varepsilon}_{1i}$ and $\boldsymbol{\varepsilon}_{2i}$ for $i=1, \dots, M$; $\boldsymbol{\varepsilon}_3$ and $\boldsymbol{\varepsilon}_4$) is denoted by $\boldsymbol{\varepsilon}$, and the parameter space of $\boldsymbol{\varepsilon}$ is unconstrained. Subscripts will be used when necessary to refer to categories. For example, the equation $\tau_{ik} = c_{1ik} + \mathbf{U}_{1ik} \boldsymbol{\varepsilon}_{1i}$ places a constraint on the intercept parameter of category k of item i , where \mathbf{U}_{1ik} is the row of \mathbf{U}_{1i} that refers to this category.

For illustration, consider a two-dimensional model in which factors are standardized and their correlation is ρ . Then $\boldsymbol{\Delta} = \mathbf{0}$, and

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (\text{B3})$$

Vector ω takes the form $\omega = \text{vech} \boldsymbol{\Omega} = (1, \rho, 1)'$, and the equation $\omega = \mathbf{c}_3 + \mathbf{U}_3 \boldsymbol{\varepsilon}_3$ reduces to:

$$\begin{pmatrix} 1 \\ \rho \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} (\boldsymbol{\varepsilon}_3) \quad (\text{B4})$$

To further clarify notation, consider the constraints in $\boldsymbol{\Lambda}_i$ that appear in equation (9). In that example, the constraint $\boldsymbol{\Lambda}_i = \mathbf{W}_i \boldsymbol{\zeta}$ reads

$$\begin{pmatrix} \lambda_{i11} & \lambda_{i12} \\ \lambda_{i21} & \lambda_{i22} \\ \lambda_{i32} & \lambda_{i32} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \varsigma_{i11} & \varsigma_{i12} \\ \varsigma_{i21} & \varsigma_{i22} \end{pmatrix}. \quad (\text{B5})$$

The equation $\boldsymbol{\lambda}_i = \mathbf{c}_{2i} + \mathbf{U}_{2i} \boldsymbol{\varepsilon}_{2i}$ is obtained from equation (B5) by converting the

parameter matrices into vectors. The result is

$$\begin{pmatrix} \lambda_{i11} \\ \lambda_{i21} \\ \lambda_{i31} \\ \lambda_{i12} \\ \lambda_{i22} \\ \lambda_{i32} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \varsigma_{i11} \\ \varsigma_{i21} \\ \varsigma_{i12} \\ \varsigma_{i22} \end{pmatrix}. \quad (\text{B6})$$

B.3. Estimating the parameters of the structural and measurement models

The maximum likelihood estimate is found by solving the estimation equation $\partial l(\boldsymbol{\varepsilon}) / \partial \boldsymbol{\varepsilon}' = \mathbf{0}$ using a marginal likelihood/EM algorithm. The EM iterates between an E and an M step until convergence. The E step consists of computing two quantities:

1. The posterior distribution $f(\boldsymbol{\theta} | \mathbf{r}_t) = f(\mathbf{r}_t, \boldsymbol{\theta}) / P_t$, the value of which is computed for each response pattern, \mathbf{r}_t , and for the values of $\boldsymbol{\theta}$ used in the numerical integration procedure.
2. The posterior variance–covariance matrix of $\boldsymbol{\theta}$ weighted over response patterns:

$$\mathbf{C} = \sum_{t=1}^T n_t \text{Var}(\boldsymbol{\theta} | \mathbf{r}_t) = \sum_{t=1}^T n_t \int_{R^D} \boldsymbol{\theta} \boldsymbol{\theta}' f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}. \quad (\text{B7})$$

The M step consists of solving the equation $\partial l(\boldsymbol{\varepsilon}) / \partial \boldsymbol{\varepsilon}' = \mathbf{0}$ while keeping the quantities $f(\boldsymbol{\theta} | \mathbf{r}_t)$ and \mathbf{C} fixed to the values obtained in the E step. Let $j(\mathbf{r}_t) = \partial \log f(\mathbf{r}_t, \boldsymbol{\theta}) / \partial \boldsymbol{\varepsilon}'$ be the gradient (row) vector of $\log f(\mathbf{r}_t, \boldsymbol{\theta})$. Then, the first-order derivative of the log likelihood is given by:

$$\frac{\partial l(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}'} = \sum_{t=1}^T n_t \int_{R^D} j(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}. \quad (\text{B8})$$

The first-order derivatives of the log likelihood are the posterior expectations of $j(\mathbf{r}_t)$,

that is, $\partial l(\boldsymbol{\varepsilon}) / \partial \boldsymbol{\varepsilon}' = \sum_{t=1}^T n_t \int_{R^D} j(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}$, and are given by:

$$\frac{\partial}{\partial \boldsymbol{\varepsilon}_{1i}} l(\boldsymbol{\varepsilon}) = \sum_{t=1}^T n_t \int_{R^D} \left(\mathbf{U}_{1ir_i} - \sum_{k=1}^{K_i} P_{ik} \mathbf{U}_{1ik} \right) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}, \quad (\text{B9})$$

$$\frac{\partial}{\partial \boldsymbol{\varepsilon}_{2i}} l(\boldsymbol{\varepsilon}) = \sum_{t=1}^T n_t \int_{R^D} \left(\boldsymbol{\theta} \mathbf{U}_{2ir_i} - \sum_{k=1}^{K_i} P_{ik} \boldsymbol{\theta} \mathbf{U}_{2ik} \right) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}, \quad (\text{B10})$$

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\varepsilon}_3} l(\boldsymbol{\varepsilon}) &= \sum_{t=1}^T n_t \int_{R^D} \frac{1}{2} (\text{vec} \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\theta} \boldsymbol{\theta}' - \boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon}' \mathbf{U}_3 f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &= \frac{1}{2} (\text{vec} \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} (\mathbf{C} - n\boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon}') \mathbf{U}_3, \end{aligned} \quad (\text{B11})$$

$$\begin{aligned} \text{and } \frac{\partial}{\partial \boldsymbol{\varepsilon}_4} l(\boldsymbol{\varepsilon}) &= \sum_{t=1}^T n_t \int_{R^D} (\text{vec} \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\theta} \boldsymbol{\theta}' - \boldsymbol{\Sigma})' \mathbf{U}_4 f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &= (\text{vec} \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} (\mathbf{C} - n\boldsymbol{\Sigma}))' \mathbf{U}_4, \end{aligned} \quad (\text{B12})$$

where n is the sample size. For the interested readers, the operators vec and vech are described in Harville (2008).

B.4. Information matrix

The asymptotic variance–covariance matrix of the estimate of $\boldsymbol{\varepsilon}$ is obtained by inverting the observed information matrix, which can be computed from the vector $j(\mathbf{r}_t)$ and the Hessian matrix: $\mathbf{H}(\mathbf{r}_t) = \partial^2 \log f(\mathbf{r}_t, \boldsymbol{\theta}) / \partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}'$. The observed information matrix is given by:

$$\begin{aligned} I(\boldsymbol{\varepsilon}) &= - \sum_{t=1}^T n_t \int_{R^D} \mathbf{H}(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &\quad - \sum_{t=1}^T n_t \int_{R^D} j'(\mathbf{r}_t) j(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &\quad + \sum_{t=1}^T n_t \left[\int_{R^D} j'(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \right] \left[\int_{R^D} j(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \right]. \end{aligned} \quad (\text{B13})$$

The inverse of $I(\boldsymbol{\varepsilon})$ is the asymptotic variance–covariance of the maximum likelihood

estimate of $\boldsymbol{\varepsilon}$.

The first term in the formula for the observed information matrix can be written as

$-\mathbf{D}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon})$, where:

$$\begin{aligned}\mathbf{D}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) &= \sum_{t=1}^T n_t \int_{R^D} \mathbf{H}(\mathbf{r}_t) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &= \sum_{t=1}^T n_t \int_{R^D} \left(\frac{\partial^2}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}'} \log f(\mathbf{r}_t, \boldsymbol{\theta}) \right) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}.\end{aligned}\tag{B14}$$

Let \mathbf{K} be a commutation matrix and \mathbf{N} be the matrix in Magnus and Neudecker (1999, Chapter 3.7); \mathbf{K} and \mathbf{N} are of the order $D \times D$. The elements of $\mathbf{D}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon})$ are given by:

$$\begin{aligned}\mathbf{D}(\boldsymbol{\varepsilon}_{1i}, \boldsymbol{\varepsilon}_{1i}) &= -\sum_{t=1}^T n_t \int_{R^D} \sum_{k=1}^{K_i} \mathbf{U}_{1ik}' \mathbf{U}_{1ik} P_{ik} f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &\quad + \sum_{t=1}^T n_t \int_{R^D} \left(\sum_{k=1}^{K_i} \mathbf{U}_{1ik}' P_{ik} \right) \left(\sum_{k=1}^{K_i} \mathbf{U}_{1ik} P_{ik} \right) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}.\end{aligned}\tag{B15}$$

$$\begin{aligned}\mathbf{D}(\boldsymbol{\varepsilon}_{2i}, \boldsymbol{\varepsilon}_{2i}) &= -\sum_{t=1}^T n_t \int_{R^D} \sum_{k=1}^{K_i} \mathbf{U}_{2ik}' \boldsymbol{\theta}' \boldsymbol{\theta} \mathbf{U}_{2ik} P_{ik} f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &\quad + \sum_{t=1}^T n_t \int_{R^D} \left(\sum_{k=1}^{K_i} \mathbf{U}_{2ik}' \boldsymbol{\theta}' P_{ik} \right) \left(\sum_{k=1}^{K_i} \boldsymbol{\theta} \mathbf{U}_{2ik} P_{ik} \right) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}.\end{aligned}\tag{B16}$$

$$\begin{aligned}\mathbf{D}(\boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_3) &= \frac{1}{2} n \mathbf{U}_3' (\boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon} \otimes \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon})' \mathbf{U}_3 \\ &\quad - \frac{1}{2} n \mathbf{U}_3' (\boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \mathbf{C} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon} \otimes \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon})' \mathbf{U}_3 \\ &\quad - \frac{1}{2} n \mathbf{U}_3' (\boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon} \otimes \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \mathbf{C} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon})' \mathbf{U}_3.\end{aligned}\tag{B17}$$

$$\mathbf{D}(\boldsymbol{\varepsilon}_4, \boldsymbol{\varepsilon}_4) = -n \mathbf{U}_4' (\boldsymbol{\Upsilon} \otimes \boldsymbol{\Upsilon})' \mathbf{K} \mathbf{U}_4 - \mathbf{U}_4' (\mathbf{C} \otimes \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon})' \mathbf{U}_4.\tag{B18}$$

$$\begin{aligned}\mathbf{D}(\boldsymbol{\varepsilon}_{1i}, \boldsymbol{\varepsilon}_{2i}) &= -\sum_{t=1}^T n_t \int_{R^D} \sum_{k=1}^{K_i} \mathbf{U}_{1ik}' \boldsymbol{\theta} \mathbf{U}_{2ik} P_{ik} f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta} \\ &\quad + \sum_{t=1}^T n_t \int_{R^D} \left(\sum_{k=1}^{K_i} \mathbf{U}_{1ik}' P_{ik} \right) \left(\sum_{k=1}^{K_i} \boldsymbol{\theta} \mathbf{U}_{2ik} P_{ik} \right) f(\boldsymbol{\theta} | \mathbf{r}_t) d\boldsymbol{\theta}.\end{aligned}\tag{B19}$$

$$\mathbf{D}(\boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4) = -\mathbf{U}_3' \mathbf{N} (\boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \mathbf{C} \otimes \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon})' \mathbf{U}_4.\tag{B20}$$

All the other derivatives vanish.

B5. Adaptive Gauss-Hermite quadrature

Multidimensional integrals are evaluated numerically by using an adaptive Gauss-Hermite quadrature procedure (Liu & Pierce, 1994), first applied to item-response modeling by Schilling and Bock (2005). Adaptive GH quadrature evaluates the integral in equation (15) by selecting the quadrature nodes and weights that optimize the precision of the numerical approximation for each response pattern. The marginal probability, P_t , of response pattern \mathbf{r}_t is approximated using a specific grid of $\boldsymbol{\theta}$ values in the numerical integration procedure than may be different from the grid for other response patterns. The grid is obtained by constructing a multivariate normal approximation for the integrand $f(\mathbf{r}_t, \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}$ be the value of $\boldsymbol{\theta}$ that maximizes $f(\mathbf{r}_t, \boldsymbol{\theta})$ and $I(\boldsymbol{\theta})$ the information matrix obtained from the second derivative in equation (B2), that is:

$$I(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log f(\mathbf{r}, \boldsymbol{\theta}). \quad (\text{B21})$$

The multivariate normal approximation to $f(\mathbf{r}_t, \boldsymbol{\theta})$ is given by:

$$h_t(\boldsymbol{\theta}) = \frac{|I(\boldsymbol{\theta})|^{1/2}}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' I(\boldsymbol{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right). \quad (\text{B22})$$

Equation (15) can now be rewritten as:

$$P_t = \int_{R^D} \left(\frac{f(\mathbf{r}_t, \boldsymbol{\theta})}{h_t(\boldsymbol{\theta})} \right) h_t(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (\text{B23})$$

and the quadrature nodes and weights are selected to optimize the precision of the evaluation of $h_t(\boldsymbol{\theta})$.

Nodes and weights are found by reducing the problem to iterated one-dimensional

Gauss-Hermite integration. This is done by performing the change of variable

$\mathbf{x} = \mathbf{L}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) / \sqrt{2}$, where \mathbf{L} is the lower triangular Cholesky factor of $I(\boldsymbol{\theta})$. Then,

$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \sqrt{2}\mathbf{L}\mathbf{x}$, differentiating: $d\boldsymbol{\theta} = 2^{D/2} |\mathbf{L}| d\mathbf{x}$, and applying the change of variable:

$$\begin{aligned} h_t(\boldsymbol{\theta}) &= \frac{|I(\boldsymbol{\theta})|^{1/2}}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' I(\boldsymbol{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right) \\ &= \frac{|I(\boldsymbol{\theta})|^{1/2}}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' (\mathbf{L}^{-1})' \mathbf{L}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right) \\ &= \frac{|I(\boldsymbol{\theta})|^{1/2}}{(2\pi)^{D/2}} \exp(-\mathbf{x}' \mathbf{x}). \end{aligned} \quad (\text{B24})$$

Inserting equation (B24) in (B23), it follows that the marginal probability can be evaluated by iterated one-dimensional Gauss-Hermite quadrature:

$$\begin{aligned} P_t &= \int \left(\frac{f(\mathbf{r}_t, \boldsymbol{\theta})}{h_t(\boldsymbol{\theta})} \right) h_t(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= 2^{D/2} |\mathbf{L}| \int_{R^D} f(\mathbf{r}_t, \boldsymbol{\theta}) \exp(\mathbf{x}' \mathbf{x}) \exp(-\mathbf{x}' \mathbf{x}) d\mathbf{x} \\ &\approx 2^{D/2} |\mathbf{L}| \sum_{x_1} \cdots \sum_{x_D} f(\mathbf{r}_t, \boldsymbol{\theta}) \exp(\mathbf{x}' \mathbf{x}) w_{x_1} \cdots w_{x_D} \\ &= 2^{D/2} |\mathbf{L}| \sum_{x_1} \cdots \sum_{x_D} f(\mathbf{r}_t, \boldsymbol{\theta}) w_{x_1}^* \cdots w_{x_D}^*, \end{aligned} \quad (\text{B25})$$

where $w_{x_i}^* = \exp(x_i^2) w_{x_i}$. The values x_1, \dots, x_D and w_{x_1}, \dots, w_{x_D} are the quadrature nodes and weights for a one-dimensional Gauss-Hermite quadrature that are tabulated, for example, in Abramowitz and Stegun (1965).

B6. References

- Abramowitz, M. & Stegun, I. A. (1965). *Handbook of mathematical functions*. New York. Dover Publications.
- Liu, Q. & Pierce, D. A. (1994). A note on G-H quadrature. *Biometrika*, 81, 624-629.

Appendix C. Conditions and results of the simulation study

A computer program was written in the C++ language using the Borland Builder 6 compiler. The simulations were run in a computer equipped with an Intel Xeon E5-1620 processor clocked at 3.60 GHz, 10 MB cache, and 4 cores, with 12 GB RAM, running Microsoft Windows 7 on 64 bits.

After some pilot runs with models of varying complexity, it was clear that the numerical integration procedures are computationally demanding and time-consuming. For these reasons and because the simulation means repeating the estimation algorithm for thousands of artificial data matrices, relatively simple conditions were used in the simulations to keep the computational demands within a manageable limit. Although the conditions of the simulation involve few dimensions, factors, and categories, they are illustrative of the relative performance of the estimation method for the different types of parameters. In the first simulation, a model with two orthogonal factors, each measured by 15 dichotomous items, was assumed. True item and individual parameter values were sampled from a standard normal distribution. Table C-1 contains true item parameters.

The three integration methods were compared. The number of quadrature points for each dimension was 3, 5, and 7 so the total number of points for numerical integration was 9, 25, and 49, respectively. These figures were taken from (Schilling & Bock, 2005), who recommend as few as two points for six or more factors. Thus, the condition with three quadrature points is informative about the expected precision in the worst case, whereas the other conditions are more realistic for a model with two factors.

Table C-1. True parameter values for the first simulation

Factor 1					
Parameters λ			Parameters τ		
Items 1-5	Items 6-10	Items 11-15	Items 1-5	Items 6-10	Items 11-15
1.309	-0.002	-0.004	-0.081	-0.713	-0.460
0.665	-0.481	-0.116	-3.294	-1.203	-1.961
0.060	2.769	0.363	-0.838	-0.482	1.243
-0.740	-0.356	1.300	1.168	-0.663	0.424
1.369	-0.881	0.058	1.599	-1.787	-1.302
Factor 2					
Parameters λ			Parameters τ		
Items 1-5	Items 6-10	Items 11-15	Items 1-5	Items 6-10	Items 11-15
-1.263	-0.999	1.749	-1.500	0.177	-1.436
-0.609	-0.093	-0.327	-0.245	-0.686	-1.288
0.662	0.008	1.177	0.444	-0.452	0.828
-0.413	-0.411	0.291	-0.600	0.944	0.355
1.540	-0.174	-0.838	-1.504	-0.846	0.004

Note: The true model has two orthogonal factors, each measured with 15 one-dimensional dichotomous items.

The simulated sample sizes were 1000 and 2000. One hundred simulated samples were taken for each of the 18 conditions of the experimental design (three integration methods \times three numbers of quadrature points \times two sample sizes). The starting values were taken at random from a uniform (-1, 1) distribution. The convergence criterion was a length of the gradient vector smaller than 0.01. The recovery of individual and item parameters was measured by the root-mean-squared error between true and estimated values; Table C-2 shows the results.

The recovery of intercept parameters has similar precision to the three integration methods. However, adaptive-GH was more accurate for the recovery of scale parameters, followed by MCEM, whereas static-GH was the least precise method. The Bayesian MAP estimate of factor scores was computed because it is used in the adaptive-GH procedure. No important difference was found in the recovery of factor scores between the simulation conditions.

Table C-2. Recovery of the parameters in the first simulation

Method	Points	$n = 1000$			$n = 2000$		
		λ	τ	θ	λ	τ	θ
Static GH	3	.176	.094	.592	.141	.069	.599
	5	.163	.093	.604	.149	.061	.587
	7	.144	.093	.589	.102	.064	.602
	Mean	.161	.094	.595	.129	.065	.596
Adaptive GH	3	.132	.097	.611	.094	.066	.583
	5	.128	.095	.590	.095	.066	.590
	7	.132	.095	.589	.084	.065	.590
	Mean	.131	.095	.597	.091	.065	.587
MCEM	3	.151	.097	.583	.099	.062	.593
	5	.144	.094	.603	.102	.066	.594
	7	.148	.095	.604	.102	.069	.596
	Mean	.148	.095	.597	.101	.065	.594

Note: The table contains the root mean squared error between 100 estimated parameters and the true values in Table C-1.

A second simulation was performed to investigate recovery with a more sophisticated latent structure. The measurement model was simpler, with fewer items per dimension, and the structural model was more involved. The theoretical model has two correlated exogenous factors and one endogenous factor related by the equation $\theta_1 = \delta_1\theta_2 + \delta_2\theta_3 + \mathcal{G}$, $Corr(\theta_2, \theta_3) = \omega$, and $Var(\mathcal{G}) = 1$ for identifiability. The matrices in the structural model are then

$$\mathbf{\Delta} = \begin{pmatrix} 0 & \delta_1 & \delta_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{\Omega} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \omega \\ 0 & \omega & 1 \end{pmatrix}. \quad (\text{C1})$$

Each factor was measured using six dichotomous variables. Item parameters were taken at random from a standard normal distribution, whereas a uniform (0, 1) distribution was used for ω and δ . In this model there is at least one free parameter for each of the four types of parameters needed to specify the model. The conditions of the simulation—regarding integration methods, quadrature points, and sample size—are the same as in the first. True parameter values appear in Table C-3.

Table C-3. True parameter values for the second simulation

Factor 1		Factor 2		Factor 3	
λ	τ	λ	τ	λ	τ
0.41		0.11		0.61	
0.78	1.44	0.21	-1.44	0.22	-0.45
0.30	-0.90	0.26	-1.45	1.02	-0.61
0.31	-1.12	0.72	1.10	.039	0.11
0.73	0.00	0.36	-0.80	2.36	-0.42
1.89	1.30	1.24	-0.85	0.04	0.65
$\omega = 0.25$		$\delta_1 = 0.36$		$\delta_2 = -0.14$	

Note: Empty cells are structural zeros. Factor 1 is an endogenous factor, and factors 2 and 3 are exogenous. The last row of the table shows the correlation between exogenous factors and the latent regression weights.

Table C-4 contains the root-mean-squared error. In the second simulation, the advantage of adaptive-GH with respect to static-GH is smaller than in Simulation 1 for the small sample and vanishes for the large sample. MCEM was less precise than the other methods.

A third simulation was run to investigate recovery for a polytomous model. The theoretical model consists of 20 items with three categories that measure a bifactor model with a general factor and two group factors. Σ is an identity matrix. Identification constraints were as in Sect. 3.1. True parameter values appear in Table C-5.

Table C-6 presents the recovery of parameters, including the RMSE for factor intercepts and slopes and the mean estimation time for each condition. The results show that static-GH is surpassed by the other methods and that the recovery of the general factor is better than that of the group factors for adaptive-GH and MCEM. The precision of these two methods depends on a normal approximation to $f(\boldsymbol{\theta}|\mathbf{r})$. The increase in the number of items per factor improves the normal approximation and, consequently, the precision of the adaptive-GH and MCEM methods. The correlations between true and

estimated values of θ were about 0.91 for factor 1, 0.80 for factor 2 and 0.85 for factor 3, with small changes across conditions. The number of items per factor seems to be influential in these correlations.

Table C- 4. Recovery of the parameters in the second simulation

Method	Points	$n = 1000$				$n = 2000$			
		ω	δ	λ	τ	ω	δ	λ	τ
Static GH	3	.084	.196	.241	.192	.055	.200	.152	.121
	5	.086	.216	.234	.176	.055	.201	.143	.104
	7	.101	.207	.249	.184	.068	.208	.149	.107
	Mean	.090	.206	.241	.184	.059	.203	.148	.111
Adaptive GH	3	.082	.211	.216	.173	.055	.200	.165	.126
	5	.093	.214	.209	.173	.061	.194	.146	.120
	7	.097	.211	.257	.177	.054	.205	.150	.110
	Mean	.091	.212	.228	.174	.057	.199	.153	.119
MCEM	3	.091	.189	.277	.221	.063	.186	.165	.166
	5	.094	.224	.280	.285	.064	.238	.167	.224
	7	.087	.201	.265	.177	.062	.183	.161	.116
	Mean	.091	.207	.275	.234	.063	.203	.164	.169

Note: The table contains the root mean squared error between 100 estimated parameters and the true values in Table C-3.

The estimation time increases exponentially with the number of points per dimension. The slowest simulations, with three factors and seven points, required almost three hours to run for each sample. The estimation time could be reduced by optimizing the computer code. Some optimizations are more suitable for some methods than for others, and the relative speed of the methods may be altered. For example, static-GH may be largely

optimized because it is based on a fixed grid of weights and values of θ , and many partial results can be stored in a table that is reused for all response patterns. This optimization, which was not implemented in the present study, is unfeasible for adaptive-GH because the grid changes during iterations and from one response pattern to another. However, the exponential increase in time within each method seems unavoidable.

Table C-5. True parameter values for the third simulation

Item			Factor 1		Factor 2		Factor 3	
	τ_{i1}	τ_{i2}	λ_{i11}	λ_{i21}	λ_{i12}	λ_{i22}	λ_{i13}	λ_{i23}
1	-0.56	-0.18	-0.40	0.92	1.12	-0.87		
2	0.15	-1.97	-0.25	-0.68	0.19	0.56		
3	0.33	-0.70	2.07	-0.64	-1.17	0.79		
4	-0.81	1.92	0.34	2.15	-0.10	-0.44		
5	0.95	-0.71	0.86	-1.19	0.56	-0.27		
6	-0.48	-0.58	0.29	0.82	1.13	-0.19		
7	-0.46	0.45	0.55	-0.36	-0.59	-0.07		
8	0.90	-0.53	1.02	-1.17	-0.83	-1.69		
9	-1.54	-0.32	-0.50	0.32	0.18	0.94		
10	-0.81	-1.08	2.23	2.57	0.06	1.33		
11	0.70	1.57	0.53	2.46			-2.11	-0.82
12	2.25	-0.22	-0.44	0.63			1.21	0.40
13	2.05	-0.05	1.21	1.32			-0.79	-1.52
14	0.94	0.80	1.52	0.07			1.17	2.26
15	0.47	-0.04	-0.75	-0.80			0.45	0.96
16	0.34	-1.23	2.74	0.40			0.46	-0.50
17	1.27	2.15	0.43	0.42			-0.06	-1.77
18	0.29	-0.25	0.00	-0.37			-0.67	0.37
19	0.30	-0.15	0.62	-0.04			0.01	-0.06
20	-0.65	-1.04	0.55	0.86			2.65	1.14

Note: The true model is a bifactor model. Factor 1 is a common factor and factors two and three are specific ones. The items have three response categories.

The main difficulty in conducting the simulation study was the long estimation time for some conditions. Although parameter estimates are unbounded in the simulation, no problem of convergence arose because models are relatively simple, with few categories per item and large samples. The gradient of the log likelihood was monitored and it goes

to zero for all conditions. However, the stopping criterion is based on parameter changes because, near the maximum of the likelihood, gradient reduces very slow and estimates changes little during iterations. The speed of convergence varies between samples within each condition. The number of iterations for estimation was typically about 40, although convergence required only about 10 iterations in some samples and more than 120 in others.

Table C-6. Recovery of the parameters in the third simulation

Method	Points	$n = 1000$						$n = 2000$					
		$\lambda_{..1}$	$\lambda_{..2}$	$\lambda_{..3}$	$\tau_{..}$	θ	T	$\lambda_{..1}$	$\lambda_{..2}$	$\lambda_{..3}$	$\tau_{..}$	θ	T
Static	3	.295	.189	.213	.146	.530	5.3	.291	.145	.171	.102	.532	8.0
GH	5	.196	.160	.187	.143	.530	32.3	.180	.125	.120	.106	.525	47.8
	7	.163	.159	.181	.148	.533	66.0	.128	.120	.123	.103	.526	170.5
	Mean	.210	.167	.191	.146	.531	34.5	.203	.130	.138	.104	.528	75.4
Adaptive	3	.157	.159	.194	.145	.533	2.3	.108	.112	.127	.099	.532	3.1
GH	5	.153	.156	.184	.146	.535	15.1	.105	.116	.127	.099	.533	24.9
	7	.155	.157	.189	.148	.534	43.7	.105	.118	.121	.098	.529	70.0
	Mean	.155	.157	.189	.146	.533	20.4	.106	.115	.125	.099	.531	32.7
MCEM	3	.156	.162	.185	.142	.529	2.3	.105	.112	.119	.095	.531	2.8
	5	.150	.159	.182	.141	.533	12.5	.101	.113	.121	.098	.531	14.7
	7	.153	.163	.176	.141	.537	35.7	.102	.116	.121	.100	.530	71.2
	Mean	.153	.161	.182	.141	.533	16.8	.103	.113	.120	.097	.530	29.6

Note: The table contains the root mean squared (RMSE) error between 100 estimated parameters and the true values in Table C-5. Columns labeled $\lambda_{..1}$, $\lambda_{..2}$, $\lambda_{..3}$ and $\tau_{..}$ contain the RMSE for each parameter type. T is the average estimation time in minutes.

The number of quadrature points is below the recommendation of Schilling and Bock

(2005) for some of the conditions of the simulations. This is so because running the simulations with more points would have demanded a prohibitive amount of computational resources. However, results showed that recovery did not improved much by increasing the number of points. Apparently, relatively simple models do not need a large number of points. The simplicity of simulated models avoids convergence problems. However, these problems may easily arise from the combination of small samples and complex models.

Appendix D. Taylor series approximation to the correlation between categories

Let Y_{ir} be an indicator variable with value 1 for category r of item i and 0 otherwise.

Consider a two-dimensional model. A Taylor series approximation about $\theta_2=0$ gives the following approximation to the probability function of Y_{ir} (Ip, 2010):

$$P_{ir} = P_{ir\theta_1} + \left[\frac{\partial}{\partial \theta_2} P_{ir} \right]_{\theta_2=0} \times \theta_2 + O(\theta_2^2), \quad (D1)$$

where $P_{ir\theta_1}$ is the probability function in equation (1) evaluated as a function of θ_1 with

$\theta_2=0$, and

$$h_{ir}(\theta_1) \stackrel{\text{def}}{=} \left[\frac{\partial}{\partial \theta_2} P_{ir} \right]_{\theta_2=0} = P_{ir\theta_1} \left(\lambda_{ir2} - \sum_k \lambda_{ik2} P_{ik\theta_1} \right). \quad (D2)$$

By the law of total variance, the variance of Y_{ir} for a fixed value of θ_1 is given by

$\text{Var}(Y_{ir}) = \text{E}(\text{Var}(Y_{ir} | \theta_2)) + \text{Var}(\text{E}(Y_{ir} | \theta_2))$, where the expectation is taken with respect

to the distribution $f(\theta_2 | \theta_1)$. Taking into account that the expectation of Y_{ir} is P_{ir} and

ignoring the second- and higher-order terms in the Taylor series of P_{ir} , the term

$\text{Var}(\text{E}(Y_{ir} | \theta_2))$ is approximated by

$$\text{Var}(\text{E}(Y_{ir} | \theta_2)) \approx h_{ir}^2(\theta_1) \text{Var}(\theta_2 | \theta_1). \quad (D3)$$

Assuming that $f(\theta_1, \theta_2)$ is bivariate normal $(\mathbf{0}, \Sigma)$, the conditional moments of θ_2 are

$\text{E}(\theta_2 | \theta_1) = \rho \theta_1 \sigma_2 / \sigma_1$ and $\text{Var}(\theta_2 | \theta_1) = \sigma_2^2 (1 - \rho^2)$, where $\rho = \sigma_{12} / (\sigma_1 \sigma_2)$. Then

$$\text{Var}(\text{E}(Y_{ir} | \theta_2)) \approx h_{ir}^2(\theta_1) \sigma_2^2 (1 - \rho^2). \quad (D4)$$

The first term in the expression for $\text{Var}(Y_{ir})$ is given by

$E(\text{Var}(Y_{ir} | \theta_2)) = \int P_{ir}(1 - P_{ir}) f(\theta_2 | \theta_1) d\theta_2$. Let $w_{ir} = P_{ir}(1 - P_{ir})$; a Taylor series approximation to w_{ir} about $\theta_2=0$ results in

$$w_{ir} = w_{ir\theta_1} + \left[\frac{\partial}{\partial \theta_2} w_{ir} \right]_{\theta_2=0} \times \theta_2 + O(\theta_2^2), \quad (\text{D5})$$

where $w_{ir\theta_1}$ is the w_{ir} evaluated with $\theta_2=0$, and

$$\frac{\partial}{\partial \theta_2} w_{ir} = \left(\lambda_{ir2} - \sum_k \lambda_{ik2} P_{ik} \right) P_{ir} (1 - 2P_{ir}) \quad (\text{D6})$$

Then

$$\begin{aligned} E(\text{Var}(Y_{ir} | \theta_2)) &\approx w_{ir\theta_1} + \left(\lambda_{ir2} - \sum_k \lambda_{ik2} P_{ik\theta_1} \right) P_{ir\theta_1} (1 - 2P_{ir\theta_1}) E(\theta_2 | \theta_1) \\ &= P_{ir\theta_1} \left(1 - P_{ir\theta_1} + \left(\lambda_{ir2} - \sum_k \lambda_{ik2} P_{ik\theta_1} \right) P_{ir\theta_1} (1 - 2P_{ir\theta_1}) \frac{\rho \theta_1 \sigma_2}{\sigma_1} \right) \end{aligned} \quad (\text{D7})$$

By the law of total covariance, the covariance between Y_{ir} and Y_{js} ($i \neq j$) is given by

$\text{Cov}(Y_{ir}, Y_{js}) = E(\text{Cov}(Y_{ir}, Y_{js} | \theta_2)) + \text{Cov}(E(Y_{ir}, Y_{js} | \theta_2))$. The first term is zero because

$\text{Cov}(Y_{ir}, Y_{js} | \theta_2) = 0$, and a Taylor series approximation to the second term gives

$$\begin{aligned} \text{Cov}(E(Y_{ir}, Y_{js} | \theta_2)) &\approx h_{ir}(\theta_1) \sigma_2^2 (1 - \rho^2) h_{js}(\theta_1) \\ &= P_{ir\theta_1} \left(\lambda_{ir2} - \sum_k \lambda_{ik2} P_{ik\theta_1} \right) \sigma_2^2 (1 - \rho^2) P_{js\theta_1} \left(\lambda_{js2} - \sum_k \lambda_{jk2} P_{jk\theta_1} \right) \end{aligned} \quad (\text{D8})$$

The approximation to the linear correlation between categories for a fixed value of θ_1 is

computed as

$$\text{Corr}(Y_{ir}, Y_{js}) \approx \text{Cov}(Y_{ir}, Y_{js}) / \sqrt{\text{Var}(Y_{ir}) \text{Var}(Y_{js})} \quad (\text{D9})$$

Table 1. Two testlets from the exam of data analysis

The means of two independent samples have been compared using a t test. The null hypothesis is $H_0: \mu_1 = \mu_2$ and the test statistic takes the value $T = 1.56$.

ITEM 9. If $P(T > 1.56) = 0.10$ the statistical decision is:

- A) Reject H_0 .
- B) Do not reject H_0 .
- C) There is not enough information to reach a decision.

ITEM 10. Taking into account that the null hypothesis is a bilateral one:

- A) The p-value is 0.10.
- B) We shall conclude that the population means do not differ.
- C) Statements A and B are incorrect.

The mean level of stress of a sample of 20 individuals has been compared before and after an anti-stress treatment. The null hypothesis is that there is no difference between the mean before and after the treatment, and the alternative hypothesis says that there is a positive difference. The test statistic is $T = 2$. If $P(T \leq 2) = 0.98$ and the type I error rate is $\alpha = 0.05$.

ITEM 16. Shall we reject H_0 ?

- A) No, because 0.98 is higher than 0.05.
- B) Yes, because 0.02 is smaller than 0.05.
- C) No, because 0.98 is higher than 0.09.

ITEM 17. What is the correct conclusion for this contrast?

- A) The mean stress level has reduced.
- B) The intervention has reduced the mean stress level.
- C) There is no evidence that the intervention has reduced the mean stress level.

Note: Correct categories are B, C, B and A.

Table 2. Goodness of fit statistics

Model	Log Lik.	G^2	Params	ΔG^2	Δgl	p	AIC	AIC _c	BIC	T
M1	-3169.63	2231.30	141				6621.26	6810.15	7166.83	52
M2	-3175.80	2243.65	126	12.35	15	0.65	6603.61	6744.60	7091.14	40
M3	-3180.78	2253.61	120	22.31	21	0.38	6601.57	6726.21	7065.89	35
M4	-3261.20	2414.45	86	183.15	55	0.00	6694.41	6750.45	7027.17	34
M5	-3272.10	2436.24	80	204.94	61	0.00	6704.20	6751.67	7013.75	23
M6	-3304.59	2501.21	80	269.91	61	0.00	6769.17	6816.64	7078.72	1

Note: Models M1 to M6 are described in the text. Log. Lik is the log-likelihood. G^2 is the likelihood ratio test against a general multinomial alternative. Params is the number of estimated parameters. The columns ΔG^2 , Δgl and p contains the result of the likelihood ratio test of each model against M1. AIC, AIC_c, BIC are the information-theoretic statistics. T is the average estimation time in minutes.

Table 3. Parameter estimates for the MNCM

Testlet	Item	General factor				Group factor	
		τ_1	τ_2	λ_1	λ_2	λ_1	λ_2
1	1	-10.00	-4.91	-0.72	-1.76	6.25	-0.22
	2	-10.00	-5.47	-2.94	-1.50	0.93	0.67
	3	-5.55	-4.59	-1.84	-1.46	-3.53	-0.87
	4	-0.34	-2.37	-0.10	-1.88	0.16	-0.14
	5	-0.64	-10.00	-0.11	-2.32	-0.10	-8.49
2	6	-10.00	-10.00	-1.32	-0.22	-4.83	-5.96
	7	-1.80	-10.00	-0.68	-1.28	-0.17	-5.12
	8	3.97	10.00	-2.34	-0.14	-7.95	-7.32
3	9	-5.35	-0.86	-2.62	-0.20	-3.31	-0.95
	10	-1.12	-3.15	-0.15	0.37	0.06	3.41
4	11	-10.00	-3.95	-0.58	-0.81	-6.05	0.12
	12	-1.91	-10.00	-0.33	3.28	0.13	-1.75
	13	-10.00	-3.34	-0.56	-0.88	-5.38	0.09
	14	-1.95	0.01	-1.03	-0.51	-0.37	-0.31
	15	-1.51	-3.76	-0.69	-1.00	-0.20	0.36
5	16	-3.97	-5.00	-2.20	-2.92	-3.51	-1.79
	17	0.23	-3.09	-1.06	-3.51	-0.75	-6.00
6	18	-2.18	-4.06	-0.05	-1.33	-0.67	-0.27
	19	-3.99	-10.00	-0.75	-0.66	0.04	-6.35
	20	-0.92	-1.88	-0.40	-0.70	-0.43	-0.39

Note: The table contains the parameters of the two incorrect categories, the parameters of the correct option are set to zero. There are six group factors, one for each testlet.

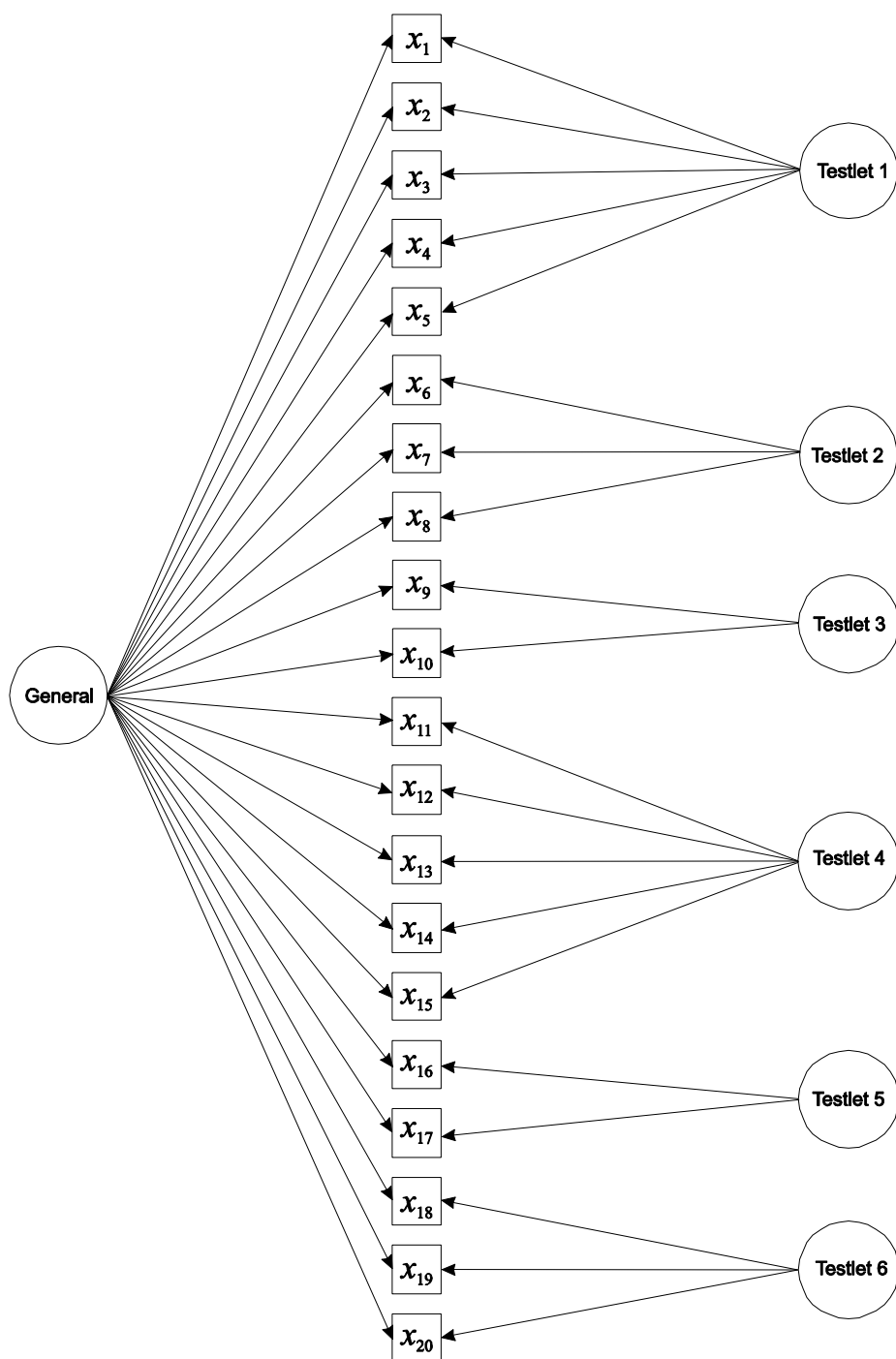


Figure 1. General structure of the factorial models fitted in the empirical example

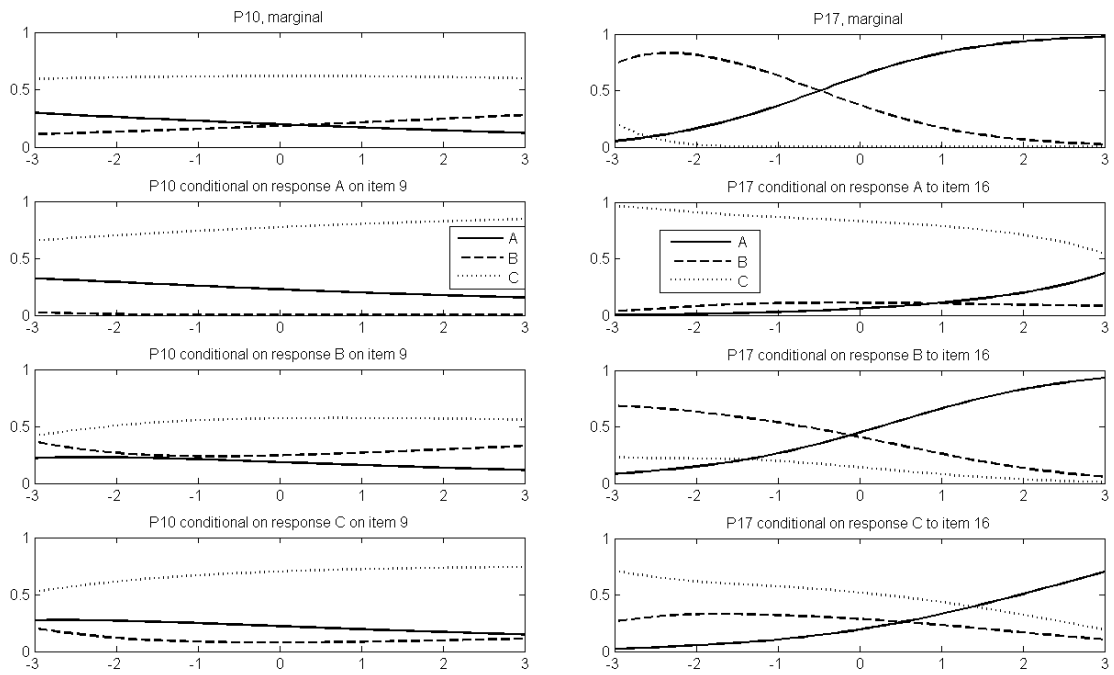


Figure 2. Marginal response functions for items 10 and 17 and response functions conditional on the response given to the preceding item in the test. The probabilities are shown as a function of the general factor, θ_1 , and the group factor has been integrated out.

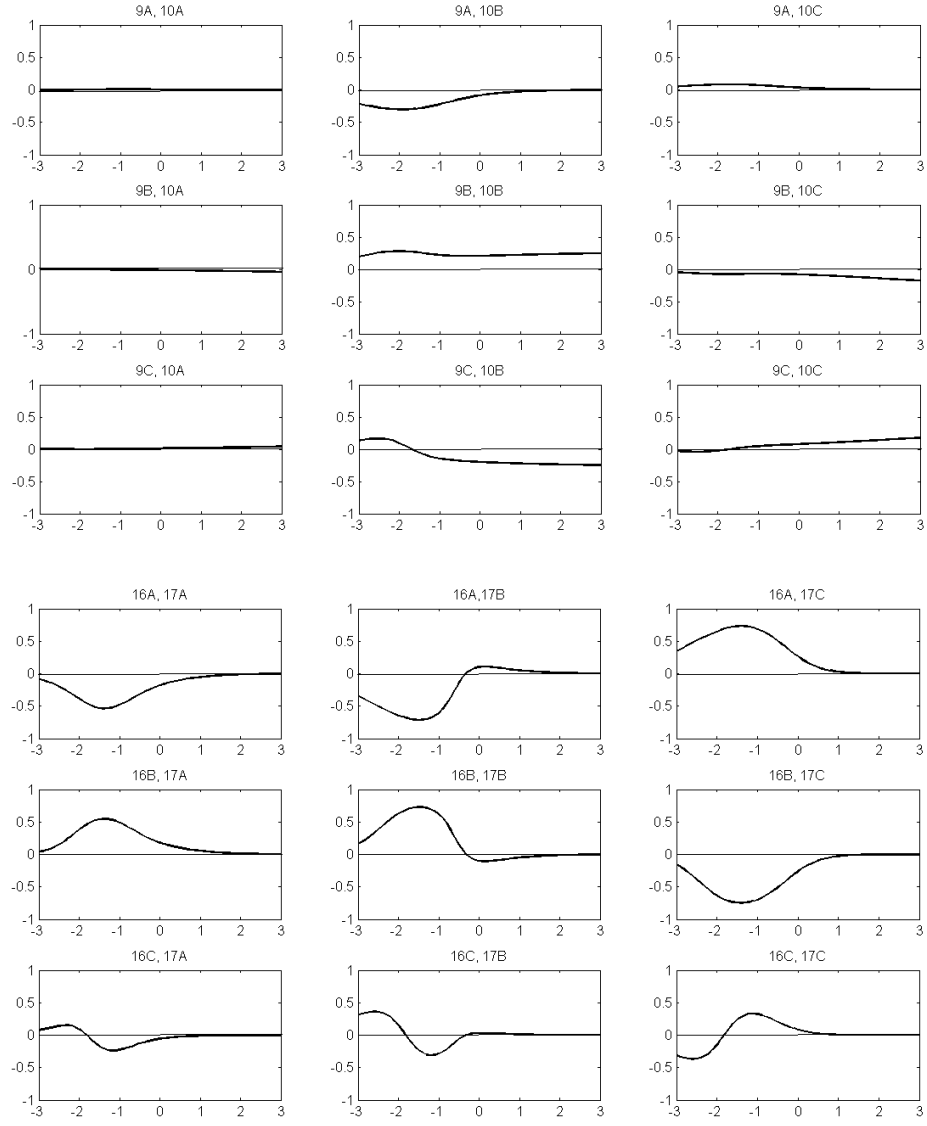


Figure 3. Correlation between categories as a function of θ_l with the group factor integrated out. The upper panel shows the correlations between items 9 and 10, and the lower panel the correlations between items 16 and 17.