

Article

Regularized Mixture Rasch Model

Alexander Robitzsch ^{1,2} 

¹ IPN—Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, 24118 Kiel, Germany; robitzsch@leibniz-ipn.de

² Centre for International Student Assessment (ZIB), Olshausenstraße 62, 24118 Kiel, Germany

Abstract: The mixture Rasch model is a popular mixture model for analyzing multivariate binary data. The drawback of this model is that the number of estimated parameters substantially increases with an increasing number of latent classes, which, in turn, hinders the interpretability of model parameters. This article proposes regularized estimation of the mixture Rasch model that imposes some sparsity structure on class-specific item difficulties. We illustrate the feasibility of the proposed modeling approach by means of one simulation study and two simulated case studies.

Keywords: mixture Rasch model; regularization; penalty; differential item functioning

1. Introduction

In education and psychological science, multivariate data of cognitive test items such as intelligence tests are frequently analyzed. The Rasch model (RM; [1–4]) is likely the most popular statistical model in applied research for analyzing a vector of random variables $X = (X_1, \dots, X_I)$ of I dichotomous item responses (i.e., $X_i \in \{0, 1\}$ for $i = 1, \dots, I$). The multivariate probability distribution $P(X = x)$ in the RM is given as

$$P(X = x) = \int \prod_{i=1}^I P_i(x_i, \theta; b_i) \phi(\theta; \mu, \sigma) d\theta \text{ for } x = (x_1, \dots, x_I) \in \{0, 1\}^I, \quad (1)$$

where $P_i(1, \theta; b_i) = \Psi(\theta - b_i)$ (which is also referred to as the item response function), $P_i(0, \theta; b_i) = 1 - P_i(1, \theta; b_i)$ and Ψ denotes the logistic distribution function. Moreover, ϕ is the density function of the normal distribution with mean μ and standard deviation σ . The latent variable θ can be thought of as an underlying unidimensional factor that represents multivariate dependencies of the discrete vector X . Notably, the normal distribution assumption in the RM could be weakened [5]. The item difficulties b_i represent a nonlinear transformation of proportion correct values of items X_i . Note that an identification constraint in (1) in the estimation of the RM must be imposed. Frequently, the mean μ is set to zero, or one fixes the mean of item difficulties to zero (i.e., $\sum_{i=1}^I b_i = 0$).

The mixture Rasch model (MRM; [6–8]) models a heterogeneous distribution for X . In a nutshell, it is assumed that the RM in each of C latent classes and the marginal distribution can be interpreted as a mixture distribution [9]. The distribution of the MRM with C latent classes is given by

$$P(X = x) = \sum_{c=1}^C p_c \int \prod_{i=1}^I P_i(x_i, \theta; b_{ic}) \phi(\theta; \mu_c, \sigma_c) d\theta, \quad (2)$$

where the non-negative mixture probabilities p_c ($c = 1, \dots, C$) add to one. The class-specific item difficulties b_{ic} in (2) indicate the difficulty (i.e., some nonlinear transformation of proportion correct values) of item X_i in latent class c . The distributional differences of latent classes are captured in the mean μ_c and standard deviation σ_c . The MRM can be interpreted as a model in which subjects are allocated into one of the C latent classes. The multivariate relationships in the vector X of items can differ across latent classes.



Citation: Robitzsch, A. Regularized Mixture Rasch Model. *Information* **2022**, *13*, 534. <https://doi.org/10.3390/info13110534>

Academic Editors: Eftim Zdravevski, Petre Lameski and Ivan Miguel Serrano Pires

Received: 30 September 2022

Accepted: 4 November 2022

Published: 10 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Like in the RM defined in Equation (1), identification constraints in the MRM defined in Equation (2) are required [10]. One can fix all class means μ_c to zero or set the mean of item difficulties within each of the classes to zero (i.e., $\sum_{i=1}^I b_{ic} = 0$ for all $c = 1, \dots, C$). The latter constraint has the advantage that differences between item parameters across latent classes can be interpreted.

After applying a standardization of class-specific item difficulties such as the above-mentioned mean centering, differences between class-specific item difficulties can be computed. The so-called latent differential item functioning (DIF; [11–13]) effects qualitatively describe the distinctive behavior of latent classes at the level of items [14]. Studying these latent DIF effects is an important exploratory step in understanding the differential performance of test takers on items [15].

The MRM has been extended to polytomous item responses [16,17] and more complex item response functions P_i [18–24]. A disadvantage of the MRM in (2) is that all item difficulties are allowed to differ across classes. In empirical data, some parameters are likely to equal each other. This is the motivation for proposing a regularized mixture Rasch model (RMRM) that presupposes that only a subset of DIF effects differs from zero. Alternatively, to put it differently, subsets of class-specific item parameters are set to be equal to each other in model estimation. This property substantially eases the model interpretation in exploratory research.

The rest of the article is structured as follows. In Section 2, we present the estimation approach for the RMRM. In Section 3, we present a simulation study for two latent classes. In Sections 4 and 5, we present two simulated case studies that involve two or three latent classes with a particular structure of DIF effects, respectively. Finally, the paper closes with a discussion in Section 6.

2. Regularized Mixture Models

In this section, we present the estimation of the RMRM. Regularized estimation recently became popular in psychometrics, such as item response modeling [25,26], structural equation modeling [27,28], and structured latent class analysis [29–31]. The MRM involves C latent classes. The allocation of persons (or subjects) to latent classes is unknown. If it were known, a multiple group RM with known (i.e., manifest) group allocation would result. The investigation of known demographic groups, such as gender or language groups, is an important topic in educational measurement. Moreover, regularization techniques were recently discussed for manifest DIF detection in the RM [32–38].

The main idea of using regularization techniques (see [39] for an overview) for the MRM is that by subtracting an appropriate penalty term from the log-likelihood function, some simplified structure on DIF effects is imposed. Let $\mathbf{X} = (x_{pi})_{pi}$ denote the matrix of dichotomous item responses. The marginal log-likelihood function in the MRM is given by

$$l(\mathbf{b}, \gamma; \mathbf{X}) = \sum_{p=1}^N \log \left(\sum_{c=1}^C p_c \int \prod_{i=1}^I P_i(x_{pi}, \theta; b_{ic}) \phi(\theta; \mu_c, \sigma_c) d\theta \right). \tag{3}$$

In practice, the integration in (3) can be substituted by a summation and evaluating θ at a finite grid θ_t for $t = 1, \dots, T$:

$$l(\mathbf{b}, \gamma; \mathbf{X}) = \sum_{p=1}^N \log \left(\sum_{c=1}^C p_c \sum_{t=1}^T \prod_{i=1}^I P_i(x_{pi}, \theta_t; b_{ic}) \omega(\theta_t; \mu_c, \sigma_c) \right), \tag{4}$$

where ω is a discrete analog of the normal density. The latent class probabilities p_c can be represented by logistically transformed parameters q_c :

$$p_c = \frac{\exp(q_c)}{\sum_{d=1}^C \exp(q_d)} \text{ for } c = 1, \dots, C, \tag{5}$$

and one sets $q_1 = 0$.

Regularization techniques use penalty functions to control the variability in subsets of model parameters. For a scalar parameter x , the lasso penalty is defined as

$$\mathcal{P}_{\text{Lasso}}(x, \lambda) = \lambda|x|, \tag{6}$$

where λ is a non-negative regularization parameter. It is known that the lasso penalty induces bias in estimated parameters. To circumvent this issue, the smoothly clipped absolute deviation (SCAD; [40]) penalty has been proposed. It is defined by

$$\mathcal{P}_{\text{SCAD}}(x, \lambda) = \begin{cases} \lambda|x| & \text{if } |x| < \lambda \\ -(x^2 - 2a\lambda|x|^2 + \lambda^2)(2(a - 1))^{-1} & \text{if } \lambda \leq |x| \leq a\lambda \\ (a + 1)\lambda^2 & \text{if } |x| > a\lambda \end{cases} \tag{7}$$

with $a = 3.7$.

2.1. Two Alternative Approaches to Regularizing the Mixture Rasch Model

We can estimate the RMRM in two variants of applying regularization. In the first approach, we use overidentified item parameters b_{ic} and the fused regularization technique [41,42]. Let \mathbf{b} denote the vector of all class-specific item parameters and γ the vector of all distribution parameters. The following estimation function H (i.e., the negative of a regularized likelihood function) is minimized

$$H(\mathbf{b}, \gamma; \mathbf{X}) = -l(\mathbf{b}, \gamma; \mathbf{X}) + N \sum_{i=1}^I \sum_{c=1}^{C-1} \sum_{c'=c+1}^C \mathcal{P}_{\text{SCAD}}(b_{ic} - b_{ic'}, \lambda). \tag{8}$$

Note that fused regularization (8) penalizes the presence of many nonvanishing item parameter differences $b_{ic} - b_{ic'}$. With a regularization parameter $\lambda = 0$, differences $b_{ic} - b_{ic'}$ in item difficulties are unpenalized. With increasing values of λ , the penalty contribution in the estimation function H becomes larger. Eventually, with sufficiently large λ values, item difficulties b_{ic} and $b_{ic'}$ are fused; that is, they receive the same estimate.

Moreover, note that sample size N is multiplied with the penalty function in (8). We prefer this choice because optimal values of the regularization parameter λ are less sample size dependent in this case. Moreover, optimal λ values can be more easily compared across different sample sizes.

It should be noted that in model estimation, the regularization parameter λ in (8) has to be fixed. In practice, the regularization parameter λ also has to be estimated. Hence, the minimization is performed on a grid of λ values (e.g., $\lambda = 0.01, 0.02, \dots, 0.50$), and the model that is selected is optimal with respect to some criterion. Typical criteria are the cross-validated log-likelihood, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) [39]. See [43] for model selection for the (nonregularized) MRM.

The second estimation approach relies on the ordinary regularization of latent DIF effects. The latent DIF effects are included by using an overidentified model with common item parameters b_{i0} and latent DIF effects e_{ic} by relying on the decomposition

$$b_{ic} = b_{i0} + e_{ic}. \tag{9}$$

Note that the difference in item difficulties for classes c and c' are given as

$$b_{ic} - b_{ic'} = e_{ic} - e_{ic'} \tag{10}$$

Hence, latent DIF effects quantify differences between item difficulties across latent classes after introducing an implicit identification constraint for determining means μ_c of latent classes $c = 1, \dots, C$ (while fixing $\mu_1 = 0$). Using latent DIF effects in the second approach (9) instead of considering the regularization of differences in item difficulties

in (8) possesses advantages if the focus of the analysis lies in the detection and assessment of latent DIF effects.

The estimation function based on the decomposition (9) is defined by

$$H(\mathbf{b}_0, \mathbf{e}, \gamma; \mathbf{X}) = -l(\mathbf{b}_0, \mathbf{e}, \gamma; \mathbf{X}) + N \sum_{i=1}^I \sum_{c=1}^C \mathcal{P}_{\text{SCAD}}(e_{ic}, \lambda), \quad (11)$$

where \mathbf{b}_0 denotes the vector of all common item parameters b_{i0} .

The special case of two latent classes in the MRM requires further attention. In this case, only one DIF effect e_i must be included in the model by relying on the decomposition

$$\begin{aligned} b_{i1} &= b_{i0} - e_i/2 \\ b_{i2} &= b_{i0} + e_i/2 \end{aligned} \quad (12)$$

In general, fused regularization will impose a bit more structured solutions for more than three latent classes if there are clusters of latent classes with the same DIF effect at the level of each item. In contrast, SCAD regularization (11) only presupposes one item-specific cluster of latent classes with zero DIF effects, and all other DIF effects differ from zero but do not potentially merge into another cluster of latent classes with the same DIF effect. Whether the more general structure of fused regularization is advantageous in Rasch mixture with at least four classes is an empirical question in concrete applications.

2.2. Estimation

The regularized likelihood functions can be optimized using marginal maximum likelihood estimation and the expectation maximization (EM) algorithm [26,31,44]. The EM algorithm alternates between the E-step and the M-step. The E-step computation is identical to the estimation in nonregularized item response models. In the M-step, the maximization of the regularized expected log-likelihood function involving expected counts is carried out. The difference in regularized estimation is that the optimization function becomes nondifferentiable because the SCAD penalty is nondifferentiable. The optimization of nondifferentiable optimization can be performed using gradient descent [39] approaches or by replacing the nondifferentiable optimization functions with differentiable approximating functions [31,42,45,46]. In our experience, the latter approach is quite satisfactory in applications.

As usually encountered in mixture models, the maximum likelihood optimization function is often prone to local optima. Hence, it is recommended to estimate the RMRM with a sufficiently large number of random starting values to ensure that the estimated solution corresponds to the global optimum of the likelihood function (see [47]).

The sketched EM algorithm can be practically implemented in the general estimation function `xxirt()` in the R package `sirt` [48]. This function is used in the simulation study and the two case studies in this paper.

2.3. Computation of Standard Errors

The computation of standard errors in regularized ML estimation is an active area of research [39]. In the simulation and case studies in this article, standard errors are computed based on nonparametric bootstrap [49]. The estimated model parameter of interest γ depends on a data-driven determined regularization parameter $\hat{\lambda}_{\text{opt}}$ that is determined by the AIC or the BIC criterion.

The bootstrap can either determine the optimal regularization parameter or one applies regularized ML using the optimal $\hat{\lambda}_{\text{opt}}$ parameter obtained from the original sample. Typically, the former introduces additional variability. In a preliminary analysis in Simulated Case Study 2, it turned out that the average chosen λ parameter in bootstrap samples was substantially larger than the regularization parameter $\hat{\lambda}_{\text{opt}}$ from the original sample. For this reason, we only report results of standard errors in bootstrap samples that use the fixed regularization parameter $\hat{\lambda}_{\text{opt}}$.

Furthermore, it is vital to implement a test of statistical significance for regularized latent DIF effects e_i or differences in class-specific item difficulties. It has been suggested to report the proportion of bootstrap samples p_{boot} in which a regularized DIF effect was estimated equal to zero [39]. Values of p_{boot} that are sufficiently close to zero indicate latent DIF effects e_i that significantly differ from zero.

3. Simulation Study 1: Simulation Study Involving Two Latent Classes

In this section, results from a simulation study of an RMRM with two latent classes are presented.

3.1. Method

The simulated datasets consisted of $I = 20$ items with two latent classes that followed an MRM. The class-specific item difficulties b_{ic} were decomposed into common item difficulties b_{i0} and DIF effects e_i according to Equation (12). The common item difficulties of the 20 items had equidistant values between -2.0 and 2.0 .

Four out of twenty items had DIF effects that differed from zero. Items 6, 8, and 17 had a positive DIF effect δ , while item 11 had a negative DIF effect $-\delta$. In the simulation study, the size of the DIF effect δ was either 0.5 or 1.0.

For identification, the mean μ_1 of the first latent class was set to zero. The standard deviation σ_1 of the first latent class was set to 1.0. For the second latent class, $\mu_2 = 0.5$ and $\sigma_2 = 0.8$ were chosen throughout the simulation. The class probabilities were fixed to $p_1 = 0.7$ and $p_2 = 0.3$.

Moreover, we varied the sample size N in the simulation. We chose sample sizes of 1000, 2500, and 5000 to cover a range of moderate to large sample sizes.

To avoid label switching issues in estimating the RMRM, we utilized a weak prior distribution on the logistically transformed the second latent class probability q_2 (i.e., $p_2 = \Psi(q_2)$). The prior $\pi(q_2)$ was chosen as the normal distribution $N(-0.7, 0.4)$, meaning that the second class was the smaller one to avoid label switching issues when estimating the mixture Rasch model. The model parameters \mathbf{b}_0 (i.e., common item difficulties b_{i0} for $i = 1, \dots, I$), \mathbf{e} (i.e., all DIF effects e_i for $i = 1, \dots, I$), and γ (i.e., $\sigma_1, \mu_2, \sigma_2$, and q_2) were obtained by minimizing the penalized likelihood function:

$$H(\mathbf{b}_0, \mathbf{e}, \gamma; \mathbf{X}) = -l(\mathbf{b}_0, \mathbf{e}, \gamma; \mathbf{X}) + N \sum_{i=1}^I \mathcal{P}_{\text{SCAD}}(e_i, \lambda) - \pi(q_2). \quad (13)$$

In total, 5000 replications were simulated in the 2 (DIF effects) \times 3 (sample size) = 6 conditions.

The following values of the regularization parameter λ were chosen in a decreasing order while using the obtained estimates from the previous estimation as starting values: 1.00, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, 0.50, 0.48, 0.46, 0.44, 0.42, 0.40, 0.38, 0.36, 0.34, 0.32, 0.30, 0.29, 0.28, 0.27, 0.26, 0.25, 0.24, 0.23, 0.22, 0.21, 0.20, 0.19, 0.18, 0.17, 0.16, 0.15, 0.14, 0.13, 0.12, 0.11, 0.10, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.009, 0.008, 0.007, 0.006, and 0.005. DIF effect estimates e_i were considered as zero if their absolute values did not exceed 0.02. In the Results Section 3.2, we report model estimates for fixed λ values of 0.05, 0.10, and 0.15, as well as parameter estimates resulting from models with minimum AIC or BIC values. The performance of parameter estimates was assessed by bias and root mean square error (RMSE).

The whole simulation was conducted in the R software [50] using the `xxirt()` function in the R package `sirt` [48]. The code for the data simulation and model estimation can be found at <https://osf.io/wrs5k/> (accessed on 30 September 2022).

3.2. Results

In Table 1, the average number of detected DIF effects (i.e., estimated to be different from zero) is presented. Four out of twenty items had DIF effects different from zero.

Interestingly, the number of detected DIF effects was substantially underestimated if the BIC was used as the regularization parameter selection criterion, except in the condition of large DIF effects of 1 (i.e., $|DIF| = 1$) and a large sample size of $N = 5000$ with an average of 3.9 detected DIF effects. In particular, model selection based on BIC had worse performance in the case of a small DIF effect of $|DIF| = 0.5$. In contrast, the model estimated with the AIC selection criterion had, on average, 5 to 7 detected DIF effects, which was slightly higher than the number of true DIF effects.

Table 1. Simulation Study 1: Average number of detected DIF effects e_i .

| DIF | N | Choice of λ | | | | |
|-----|------|---------------------|-----|------|-----|------|
| | | AIC | BIC | 0.05 | 0.1 | 0.15 |
| 0.5 | 1000 | 6.4 | 0.4 | 12.5 | 9.1 | 5.3 |
| | 2500 | 5.3 | 0.2 | 8.6 | 5.7 | 3.4 |
| | 5000 | 5.1 | 0.3 | 7.7 | 4.4 | 2.0 |
| 1 | 1000 | 6.9 | 0.9 | 13.0 | 9.4 | 5.6 |
| | 2500 | 6.7 | 2.5 | 9.8 | 6.4 | 4.4 |
| | 5000 | 6.3 | 3.9 | 7.9 | 4.9 | 4.0 |

Note. $|DIF|$ = absolute value of DIF effects e_i ; N = sample size; λ = regularization parameter. Note that 4 out of 20 items had DIF effects different from zero.

If the RMRM is estimated with a fixed regularization parameter λ , the average number of detected DIF effects decreases with increasing sample sizes. Overall, model selection based on AIC might be preferred over BIC if it is more critical not to detect true DIF effects.

Table 2 presents type I error for non-DIF effects e_i (i.e., e_i was zero in the simulated data) and power rates for DIF effects e_i (i.e., e_i had values different from zero). It turned out that type I error rates were relatively high if the AIC was used as the model selection criterion (Min = 14.6, Max = 30.7). Type I error rates for the model selection based on BIC ranged between 0.4 and 1.7. These low type I error rates for the BIC criterion come at the price that BIC has very low power for detecting DIF effects if the true DIF effect is small (i.e., $|DIF| = 0.5$) or in not too large sample sizes (i.e., $N = 1000$). Interestingly, for the smallest sample size of $N = 1000$ and small DIF effects, type I error rates and power rates were very close to each other (i.e., based on AIC: type I error rate was 30.7 and the power rate was 38.3), but power rates improve in large sample sizes or for large DIF effects.

Table 2. Simulation Study 1: Average type I error rates for items with no DIF effects and average power rates for items with DIF effects.

| DIF | N | Type I Error Rate for Non-DIF Effects e_i | | | | | Power Rate for DIF Effects e_i | | | | |
|-----|------|---|-----|------|------|------|----------------------------------|------|-------|------|------|
| | | Choice of λ | | | | | Choice of λ | | | | |
| | | AIC | BIC | 0.05 | 0.1 | 0.15 | AIC | BIC | 0.05 | 0.1 | 0.15 |
| 0.5 | 1000 | 30.7 | 1.6 | 61.1 | 44.1 | 25.3 | 38.3 | 2.7 | 69.0 | 52.3 | 31.1 |
| | 2500 | 21.3 | 0.5 | 37.1 | 23.2 | 13.3 | 46.8 | 2.7 | 67.4 | 48.6 | 31.2 |
| | 5000 | 17.5 | 0.4 | 29.6 | 14.8 | 5.9 | 58.5 | 5.4 | 74.4 | 52.0 | 27.1 |
| 1 | 1000 | 25.6 | 1.7 | 58.5 | 38.4 | 19.7 | 70.0 | 16.5 | 90.1 | 80.5 | 61.6 |
| | 2500 | 17.6 | 1.0 | 36.4 | 16.1 | 5.8 | 95.9 | 58.0 | 98.9 | 95.3 | 87.5 |
| | 5000 | 14.6 | 0.6 | 24.5 | 5.5 | 0.8 | 99.9 | 95.4 | 100.0 | 99.6 | 96.0 |

Note. $|DIF|$ = absolute value of DIF effects e_i ; N = sample size; λ = regularization parameter.

Figure 1 displays the optimal regularization parameter λ_{opt} as a function of sample size, the size of DIF effects, and the chosen information criterion AIC or BIC. It can be seen that λ_{opt} values were generally smaller when based on AIC instead of BIC. Moreover, the optimal regularization parameter decreases with a larger sample size. It is evident that there is substantial variability in the estimated λ_{opt} values in repeated samples. It can be

seen that the largest λ_{opt} value was frequently obtained for BIC (i.e., for a small DIF effect of $|DIF| = 0.5$ or $N = 1000$). In this case, all latent DIF effects were regularized.

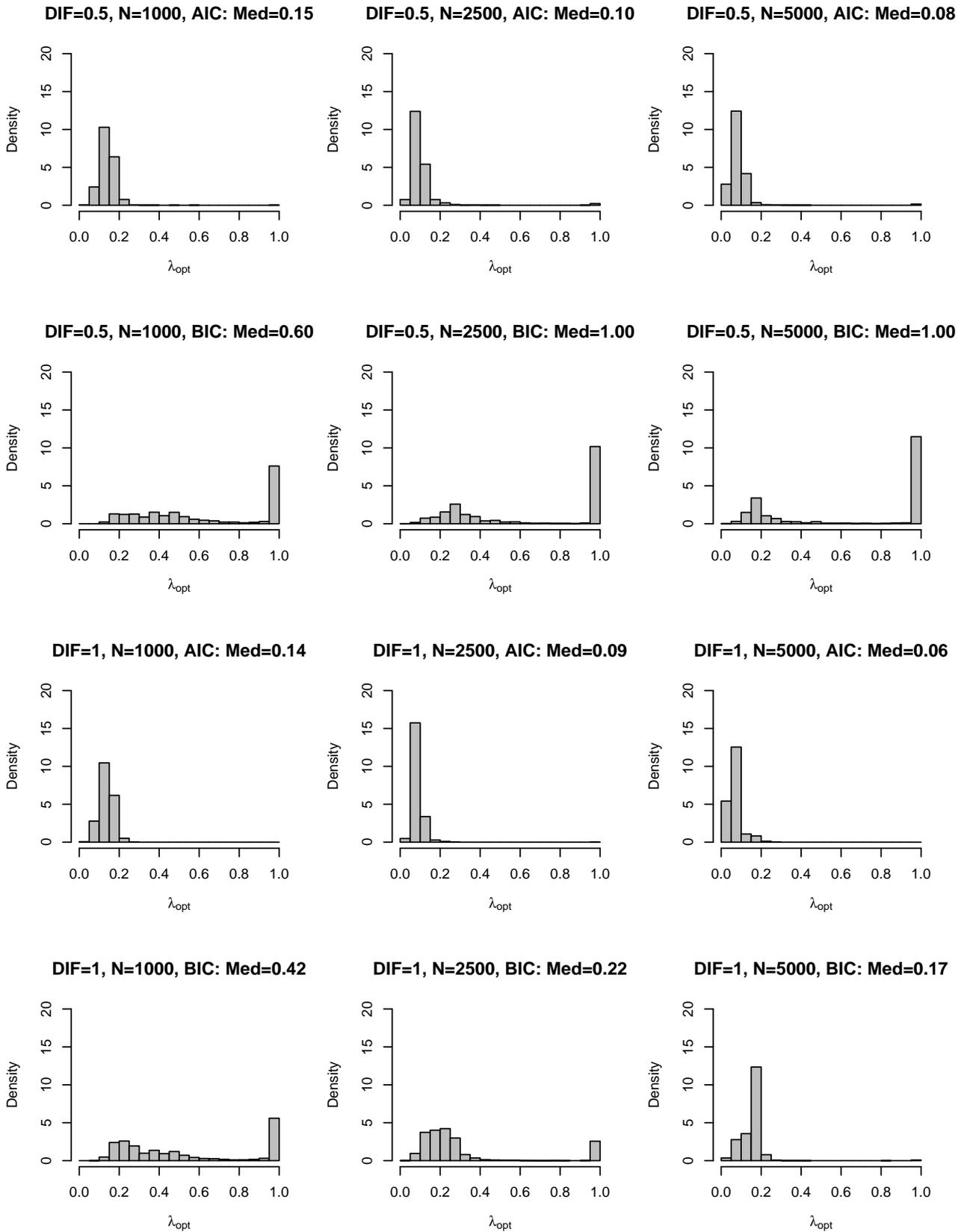


Figure 1. Simulation Study 1: Empirical histograms of optimal regularization parameters λ_{opt} as a function of sample size N , size of DIF effects and the chosen information criterion (i.e., AIC or BIC).

Table 3 displays average absolute biases and RMSE for different parameters or averaged across groups of parameters. In general, bias and RMSE were reduced in larger samples and were smaller in the presence of large DIF effects than for small DIF effects. Interestingly, and in coherence with a statement in [51], bias and RMSE for model parameters can be smaller for a fixed regularization parameter (i.e., for $\lambda = 0.10$) compared with model selection based on AIC or BIC. The property of thresholding parameter estimates to zero is helpful in parameter selection (i.e., detecting DIF effects) but has disadvantages for statistical frequentist properties of bias and RMSE. It remains to be investigated whether a noticeable increase in type I error rates for a fixed regularization parameter λ is of concern in applications of the RMRM. Overall, bias and RMSE were smaller if the model selection was carried out based on AIC instead of BIC.

Table 3. Simulation Study 1: Average absolute bias (bias) and root mean square error of model parameters.

| Par | DIF | N | Bias | | | | | RMSE | | | | |
|------------|-----|------|---------------------|------|------|------|------|---------------------|------|------|------|------|
| | | | Choice of λ | | | | | Choice of λ | | | | |
| | | | AIC | BIC | 0.05 | 0.1 | 0.15 | AIC | BIC | 0.05 | 0.1 | 0.15 |
| b_i | 0.5 | 1000 | 0.04 | 0.05 | 0.03 | 0.03 | 0.04 | 0.17 | 0.17 | 0.16 | 0.16 | 0.17 |
| | | 2500 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 |
| | | 5000 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |
| | 1 | 1000 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.16 | 0.17 | 0.16 | 0.16 | 0.17 |
| | | 2500 | 0.04 | 0.05 | 0.04 | 0.04 | 0.04 | 0.10 | 0.11 | 0.10 | 0.10 | 0.10 |
| | | 5000 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.08 | 0.07 | 0.08 | 0.08 | 0.07 |
| μ_2 | 0.5 | 1000 | 0.15 | 0.19 | 0.12 | 0.13 | 0.16 | 0.34 | 0.37 | 0.30 | 0.31 | 0.34 |
| | | 2500 | 0.09 | 0.11 | 0.08 | 0.09 | 0.10 | 0.28 | 0.29 | 0.27 | 0.27 | 0.28 |
| | | 5000 | 0.05 | 0.07 | 0.05 | 0.05 | 0.07 | 0.22 | 0.23 | 0.21 | 0.21 | 0.22 |
| | 1 | 1000 | 0.09 | 0.13 | 0.07 | 0.08 | 0.10 | 0.29 | 0.33 | 0.26 | 0.27 | 0.30 |
| | | 2500 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.17 | 0.18 | 0.16 | 0.17 | 0.17 |
| | | 5000 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| σ_1 | 0.5 | 1000 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.10 | 0.10 | 0.09 | 0.09 | 0.10 |
| | | 2500 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 |
| | | 5000 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| | 1 | 1000 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 | 0.09 | 0.10 | 0.09 | 0.09 | 0.09 |
| | | 2500 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.06 | 0.05 | 0.05 | 0.05 |
| | | 5000 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| σ_2 | 0.5 | 1000 | 0.06 | 0.05 | 0.06 | 0.06 | 0.06 | 0.16 | 0.16 | 0.15 | 0.15 | 0.16 |
| | | 2500 | 0.05 | 0.03 | 0.05 | 0.05 | 0.05 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| | | 5000 | 0.03 | 0.03 | 0.04 | 0.04 | 0.03 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| | 1 | 1000 | 0.04 | 0.03 | 0.04 | 0.04 | 0.04 | 0.14 | 0.15 | 0.14 | 0.14 | 0.14 |
| | | 2500 | 0.03 | 0.02 | 0.03 | 0.03 | 0.03 | 0.08 | 0.09 | 0.08 | 0.08 | 0.08 |
| | | 5000 | 0.03 | 0.02 | 0.03 | 0.03 | 0.02 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |
| p_2 | 0.5 | 1000 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 | 0.03 | 0.04 | 0.04 | 0.04 |
| | | 2500 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 | 0.03 | 0.04 | 0.04 | 0.04 |
| | | 5000 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 |
| | 1 | 1000 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 | 0.04 | 0.05 | 0.05 | 0.04 |
| | | 2500 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| | | 5000 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 |

Table 3. *Cont.*

| Par | DIF | N | Bias | | | RMSE | | | | | | |
|----------------|-----|------|---------------------|------|------|---------------------|------|------|------|------|------|------|
| | | | Choice of λ | | | Choice of λ | | | | | | |
| | | | AIC | BIC | 0.05 | 0.1 | 0.15 | AIC | BIC | 0.05 | 0.1 | 0.15 |
| e_i (no DIF) | 0.5 | 1000 | 0.04 | 0.01 | 0.03 | 0.04 | 0.05 | 0.57 | 0.20 | 0.60 | 0.59 | 0.56 |
| | | 2500 | 0.03 | 0.00 | 0.02 | 0.03 | 0.03 | 0.36 | 0.10 | 0.38 | 0.37 | 0.34 |
| | | 5000 | 0.02 | 0.00 | 0.02 | 0.02 | 0.02 | 0.25 | 0.07 | 0.27 | 0.25 | 0.21 |
| | 1 | 1000 | 0.05 | 0.01 | 0.04 | 0.04 | 0.05 | 0.46 | 0.19 | 0.50 | 0.48 | 0.45 |
| | | 2500 | 0.01 | 0.00 | 0.01 | 0.02 | 0.01 | 0.23 | 0.09 | 0.26 | 0.23 | 0.18 |
| | | 5000 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.15 | 0.05 | 0.17 | 0.11 | 0.05 |
| e_i (DIF) | 0.5 | 1000 | 0.22 | 0.46 | 0.14 | 0.17 | 0.25 | 0.62 | 0.54 | 0.59 | 0.60 | 0.63 |
| | | 2500 | 0.15 | 0.47 | 0.09 | 0.14 | 0.22 | 0.47 | 0.51 | 0.42 | 0.47 | 0.51 |
| | | 5000 | 0.13 | 0.45 | 0.09 | 0.15 | 0.27 | 0.40 | 0.50 | 0.35 | 0.41 | 0.48 |
| | 1 | 1000 | 0.26 | 0.77 | 0.19 | 0.22 | 0.31 | 0.67 | 0.95 | 0.56 | 0.61 | 0.72 |
| | | 2500 | 0.08 | 0.37 | 0.07 | 0.09 | 0.12 | 0.37 | 0.68 | 0.34 | 0.37 | 0.44 |
| | | 5000 | 0.06 | 0.08 | 0.06 | 0.06 | 0.07 | 0.23 | 0.29 | 0.23 | 0.23 | 0.28 |

Note. Par = parameter group; |DIF| = absolute value of DIF effects e_i ; N = sample size; λ = regularization parameter.

Table 3 only contains a number of selected values of the regularization parameter λ . In Figure 2, the RMSE of parameters μ_2 , σ_2 , and p_2 are displayed and compared with the RMSE based on the optimal regularization parameter obtained from AIC or BIC. It can be seen that small fixed λ values were competitive with optimal regularization parameters in terms of RMSE. The situation slightly differs for the p_2 parameter. For moderate sample sizes $N = 1000$ or $N = 2500$, very large λ values near to one led to the lowest RMSE values.

Finally, Figure 3 presents the RMSE for parameter groups b_i (parameters “b”), latent DIF effects e_i with a true value of zero (i.e., non-DIF effects; parameters “e_nodif”), and latent DIF effects e_i with a true value different from zero (parameters “e_dif”) for selected values of the regularization parameter λ . It can be beneficial for item difficulties b_i in terms of RMSE if small fixed λ values are chosen. Obviously, using a large λ value for non-DIF effects is advantageous because these parameters would be correctly regularized. However, a large fixed λ value comes at the price of not detecting true DIF effects. To sum up, these findings illustrate that choosing a fixed λ value could outperform AIC- or BIC-based regularized estimation if RMSE were the statistical criterion that would drive the estimator choice.

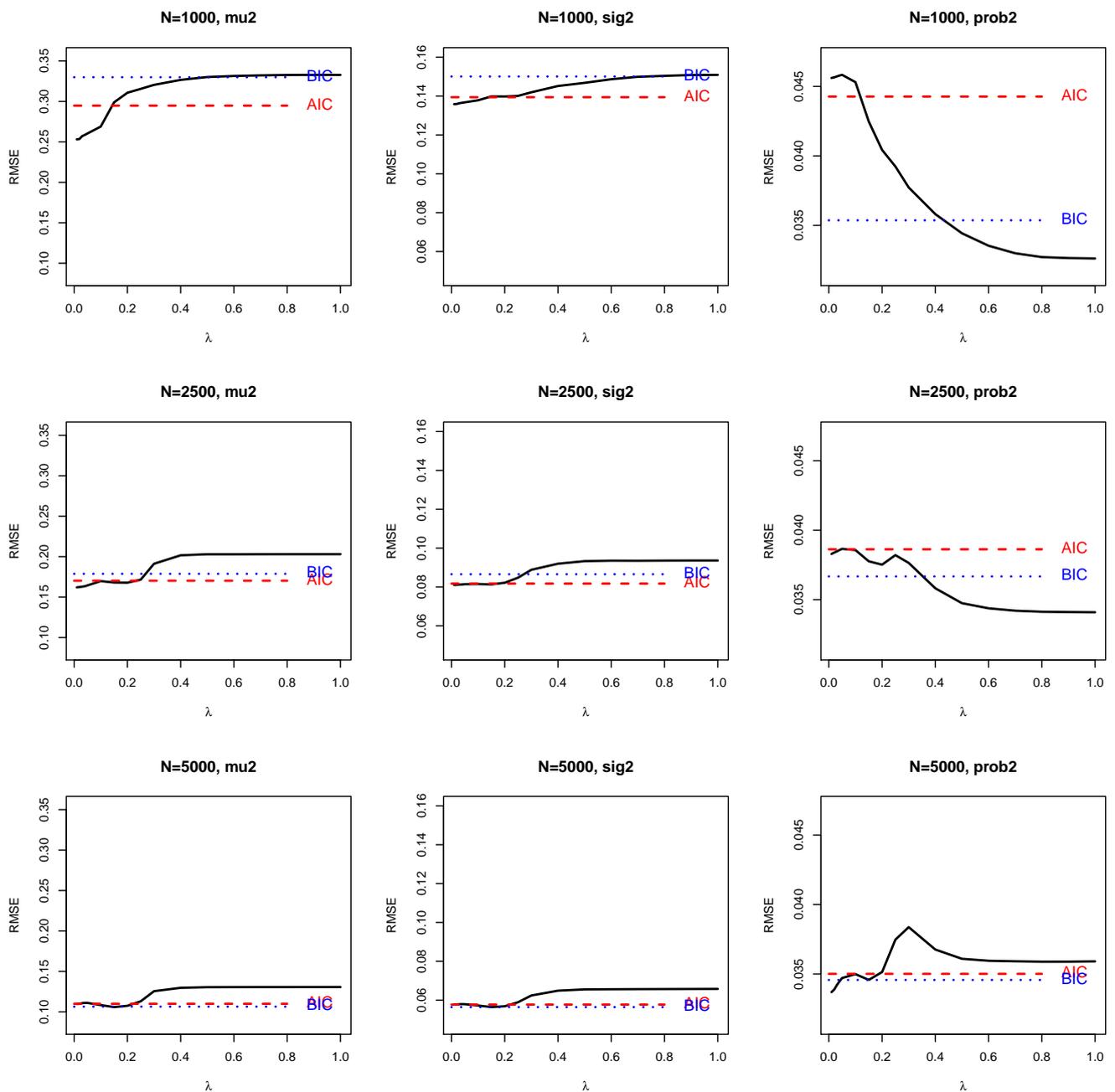


Figure 2. Simulation Study 1: RMSE for parameters μ_2 (mu2), σ_2 (sig2), and p_2 (prob2) as a function of a fixed regularization parameter λ and optimal regularization parameters obtained from AIC or BIC.

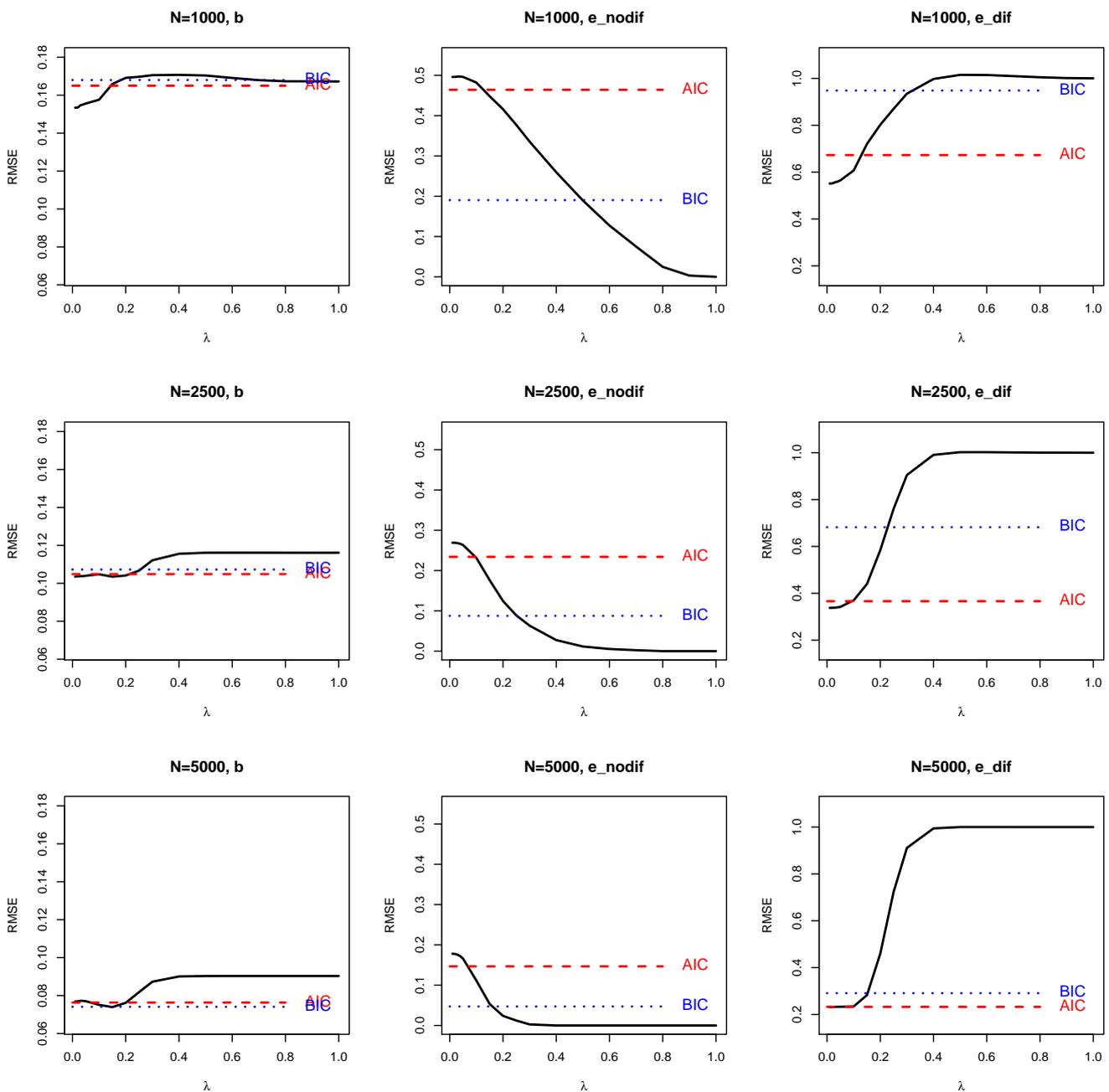


Figure 3. Simulation Study 1: Average RMSE for parameter groups of item difficulties b_i (b), DIF effects e_i with true values of zero (e_nodif), DIF effects e_i with true values different from zero (e_dif) as a function of a fixed regularization parameter λ and optimal regularization parameters obtained from AIC or BIC.

4. Simulated Case Study 2: Illustrative Example with a Nonspeeded and a Speeded Latent Class

In a linear fixed administered test, there is a fixed order of test items. Frequently, items at later test positions are prone to position effects; that is, they are difficultly compared with the situation if they were administered at earlier test positions. Similarly, test takers can show a performance decline [52–54]. This means that persons show lower performance at the end of the test compared with the beginning of the test. Importantly, the extent of performance decline can vary across persons [55,56].

Performance decline can occur if the test is speeded; that is, not all test takers reach the end of the test due to low item processing, limited testing time, or a lack of motivation.

MRMs have been proposed for handling speededness effects [57]. Bolt [57] proposed to use a MRM with two latent classes. The first class refers to the nonspeeded class of test takers, while the second class refers to the speeded class of test takers. The speeded class is typically characterized by increased item difficulties for items at the end of the test [57]. In this simulated case study, we assume two latent classes in the MRM, where the class-specific item difficulties are modeled as

$$\begin{aligned} b_{i1} &= b_{i0} \\ b_{i2} &= b_{i0} + e_i \end{aligned} \tag{14}$$

In the simulated dataset, we used item parameters adapted from [57]. The item difficulties are shown in Figure 4 and numerically presented in Table 4. In total, there are 26 test items. Only items 19 to 26 were prone to speededness effects and had DIF effects e_i larger than zero, while items 1 to 18 had equal item difficulties in the two latent classes (i.e., they had no DIF effects). The nonspeeded class had a class probability of $p_1 = 0.75$, and the speeded class had a probability $p_2 = 0.25$. The means of the two classes in the MRM were $\mu_1 = 0$ and $\mu_2 = -0.4$, respectively. Hence, the speeded class had a lower ability on average. Moreover, the standard deviations were set to $\sigma_1 = 1.1$ and $\sigma_2 = 1.4$, respectively.

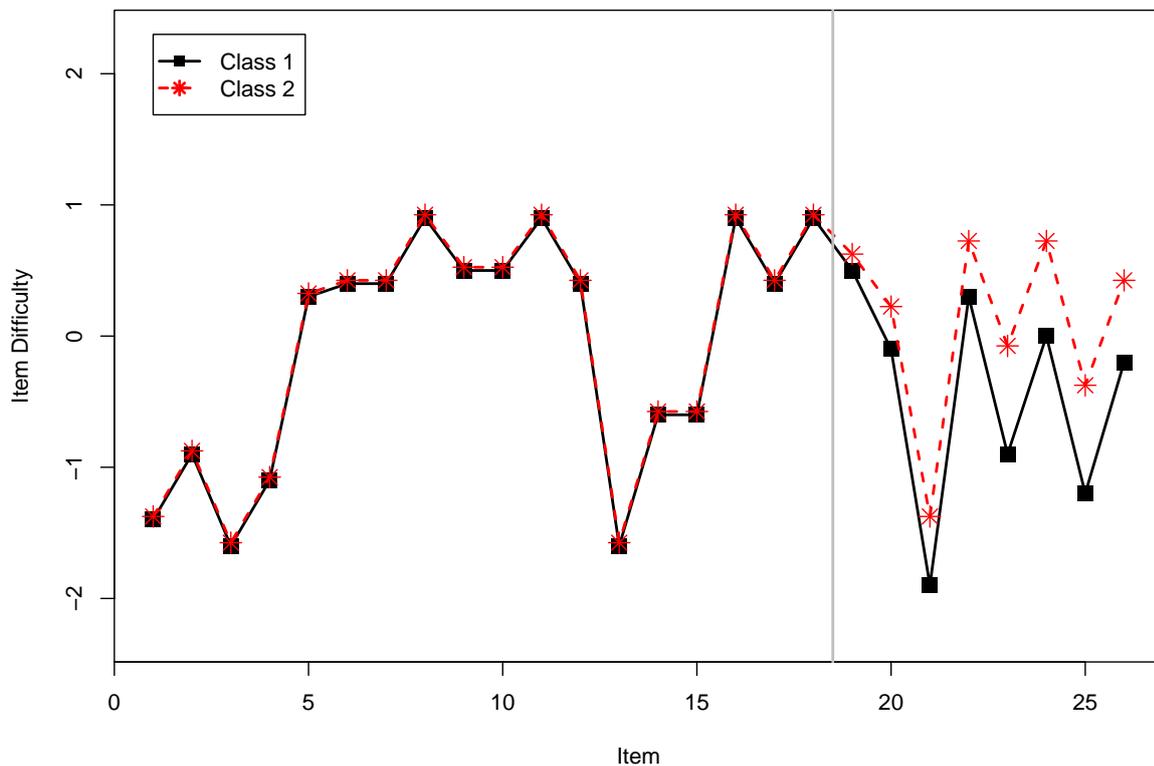


Figure 4. Simulated Case Study 2: True item difficulties b_{ic} .

Table 4. Simulated Case Study 2: True and estimated item parameters.

| Item | b_i | | | e_i | | p_{boot} |
|------|-------|-------|------|-------|-------|------------|
| | True | Est | SE | True | Est | |
| 1 | −1.4 | −1.31 | 0.16 | 0.0 | 0.00 | 0.74 |
| 2 | −0.9 | −0.85 | 0.15 | 0.0 | 0.00 | 0.81 |
| 3 | −1.6 | −1.59 | 0.16 | 0.0 | 0.00 | 0.82 |
| 4 | −1.1 | −1.02 | 0.17 | 0.0 | 0.00 | 0.77 |
| 5 | 0.3 | 0.32 | 0.20 | 0.0 | 0.00 | 0.59 |
| 6 | 0.4 | 0.44 | 0.17 | 0.0 | 0.00 | 0.77 |
| 7 | 0.4 | 0.50 | 0.15 | 0.0 | 0.00 | 0.86 |
| 8 | 0.9 | 0.95 | 0.15 | 0.0 | 0.00 | 0.83 |
| 9 | 0.5 | 0.56 | 0.21 | 0.0 | 0.00 | 0.63 |
| 10 | 0.5 | 0.58 | 0.18 | 0.0 | 0.00 | 0.81 |
| 11 | 0.9 | 0.94 | 0.15 | 0.0 | 0.00 | 0.84 |
| 12 | 0.4 | 0.56 | 0.29 | 0.0 | −0.34 | 0.18 |
| 13 | −1.6 | −1.68 | 0.17 | 0.0 | 0.25 | 0.65 |
| 14 | −0.6 | −0.75 | 0.27 | 0.0 | 0.49 | 0.20 |
| 15 | −0.6 | −0.54 | 0.17 | 0.0 | 0.00 | 0.85 |
| 16 | 0.9 | 1.01 | 0.20 | 0.0 | 0.00 | 0.60 |
| 17 | 0.4 | 0.53 | 0.20 | 0.0 | −0.27 | 0.72 |
| 18 | 0.9 | 1.04 | 0.22 | 0.0 | −0.24 | 0.36 |
| 19 | 0.5 | 0.59 | 0.16 | 0.1 | 0.00 | 0.65 |
| 20 | −0.1 | 0.03 | 0.15 | 0.3 | 0.00 | 0.87 |
| 21 | −1.9 | −1.75 | 0.18 | 0.5 | 0.00 | 0.85 |
| 22 | 0.3 | 0.22 | 0.18 | 0.4 | 0.43 | 0.67 |
| 23 | −0.9 | −0.80 | 0.25 | 0.8 | 0.40 | 0.35 |
| 24 | 0.0 | 0.01 | 0.22 | 0.7 | 0.29 | 0.23 |
| 25 | −1.2 | −1.42 | 0.27 | 0.8 | 1.03 | 0.23 |
| 26 | −0.2 | −0.22 | 0.23 | 0.6 | 0.57 | 0.31 |

Note. b_i = item difficulty; e_i = DIF effect; True = true item parameters; Est = estimated item parameters; SE = standard error estimated by nonparametric bootstrap; p_{boot} = bootstrap probability of obtaining an estimate equal to zero.

A dataset of a sample size $N = 6000$ was generated. We estimated an RMRM using the parameterization (14). We used the identification constraint $\mu_1 = 0$. The regularization parameter λ was specified on an equidistant grid of values between 0.50 and 0.01 with decrements of 0.01. Replication material and the dataset can be found at <https://osf.io/wrs5k/> (accessed on 30 September 2022).

For illustrating standard error computation, we used a nonparametric bootstrap with 500 bootstrap samples. We determined the standard error by using the robust scale parameter median absolute deviation (MAD implemented with the R function `stats::mad()`; [58]) of bootstrap parameter estimates to diminish the potential effect of outliers.

Figure 5 displays the AIC as a function of increasing values of the regularization parameter λ . It turned out that $\lambda = 0.06$ provided the least AIC value. Therefore, we report results based on this regularization parameter.

The regularization paths for DIF effects e_i are displayed in Figure 6. With increasing values of λ , fewer DIF effects were estimated as nonzero. For example, for $\lambda = 0.20$, only one estimated DIF effect differed from zero.

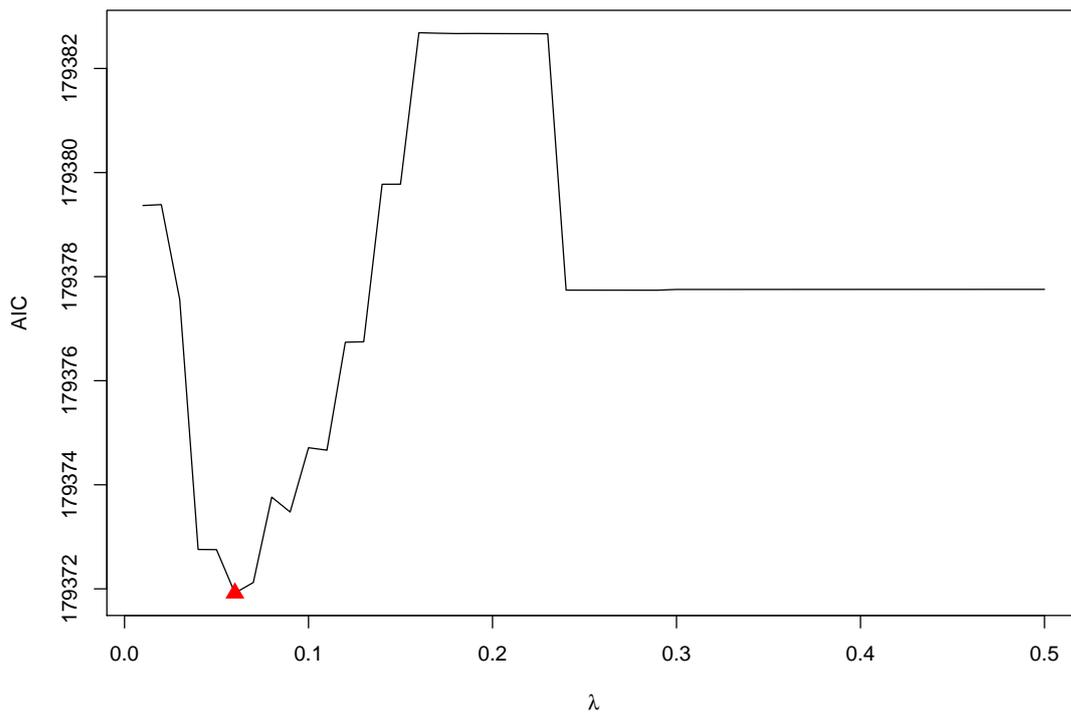


Figure 5. Simulated Case Study 2: AIC as a function of the regularization parameter λ . The red triangle corresponds to the optimale λ value with minimal AIC.

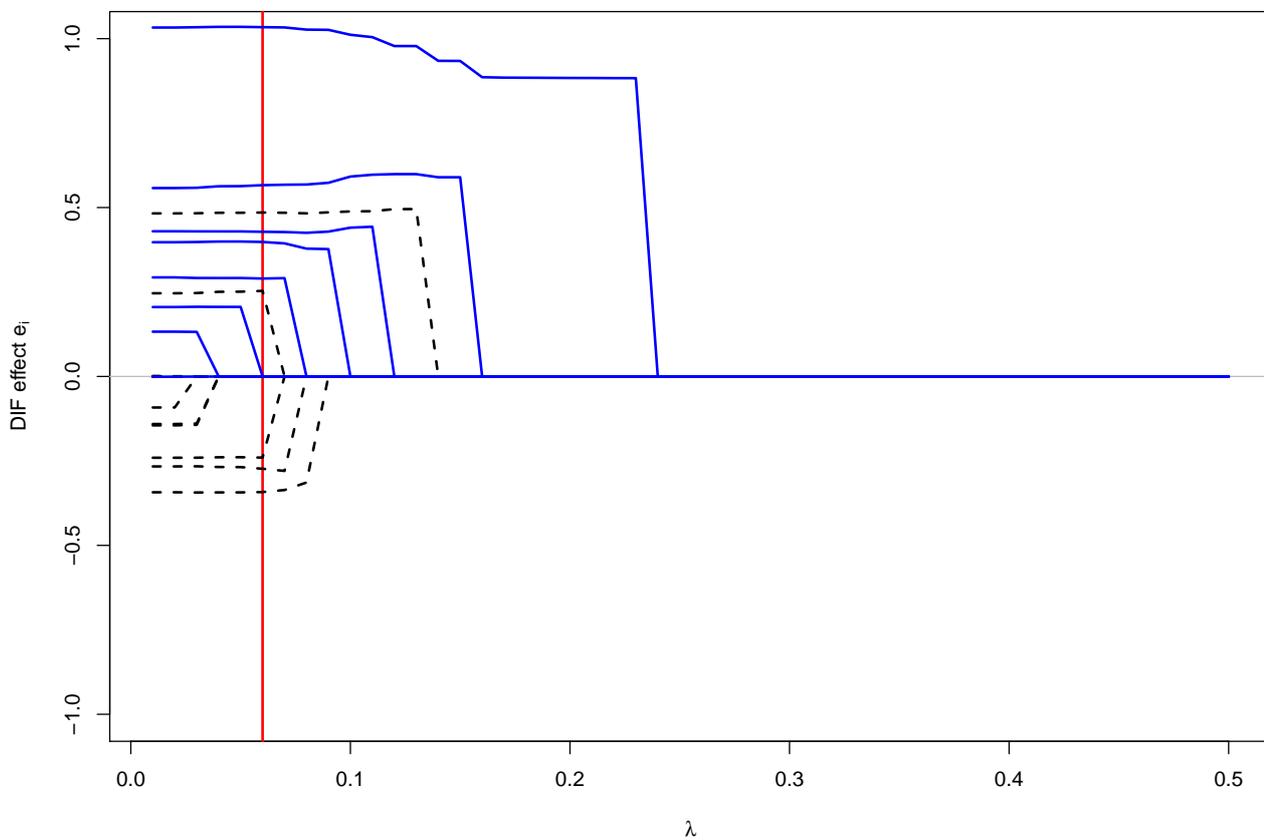


Figure 6. Simulated Case Study 2: DIF effects e_i as a function of the regularization parameter λ . Regularization paths for DIF effects are printed in blue solid lines, while paths for non-DIF effects are shown in black dashed lines.

The standard deviation of the first class was estimated as $\hat{\sigma}_1 = 0.94$ (SE = 0.10), which somehow differed from the true value $\sigma_1 = 1.1$. The second speeded latent class had the following estimated parameters, which closely resembled the data-generating parameters: $\hat{p}_2 = 0.36$ (SE = 0.16, true: $p_2 = 0.25$), $\hat{\mu}_2 = -0.15$ (SE = 0.33, true: $\mu_2 = -0.4$), and $\hat{\sigma}_2 = 1.38$ (SE = 0.18, true: $\sigma_2 = 1.40$).

The estimated item parameters are shown in Table 4. It can be seen that for the 8 DIF items, 5 DIF effects e_i were correctly estimated as different from zero, while 3 DIF items had estimated DIF effects of 0. Notably, 5 non-DIF items (i.e., items 12, 13, 14, 17, and 18) had estimated DIF effects different from zero. Overall, the estimated common item difficulties \hat{b}_{0i} were close to the data-generating values. In accordance with findings in Simulation Study 1, the detection of DIF effects based on the BIC was less satisfactory than based on the AIC. Based on this illustrative study, it turned out that bootstrap probabilities p_{boot} (see Section 2.3) were substantially larger than 0.05 for true latent DIF effects that were estimated as different from zero.

5. Simulated Case Study 3: Illustrative Example Involving Three Latent Classes

In Simulated Case Study 3, we simulate data from an MRM with three latent classes. Bolt [59] presented an application in which sparse DIF effects occur. Figure 7 shows the data-generating item difficulties for the simulated dataset in this Simulated Case Study 3 that were adapted from [59]. It can be seen that many of the class-specific item difficulties were equal to each other. RMRM can be used to effectively estimate Rasch mixtures under some sparsity assumptions of DIF effects.

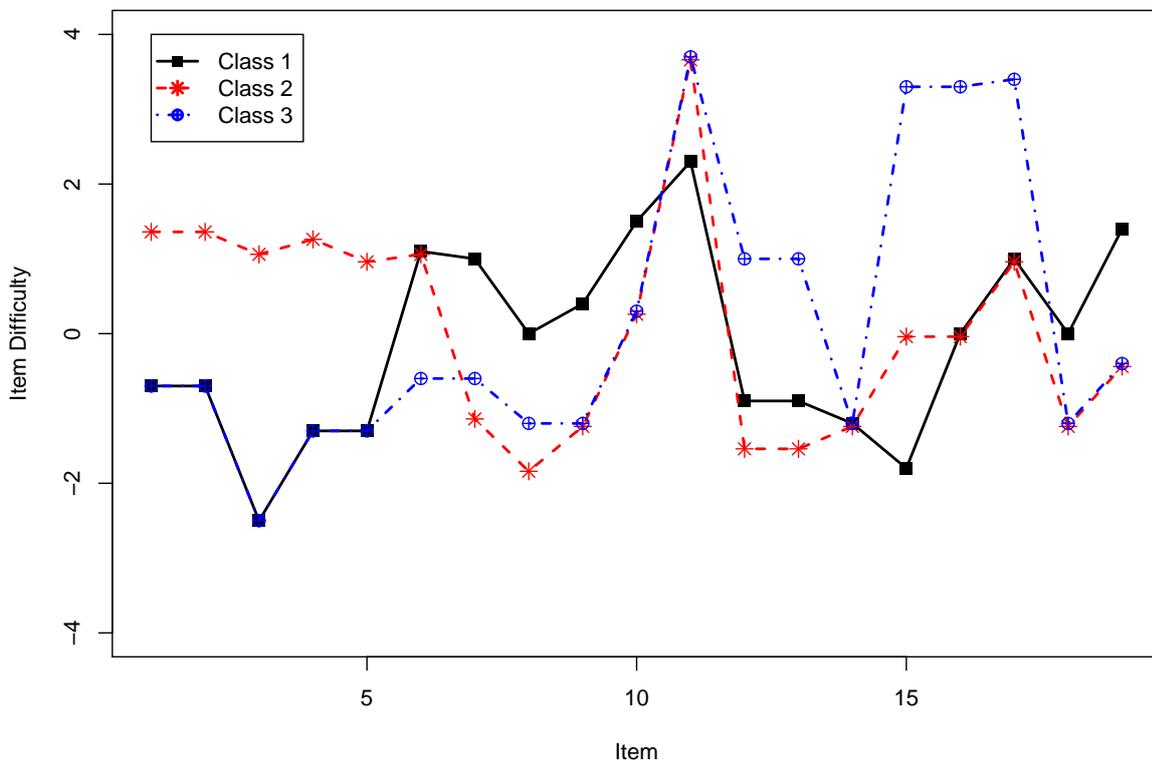


Figure 7. Simulated Case Study 3: True item difficulties b_{ic} .

The simulated dataset had a sample size $N = 5000$ and $I = 19$ items. The data-generating item parameters can be found in Table 5. The class-specific distribution parameters were $\mu_1 = 0, \sigma_1 = 1, p_1 = 0.45$ for Class 1, $\mu_2 = 0.8, \sigma_2 = 0.7, p_2 = 0.35$ for Class 2, and $\mu_3 = -0.5, \sigma_3 = 1.2, p_3 = 0.2$ for Class 3.

Table 5. Simulated Case Study 3: True and estimated item parameters.

| Item | True | | | Est Fused Reg b_{ic} | | | Est Reg e_{ic} | | |
|------|-------------|-------------|-------------|------------------------|--------------|--------------|------------------|--------------|--------------|
| | b_{i1} | b_{i2} | b_{i3} | b_{i1} | b_{i2} | b_{i3} | b_{i1} | b_{i2} | b_{i3} |
| 1 | -0.7 | 1.4 | -0.7 | -0.74 | 1.47 | -0.74 | -0.72 | 1.44 | -0.72 |
| 2 | -0.7 | 1.4 | -0.7 | -0.70 | 1.44 | -0.70 | -0.68 | 1.41 | -0.68 |
| 3 | -2.5 | 1.1 | -2.5 | -2.62 | 1.13 | -2.19 | -2.61 | 1.10 | -2.18 |
| 4 | -1.3 | 1.3 | -1.3 | -1.29 | 1.33 | -1.29 | -1.27 | 1.30 | -1.27 |
| 5 | -1.3 | 1.0 | -1.3 | -1.32 | 1.07 | -1.32 | -1.30 | 1.05 | -1.30 |
| 6 | 1.1 | 1.1 | -0.6 | 1.17 | 1.17 | -0.57 | 1.17 | 1.17 | -0.56 |
| 7 | 1.0 | -1.1 | -0.6 | 1.03 | -1.21 | -0.64 | 1.04 | -1.24 | -0.62 |
| 8 | 0.0 | -1.8 | -1.2 | -0.05 | -1.73 | -1.34 | -0.05 | -1.76 | -1.31 |
| 9 | 0.4 | -1.2 | -1.2 | 0.42 | -1.16 | -1.16 | 0.42 | -1.16 | -1.16 |
| 10 | 1.5 | 0.3 | 0.3 | 1.45 | 0.30 | 0.30 | 1.45 | 0.29 | 0.29 |
| 11 | 2.3 | 3.7 | 3.7 | 2.34 | 3.79 | 3.79 | 2.34 | 3.81 | 3.81 |
| 12 | -0.9 | -1.5 | 1.0 | -0.84 | -1.41 | 1.05 | -0.83 | -1.43 | 1.08 |
| 13 | -0.9 | -1.5 | 1.0 | -0.92 | -1.42 | 1.15 | -0.91 | -1.44 | 1.17 |
| 14 | -1.2 | -1.2 | -1.2 | -1.16 | -1.16 | -1.16 | -1.15 | -1.15 | -1.15 |
| 15 | -1.8 | 0.0 | 3.3 | -1.70 | 0.19 | 2.89 | -1.69 | 0.16 | 3.06 |
| 16 | 0.0 | 0.0 | 3.3 | 0.07 | 0.07 | 3.21 | 0.07 | 0.07 | 3.24 |
| 17 | 1.0 | 1.0 | 3.4 | 1.00 | 1.00 | 3.20 | 0.99 | 0.99 | 3.22 |
| 18 | 0.0 | -1.2 | -1.2 | 0.02 | -1.26 | -1.26 | 0.02 | -1.26 | -1.26 |
| 19 | 1.4 | -0.4 | -0.4 | 1.49 | -0.48 | -0.48 | 1.49 | -0.63 | -0.33 |

Note. True = true item parameters; Est Fused Reg b_{ic} = Estimated item parameters using fused regularization for item difficulties b_{ic} ; Est Reg e_{ic} = Estimated item parameters using SCAD regularization for DIF effects e_{ic} ; b_{ic} = item difficulty of item i in class c . Correctly detected DIF effects are printed in black bold font. Incorrectly detected DIF effects are printed in red bold font.

We estimated the RMRM in two variants. First, we applied fused regularization to item parameter differences $b_{ic} - b_{ic'}$ (see Equation (8)). Second, we used SCAD regularization for class-specific DIF effects e_{ic} based on the decomposition (9) and the regularized likelihood function (11). We used the identification constraint $\hat{\mu}_1 = 0$ in model estimation. Replication material can be found at <https://osf.io/wrs5k/> (accessed on 30 September 2022). We did not carry out a bootstrap to compute standard errors because it would require some computational effort, and our primary interests were only interpretational purposes.

The optimal regularization parameter λ was chosen using the least AIC value. It was $\lambda = 0.07$ for the two estimation approaches. Overall, it turned out that the estimated model parameters were very close in the two estimation approaches. The estimated distribution parameters for fused regularization were $\hat{\mu}_1 = 0$, $\hat{\sigma}_1 = 1.03$, $\hat{\rho}_1 = 0.53$ for Class 1, $\hat{\mu}_2 = 0.83$, $\hat{\sigma}_2 = 0.73$, $\hat{\rho}_2 = 0.33$ for Class 2, and $\hat{\mu}_3 = -0.56$, $\hat{\sigma}_3 = 1.12$, $\hat{\rho}_3 = 0.14$ for Class 3. The estimated distribution parameters for SCAD regularization for DIF effects e_{ic} were $\hat{\mu}_1 = 0$, $\hat{\sigma}_1 = 1.03$, $\hat{\rho}_1 = 0.53$ for Class 1, $\hat{\mu}_2 = 0.81$, $\hat{\sigma}_2 = 0.74$, $\hat{\rho}_2 = 0.33$ for Class 2, and $\hat{\mu}_3 = -0.54$, $\hat{\sigma}_3 = 1.12$, $\hat{\rho}_3 = 0.14$ for Class 3.

The estimated class-specific item difficulties are displayed in Table 5. It is evident that the pattern of true DIF effects was perfectly detected by fused and SCAD regularization. Moreover, one item (item 3) or two items (items 3 and 19) were detected to possess additional DIF effects that were not simulated for fused and SCAD regularization, respectively. Overall, the item parameter differences between the two estimation approaches were negligible.

6. Discussion

In this article, we proposed a regularized estimation approach of the mixture Rasch model. By putting a regularization penalty on differences in class-specific item difficulties or on latent DIF effects, the interpretability of latent classes in the mixture Rasch model is substantially eased. The regularization technique enables the automatic detection of latent DIF effects and provides a parsimonious model selection.

In the simulation study that involves two latent classes, model selection based on AIC tended to outperform model selection based on BIC. With AIC, there is a tendency to overestimate the number of DIF effects. At the same time, model selection based on BIC substantially underestimates the number of DIF effects. Which of the two criteria should be used in practice depends on the choice of how large type I error rates for non-DIF effects should be tolerated while guaranteeing sufficiently large power rates for the detection of DIF effects. In our view, AIC should be preferred because BIC would result in too many true DIF effects that remain undetected.

We presented two case studies to illustrate the potential of regularized mixture Rasch models. With sufficiently large sample sizes and using AIC for model selection, we successfully recovered the data-generating structure of DIF effects. Our observation that BIC should not be universally preferred over AIC was also confirmed in other research on the mixture Rasch model [43]. Moreover, we could also replicate this finding for other classes of item response models in our research [60].

We limited our simulation study only to sample sizes larger than 1000. Much smaller sample sizes might be interesting in applied research. However, we think that the maximum likelihood estimation of mixture models should involve large sample sizes (say, at least larger than 500) to ensure a sufficiently stable estimation of model parameters. Investigating the limits of applying the regularized mixture Rasch model might be an interesting topic of future research.

The computation of standard errors by nonparametric bootstrap has only been illustrated in Simulated Case Study 2. In future research, different standard error computation methods for estimating regularized mixture Rasch models might be investigated.

As with any newly proposed statistical technique, the future will tell whether the regularization approach can prove helpful in empirical applications. We think that this technique provides a means for obtaining more interpretable and less variable class-specific item parameter estimates. Likely, the regularization approach can also be applied to other classes of mixture latent variable models, such as the two- or three-parameter mixture logistic item response model or factor mixture models.

In conclusion, we believe that regularized mixture Rasch models can be used in exploratory analysis in the same way as nonregularized mixture Rasch models. We recognize the primary potential of regularization in obtaining more structured (and more stable) results if the true class-specific item difficulties follow a sparsity assumption. This assumption might not be realistic in all applications. However, one can at least include the regularized mixture Rasch model in the researcher's toolbox for analyzing dichotomous item responses.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets can be found at <https://osf.io/wrs5k/> (accessed on 30 September 2022).

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-----|--------------------------------|
| AIC | Akaike information criterion |
| BIC | Bayesian information criterion |
| DIF | differential item functioning |
| EM | expectation maximization |
| MRM | mixture Rasch model |
| RM | Rasch model |

| | |
|------|-------------------------------------|
| RMRM | regularized mixture Rasch model |
| RMSE | root mean square error |
| SCAD | smoothly clipped absolute deviation |

References

- Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Danish Institute for Educational Research: Copenhagen, Denmark, 1960.
- von Davier, M. The Rasch model. In *Handbook of Item Response Theory, Volume 1: Models*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 31–48. [[CrossRef](#)]
- Debelak, R.; Strobl, C.; Zeigenfuss, M.D. *An Introduction to the Rasch Model with Examples in R*; CRC Press: Boca Raton, FL, USA, 2022. [[CrossRef](#)]
- Robitzsch, A. A comprehensive simulation study of estimation methods for the Rasch model. *Stats* **2021**, *4*, 814–836. [[CrossRef](#)]
- Xu, X.; von Davier, M. *Fitting the Structured General Diagnostic Model to NAEP Data*; Research Report No. RR-08-28; Educational Testing Service: Princeton, NJ, USA, 2008. [[CrossRef](#)]
- Rost, J. Rasch models in latent classes: An integration of two approaches to item analysis. *Appl. Psychol. Meas.* **1990**, *14*, 271–282. [[CrossRef](#)]
- von Davier, M.; Rost, J. Mixture distribution item response models. In *Handbook of Statistics, Vol. 26: Psychometrics*; Rao, C.R., Sinharay, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; Volume 26, pp. 643–661. [[CrossRef](#)]
- Frick, H.; Strobl, C.; Leisch, F.; Zeileis, A. Flexible Rasch mixture models with package psychomix. *J. Stat. Softw.* **2012**, *48*, 1–25. [[CrossRef](#)]
- von Davier, M. *Mixture Distribution Diagnostic Models*; (Research Report No. RR-07-32); Educational Testing Service: Princeton, NJ, USA, 2007. [[CrossRef](#)]
- Paek, I.; Cho, S.J. A note on parameter estimate comparability: Across latent classes in mixture IRT modeling. *Appl. Psychol. Meas.* **2015**, *39*, 135–143. [[CrossRef](#)]
- Bulut, O.; Suh, Y. Detecting multidimensional differential item functioning with the multiple indicators multiple causes model, the item response theory likelihood ratio test, and logistic regression. *Front. Educ.* **2017**, *2*, 51. [[CrossRef](#)]
- Holland, P.W.; Wainer, H. (Eds.) *Differential Item Functioning: Theory and Practice*; Lawrence Erlbaum: Hillsdale, NJ, USA, 1993. [[CrossRef](#)]
- Penfield, R.D.; Camilli, G. Differential item functioning and item bias. In *Handbook of Statistics, Volume 26: Psychometrics*; Rao, C.R., Sinharay, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 125–167. [[CrossRef](#)]
- Cho, S.J.; Suh, Y.; Lee, W. An NCME instructional module on latent DIF analysis using mixture item response models. *Educ. Meas.* **2016**, *35*, 48–61. [[CrossRef](#)]
- Frick, H.; Strobl, C.; Zeileis, A. Rasch mixture models for DIF detection: A comparison of old and new score specifications. *Educ. Psychol. Meas.* **2015**, *75*, 208–234. [[CrossRef](#)] [[PubMed](#)]
- Rost, J. A logistic mixture distribution model for polychotomous item responses. *Br. J. Math. Stat. Psychol.* **1991**, *44*, 75–92. [[CrossRef](#)]
- von Davier, M.; Rost, J. Polytomous mixed Rasch models. In *Rasch Models*; Fischer, G.H., Molenaar, I.W., Eds.; Springer: New York, NY, USA, 1995; pp. 371–379. [[CrossRef](#)]
- Choi, Y.J.; Alexeev, N.; Cohen, A.S. Differential item functioning analysis using a mixture 3-parameter logistic model with a covariate on the TIMSS 2007 mathematics test. *Int. J. Test.* **2015**, *15*, 239–253. [[CrossRef](#)]
- Formann, A.K.; Kohlmann, T. Structural latent class models. *Sociol. Methods Res.* **1998**, *26*, 530–565. [[CrossRef](#)]
- Formann, A.K.; Kohlmann, T. Three-parameter linear logistic latent class analysis. In *Applied Latent Class Analysis*; Hagenaaers, J.A., McCutcheon, A.L., Eds.; Cambridge University Press: Cambridge, UK, 2002; pp. 183–210.
- Muthén, B.; Asparouhov, T. Item response mixture modeling: Application to tobacco dependence criteria. *Addict. Behav.* **2006**, *31*, 1050–1066. [[CrossRef](#)] [[PubMed](#)]
- Revuelta, J. Estimating the π^* goodness of fit index for finite mixtures of item response models. *Br. J. Math. Stat. Psychol.* **2008**, *61*, 93–113. [[CrossRef](#)] [[PubMed](#)]
- Sen, S.; Cohen, A.S. Applications of mixture IRT models: A literature review. *Meas. Interdiscip. Res. Persp.* **2019**, *17*, 177–191. [[CrossRef](#)]
- Smit, A.; Kelderman, H.; van der Flier, H. The mixed Birnbaum model: Estimation using collateral information. *Methods Psychol. Res. Online* **2000**, *5*, 31–43.
- Chen, Y.; Li, X.; Liu, J.; Ying, Z. Robust measurement via a fused latent and graphical item response theory model. *Psychometrika* **2018**, *83*, 538–562. [[CrossRef](#)] [[PubMed](#)]
- Sun, J.; Chen, Y.; Liu, J.; Ying, Z.; Xin, T. Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika* **2016**, *81*, 921–939. [[CrossRef](#)]
- Huang, P.H.; Chen, H.; Weng, L.J. A penalized likelihood method for structural equation modeling. *Psychometrika* **2017**, *82*, 329–354. [[CrossRef](#)]
- Jacobucci, R.; Grimm, K.J.; McArdle, J.J. Regularized structural equation modeling. *Struct. Equ. Model.* **2016**, *23*, 555–566. [[CrossRef](#)]

29. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Regularized latent class analysis with application in cognitive diagnosis. *Psychometrika* **2017**, *82*, 660–692. [CrossRef]
30. Robitzsch, A.; George, A.C. The R package CDM for diagnostic modeling. In *Handbook of Diagnostic Classification Models*; von Davier, M., Lee, Y.S., Eds.; Springer: Cham, Switzerland, 2019; pp. 549–572. [CrossRef]
31. Robitzsch, A. Regularized latent class analysis for polytomous item responses: An application to SPM-LS data. *J. Intell.* **2020**, *8*, 30. [CrossRef]
32. Belzak, W.; Bauer, D.J. Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychol. Methods* **2020**, *25*, 673–690. [CrossRef] [PubMed]
33. Bauer, D.J.; Belzak, W.C.M.; Cole, V.T. Simplifying the assessment of measurement invariance over multiple background variables: Using regularized moderated nonlinear factor analysis to detect differential item functioning. *Struct. Equ. Model.* **2020**, *27*, 43–55. [CrossRef] [PubMed]
34. Chen, Y.; Li, C.; Xu, G. DIF statistical inference and detection without knowing anchoring items. *arXiv* **2021**, arXiv:2110.11112. [CrossRef]
35. Gürer, C.; Draxler, C. Penalization approaches in the conditional maximum likelihood and Rasch modelling context. *Br. J. Math. Stat. Psychol.* **2022**. [CrossRef] [PubMed]
36. Liang, X.; Jacobucci, R. Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net. *Struct. Equ. Model.* **2020**, *27*, 722–734. [CrossRef]
37. Tutz, G.; Schauberger, G. A penalty approach to differential item functioning in Rasch models. *Psychometrika* **2015**, *80*, 21–43. [CrossRef]
38. Schauberger, G.; Mair, P. A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behav. Res. Methods* **2020**, *52*, 279–294. [CrossRef]
39. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press: Boca Raton, FL, USA, 2015. [CrossRef]
40. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
41. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 91–108. [CrossRef]
42. Tutz, G.; Gertheiss, J. Regularized regression for categorical data. *Stat. Model.* **2016**, *16*, 161–200. [CrossRef]
43. Sen, S.; Cohen, A.S.; Kim, S.H. Model selection for multilevel mixture Rasch models. *Appl. Psychol. Meas.* **2019**, *43*, 272–289. [CrossRef] [PubMed]
44. Chen, Y.; Liu, J.; Xu, G.; Ying, Z. Statistical analysis of Q-matrix based diagnostic classification models. *J. Am. Stat. Assoc.* **2015**, *110*, 850–866. [CrossRef]
45. Battauz, M. Regularized estimation of the nominal response model. *Multivar. Behav. Res.* **2020**, *55*, 811–824. [CrossRef] [PubMed]
46. Oelker, M.R.; Tutz, G. A uniform framework for the combination of penalties in generalized structured models. *Adv. Data Anal. Classif.* **2017**, *11*, 97–120. [CrossRef]
47. Asparouhov, T.; Muthén, B. *Random Starting Values and Multistage Optimization*. Technical Report. 2019. Available online: <https://bit.ly/3SCLTjt> (accessed on 30 September 2022).
48. Robitzsch, A. *sirt: Supplementary Item Response Theory Models*. R Package Version 3.12-66. 2022. Available online: <https://CRAN.R-project.org/package=sirt> (accessed on 17 May 2022).
49. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994. [CrossRef]
50. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2022. Available online: <https://www.R-project.org/> (accessed on 11 January 2022).
51. Liu, X.; Wallin, G.; Chen, Y.; Moustaki, I. Rotation to sparse loadings using L^p losses and related inference problems. *arXiv* **2022**, arXiv:2206.02263. [CrossRef]
52. Alexandrowicz, R.; Matschinger, H. Estimation of item location effects by means of the generalized logistic regression model: A simulation study and an application. *Psychol. Sci.* **2008**, *50*, 64–74.
53. Jin, K.Y.; Wang, W.C. Item response theory models for performance decline during testing. *J. Educ. Meas.* **2014**, *51*, 178–200. [CrossRef]
54. List, M.K.; Robitzsch, A.; Lüdtke, O.; Köller, O.; Nagy, G. Performance decline in low-stakes educational assessments: Different mixture modeling approaches. *Large-Scale Assess. Educ.* **2017**, *5*, 15. [CrossRef]
55. Debeer, D.; Janssen, R. Modeling item-position effects within an IRT framework. *J. Educ. Meas.* **2013**, *50*, 164–185. [CrossRef]
56. Hartig, J.; Buchholz, J. A multilevel item response model for item position effects and individual persistence. *Psych. Test Assess. Model.* **2012**, *54*, 418–431.
57. Bolt, D.M.; Cohen, A.S.; Wollack, J.A. Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *J. Educ. Meas.* **2002**, *39*, 331–348. [CrossRef]
58. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics: Theory and Methods*; Wiley: New York, NY, USA, 2006. [CrossRef]
59. Bolt, D.M.; Kim, J.S.; Blanton, M.; Knuth, E. Applications of item response theory in mathematics education research. *J. Res. Math. Educ.* **2016**, *15*, 31–52.
60. Robitzsch, A. Four-parameter guessing model and related item response models. *Preprints* **2022**, 2022100430. [CrossRef]