Marginalized Maximum a Posteriori Estimation for the 4-Parameter Logistic Model under a Mixture Modeling Framework

1st Xiangbin Meng[1*], 2nd Gongjun Xu[2] , 3rd Jiwei Zhang[1], 4th Jian Tao[1]

[1] 1st School of Mathematics and Statistics, KLAS, Northeast Normal University, Changchun, Jilin, China.

[2] 2nd Department of Statistics, University of Michigan, 1085 South University, Ann Arbor, MI, 48109.

*Corresponding author information: Xiangbin Meng, Northeast Normal University, 5268 Renmin Street, Changchun, Jilin Province, China,(e-mail:mengxb600@nenu.edu.cn).

Abstract:

The 4-parameter logistic model (4PLM) has recently gained great interests in various applications. Motivated by recent studies that re-express the 4-parameter model to be a mixture model with two levels of latent variables, this paper develops a new Expectation-Maximization (EM) algorithm for marginalized maximum a posterior (MMAP) estimation of the 4PLM parameters. The mixture modelling framework of the 4PLM not only makes the proposed EM algorithm more easily to be implemented in practice, but also provides a natural connection with the popular cognitive diagnosis models. Simulation studies were constructed to show the good performance of the proposed estimation method and to investigate the impact of the additional upper asymptote parameter on the estimation of other parameters. Moreover, a real data set was analyzed by the 4PLM to show its outperformance over the 3-parameter logistic model (3PLM).

Keywords:

4-parameter logistic model, marginalized maximum a posteriori estimation, Expectation-Maximization algorithm, mixture model.

Acknowledgements:

The authors are greatly indebted to the editor and two anonymous reviewers for their valuable comments and suggestions. This research was supported by the National Natural Science Foundation of China (11571069), the Jilin Province Science and Technology Department(201705200054JH), the Fundamental Research Funds for the Central Universities, and National Science Foundation (1659328, 1712717).

Marginalized Maximum a Posteriori Estimation for the 4-Parameter Logistic Model under a Mixture Modeling Framework

Abstract

The 4-parameter logistic model (4PLM) has recently gained great interests in various applications. Motivated by recent studies that reexpress the 4-parameter model to be a mixture model with two levels of latent variables, this paper develops a new Expectation-Maximization (EM) algorithm for marginalized maximum a posterior (MMAP) estimation of the 4PLM parameters. The mixture modeling framework of the 4PLM not only makes the proposed EM algorithm more easily to be implemented in practice, but also provides a natural connection with the popular cognitive diagnosis models. Simulation studies were constructed to show the good performance of the proposed estimation method and to investigate the impact of the additional upper asymptote parameter on the estimation of other parameters. Moreover, a real data set was analyzed by the 4PLM to show its outperformance over the 3-parameter logistic model (3PLM).

Key words: 4-parameter logistic model, marginalized maximum a posteriori estimation, Expectation-Maximization algorithm, mixture model.

1 Introduction

The 4PLM was proposed by Barton and Lord (1981), who introduced an upper asymptote parameter, d, that is slightly less than 1, to model the uncertainty of a high-ability examinee missing an easy item. The limitation of Barton and Lord's modeling approach is that all items in a test share a common upper asymptote parameter, and they did not estimate the fourth parameter but rather fitted the model with some fixed values for d. Recent studies (Rouse et al, 1999; Linacre, 2004; Rupp, 2003; Tavares et al., 2004; Waller & Reise, 2010) demonstrated that, in most cases, the upper asymptote varies across items in a test. The formulation of the 4PLM that allows the upper asymptote parameter to be item-specific is therefore considered more appropriate, which is,

$$p_j\left(\theta_i\right) = P\left(U_{ij} = 1\left|\theta_i, \xi_j\right.\right) = c_j + \left(d_j - c_j\right)\frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}. \tag{1}$$

where U_{ij} denotes the observed dichotomous response of examinee i $(i = 1, ..., N)$ to item j $(j = 1, ..., M)$ with $U_{ij} = 1$ denoting the correct response and $U_{ij} = 0$ otherwise; $\theta_i \in (-\infty, +\infty)$ is the ability parameter; $\xi_j = \{a_j, b_j, c_j, d_j\}$ is the item parameter set for the jth item with $a_j \in (0, +\infty)$, $b_j \in (-\infty, +\infty)$, $c_j \in [0, 1]$, and $d_j \in (c_j, 1]$ being the discrimination, difficulty, guessing, and upper asymptote parameters, respectively. The parameter d_j is the maximum probability of endorsing item j, then $1 - d_j$ can be considered as the slipping probability of a student who can answer correctly but miss the item. Here, N and M are used to denote the number of the examinees (sample size) and the number of the items (test length).

The difficulties in parameter estimation and a lack of evidence supporting the need likely result in that the 4PLM was not widely applied for a long time (Loken & Rulison, 2010). In recent years, researchers are showing renewed interest in the 4PLM. For instance, Liao et al. (2012) and Rulison and Loken (2009) argued that the 4PLM can improve the accuracy of ability estimation by taking into account examinees' early careless errors in CAT. Reise and Waller (2003) and Waller and Reise (2010) demonstrated that the item response model including an upper asymptote parameter may be more appropriate for measuring psychopathology traits than the 3PLM or 2PLM, since the situation of a high-trait subject who is reluctant to self-report attitudes is very common in psychopathology measurement. Ogasawara (2012) gave the asymptotic distribution of the ability estimation under the 4PLM, and Magis (2013) derived the maximum value of the information function. Furthermore, several methods on the estimation of the parameters in the 4-parameter model have been proposed. For instance, Loken and Rulison (2010) employed a Bayesian approach with the Markov Chain Monte Carlo (MCMC) sampler to estimate the 4PLM parameters. Feuerstahler and Waller (2014) employed the marginal maximum likelihood (MML) method to recover the 4PLM using the R package "mirt". In comparison to the Bayesian estimation method calculated with the MCMC sampler algorithm, the MML method requires shorter computation time, but it may not be stable and the deviant values may be produced in many cases (Baker & Kim, 2004). To overcome this disadvantage of the MML estimation, Mislevy (1986) proposed the Bayesian modal (BM) estimation for the 3PLM. The BM estimation can be considered as a MMAP estimation, it employs an augmented optimization objective that includes the likelihood and some prior beliefs for the item parameters, these priors were used to prevent deviant parameter estimates from occurring. In fact, the BM estimation can be seen as a regulation of the MML estimation, while the MML estimation is a special case of the BM estimation that assume the uniform prior distributions of parameters. Waller and Feuerstahler (2017) recently applied the BM estimation as implemented in the R package "mirt" for the 4PLM.

In addition to the above researches on estimating the 4PLM, mixture modeling approaches have been developed by introducing latent variables to deal with the response process. For instance, Béguin and Glas (2001), San Martin, del Pino and DeBoeck (2006), and von Davier (2009) interpreted the 3PLM from a two response strategies, guessing and non-guessing, by revising the 3PLM to be a mixture model. Recently, Culpepper (2016, 2017) further developed a mixture modeling approach to reformulate the 4-parameter normal ogive model (4PNOM) and multidimensional 4PNOM. To estimate the model parameters, the existing works mostly focused on the Bayesian estimation with MCMC sampling procedure and computationally may be time consuming, especially for large data sets. Motivated by the mixture modeling specification in these researches, this paper proposes a computationally efficient EM algorithm to compute the MMAP estimates of the 4PLM parameters.

The rest of the article is organized as follows. Section 2 reviews the mixture modeling reformulation of the 4PLM and discuss the relationship between the 4PLM and cognitive diagnosis model (CDM). Section 3 presents the derivations of the EM algorithm for the MMAP estimation of the 4PLM under the mixture modeling framework. Section 4 reports three simulation studies that were constructed to evaluate the performance of the proposed method. Section 5 presents an application of the 4PLM to an empirical dataset. Finally, we provide further discussions on some future research directions in Section 6.

2 An alternative expression of the 4PLM from the two response processes: guessing versus slipping

From Equation (1), the probability of a correct response in the 4PLM is equivalent to,

$$P(U_{ij} = 1|\theta_i, \xi_j) = c_j \times \left(1 - p_j^*(\theta_i)\right) + d_j \times p_j^*(\theta_i), \tag{2}$$

where

$$p_j^*(\theta_i) = \frac{\exp\left[a_j(\theta_i - b_j)\right]}{1 + \exp\left[a_j(\theta_i - b_j)\right]}, \tag{3}$$

is the 2-parameter Logistic model (2PLM).

Following the mixture framework of conceptualizing the process of ability-based responding and guessing behaviors for 3PLM in von Davier (2009) and the study of 4PNOM in Culpepper (2016), we present an alternative expression of the 4PLM using a mixture model. Specifically, we introduce an unobserved latent variable $W_{ij} \in \{0,1\}$ to characterize the two random response status of an examinee: $W = 1$ indicates the examinee is "capable" to answer the item based on his/her ability and $W = 0$ otherwise. Following the 4PLM representation in (2) and (3), we let W_{ij} follow a Bernoulli distribution

$$W_{ij} \mid \theta_i, \xi_j \sim \text{Bernoulli}(p_j^*(\theta_i)), \tag{4}$$

where $p_j^*(\theta_i)$ is specified in (3), indicating that a higher ability θ_i leads to a higher chance of having $W_{ij} = 1$. When $W_{ij} = 1$, the conditional probability of the response U_{ij} is specified as

$$U_{ij} \mid W_{ij} = 1, \xi_j \sim \text{Bernoulli}(d_j), \tag{5}$$

where $1 - d_j$ corresponds to the slipping probability of making an mistake though the examinee is "capable" of answering item j. On the other hand, when $W_{ij} = 0$, that is the ith examinee does not know the correct answer of the jth item, the conditional distribution of U_{ij} is,

$$U_{ij} \mid W_{ij} = 0, \xi_j \sim \text{Bernoulli}(c_j), \tag{6}$$

where c_j is the guessing probability of a correct response.

We next show that the mixture model specification in (4)–(6) is equivalent to the 4PLM given in (2). Based on the above distributions in (4)–(6), the joint probability distribution of U_{ij} and W_{ij} (conditionally on θ_i and ξ_j) can be given as,

$$\begin{aligned} p_{(U_{ij},W_{ij})}(u_{ij}, w_{ij} \mid \theta_i, \xi_j) &= p_{U_{ij}|W_{ij},\theta_i,\xi_j}(u_{ij}|w_{ij}) p_{W_{ij}|\theta_i,\xi_j}(w_{ij} \mid \theta_i, \xi_j) \\ &= d_j^{w_{ij}u_{ij}}(1-d_j)^{w_{ij}(1-u_{ij})} c_j^{(1-w_{ij})u_{ij}}(1-c_j)^{(1-w_{ij})(1-u_{ij})} \\ &\quad \times p_j^*(\theta_i)^{w_{ij}} \left[1 - p_j^*(\theta_i)\right]^{1-w_{ij}}. \end{aligned} \tag{7}$$

Hence, the marginal probability distribution of U_{ij} over W_{ij} can be given by,

$$\begin{aligned} p_{U_{ij}}(u_{ij} \mid \theta_i, \xi_j) &= \sum_{w_{ij}=1,0} p_{(U_{ij},W_{ij})}(u_{ij}, w_{ij} \mid \theta_i, \xi_j) \\ &= d_j^{u_{ij}}(1-d_j)^{(1-u_{ij})} p_j^*(\theta_i) + c_j^{u_{ij}}(1-c_j)^{(1-u_{ij})}(1-p_j^*(\theta_i)), \end{aligned} \tag{8}$$

which is a two-class mixture Bernoulli distribution. From Equation (8), we have the marginal probability of $U_{ij} = 1$,

$$p_{U_{ij}}(u_{ij} = 1 \mid \theta_i, \xi_j) = p_j^*(\theta_i) \cdot d_j + (1 - p_j^*(\theta_i)) \cdot c_j, \tag{9}$$

which is the same as the 4PLM given in (2).

The above derivations demonstrate that the 4PLM can be considered as a two-strategies mixture model. What's more, the mixture model framework offers a new sight to understand the 4PLM and naturally connects it with the cognitive diagnosis models as shown in Remark 1.

Remark 1 (Connection to CDMs) *From the cognitive diagnosis models (CDMs) literature,* W_{ij} *can also be interpreted as the ideal response variable, where* $W_{ij} = 1$ *indicates* i*th examinee is capable to answer item* j *and* $W_{ij} = 0$ *otherwise. Then the distribution of* U_{ij} *specified in* (5) *and* (6) *is the same as the DINA model specification, where* c_j *here corresponds to the guessing parameter and* $1 - d_j$ *corresponds to the slipping parameter. Moreover, we show that the 4PLM can also be viewed as a generalization of the Higher-Order DINA model (de la Torre and Douglas, 2004) with only one latent attribute. In particular, consider a cognitive diagnosis test with only one latent attribute* $A \in \{0, 1\}$*. Then the Q-matrix is* $J \times 1$ *and we set* $Q = (1, \ldots, 1)'_{J\times 1}$*, that is, all items require the attribute* A*. Note that in this special case, the ideal responses of an examinee to all items are all the same. Let* A_i *be the* i*th examinee's latent attribute and the common ideal responses to all items are* $I(A_i = 1) = A_i$*. The Higher-Order DINA model assumes the* *probability of* $A_i = 1$ *is from a 2PLM that*

$$P(A_i = 1 \mid \theta_i, \lambda) = \frac{\exp\left[\lambda_0(\theta_i - \lambda_1)\right]}{1 + \exp\left[\lambda_0(\theta_i - \lambda_1)\right]}. \tag{10}$$

where θ_i *denotes a latent variable representing general ability in the studied domain and* λ*'s are regression parameters. Furthermore, given* $I(A_i = 1) = A_i$*, the* i*th examinee's response* U_{ij} *to the* j*th item follows the same models in* (5) *and* (6) *under the Higher-Order DINA model. Therefore, the only difference between the 4PLM and the one-attribute Higher-Order DINA model lies in how they model the ideal responses (*W_{ij} *and* A_i*, respectively). Comparing the model setup of the ideal responses between the Higher-Order DINA model in* (10) *and the 4PLM in* (2)*, we can see that* (10) *can be considered as a special case of* (2) *with all* a_j*'s replaced by a common parameter* λ_0*,* b_j*'s replaced by* λ_1*, and* W_{ij} *replaced by a common variable* A_i *not depending on* j*. From this perspective, the the one-attribute Higher-Order DINA model can be viewed as a special case of the 4PLM. More generally, we may consider the multi-attribute Higher-Order DINA model as a sub-model of the multi-dimentional 4PLM.*

3 The MMAP estimation for the 4PLM with an EM algorithm

Under the mixture model framework, we develop an EM algorithm for calculating the MMAP estimation for the item parameters in the 4PLM. In the following, we first specify the prior distributions on the 4PLM parameters and then derive the formula of the EM algorithm to calculate the MMAP estimators of the 4PLM item parameters.

We first introduce some notations. Let $\boldsymbol{u}_{i\cdot} = (u_{i1}, ..., u_{iM})$ denote the observed response vector of examinee i, $\boldsymbol{u}_{\cdot j} = (u_{1j}, ..., u_{Nj})'$ denote the observed response vector of item j, and $\boldsymbol{u} = (\boldsymbol{u}_{\cdot 1}, ..., \boldsymbol{u}_{\cdot M})$ denote the realized response matrix. Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_N)$ be the ability parameter vector of all N examinees, $\xi_j = (a_j, b_j, c_j, d_j)$ be the item parameter vector of item j, and $\boldsymbol{\xi} = (\xi_1, ..., \xi_M)$ be all the item parameters of all M items.

The prior distribution for the ability variable, θ_i, is specified to be a normal distribution, $\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$. This is the standard choice in calculating the MML or MMAP estimates of the parameters in IRT models. For the discrimination parameter a_j, we firstly transform $a_j = e^{\alpha_j}$, then a norm prior is assigned for α_j, $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$. The prior for b_j is a norm distribution, $b_j \sim N(\mu_b, \sigma_b^2)$. The prior for c_j is a Beta prior, $c_j \sim Beta(s_c, t_c)$. These prior distributions are commonly used in the applications of the IRT models. Finally, we assign a truncated Beta prior for d_j, $d_j \mid c_j \sim Beta(s_d, t_d) I(c_j < d_j)$, since $d_j > c_j$. Such truncated prior has been used in Culpepper (2016) to enforce the monotonicity condition. Here $\Omega := \{\mu_\theta, \sigma_\theta^2, \mu_\alpha, \sigma_\alpha^2, s_c, t_c, s_d, t_d\}$ are hyper-parameters to be prespecified in practice.

According to Bayes theorem, the joint posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ is, $p(\boldsymbol{\xi}, \boldsymbol{\theta} \mid \boldsymbol{u}, \Omega, \boldsymbol{\tau}) \propto L(\boldsymbol{u} \mid \boldsymbol{\xi}, \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \boldsymbol{\tau}) f(\boldsymbol{\xi} \mid \Omega)$, where

$$L(\boldsymbol{u} \mid \boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^{N} \prod_{j=1}^{M} p_j(\theta_i)^{u_{ij}} (1 - p_j(\theta_i))^{1-u_{ij}}$$

is the likelihood of the observed response data $\boldsymbol{u}$, and

$$f(\boldsymbol{\theta} \mid \boldsymbol{\tau}) = \prod_{i=1}^{N} f(\theta_i|\tau), \quad f(\boldsymbol{\xi} \mid \Omega) = \prod_{j=1}^{N} f(\xi_j \mid \Omega),$$

are the prior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, respectively.

As known in the literatures (Neyman & Scott, 1948; Baker & Kim, 2004), direct jointly estimation of the persons' ability parameters θ_i's and item parameters often lead to inconstant estimators, therefore θ_i's are generally needed to be integrated over to estimate the item parameters. Then we have the corresponding marginal distribution as

$$p(\boldsymbol{\xi} \mid \boldsymbol{u}, \Omega, \boldsymbol{\tau}) = \int p(\boldsymbol{\xi}, \boldsymbol{\theta} \mid \boldsymbol{u}, \Omega, \boldsymbol{\tau}) \, d\boldsymbol{\theta}, \tag{11}$$

and the modes of the marginal posterior $p(\boldsymbol{\xi}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau})$,

$$\hat{\boldsymbol{\xi}} = \underset{\boldsymbol{\xi}\in\Theta_{\boldsymbol{\xi}}}{\arg\max} p(\boldsymbol{\xi}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}), \tag{12}$$

are defined as the MMAP estimates of $\boldsymbol{\xi}$.

From Equation (7), if the latent variables $\boldsymbol{W} = \{W_{ij}, i = 1, ..., N; j = 1, ..., M\}$ were observed, the 4PLM could be divided into two Bernoulli models, and the calculation of the estimators of $\boldsymbol{\xi}$ would be straightforward. Specifically, let $\boldsymbol{z} = (\boldsymbol{u}, \boldsymbol{W}, \boldsymbol{\theta})$ be the the complete data. The likelihood of $\boldsymbol{z}$ is

$$\begin{aligned} L(\boldsymbol{z} \mid \boldsymbol{\xi}) &= \prod_{i=1}^{N}\prod_{j=1}^{M} d_j^{W_{ij}u_{ij}}(1-d_j)^{W_{ij}(1-u_{ij})} c_j^{(1-W_{ij})u_{ij}}(1-c_j)^{(1-W_{ij})(1-u_{ij})} \\ &\quad \times p_j^*(\theta_i)^{W_{ij}}(1-p_j^*(\theta_i))^{1-W_{ij}} f(\theta_i\,|\boldsymbol{\tau}). \end{aligned} \tag{13}$$

The marginal posterior distribution $p(\boldsymbol{\xi}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau})$ in (11) can be calculated by,

$$p(\boldsymbol{\xi}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}) = \int\int p(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau})\, d\boldsymbol{W} d\boldsymbol{\theta},$$

where

$$p(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}) \propto L(\boldsymbol{z}\,|\boldsymbol{\xi})\, f(\boldsymbol{\xi}\,|\Omega). \tag{14}$$

With the unobserved $\boldsymbol{W}$ in practice, we propose an EM interaction procedures under the complete data ($\boldsymbol{z}$) for calculating the MMAP estimators of $\boldsymbol{\xi}$ in Equation (12). Let $\boldsymbol{\xi}^{(t)}$ be the current values for $\boldsymbol{\xi}$ at the tth iteration, and the EM algorithm performs the following two steps:

E-step: Given $\boldsymbol{\xi}^{(t)}$ and $\boldsymbol{u}$ calculate the conditional distribution of the latent variables $\boldsymbol{W}$ and $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{W}, \boldsymbol{\theta}|\boldsymbol{u}, \boldsymbol{\xi}^{(t)})$, and then use $p(\boldsymbol{W}, \boldsymbol{\theta}|\boldsymbol{u}, \boldsymbol{\xi}^{(t)})$ to calculate the corresponding expectation of $\ln p(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau})$, i.e.,

$$Q\left(\boldsymbol{\xi}, \boldsymbol{\xi}^{(t)}\right) = E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^{(t)}}\left\{\ln p(\boldsymbol{z}, \boldsymbol{\xi}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau})\right\}. \tag{15}$$

M-step: Update the parameter estimate $\boldsymbol{\xi}^{(t+1)}$ by maximizing $Q(\boldsymbol{\xi}, \boldsymbol{\xi}^{(t)})$, i.e.,

$$\boldsymbol{\xi}^{(t+1)} = \arg\max Q\left(\boldsymbol{\xi}, \boldsymbol{\xi}^{(t)}\right).$$

We next describe the details in the E- and M-steps. From Equations (13) and (14),

$$\begin{aligned} \ln p\left(\boldsymbol{\xi}, \boldsymbol{z} \,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}\right) &= \ln L(\boldsymbol{z} \mid \boldsymbol{\xi}) + \sum_{j=1}^{M} \ln f(\xi_j \mid \Omega) \\ &= L_1(c, d) + L_2(\alpha, b) + \sum_{i=1}^{N} \ln f(\theta_i|\tau) + \sum_{j=1}^{M} \ln f(\xi_j \mid \Omega), \end{aligned} \quad (16)$$

where

$$\begin{aligned} L_1(c, d) &= \sum_{i=1}^{N}\sum_{j=1}^{M}\{W_{ij}u_{ij}\ln d_j + W_{ij}(1-u_{ij})\ln(1-d_j) + (1-W_{ij})u_{ij}\ln c_j \\ &\quad +(1-W_{ij})(1-u_{ij})\ln(1-c_j)\}, \\ L_2(\alpha, b) &= \sum_{i=1}^{N}\sum_{j=1}^{M} W_{ij}\ln p_j^*(\theta_i) + (1-W_{ij})\ln(1-p_j^*(\theta_i)). \end{aligned}$$

From Equation (16), we note that the estimators of (c_j, d_j) and (α_j, b_j) can be calculated separately with respect to $L_1(c, d)$ and $L_2(\alpha, b)$ in E- and M-steps. Since $L_1(c, d)$ is a linear function of W_{ij}, the E-step is done by simply replace W_{ij} with $E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^{(t)}}(W_{ij})$. In M-step, the estimators of c_j and d_j can then be calculated as

$$\begin{aligned} c_j^{(t+1)} &= \frac{\sum_{i=1}^{N}\left(1 - E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^{(t)}}(W_{ij})\right)u_{ij} + s_c - 1}{\sum_{i=1}^{N}\left(1 - E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^{(t)}}(W_{ij})\right) + s_c + t_c - 2}, \\ d_j^{(t+1)} &= d_j^* \times \mathbf{I}(d_j^* > c_j^{(t+1)}) + (c_j + \delta) \times \left[1 - \mathbf{I}(d_j^* > c_j^{(t+1)}))\right], \end{aligned} \quad (17)$$

where

$$d_j^* = \frac{\sum_{i=1}^{N}\left(E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^{(t)}}(W_{ij})\right)u_{ij} + s_d - 1}{\sum_{i=1}^{N}\left(E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^{(t)}}(W_{ij})\right) + s_d + t_d - 2}, \quad (18)$$

and $\mathbf{I}(d_j^* > c_j^{(t+1)})$ is the indicative function of $d_j^* > c_j^{(t+1)}$. Note that to impose the restriction that $d_j > c_j$, $d_j^{(t+1)}$ is assigned to be $c_j^{(t+1)} + \delta$ for a small $\delta > 0$ when $d_j^* \le c_j^{(t+1)}$.

Based on Equations (7) and (8), we have,

$$E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left[W_{ij}\right] = \int \left[\frac{d_j \cdot p_j^*(\theta_i)}{p_j(\theta_i)}\right]^{u_{ij}} \left[\frac{(1-d_j)\cdot p_j^*(\theta_i)}{1-p_j(\theta_i)}\right]^{1-u_{ij}} p\left(\theta_i \middle| \boldsymbol{u}_i, \boldsymbol{\xi}^{(t)}\right) d\theta_i,$$

where $p_j^*(\cdot)$ is defined in (3). A quadrature approximation method is used to compute the integrals in the E-step. In particular, define a grid of K equally-spaced points, $x_k\,(k = 1, ..., K)$, specified for θ, and the associated weights $A\,(x_k)$ is assigned by $f\,(x_k\,|\boldsymbol{\tau}\,)\times (x_{k+1} - x_k)$. The posterior probability of x_k can be given by,

$$p\left(x_k \middle| \boldsymbol{u}_i, \boldsymbol{\xi}^{(t)}\right) \cong \frac{\prod_{j=1}^{M} p_j^{(t)}\,(x_k)^{u_{ij}}\, q_j^{(t)}\,(x_k)^{1-u_{ij}}\, A\,(x_k)}{\sum_{k=1}^{K}\prod_{j=1}^{M} p_j^{(t)}\,(x_k)^{u_{ij}}\, q_j^{(t)}\,(x_k)^{1-u_{ij}}\, A\,(x_k)}, \tag{19}$$

where

$$p_j^{(t)}(x_k) = c_j^{(t)} - (d_j^{(t)} - c_j^{(t)})\frac{\exp\,(e^{\alpha_j^{(t)}}(x_k - b_j^{(k)}))}{1+\exp\,(e^{\alpha_j^{(t)}}(x_k - b_j^{(k)}))}$$

and $q_j^{(t)}(x_k) = 1 - p_j^{(t)}(x_k)$. Then $E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left[W_{ij}\right]$ can be approximately calculated by,

$$E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left[W_{ij}\right] \cong \sum_{k=1}^{K}\left[\frac{d_j^{(t)}\cdot p_j^{*(t)}(x_k)}{p_j^{(t)}(x_k)}\right]^{u_{ij}} \left[\frac{(1-d_j^{(t)})\cdot p_j^{*(t)}(x_k)}{1-p_j^{(t)}(x_k)}\right]^{1-u_{ij}} p(x_k \mid \boldsymbol{u}_i, \boldsymbol{\xi}_j^{(t)}),$$

where $i = 1, ..., N$, $j = 1, ..., M$. Finally, plug them into the Equations (17) and (18), the revise estimators, $c_j^{(t+1)}$ and $d_j^{(t+1)}$, can be approximately calculated.

In the M-step, the estimation equations for α_j and b_j can be approximated by

$$\frac{\partial E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left(\ln p\left(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}\,\right)\right)}{\partial \alpha_j} \cong \sum_{k=1}^{K}(x_k - b_j)(\hat{N}(x_k) - \hat{R}(x_k)p_j^*(x_k)) - \frac{\alpha_j - \mu_\alpha}{\sigma_\alpha} = 0, \tag{20}$$

$$\frac{\partial E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left(\ln p\left(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}\,\right)\right)}{\partial b_j} \cong -e^{e^{\alpha_j}}\sum_{k=1}^{K}(\hat{N}(x_k) - \hat{R}(x_k)p_j^*(x_k)) - \frac{b_j - \mu_b}{\sigma_b} = 0, \tag{21}$$

where

$$\hat{N}(x_k) = \sum_{i=1}^{N}\left[\frac{d_j^{(t)}\cdot p_j^{*(t)}(x_k)}{p_j^{(t)}(x_k)}\right]^{u_{ij}} \left[\frac{(1-d_j^{(t)})\cdot p_j^{*(t)}(x_k)}{1-p_j^{(t)}(x_k)}\right]^{1-u_{ij}} p(x_k \mid \boldsymbol{u}_i, \boldsymbol{\xi}_j^{(t)}),$$

$$\hat{R}(x_k) = \sum_{i=1}^{N} p\left(x_k \middle| \boldsymbol{u}_i, \boldsymbol{\xi}^{(t)}\right),$$

and $p(x_k|\boldsymbol{u}_i, \boldsymbol{\xi}^{(t)})$ is calculated as in (19). To solve the non-linear equations (20) and (21), a Newton-Raphson iterate algorithm is used and the detailed calculation procedure and the corresponding MATLAB code are presented in the Appendix.

4 Monte Carlo Simulation

This section reports three simulation studies to display the performance of the proposed MMAP estimation. Specifically, the first simulation study was to investigate the influences of the prior distributions on the performance of the MMAP estimation; The second simulation was constructed to study the relationship between the d parameter and the properties of MMAP estimation; The third simulation was performed to compare the performances of the proposed MMAP\EM method with the existing BM estimation procedure implemented in R package "mirt" (Waller & Feuerstahler, 2017).

4.1 Simulation Study 1

In this simulation, the test length was $M = 20$ and the true values of a_j, b_j and c_j $(j = 1, \ldots, M)$ were randomly drawn from a large scale achievement test that was analyzed in Wang, Chang, and Douglas (2013). Following a similar setup to that of Loken and Rulison (2010), the parameters $d_j\,(j = 1, \ldots, M)$ were randomly generated from a truncated Beta distribution, $d_j \sim Beta\,(8, 2)$ with the constrain of $d_j > c_j$. The true values of these item parameters are shown in the left four columns of Table 2. The examinees' ability variables, θ_i $(i = 1, \ldots, N)$, were randomly drawn from the standard normal distribution, $\theta_i \sim N\,(0, 1)$. As the sample size is an important data characteristic determining the properties of the item parameter estimation, we generated response data with three sample sizes of $N = \{1000, 5000, 10000\}$.

To investigate the influence of the prior distributions of parameters a, b, c, and d, the MMAP estimation was implemented under three groups of priors, please refer to Table 1. Specifically, among the three groups of priors, those in the first line (denoted as MMAP1) provide the strongest prior information. The distributions shown in the third line (denoted as MMAP3) are the weakest informative priors, where $Beta(1,1)$ is the uniform distribution on $[0,1]$, and $N(0,10^2)$ is close to non-informative prior. That is, the MMAP estimators calculated under this group of priors can be considered as an approximation of the MML estimators. The prior distributions shown in the middle line (denoted as MMAP2) are weaker than the MMAP1 but stronger than the MMAP3.

To reduce the Monte Carlo error, 500 replications of the response data sets were randomly generated, and the MMAP estimates were calculated for each of the 500 data sets. The number of quadrature points in the MMAP estimation was set to be 20, and both the convergence criterions for the EM algorithm and the N-R iterations were specified to be 0.001. Finally, the root mean squared error (RMSE) and mean error(ME) were calculated across the 500 replications to evaluate the accuracy and bias of the MMAP estimators. The RMSE is defined as

$$\mathrm{RMSE}\left(\delta_j\right) = \sqrt{G^{-1}\sum\nolimits_{g=1}^{G}\left(\hat{\delta}_{gj}-\delta_j\right)^2}, \tag{22}$$

and the ME is defined as,

$$\mathrm{ME}\left(\delta_j\right) = G^{-1}\sum\nolimits_{g=1}^{G}\left(\hat{\delta}_{gj}-\delta_j\right), \tag{23}$$

where δ_j is the item parameter (any α_j, b_j, c_j, d_j) of interest, $\hat{\delta}_{gj}$ denotes the estimate of δ_j in the g-th repetition, and G is the number of replications ($G = 500$ in this study).

In this simulation, there was not any deviant parameter estimate or any unsuccessful iteration, even in the case of the weakly informative priors given in MMAP3. We consider that the proposed estimation method based on the mixture model interpretation is helpful for improving the convergence rate of the EM algorithm. Furthermore, the implementation of the EM procedure was generally fast. For instance, the average calculation time did not exceed 0.8, 2.5 and 10.0 seconds under the three sample sizes $N = \{1000, 5000, 10000\}$. (The PC information: Intel Core i5-8200 CPU(1.6 GHz)),RAM(8G)). Tables 2–4 show the obtained values of RMSEs for the MMAP estimators with the three prior specifications (MMAP1, MMAP2 and MMAP3) across the three sample sizes (N =1000, 5000 and 10000). Observing these results, the following trends can be observed.

1. Under the sample size of $N = 1000$, there are slight differences in the values of RMSE of the MMAP estimators under the three groups of priors (MMAP1, MMAP2, and MMAP3). Overall, the MMAP3 estimators displayed larger values of RMSE than that of the MMAP1 and MMAP2 estimators. However, as the sample size increased, the differences in the RMSE of the three estimators become much smaller. For instance, under the sample sizes of $N = 5000$ and 10000, the differences in RMSE of the three MMAP estimators were negligible for most item parameters. The same phenomenon was observed on the values of ME (the values of ME are not reported here due to space limitation). This suggests that when the number of examinees is large, the MMAP estimators are mainly determined by the response data and the specification of the prior distributions is not less crucial. On the other hand, when the sample size is small, the prior information will have a larger impact on the performance of the MMAP estimation, so to avoid the subjective error from the misspecification of prior distributions, weakly informative or non-informative priors may be recommended in practice. Additionally, we also calculated the BM estimates of the 4PLM by implementing the "mirt" package. The results showed that the BM estimators with informative priors perform similar to our method, while the BM with the non-informative priors not only displayed lower accuracy but also suffered from unsuccessful convergences frequently. It can be considered that the mixture strategies framework of the 4PLM is helpful for the convergence of the EM algorithm. The results of BM estimation were not reported here as they

are not the main focus of this simulation study and more comparisons between our method and the BM estimation are provided in Simulation Study 3.

2. It can be observed that the RMSE(d) of the items $j = \{4, 7, 8, 12, 19\}$ are much larger than those of the other items. The common characters of these items are: their a-parameters were much lower than the other items, as well as the b- and d-parameters were relative larger. This phenomenon was also observed in Culpepper(2016). Inspired by the research of Lord (1975) and Mislevy (1986), which verified under the 3PLM that the estimation accuracy of c_j and $b_j - 2/a_j$ are positively correlated, we may explain this phenomenon by a negative correlation between the estimation accuracy of d_j and the value of $b_j + 2/a_j$ under the 4PLM. Heuristically, a larger value of $b_j + 2/a_j$ implies fewer examinees satisfying $a_j(\theta_i - b_j) > 2$, and therefore less information on d_j is provided by the responses, which then reduces the estimation accuracy of d_j. The scatter plots with the Pearson correlation coefficients were created to display the influence of $b_j + 2/a_j$ on the estimation errors and biases of the MMAP estimators of d, seeing Figure 1. It can be found that across the three sample sizes, both the RMSE(d) and absolute ME(d) were positively correlated with $b_j + 2/a_j$, and the correlations increase with the sample size. These results demonstrated that the higher the difficult and the lower the discrimination, the poorer the estimation accuracy for the d parameter in terms of both mean squared error and bias.

4.2 Simulation Study 2

The main purpose of this simulation is to investigate the impact of the d parameter on the performance of the MMAP estimation. An artificial test with 4 levels of d, $d \in \{0.65, 0.75, 0.85, 0.95\}$, were constructed, where each d-level included 5 items and the test length was $M = 20$. To produce a controlled experiment, the values of a, b and c were identical for all items with $a = 1.0, b = 0.0$, and $c = 0.2$. Following to the simulation study 1, the sample sizes were set to be $N = \{1000, 5000, 10000\}$, and the examinees' ability parameter θ's were randomly drawn from $N(0, 1)$. Additionally, 500 response data sets were randomly generated, and the MMAP estimate were calculated with the three groups of priors in Table 1. Finally, the RMSE and ME of the MMAP estimates were calculated to display the properties (efficiency and bias) of the estimator. Because the trends on the MMAP estimators with the three groups of priors were consistent, we only report the results under the priors of MMAP1 here.

Figures 2 and 3 show the values of RMSE and ME for the MMAP estimators of a, b, c and d at the four different levels of d. Observing these plots, the following trends can be found.

1. For the a-parameters and b-parameters, it can be seen that the values of RMSE(a) and RMSE(b) at $d = \{0.75, 0.85\}$ were smaller than $d = \{0.65, 0.95\}$. Similarly, the values of ME(a) were closer to 0 (smaller biases) for $d = \{0.75, 0.85\}$ than $d = \{0.65, 0.95\}$. This indicates the parameters a and b are more difficult to estimate when d takes more extremal values.

2. For the c-parameters. it can be seen that the relationships between d and RMSE(c) were the weakest among the four types of item parameters, and the highest values were not larger than 0.05. The values of ME(c) were very close to 0. These results demonstrated that the parameter of d have the smallest impact on the MMAP estimator of c.

3. For the d-parameters, the RMSE(d) displays substantial differences under the four levels of d, for the two middle levels of d, $d = \{0.75, 0.85\}$, the RMSE(d) were smaller than that of the two sides levels of d, $d = \{0.65, 0.95\}$ and had smaller biases. It suggests that the estimators of the middle d values are more accurate than that of

the extreme d values.

4.3 Simulation Study 3

Many researchers have studied the application of 4PLM to the psychopathology testing (Culpepper, 2016; Reise & Waller, 2003; Waller & Reise, 2010), where subjects with higher levels of psychopathology may be reluctant to self-report attitudes, behaviors, and/or experiences. Therefore, in this simulation, we compared the performance of the proposed MMAP estimation with that of the BM estimation for estimating the 4PLM with a set of psychopathology items. Following Culpepper (2016) and Waller and Feuerstahler (2017), this study generated responses based on the 4PLM with the $M = 23$ psychopathology item parameters from Waller and Reise (2010) as the true values; please refer to Table 5. The same as the above two simulation studies, the examinees' abilities (θ's) were randomly drawn from $N(0, 1)$, and three sample sizes of $N = \{1000, 5000, 10000\}$ were considered.

The MMAP estimates were calculated with the informative prior distributions that were given in the MMAP1 of Table1. In the "mirt" R library, the logistic model was design by a slope-threshold parameterizations, that is $1.7a_i$ and $1.7a_ib_i$ were estimated instead of directly estimating a_i and b_i. According to Waller and Feuerstahler (2017), the priors for $1.7a$ and $1.7ab$ were set to be $1.7a \sim LN(1, 1^2)$ and $1.7ab \sim N(0, 2^2)$. In addition, the prior distributions for c and d were set to be logistic(c) $\sim N(-1.2, 0.5^2)$ and logistic(d) $\sim N(1.2, 0.5^2)$, which are approximately equal to $Beta(5, 17)$ and $Beta(17, 5)$; please see Figure 4. To sum up, the prior distributions for the two estimation methods were very close. The MMAP and BM estimations of the 4PLM were calculated across 500 replications, and the RMSE were calculated to evaluate the properties of the estimators; please see Figures 5–7.

From these plots, it can be observed that, for most of the 23 items, the MMAP estimators of the item parameters (a, b, c, d) provided lower values of RMSEs than those of the BM estimators across the three sample sizes. It is indicated that the accuracy of the MMAP estimators were superior to that of the BM estimator. It is obviously that the RMSEs of MMAP and BM estimators both display decreasing trends as the sample size increased. That is, increasing sample size can improve the estimation accuracy, which is expected. Finally, the differences between the RMSEs of the MMAP and BM estimators were still exist under the sample size of $N = 10000$, but the superiorities of MMAP estimator were weaken, especially for the parameters of c, the two estimators were extremely close.

5 Empirical study

This section demonstrated an application of the 4PLM with an empirical example. The data set is from a state reading assessment test that was previously analyzed in Tao, Shi and Chang (2012). The dataset includes 50 dichotomous items and the sample size is $N = 2000$. In our study, the response data of the 50 dichotomous items was fitted by the 4PLM. The item parameters were estimated using the MMAP method, and the examinees' abilities were estimated using the Warm's weighted maximum likelihood estimation (WMLE). The Warm's WMLE has been proved to be superior to the ML and EAP estimates by many studies (Penfield & Bergeron, 2005; Warm,1989; Wand & Wang, 2001; Meng, Tao & Chen, 2016). The results of the parameter estimation and the model fitting evaluation are reported in the following.

5.1 Results of the item parameter estimation

The item parameter estimates from the 3PLM and 4PLM are presented in Table 6. It can be observed that the estimates of the parameters ($a, b,$ and c) in the two models (3PLM and 4PLM) are close for most items, while for the items with lower level of d, the differences between the estimates are more substantial. For instance, for items $j = 5, 9, 18$ and 50, their a parameters estimated from the 3PLM are extremely small, while the estimates from the 4PLM are much larger. This may be because that there were a large proportion of examinees slipping their responses to these items, resulting in the 3PLM underestimate their discrimination; see also the model fitting evaluation results given in Table 6 to be discussed in the next subsection.

The Pearson correlation coefficients between the parameters estimates of the 3PLM and the 4PLM are obtained: $r_{a^{(3PL)},a^{(4PL)}} = 0.68$, $r_{b^{(3PL)},b^{(4PL)}} = 0.94$, $r_{c^{(3PL)},c^{(4PL)}} = 0.88$, and the corresponding scatter plots are shown in the left column of Figure 8. We also illustrate the differences of the distributions of a, b and c between the 3PLM and the 4PLM by estimating their kernel density curves across the test; please see the right column in Figure 8. The estimates of a, b and c in 4PLM are highly correlated with those in 3PLM. Furthermore, it can be observed that the a parameter of the 4PLM was consistently higher than that of the 3PLM for each item, but the b parameter presented the opposite trend. This phenomenon has also been found in Loken and Rulison (2010). The reason for this may be that the upper asymptote less than 1 results in the response function does not have to flatten out to accommodate the poorly fitting responses (Loken & Rulison, 2010).

Finally, we compare the performances of the 4PLM and the 3PLM on estimating the examinees' abilities θ's. The scatter plot between the estimates of θ's from the 3PLM and the 4PLM and their kernel probability density function curves are presented in Figure 9. It can be seen that the estimates of θ's from the two models are highly correlated with their

Pearson correlation $r_{\theta(3PL),\theta(4PL)} = 0.98$. However, when $\theta > 1.0$, the estimates of θ's from the 4PLM are a little larger than that from the 3PLM. This indicates that the 3PLM is likely to underestimate the high-ability examinees. Furthermore, from the kernel density curves, it can be observed that the two curves of θ's are mostly overlapped, except for the right tail, where 3PLM may fail to capture the behaviors of the high ability students. It would be interesting to further investigate whether the result obtained in the empirical study still holds in general and how it would impact test taking strategies if 4PLM is known to be the scoring model beforehand. We would like to leave this interesting topic for future study.

5.2 Assessing model data fit

Assessing model fit is a routine and important procedure in IRT domain. IRT models can be implemented effectively for analyzing educational and psychological test data only when the fit of the model is met at least to a reasonable degree. In this study, the fit of the model to data was evaluated at the test and item levels respectively.

At the test level, the Chi-Square statistic, -2Log-Likelihood (-2logL) and AIC (Akaike,1973) were calculated. The test Chi-Square statistic is defined as,

$$\chi^2_{test} = \sum\nolimits_{h=1}^{H} \frac{(f_{oh} - f_{eh})^2}{f_{eh}},$$

where f_{oh} and f_{eh} is the observed and expected frequency of score h, $(h = 0, 1, ..., 50)$. The obtained results are displayed in Table 8. It can be seen that the three test-model fitting indexes consistently support that the 4PLM fits the data better than the 3PLM.

Moreover, to display the difference between the observed and the model predicted number-correct score distributions, the test fitting plot (Hambleton & Traub, 1973; Swaminathan, et al., 2006) is reported in Figure 10. It can be observed that the differences of the lines between the two models are very small for the test takers with test scores ≤ 40, but when the test scores > 40, the fitting frequency curve of the 4PLM is much closer to the observed score distribution than that of the 3PLM. That is, the 4PLM can better describe the data of the high-scores by modeling the slipping behaviors.

Following one reviewer's suggestion, we also fitted the 4PLM with several fixed upper asymptotes that are less than 1. We calculated the fitting indexes of the 4PLM under fixed $d = 0.98, 0.95$ and 0.90. The obtained results of the model-data fitting assessment are given in the bottom panel of Table 8. All the model indexes consistently support that the fitting of the 4PLM (without specifying d) is the best among all the considered models. This suggests that the 4PLM is a better choice in practice than the 4PLM with a fixed upper asymptote.

At the item level, the Pearson Chi-Square fit statistic (Hambleton & Han, 2005; Hambleton et al., 1991; Rogers & Hattie, 1987),

$$\chi^2_{item} = \sum\nolimits_{t=1}^{T} N_t \frac{(O_t - E_t)^2}{E_t\,(1 - E_t)},$$

and the likelihood ratio statistic (Mislevy & Bock, 1990; McKinley & Mills, 1985) provided in BILOG-MG,

$$G^2 = 2\sum\nolimits_{t=1}^{T} N_t \left(O_t \ln \frac{O_t}{E_t} + (1 - O_t) \ln \frac{1 - O_t}{1 - E_t} \right),$$

were calculated for assessing the model fitting. Here O_t denotes the observed proportion correct in trait interval t, E_t denotes the expected proportion correct in the interval under the given model, N_t is the number of persons in the interval, and T is the number of the trait intervals. In this study, $T = 15$ equal size intervals between -2.5 and 2.5 were chosen and the mean of the probabilities of a correct response was calculated as the expected. The obtained results are shown in Table 7. It can be found that the values of χ^2_{item} and G of the 4PLM are smaller than that of the 3PLM for most items, and the number of significant χ^2_{item} and G^2 statistics of the 4PLM is fewer, indicating that the 4PLM fits the data better than that of the 3PLM.

To further illustrate, we use graphical display to examine the discrepancy between observed and expected proportions (Swaminathan, et al., 2006). For illustration purposes, the fitting plot of item 5 is displayed in Figure 11. It shows that the upper asymptote of the probability of correct response gets close to 0.85 rather than approaching 1, as the ability level increases. Hence, the fitting of the 3PLM for this item shows serious deviation while the 4PLM can better captures the response behavior on this item.

6 Discussion

In this paper, we utilize a mixture model representation of the 4PLM and propose a MMAP approach for estimating the 4PLM with an EM algorithm. The mixture modeling revision of the 4PLM not only made the EM algorithm more easily to be implemented but also provided a natural connection with the popular cognitive diagnosis models. Three simulation studies were conducted to investigate the properties of the MMAP\EM estimation under various conditions. The first simulation study was designed to investigate the impacts of prior distributions on the accuracy of the MMAP estimation. The simulation results demonstrated that the accuracy of the MMAP estimators under different specifications of priors were almost equivalent when the sample size as large as $N = 5000$ and 10000. In the case of a smaller sample size $N = 1000$, the prior information has a larger impact on the MMAP estimation. This phenomenon is consistent with the simulation results reported in the discussion of Culpepper (2016). Thus uninformative priors should be recommended in the case of the sample size is small and a accurate prior information can not be obtained beforehand. The second simulation was to study the influences of the upper asymptote parameter d on the MMAP estimation. The results of this simulation demonstrated that the parameter d displayed substantial impacts on the MMAP estimates of a, b and d, where extreme values of d leaded to the decrease of the accuracy of MMAP estimators, but the influences of d on c were weaker. The goal of the third simulation was to compare the performance of the MMAP estimation with the BM estimation in Waller and Feuerstahler (2017). The obtained results suggested that our MMAP estimators are more accurate than that of the BM estimators across different sample sizes. Finally, a real data from a state reading assessment testing was analyzed using the 4PLM. The obtained results suggested that the upper asymptote parameter was needed, and in comparison with the 3PLM, the 4PLM can better fit this data. Additionally, the relationships of the common parameters estimators of the two models (3PLM and 4PLM) were investigated in this empirical study, which further illustrates the outperformance of the 4PLM.

There are several issues to be pursued in the future. First, it is interesting to study the MMAP estimation based on a hierarchical prior distribution that jointly models all the item parameters. The more flexible priors would allow to reduce the subjective error when specifying the prior distributions. On the other hand, this is also likely to increase the computational complexity which may result in a decrease in the accuracy of the parameter estimation. Second, the results of the empirical study demonstrated that scaling the high-ability examinees based on the 4PLM is more accurate than 3PLM. It is needed to further study the estimation performance under different simulation conditions. Furthermore, it would be interesting to study how it would impact test takers' strategies to answer items if the scoring model (such as 4PLM or 3PLM) is known beforehand. This is an important issue in practice and will be studied in the future. Third, the distribution of the ability parameter θ is specified to be the standard normal distribution in this study, which is commonly used in IRT. However, this assumption is likely to fail in practice, as suggested by the kernel density curves in Figure 9. It would be interesting to apply the joint likelihood estimation approach to estimate item parameters and θ simultaneously, which relaxes the normality assumption of θ. On the other hand, as known in the literature that joint estimation may suffer from inconsistency estimation issue when the number of items are not large enough. Therefore we would like to leave this interesting topic as a future study.

Table 1: The prior distributions of item parameters in the 4PLM.

	Prior (α)	Prior (b)	Prior (c)	Prior (d)
MMAP 1	$(\mu_\alpha = 0, \sigma_\alpha^2 = 1^2)$	$(\mu_b = 0, \sigma_b^2 = 1^2)$	$(s_c = 5, t_c = 17)$	$(s_d = 17, t_d = 5)$
MMAP 2	$(\mu_\alpha = 0, \sigma_\alpha^2 = 5^2)$	$(\mu_b = 0, \sigma_b^2 = 5^2)$	$(s_c = 3, t_c = 9)$	$(s_d = 9, t_d = 3)$
MMAP 3	$(\mu_\alpha = 0, \sigma_\alpha^2 = 10^2)$	$(\mu_b = 0, \sigma_b^2 = 10^2)$	$(s_c = 1, t_c = 1)$	$(s_d = 1, t_d = 1)$

Table 2: The values of RMSE for the MMAP estimators of the 4PLM item parameters, Sample Size ($N = 1000$).

	True Values				MMAP1				MMAP2				MMAP3			
Item	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	0.92	−0.48	0.16	0.85	0.28	0.13	0.04	0.03	0.30	0.15	0.04	0.03	0.35	0.26	0.09	0.05
2	0.93	0.75	0.18	0.82	0.20	0.19	0.03	0.04	0.23	0.20	0.04	0.04	0.23	0.23	0.04	0.08
3	1.22	0.23	0.16	0.95	0.36	0.14	0.03	0.06	0.33	0.14	0.04	0.05	0.27	0.13	0.04	0.04
4	0.65	1.77	0.18	0.87	0.23	0.45	0.03	0.11	0.23	0.41	0.03	0.10	0.25	0.54	0.03	0.16
5	1.35	2.16	0.24	0.77	0.33	0.25	0.02	0.02	0.40	0.27	0.02	0.03	0.44	0.38	0.02	0.20
6	1.09	1.64	0.12	0.89	0.39	0.26	0.02	0.10	0.41	0.23	0.02	0.10	0.39	0.25	0.02	0.12
7	0.49	1.60	0.17	0.90	0.18	0.60	0.02	0.14	0.17	0.55	0.03	0.14	0.24	0.65	0.04	0.18
8	0.74	1.46	0.11	0.91	0.25	0.41	0.03	0.13	0.27	0.39	0.03	0.13	0.29	0.42	0.03	0.14
9	0.86	0.16	0.13	0.92	0.39	0.14	0.06	0.08	0.39	0.16	0.06	0.07	0.36	0.17	0.07	0.06
10	0.72	0.45	0.18	0.88	0.21	0.25	0.03	0.07	0.21	0.25	0.04	0.07	0.24	0.27	0.05	0.07
11	1.31	1.23	0.16	0.93	0.32	0.24	0.02	0.11	0.33	0.21	0.03	0.10	0.31	0.17	0.03	0.07
12	1.09	1.69	0.14	0.91	0.40	0.28	0.02	0.12	0.39	0.26	0.02	0.12	0.38	0.27	0.02	0.14
13	1.07	0.61	0.05	0.86	0.48	0.12	0.05	0.07	0.45	0.13	0.05	0.07	0.26	0.17	0.04	0.08
14	1.09	0.78	0.19	0.88	0.26	0.22	0.03	0.06	0.29	0.19	0.04	0.05	0.26	0.20	0.04	0.06
15	1.23	0.89	0.20	0.84	0.25	0.17	0.03	0.04	0.28	0.16	0.03	0.04	0.27	0.17	0.03	0.07
16	0.97	1.88	0.08	0.81	0.40	0.22	0.02	0.04	0.39	0.21	0.02	0.04	0.40	0.30	0.02	0.14
17	0.61	0.17	0.05	0.87	0.37	0.18	0.12	0.08	0.37	0.21	0.11	0.08	0.40	0.28	0.13	0.09
18	0.60	1.14	0.10	0.86	0.29	0.26	0.05	0.09	0.28	0.28	0.05	0.09	0.32	0.32	0.05	0.12
19	0.79	1.89	0.25	0.91	0.27	0.64	0.07	0.16	0.29	0.54	0.07	0.14	0.23	0.62	0.05	0.19
20	0.68	0.56	0.18	0.92	0.23	0.32	0.03	0.10	0.24	0.32	0.04	0.10	0.25	0.27	0.05	0.09

Table 3: The values of RMSE for the MMAP estimators of the 4PLM item parameters, Sample Size ($N = 5000$).

	True Values				MMAP1				MMAP2				MMAP3			
Item	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	0.92	−0.48	0.16	0.85	0.11	0.09	0.04	0.02	0.11	0.10	0.03	0.02	0.12	0.13	0.04	0.02
2	0.93	0.75	0.18	0.82	0.10	0.13	0.03	0.02	0.12	0.13	0.03	0.03	0.12	0.12	0.03	0.03
3	1.22	0.23	0.16	0.95	0.16	0.08	0.02	0.03	0.13	0.08	0.02	0.02	0.12	0.07	0.02	0.02
4	0.65	1.77	0.18	0.87	0.08	0.42	0.03	0.14	0.09	0.47	0.03	0.14	0.08	0.52	0.01	0.16
5	1.35	2.16	0.24	0.77	0.30	0.12	0.01	0.02	0.32	0.12	0.01	0.04	0.29	0.19	0.01	0.14
6	1.09	1.64	0.12	0.89	0.23	0.18	0.01	0.11	0.23	0.19	0.01	0.11	0.23	0.21	0.01	0.12
7	0.49	1.60	0.17	0.90	0.07	0.62	0.02	0.15	0.07	0.62	0.02	0.15	0.09	0.67	0.02	0.17
8	0.74	1.46	0.11	0.91	0.10	0.37	0.02	0.13	0.09	0.35	0.02	0.13	0.10	0.37	0.02	0.13
9	0.86	0.16	0.13	0.92	0.18	0.09	0.04	0.04	0.14	0.08	0.03	0.04	0.13	0.09	0.03	0.03
10	0.72	0.45	0.18	0.88	0.08	0.18	0.02	0.05	0.08	0.17	0.02	0.04	0.08	0.17	0.02	0.04
11	1.31	1.23	0.16	0.93	0.23	0.14	0.01	0.07	0.22	0.11	0.01	0.06	0.19	0.09	0.01	0.05
12	1.09	1.69	0.14	0.91	0.20	0.22	0.01	0.12	0.20	0.22	0.01	0.12	0.21	0.24	0.01	0.14
13	1.07	0.61	0.05	0.86	0.23	0.08	0.03	0.04	0.22	0.08	0.03	0.04	0.19	0.08	0.02	0.04
14	1.09	0.78	0.19	0.88	0.12	0.13	0.02	0.03	0.13	0.12	0.02	0.03	0.15	0.11	0.03	0.03
15	1.23	0.89	0.20	0.84	0.18	0.09	0.02	0.02	0.19	0.09	0.03	0.03	0.20	0.09	0.03	0.04
16	0.97	1.88	0.08	0.81	0.18	0.16	0.01	0.07	0.19	0.18	0.01	0.09	0.22	0.27	0.01	0.14
17	0.61	0.17	0.05	0.87	0.24	0.13	0.09	0.06	0.22	0.13	0.09	0.06	0.22	0.15	0.09	0.06
18	0.60	1.14	0.10	0.86	0.16	0.24	0.03	0.10	0.15	0.25	0.03	0.10	0.15	0.28	0.03	0.11
19	0.79	1.89	0.25	0.91	0.26	0.60	0.07	0.17	0.27	0.57	0.07	0.17	0.25	0.62	0.06	0.19
20	0.68	0.56	0.18	0.92	0.11	0.24	0.02	0.07	0.10	0.24	0.02	0.07	0.10	0.22	0.02	0.07

Table 4: The values of RMSE for the MMAP estimators of the 4PLM item parameters, Sample Size ($N = 10000$).

	True Values				MMAP1				MMAP2				MMAP3			
Item	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	0.92	−0.48	0.16	0.85	0.08	0.08	0.03	0.02	0.08	0.09	0.03	0.02	0.08	0.10	0.04	0.02
2	0.93	0.75	0.18	0.82	0.10	0.09	0.02	0.02	0.11	0.09	0.03	0.02	0.11	0.09	0.03	0.02
3	1.22	0.23	0.16	0.95	0.08	0.06	0.02	0.02	0.08	0.06	0.02	0.01	0.09	0.05	0.02	0.01
4	0.65	1.77	0.18	0.87	0.07	0.43	0.03	0.13	0.07	0.49	0.03	0.14	0.06	0.52	0.03	0.15
5	1.35	2.16	0.24	0.77	0.30	0.09	0.01	0.02	0.30	0.10	0.01	0.04	0.27	0.13	0.01	0.10
6	1.09	1.64	0.12	0.89	0.16	0.18	0.01	0.10	0.16	0.19	0.01	0.10	0.17	0.19	0.01	0.10
7	0.49	1.60	0.17	0.90	0.06	0.65	0.02	0.15	0.05	0.63	0.02	0.16	0.06	0.64	0.02	0.16
8	0.74	1.46	0.11	0.91	0.06	0.35	0.02	0.12	0.06	0.34	0.02	0.13	0.07	0.35	0.02	0.12
9	0.86	0.16	0.13	0.92	0.12	0.06	0.03	0.03	0.09	0.07	0.03	0.03	0.09	0.07	0.03	0.02
10	0.72	0.45	0.18	0.88	0.06	0.15	0.02	0.04	0.05	0.16	0.02	0.03	0.06	0.13	0.02	0.03
11	1.31	1.23	0.16	0.93	0.16	0.10	0.01	0.05	0.15	0.08	0.01	0.04	0.15	0.07	0.01	0.03
12	1.09	1.69	0.14	0.91	0.14	0.22	0.01	0.12	0.14	0.23	0.01	0.12	0.17	0.23	0.01	0.12
13	1.07	0.61	0.05	0.86	0.17	0.05	0.02	0.03	0.15	0.06	0.02	0.03	0.12	0.06	0.02	0.03
14	1.09	0.78	0.19	0.88	0.10	0.08	0.02	0.02	0.11	0.07	0.02	0.02	0.12	0.07	0.02	0.02
15	1.23	0.89	0.20	0.84	0.17	0.06	0.02	0.02	0.19	0.06	0.02	0.03	0.20	0.06	0.02	0.03
16	0.97	1.88	0.08	0.81	0.14	0.17	0.01	0.08	0.15	0.20	0.01	0.11	0.19	0.25	0.01	0.13
17	0.61	0.17	0.05	0.87	0.20	0.11	0.08	0.06	0.19	0.10	0.08	0.05	0.19	0.12	0.08	0.05
18	0.60	1.14	0.10	0.86	0.13	0.23	0.02	0.09	0.12	0.24	0.02	0.09	0.13	0.24	0.02	0.09
19	0.79	1.89	0.25	0.91	0.27	0.60	0.07	0.17	0.26	0.59	0.07	0.17	0.24	0.62	0.07	0.18
20	0.68	0.56	0.18	0.92	0.08	0.20	0.01	0.06	0.08	0.20	0.02	0.06	0.07	0.18	0.02	0.05

Table 5: The values of item parameters for the psychopathology item in Waller and Reise (2010).

	Item parameters					Item parameters			
Item	a	b	c	d	Item	a	b	c	d
1	1.91	−0.28	0.04	0.52	14	0.84	0.72	0.04	0.75
2	1.95	−0.16	0.02	0.48	15	1.13	0.15	0.03	0.61
3	1.50	0.05	0.02	0.60	16	0.79	1.19	0.04	0.73
4	1.12	0.06	0.02	0.63	17	1.27	0.48	0.01	0.84
5	0.89	0.45	0.04	0.82	18	0.94	1.37	0.09	0.94
6	1.08	−0.50	0.06	0.83	19	0.84	1.44	0.02	0.82
7	1.16	−0.47	0.07	0.71	20	1.14	1.52	0.00	0.82
8	1.10	0.01	0.04	0.73	21	1.10	0.25	0.02	0.93
9	0.78	0.45	0.05	0.57	22	0.72	0.53	0.24	0.95
10	1.23	0.19	0.01	0.90	23	0.88	1.56	0.06	0.91
11	1.34	0.41	0.02	0.85					
12	1.54	−0.48	0.06	0.59					
13	1.16	0.18	0.02	0.40					

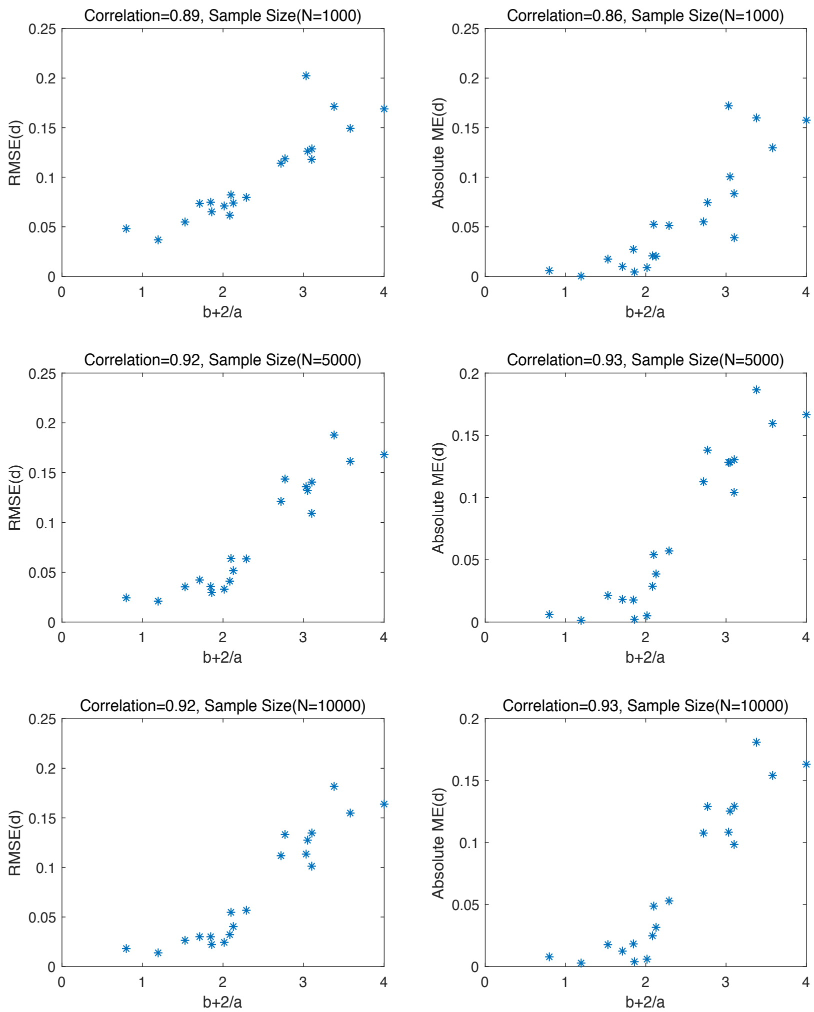

Figure 1: The left column show the scatter plots between $b+2/a$ and RMSE of the MMAP estimators for the d-parameter across the sample sizes of $N = \{1000, 5000, 10000\}$; The right column show the scatter plots between $b + 2/a$ and absolute ME of the MMAP estimators for the d-parameter across the sample sizes of $N = \{1000, 5000, 10000\}$.

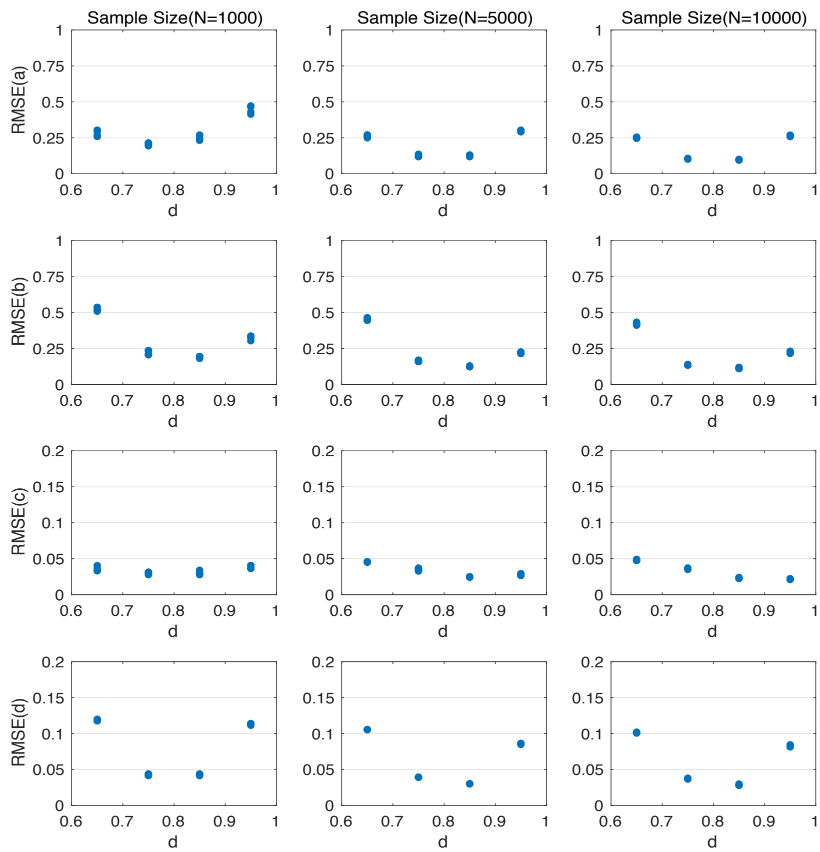

Figure 2: The values of RMSE of the MMAP estimators for the 4PLM item parameters under the four levels of $d = \{0.65, 0.75, 0.85, 0.95\}$ and the three sample sizes of $N = \{1000, 5000, 10000\}$.

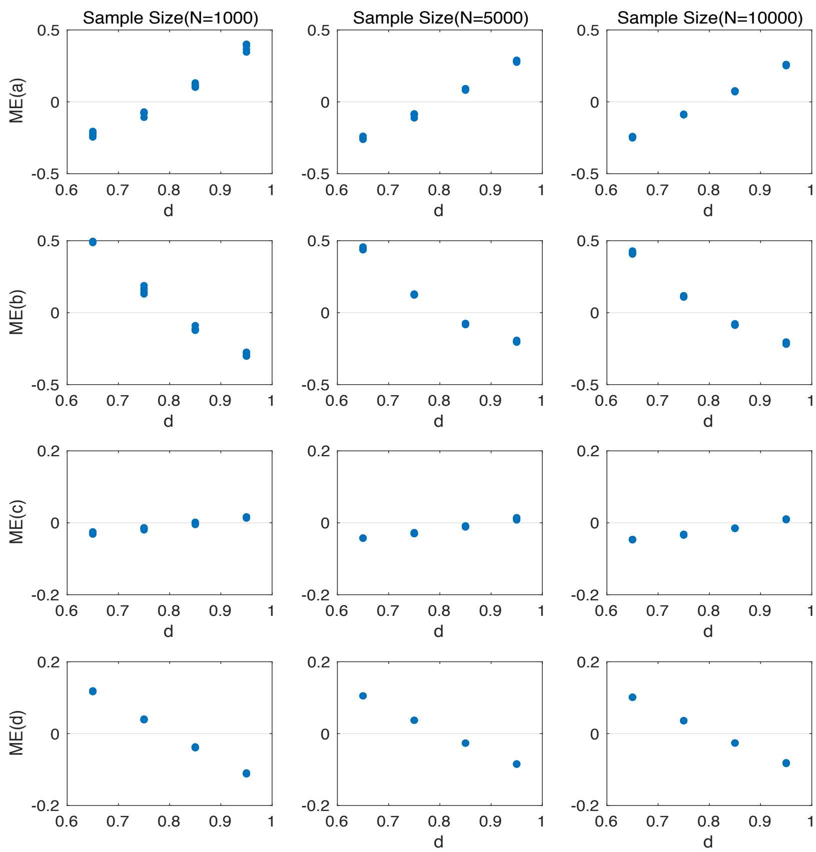

Figure 3: The values of ME of the MMAP estimators for the 4PLM item parameters under the four levels of $d = \{0.65, 0.75, 0.85, 0.95\}$ and the three sample sizes of $N = \{1000, 5000, 10000\}$.

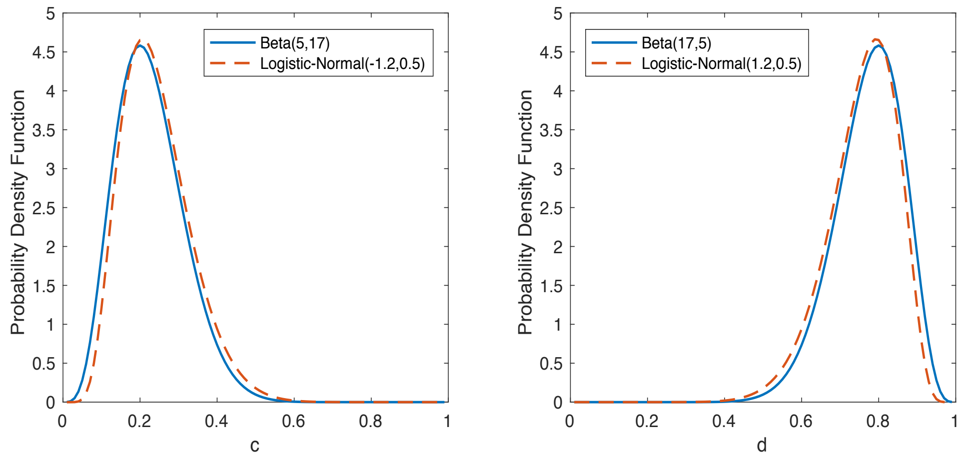

Figure 4: The probability density function curves for the distributions of Beta(5,17) and Logistic-Normal(-1.2,0.5), Beta(17,5) and Logistic-Normal(1.2,0.5).

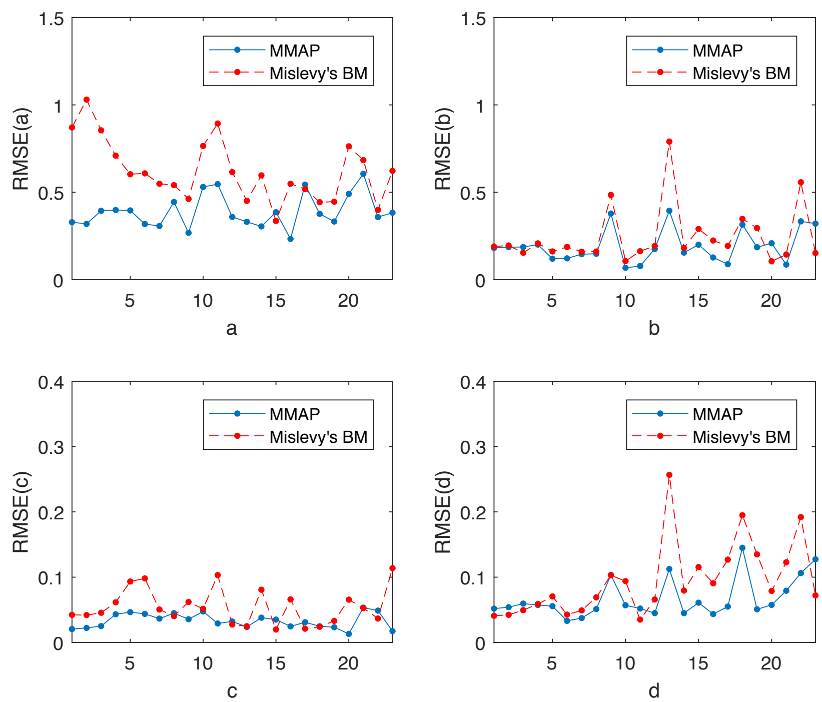

Figure 5: The values of RMSE for the MMAP and BM estimators of the 4PLM item parameters under the sample size is $N = 1000$.

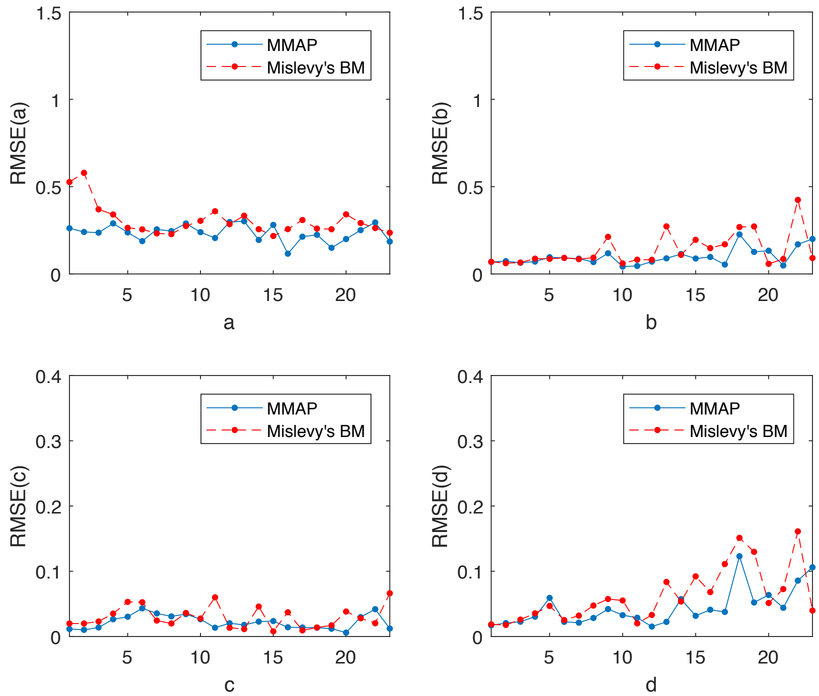

Figure 6: The values of RMSE for the MMAP and BM estimators of the 4PLM item parameters under the sample size is $N = 5000$.

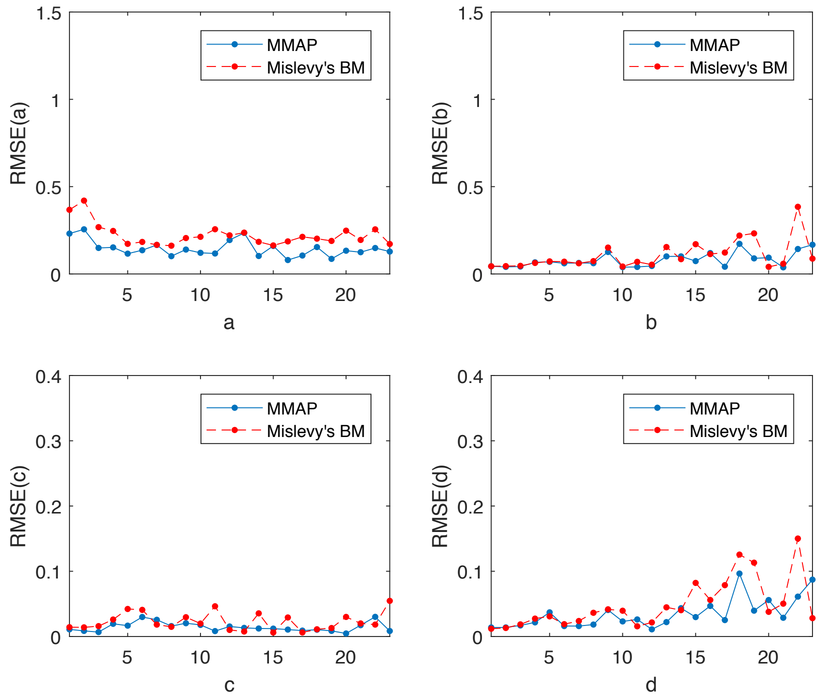

Figure 7: The values of RMSE for the MMAP and BM estimators of the 4PLM item parameters under the sample size is $N = 10000$.

Table 6: The item parameter estimates of 4PLM and 3PLM for the empirical data.

	4PLM				3PLM				4PLM				3PLM		
Item	a	b	c	d	a	b	c	Item	a	b	c	d	a	b	c
1	0.95	−0.96	0.07	0.97	0.88	−0.98	0.05	26	1.58	0.07	0.07	0.96	1.39	0.17	0.05
2	1.20	−0.88	0.07	0.91	0.75	−0.81	0.04	27	0.99	−0.23	0.15	0.95	0.84	−0.16	0.13
3	0.91	−0.38	0.08	0.93	0.74	−0.28	0.05	28	1.15	−0.63	0.09	0.94	0.84	−0.58	0.05
4	1.01	−0.79	0.07	0.96	0.83	−0.76	0.05	29	1.31	0.47	0.23	0.93	1.21	0.64	0.23
5	1.79	−1.31	0.08	0.84	0.47	−1.49	0.05	30	1.25	0.12	0.15	0.94	1.16	0.28	0.15
6	1.50	−1.29	0.07	0.98	1.17	−1.38	0.04	31	1.41	−0.31	0.16	0.94	1.06	−0.23	0.12
7	1.97	−1.04	0.08	0.98	1.43	−1.12	0.04	32	0.68	0.04	0.11	0.87	0.55	0.31	0.08
8	0.71	−0.89	0.11	0.92	0.53	−0.81	0.06	33	1.01	−0.96	0.13	0.98	0.85	−1.05	0.07
9	0.87	0.31	0.13	0.76	0.57	0.83	0.08	34	1.08	−0.61	0.06	0.91	0.76	−0.45	0.04
10	1.13	−0.28	0.10	0.93	0.88	−0.19	0.06	35	1.58	−0.26	0.20	0.92	1.00	−0.19	0.13
11	1.44	−0.66	0.11	0.97	1.11	−0.67	0.06	36	1.54	−0.35	0.14	0.87	0.79	−0.23	0.05
12	1.09	0.14	0.08	0.88	0.80	0.36	0.06	37	1.25	−0.39	0.11	0.96	1.02	−0.36	0.07
13	1.91	−0.43	0.19	0.97	1.45	−0.41	0.16	38	0.97	−0.29	0.11	0.92	0.73	−0.19	0.07
14	1.28	−0.78	0.07	0.90	0.77	−0.68	0.03	39	0.70	0.52	0.11	0.85	0.61	0.84	0.08
15	0.97	−0.84	0.10	0.97	0.81	−0.86	0.06	40	1.07	0.14	0.13	0.94	0.99	0.29	0.13
16	1.27	−0.88	0.08	0.94	0.87	−0.87	0.04	41	1.31	0.26	0.09	0.93	1.17	0.39	0.08
17	0.85	−0.11	0.07	0.83	0.62	0.24	0.04	42	1.33	−0.89	0.09	0.96	0.96	−0.92	0.04
18	1.17	−0.96	0.09	0.70	0.37	−0.24	0.05	43	0.79	−0.06	0.07	0.87	0.63	0.19	0.04
19	1.54	−0.32	0.12	0.96	1.22	−0.25	0.09	44	1.64	−0.45	0.14	0.97	1.32	−0.39	0.11
20	1.31	−0.17	0.13	0.94	1.11	−0.02	0.12	45	1.23	−0.58	0.13	0.93	0.81	−0.57	0.05
21	0.84	0.09	0.15	0.94	0.81	0.28	0.16	46	1.82	−0.53	0.09	0.94	1.15	−0.49	0.04
22	1.30	−0.55	0.09	0.97	1.08	−0.52	0.06	47	2.22	−0.57	0.09	0.98	1.81	−0.56	0.06
23	2.36	−0.38	0.18	0.98	1.76	−0.32	0.15	48	1.38	−0.67	0.11	0.98	1.21	−0.67	0.08
24	1.98	−0.90	0.07	0.94	1.01	−0.93	0.03	49	1.12	−0.39	0.09	0.91	0.81	−0.26	0.04
25	1.17	−0.67	0.12	0.96	0.91	−0.69	0.07	50	0.89	−0.60	0.10	0.70	0.39	0.22	0.05

Table 7: The item model fit indices for 4PLM and 3PLM.

	χ^2_{item}		G^2			χ^2_{item}		G^2	
Item	4PL	3PL	4PL	3PL	Item	4PL	3PL	4PL	3PL
1	21.56	23.34*	19.51	25.79*	26	14.81	6.98	20.17*	8.34
2	5.58	11.27	7.97	12.70	27	20.12*	16.00	19.96*	16.13
3	4.24	6.85	4.39	5.24	28	13.14	10.96	12.38	12.52
4	18.15	28.04*	22.11*	27.72*	29	14.34	13.91	17.99	15.78
5	5.14	48.02*	7.81	45.26*	30	17.79	9.26	21.07*	10.06
6	6.94	6.78	10.19	7.61	31	9.93	13.11	9.25	15.76
7	10.61	15.19	14.44	15.85	32	24.91*	13.46	22.22*	12.85
8	14.79	10.66	16.65	11.76	33	19.39	18.74	22.06*	22.13*
9	16.51	21.47*	15.69	19.12	34	15.58	13.68	15.06	15.71
10	24.07*	24.49*	27.83*	23.28*	35	11.48	25.55*	12.43	21.88*
11	6.57	7.66	8.98	7.58	36	14.52	28.88*	17.21	25.99*
12	11.74	9.07	12.20	10.05	37	11.13	14.85	14.14	13.36
13	11.03	14.39	14.03	13.81	38	9.37	16.36	9.80	17.22
14	7.81	15.75	7.54	18.61	39	32.99*	17.97	31.28*	16.69
15	5.40	9.42	13.17	12.03	40	16.32	14.27	18.04	15.49
16	7.10	12.50	8.01	14.12	41	9.30	17.08	12.78	20.84
17	9.28	9.28	10.38	9.75	42	12.50	22.64*	11.42	19.81*
18	21.59*	31.20*	23.35*	31.74*	43	18.41	7.75	20.29*	8.72
19	11.13	11.88	12.09	11.15	44	8.84	8.59	12.59	7.98
20	13.60	12.98	14.62	14.31	45	9.54	15.06	11.00	13.97
21	19.02	24.35*	19.01	29.91*	46	13.82	29.57*	15.33	19.78*
22	6.01	8.72	6.88	9.62	47	16.58	8.97	18.56	10.68
23	19.56	19.98	18.25	17.60	48	9.06	9.81	11.30	13.85
24	13.55	60.45*	20.77*	51.16*	49	4.32	10.38	5.87	10.39
25	8.73	6.26	9.77	8.17	50	8.97	21.56*	8.78	24.45*

Note:* denotes the valus of χ^2_{test} and G^2 are greater than the critical values at the significance level of 5%.

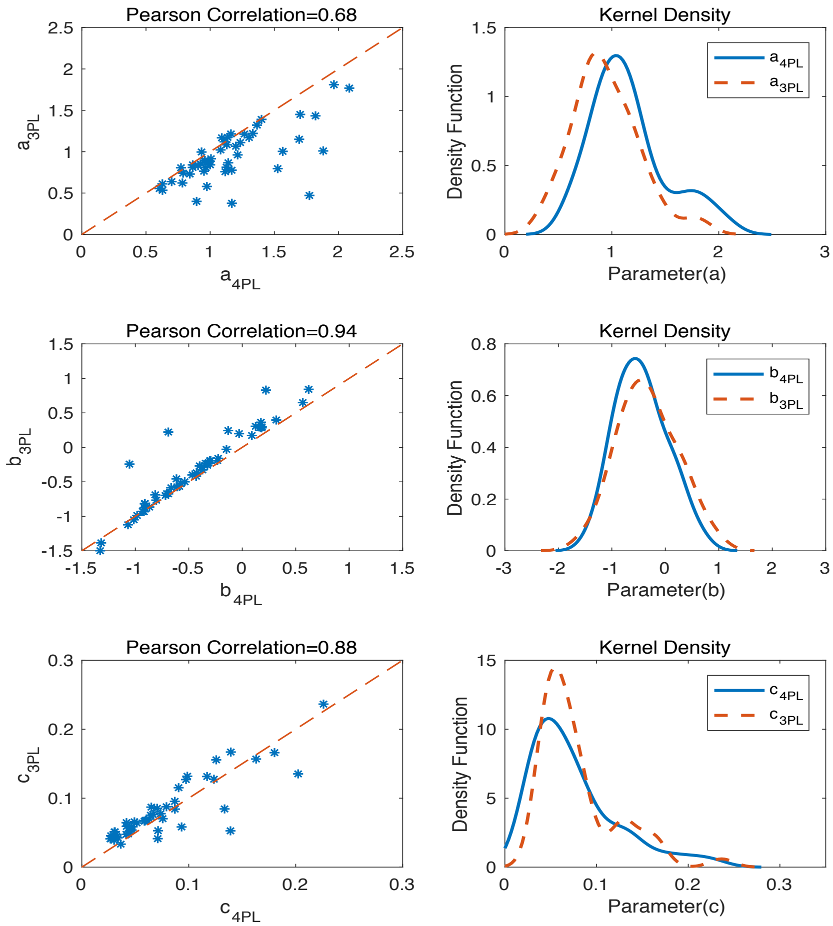

Figure 8: The three plots in the left column are the scatter plots between the estimates of the item parameter (a, b and c) in the 4PLM and the 3PLM; The three plots in the right column are the kernel probability density function curves of a, b and c under the 4PLM and the 3PLM

Table 8: The test model fit indices for the 4PLM, 3PLM, and the 4PLM with three constrained upper asymptotes ($d = 0.98, 0.95, 0.90$).

	χ^2_{test}	$-2LogL$	AIC
4PLM	99.87	104631	105031
3PLM	112.25	104896	105196
4PLM−0.98	101.20	104944	105244
4PLM−0.95	103.20	105124	105424
4PLM−0.90	301.57	105850	106150

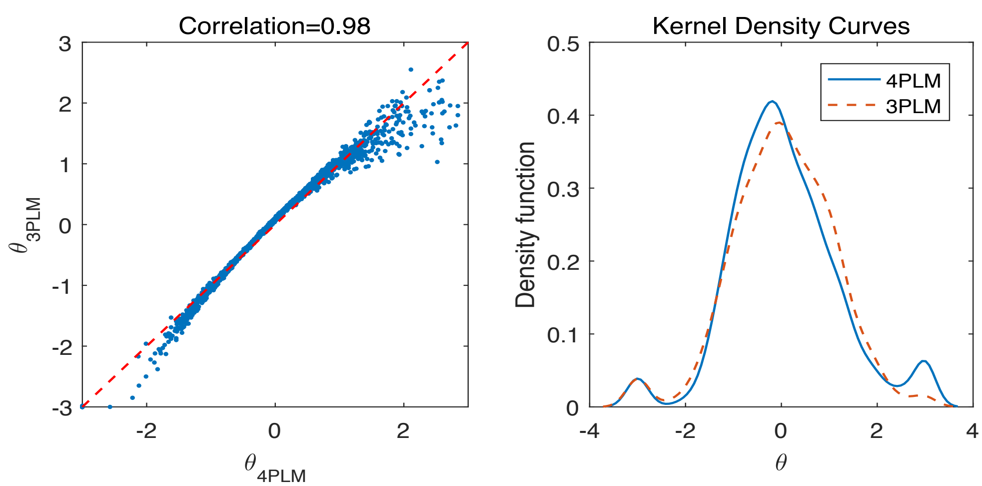

Figure 9: The scatter plots between the estimates of θ in the 4PLM and the 3PLM, and the kernel density function curves of the estimates of θ in the 4PLM and the 3PLM.

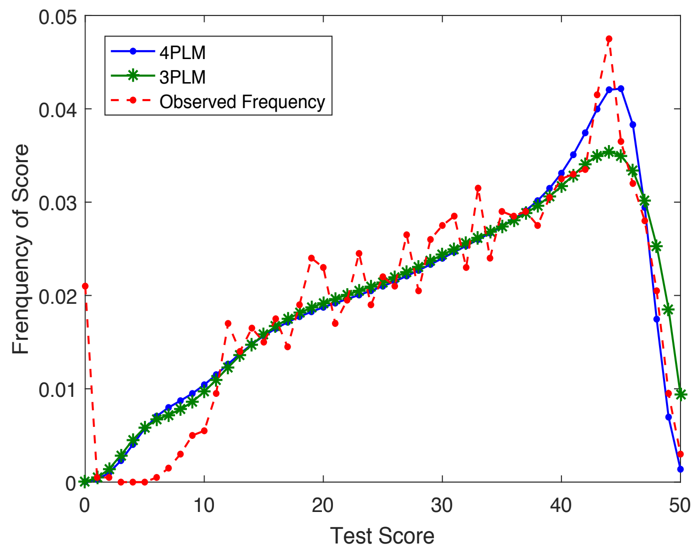

Figure 10: Observed and expected test score distributions based on 4PLM and 3PLM.

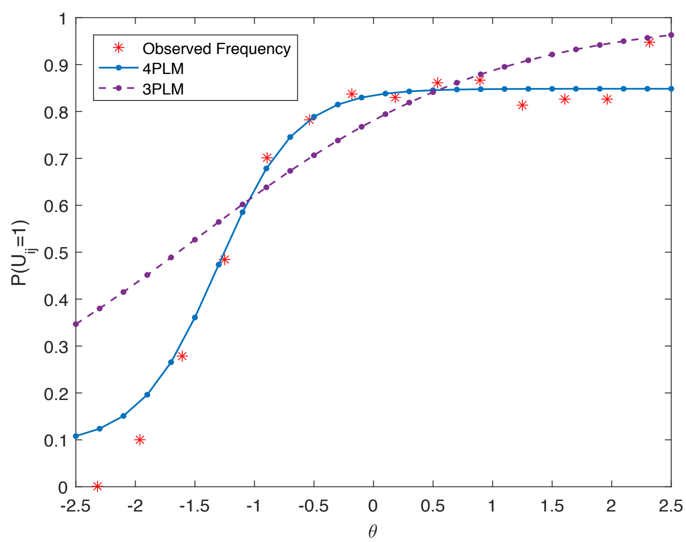

Figure 11: Observed and expected proportion of correct response on item 5 based on 4PLM and 3PLM.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N. and Csaki, F. (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Academiai Kiado.

Baker, F. B., & Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker.

Barton, M. A., & Lord, F. M. (1981). *An upper asymptote for the three-parameter logistic item-response model* (Tech.Rep. No. No. 80-20). Educational Testing Service.

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

Culpepper, S. A. (2017). The Prevalence and Implications of Slipping on Low-Stakes, Large-Scale Assessments. *Journal of Educational and Behavioral Statistics*, *42*, 706-725.

Culpepper, S. A. (2016). Revisting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, *81*, 1142–1163.

de la Torre, J., & Douglas, J.A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333-353.

Feuerstahler, L. M., & Waller, N. G. (2014). Estimation of the 4-parameter model with marginal maximum likelihood. *Multivariate Behavioral Research*, *49*, 285–285.

Hambleton, R.K., & Han, N. (2005). Assessing the fit of IRT models to educational and psychological test data: A five step plan and several graphical displays. In: Lenderking, W.R., Revicki, D. (Eds.), Advances in Health Outcomes Research Methods, Measurement, Statistical Analysis, and Clinical Applications, Degnon Associates, Washington.

Hambleton, R. K.,& Traub, R. E. (1973). Analysis of empirical data using two logistic latent trait models. *British Journal of Mathematical and Statistical Psychology*, *24*, 273–281.

Hambleton,R.K., Swaminathan,H., & Rogers,H.J.(1991). Fundamentals of Item Response Theory. Sage, Newbury Park, CA.

Liao, W., Ho, R., Yen, Y., & Cheng, H. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. Social Behavior and Personality: *An International Journal*, *40*, 1679–1694.

Linacre, J. M. (2004). Discrimination, guessing and carelessness: Estimating IRT parameters with Rasch. *Rasch Measurement Transactions*, *18*, 959–960.

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *British Journal of Mathematical and Statistical Psychology*, *63*, 509–525.

Lord, F. M. (1975). Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters (RB-75-33). Princeton, NJ: Educational Testing Service.

Magis, D. (2013). A note on the item information function of the four-parameter logistic model. *Applied Psychological Measurement*, *37*, 304–315.

McKinley, R., & Mills, C. (1985). A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement*, 9, 49-57.

Meng, X. B., Tao, J., & Chen, S. L. (2016). Warm's weighted maximum likelihood estimation of latent trait in the four-parameter logistic model. *Acta Psychologica Sinica*, *48*, 1047–1056.

Mislevy, R. J. (1986). Bayes model estimation in item response models. *Psychometrika*, *51*, 177–195.

Mislevy, R.J.,& Bock, R.D. (1990). BILOG: Item analysis and test scoring with binary logistic models [Computer program]. Scientific Software, Chicago.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations *Econometrica*, *16*, 1–32.

Ogasawara, H. (2012).Asymptotic expansions for the ability estimator in item response theory. *Computational Statistics*, *27*, 661–683.

Penfield, R. D & Bergeron, J. M. (2005). Applying a weighted maximum likelihood latent trait estimator to the generalized partial credit model. *Applied Psychological Measurement*, *29*, 218–233.

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods*, 8, 164–184.

Rouse, S. V., Finger, M. S., & Butcher, J. N. (1999). Advances in clinical personality measurement: An item response theory analysis of the MMPI-2 PSY-5 scales. *Journal of Personality Assessment*, *72*, 282–307.

Rogers, H., & Hattie, J. (1987). A Monte Carlo investigation of several person and item fit statistics for item response models. Applied Psychological Measurement 11, 47C57.

Rulison, K. L., & Loken, E. (2009). I've fallen and I can't get up: Can high-ability students recover from early mistakes in CAT? *Applied Psychological Measurement*, *33*, 83–101.

Rupp, A. A. (2003). Item response modeling with BILOG-MG and MULTILOG for Windows. *International Journal of Testing*, *3*, 365–384.

San Martn, E., del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, *30*, 183–203.

Swaminathan, H., Hambleton, R.K., & Rogers,H.J.(2006) Assessing the Fit of Item Response Theory Models. *Handbook of Statistics,Volume 26, Psychometrics*, pp: 683-715.

Tao, J., Shi, N. Z., & Chang, H. H. (2012). Item-weighted likelihood method for ability estimation in tests composed of both dichotomous and polytomous items. *Journal of Educational and Behavioral Statistics*, *37*, 298–315.

Tavares, H. R., de Andrade, D. F., & Pereira, C. A. (2004). Detection of determinant genes and diagnostic via item response theory. *Genetics and Molecular Biology*, *27*, 679–685.

Von Davier, M.(2009). Is There Need for the 3PL Model? Guess What?. *Measurement: Interdisciplinary Research* & *Perspective*, *7*: 110-114.

Wang, C., Chang, H.-H., & Douglas, J. A. (2013). The linear transformation model with frail-ties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology*, 66, 144-168.

Waller, N. G.,& Feuerstahler, S. (2017). Bayesian Modal Estimation of the Four-Parameter Item Response Model in Real, Realistic, and Idealized Data Sets. *Multivariate Behavioral Research*, 1-21.

Waller, N. G., & Reise, S. P. (2010). Measuring psychopathology with nonstandard item response theory models: Fitting the four-parameter model to the Minnesota

Multiphasic Personality Inventory. InS. Embretson (Ed.), *Measuring psychological constructs: Advances in model based approaches.* Washington, DC: American Psychological Association.

Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*, *25*, 317–331.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427–450.

Appendix A: The Newton-Raphson interaction for solving equations 20 and 21.

Let $\alpha_j^{(r)}$ and $b_j^{(r)}$ be the current estimates, then the next estimates are given by,

$$\begin{pmatrix} \alpha_j^{(r+1)} \\ b_j^{(r+1)} \end{pmatrix} = \begin{pmatrix} \alpha_j^{(r)} \\ b_j^{(r)} \end{pmatrix} - \begin{pmatrix} L_{\alpha_j\alpha_j}(\alpha_j^{(r)}, b_j^{(r)}) & L_{\alpha_j b_j}(\alpha_j^{(r)}, b_j^{(r)}) \\ L_{\alpha_j b_j}(\alpha_j^{(r)}, b_j^{(r)}) & L_{b_j b_j}(\alpha_j^{(r)}, b_j^{(r)}) \end{pmatrix}^{-1} \begin{pmatrix} L_{\alpha_j}(\alpha_j^{(r)}, b_j^{(r)}) \\ L_{b_j}(\alpha_j^{(r)}, b_j^{(r)}) \end{pmatrix}, \tag{A.1}$$

where

$$L_{\alpha_j}(\alpha_j, b_j) = \frac{\partial E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left(\ln p\left(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}\,\right)\right)}{\partial \alpha_j},$$

$$L_{b_j}(\alpha_j, b_j) = \frac{\partial E_{\boldsymbol{W},\boldsymbol{\theta}|\boldsymbol{u},\boldsymbol{\xi}^t}\left(\ln p\left(\boldsymbol{\xi}, \boldsymbol{z}\,|\boldsymbol{u}, \Omega, \boldsymbol{\tau}\,\right)\right)}{\partial b_j},$$

are given in Equations (20) and (21), and

$$L_{\alpha_j\alpha_j}(\alpha_j, b_j) = \frac{\partial L_{\alpha_j}(\alpha_j, b_j)}{\partial \alpha_j} = -e^{2\alpha_j} \sum_{i=k}^{K} \left[\hat{R}(x_k)(x_k - b_j)^2 p_j^*(x_k)(1 - p_j^*(x_k))\right] - \frac{1}{\sigma_\alpha}, \tag{A.2}$$

$$L_{b_j b_j}(\alpha_j, b_j) = \frac{\partial L_{b_j}(\alpha_j, b_j)}{\partial b_j} = -e^{2\alpha_j} \sum_{i=k}^{K} \left[\hat{R}(x_k) p_j^*(x_k)(1 - p_j^*(x_k))\right] - \frac{1}{\sigma_b}, \tag{A.3}$$

$$\begin{aligned} L_{b_j\alpha_j}(\alpha_j, b_j) &= L_{\alpha_j b_j}(\alpha_j, b_j) \\ &= \frac{\partial L_{\alpha_j}(\alpha_j, b_j)}{\partial b_j} = e^{2\alpha_j} \sum_{i=k}^{K} \left[\hat{R}(x_k)(x_k - b_j) p_j^*(x_k)(1 - p_j^*(x_k))\right]. \end{aligned} \tag{A.4}$$

where $p_j^*(\cdot)$ is defined in (3).

Appendix B: Matlab code of MMAP\EM for 4PLM

```
function [R_a, R_b, R_c, R_d]=MMAP(u, n, prior_a, prior_b, prior_c, prior_d)
%   u: is the response matrix
%   prior_a: is the prior of a
%   prior_b: is the prior of b
%   prior_c: is prior of c
%   prior_d: is prior of d
%   M: is the number of test takers
%   N: is the number of items
%   ntime: is number of the Fisher-Scoring iteration
%   NTIME: is number of the EM algorithm
%   a0: is initial value of a parameter
%   b0: is initial value of b parameter
%   c0: is initial value of c parameter
%   d0: is initial value of d parameter
%   n: is the number of the quadrature points
%   x: is quadrature points
indice=1;
INDICE=1;
ntime=0;
NTIME=0;
[M,N]=size(u);
x=linspace(-4,4,n);
x1=x';
d=x1(2)-x1(1);
Ak=normpdf(x1,0,1)*d;
r0=identify(u);
a0=r0./sqrt(1-r0.^2);
a0=log(a0)*0;
b0=sum(u)./M;
b0=-norminv(b0,0,1)./r0;
c0=0*a0+0.25;
d0=0*a0+0.8;
P=@(a,b,c,d,x)c+(d-c)./(1+exp(-a.*(x-b)));
% ----------------------------------------
MM=ones(M,1);
a1=MM*a0;
b1=MM*b0;
c1=MM*c0;
d1=MM*d0;
amu=prior_a(1);
```

```
asigma2=prior_a(2);
bmu=prior_b(1);
bsigma2=prior_b(2);
for k=1:n
    p=P(exp(a1),b1,c1,d1,x(k));
    L=p.^u.*((1-p).^(1-u));
    LL(:,k)=prod(L,2)*Ak(k);
end
LL0=sum(LL,2);
LH=sum(log(LL0));
% E-step and M-step
NM=ones(1,n);
while INDICE==1 && NTIME<Niteration
    LL1=LL0*nn;
    h=LL./LL1;
    f=sum(h);
    for i=1:N
        U=(u(:,i))*NM;
        p=MM*P(exp(a0(i)),b0(i),c0(i),d0(i),x);
        ppp=(p-c0(i))/(d0(i)-c0(i));
        pz=(d0(i)*ppp./p).*U+((1-d0(i))*ppp./(1-p)).*(1-U);
        PZ=pz.*h;
        r=sum(PZ);
        c0(i)=(sum(u(:,i).*(1-sum(PZ,2)))+prior_c(1))/(sum(1-sum(PZ,2))+prior_c(2));
        S=(sum(sum(PZ,2).*u(:,i))+prior_d(1))/(sum(sum(PZ,2))+prior_d(2));
        if S>c0(i)
            d0(i)=S;
        else
            d0(i)=c0(i)+0.1;
        end
        at=a0(i);
        bt=b0(i);
        while indice==1 && ntime<50
            Pi=P(exp(at),bt,0,1,x);
            w=Pi.*(1-Pi);
            la1=exp(at)*sum((x-bt).*(r-f.*Pi))-(at-amu)/asigma2;
            lb1=-exp(at)*sum(r-f.*Pi)-(bt-bmu)/bsigma2;
            laa=-exp(2*at)*sum((f.*(x-bt).^2.*w))-1/asigma2;
            lbb=-exp(2*at)*sum((f.*w))-1/bsigma2;
            lab=exp(2*at)*sum((x-bt).*f.*w) ;
            res=[at;bt]-[laa,lab;lab,lbb]^(-1)*[la1;lb1];
            at1=res(1);
            bt1=res(2);
            if norm([at1-at;bt1-bt],2)<0.0001
```

```
                indice=0;
            else
                at=at1;
                bt=bt1;
                ntime=ntime+1;
            end
        end
        ntime=1;
        indice=1;
        a0(i)=at1;
        b0(i)=bt1;
    end
    a1=MM*a0;
    b1=MM*b0;
    c1=MM*c0;
    d1=MM*d0;
    for k=1:n
        p=P(exp(a1),b1,c1,d1,x(k));
        L=p.^u.*((1-p).^(1-u));
        LL(:,k)=prod(L,2)*Ak(k);
    end
    LL0=sum(LL,2);
    LH1=sum(log(LL0));
    if abs(LH-LH1)<10^(-3)
        INDICE=0;
    else
        NTIME=NTIME+1;
        LH=LH1;
    end
    RL(NTIME)=LH1;
end
Ra=exp(a0);
Rb=b0;
Rc=c0;
Rd=d0;
```