Brazilian Journal of Probability and Statistics
2010, Vol. 24, No. 3, 415–433
DOI: 10.1214/09-BJPS102

An estimation method for latent traits and population parameters in Nominal Response Model

Caio L. N. Azevedo[a] **and Dalton F. Andrade**[b]

[a]*University of Campinas*
[b]*Universidade Federal de Santa Catarina*

Abstract. The nominal response model (NRM) was proposed by Bock [*Psychometrika* **37** (1972) 29–51] in order to improve the *latent trait (ability)* estimation in multiple choice tests with nominal items. When the item parameters are known, expectation a posteriori or maximum a posteriori methods are commonly employed to estimate the latent traits, considering a standard symmetric normal distribution as the latent traits prior density. However, when this item set is presented to a new group of examinees, it is not only necessary to estimate their latent traits but also the population parameters of this group. This article has two main purposes: first, to develop a Monte Carlo Markov Chain algorithm to estimate both latent traits and population parameters concurrently. This algorithm comprises the Metropolis–Hastings within Gibbs sampling algorithm (MHWGS) proposed by Patz and Junker [*Journal of Educational and Behavioral Statistics* **24** (1999b) 346–366]. Second, to compare, in the latent trait recovering, the performance of this method with three other methods: maximum likelihood, expectation a posteriori and maximum a posteriori. The comparisons were performed by varying the total number of items (NI), the number of categories and the values of the mean and the variance of the latent trait distribution. The results showed that MHWGS outperforms the other methods concerning the latent traits estimation as well as it recoveries properly the population parameters. Furthermore, we found that NI accounts for the highest percentage of the variability in the accuracy of latent trait estimation.

1 Introduction

The nominal response model (NRM) was proposed by Bock (1972) in order to improve the latent trait estimation in nominal tests. Let us suppose that a test consisting of I items, with m_i categories each one, is administered to n examinees, and that a random variable $Y_{ijh}, i = 1, 2, \dots, I, j = 1, 2, \dots, n, h = 1, 2, \dots, m_i$, which indicates the category chosen by subject j to item i, by assuming value 1 for this category and 0 for the all remaining ones, is observed. The NRM, which

---

Key words and phrases. Nominal response model, latent trait, population parameters, MCMC simulation, Metropolis–Hastings.

Received September 2008; accepted March 2009.

represents the probability of such answer, is given by

$$\begin{aligned} P_{ijh} &= P(Y_{ijh} = 1|\theta_j, \boldsymbol{\zeta}_i) \\ &= \frac{\exp[a_{ih}(\theta_j - b_{ih})]}{\sum_{s=1}^{m_i} \exp[a_{is}(\theta_j - b_{is})]} = \frac{\exp(d_{ih} + a_{ih}\theta_j)}{\sum_{s=1}^{m_i} \exp(d_{is} + a_{is}\theta_j)}, \end{aligned} \tag{1.1}$$

where:

θ_j: latent trait of subject j,
$\boldsymbol{\zeta}_i$: $(\mathbf{d}_i', \mathbf{a}_i')' = (d_{i1}, \ldots, d_{im_i}, a_{i1}, \ldots, a_{im_i})'$,
a_{ih}: slope (discrimination) parameter of the category h of item i,
b_{ih}: difficulty parameter of the category h of item i,
$d_{ih} = -a_{ih}b_{ih}$: intercept parameter of the category h of item i.

Discussion about the interpretations of the model and the item parameters can be found, for example, in Bock (1972), De Ayala (1992), DeMars (2003), Azevedo (2003) and Baker and Kim (2004).

Most of the articles that deal with estimation in NRM are concerned with item parameter estimation, under different conditions. See De Ayala and Sava-Bolesta (1999), Bolt, Cohen and Wollack (2001), Wollack et al. (2002) and DeMars (2003), for instance. Latent traits estimation is discussed in the works of De Ayala (1989), De Ayala (1992) and Baker and Kim (2004). The latter discusses also the estimation of the population parameters via marginal maximum likelihood (MML), but not jointly with the latent traits; see also Bock and Aitkin (1981). The first two aforementioned articles use expectation a posteriori (EAP) method to estimate the latent traits, while Baker and Kim (2004) consider maximum likelihood (ML). The maximum a posteriori (MAP) method is implemented in the Multilog program [Thissen, Chen and Bock (2003)]. In both EAP and MAP methods it is assumed a standard normal distribution for the latent traits.

In this work we are concerning with the situation where the item parameters in the NRM are known in some metric [see, e.g., Andrade and Tavares (2005)], and we want to estimate the latent traits and the population parameters of a group of examinees, different from that one used to calibrate the item parameters. In this case, the population parameters are free to be estimated.

This paper has two goals: first, to develop a Metropolis–Hastings within Gibbs sampling algorithm to estimate jointly the latent traits and the population parameters. Second, to compare, in the latent trait recovering, the performance of this method with three others: ML, EAP and MAP. The comparisons are performed by varying the total number of items (NI), the number of categories (NC) and the values of the mean and the variance of the latent trait distribution. The NI and NC are known to have influence in the estimation accuracy, according to De Ayala and Sava-Bolesta (1999), Wollack et al. (2002) and DeMars (2003). Furthermore, we want to verify the impact of different values of the mean and variance of the latent trait distribution. We believe that the latent traits will be better estimated when one uses information about the population parameters.

In Section 2 we present the MCMC algorithm and the other methods. In Section 3 we perform a simulation study to compare the aforementioned methods and in Section 4 we outline some comments and conclusions.

2 MCMC estimation and the other methods

MCMC algorithms are powerful tools to make Bayesian inference; see Gamerman and Lopes (2006) for details. One of the most used algorithms of this class is the Gibbs sampling. This procedure calculates, empirically, joint posterior distributions through the so-called full conditional distributions. In many situations it is not possible to obtain these distributions analytically. This is the case of NRM. A way of avoiding such problem is to use some auxiliary algorithm as the Metropolis–Hastings or the adaptive rejection; see, for example, Patz and Junker (1999a) and Ghosh et al. (2000). Wollack et al. (2002) proposed a MCMC Gibbs sampling with adaptive rejection sampling algorithm to fit the NRM under standard normal latent distribution by using WinBugs package; see Lunn et al. (2000). On the other hand, Patz and Junker (1999a) developed a Metropolis–Hastings within Gibbs sampling algorithm for the one, two and three parameter logistic models and for the generalized partial credit model. We consider such approach for the NRM, henceforth MHWGS approach.

First, let us calculate the joint posterior distribution. Considering the usual assumptions of conditional independence, [see, e.g., Baker and Kim (2004)], the likelihood, for the latent trait of subject j, is given by

$$L(\theta_j|\mathbf{y}_{\cdot j \cdot}, \boldsymbol{\zeta}) = \prod_{i=1}^{I} \prod_{h=1}^{m_i} P_{ijh}^{y_{ijh}}, \tag{2.1}$$

where P_{ijh} is as described in (1.1), $\mathbf{y}_{\cdot j \cdot} = (y_{1j1}, \dots, y_{1jm_1}, \dots, y_{Ij1}, \dots, y_{Ijm_I})'$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1, \dots, \boldsymbol{\zeta}'_I)'$. For Bayesian inference we assume the following prior:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\eta}_\theta) &= p(\boldsymbol{\theta}|\boldsymbol{\eta}_\theta) p(\boldsymbol{\eta}_\theta) \\ &= \left\{ \prod_{j=1}^{n} p(\theta_j|\boldsymbol{\eta}_\theta) \right\} p(\mu_\theta) p(\psi_\theta), \end{aligned}$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and $\boldsymbol{\eta}_\theta = (\mu_\theta, \psi_\theta)$. The prior for θ_j is assumed to be

$$\theta_j|\boldsymbol{\eta}_\theta \sim N(\mu_\theta, \psi_\theta). \tag{2.2}$$

For the population parameters, natural choices, which lead to conditional conjugate families [see Gelman (2006)] are

$$\mu_\theta \sim N(\mu_\mu, \psi_\mu), \tag{2.3}$$

$$\psi_\theta \sim IG(\nu_0/2, \kappa_0/2). \tag{2.4}$$

Also, the prioris (2.3) and (2.4) are proper and this ensures that the posterior of $\boldsymbol{\eta}_\theta$ is also proper; see Gelman et al. (2004), for example. Therefore, from (2.1), (2.2), (2.3) and (2.4) it follows that the joint posterior distribution is given by

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\eta}_\theta | \mathbf{y}_{\cdots}) &\propto L(\boldsymbol{\theta} | \mathbf{y}_{\cdot j \cdot}, \boldsymbol{\zeta}) p(\boldsymbol{\theta}, \boldsymbol{\eta}_{\boldsymbol{\theta}}) \\
&\propto \left\{ \prod_{i=1}^{I} \prod_{j=1}^{n} \prod_{h=1}^{m_i} P_{ijh}^{y_{ijh}} \right\} \left\{ \prod_{j=1}^{n} \exp\left[-\frac{(\theta_j - \mu_\theta)^2}{2\psi_\theta} \right] \right\} \psi_\theta^{-n/2} \qquad (2.5) \\
&\quad \times \left\{ \exp\left[-\frac{(\mu_\theta - \mu_\mu)^2}{2\psi_\mu} \right] \right\} \left\{ \psi_\theta^{-(\nu_0/2+1)} \exp\left(-\frac{\kappa_0}{2\psi_\theta} \right) \right\}.
\end{aligned}
$$

The distribution (2.5) has an intractable form. Also, the full conditional distribution of the latent traits is not known. However, the full distributions of the population parameters are known and easy to sample from. Therefore, a hybrid MCMC algorithm can be used to simulate from (2.5). This algorithm is composed by a Metropolis–Hastings within Gibbs sampling step for the latent traits and two Gibbs sampling steps to estimate $\boldsymbol{\eta}_\theta$, which is exactly the MHWGS algorithm proposed by Patz and Junker (1999b). A kernel function [see Patz and Junker (1999a)] is necessary to draw from the latent traits. Following Patz and Junker (1999a) we choose

$$
q(\theta_j^{(t)}, \theta_j^{(t-1)}) \sim N(\theta_j^{(t-1)}, \sigma_\theta^2),
$$

where $\theta_j^{(t)}$ is the current simulated value of θ and $\theta_j^{(t-1)}$ is the simulated value of the former iteration. Denoting $(\cdot)$ the set of all other parameters, the hybrid MCMC algorithm (MHWGS) is defined as follows:

(1) Draw $\theta_j^{(t)}$ from $\theta_j|(\cdot) \sim p(\theta_j | \boldsymbol{\zeta}, \mathbf{y}_{\cdots})$, for $j = 1, \ldots, n$ mutually independent:

   (a) Draw $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \sigma_\theta^2)$.

   (b) Accept $\theta_j^{(t)} = \theta_j^{(*)}$ with probability:

$$
\begin{aligned}
&\pi_j(\theta_j^{(t-1)}, \theta_j^{(*)}) \\
&\quad = \min\left\{ \frac{L(\theta_j^{(*)} | \mathbf{y}_{\cdot j \cdot}, \boldsymbol{\zeta}) \exp\{-(\theta_j^{(*)} - \mu_\theta^{(t-1)})^2/(2\sigma_{\theta_j}^2)\}}{L(\theta_j^{(t-1)} | \mathbf{y}_{\cdot j \cdot}, \boldsymbol{\zeta}) \exp\{-(\theta_j^{(t-1)} - \mu_\theta^{(t-1)})^2/(2\sigma_{\theta_j}^2)\}}, 1 \right\},
\end{aligned}
$$

   otherwise, set $\theta_j^{(t)} = \theta_j^{(t-1)}$.

(2) Draw $\mu_\theta^{(t)}$ from $\mu_\theta|(\cdot) \sim N(\widehat{\psi}_\mu^{(t)} \widehat{\mu}_\theta^{(t)}, \widehat{\psi}_\mu^{(t)})$, where:

$$
\widehat{\mu}_\theta^{(t)} = \frac{1}{\psi_\theta^{(t-1)}} \sum_{j=1}^{n} (\theta_j^{(t)}) + \frac{\mu_\mu}{\psi_\mu}, \qquad \widehat{\psi}_\mu^{(t)} = \left(\frac{n}{\psi_\theta^{(t-1)}} + \frac{1}{\psi_\mu} \right)^{-1}.
$$

(3) Draw $\psi_\theta^{(t)}$ from $\psi_\theta|(\cdot) \sim IG(\widehat{\nu}^{(t)}, \widehat{\kappa}^{(t)})$, where:

$$\widehat{\nu}^{(t)} = \frac{1}{2}\left[\sum_{j=1}^{n}(\theta_j^{(t)} - \mu_\theta^{(t)})^2 + \nu_0\right],$$

$$\widehat{\kappa}^{(t)} = \frac{n+\kappa_0}{2}.$$

Giving suitable starting values $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\eta}_\theta^{(0)})$, the iteration of the three steps above comprises the MHWGS algorithm. Therefore, any quantity, as the mean and the variance of the posterior distributions can be calculated, based on the generated samples. These values can be used as estimates and measures of precision of such estimates.

The ML, for the NRM, is well described in the literature; see Baker and Kim (2004), for example. Therefore it will be not presented in this work. To describe the other two Bayesian methods we will consider, as usual, a standard normal distrubition as the prior, that is, $\boldsymbol{\eta}_\theta = (0, 1)$ in (2.2). Therefore, the posterior distribution of θ_j is given by

$$\begin{aligned} p(\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta}) &\propto L(\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})p(\theta_j|\boldsymbol{\eta}_{\boldsymbol{\theta}}) \equiv L(\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})p(\theta_j) \\ &\propto \left\{\prod_{i=1}^{I}\prod_{h=1}^{m_i} P_{ijh}^{y_{ijh}}\right\}\left\{\exp\left[-\frac{\theta_j^2}{2}\right]\right\}. \end{aligned} \tag{2.6}$$

The EAP estimates is the expectation of (2.7), that is,

$$\mathbf{E}[\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta}] = \frac{\int_{\mathbf{R}} \theta L(\theta|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})p(\theta)\,d\theta}{\int_{\mathbf{R}} L(\theta|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})p(\theta)\,d\theta}.$$

However, notice that it is not possible to obtain such expectation analytically. Numerical methods, as the Gaussian quadrature integration [see, e.g., Stroud and Secrest (1980)] should be employed in order to obtain the EAP estimates. Considering such approach, the EAP, in terms of Gaussian quadrature points, is given by

$$\mathbf{E}[\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta}] = \overline{\theta}_j \approx \frac{\sum_{q=1}^{Q} \overline{\theta}_q L(\overline{\theta}_q|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})A_q}{\sum_{q=1}^{Q} L(\overline{\theta}_q|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})A_q}, \tag{2.7}$$

where $\overline{\theta}_q, q = 1, \dots, Q$ are the quadrature points and A_l the associated quadrature weights. The accuracy of EAP can be evaluated through the variance a posteriori:

$$\mathrm{Var}[\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta}] = \frac{\int_{\mathbf{R}}\{\theta - \mathbf{E}[\theta_j|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta}]\}^2 L(\theta|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})p(\theta)\,d\theta}{\int_{\mathbf{R}} L(\theta|\mathbf{y}_{\cdot j\cdot}, \boldsymbol{\zeta})p(\theta)\,d\theta}. \tag{2.8}$$

Again, no explicit solutions are avaible and numerical methods should be used. Therefore, the equation (2.8), in terms of quadrature points, becomes

$$\mathrm{Var}[\theta_j|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta}] \approx \frac{\sum_{q=1}^{Q}(\overline{\theta}_q-\overline{\theta}_j)^2 L(\overline{\theta}_q|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta})A_q}{\sum_{q=1}^{Q} L(\overline{\theta}_q|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta})A_q},$$

where $\overline{\theta}_j$ is given by (2.7). The MAP is obtained by maximizing (2.7) with respect θ_j, which is equivalent to solve the following equation:

$$\begin{aligned} S(\theta_j) &= \frac{\partial \ln p(\theta_j|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta})}{\partial\theta_j} = \frac{\partial \ln L(\theta_j|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta})}{\partial\theta_j} + \frac{\partial \ln p(\theta_j)}{\partial\theta_j} \\ &= \sum_{i=1}^{I}\boldsymbol{\alpha}_i'\mathbf{T}_i[\mathbf{y}_{ij.}-\mathbf{P}_{ij.}]-\theta_j, \end{aligned} \tag{2.9}$$

where $\boldsymbol{\alpha}_i$ and $\mathbf{T}_i$ are appropriate matrices; see Baker and Kim (2004) or Azevedo (2003), for example. Notice that it is not possible to solve equation (2.9) analytically. Therefore, some numerical method should be employed; for example, Newton–Raphson or Fisher Scoring, see Rao (2002), for example. To accomplish for that, we need the Hessian function, which is given by

$$\begin{aligned} H(\theta_j) &= \frac{\partial^2 \ln p(\theta_j|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta})}{\partial\theta_j^2} = \frac{\partial^2 \ln L(\theta_j|\mathbf{y}_{\cdot j\cdot},\boldsymbol{\zeta})}{\partial\theta_j^2} + \frac{\partial^2 \ln p(\theta_j)}{\partial\theta_j^2} \\ &= -\sum_{i=1}^{I}\{\boldsymbol{\alpha}_i'\mathbf{T}_i\mathbf{W}_{ij}\mathbf{T}_i'\boldsymbol{\alpha}_i\}-1, \end{aligned} \tag{2.10}$$

where $\mathbf{W}_{ij}$ is a suitable matrix; see Baker and Kim (2004) or Azevedo (2003), for example. The first components of equations (2.9) and (2.10) require some algebra to be developed. The details can be found in Bock (1972), Baker and Kim (2004) or Azevedo (2003), for example. Furthermore, notice that the Fisher information is $I(\theta_j) = -H(\theta_j)$. Therefore, the Newton–Raphson and Fisher scoring algorithms are equivalent. Hence, given a suitable starting value, say $\widehat{\theta}_j^{(0)}$, the iterative process is given by

$$\widehat{\theta}_j^{(t)} = \widehat{\theta}_j^{(t-1)} + [I(\widehat{\theta}_j^{(t-1)})]^{-1}[S(\widehat{\theta}_j^{(t-1)})],$$

$t = 1, 2, \ldots$, until a convergence criterium be achieved.

The ML and MAP methods are both implemented in the commercial package Multilog; see http://www.ssicentral.com/. However, for practical reasons, all the methods were implemented in a program developed by the authors using the object-oriented statistical system Ox^{tm}; see http://www.doornik.com/products.html#Ox, and it is avaliable upon request from the authors. In the next section we will present a simulation study.

3 Simulation study

To compare the performance of the four estimation methods we conducted a Monte Carlo simulation study, according to Harwell et al. (1996). That is, considering replication of datasets and using suitable statistics to measure the goodness of the estimates. The number of replications was $R = 20$. This choice was based on previous works; see De Ayala and Sava-Bolesta (1999), DeMars (2003) and Wollack et al. (2002). The factors (and their levels) considered were: the number of items (NI) (20, 30, 40), the number of categories per item (NC) (4, 5), the values of the population means $(-2, 0, 2)$ and the values of the population variance $(0.5, 1.0, 1.5)$. Hence, there are $3 \times 2 \times 3 \times 3 = 54$ combinations. Once that our interest lies on the main factors (and not in the possible interactions) and also because the large time demanded by the MCMC methods, we considered a fractional factorial design using 18 of the 54 combinations. These combinations were generated in order to ensure the estimability of the main effects and they are presented in Table 1. In all simulations we set $n = 600$ examinees.

We considered a set of six tests varying in number of items: 20, 30, 40, and in number of categories (alternatives) per item: 4 and 5. Tables 2 and 3 present the items used to build the tests. The 20 items tests were built using the first 20 items and in the 40 item tests we considered all the items. The 30 items tests were built considering the first 21 and the items: 23, 25, 27, 29, 32, 34, 36, 38 and 40. This allows us to have tests in which the item parameteres range properly in terms of difficult and discrimination. That is, it is possible to cover the latent trait range and also to identify examinees with different abilities.

As hyperparameters of priors (2.3) and (2.4), we considered $(\mu_\mu = 0, \psi_\mu = 100)$ and $(\kappa_0 = 4.2, \nu_0 = 2.2)$. That is, the μ_θ prior is flat while the ψ_θ is moderately informative. For proposal variance we used $(\sigma_\theta^2 = 0.8^2)$. One of the most important aspects of the MCMC methods is to verify the convergence of the simulated values to the posterior densities of interest. There are several suggestions in the literature but no agreement about the most suitable one; see Gamerman and Lopes (2006), for example. In this work we considered: the monitoring of the chains generated by three different sets of starting values, trace plots and

Table 1 *Factor level combinations of the fractional factorial design*

NI	*NC*	μ_θ	ψ_θ	*NI*	*NC*	μ_θ	ψ_θ	*NI*	*NC*	μ_θ	ψ_θ
20	4	−2.0	1.5	30	4	−2.0	1.0	40	4	−2.0	0.5
20	4	2.0	0.5	30	4	2.0	1.5	40	4	2.0	1.0
20	4	0.0	1.0	30	4	0.0	0.5	40	4	0.0	1.5
20	5	−2.0	1.5	30	5	−2.0	1.0	40	5	−2.0	0.5
20	5	2.0	0.5	30	5	2.0	1.5	40	5	2.0	1.0
20	5	0.0	1.0	30	5	0.0	0.5	40	5	0.0	1.5

Table 2 *Item parameters used in the simulation studies: 4 alternatives tests*

Item	b	a	Item	b	a	Item	b	a	Item	b	a
1	−2.76	−0.50	11	0.57	−0.60	21	−2.22	−0.60	31	0.47	−0.70
1	−3.80	−0.20	11	−0.40	−0.70	21	−3.80	−0.20	31	−0.40	−0.70
1	−3.40	0.10	11	−0.20	0.40	21	−3.40	0.10	31	−0.20	0.40
1	−3.00	0.60	11	0.16	0.90	21	−2.50	0.70	31	0.13	1.00
2	−2.63	−1.20	12	0.41	−1.70	22	−2.07	−1.10	32	0.36	−1.60
2	−2.90	−0.40	12	−0.20	−0.30	22	−2.80	−0.40	32	−0.20	−0.30
2	−2.80	0.20	12	0.10	0.80	22	−2.40	0.20	32	0.10	0.80
2	−2.68	1.40	12	0.47	1.20	22	−2.24	1.30	32	0.39	1.10
3	−1.96	−0.60	13	0.97	−0.50	23	−1.44	−0.70	33	0.79	−0.60
3	−2.80	−0.70	13	0.10	−0.20	23	−2.80	−0.70	33	0.10	−0.20
3	−2.50	0.40	13	0.30	0.10	23	−2.50	0.40	33	0.30	0.10
3	−2.37	0.90	13	0.79	0.60	23	−1.97	1.00	33	0.66	0.70
4	−2.12	−1.70	14	1.30	−1.20	24	−1.89	−1.60	34	1.09	−1.10
4	−2.60	−0.30	14	0.40	−0.40	24	−2.60	−0.30	34	0.40	−0.40
4	−2.40	0.80	14	0.80	0.20	24	−2.40	0.80	34	0.80	0.20
4	−2.05	1.20	14	1.11	1.40	24	−1.71	1.10	34	0.92	1.30
5	−1.59	−0.50	15	1.86	−0.60	25	−1.26	−0.60	35	1.46	−0.70
5	−2.20	−0.20	15	0.80	−0.70	25	−2.20	−0.20	35	0.80	−0.70
5	−1.90	0.10	15	1.00	0.40	25	−1.90	0.10	35	1.00	0.40
5	−1.74	0.60	15	1.42	0.90	25	−1.44	0.70	35	1.18	1.00
6	−1.32	−1.20	16	1.62	−1.70	26	−1.03	−1.10	36	1.41	−1.60
6	−1.80	−0.40	16	1.00	−0.30	26	−1.80	−0.40	36	1.00	−0.30
6	−1.60	0.20	16	1.20	0.80	26	−1.60	0.20	36	1.20	0.80
6	−1.42	1.40	16	1.74	1.20	26	−1.18	1.30	36	1.45	1.10
7	−0.73	−0.60	17	2.30	−0.50	27	−0.51	−0.70	37	1.86	−0.60
7	−1.60	−0.70	17	1.20	−0.20	27	−1.60	−0.70	37	1.20	−0.20
7	−1.40	0.40	17	1.60	0.10	27	−1.40	0.40	37	1.60	0.10
7	−1.11	0.90	17	2.05	0.60	27	−0.92	1.00	37	1.71	0.70
8	−0.80	−1.70	18	2.57	−1.20	28	−0.71	−1.60	38	2.18	−1.10
8	−1.30	−0.30	18	1.60	−0.40	28	−1.30	−0.30	38	1.20	−0.40
8	−1.00	0.80	18	2.00	0.20	28	−1.00	0.80	38	1.60	0.20
8	−0.79	1.20	18	2.37	1.40	28	−0.66	1.10	38	1.97	1.30
9	−0.32	−0.50	19	3.15	−0.60	29	−0.26	−0.60	39	2.64	−0.70
9	−1.00	−0.20	19	2.00	−0.70	29	−1.00	−0.20	39	1.70	−0.70
9	−0.80	0.10	19	2.20	0.40	29	−0.80	0.10	39	2.00	0.40
9	−0.47	0.60	19	2.68	0.90	29	−0.39	0.70	39	2.24	1.00
10	0.01	−1.20	20	2.99	−1.70	30	0.06	−1.10	40	2.41	−1.60
10	−0.80	−0.40	20	2.00	−0.30	30	−0.80	−0.40	40	1.90	−0.30
10	−0.40	0.20	20	2.60	0.80	30	−0.40	0.20	40	2.10	0.80
10	−0.16	1.40	20	3.00	1.20	30	−0.13	1.30	40	2.50	1.10

Geweke statistics. All these procedures were implemented in the object-oriented statistical system Ox^{tm}. In the first starting values set, the latent traits were drawn from a N(0, 1) distribution and μ_θ and ψ_θ were fixed equal to 0 and 1, respectively. In the second set, the latent traits were all fixed to 0, μ_θ was drawn from N(0, 1) and ψ_θ was drawn from U(0, 2). Finally, in the third set, the standardized

Table 3 *Item parameters used in the simulation studies: 5 alternatives tests*

Item	b	a	Item	b	a	Item	b	a	Item	b	a
1	−2.50	−0.40	11	0.32	−0.90	21	−1.94	−0.50	31	0.57	−0.90
1	−3.90	−0.20	11	−0.70	−0.20	21	−3.80	−0.20	31	−0.80	−0.70
1	−3.60	−0.10	11	−0.40	−0.20	21	−3.60	−0.10	31	−0.50	0.20
1	−3.40	0.10	11	−0.20	0.40	21	−3.40	0.10	31	−0.20	0.40
1	−3.00	0.60	11	0.16	0.90	21	−2.50	0.70	31	0.13	1.00
2	−2.40	−0.80	12	0.64	−1.20	22	−1.74	−1.00	32	0.63	−1.00
2	−3.00	−0.50	12	−0.20	−0.60	22	−3.00	−0.60	32	−0.20	−0.60
2	−2.70	−0.30	12	0.00	−0.20	22	−2.70	−0.30	32	0.00	−0.30
2	−2.40	0.20	12	0.10	0.80	22	−2.40	0.60	32	0.10	0.80
2	−2.68	1.40	12	0.47	1.20	22	−2.24	1.30	32	0.39	1.10
3	−2.19	−0.90	13	1.36	−0.40	23	−1.59	−0.90	33	1.06	−0.50
3	−3.00	−0.20	13	−0.20	−0.20	23	−3.00	−0.70	33	−0.20	−0.20
3	−2.80	−0.20	13	0.00	−0.10	23	−2.80	0.20	33	0.00	−0.10
3	−2.50	0.40	13	0.30	0.10	23	−2.50	0.40	33	0.30	0.10
3	−2.37	0.90	13	0.79	0.60	23	−1.97	1.00	33	0.66	0.70
4	−1.85	−1.20	14	1.93	−0.80	24	−1.34	−1.00	34	1.26	−1.00
4	−2.77	−0.60	14	0.10	−0.50	24	−2.80	−0.60	34	0.40	−0.60
4	−2.50	−0.20	14	0.40	−0.30	24	−2.60	−0.30	34	0.60	−0.30
4	−2.40	0.80	14	0.80	0.20	24	−2.40	0.80	34	0.80	0.60
4	−2.05	1.20	14	1.11	1.40	24	−1.71	1.10	34	0.92	1.30
5	−1.49	−0.40	15	1.55	−0.90	25	−1.06	−0.50	35	1.66	−0.90
5	−2.20	−0.20	15	0.60	−0.20	25	−2.30	−0.20	35	0.30	−0.70
5	−2.00	−0.10	15	0.80	−0.20	25	−2.10	−0.10	35	0.60	0.20
5	−1.90	0.10	15	1.00	0.40	25	−1.90	0.10	35	1.00	0.40
5	−1.74	0.60	15	1.42	0.90	25	−1.44	0.70	35	1.18	1.00
6	−1.02	−0.80	16	2.02	−1.20	26	−0.75	−1.00	36	1.70	−1.00
6	−1.90	−0.50	16	0.70	−0.60	26	−2.00	−0.60	36	0.90	−0.60
6	−1.80	−0.30	16	1.00	−0.20	26	−1.80	−0.30	36	1.00	−0.30
6	−1.60	0.20	16	1.20	0.80	26	−1.60	0.60	36	1.20	0.80
6	−1.42	1.40	16	1.74	1.20	26	−1.18	1.30	36	1.44	1.10
7	−0.98	−0.90	17	2.58	−0.40	27	−0.49	−0.90	37	2.01	−0.50
7	−1.80	−0.20	17	1.10	−0.20	27	−2.00	−0.70	37	1.10	−0.20
7	−1.60	−0.20	17	1.40	−0.10	27	−1.80	0.20	37	1.30	−0.10
7	−1.40	0.40	17	1.60	0.10	27	−1.40	0.40	37	1.60	0.10
7	−1.11	0.90	17	2.05	0.60	27	−0.92	1.00	37	1.71	0.70
8	−0.62	−1.20	18	2.97	−0.80	28	−0.24	−1.00	38	2.62	−1.00
8	−1.30	−0.60	18	1.60	−0.50	28	−1.50	−0.60	38	1.20	−0.60
8	−1.10	−0.20	18	1.80	−0.30	28	−1.30	−0.30	38	1.40	−0.30
8	−1.00	0.80	18	2.00	0.20	28	−1.00	0.80	38	2.00	0.60
8	−0.79	1.20	18	2.37	1.40	28	−0.66	1.10	38	1.97	1.30
9	0.02	−0.40	19	2.81	−0.90	29	−0.07	−0.50	39	2.84	−0.90
9	−1.30	−0.20	19	1.80	−0.20	29	−1.10	−0.20	39	1.20	−0.70
9	−1.10	−0.10	19	2.00	−0.20	29	−1.00	−0.10	39	1.80	0.20
9	−0.80	0.10	19	2.20	0.40	29	−0.80	0.10	39	2.00	0.40
9	−0.47	0.60	19	2.68	0.90	29	−0.39	0.70	39	2.24	1.00
10	0.35	−0.80	20	3.35	−1.20	30	0.58	−1.00	40	2.88	−1.00
10	−0.80	−0.50	20	2.00	−0.60	30	−1.20	−0.60	40	1.60	−0.60
10	−0.60	−0.30	20	2.30	−0.20	30	−0.90	−0.30	40	1.70	−0.30
10	−0.40	0.20	20	2.60	0.80	30	−0.40	0.60	40	2.00	0.80
10	−0.16	1.40	20	3.00	1.20	30	−0.13	1.30	40	2.50	1.10

scores were used as the latent traits starting values. For μ_θ and ψ_θ we calculated the sample mean and variance of the standardized scores plus values generated from a N(0, 1) distribution. We simulated a set of responses for the combination $(NI = 20, NC = 4, \mu_\theta = -2, \psi_\theta = 1.5)$ and applied the convergence assessment procedures aforementioned. The Geweke statistics were calculated considering independent samples of size 100 after 2000 iterations. They showed that the convergence occurred for all considered parameters (the population parameters and some latent traits randomly chosen). Also, the observation of trace plots indicated that a burn-in of 2000 is enough to draw from the posterior densities. The Figure 1 shows the autocorrelations and trace plots of the generated chains for one latent trait (randomly chosen), the mean and the variance obtained from different starting values. It shows that the autocorrelations become negligible after a lag of 20. Also, the initial states of the chains do not affect the convergence. Therefore we decided to consider a burn-in of 2000, simulating more 18,000 values after that and retaining every 20 values. Hence, we had 900 values to estimate the posterior densities.

To compare the performance of the estimation methods we considered the following statistics for each latent trait:

Corr: mean of the correlation between $\overline{\theta}_j$ and θ_j among the examinees.
Bias: $\frac{1}{n}\sum_{j=1}^{n}(\theta_j - \widehat{\overline{\theta}}_j)$.
Var: $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{R}\sum_{r=1}^{R}(\widehat{\theta}_{jr} - \widehat{\overline{\theta}}_j)^2$.

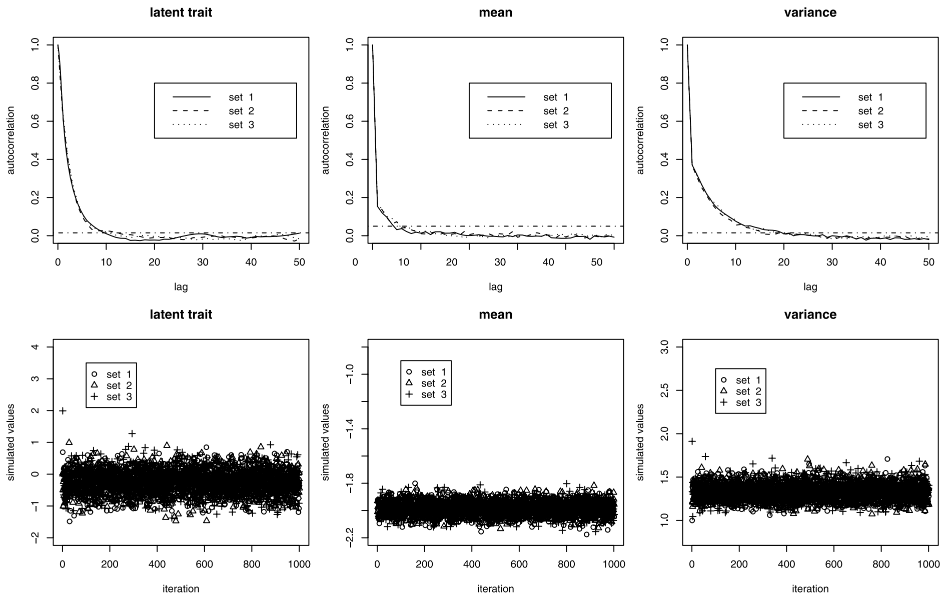

Figure 1 *Autocorrelations and trace plots for chains generated from different starting values.*

MSE: $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{R}\sum_{r=1}^{R}(\theta_j - \widehat{\theta}_{jr})^2$.
RMSE: the square root of MSE.

Where $\widehat{\theta}_{jr}$ is the estimate of the latent trait of examinee j obtained from the dataset r and $\overline{\widehat{\theta}}_j = \frac{1}{R}\sum_{j=1}^{R}\widehat{\theta}_{jr}$.

Table 4 presents the aforementioned statistics for each estimation method in each factors level combination. The results indicate that the MCMC performs equally or better than the other methods. When the true population parameters match with that considered in the usual methods, the four estimation procedures produce quite similar results. However, when this is not true, the MCMC produce more accurate results. This is even more clear when the true population parameters are far from (0, 1). Hence, we concluded that estimating the population parameters properly, improves the latent traits estimates. Also, the MHWGS approach presented the most accurate results.

Figures 2, 3 and 4 present the MSE per latent trait ranges for different combinations of level factors. One can see that, even though the MHWGS was the most accurate estimation procedure, the MSE of the estimates of the other methods are smaller than that of the MHWGS ones, for some latent trait ranges. Therefore the

Table 4 *Statistics for the latent trait estimation*

NI	*NC*	μ_θ	ψ_θ	E.M.	Corr	Var	Bias	RMSE
20	4	−2.0	1.5	MV	0.988	0.571	−0.134	0.813
				MAP	0.980	0.101	0.405	0.648
				EAP	0.981	0.105	0.388	0.633
				MHWGS	0.989	0.159	−0.010	0.462
20	4	2.0	0.5	MV	0.987	0.240	0.041	0.509
				MAP	0.990	0.110	−0.322	0.491
				EAP	0.990	0.113	−0.312	0.486
				MHWGS	0.990	0.092	0.003	0.369
20	4	0.0	1.0	MV	0.995	0.164	−0.001	0.417
				MAP	0.995	0.113	0.000	0.379
				EAP	0.995	0.114	−0.001	0.379
				MHWGS	0.995	0.115	−0.001	0.379
20	5	−2.0	1.5	MV	0.980	1.416	−0.113	1.234
				MAP	0.977	0.113	0.474	0.729
				EAP	0.978	0.116	0.455	0.712
				MHWGS	0.987	0.177	0.004	0.502
20	5	2.0	0.5	MV	0.987	0.233	0.031	0.499
				MAP	0.987	0.124	−0.353	0.531
				EAP	0.986	0.126	−0.345	0.526
				MHWGS	0.987	0.098	0.003	0.391
20	5	0.0	1.0	MV	0.994	0.182	−0.003	0.440
				MAP	0.995	0.118	0.003	0.395
				EAP	0.995	0.120	0.001	0.395
				MHWGS	0.995	0.120	0.001	0.395

Table 4 *Continued.*

NI	*NC*	μ_θ	ψ_θ	E.M.	Corr	Var	Bias	RMSE
30	4	−2.0	1.0	MV	0.991	0.260	−0.065	0.539
				MAP	0.987	0.086	0.304	0.499
				EAP	0.987	0.088	0.287	0.487
				MHWGS	0.991	0.108	0.002	0.380
30	4	2.0	1.5	MV	0.995	0.241	0.052	0.517
				MAP	0.983	0.073	−0.284	0.511
				EAP	0.983	0.074	−0.273	0.502
				MHWGS	0.992	0.104	−0.013	0.376
30	4	0.0	0.5	MV	0.995	0.100	−0.008	0.324
				MAP	0.995	0.080	−0.002	0.299
				EAP	0.995	0.081	−0.006	0.300
				MHWGS	0.995	0.066	−0.002	0.295
30	5	−2.0	1.0	MV	0.986	0.519	−0.059	0.747
				MAP	0.988	0.098	0.346	0.538
				EAP	0.988	0.101	0.327	0.525
				MHWGS	0.992	0.127	−0.001	0.404
30	5	2.0	1.5	MV	0.995	0.206	0.045	0.476
				MAP	0.990	0.084	−0.279	0.490
				EAP	0.989	0.084	−0.273	0.487
				MHWGS	0.995	0.112	0.004	0.371
30	5	0.0	0.5	MV	0.993	0.121	−0.008	0.358
				MAP	0.993	0.093	−0.001	0.326
				EAP	0.993	0.094	−0.006	0.326
				MHWGS	0.993	0.075	−0.001	0.319
40	4	−2.0	0.5	MV	0.994	0.130	−0.035	0.374
				MAP	0.991	0.070	0.193	0.355
				EAP	0.991	0.072	0.179	0.348
				MHWGS	0.991	0.066	0.000	0.298
40	4	2.0	1.0	MV	0.995	0.155	0.032	0.412
				MAP	0.990	0.065	−0.212	0.393
				EAP	0.990	0.066	−0.202	0.387
				MHWGS	0.994	0.078	−0.008	0.316
40	4	0.0	1.5	MV	0.999	0.082	0.000	0.295
				MAP	0.998	0.063	0.003	0.285
				EAP	0.998	0.064	0.002	0.284
				MHWGS	0.998	0.069	0.001	0.280
40	5	−2.0	0.5	MV	0.992	0.146	−0.031	0.396
				MAP	0.990	0.079	0.232	0.395
				EAP	0.990	0.082	0.215	0.386
				MHWGS	0.991	0.075	0.003	0.320
40	5	2.0	1.0	MV	0.997	0.118	0.024	0.354
				MAP	0.995	0.072	−0.210	0.379
				EAP	0.994	0.072	−0.202	0.375
				MHWGS	0.996	0.083	0.005	0.312
40	5	0.0	1.5	MV	0.998	0.095	−0.005	0.316
				MAP	0.997	0.072	0.001	0.311
				EAP	0.997	0.073	−0.001	0.310
				MHWGS	0.998	0.079	−0.002	0.304

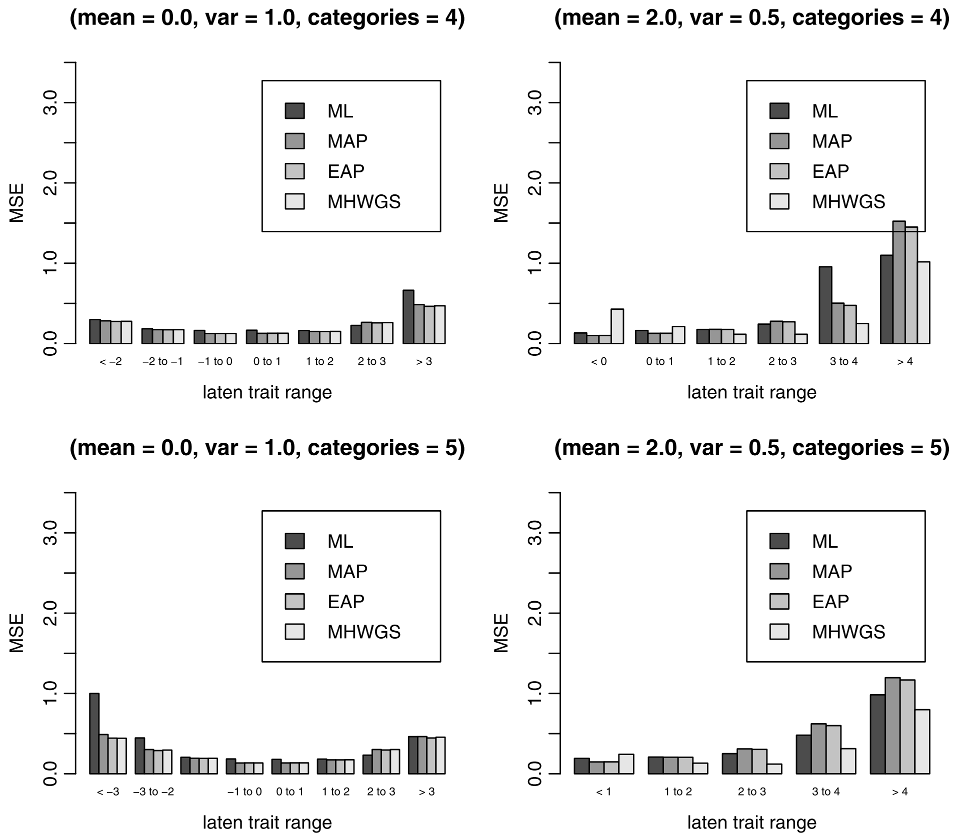

Figure 2 *MSE per latent trait range for some 20 items tests.*

MHWGS is not uniformly better than the other three methods (concerning the latent trait values). Also, we can notice that the highest MSE occurs for the more extreme latent trait values. Furthermore, when the population parameters are far from (0, 1) values, the MSE tend to be higher. Anyway, in general, one can say that MHWGS algorithm outperforms the other methods.

An ANOVA was calculated for the fractional factorial design by considering as the response variable, ln(RMSE); see De Ayala and Sava-Bolesta (1999) and DeMars (2003). Since that there are many observations (number of examinees) by each combination of factor levels, any statistical test would be significant. Also, some usual ANOVA assumptions as: normality and homogeneity cannot hold in considering as the response, ln(RMSE). Therefore, the statistic ω^2 was considered instead of the F one, without performing any statistical test. The ω^2 statistic, for a specific factor, is given by

$$\omega^2 = \frac{SS_{\text{factor}} - df_{\text{factor}} MS_{\text{error}}}{MS_{\text{error}} + SS_{\text{total}}},$$

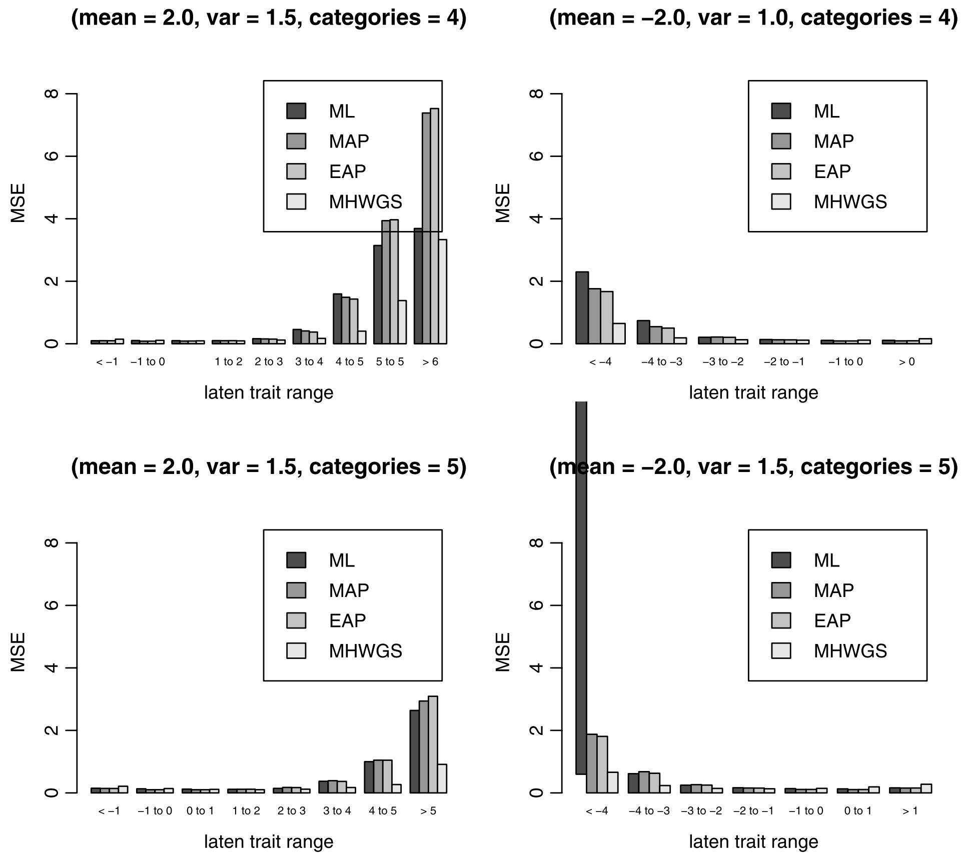

Figure 3 *MSE per latent trait range for some 30 items tests.*

where $SS_.$ is the sum of squares, $MS_.$ is the mean square and $df_.$ are the degrees of freedom. This allows us to evaluate the contribution of each main effect in the difference of the accuracy of the estimates; see De Ayala and Sava-Bolesta (1999) and DeMars (2003), for example. Table 5 presents such results. We see that NI accounts for the highest percentage of the variability of the ln(RMSE), while the other factors (NC, mean and variance) have a small impact. The remaining effects, including possible interactions and other factors which could be considered as the sample size ratio [see DeMars (2003)] have, all together, a reasonable influence (they are into the error term). In the design that was considered it is not possible to identify the contribution of each one.

By inspecting Table 6, one can see that the population parameters were all well recovered by the MHWGS approach. Table 7 presents the number of iterations necessary to obtain the convergence in the iterative process for MV and MAP as well as the precision achieved. Also, the time spent by the estimation process is

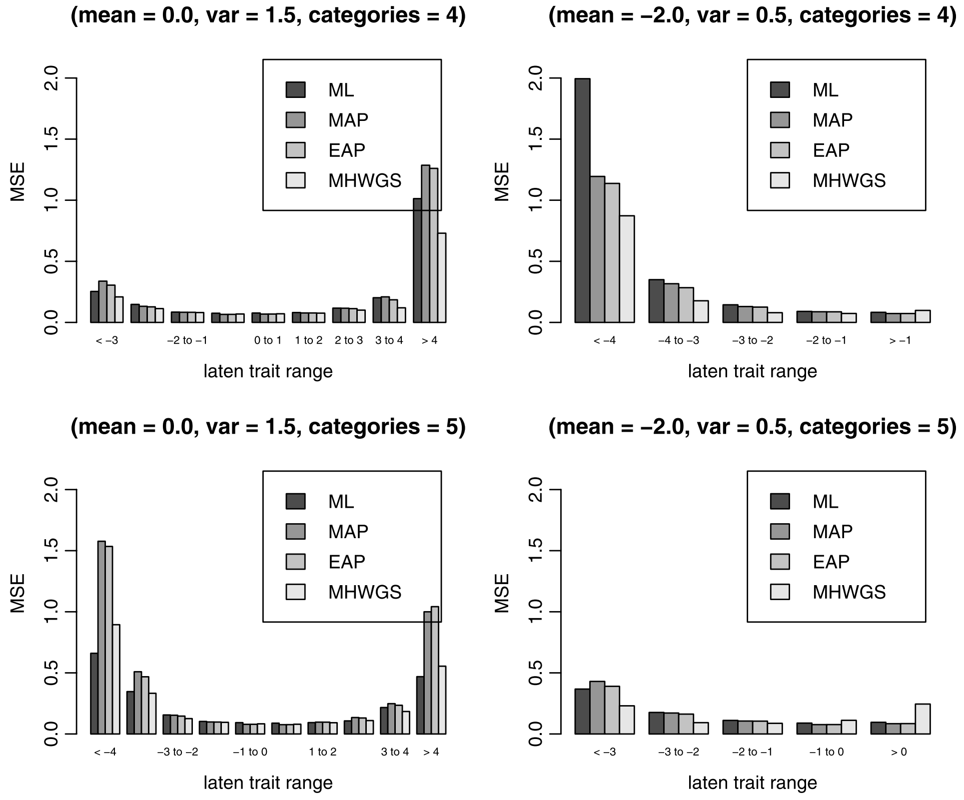

Figure 4 *MSE per latent trait range for some 40 items tests.*

Table 5 *Anova for the latent trait estimation*

Source of variation	*SS*	*df*	*MS*	ω^2
NI	161.64	2	80.82	0.216
NC	9.50	1	9.50	0.013
μ_θ	41.37	2	20.68	0.055
ψ_θ	24.70	2	12.35	0.033
Error	509.08	10792	0.05	–
Total	746.28	–	–	–

displayed for the two former methods and MHWGS algorithm. It is clear that MAP requires less iterations and less time of computation. Furthermore, we can see that the MHWGS algorithm requires much more time than the other methods.

Table 6 *Results of the population parameter estimation for MHWGS algorithm*

NI	*NC*	μ_θ	ψ_θ	Statistic	est. μ_θ	est. ψ_θ
20	4	−2.0	1.5	mean	−2.010	1.516
				meanse	0.055	0.110
				var	< 0.001	0.003
20	4	2.0	0.5	mean	2.003	0.506
				meanse	0.034	0.042
				var	< 0.001	0.001
20	4	0.0	1.0	mean	0.000	0.998
				meanse	0.044	0.069
				var	< 0.001	0.001
20	5	−2.0	1.5	mean	−1.996	1.481
				meanse	0.055	0.111
				var	< 0.001	0.003
20	5	2.0	0.5	mean	2.003	0.503
				meanse	0.035	0.043
				var	< 0.001	0.001
20	5	0.0	1.0	mean	0.001	0.987
				meanse	0.044	0.070
				var	< 0.001	0.001
30	4	−2.0	1.0	mean	−1.998	0.992
				meanse	0.044	0.071
				var	< 0.001	0.001
30	4	2.0	1.5	mean	1.987	1.460
				meanse	0.052	0.100
				var	< 0.001	0.002
30	4	0.0	0.5	mean	−0.002	0.496
				meanse	0.032	0.035
				var	< 0.001	0.001
30	5	−2.0	1.0	mean	−2.001	1.026
				meanse	0.045	0.075
				var	< 0.001	0.001
30	5	2.0	1.5	mean	2.004	1.505
				meanse	0.052	0.100
				var	< 0.001	0.001
30	5	0.0	0.5	mean	−0.001	0.502
				meanse	0.033	0.037
				var	< 0.001	0.000

4 Concluding remarks

The proposed MHWGS algorithm recovers both latent traits and population parameters properly. Its computational implementation is straightforward and can be extended to other situations such as: unknown item parameters and different latent traits distributions. It is clear that estimating properly the population parameters,

Table 6 *Continued.*

NI	*NC*	μ_θ	ψ_θ	Statistic	est. μ_θ	est. ψ_θ
40	4	−2.0	0.5	mean	−2.000	0.502
				meanse	0.032	0.037
				var	< 0.001	< 0.001
40	4	2.0	1.0	mean	1.992	0.980
				meanse	0.043	0.066
				var	< 0.001	0.001
40	4	0.0	1.5	mean	0.001	1.487
				meanse	0.051	0.093
				var	< 0.001	0.001
40	5	−2.0	0.5	mean	−1.997	0.510
				meanse	0.033	0.039
				var	< 0.001	0.001
40	5	2.0	1.0	mean	2.005	1.019
				meanse	0.043	0.068
				var	< 0.001	0.001
40	5	0.0	1.5	mean	−0.002	1.472
				meanse	0.051	0.093
				var	< 0.001	0.001

mean = mean of the estimates, meanse = mean of the standard errors, var = variance of the estimates.

Table 7 *Time spent in seconds (TS), precision achieved (PA) and number of required iterations (NRI) in the simulation study*

				MV			MAP			MCMC
Items	Cat	μ_θ	ψ_θ	*PA*	*NRI*	*ST*	*PA*	*NRI*	*ST*	*ST*
20	4	−2.0	1.5	0.000009	52.00	6.65	0.000009	41.10	5.30	1539.69
20	4	2.0	0.5	0.000607	97.55	10.43	0.000009	48.15	5.22	1316.58
20	4	0.0	1.0	0.000009	52.00	6.65	0.000009	41.10	5.30	1539.69
20	5	−2.0	1.5	0.000007	27.20	4.43	0.000008	29.40	4.76	1810.29
20	5	2.0	0.5	0.003033	100.00	20.24	0.000010	70.25	14.27	2392.73
20	5	0.0	1.0	0.000008	25.30	3.56	0.000008	31.65	4.38	1653.41
30	4	−2.0	1.0	0.000332	89.20	15.30	0.000008	31.45	5.55	2095.55
30	4	2.0	1.5	0.000122	99.90	29.92	0.000010	76.95	23.11	3511.32
30	4	0.0	0.5	0.000008	23.50	4.87	0.000008	36.65	7.47	2396.07
30	5	−2.0	1.0	0.000008	24.80	6.49	0.000008	30.65	7.73	3028.68
30	5	2.0	1.5	0.000063	74.70	9.01	0.000009	36.95	4.52	1475.86
30	5	0.0	0.5	0.000007	19.90	5.05	0.000008	31.10	7.74	2848.61
40	4	−2.0	0.5	0.000031	71.15	19.39	0.000009	32.10	8.90	3221.85
40	4	2.0	1.0	0.001557	100.00	27.32	0.000009	80.80	22.21	3276.54
40	4	0.0	1.5	0.000025	69.55	22.28	0.000009	54.40	17.29	3708.44
40	5	−2.0	0.5	0.000007	20.60	8.84	0.000008	26.25	11.16	4719.00
40	5	2.0	1.0	0.000009	76.85	28.34	0.000009	59.10	21.91	4282.20
40	5	0.0	1.5	0.000009	49.55	20.81	0.000009	42.30	18.08	4807.37

better latent traits estimates are obtained. Furthermore, we notice that the number of items accounts for the highest percentage in the variability of the accuracy in the latent traits estimation. The other factors: number of categories and the values of the population parameters, account for a small percentage of this variability. We need to point out that, the EAP and MAP procedure may be improved, if one use suitable estimates of population parameters, as those that can be obtained by using the MML approach; see Baker and Kim (2004). As far we know, there are no programs which allow to consider such approach for the NRM.

Acknowledgments

This research was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) from Brazil, through a Master's program scholarship. It is part of the Master's Dissertation written by the first author under the guidance of the second.

References

Andrade, D. F. and Tavares, H. R. (2005). Item response theory for longitudinal data: Population parameter estimation. *Journal of Multivariate Analysis* **95** 1–22. MR2164119

Azevedo, C. L. N. (2003). Métodos de estimação na Teoria da Resposta ao Item. Master's Dissertation, University of São Paulo, in Portuguese.

Baker, F. B. and Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques*, 2nd ed. Marcel Dekker, New York. MR2086862

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 and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46** 443–459. MR0668311

Bolt, D. M., Cohen, A. S. and Wollack, J. A. (2001). A mixture item response model for multiple choice data. *Journal of Educational and Behavorial Statistics* **26** 381–409.

De Ayala, R. J. (1989). A comparison of the nominal response model and three-parameter logistic model in computerized adaptive testing. *Educational and Psychological Measurement* **49** 789–805.

De Ayala, R. J. (1992). The nominal response model in computerized adaptive testing. *Applied Psychological Measurement* **16** 327–343.

De Ayala, R. J. and Sava-Bolesta, M. (1999). Item Parameter Recovery for the Nominal Response Model. *Applied Psychological Measurement* **23** 3–19.

DeMars, C. E. (2003). Sample Size and the Recovery of Nominal Response Model Item Parameters. *Applied psychological measurement* **27** 275–288. MR1977220

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London. MR2260716

Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis* **1** 515–533. MR2221284

Gelman, A. and Carlin, J. B. and Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman and Hall, London. MR2027492

Ghosh, M., Ghosh, A., Chen, M.-H. and Agresti, A. (2000). Noninformative priors for one-parameter item response models. *Journal of Statistical Planning and Inference* **88** 99–115. MR1767562

Harwell, M., Stone, C. A., Hsu, T.-C. and Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement* **20** 101–125.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10** 325–337.

Patz, R. J. and Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics* **24** 146–178.

Patz, R. J. and Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: multiple item types, missing data and rated responses. *Journal of Educational and Behavioral Statistics* **24** 346–366.

Rao, C. R. (2002). *Linear Statistical Inference and Its Applications*. Wiley, New York.

Stroud, A. H. and Secrest, D. (1980). *Gaussian Quadrature Formulas*. Prentice-Hall, Englewood, NJ. MR0202312

Thissen, D., Chen, W. H. and Bock, D. (2003). In *IRT from SSI* (M. Du Toiti, ed.). Scientific Software International, Chicago, IL.

Wollack, J. A., Bolt, D. M., Cohen, A. S. and Lee, Y.-S. (2002). Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and Markov chain Monte Carlo estimation. *Applied Psychological Measurement* **26** 339–352. MR1917876

University of Campinas
Cidade Universitária “Zeferino Vaz”
Distr. Barão Geraldo
Campinas
São Paulo 13083-859
Brasil
E-mail: cnaber@ime.unicamp.br

Departamento de Informática e Estatística
Universidade Federal de Santa Catarina
Florianopólis/SC, 57072–970
Brazil
E-mail: dandrade@inf.ufsc.br
URL: http://www.inf.ufsc.br/~dandrade