PSYCHOMETRIKA—VOL. 89, NO. 4, 1337–1365
DECEMBER 2024
https://doi.org/10.1007/s11336-024-09998-x

BAYESIAN ADAPTIVE LASSO FOR DETECTING ITEM–TRAIT RELATIONSHIP AND DIFFERENTIAL ITEM FUNCTIONING IN MULTIDIMENSIONAL ITEM RESPONSE THEORY MODELS

NA SHAN

NORTHEAST NORMAL UNIVERSITY

PING- FENG XU

NORTHEAST NORMAL UNIVERSITY

SHANGHAI ZHANGJIANG INSTITUTE OF MATHEMATICS

In multidimensional tests, the identification of latent traits measured by each item is crucial. In addition to item–trait relationship, differential item functioning (DIF) is routinely evaluated to ensure valid comparison among different groups. The two problems are investigated separately in the literature. This paper uses a unified framework for detecting item–trait relationship and DIF in multidimensional item response theory (MIRT) models. By incorporating DIF effects in MIRT models, these problems can be considered as variable selection for latent/observed variables and their interactions. A Bayesian adaptive Lasso procedure is developed for variable selection, in which item–trait relationship and DIF effects can be obtained simultaneously. Simulation studies show the performance of our method for parameter estimation, the recovery of item–trait relationship and the detection of DIF effects. An application is presented using data from the Eysenck Personality Questionnaire.

Key words: Bayesian adaptive Lasso, item–trait relationship, differential item functioning, multidimensional item response theory model, regularization.

1. Introduction

In modern psychological and educational tests, multiple latent traits are often assessed collectively from a bundle of item responses. To model the probability of an item response as a function of an individual's multiple latent traits and item characteristics, a variety of multidimensional item response theory (MIRT) models have been proposed (Reckase, 2009) . Most MIRT models are confirmatory, i.e., the latent traits associated with each item are pre-specified by prior knowledge (Janssen & De Boeck, 1999; Mckinley, 1989) . Various estimation methods have been developed for confirmatory MIRT models, including marginal maximum likelihood estimation (Bock et al., 1988) and Bayesian estimation (Béguin & Glas, 2001) . However, if the item–trait relationship in the confirmatory analysis is misspecified, model lack of fit and erroneous parameter estimation will occur (da Silva et al., 2019; Jin & Wang, 2014) .

A conventional approach to explore the item–trait relationship is exploratory item factor analysis (IFA; Bock et al., 1988), which is data driven and could avoid the problems caused by the erroneous item–trait specification. Exploratory IFA aims to identify the optimal number of latent traits as well as the entire item–trait relationship. Nevertheless, exploratory IFA cannot be applied without drawbacks. Since little prior knowledge or constraints on the null relations among items and latent traits are utilized in exploratory IFA, the resulting estimation may include

Correspondence should be made to Na Shan, School of Psychology & Key Laboratory of Applied Statistics of MOE, Northeast Normal University, 5268 Renmin Street, Changchun, Jilin, China. Email: shanna1981@126.com

redundant parameters. Previous studies have shown that unnecessary model parameters can yield less efficient estimators and lower the generalizability of exploratory IFA (Browne & Cudeck, 1989; Huang et al., 2017) .

The confirmatory and exploratory approaches lie on two ends of the input of item–trait relationship in MIRT models. To be more flexible on the substantive continuum, latent variable selection using regularization approaches has been developed on the basis of the confirmatory analysis. Sun et al. (2016) proposed a sparse estimation of the item–trait relationship in MIRT models by using the expectation–maximization (EM) algorithm to maximize the L_1 penalized log-likelihood. Chen (2020) used the Bayesian Lasso to estimate within-item dimensionality (loading) and residual structure in MIRT models under a partially confirmatory framework. Further developments of latent variable selection in MIRT models can be seen in Xu et al. (2022) and Zhang and Chen (2022). With the same identifiability conditions given in Sun et al. (2016), Xu et al. (2022) optimized the L_0 penalized log-likelihood by updating the model (i.e., item–trait relationship) and the model parameters simultaneously in each iteration, and the estimation accuracy of the item–trait relationship is improved. Zhang and Chen (2022) gave a quasi-Newton stochastic proximal algorithm for maximizing an objective function based on a marginal likelihood/pseudo-likelihood, possibly with constraints and/or penalties on parameters, and their method can enhance the computational efficiency of the L_1 penalized log-likelihood proposed by Sun et al. (2016).

The latent variable selection methods in MIRT models can identify the sparsity of item–trait relationship, but the above studies do not incorporate individual characteristics, such as gender and age. In heterogeneous populations, differential item functioning (DIF) is routinely examined to judge whether item responses are related to individual characteristics. Generally, DIF refers to the condition in which persons from different groups with the same latent traits have unequal probabilities of endorsing an item. As a result of DIF, a biased item provides either a constant advantage for a particular group (i.e., uniform DIF) or an advantage varying in magnitude and/or direction across the latent trait continuum (i.e., non-uniform DIF). If either type of DIF is present but not correctly addressed, biased estimates and specious treatment differences will arise, and the fairness of test is threatened (Bauer, 2017; Millsap & Everson, 1993; Teresi et al., 2008) .

In multidimensional tests with confirmatory item–trait relationship, several approaches have been proposed for DIF detection, and they are mostly multidimensional extensions of unidimensional DIF detection approaches, such as multidimensional SIBTEST (Stout et al., 1997) , multidimensional differential item functioning of items and tests (Oshima et al., 1997) , logistic regression (Mazor et al., 1998) , item response theory likelihood ratio (IRT-LR) test (Suh & Cho, 2014) and multiple indicators multiple causes (MIMIC) model (Lee et al., 2017) . These approaches have in common that a test statistic is performed for each item separately and the item is regarded as DIF if the test statistic exceeds a critical threshold. When DIF test statistics are separately computed for each item, some problems such as multiple testing and a contaminated anchor set may arise (Kim & Oshima, 2013; Woods, 2009) .

In recent years, regularization methods have been proposed for DIF detection, where DIF effects are simultaneously examined for all items on the basis of a statistical model (e.g., IRT model). Magis et al. (2015) used the Lasso (least absolute shrinkage and selection operator; Tibshirani, 1996) approach for identifying DIF in a logistic regression model and found the Lasso method outperformed the logistic regression and Mantel–Haenszel methods in terms of false positive and true positive rates for small samples. Tutz and Schauberger (2015) and Schauberger and Mair (2020) both introduced multiple DIF-inducing covariates, and then computed the penalized maximum likelihood estimators for simultaneously detecting DIF effects from different covariates in Rasch models and generalized partial credit models, respectively. Belzak and Bauer (2020) investigated Lasso regularization for identifying DIF in two-parameter logistic (2PL) models, and found the Lasso regularization had better control of type I error than the likelihood ratio test method when DIF was pervasive and sample size was large. Furthermore, Bayesian regularization

methods with a variety of penalized priors have been investigated for DIF detection in moderated nonlinear factor analysis models, and Lasso and spike-and-slab priors were found to outperform the other priors (Bauer et al., 2020; Brandt et al., 2023; Chen et al., 2022). The above regularization approaches are all aimed at unidimensional DIF detection.

For identifying DIF in the simple-structure multidimensional 2PL models, Wang et al. (2023) found that the adaptive Lasso outperformed the Lasso, and both regularization methods performed better than the likelihood ratio test in most conditions. In Wang et al. (2023)'s study, the simple structure means that each item simply measures one latent trait, and the item–trait structure is confirmatory and known in advance. In practical applications, some items may correlated with more than one latent trait in a test. As shown in Asparouhov and Muthén (2009), when nonzero cross-loadings are misspecified as zero in confirmatory factor analysis (CFA), it will result in substantial bias in the rest of the parameter estimates (i.e., overestimated factor correlations) as well as poor confidence interval coverage. In order to add modeling flexibility and reduce the bias of parameter estimates from the misspecification of factor loadings in a confirmatory measurement model, exploratory structural equation modeling (ESEM) is introduced by Asparouhov and Muthén (2009). In an ESEM model, an exploratory factor analysis (EFA) measurement model is used, instead of a CFA measurement model in a structural equation model. Examples of ESEM models include but are not limited to multiple-group EFA with measurement invariance testing, and test-retest (longitudinal) EFA.

In an MIRT model, when the item–trait relationship is not correctly specified (e.g., small cross-loadings are misspecified as zero for a simple structure), what impact will it have on subsequent parameter estimation and DIF detection? Consider a simple example with two latent traits, each of which was measured by five test items. The item discriminations were set with two cross-loadings 0.3 for each latent trait. Two groups of persons were investigated, and small uniform DIF effects were assumed for items 4 and 8. More details about the example can be seen in Section "A heuristic simple example". We found that eliminating all small cross-loadings and using a confirmatory item–trait structure resulted in substantial bias in the estimates of discriminations, DIF parameters and trait correlation as well as poor confidence interval coverage. Two exploratory methods for identifying the item–trait structure were also examined in the example. One was first identifying the item–trait structure by the EML1 method given by Sun et al. (2016) and then used the structure as confirmatory in the subsequent DIF detection; the other was our proposed method for simultaneously detecting item–trait relationship and DIF effects. Our proposed method had the smallest bias and highest credible interval coverage for the estimates of discriminations, DIF parameters and trait correlation, as shown in Table 1.

Given the effectiveness of Bayesian regularization methods for analyzing complex models and data types in psychological and behavioral studies (Brandt et al., 2023; Chen et al., 2021, 2022; Feng et al., 2017; Pan et al., 2017) , we propose a Bayesian adaptive Lasso approach for simultaneously detecting item–trait relationship and DIF effects in MIRT models. By incorporating DIF-inducing covariates in MIRT models, the detection of item–trait relationship and DIF effects can be solved jointly latent/observed variables and their interactions as variable selection for latent/observed variables and their interactions. The contribution of this study is twofold. First, we will explore the simultaneous detection of item–trait relationship and DIF effects in the context of MIRT models. Compared to Wang et al. (2023)'s study, the main difference is that they used the simple-structure multidimensional 2PL models, where the item–trait relationship was confirmatory and no cross-loadings were allowed. Our proposed method explores the item–trait relationship for non-anchor items, which can load on more than one latent trait. Second, we use Bayesian adaptive Lasso (a type of Bayesian regularization method) to estimate item discriminations in addition to DIF parameters. In Belzak and Bauer (2020) and Chen et al. (2022), no regularization was used for the baseline item discriminations, since they focused on unidimensional factor models for DIF analysis. In addition, we study DIF effects for both categorical and

metric covariates, extending the multiple types of DIF-inducing covariates investigated in the unidimensional factor models.

The rest of the article is organized as follows. First, the two-parameter compensatory MIRT models incorporating DIF effects are introduced. Then, we describe the Bayesian estimation with the adaptive Lasso for the proposed models. Next, a comprehensive simulation study is conducted and a real data analysis is reported. Finally, we conclude the article with discussion.

2. MIRT Models Incorporating DIF Effects

In multidimensional tests that intentionally measure two or more latent traits, MIRT models are often used to model the response probability of an item as a function of item characteristics and individual's multiple latent traits (Reckase, 2009) . Consider a test containing J items and K latent traits. There are N persons, who respond to all J items. In this paper, all responses are dichotomous. Let y_{ij} be the response of person i to item j, with $y_{ij} = 1$ denoting a correct response and $y_{ij} = 0$ otherwise. Following the notation of Wang et al. (2023), a two-parameter compensatory MIRT model incorporating DIF effects can be described as:

$$p(y_{ij} = 1|\boldsymbol{\theta}_i) = F(\boldsymbol{a}_j^T\boldsymbol{\theta}_i + d_j + \boldsymbol{x}_i^T\boldsymbol{\beta}_j + \boldsymbol{x}_i^T\boldsymbol{\gamma}_j\boldsymbol{\theta}_i), \tag{1}$$

where $p(y_{ij} = 1|\boldsymbol{\theta}_i)$ is the probability of a correct response for person i to item j, $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{iK})^T$ is a K-dimensional vector of latent traits for person i, $F : \boldsymbol{R} \rightarrow [0, 1]$ is a pre-specified non-decreasing function, $\boldsymbol{a}_j = (a_{j1}, \cdots, a_{jK})^T$ is a K-dimensional vector of discriminations for item j, d_j is an intercept of item j, $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{iP})^T$ is a P-dimensional covariate vector for person i that can contain both categorical variable (i.e., gender) and metric variable (i.e., age), $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{jP})^T$ is a P-dimensional vector of regression coefficients implying the main effects of each covariate on item j, and $\boldsymbol{\gamma}_j = (\gamma_{jpk})$ is a P-by-K matrix of regression coefficients with element γ_{jpk} denoting the interaction effect of the pth covariate and the kth latent trait on item j. For the illustration of DIF effects on an item, take gender as a covariate. If the β coefficient of gender is unequal to zero, this represents a consistent advantage for males or females (i.e., uniform DIF) on that item; if the γ coefficient of the interaction between gender and a latent trait is unequal to zero, this implies a varying advantage across the latent trait for males or females (i.e., non-uniform DIF).

In Eq. (1), item j is related to latent trait k if $a_{jk} \neq 0$. The latent trait vector $\boldsymbol{\theta}_i$ follows a multivariate normal distribution of $\boldsymbol{\theta}_i \sim \text{MVN}(\boldsymbol{\alpha}_i, \boldsymbol{\Psi}_i)$, where both the mean vector and the covariance matrix are person-specific. Following the models given by Bauer (2017), the mean of each latent trait k $(k = 1, \cdots, K)$ can be represented as

$$\alpha_{ki} = \alpha_{k0} + \boldsymbol{\Upsilon}_k^{'}\boldsymbol{x}_i, \tag{2}$$

where α_{k0} is the baseline mean when $\boldsymbol{x}_i = \boldsymbol{0}$, and $\boldsymbol{\Upsilon}_k$ is a P-dimensional vector that captures the linear dependence on $\boldsymbol{x}_i$. For the covariance matrix $\boldsymbol{\Psi}_i$, it can be rewritten as $\boldsymbol{\Psi}_i = \boldsymbol{\Delta}_i\boldsymbol{\Omega}_i\boldsymbol{\Delta}_i$, where $\boldsymbol{\Omega}_i$ is the correlation matrix, and $\boldsymbol{\Delta}_i$ is a diagonal matrix consisting of standard deviations. The standard deviation of each latent trait k $(k = 1, \cdots, K)$ can be expressed as a log-linear function of $\boldsymbol{x}_i$ (Chen et al., 2022) :

$$\Delta_{(kk)i} = \Delta_{(kk0)} \exp(\boldsymbol{\eta}_{(kk)}^{'}\boldsymbol{x}_i), \tag{3}$$

where $\Delta_{(kk0)}$ is the baseline standard deviation when $\boldsymbol{x}_i = \mathbf{0}$, and $\boldsymbol{\eta}_{(kk)}$ is a P-dimensional vector indicating the differences in the standard deviation as a function of $\boldsymbol{x}_i$. For each off-diagonal correlation in $\boldsymbol{\Omega}_i$, its Fisher's z-transformation can be modeled as a linear moderation function of $\boldsymbol{x}_i$, and the details can be found in Bauer (2017). In the following, $\boldsymbol{\Omega}_i$ is assumed to be constant across persons for simplicity, i.e., $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}$. Any nonzero elements in $\boldsymbol{\Upsilon}_k$ or $\boldsymbol{\eta}_{(kk)}$ indicate differences in the distribution of the individual latent traits. Such differences, also called impacts, may exist regardless of whether there is DIF or not.

To identify our model defined in Eqs. (1)-(3), some assumptions need to be satisfied. Extending Sun et al. (2016)'s conditions for latent variable selection and Wang et al. (2023)'s conditions for multidimensional DIF detection, the identifiability conditions of our model are as follows:

(1) the N-by-$(1 + P)$ matrix with rows $(\mathbf{1}, \boldsymbol{x}_1^T), \cdots, (\mathbf{1}, \boldsymbol{x}_N^T)$ is full rank.
(2) $\boldsymbol{\theta}_i$ has mean vector $\mathbf{0}$ when $\boldsymbol{x}_i = \mathbf{0}$, i.e., $\alpha_{k0} = 0$ for $k = 1, \cdots, K$.
(3) there are K DIF-free (anchor) items, loading on each dimension separately with unity loadings.

Condition (1) is in line with Wang et al. (2023) for multidimensional DIF detection. Condition (2) and the fixed loadings for each dimension in condition (3) are used to constrain the scale of baseline latent traits. Following Wang et al. (2023), Eq. (1) is identifiable when there are K DIF-free items, one for each dimension separately. Furthermore, latent variable selection in MIRT models requires K items that load on each dimension separately (Sun et al., 2016). For the third condition, without loss of generality, we assume that the first K items are DIF-free and load on each of the K dimensions separately with unity loadings, i.e., $a_{jj} = 1$ and $a_{jl} = 0$ for $1 \leq j \neq l \leq K$. Under the identifiability conditions, there are $(J-K)K$ item discriminations and J item intercepts for estimation. In addition, there are totally $(J-K)P + (J-K)KP$ additional parameters introduced to the conventional MIRT models, representing the possible uniform and non-uniform DIF effects of covariates on non-anchor items.

3. Model Estimation by Bayesian Adaptive Lasso

Regularization methods have been well developed in statistics and machine learning (Hastie et al., 2009; Wellner & Zhang, 2012; Tibshirani et al., 2021). Tibshirani (1996) introduced the famous Lasso estimates for linear regression, which are least squares estimates with the L_1 norm penalty. The L_1 penalty shrinks more weakly related coefficients to zero faster and results in sparse estimates. The Bayesian version of Lasso is later proposed by Park and Casella (2008). From a Bayesian perspective, the Lasso estimates can be interpreted as the posterior modes with a Laplace prior assigned to all coefficients. Since the Lasso procedure imposes the same penalty for all coefficients, it may lead to appreciable bias for the resulting estimates. To solve this problem, Zou (2006) developed the adaptive Lasso procedure, which uses adaptive weights for penalizing different coefficients. The adaptive Lasso imposes relatively higher penalties for zero coefficients and lower penalties for nonzero coefficients, so it shrinks zero coefficients more efficiently and produces better estimation for nonzero coefficients than Lasso does. The Bayesian adaptive Lasso is proposed by Leng et al. (2014) with independent Laplace priors imposed on different coefficients. Furthermore, many other regularization methods have been studied with sparsity as a primary driving force (Fan & Li, 2001; Polson & Sokolov, 2019; Tibshirani et al., 2021; Zhang, 2010).

Recently, the idea of regularization is introduced to the fields of psychometrics, clinical psychology, psychiatry and so on (Dwyer et al., 2018; Epskamp & Fried, 2018). In addition to the regularization methods used in latent variable selection and DIF detection, regularization especially Bayesian regularization has been successfully developed in structural equation modeling

(Chen et al., 2021; Huang, 2018; Jacobucci et al., 2016; Pan et al., 2017; Serang et al., 2017) . Compared with frequentist regularization, Bayesian regularization is highly efficient and easy to implement for complex models and data types (Alhamzawi et al., 2012; Feng et al., 2017) . Due to the advantages of Bayesian adaptive Lasso, we use it for the simultaneous detection of item–trait relationship and DIF effects in MIRT models.

3.1. Bayesian Adaptive Lasso

In the framework of frequentist statistics, regularization is a general approach for reducing the complexity of a model for meaningful interpretation. By adding a penalty term to the usual likelihood, regularization approaches can shrink unimportant model parameters to exactly zero. Suppose the observed data are denoted by $\boldsymbol{y}$, and the set of parameters in a model M is denoted by $\boldsymbol{\beta}$ with elements β_k $(k = 1, \cdots, r)$. The adaptive Lasso approach uses the following objective function:

$$PL(\boldsymbol{\beta}|M) = log(p(\boldsymbol{y}|\boldsymbol{\beta}, M)) + \sum_{k=1}^{r} \lambda_k|\beta_k| = LL(\boldsymbol{\beta}|M) + \sum_{k=1}^{r} \lambda_k|\beta_k|,$$

where $PL(\boldsymbol{\beta}|M)$ and $LL(\boldsymbol{\beta}|M)$ are respectively the penalized and the usual log-likelihoods based on model M, and $\lambda_k \geq 0$ is a penalty parameter for β_k. A larger λ_k tends to increase the penalty for β_k. Using adaptive weights for penalizing different coefficients, the adaptive Lasso can shrink zero coefficients more efficiently and produce better estimation for nonzero coefficients than Lasso (Zou, 2006) .

The crucial quantity for Bayesian statistics is the posterior distribution $p(\boldsymbol{\beta}|\boldsymbol{y}, M) \propto p(\boldsymbol{y}|\boldsymbol{\beta}, M) \times p(\boldsymbol{\beta}|M)$, where $p(\boldsymbol{\beta}|M)$ is the prior distribution. Compared with a frequentist approach, the prior $p(\boldsymbol{\beta}|M)$ is an important connection to a regularization approach such as the adaptive Lasso. Following Leng et al. (2014), the adaptive Lasso estimates can be interpreted under the Bayesian framework when β_ks are assigned independent Laplace priors $\frac{\lambda_k}{2} e^{-\lambda_k|\beta_k|}$. For a small value of λ_k, the Laplace distribution is wide and no shrinkage is imposed. As the value of λ_k increases, the probability density function tends to be more concentrated around zero, leading to a larger penalty (Pan et al., 2017) . Moreover, the Bayesian framework provides a flexible way of estimating the penalty parameters, and hyperpriors can be used for the λ_ks. Specifically, λ_ks for Bayesian adaptive Lasso are assigned with the Gamma priors $\lambda_k \sim \text{Gamma}(\alpha_{k0}, \delta_{k0})$ $(k = 1, \cdots, r)$, where α_{k0} and δ_{k0} are hyperparameters with pre-assigned values. Following the suggestions of previous studies (Brandt et al., 2023; Chen et al., 2022; Feng et al., 2017) , dispersed hyperpriors are often adopted.

3.2. Bayesian Model Implementation

To implement Bayesian adaptive Lasso for identifying item–trait relationship and DIF effects, independent Laplace priors are assigned to the discriminations and DIF parameters of the last $J - K$ items. For the other parameters, commonly used priors are adopted for convenience. The priors and hyperpriors are given below.

For each element of the discrimination vectors $\boldsymbol{a}_{K+1}, \cdots, \boldsymbol{a}_J$, independent Laplace priors are assigned and expressed as:

$$p(\boldsymbol{a}_{K+1}, \cdots, \boldsymbol{a}_J) \propto \exp\left(-\sum_{j=K+1}^{J}\sum_{k=1}^{K} \lambda_{ajk}|a_{jk}|\right),$$

where λ_{ajk} is the penalty parameter for a_{jk}.

For each intercept d_j $(j = 1, \cdots, J)$, the normal prior is adopted as:

$$d_j \sim N(\mu_{dj0}, \sigma^2_{dj0}),$$

where μ_{dj0} and σ^2_{dj0} are hyperparameters with pre-assigned values, denoting the mean and variance of the normal distribution.

For each uniform DIF parameter in $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J$, independent Laplace priors can be expressed as:

$$p(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J) \propto \exp\left(-\sum_{j=1}^{J}\sum_{p=1}^{P}\lambda_{\beta jp}|\beta_{jp}|\right),$$

where $\lambda_{\beta jp}$ is the penalty parameter for β_{jp}.

For each non-uniform DIF parameter in $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_J$, independent Laplace priors can be expressed as:

$$p(\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_J) \propto \exp\left(-\sum_{j=1}^{J}\sum_{p=1}^{P}\sum_{k=1}^{K}\lambda_{\gamma jpk}|\gamma_{jpk}|\right),$$

where $\lambda_{\gamma jpk}$ is the penalty parameter for γ_{jpk}.

For the penalty parameters λ_{ajk}, $\lambda_{\beta jp}$ and $\lambda_{\gamma jpk}$, the Gamma priors can be assigned as:

$$\begin{aligned}
\lambda^2_{ajk} &\sim \text{Gamma}\left(\alpha_{ajk0}, \delta_{ajk0}\right), \\
\lambda^2_{\beta jp} &\sim \text{Gamma}\left(\alpha_{\beta jp0}, \delta_{\beta jp0}\right), \\
\lambda^2_{\gamma jpk} &\sim \text{Gamma}\left(\alpha_{\gamma jpk0}, \delta_{\gamma jpk0}\right),
\end{aligned}$$

where α_{ajk0}, δ_{ajk0}, $\alpha_{\beta jp0}$, $\delta_{\beta jp0}$, $\alpha_{\gamma jpk0}$ and $\delta_{\gamma jpk0}$ are hyperparameters whose values are pre-assigned.

For each $\Delta_{(kk0)}$ $(j = 1, \cdots, K)$, the half-Cauchy prior is assigned as (Gelman, 2006) :

$$\Delta_{(kk0)} \sim C^+(0, \iota_{k0}),$$

where ι_{k0} is the hyperparameter for the half-Cauchy distribution.

The LKJ correlation distribution is used for the prior of $\boldsymbol{\Omega}$ with the density (Lewandowski et al., 2009)

$$\text{LkjCholesky}(\boldsymbol{\Omega}) \propto \det(\boldsymbol{\Omega})^{\nu-1},$$

where $\det(\cdot)$ denotes the determinant and ν is the shape parameter.

For each element Υ_{kp} in $\boldsymbol{\Upsilon}_k$, the normal prior is adopted as:

$$\Upsilon_{kp} \sim N(\mu_{\Upsilon_{kp}0}, \sigma^2_{\Upsilon_{kp}0}),$$

where $\mu_{\Upsilon_{kp}0}$ and $\sigma^2_{\Upsilon_{kp}0}$ are hyperparameters with pre-assigned values.

For each element $\eta_{(kk)p}$ in $\boldsymbol{\eta}_{(kk)}$ $(k = 1, \cdots, K)$, the normal prior is adopted as:

$$\eta_{(kk)p} \sim N(\mu_{\eta_{(kk)p}0}, \sigma^2_{\eta_{(kk)p}0}),$$

where $\mu_{\eta_{(kk)p}0}$ and $\sigma^2_{\eta_{(kk)p}0}$ are hyperparameters with pre-assigned values.

With the prior and hyperprior distributions given above, the posterior inference can be conducted by sampling from the joint posterior distribution, and the posterior means are used to estimate the unknown parameters. Though the joint posterior distribution is intractable in general, the Bayesian inference can be feasibly implemented in an available Bayesian software package, such as Stan (Carpenter et al., 2017) or Jags (Plummer, 2017) . In our study, the rstan package (Carpenter et al., 2017; Stan Development, 2023) in R (R Core Team, 2022) was used to implement the Bayesian adaptive Lasso estimation. When posterior means are used as estimates, the Bayesian adaptive Lasso does not shrink any parameter to exactly zero, and a variable selection criterion should be applied for determining the significance of the unknown parameters. As proposed by Brandt et al. (2023), the 95% posterior credible intervals (CIs) were used in this paper.

4. A Simple Heuristic Example

In this section, a simple hypothetical example is provided to illustrate the motivation of our study. Consider two latent traits, each of which was measured by five test items. Two groups of persons were investigated and coded by a binary covariate, with 0 for the reference group and 1 for the focal group. The mean vector of latent traits in the reference group was set as $(0, 0)'$, and the mean vector of latent traits in the focal group was set as $(0.5, -0.5)'$. For both groups, the variances of latent traits were 1 and the correlation between latent traits was 0.5. The item intercepts were all set at 0 for simplicity, and the item discriminations were given as

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0.3 & 0.3 \\ 0 & 1 & 0 & 0 & 0.3 & 0.3 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

We assumed that items 4 and 8 had uniform DIF effects with $\beta_{41} = 0.3$ and $\beta_{81} = 0.3$. The sample size was 500, divided evenly into the two groups. Data were generated with 50 replications.

Our proposed model was compared with two alternative models. The first one used a confirmatory simple-structure MIRT model for DIF detection, with small cross-loadings fixed to 0; the second one first identified the item–trait structure by the EML1 method given by Sun et al. (2016), and then used the structure as confirmatory for DIF detection. In the three models, except for the item discriminations, the other model parameters were estimated using the same Bayesian priors.

To evaluate the performance of parameter estimation, the mean absolute bias and CI coverage were computed for each model, as shown in Table 1. The formal is the average absolute values of bias across converged replications and interested parameters. The latter is calculated as the number of converged replications where the equal-tailed 95% CIs covered the true values of the interested parameters divided by the total number of converged replications and interested parameters. From these results, we found that eliminating small cross-loadings in item–trait structure resulted in substantial bias in the estimates of item discriminations, DIF parameters and trait correlation as well as poor CI coverage. When the item–trait structure was first identified by the EML1

TABLE 1.
Mean absolute bias (CI coverage) of parameter estimates in the simple example

Model	a_j	d_j	β_j	Υ_k	$\Delta_{(kk0)}$	$\eta_{(kk)}$	$\Omega_{(12)}$
Confirmatory DIF	0.190(0.644)	0.013(0.962)	0.124(0.903)	0.046(0.950)	0.040(0.920)	0.013(0.960)	0.153(0.140)
EML1 DIF	0.094(0.906)	0.014(0.966)	0.040(0.975)	0.021(0.950)	0.035(0.950)	0.020(0.970)	0.023(0.960)
our joint DIF	0.031(0.974)	0.016(0.967)	0.016(0.990)	0.010(0.939)	0.056(0.990)	0.021(0.980)	0.004(0.980)

method, most mean absolute bias decreased and the CI coverage improved. Our proposed model performed best among the three models, with the smallest bias and highest CI coverage for item discriminations, DIF parameters and trait correlation.

5. Simulation Studies

Two simulation studies were conducted to evaluate the empirical performance of the Bayesian adaptive Lasso for uniform DIF (study 1) and non-uniform DIF (study 2) conditions. For both studies, the model defined in Eqs. (1)-(3) was used. The total number of items J was fixed at 15, and the number of latent traits K was fixed at 2. Table 2 gives the two discriminations for each item that reflected a common range of them, and the item intercepts were generated from the standard normal distribution (Wang et al., 2023) . Four covariates x_{i1}, x_{i2}, x_{i3} and x_{i4} were considered. x_{i1} and x_{i2}, having DIF effects on some items, were independently generated from the standard normal distribution and the Bernoulli distribution with a success probability 0.5. x_{i3} and x_{i4}, having no DIF effects on any items, were jointly generated from a multivariate normal distribution with a mean vector $\mathbf{0}$ and a correlation matrix with off-diagonal elements 0.5. The baseline means of latent traits were $\alpha_{10} = \alpha_{20} = 0$ for identification, and the mean impacts were set at $\boldsymbol{\Upsilon}_1 = (0, 0.5, 0, 0)'$ and $\boldsymbol{\Upsilon}_2 = (0, -0.5, 0, 0)'$, indicating the latent mean differences only related to the second covariate. The baseline standard deviations were set at $\Delta_{(110)} = \Delta_{(220)} = 1$, and no standard deviation impacts were set with $\boldsymbol{\eta}_{(11)} = \boldsymbol{\eta}_{(22)} = (0, 0, 0, 0)'$. The correlation between two latent traits $\boldsymbol{\Omega}_{(12)}$ was set at 0.5, reflecting a moderate degree of correlation.

Three factors were manipulated: (a) the sample size N, (b) the percentage of DIF items, and (c) the magnitude of DIF. Two levels of sample size were evaluated: $N = 500$ and $N = 1000$, which was in line with previous studies (Sun et al., 2016; Xu et al., 2022). Two percentages of DIF items (20% and 60%) and two levels of the magnitude of DIF (small and large) were considered, and these choices were similar to the study of Wang et al. (2023).

We evaluated the performance of our method in terms of (1) the accuracy of parameter estimation, (2) the correct rate (CR), false positive rate (FPR) and false negative rate (FNR) for latent variable selection, and (3) the true positive rate (TPR) and FPR for DIF detection. The results were computed on the basis of 50 replications for each condition. DIF effects were kept constant across replications with a given condition, which can avoid the mixture of within- and between-condition variability of DIF effects (Belzak and Bauer, 2020; Wang et al., 2023). For the accuracy of parameter estimation, the mean-squared error (MSE) for each parameter is computed as

$$\text{MSE}(\kappa) = \frac{1}{Z}\sum_{z=1}^{Z}(\hat{\kappa}^z - \kappa)^2,$$

where $\hat{\kappa}^z$ denotes an estimate of κ based on the zth converged replication, and Z is the number of converged replications. For summarizing our simulation results, MSE is displayed by each parameter type below. For example, the MSE for item discriminations is the average MSE of all estimated item discrimination parameters. The CR for latent variable selection is defined by the recovery of the unknown elements in the incidence matrix $\Xi = (\xi_{jk})$, where $\xi_{jk} = I(a_{jk} \neq 0)$, and it is given as

$$\text{CR} = \frac{1}{Z(J-K)K}\sum_{z=1}^{Z}\sum_{j=K+1}^{J}\sum_{k=1}^{K} I(\hat{\xi}_{jk}^z = \xi_{jk}),$$

TABLE 2.
Simulated true item parameters

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a_1	1.00	0	1.40	1.20	0.80	0.60	0	0	0	0	0.60	0.80	1.00	1.20	1.40
a_2	0	1.00	0	0	0	0	1.40	1.20	0.80	0.60	1.40	1.20	1.00	0.80	0.60
d	-1.48	1.58	-0.96	-0.92	-2.00	-0.27	-0.32	-0.63	-0.11	0.43	-0.78	-1.29	-0.78	0.01	-0.15

where $\hat{\xi}_{jk}^{z}$ is an estimate of the true ξ_{jk} based on the zth converged replication. FPR for latent variable selection refers to the ratio of incorrectly detected nonzero incidence relations among all true zero incidence relations and converged replications, and the FNR for latent variable selection refers to the ratio of incorrectly detected zero incidence relations among all true nonzero incidence relations and converged replications. For DIF detection, in order to avoid the mixture of the impact of different covariates on DIF, TPR and FPR are calculated in terms of item–covariate combinations (Chen et al., 2022; Schauberger & Mair, 2020) . Specifically, TPR and FPR are calculated as the proportions of item–covariate combinations in which a covariate is detected as having significant uniform or non-uniform DIF parameters for an item across all converged replications and item–covariate combinations that do or do not have DIF, respectively.

In the simulation studies, data generation and parameter estimation were all implemented in R statistical programming software. The R codes are available at https://github.com/Shann285/LdDIFMIRT. We ran all R codes on the Windows 10 64-bit platform with an Inter(R) Core(TM) i9-9900 CPU at 3.10 GHz and 32 GB memory. Our Bayesian models were fitted with 3 chains of Hamiltonian Markov Chain Monte Carlo (MCMC) samples using the R package rstan. Each Hamiltonian MCMC chain had 4000 iterations with the first 2000 iterations as a burn-in period. The convergence of the chains was monitored by zero divergent transitions in the sampling process and "Rhat" indices less than 1.05. The convergence rates varied depending on the data and prior assignments, and they can be improved by adding the number of iterations and thinning (Chen et al., 2022).

5.1. Simulation Study 1

In this study, only uniform DIF effects were considered, i.e., all γ coefficients were fixed at 0. The β coefficients were set as 0.3 and 0.6 for small and large magnitude of DIF, respectively. Specifically, β_{41}, β_{82}, $\beta_{13,1}$ and $\beta_{13,2}$ were equal to 0.3 (or 0.6) for the 20% DIF condition, and β_{31}, β_{41}, β_{51}, β_{72}, β_{82}, β_{92}, $\beta_{12,1}$, $\beta_{12,2}$, $\beta_{13,1}$, $\beta_{13,2}$, $\beta_{14,1}$ and $\beta_{14,2}$ were equal to 0.3 (or 0.6) for the 60% DIF condition. These choices were similar to those used in Wang et al. (2023).

Following the suggestions of the existing literature (Feng et al., 2017; Pan et al., 2017; Chen et al., 2022; Brandt et al., 2023), the prior and hyperprior distributions were chosen as follows: the normal priors used the normal distribution $N(0, 2^2)$, the hyperpriors for the penalty parameters were the Gamma distribution Gamma(9, 3), the half-Cauchy distribution was $C^{+}(0, 2.5)$, and the LKJ correlation distribution was set with $\nu = 2$. The initial values were generated similarly to those used in Chen et al. (2022). DIF in an item due to a specific covariate was assumed if the 95% CI for the respective element in β did not include zero.

The Bayesian adaptive Lasso for uniform DIF detection achieved preferable convergence rates, all above 95%. Though non-convergence might be modified with further adjustments, we did not do this and simply used the converged replications for the results. The running times for different conditions varied, mainly depending on the sample size. For the sample size $N = 500$, the average CPU times were less than 1000 s for each condition. The average CPU times increased to more than 2000 s when the sample size was $N = 1000$. The specific values of the average CPU times are shown in Table 8 of Appendix A.

Figure 1 shows MSE as a combination of squared bias and variance for estimating item discriminations a, item intercepts d, uniform DIF parameters β, mean impacts Υ_k, baseline standard deviations $\Delta_{(kk0)}$, standard deviation impacts $\eta_{(kk)}$, and the correlation between latent traits denoted as $\Omega_{(12)}$. And the MSEs for each parameter estimate are provided in Table 9 of Appendix A. We found that most model parameters could be recovered well. When DIF percentage was 60%, the bias of most estimates increased. In contrast, the magnitude of DIF had little influence on the estimates. The estimates of mean impacts, baseline standard deviations, standard deviation impacts and the correlation changed little under different percentage and magnitude of DIF effects.

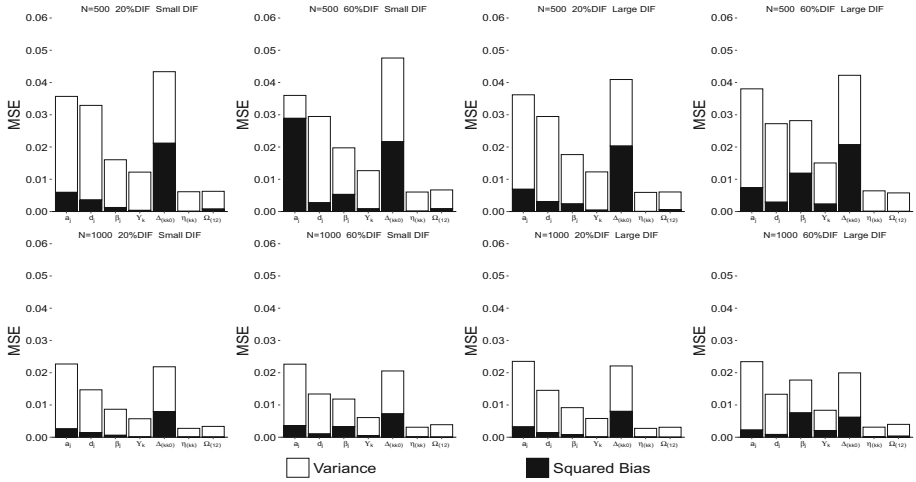

FIGURE 1.
MSEs of the model parameter estimates in study 1.

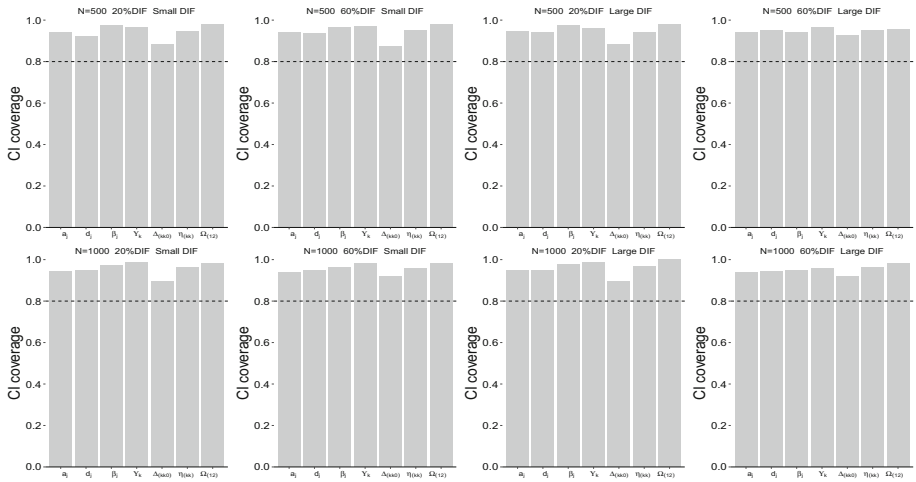

FIGURE 2.
CI coverage for different parameters in study 1.

As sample size increased, most MSEs reduced. The CI coverage rates were calculated to evaluate the uncertainty of the estimates for the population parameters, as shown in Fig. 2. The coverage rates under all conditions were above 80%, which were similar to the results of Brandt et al. (2023) and Chen et al. (2022).

Table 3 summarizes the results for recovering the incidence matrix and detecting DIF effects over 50 independent datasets in each simulated condition. For the incidence matrix Ξ, the CRs, FPRs and FNRs were calculated in each condition. The CRs were all above 0.98, and the FPRs and FNRs did not exceed 0.05. The percentage and magnitude of DIF had little impacts on the recovery of Ξ. For DIF detection, TPRs and FPRs are shown at the bottom of Table 3. Consistent with previous research (Belzak and Bauer, 2020; Schauberger and Mair, 2020), small magnitude of DIF effects was difficult to detect. The TPRs reduced when DIF percentage was 60%. All TPRs grew as sample size increased. Our method produced acceptable FPRs for any study conditions,

TABLE 3.
Results of latent variable selection and DIF detection in study 1

		Small DIF				Large DIF			
		20% DIF		60% DIF		20% DIF		60% DIF	
		$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
Ξ	CR	0.982	0.988	0.982	0.984	0.986	0.989	0.982	0.987
	FPR	0.036	0.029	0.034	0.044	0.029	0.029	0.026	0.036
	FNR	0.009	0.005	0.010	0.003	0.008	0.003	0.015	0.002
DIF	TPR	0.422	0.568	0.302	0.467	0.734	0.969	0.635	0.781
	FPR	0.023	0.026	0.024	0.029	0.023	0.025	0.033	0.034

which was in concert with previous findings that regularization methods have good control of type I errors (Brandt et al., 2023; Chen et al., 2022; Wang et al., 2023).

5.2. *Simulation Study 2*

In the second study, non-uniform DIF effects were evaluated for DIF items. The item parameters were the same as simulation study 1. For DIF parameters, β were set as 0.3 and 0.6 for small and large magnitude DIF, and γ were set as -0.3 and -0.6 for small and large magnitude DIF. For the 20% DIF condition, β_{41}, β_{82}, $\beta_{13,1}$ and $\beta_{13,2}$ were equal to 0.3 (or 0.6), and γ_{411}, γ_{822}, $\gamma_{13,11}$ and $\gamma_{13,22}$ were equal to -0.3 (or -0.6). For the 60% DIF condition, β_{31}, β_{41}, β_{51}, β_{72}, β_{82}, β_{92}, $\beta_{12,1}$, $\beta_{12,2}$, $\beta_{13,1}$, $\beta_{13,2}$, $\beta_{14,1}$ and $\beta_{14,2}$ were equal to 0.3 (or 0.6), and γ_{311}, γ_{411}, γ_{511}, γ_{722}, γ_{822}, γ_{922}, $\gamma_{12,11}$, $\gamma_{12,22}$, $\gamma_{13,11}$, $\gamma_{13,22}$, $\gamma_{14,11}$ and $\gamma_{14,22}$ were equal to -0.3 (or -0.6). The above choices were similar to the study of Wang et al. (2023).

Different from the uniform DIF models considered in study 1, the non-uniform DIF models had unknown γ coefficients for non-anchor items. Since the non-uniform DIF models had more DIF parameters than the uniform DIF models, stronger penalty was used to achieve adequate convergence rates. The hyperpriors of the penalty parameters were set with Gamma(27, 3). The other priors and initial values were the same as those used in study 1.

The Bayesian adaptive Lasso for non-uniform DIF models achieved reasonable convergence rates, ranging from 84% to 98% with an average above 92%. Convergence below 90% occurred in the small DIF magnitude and high DIF percentage conditions, which were in concert with Chen et al. (2022). Only the converged replications were used for evaluating the estimated results. The running times for non-uniform DIF models all exceeded 3500 s. For the sample size $N = 500$, the average CPU times were about an hour. When the sample size was $N = 1000$, the average CPU times were nearly two hours. The average CPU times for non-uniform DIF conditions are also shown in Table 8 of Appendix A.

Figure 3 shows the MSEs of the estimated parameters for non-uniform DIF conditions. And the MSEs for each parameter estimate can be found at https://github.com/Shann285/LdDIFMIRT. The MSEs of item discriminations were larger than those of other parameters and the MSEs in study 1, reflecting the increased uncertainty due to the unknown γ coefficients. The bias of most estimates increased when DIF percentage was 60%. The magnitude of DIF had no obvious impact on the estimates when DIF percentage was low, but may lead to larger bias when DIF was pervasive. Most MSEs reduced as sample size increased. The CI coverage rates for all conditions in study 2 were above 80%, as shown in Fig. 4.

Table 4 presents the results for the incidence matrix and DIF detection under the non-uniform DIF conditions. For recovering the incidence matrix, the overall CRs were above 94% and the

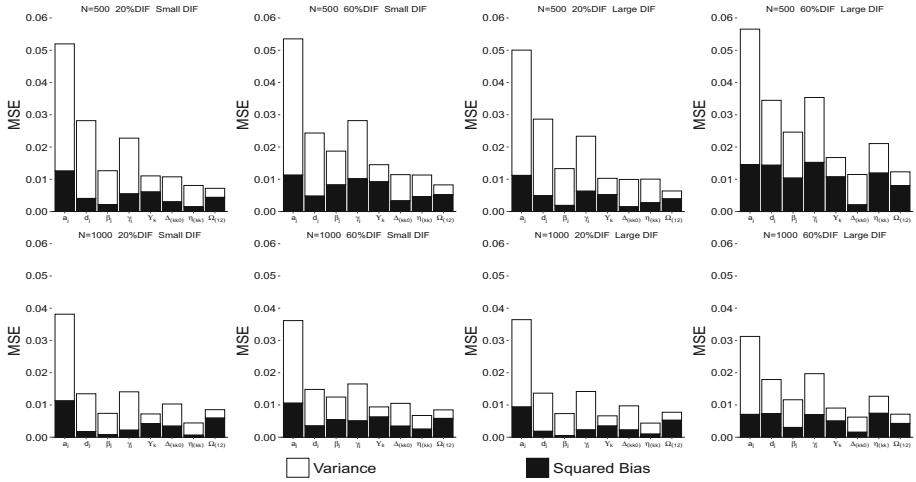

FIGURE 3.
MSEs of the model parameter estimates in study 2.

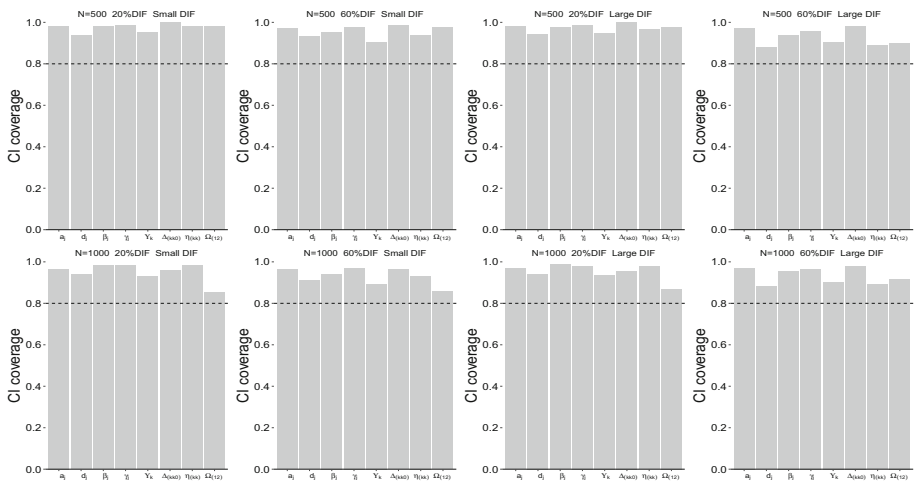

FIGURE 4.
CI coverage for different parameters in study 2.

FPRs did not exceed 0.02, but the FNRs were slightly above 0.05 when the DIF percentage was 60%. Small magnitude of DIF effects was more difficult to detect in the non-uniform DIF conditions, and this may be due to the increased number of model parameters. In contrast, the TPRs for large DIF conditions were larger than those in study 1, indicating that large γ coefficients were helpful for identifying DIF items. Similar to the results of study 1, the TPRs reduced when DIF was pervasive, and all TPRs grew as sample size increased. We had acceptable control of the FPRs, slightly exceeding 0.05 in the 60% DIF percentage.

TABLE 4.
Results of latent variable selection and DIF detection in study 2

		Small DIF				Large DIF			
		20% DIF		60% DIF		20% DIF		60% DIF	
		$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
Ξ	CR	0.967	0.988	0.962	0.984	0.971	0.987	0.943	0.985
	FPR	0.003	0.013	0.003	0.018	0.005	0.016	0.003	0.013
	FNR	0.046	0.012	0.053	0.015	0.040	0.011	0.082	0.016
DIF	TPR	0.296	0.511	0.138	0.278	0.864	1.000	0.682	0.948
	FPR	0.034	0.039	0.040	0.057	0.040	0.045	0.057	0.060

6. Real Data Analysis

A real data set from the Eysenck Personality Questionnaire (EPQ) data given in Eysenck and Barrett (2013) was used to further illustrate the performance of our method. Three factors of Psychoticism (P), Extraversion (E) and Neuroticism (N) were initially investigated by Xu et al. (2022) from this data. Since the psychometric weaknesses in the P scale of the EPQ, only Extraversion (E) and Neuroticism (N) were focused in our analysis. In line with Xu et al. (2022), two items in E were deleted, because their corrected item–total correlation values were less than 0.2 (Kline, 1986). As a result, 42 items were selected, including 19 items corresponding to E and 23 items corresponding to N. The initial design of EPQ is confirmative and each item is associated to only one factor. The used items and their original indices are listed in Table 5.

Two covariates were considered for detecting DIF effects, among of which age was a continuous variable representing age of the person and gender was a binary categorical variable. Moreover, only the ages of 18, 19, 20 and 21 were included, since the number of persons in other age groups was small. After eliminating persons with missing data, our analysis was based on 843 individuals from Canada. The model defined in Eqs. (1)-(3) with $K = 2$ and $P = 2$ was applied to analyze the real data. In addition, the person-specific correlation between latent traits was modeled as

$$\boldsymbol{\Omega}_{(12)i} = \frac{\exp(2(\omega_{(12)0} + \boldsymbol{\omega}_{(12)}^{'}\boldsymbol{x}_i)) - 1}{\exp(2(\omega_{(12)0} + \boldsymbol{\omega}_{(12)}^{'}\boldsymbol{x}_i)) + 1},$$

which indicated that the Fisher's z-transformation of $\boldsymbol{\Omega}_{(12)i}$ was a linear function of $\boldsymbol{x}_i$ (Bauer, 2017). The priors of the above model parameters also used the normal distribution $N(0, 2^2)$, and the other priors and initial values for the real data analysis were similar to those used in the simulation studies. For model identification, following the study of Xu et al. (2022), items 1 and 20 were designated for E and N separately, and they were assumed as DIF-free items. Both uniform and non-uniform DIF models were fitted, respectively. For each model estimation, three chains of Hamiltonian MCMC samples were used, and each chain had 5000 iterations with the first 2500 iterations as burn-in. The convergence of the chains was checked. We only reported the results for uniform DIF detection, as the vast majority of the γ coefficients in the non-uniform DIF model were not significant. The running time for the uniform DIF detection was nearly two hours.

Table 6 shows the estimated item discriminations and intercepts after rescaling the baseline latent trait variances to be unity. We found that most items remained associated with one single

TABLE 5.
The Eysenck Personality Questionnaire with items for E and N

1	E5	Are you a talkative person?
2	E10	Are you rather lively?
3	E14	Can you usually let yourself go and enjoy yourself at a lively party?
4	E17	Do you enjoy meeting new people?
5(R)	E21	Do you tend to keep in the background on social occasions?
6	E25	Do you like going out a lot?
7(R)	E29	Do you prefer reading to meeting people?
8	E32	Do you have many friends?
9	E36	Would you call yourself happy-go-lucky?
10	E40	Do you usually take the initiative in making new friends?
11(R)	E42	Are you mostly quiet when you are with other people?
12	E45	Can you easily get some life into a rather dull party?
13	E49	Do you like telling jokes and funny stories to your friends?
14	E52	Do you like mixing with people?
15	E60	Do you like doing things in which you have to act quickly?
16	E64	Do you often take on more activities than you have time for?
17	E70	Can you get a party going?
18	E82	Do you like plenty of bustle and excitement around you?
19	E86	Do other people think of you as being very lively?
20	N3	Does your mood often go up and down?
21	N7	Do you ever feel “just miserable" for no reason?
22	N12	Do you often worry about things you should not have done or said?
23	N15	Are you an irritable person?
24	N19	Are your feelings easily hurt?
25	N23	Do you often feel “fed-up"?
26	N27	Are you often troubled about feelings of guilt?
27	N31	Would you call yourself a nervous person?
28	N34	Are you a worrier?
29	N38	Do you worry about awful things that might happen?
30	N41	Would you call yourself tense or “highly-strung"?
31	N47	Do you worry about your health?
32	N54	Do you suffer from sleeplessness?
33	N58	Have you often felt listless and tired for no reason?
34	N62	Do you often feel life is very dull?
35	N66	Do you worry a lot about your looks?
36	N68	Have you ever wished that you were dead?
37	N72	Do you worry too long after an embarrassing experience?
38	N75	Do you suffer from “nerves"?
39	N77	Do you often feel lonely?
40	N80	Are you easily hurt when people find fault with you or the work you do?
41	N84	Are you sometimes bubbling over with energy and sometimes very sluggish?
42	N88	Are you touchy about some things?

“R" Denotes the negatively worded items in the original questionnaire.

TABLE 6.
The estimated item discriminations and intercepts for the real data

	A		*d*		*A*		*d*		*A*		*d*
1	1.044*	0.000	0.689	15	0.463*	-0.154*	0.222	29	-0.027	0.734*	0.293
2	0.933*	0.003	1.388	16	0.315*	0.119*	0.323	30	0.034	0.855*	-0.895
3	0.841*	-0.097	1.180	17	1.105*	-0.051	0.445	31	0.077	0.469*	0.193
4	0.665*	-0.167*	1.555	18	0.654*	-0.085	0.865	32	0.004	0.490*	-0.264
5	0.930*	-0.201*	0.460	19	1.007*	-0.048	0.872	33	-0.048	0.660*	0.487
6	0.519*	0.029	0.794	20	0.000	0.807*	0.523	34	-0.173*	0.521*	-0.638
7	0.643*	-0.086	1.259	21	-0.015	0.562*	0.566	35	0.079	0.621*	0.459
8	0.641*	-0.042	1.343	22	-0.042	0.821*	1.309	36	-0.132*	0.388*	0.018
9	0.528*	-0.273*	-0.025	23	0.031	0.531*	-0.738	37	−0.136*	0.811*	0.433
10	0.872*	-0.085	0.355	24	-0.047	0.825*	0.718	38	-0.031	1.012*	-0.529
11	1.028*	-0.056	0.612	25	-0.086	0.866*	0.270	39	-0.160*	0.689*	-0.157
12	1.138*	0.011	-0.020	26	-0.022	0.871*	0.219	40	-0.124	0.720*	0.582
13	0.457*	-0.074	1.253	27	-0.174*	0.673*	-0.343	41	-0.014	0.526*	1.157
14	1.067*	-0.148	1.840	28	-0.045	1.117*	0.715	42	0.077	0.465*	1.223

$*$ denotes the significance of item discriminations.

trait. There are more items associated with both latent traits than those found in Xu et al. (2022), and most of the cross-loadings were sensible. For example, item 5 (E21) was also related to neuroticism, the same as the results of Sun et al. (2016) and Xu et al. (2022). Item 9 (E36) and item 15 (E60) were also related to neuroticism, and these were in line with Xu et al. (2022). Item 27 (N31) was also related to extraversion and it was consistent with Sun et al. (2016). Moreover, item 39 (N77 'Do you often feel lonely?') was newly found to be related to both extraversion and neuroticism, which was in accordance with Buecker et al.'s (2020) findings that extraversion and neuroticism were significantly related to loneliness, and the average lonely person was rather introverted and neurotic than the average non-lonely person. Moreover, the mean impact of age on the trait N was −0.243, indicating that the average neuroticism in males was significantly lower than that in females. But the other impacts were not significant.

For DIF detection, our results are compared with a commonly used IRT-LR test (Suh and Cho, 2014), where both age and gender were considered as grouping covariates. The IRT-LR test was implemented by the R package *mirt*, and items 1 and 20 were also assigned as anchor items for DIF detection in IRT-LR. The results of Bayesian adaptive Lasso and IRT-LR for DIF detection are provided in Table 7. Most DIF items identified by Bayesian adaptive Lasso were also identified as DIF by IRT-LR. In addition, IRT-LR identified more DIF items, especially for gender. As pointed out by previous studies (Belzak and Bauer, 2020; Wang et al., 2023), IRT-LR leads to high false positive rates when DIF is pervasive.

7. Discussion

Regularization methods for latent variable selection or DIF detection come into use about a decade ago. For either of the two purposes, regularization methods often outperform the corresponding conventional methods. In frequentist statistics, the success of the regularization methods depends on choosing the regularization (penalty) parameters, and some criteria, such as Bayesian information criterion (BIC) and cross-validation (CV), can be used to select the optimal regularization parameters for model fitting. From the view of Bayesian statistics, the regularization parameters can be considered as random and assigned with appropriate prior distributions. By

TABLE 7.
Comparisons of DIF detection for BaLasso and IRT-LR in the real data

	Age			Age			Gender			Gender	
Item	BaLasso	IRT-LR	Item	BaLasso	IRT-LR	Item	BaLasso	IRT-LR	Item	BaLasso	IRT-LR
1	✓	✓	22	✓	✗	1	✓	✓	22	✓	✗
2	✓	✓	23	✓	✓	2	✓	✓	23	✓	✓
3	✓	✗	24	✓	✗	3	✓	✓	24	✗	✗
4	✓	✓	25	✓	✗	4	✗	✗	25	✓	✓
5	✓	✓	26	✓	✓	5	✓	✓	26	✓	✗
6	✓	✓	27	✗	✗	6	✓	✓	27	✓	✗
7	✓	✓	28	✓	✓	7	✓	✓	28	✗	✗
8	✓	✓	29	✓	✓	8	✓	✓	29	✗	✗
9	✓	✓	30	✓	✓	9	✓	✓	30	✓	✓
10	✓	✓	31	✓	✓	10	✓	✓	31	✗	✓
11	✓	✓	32	✗	✗	11	✓	✓	32	✓	✓
12	✓	✓	33	✓	✓	12	✗	✗	33	✓	✓
13	✓	✓	34	✗	✗	13	✓	✓	34	✗	✓
14	✓	✓	35	✓	✓	14	✓	✓	35	✓	✓
15	✓	✓	36	✓	✓	15	✗	✗	36	✗	✗
16	✓	✓	37	✓	✓	16	✗	✓	37	✓	✗
17	✓	✓	38	✓	✓	17	✗	✗	38	✓	✗
18	✓	✓	39	✓	✓	18	✓	✓	39	✓	✓
19	✓	✓	40	✓	✓	19	✓	✓	40	✗	✗
20	✓	✓	41	✓	✓	20	✓	✓	41	✗	✗
21	✓	✓	42	✗	✓	21	✗	✗	42	✓	✗

✓ and ✗ denote the items identified as DIF free and DIF for the corresponding covariates, respectively.

incorporating DIF-inducing covariates in MIRT models, we propose a Bayesian adaptive Lasso approach for simultaneously detecting item–trait relationship and DIF effects.

Our simulation studies showed that our proposed method can produce good parameter estimates, and performed well for the recovery of item–trait relationship. For uniform DIF detection, our method had acceptable TPRs and good control of FPRs. These results are similar to the studies of Bauer et al. (2020) and Wang et al. (2023). For non-uniform DIF detection, FPRs inflated a little than the uniform DIF conditions, since the non-uniform DIF models include more model parameters to be estimated. Moreover, it should be noted that both Tables 3 and 5 show slightly increased FPRs when the sample size increased. Though this phenomenon is similar to some existing researches (Belzak and Bauer, 2020; Brandt et al., 2023; Chen et al., 2022; Wang et al., 2023), the model (variable) selection consistency in latent variable models, especially in item response theory models, needs to be further investigated and theoretically justified in future. In addition, the DIF effects of multiple covariates are simultaneously detected in a multidimensional latent trait model, and it is beneficial for alleviating the problems caused by multicollinearity, where using a method repeatedly for different covariates is not appropriate.

It is meaningful to investigate how our methods perform when no DIF effects exist. Using the same data generation settings as in our simulation studies except for the zero DIF effects, 50 replications were generated with the sample size $N = 500$. With the same priors and initial values as those in the simulation studies, the uniform and non-uniform DIF models were fitted, respectively. The convergence rates were 98% and 94% for the uniform and non-uniform DIF models. The MSE compositions and CI coverage rates are shown in Fig. 5 of Appendix A, which indicated good recovery of the model parameters. The CRs, FPRs and FNRs for the incidence

matrix were satisfactory, with 0.982, 0.038 and 0.009 for uniform DIF model, and 0.976, 0.003 and 0.033 for non-uniform DIF model. Two models produced well-behaved FPRs for DIF detection, with 0.020 and 0.030 for the uniform and non-uniform DIF models, respectively.

The current study also has some limitations and can be further improved in several aspects. First, our models are complex, especially for the non-uniform DIF conditions. Their computational costs using Bayesian adaptive Lasso are high and the running times are long. It will be important to improve the computational efficiency of our procedures. Second, in order to distinguish different latent traits and place different persons on a common metric, we need to designate K DIF-free items, loading on one dimension separately. These constraints are based on empirical knowledge of the items and may affect the estimation results. When DIF percentage is high, finding the right anchor items may not be easy (Wang et al., 2023). Third, our method can be developed easily to allow for the inclusion of missing data. Standard DIF detection methods are sensitive to missing data and the results for DIF detection are affected by different imputation methods. However, Bayesian method can handle missing data by sampling from posterior distribution, and no imputation is needed. Fourth, other penalty functions or regularized priors can be studied. Several nonconvex penalties, such as SCAD (smoothly clipped absolute deviation; Fan & Li, 2001) and MCP (minimax concave penalty; Zhang, 2010), are well-known. But their performance for simultaneously detecting item–trait relationship and DIF effects is lack of an in-depth study. Furthermore, regularized priors with different types of mixture distributions should be thoroughly investigated. Finally, since the indeterminacy of item–trait relationship, the interactions involving latent traits are very complicated. Further studies for distinguishing the incidence of latent traits and the discriminatory power of covariates need to be investigated.

Acknowledgements

We thank the editor, associate editor, and three anonymous referees for their careful review and valuable comments. This research is partially supported by the National Natural Science Foundation of China (No. 11871013) and the Natural Science Foundation of Jilin Province (No. 20210101152JC).

Author Contributions Na Shan contributed to conceptualization, methodology, writing—original draft, and writing—review and editing. Ping-Feng Xu contributed to supervision and methodology.

Data Availability Data sharing is not applicable to this paper as no new data were created or analyzed in this study.

Declarations

Conflict of interest All authors declare no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

APPENDIX

Appendix A. Additional tables and figures.

TABLE 8.
Average CPU times in seconds for all conditions in studies 1 and 2

	Small DIF				Large DIF			
	20% DIF		60% DIF		20% DIF		60% DIF	
	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
Uniform DIF	974.273	2406.643	990.339	2456.201	969.788	2324.013	994.945	2642.451
Non-uniform DIF	3616.052	7499.925	3641.076	7116.344	3660.076	7442.180	3594.443	7287.310

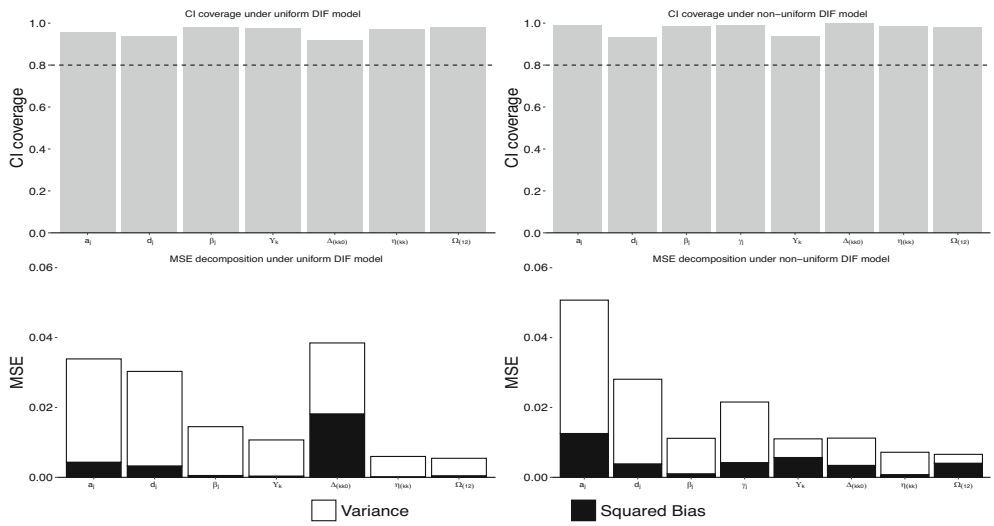

FIGURE 5.
CI coverage and MSE decompositions for the simulation with no DIF effects and sample size 500.

TABLE 9.
MSEs for each parameter estimate in study 1

	Small DIF				Large DIF			
	20% DIF		60% DIF		20% DIF		60% DIF	
parameter	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
a_{31}	0.058	0.037	0.059	0.027	0.056	0.030	0.046	0.033
a_{32}	0.040	0.027	0.043	0.030	0.042	0.033	0.042	0.033
a_{41}	0.039	0.021	0.031	0.022	0.034	0.022	0.034	0.020
a_{42}	0.018	0.008	0.016	0.007	0.018	0.007	0.017	0.008
a_{51}	0.022	0.012	0.022	0.012	0.022	0.012	0.025	0.010
a_{52}	0.012	0.009	0.013	0.011	0.015	0.010	0.018	0.010
a_{61}	0.008	0.004	0.011	0.006	0.008	0.005	0.008	0.008
a_{62}	0.009	0.004	0.010	0.005	0.010	0.005	0.008	0.004
a_{71}	0.034	0.019	0.029	0.021	0.033	0.023	0.029	0.024
a_{72}	0.025	0.023	0.020	0.023	0.021	0.027	0.042	0.025
a_{81}	0.032	0.026	0.034	0.030	0.030	0.032	0.039	0.022
a_{82}	0.044	0.024	0.031	0.022	0.042	0.026	0.041	0.035
a_{91}	0.074	0.040	0.064	0.044	0.098	0.043	0.074	0.038
a_{92}	0.011	0.009	0.013	0.009	0.011	0.008	0.013	0.015
$a_{10,1}$	0.012	0.008	0.012	0.010	0.009	0.007	0.009	0.010
$a_{10,2}$	0.014	0.012	0.013	0.014	0.013	0.010	0.012	0.007
$a_{11,1}$	0.011	0.003	0.007	0.004	0.011	0.003	0.007	0.004
$a_{11,2}$	0.092	0.064	0.098	0.043	0.081	0.053	0.072	0.053
$a_{12,1}$	0.065	0.043	0.067	0.042	0.065	0.044	0.062	0.026
$a_{12,2}$	0.027	0.014	0.030	0.017	0.026	0.015	0.028	0.018
$a_{13,1}$	0.020	0.011	0.022	0.012	0.020	0.013	0.023	0.009
$a_{13,2}$	0.075	0.053	0.088	0.055	0.076	0.054	0.087	0.066
$a_{14,1}$	0.078	0.038	0.083	0.041	0.079	0.041	0.075	0.045

TABLE 9.
continued

	Small DIF				Large DIF			
	20% DIF		60% DIF		20% DIF		60% DIF	
parameter	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
$a_{14,2}$	0.043	0.033	0.040	0.034	0.051	0.041	0.074	0.036
$a_{15,1}$	0.032	0.024	0.044	0.023	0.036	0.026	0.054	0.024
$a_{15,2}$	0.033	0.024	0.035	0.025	0.031	0.022	0.051	0.024
d_1	0.035	0.015	0.033	0.017	0.037	0.015	0.036	0.022
d_2	0.049	0.042	0.046	0.031	0.039	0.042	0.029	0.021
d_3	0.029	0.014	0.025	0.010	0.029	0.013	0.020	0.010
d_4	0.042	0.013	0.040	0.012	0.029	0.011	0.032	0.016
d_5	0.080	0.025	0.062	0.024	0.063	0.025	0.073	0.022
d_6	0.011	0.004	0.011	0.003	0.010	0.004	0.008	0.004
d_7	0.014	0.006	0.016	0.006	0.015	0.006	0.018	0.009
d_8	0.016	0.008	0.015	0.009	0.018	0.008	0.022	0.008
d_9	0.012	0.004	0.012	0.004	0.012	0.004	0.008	0.007
d_{10}	0.009	0.006	0.009	0.007	0.009	0.006	0.008	0.005
d_{11}	0.041	0.020	0.051	0.022	0.039	0.018	0.045	0.014
d_{12}	0.046	0.024	0.023	0.014	0.049	0.027	0.042	0.020
d_{13}	0.041	0.015	0.041	0.015	0.026	0.015	0.024	0.012
d_{14}	0.033	0.011	0.024	0.014	0.034	0.012	0.022	0.016
d_{15}	0.034	0.013	0.034	0.013	0.033	0.013	0.021	0.014

TABLE 9.
continued

	Small DIF				Large DIF			
	20% DIF		60% DIF		20% DIF		60% DIF	
parameter	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
β_{31}	0.014	0.006	0.019	0.011	0.013	0.005	0.030	0.011
β_{32}	0.024	0.016	0.028	0.016	0.027	0.014	0.033	0.020
β_{33}	0.011	0.006	0.012	0.008	0.012	0.006	0.013	0.008
β_{34}	0.010	0.007	0.011	0.008	0.009	0.007	0.013	0.007
β_{41}	0.008	0.006	0.011	0.009	0.009	0.009	0.017	0.010
β_{42}	0.032	0.015	0.033	0.014	0.022	0.018	0.033	0.015
β_{43}	0.011	0.006	0.011	0.006	0.008	0.007	0.009	0.009
β_{44}	0.007	0.006	0.005	0.005	0.007	0.005	0.012	0.006
β_{51}	0.007	0.007	0.015	0.010	0.007	0.007	0.021	0.008
β_{52}	0.039	0.020	0.016	0.029	0.038	0.019	0.028	0.014
β_{53}	0.013	0.008	0.013	0.006	0.013	0.008	0.010	0.005
β_{54}	0.014	0.008	0.012	0.007	0.013	0.008	0.013	0.007
β_{61}	0.005	0.002	0.006	0.003	0.005	0.002	0.005	0.003
β_{62}	0.018	0.007	0.017	0.007	0.019	0.007	0.015	0.010
β_{63}	0.005	0.003	0.005	0.003	0.005	0.003	0.005	0.004
β_{64}	0.004	0.003	0.005	0.003	0.004	0.003	0.005	0.004
β_{71}	0.008	0.007	0.009	0.006	0.009	0.007	0.013	0.006
β_{72}	0.037	0.021	0.073	0.041	0.042	0.025	0.128	0.093
β_{73}	0.007	0.007	0.008	0.009	0.008	0.008	0.013	0.006
β_{74}	0.012	0.007	0.008	0.006	0.013	0.006	0.014	0.008
β_{81}	0.009	0.005	0.009	0.004	0.008	0.004	0.011	0.006
β_{82}	0.050	0.019	0.078	0.033	0.076	0.027	0.133	0.087
β_{83}	0.008	0.006	0.007	0.006	0.006	0.005	0.010	0.007
β_{84}	0.012	0.006	0.013	0.005	0.009	0.005	0.006	0.005
β_{91}	0.008	0.002	0.007	0.002	0.008	0.002	0.004	0.003

TABLE 9.
continued

	Small DIF				Large DIF			
	20% DIF		60% DIF		20% DIF		60% DIF	
parameter	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
β_{92}	0.033	0.012	0.044	0.024	0.035	0.011	0.062	0.038
β_{93}	0.008	0.005	0.006	0.005	0.009	0.005	0.006	0.002
β_{94}	0.006	0.003	0.006	0.003	0.007	0.003	0.007	0.003
$\beta_{10,1}$	0.006	0.002	0.006	0.002	0.005	0.002	0.006	0.002
$\beta_{10,2}$	0.022	0.011	0.029	0.014	0.023	0.013	0.031	0.020
$\beta_{10,3}$	0.005	0.003	0.005	0.003	0.005	0.003	0.005	0.003
$\beta_{10,4}$	0.004	0.003	0.004	0.003	0.004	0.003	0.004	0.003
$\beta_{11,1}$	0.011	0.007	0.013	0.008	0.011	0.007	0.017	0.010
$\beta_{11,2}$	0.054	0.024	0.094	0.043	0.066	0.026	0.117	0.082
$\beta_{11,3}$	0.017	0.008	0.019	0.008	0.016	0.007	0.013	0.006
$\beta_{11,4}$	0.014	0.007	0.015	0.008	0.015	0.007	0.007	0.010
$\beta_{12,1}$	0.011	0.007	0.013	0.012	0.010	0.008	0.029	0.011
$\beta_{12,2}$	0.048	0.015	0.087	0.041	0.054	0.016	0.125	0.087
$\beta_{12,3}$	0.009	0.006	0.011	0.004	0.010	0.006	0.009	0.005
$\beta_{12,4}$	0.009	0.010	0.009	0.008	0.010	0.010	0.010	0.006
$\beta_{13,1}$	0.013	0.006	0.014	0.009	0.013	0.010	0.012	0.008
$\beta_{13,2}$	0.031	0.023	0.046	0.035	0.064	0.021	0.125	0.073
$\beta_{13,3}$	0.009	0.004	0.010	0.003	0.008	0.005	0.007	0.005
$\beta_{13,4}$	0.013	0.007	0.014	0.006	0.010	0.009	0.013	0.006
$\beta_{14,1}$	0.013	0.005	0.013	0.008	0.013	0.006	0.023	0.012

TABLE 9.
continued

	Small DIF				Large DIF			
	20% DIF		60% DIF		20% DIF		60% DIF	
parameter	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
$\beta_{14,2}$	0.042	0.019	0.039	0.047	0.045	0.022	0.094	0.067
$\beta_{14,3}$	0.013	0.005	0.011	0.004	0.013	0.005	0.012	0.009
$\beta_{14,4}$	0.007	0.008	0.009	0.007	0.007	0.008	0.013	0.013
$\beta_{15,1}$	0.015	0.007	0.024	0.012	0.016	0.008	0.029	0.010
$\beta_{15,2}$	0.027	0.023	0.032	0.025	0.037	0.025	0.073	0.055
$\beta_{15,3}$	0.013	0.006	0.012	0.006	0.014	0.006	0.013	0.007
$\beta_{15,4}$	0.011	0.010	0.011	0.010	0.012	0.009	0.013	0.010
Υ_{11}	0.008	0.004	0.010	0.006	0.008	0.004	0.012	0.005
Υ_{12}	0.022	0.009	0.021	0.009	0.023	0.009	0.026	0.014
Υ_{13}	0.006	0.004	0.007	0.004	0.006	0.004	0.006	0.004
Υ_{14}	0.006	0.004	0.006	0.003	0.006	0.003	0.008	0.005
Υ_{21}	0.006	0.003	0.006	0.003	0.006	0.003	0.009	0.004
Υ_{22}	0.029	0.015	0.033	0.017	0.029	0.015	0.036	0.026
Υ_{23}	0.011	0.003	0.009	0.003	0.011	0.003	0.010	0.005
Υ_{24}	0.009	0.004	0.009	0.004	0.009	0.004	0.013	0.005
$\Delta_{(110)}$	0.024	0.016	0.027	0.018	0.027	0.015	0.035	0.022
$\Delta_{(220)}$	0.062	0.028	0.068	0.023	0.055	0.029	0.050	0.017
$\eta_{(11)1}$	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.001
$\eta_{(11)2}$	0.014	0.007	0.014	0.008	0.013	0.007	0.014	0.009
$\eta_{(11)3}$	0.004	0.002	0.004	0.001	0.005	0.002	0.005	0.002
$\eta_{(11)4}$	0.005	0.002	0.005	0.002	0.005	0.002	0.006	0.002
$\eta_{(22)1}$	0.004	0.001	0.004	0.001	0.004	0.001	0.004	0.001
$\eta_{(22)2}$	0.011	0.006	0.010	0.007	0.010	0.006	0.012	0.006
$\eta_{(22)3}$	0.003	0.002	0.004	0.002	0.003	0.002	0.003	0.002
$\eta_{(22)4}$	0.004	0.001	0.004	0.001	0.004	0.001	0.004	0.001
$\Omega_{(12)}$	0.006	0.003	0.007	0.004	0.006	0.003	0.006	0.004

References

Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling, 12*(3), 279–297. https://doi.org/10.1177/1471082X1101200304

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438. https://doi.org/10.1080/10705510903008204

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22*(3), 507–526. https://doi.org/10.1037/met0000077

Bauer, D. J., Belzak, W. C. M., & Cole, V. T. (2020). Simplifying the assessment of measurement invariance over multiple background variables: Using regularized moderated nonlinear factor analysis to detect differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 43–55. https://doi.org/10.1080/10705511.2019.1642754

Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika, 66*(4), 541–561. https://doi.org/10.1007/BF02296195

Belzak, W. C. M., & Bauer, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychological Methods, 25*(6), 673–690. https://doi.org/10.1037/met0000253

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*(3), 261–280. https://doi.org/10.1177/014662168801200305

Brandt, H., Chen, S. M., & Bauer, D. J. (2023). Bayesian penalty methods for evaluating measurement invariance in moderated nonlinear factor analysis. *Psychological Methods*. https://doi.org/10.1037/met0000552

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24*(4), 445–455. https://doi.org/10.1207/s15327906mbr2404_4

Buecker, S., Maes, M., Denissen, J. J. A., & Luhmann, M. (2020). Loneliness and the big five personality traits: A meta-analysis. *European Journal of Personality, 34*, 8–28. https://doi.org/10.1002/per.2229

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chen, J. S. (2020). A partially confirmatory approach to the multidimensional item response theory with the Bayesian Lasso. *Psychometrika, 85*(3), 738–774. https://doi.org/10.1007/s11336-020-09724-3

Chen, J. S., Guo, Z. H., Zhang, L. J., & Pan, J. H. (2021). A partially confirmatory approach to scale development with the Bayesian Lasso. *Psychological Methods, 26*(2), 210–235. https://doi.org/10.1037/met0000293

Chen, S. M., Bauer, D. J., Belzak, W. M., & Brandt, H. (2022). Advantages of spike and slab priors for detecting differential item functioning relative to other Bayesian regularizing priors and frequentist Lasso. *Structural Equation Modeling: A Multidisciplinary Journal, 29*(1), 122–139. https://doi.org/10.1080/10705511.2021.1948335

da Silva, M. A., Liu, R., Huggins-Manley, A. C., & Bazán, J. L. (2019). Incorporating the Q-Matrix into multidimensional item response theory models. *Educational and Psychological Measurement, 79*(4), 665–687. https://doi.org/10.1177/0013164418814898

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816045037

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617–634. https://doi.org/10.1037/met0000167

Eysenck, S., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences, 54*(4), 485–489. https://doi.org/10.1016/j.paid.2012.09.022

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96*(456), 1348–1360. https://doi.org/10.1198/016214501753382273

Feng, X. N., Wu, H. T., & Song, X. Y. (2017). Bayesian adaptive Lasso for ordinal regression with latent variables. *Sociological Methods and Research, 46*(4), 926–953. https://doi.org/10.1177/0049124115610349

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.)*. Springer.

Huang, P. H. (2018). A penalized likelihood method for multi-group structural equation modelling. *British Journal of Mathematical and Statistical Psychology, 71*(3), 499–522. https://doi.org/10.1111/bmsp.12130

Huang, P. H., Chen, H., & Weng, L. J. (2017). A penalized likelihood method for structural equation modeling. *Psychometrika, 82*(2), 329–354. https://doi.org/10.1007/s11336-017-9566-9

Jacobucci, R., Grimm, K. J., & Mcardle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(4), 555–566. https://doi.org/10.1080/10705511.2016.1154793

Janssen, R., & De Boeck, P. (1999). Confirmatory analyses of componential test structure using multidimensional item response theory. *Multivariate Behavioral Research, 34*(2), 245–268. https://doi.org/10.1207/S15327906Mb340205

Jin, K. Y., & Wang, W. C. (2014). Item response theory models for performance decline during testing. *Journal of Educational Measurement, 51*(2), 178–200. https://doi.org/10.1111/jedm.12041

Kim, J., & Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*(3), 458–470. https://doi.org/10.1177/0013164412467033

Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. Methuen.

Lee, S., Bulut, O., & Suh, Y. (2017). Multidimensional extension of multiple indicators multiple causes models to detect DIF. *Educational and Psychological Measurement, 77*(4), 545–569. https://doi.org/10.1177/0013164416651116

Leng, C., Tran, M. N., & Nott, D. (2014). Bayesian adaptive Lasso. *Annals of the Institute of Statistical Mathematics, 66*(2), 221–244. https://doi.org/10.1007/s10463-013-0429-6

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of differential item functioning using the Lasso approach. *Journal of Educational and Behavioral Statistics, 40*(2), 111–135. https://doi.org/10.3102/1076998614559747

Mazor, K. M., Hambleton, R. K., & Clauser, B. E. (1998). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. *Applied Psychological Measurement, 22*(4), 357–367. https://doi.org/10.1177/014662169802200404

Mckinley, R. (1989). *Confirmatory analysis of test structure using multidimensional item response theory*. Technical Report No. RR-89-31. Educational Testing Service.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*(4), 297–334. https://doi.org/10.1177/014662169301700401

Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement, 34*(3), 253–272. https://doi.org/10.1111/j.1745-3984.1997.tb00518.x

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian Lasso. *Psychological Methods, 22*(4), 687–704. https://doi.org/10.1037/met0000112

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association, 103*(482), 681–686. https://doi.org/10.1198/016214508000000337

Plummer, M. (2017). Jags version 4.3.0 user manual. https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/

Polson, N. G., & Sokolov, V. (2019). Bayesian regularization: From Tikhonov to horseshoe. *WIREs Computational Statistics, 11*, e1463. https://doi.org/10.1002/wics.1463

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org

Reckase, M. D. (2009). *Multidimensional item response theory*. Springer.

Schauberger, G., & Mair, P. (2020). A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behavior Research Methods, 52*(1), 279–294. https://doi.org/10.3758/s13428-019-01224-2

Serang, S., Jacobucci, R., Brimhall, K. C., & Grimm, K. J. (2017). Exploratory mediation analysis via regularization. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(5), 733–744. https://doi.org/10.1080/10705511.2017.1311775

Stan Development Team. (2023). RStan: The R interface to Stan [R package version 2.21.8]. http://mc-stan.org/

Stout, W., Li, H., Nandakumar, R., & Bolt, D. (1997). MULTISIB - a procedure to investigate DIF when a test is intentionally multidimensional. *Applied Psychological Measurement, 21*(3), 195–213. https://doi.org/10.1177/01466216970213001

Suh, Y., & Cho, S. J. (2014). Chi-square difference tests for detecting functioning in a multidimensional IRT model: A Monte Carlo study. *Applied Psychological Measurement, 38*(5), 359–375. https://doi.org/10.1177/0146621614523116

Sun, J., Chen, Y., Liu, J., Ying, Z., & Xin, T. (2016). Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika, 81*(4), 921–939. https://doi.org/10.1007/s11336-016-9529-6

Teresi, J. A., Ramirez, M., Lai, J., & Silver, S. (2008). Occurrences and sources of differential item functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science, 50*(4), 538–612.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tibshirani, R., Friedman, J., Hastie, T., Narasimhan, B., Simon, N., & Qian, J. (2021). *glmnet: Lasso and elastic-net regularized generalized linear models*. https://www.rdocumentation.org/packages/glmnet/versions/4.1-3

Tutz, G., & Schauberger, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika, 80*(1), 21–43. https://doi.org/10.1007/s11336-013-9377-6

Wang, C., Zhu, R. Y., & Xu, G. J. (2023). Using Lasso and adaptive Lasso to identify DIF in multidimensional 2PL models. *Multivariate Behavioral Research, 58*(2), 387–407. https://doi.org/10.1080/00273171.2021.1985950

Wellner, J., & Zhang, T. (2012). Introduction to the special issue on sparsity and regularization methods. *Statistical Science, 27*(4), 447–449. https://doi.org/10.1214/12-STS409

Woods, C. M. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement, 33*(1), 42–57. https://doi.org/10.1177/0146621607314044

Xu, P. F., Shang, L., Zheng, Q. Z., Shan, N., & Tang, M. L. (2022). Latent variable selection in multidimensional item response theory models using the expectation model selection algorithm. *British Journal of Mathematical and Statistical Psychology, 75*(2), 363–394. https://doi.org/10.1111/bmsp.12261

Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics, 38*(2), 894–942. https://doi.org/10.1214/09-AOS729

Zhang, S., & Chen, Y. (2022). Computation for latent variable model estimation: A unified stochastic proximal framework. *Psychometrika, 87*(4), 1473–1502. https://doi.org/10.1007/s11336-022-09863-9

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735

Manuscript Received: 2 DEC 2023
Final Version Received: 16 JUL 2024
Published Online Date: 10 AUG 2024