stats

MDPI

Article

A Comparison of Limited Information Estimation Methods for the Two-Parameter Normal-Ogive Model with Locally Dependent Items

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 two-parameter normal-ogive (2PNO) model is one of the most popular item response theory (IRT) models for analyzing dichotomous items. Consistent parameter estimation of the 2PNO model using marginal maximum likelihood estimation relies on the local independence assumption. However, the assumption of local independence might be violated in practice. Likelihood-based estimation of the local dependence structure is often computationally demanding. Moreover, many IRT models that model local dependence do not have a marginal interpretation of item parameters. In this article, limited information estimation methods are reviewed that allow the convenient and straightforward handling of local dependence in estimating the 2PNO model. In detail, pairwise likelihood, weighted least squares, and normal-ogive harmonic analysis robust method (NOHARM) estimation are compared with marginal maximum likelihood estimation that ignores local dependence. A simulation study revealed that item parameters can be consistently estimated with limited information methods. At the same time, marginal maximum likelihood estimation resulted in biased item parameter estimates in the presence of local dependence. From a practical perspective, there were only minor differences regarding the statistical quality of item parameter estimates of the different estimation methods. Differences between the estimation methods are also compared for two empirical datasets.

Keywords: item response model; normal-ogive model; local dependence; pairwise likelihood estimation; weighted least squares estimation; NOHARM estimation

check for updates

Citation: Robitzsch, A. A Comparison of Limited Information Estimation Methods for the Two-Parameter Normal-Ogive Model with Locally Dependent Items. *Stats* **2024**, *7*, 576–591. https://doi.org/10.3390/stats7030035

Academic Editor: Dungang Liu

Received: 27 April 2024
Revised: 18 June 2024
Accepted: 19 June 2024
Published: 21 June 2024

Copyright: © 2024 by the author. 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/).

1. Introduction

Item response theory (IRT) models [1–3] are statistical models for analyzing a vector of binary (i.e., dichotomous) random variables. IRT models represent a high-dimensional contingency table by a low-dimensional latent factor variable (also referred to as a trait or ability variable).

Let $\boldsymbol{X} = (X_1, \ldots, X_I)$ be the vector of I dichotomous items $X_i \in \{0, 1\}$. A unidimensional IRT model [3,4] is a statistical model for the probability distribution $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x})$ for the vector of item responses $\boldsymbol{x} \in \{0, 1\}^I$, where

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{x}; \gamma) = \int_{-\infty}^{\infty} \prod_{i=1}^{I} P_i(x_i, \theta; \gamma_i) f(\theta) \, \mathrm{d}\theta \tag{1}$$

and γ_i are item parameters of item $i = 1, \ldots, I$. The latent variable θ with density f in (1) summarizes the vector $\boldsymbol{X}$ of item responses. Frequently, the density f is fixed and does not depend on the model parameters to be estimated. For convenience, f is often chosen as the standard normal density ϕ. The IRT model is most frequently estimated with marginal maximum likelihood (MML) estimation, in which the latent variable θ is integrated out [4,5].

The item response functions (IRFs)

$$P_i(x, \theta; \gamma_i) = \mathsf{P}(X_i = x|\theta) \text{ (for } x = 0, 1) \tag{2}$$

in (1) describe the relationship of the dichotomous item X_i with θ. Let $\gamma = (\gamma_1, \ldots, \gamma_I)$ be the model parameter of interest. The two-parameter logistic (2PL) model [6] (also referred to as the two-parameter logit model) uses the IRF

$$P_i(1, \theta; a_i, b_i) = \Psi(a_i\theta - b_i) = \frac{\exp(a_i\theta - b_i)}{1 + \exp(a_i\theta - b_i)} , \tag{3}$$

where Ψ denotes the logistic distribution function and $\gamma_i = (a_i, b_i)$. The item parameter a_i is often called item discrimination, while b_i is denoted as the item intercept. As an alternative IRT model, the IRF of the two-parameter normal-ogive (2PNO) model [7] (also referred to as the two-parameter probit model) is defined as

$$P_i(1, \theta; a_i, b_i) = \Phi(a_i\theta - b_i) = \int_{-\infty}^{a_i\theta - b_i} \phi(u)\, \mathrm{d}u , \tag{4}$$

where Φ denotes the standard normal distribution function and $\gamma_i = (a_i, b_i)$. There is a close correspondence between the 2PL and the 2PNO model (see [8–11]) because it holds that

$$|\Phi(x) - \Phi(Dx)| < 0.01 \text{ with } D = 1.701 . \tag{5}$$

Hence, item parameters from the 2PL model can be converted to item parameters of the 2PNO model and the other way around if one defines

$$\Phi(a_i\theta - b_i) = \Psi(Da_i\theta - Db_i) = \Psi(\tilde{a}_i\theta - \tilde{b}_i) \text{ with } \tilde{a}_i = Da_i \text{ and } \tilde{b}_i = Db_i . \tag{6}$$

The dimensional reduction of $\boldsymbol{X}$ to θ in the IRT model (1) is accomplished by the property of local independence [3], which means that item responses X_i are independent conditionally on the latent variable θ. In the IRT model, it is assumed that

$$\mathsf{P}(\boldsymbol{X}|\theta) = \prod_{i=1}^{I} \mathsf{P}(X_i|\theta) . \tag{7}$$

In empirical datasets containing item responses, the local independence assumption could be violated. In this case, there could exist items X_i and X_j such that

$$\mathsf{P}(X_i, X_j|\theta) \neq \mathsf{P}(X_i|\theta)\mathsf{P}(X_j|\theta) . \tag{8}$$

For example, in a reading comprehension test, students are asked to respond to items that refer to several reading texts. However, items that refer to the same reading text are likely more dependent on each other than items from different reading texts. Hence, the local dependence (i.e., deviation from local independence) of item responses of items within the same reading text can be expected. The item groups that are prone to local dependence are also referred to as testlets [12]. However, local independence holds for item responses to different testlets (i.e., reading texts).

It has been shown that item discriminations are distorted and positively biased in the case of local dependence [13]. Moreover, the neglected local dependence entails inflated reliability estimates [14,15]. Hence, parameter estimation methods must be adapted to allow consistent parameter estimates for γ_i of the IRFs.

The literature mainly treats local dependence by means of three alternative strategies. First, in testlet IRT models, additional latent variables are included to model the testlet structure [12,16]. Second, the dependency structure is covered by additional parameters within a testlet [17–21]. In this approach, a polytomous superitem is defined as one that decodes the combination of values of items within a testlet in their categories [22]. For example,

if a testlet consists of three items, the superitem possesses $2^3 = 8$ categories. The first and the second approaches have the disadvantage that item parameters do not have a marginal interpretation [14,23]. The third approach employs a marginal unidimensional IRT model and models the local dependence structure employing copula models [14,24–27]. This approach allows a marginal interpretation of item parameters. However, copula models require the specification of the type of copula distribution of the testlets and estimate testlet residual dependence parameters δ along with item parameters γ.

In this article, we are interested in estimating item parameters a_i and b_i of the 2PNO model that possess a marginal interpretation but avoid simultaneously estimating local dependence parameters δ. The 2PNO model can be equivalently formulated using underlying continuous latent variables X_i^* for dichotomous items X_i. Let the dichotomous variable X_i be defined due to the dichotomization of X_i^*

$$X_i = \mathbf{1}(X_i^* > 0) \,, \tag{9}$$

where $\mathbf{1}$ denotes the indicator function. The Gaussian copula 2PNO model with local dependence [14,25,28] relies on a vector of correlated item residuals $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_I)$. Item residuals ε and the factor variable θ are uncorrelated. Moreover, each of the components ε_i is standard normally distributed. The correlation matrix $\boldsymbol{\Sigma}^* = \mathrm{Var}(\varepsilon)$ models deviations from local independence. Define

$$X_i^* = a_i\theta - b_i - \varepsilon_i \,. \tag{10}$$

Due to the definition (9), we obtain

$$\mathsf{P}(X_i = 1|\theta) = \mathsf{P}(X_i^* > 0|\theta) = \mathsf{P}(\varepsilon_i < a_i\theta - b_i|\theta) = \Phi(a_i\theta - b_i) \,, \tag{11}$$

which is the IRF of the 2PNO model. Hence, local dependence can be induced by assuming a non-diagonal correlation matrix $\boldsymbol{\Sigma}^*$. Nevertheless, item parameters a_i and b_i retain their marginal interpretation irrespective of an assumed local dependence structure.

Purpose

This article compares different limited information estimation methods of the 2PNO model with correlated item residuals (i.e., the 2PNO normal copula model). The limited information estimation methods can handle local dependence but do not require the specification and estimation of parameters that refer to deviations from local independence. Previous research considered item response datasets under local independence. It indicated that the different limited information methods provided similar estimates and resulted in only minor efficiency loss compared to full information methods. This research investigates whether the findings also translate to item response data with local dependence. To our knowledge, no simulation studies have previously addressed this topic.

The rest of the article is structured as follows. Section 2 discusses the application of limited information methods to handle local dependence. Section 3 presents findings from a simulation study that compares different limited information methods with MML estimation that ignores local dependence. Differences in estimation methods are illustrated through two empirical datasets in Section 4. Section 5 discusses the main findings of our study. Finally, the paper closes with conclusions in Section 6.

2. Limited Information Methods for Local Dependence

In this section, we describe three limited information methods capable of handling local dependence in item responses.

2.1. Pairwise Maximum Likelihood Estimation (PML)

Pairwise (maximum) likelihood (PML; [29–33]) estimation models first-order or second-order probabilities of items. Hence, in contrast to MML, the dependence among items is

only modeled up to probabilities of item responses if they are item pairs. The marginal univariate probabilities are given by

$$L_{1,i}(x; a_i, b_i) = \mathsf{P}(X_i = x) = \int P_i(x, \theta; a_i, b_i)\phi(\theta)\,\mathrm{d}\theta\,, \tag{12}$$

where ϕ denotes the density of the standard normal distribution. PML estimation also relies on the evaluation of bivariate probabilities for an item pair (X_i, X_j)

$$L_{2,ij}(x, y; a_i, b_i, a_j, b_j) = \mathsf{P}(X_i = x, X_j = y) = \int P_i(x, \theta; a_i, b_i)P_j(y, \theta; a_j, b_j)\phi(\theta)\,\mathrm{d}\theta\,. \tag{13}$$

In PML estimation, item parameters $\gamma = (a_1, b_1, \ldots, a_I, b_I)$ are computed by maximizing a weighted sum of likelihood contributions of univariate and bivariate probabilities. In more detail, the optimization function utilized in PML is given by

$$l(\gamma) = \sum_{i=1}^{I} w_{1,i}\left(\sum_{x=0}^{1} n_{i,x}\log L_{1,i}(x; a_i, b_i)\right) + \sum_{i<j} w_{2,ij}\left(\sum_{x=0}^{1}\sum_{y=0}^{1} n_{ij,xy}\log L_{2,ij}(x, y; a_i, b_i, a_j, b_j)\right)\,, \tag{14}$$

where $n_{i,x}$ denotes the univariate frequency that item X_i takes the value $x \in \{0, 1\}$. Furthermore, $n_{ij,xy}$ denotes the bivariate frequency in the sample that item X_i has the value $x \in \{0, 1\}$ and item X_j has the value $y \in \{0, 1\}$. The weights $w_{1,i}$ and $w_{2,ij}$ can be specifically chosen [34]. For example, the choices $w_{1,i} = 1/I$ and $w_{2,ij} = 2I^{-1}(I-1)^{-1}$ ensure that the univariate and bivariate frequencies equally contribute to the optimization function l defined in (14).

PML is particularly useful in avoiding computationally demanding multidimensional integration in multidimensional IRT models or models involving longitudinal data [35–38]. In these situations, high-dimensional integrals of latent variables in MML are reduced to lower-dimensional integrals in PML estimation.

PML can also be used to estimate IRT models with locally dependent items. In this case, item pairs (X_i, X_j) that are locally dependent have an inflated associated and will bias estimated item discriminations a_i and item intercepts b_i in the 2PNO model. However, these item pairs can be essentially removed from the PML estimation by setting the corresponding weights $w_{2,ij}$ to zero. This technique has proven successful in estimating the 2PL model [39].

2.2. Weighted Least Squares Estimation (DWLS and ULS)

As an alternative to MML and PML estimation, weighted least squares (WLS) estimation for factor analysis of dichotomous and polytomous item responses have been proposed [40–44]. WLS is a two-step (or three-step) procedure that estimates thresholds and tetrachoric correlations in the first step and computes item parameters in the second step.

Assume that dichotomous items X_i have an underlying normally distributed latent variable X_i^* defined in the 2PNO model by

$$X_i^* = a_i\theta - b_i + \varepsilon_i\,, \text{and } X_i^* \sim \mathrm{N}(-b_i, a_i^2 + 1)\,, \tag{15}$$

where ε_i is the item residual, and $\mathrm{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. The proportion correct of item X_i can be computed as

$$p_i = \mathsf{P}(X_i = 1) = \mathsf{P}(X_i^* > 0) = \Phi\left(-\frac{b_i}{\sqrt{a_i^2 + 1}}\right) = \Phi(\tau_i)\,, \tag{16}$$

where τ_i is the item threshold. The threshold can be estimated by $\widehat{\tau}_i = \Phi^{-1}(\widehat{p}_i)$, where $\widehat{p}_i$ is the proportion correct of item X_i in the sample.

The correlation between the underlying latent variables X_i^* and X_j^* (i.e., the tetrachoric correlation) ρ_{ij} can be determined from the 2PNO model as

$$\rho_{ij} = \text{Cor}(X_i^*, X_j^*) = \frac{\text{Cov}(X_i^*, X_j^*)}{\sqrt{\text{Var}(X_i^*)\text{Var}(X_j^*)}} = \frac{a_i a_j}{\sqrt{a_i^2+1}\sqrt{a_j^2+1}} \,. \tag{17}$$

The tetrachoric correlation ρ_{ij} can be estimated with maximum likelihood [43] with fixed item thresholds $\widehat{\tau}_i$ and $\widehat{\tau}_j$. The involved probabilities of item pair (X_i, X_j) in the estimation are given by

$$\mathsf{P}(X_i = 1, X_j = 1) = \mathsf{P}(X_i^{**} > \tau_i, X_j^{**} > \tau_j) = \mathsf{P}(-X_i^{**} < -\tau_i, -X_j^{**} < -\tau_j) = \Phi_2(-\tau_i, -\tau_j, \rho_{ij}) \,, \tag{18}$$

where Φ_2 denotes the distribution function of the bivariate standard normal distribution with correlation ρ_{ij}. Note that X_i^{**} and X_j^{**} are underlying normally distributed standardized latent variables. More generally, it holds that

$$\mathsf{P}(X_i = x, X_j = y) = \Phi_2\big(-(2x-1)\tau_i, -(2y-1)\tau_j, (-1)^{x+y}\rho_{ij}\big) \text{ for } x, y \in \{0, 1\} \,. \tag{19}$$

The optimization function G in diagonally weighted least squares (DWLS; [30]) relies on estimated tetrachoric correlations $\widehat{\rho}_{ij}$ and is defined as

$$G(\boldsymbol{a}) = \sum_{i<j} w_{ij} \left(\widehat{\rho}_{ij} - \frac{a_i a_j}{\sqrt{(1+a_i^2)(1+a_j^2)}} \right)^2 \,. \tag{20}$$

The weights w_{ij} are chosen as the inverse of the variances of estimated tetrachoric correlations $\widehat{\rho}_{ij}$ and must be estimated in the first step (see [41]). If the test contains locally dependent items, the corresponding item pair (X_i, X_j) receives a weight of zero (i.e., $w_{ij} = 0$) in the DWLS estimation.

Unweighted least squares (ULS; [45]) estimation is obtained by choosing all weights equal to one. Again, weights w_{ij} of item pairs (X_i, X_j) that are locally dependent can be set to zero.

The minimizer of G defined in (20) provides an estimate $\widehat{\boldsymbol{a}} = (\widehat{a}_1, \ldots, \widehat{a}_I)$ of item discriminations. Item intercepts b_i are subsequently estimated by

$$\widehat{b}_i = -\sqrt{1+\widehat{a}_i^2} \cdot \widehat{\tau}_i \,. \tag{21}$$

The IRT approach and the underlying latent variable approach to dichotomous items have been shown to be (essentially) equivalent [46–48]. DWLS and ULS estimation showed only a minor efficiency loss compared to MML estimation for dichotomous and polytomous item responses [49,50]. The advantage of WLS estimation is that multidimensionality in latent variables does not pose an issue in the estimation because no multidimensional integrals appear in the estimation function. However, WLS estimation requires the estimation of tetrachoric correlations, which can also be computationally demanding for a large number of items.

2.3. NOHARM Estimation

As an alternative to MML, PML, and WLS estimation, McDonald [51–53] proposed an alternative estimation method that relies on approximating the normal-ogive IRF in the 2PNO model by a series of orthogonal Hermite polynomials of order three. This estimation method is referred to as the normal-ogive harmonic analysis robust method (NOHARM) method (see [54–56] for overviews). As for WLS estimation, the estimation of item intercepts b_i is separated from the estimation of item discriminations a_i.

Let the item threshold τ_i be defined as

$$p_i = \mathsf{P}(X_i = 1) = \Phi(\tau_i) \,. \tag{22}$$

Like in WLS estimation, the threshold τ_i can be estimated on the sample proportion correct:

$$\widehat{\tau}_i = \Phi^{-1}(\widehat{p}_i) \, . \tag{23}$$

The NOHARM estimation method starts with determining the coefficients of the Hermite series for item X_i:

$$c_{i0} = \Phi(\widehat{\tau}_i) \, , \tag{24}$$

$$c_{i1} = -\phi(\widehat{\tau}_i) \, , \tag{25}$$

$$c_{i2} = \frac{1}{\sqrt{2}} \widehat{\tau}_i \phi(\widehat{\tau}_i) \, , \text{ and} \tag{26}$$

$$c_{i3} = -\frac{1}{\sqrt{6}} (\widehat{\tau}_i^2 - 1) \phi(\widehat{\tau}_i) \, . \tag{27}$$

Due to the orthogonality of the Hermite polynomials, the pairwise probabilities passing items X_i and X_j can be approximated by

$$\mathsf{P}(X_i = 1, X_j = 1) = p_{ij}(a_i, a_j) = \sum_{k=0}^{3} c_{ik} c_{jk} \gamma_{ij}^k \text{ with } \gamma_{ij} = \frac{a_i a_j}{\sqrt{(1 + a_i^2)(1 + a_j^2)}} \, . \tag{28}$$

Using (28), estimates of the item discriminations can be obtained by minimizing

$$G(\boldsymbol{a}) = \sum_{i<j} w_{ij} \left(\widehat{p}_{ij} - p_{ij}(a_i, a_j) \right)^2 \, . \tag{29}$$

The item intercepts b_i can be obtained in the same way as in (21). The originally proposed NOHARM procedure fixes all weights w_{ij} in (28) equally to one. In the case of the local dependence of item pairs (X_i, X_j), the corresponding weights w_{ij} can be set to zero. In this article, we also investigate a weighted NOHARM (WNOHARM) estimation method in which we define the weights as the variance of the estimated proportions $\widehat{p}_{ij}$ (i.e., $w_{ij} = N \widehat{p}_{ij}^{-1} (1 - \widehat{p}_{ij})^{-1}$). To our knowledge, the WNOHARM method has not been investigated previously in the literature. In the literature, it has been demonstrated that NOHARM produces similar results to MML, DWLS, and ULS estimation [57–61].

3. Simulation Study

3.1. Methodology

In this simulation study, item responses under local dependence are simulated according to the 2PNO model with a normal copula model for item residuals [25]. The factor variable θ is assumed to be standard normally distributed. Two test lengths for the 2PNO model with local dependence are simulated: $I = 10$ and $I = 20$ items. The item discriminations a_i in the case of $I = 10$ items are 1.0, 0.9, 0.8, 0.6, 0.6, 1.1, 0.9, 1.0, 0.8, and 0.7. The item intercepts b_i are chosen as -0.9, -0.7, 1.3, -1.9, -1.0, -1.3, -0.2, -1.3, -0.1, and 0.2. The 10 items are arranged into three testlets containing three items each and a single item. The item residuals ε_i are standard normally distributed with a correlation matrix

$$\boldsymbol{\Sigma}^* = \left(\begin{array}{ccc|ccc|ccc|c} 1 & & & & & & & & & \\ \delta & 1 & & & & & & & & \\ \delta & \delta & 1 & & & & & & & \\ \hline 0 & 0 & 0 & 1 & & & & & & \\ 0 & 0 & 0 & \delta & 1 & & & & & \\ 0 & 0 & 0 & \delta & \delta & 1 & & & & \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \delta & 1 & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \delta & \delta & 1 & \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right). \tag{30}$$

The local dependence parameter δ is chosen as 0, 0.4, and 0.8, indicating local independence, moderate local dependence, and strong local dependence, respectively. In the case of $I = 20$ items, the item parameters from the case of 10 items are duplicated.

The sample sizes $N = 500$, 1000, and 2000 are employed in this simulation. In smaller sample sizes, estimation issues of the 2PNO model can be expected.

Each simulated dataset is analyzed with the 2PNO model assuming a standard normal distribution for the factor variable θ. Six different estimation methods are investigated. First, MML estimation ignores the local dependence. Next to MML, five different limited information estimation methods (PML, DWLS, ULS, WNOHARM, NOHARM) are specified that take the local dependence structure into account. In all the approaches, the weights w_{ij} referring to item pairs (X_i, X_j) are set to zero if there is local dependence between X_i and X_j.

In total, 1500 replications in each of the 2 (number of items I) $\times$ 3 (extent of local dependence δ) $\times$ 3 (sample size N) $=$ 18 cells of the simulation study are conducted. We compute the empirical bias and the empirical root mean square error (RMSE) for all estimated item discriminations a_i and item intercepts b_i. For each item parameter, the RMSE of an estimation method is divided by the RMSE of the PML estimation method and multiplied by 100. By doing so, the estimation methods are compared relative to the performance of PML estimation. To summarize the simulation results, we compute the average absolute bias and the average relative RMSE across all item discriminations and item intercepts, respectively.

All estimation methods and the analysis of the findings of this simulation study have been carried out using the statistical software R [62]. MML and PML estimation is specified using the R function `sirt::xxirt()` in the R package sirt [63]. DWLS and ULS estimation is conducted with the `probit_wls()` function that can be found at https://osf.io/wxbfn accessed on 27 April 2024. The required tetrachoric correlations and their variances are estimated with the R package lavaan [64]. The WNOHARM and NOHARM estimation methods are specified in the R function `sirt::noharm.sirt()` in the R package sirt [63]. The functions `sirt::xxirt()`, `probit_wls()`, and `sirt::noharm.sirt()` utilize the R-internal optimizers `stats::optim()` or `stats::nlminb()`. Replication material for the simulation study can be found at https://osf.io/wxbfn accessed on 27 April 2024.

3.2. Results

Table 1 reports the average absolute bias and the average relative RMSE for estimated item discriminations and item intercepts in the 2PNO model as a function of the number of items I, the extent of local dependence δ, and the sample size N.

Table 1. Simulation Study: Average absolute bias and average relative root mean square error (RMSE) for estimated item discriminations a_i and item intercepts b_i in the two-parameter normal-ogive (2PNO) model as a function of the number of items I, the extent of local dependence δ, and sample size N.

			Average Absolute Bias						**Average Relative RMSE**					
I	***δ***	***N***	**MML**	**PML**	**DWLS**	**ULS**	**WNH**	**NH**	**MML**	**PML**	**DWLS**	**ULS**	**WNH**	**NH**
			Item discriminations a_i											
10	0	500	0.014	**0.017**	**0.022**	**0.018**	**0.021**	**0.021**	91.8	100 ‡	100.5	102.7	104.6	104.3
		1000	0.007	0.008	0.011	0.010	0.011	0.011	92.7	100 ‡	100.7	104.4	104.1	104.0
		2000	0.003	0.003	0.005	0.004	0.007	0.007	92.6	100 ‡	100.4	104.9	103.6	103.6
	0.4	500	**0.118**	**0.020**	**0.023**	**0.020**	**0.022**	**0.022**	**124.0**	100 ‡	100.4	104.2	103.7	103.2
		1000	**0.113**	0.010	0.012	0.011	0.013	0.013	**149.3**	100 ‡	100.5	104.7	103.7	103.7
		2000	**0.107**	0.006	0.006	0.006	0.009	0.009	**183.5**	100 ‡	100.5	**105.5**	103.4	103.6
	0.8	500	**0.334**	**0.025**	**0.025**	**0.024**	**0.026**	**0.026**	**274.7**	100 ‡	99.7	103.6	102.1	102.4
		1000	**0.320**	0.012	0.012	0.012	0.013	0.014	**350.0**	100 ‡	100.2	**105.4**	102.1	102.6
		2000	**0.316**	0.008	0.007	0.007	0.009	0.009	**464.0**	100 ‡	100.2	105.0	102.5	102.9
20	0	500	0.007	0.008	**0.015**	0.008	0.012	0.012	95.7	100 ‡	101.2	101.2	103.4	103.2
		1000	0.011	0.005	0.009	0.006	0.009	0.009	97.3	100 ‡	100.6	101.9	103.1	102.9
		2000	0.001	0.003	0.004	0.003	0.006	0.006	95.7	100 ‡	100.5	102.1	103.5	103.4
	0.4	500	**0.053**	0.013	**0.019**	0.013	**0.016**	**0.016**	**105.7**	100 ‡	101.4	101.5	103.3	103.1
		1000	**0.057**	0.005	0.009	0.006	0.009	0.009	**121.6**	100 ‡	100.5	101.9	102.7	102.5
		2000	**0.046**	0.004	0.005	0.003	0.007	0.007	**128.4**	100 ‡	100.3	102.2	103.3	103.3
	0.8	500	**0.111**	0.014	**0.018**	0.012	**0.016**	**0.016**	**139.9**	100 ‡	101.1	101.5	102.8	102.9
		1000	**0.119**	0.007	0.010	0.008	0.011	0.011	**178.5**	100	100.4	102.2	102.5	102.7
		2000	**0.104**	0.004	0.004	0.003	0.007	0.007	**207.7**	100 ‡	100.2	102.4	102.8	103.1
			Item intercepts b_i											
10	0	500	0.013	0.014	**0.016**	**0.015**	**0.017**	**0.016**	96.4	100 ‡	100.2	100.6	102.3	102.0
		1000	0.006	0.007	0.008	0.008	0.009	0.008	97.0	100 ‡	100.6	102.1	102.2	102.1
		2000	0.003	0.003	0.004	0.004	0.005	0.005	97.0	100 ‡	100.3	102.4	101.9	101.9
	0.4	500	**0.056**	**0.017**	**0.017**	**0.017**	**0.018**	**0.017**	**110.4**	100 ‡	100.1	101.5	102.2	101.9
		1000	**0.046**	0.008	0.009	0.010	0.010	0.010	**116.3**	100 ‡	100.6	102.5	102.2	102.2
		2000	**0.044**	0.005	0.005	0.005	0.006	0.006	**129.2**	100 ‡	100.5	102.8	102.1	102.1
	0.8	500	**0.114**	**0.022**	**0.020**	**0.020**	**0.021**	**0.021**	**157.4**	100 ‡	99.2	100.5	101.1	101.2
		1000	**0.090**	0.010	0.010	0.011	0.011	0.011	**165.7**	100 ‡	100.3	102.7	101.6	101.8
		2000	**0.101**	0.006	0.005	0.005	0.006	0.006	**211.6**	100 ‡	99.8	102.1	101.3	101.4
20	0	500	**0.015**	0.010	0.013	0.010	0.012	0.012	100.7	100 ‡	100.7	99.7	101.4	101.2
		1000	0.013	0.005	0.007	0.006	0.007	0.007	100.2	100 ‡	100.6	100.8	101.6	101.4
		2000	0.005	0.003	0.003	0.003	0.004	0.004	101.9	100 ‡	100.1	100.7	101.3	101.2
	0.4	500	**0.045**	0.013	**0.015**	0.012	0.014	0.014	**109.6**	100 ‡	100.5	99.6	101.4	101.2
		1000	**0.020**	0.006	0.007	0.006	0.007	0.007	103.8	100 ‡	100.6	100.8	101.6	101.4
		2000	**0.030**	0.004	0.004	0.003	0.005	0.005	**118.2**	100 ‡	99.9	100.6	101.3	101.1
	0.8	500	**0.078**	0.013	0.014	0.011	0.013	0.013	**127.6**	100 ‡	100.4	99.7	101.1	101.0
		1000	**0.034**	0.007	0.008	0.007	0.008	0.008	**112.2**	100 ‡	100.5	101.0	101.5	101.4
		2000	**0.059**	0.004	0.004	0.003	0.005	0.005	**156.5**	100 ‡	99.5	100.4	100.7	100.7

Note. MML = marginal maximum likelihood estimation ignoring local independence; PML = pairwise likelihood estimation; DWLS = diagonally least squares estimation; ULS = unweighted least squares estimation; WNH = weighted normal ogive harmonic analysis robust method (WNOHARM) estimation; NH = normal ogive harmonic analysis robust method (NOHARM) estimation; ‡ = PML was defined as the reference method for computing the average RMSE. Average absolute bias values larger than 0.015 and average RMSE values larger than 105.0 are printed in bold font.

All six estimation methods are unbiased in large samples if local independence holds (i.e., $\delta = 0$). Notably, small biases occur for the sample size $N = 500$. The average relative RMSE indicates an efficiency loss when using PML estimation and omitting some item pairs

from estimation compared to MML under local independence. However, the efficiency loss is decreased with a larger number of items.

In the case of local dependence (i.e., $\delta = 0.4$ and $\delta = 0.8$), MML produces strongly biased results. The biases for MML are larger for item discriminations than for item intercepts. Note that the bias in MML estimation is reduced with an increasing number of items. Across all conditions, PML and DWLS perform similarly regarding the mean (denoted by M) RMSE and slightly outperform ULS, WNOHARM, and NOHARM estimation for item discriminations a_i (PML: M = 100, DWLS: M = 100.5, ULS: M = 103.2, WNOHARM: M = 103.2, NOHARM: M = 103.2). Overall, WNOHARM resulted in similar estimates to NOHARM. However, there are slight advantages of WNOHARM compared to NOHARM only for strong local dependence in the case of $\delta = 0.8$. For item intercepts b_i, PML and DWLS have smaller advantages over ULS, WNOHARM, and NOHARM (PML: M = 100, DWLS: M = 100.3, ULS: M = 101.1, WNOHARM: M = 101.6, NOHARM: M = 101.5).

To sum up, all limited information methods could handle locally dependent item responses and result in approximately unbiased item parameter estimates. In applied research, the differences between the estimation methods can be considered negligible.

4. Empirical Examples

In this section, two datasets with item responses with a testlet structure from the R package sirt [63] are analyzed with the 2PNO model. The same estimation methods (i.e., MML, PML, DWLS, ULS, WNOHARM, and NOHARM) as in the simulation study (see Section 4) are utilized. Item pairs that refer to the same testlet receive a zero weight in the limited information estimation methods. We compute the average absolute deviation of estimated item parameters between the estimation methods as a measure of the discrepancy between the methods.

4.1. Dataset `data.read`

The first dataset `data.read` contains item responses of $N = 328$ subjects on $I = 12$ items from a reading comprehension test administered to Austrian students. The twelve items are arranged into three testlets, each testlet containing four items. Table 2 displays estimated item discriminations a_i and item intercepts b_i in the 2PNO model.

MML substantially differs in the item discriminations a_i from all limited information methods regarding the average absolute deviations (PML: 0.20, DWLS: 0.23, ULS: 0.25, WNOHARM: 0.22, NOHARM: 0.23). The average absolute deviations between the limited information methods are much smaller and range between 0.01 (WNOHARM and NOHARM) and 0.07 (ULS and NOHARM).

We also observe non-negligible differences for item intercepts b_i of MML with the five specified limited information methods (PML: 0.13, DWLS: 0.14, ULS: 0.15, WNOHARM: 0.14, NOHARM: 0.14). The differences in item intercepts between the limited information methods are small, with the largest average absolute deviation of 0.03 between ULS and NOHARM estimation.

Figure 1 presents pairwise scatter plots and correlations of estimated item discriminations a_i for the dataset `data.read`. The correlation of the MML estimates with the other estimation methods are almost zero. Obviously, one outlying observation (i.e., Item C1 with $a_i = 1.50$) impacts the correlations of the MML estimates with other methods. Moreover, WNOHARM and NOHARM are correlated at 1.00, while the correlation of ULS and NOHARM is relatively small at 0.86, which indicates non-negligible differences in the estimates.

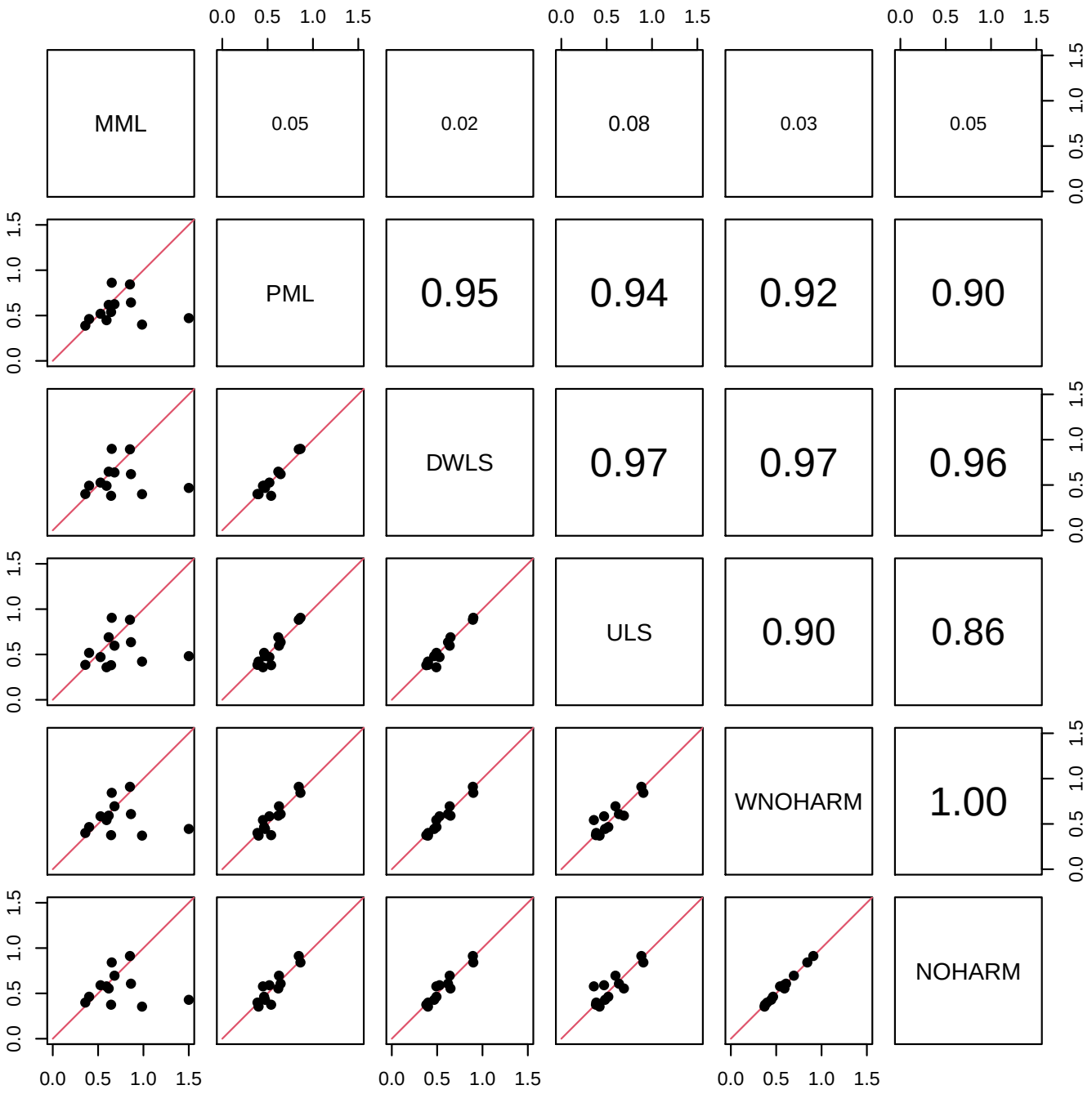


Figure 1. Dataset `data.read`: Pairwise scatter plots and correlations of estimated item discriminations a_i for between different estimation methods.

Figure 2 shows pairwise scatter plots and correlations of estimated item intercepts b_i. The correlations of the MML estimates for b_i with the other estimation methods are far from perfect at 0.95. However, the estimates of the alternative estimation methods correlate at 1.00, indicating almost identical estimates.

Table 2. Dataset `data.read`: Estimated item discriminations a_i and item intercepts b_i in the two-parameter normal-ogive (2PNO) model for different estimation methods.

		a_i						b_i					
Item	**Testlet**	**MML**	**PML**	**DWLS**	**ULS**	**WNH**	**NH**	**MML**	**PML**	**DWLS**	**ULS**	**WNH**	**NH**
A1	A	0.59	0.45	0.49	0.36	0.54	0.58	−1.20	−1.14	−1.16	−1.10	−1.18	−1.20
A2	A	0.85	0.84	0.89	0.88	0.91	0.91	−0.84	−0.83	−0.85	−0.85	−0.86	−0.86
A3	A	0.68	0.63	0.64	0.60	0.69	0.70	−0.21	−0.20	−0.20	−0.20	−0.21	−0.21
A4	A	0.53	0.52	0.53	0.47	0.58	0.59	0.11	0.11	0.11	0.11	0.12	0.12
B1	B	0.36	0.39	0.40	0.39	0.40	0.40	−0.60	−0.60	−0.61	−0.60	−0.61	−0.61
B2	B	0.40	0.46	0.49	0.52	0.47	0.46	−0.01	−0.02	−0.02	−0.02	−0.02	−0.02
B3	B	0.62	0.62	0.65	0.69	0.59	0.55	−1.56	−1.57	−1.59	−1.62	−1.55	−1.52
B4	B	0.65	0.86	0.90	0.90	0.84	0.84	−0.57	−0.63	−0.64	−0.64	−0.62	−0.62
C1	C	1.50	0.47	0.47	0.48	0.44	0.43	−2.66	−1.66	−1.65	−1.66	−1.64	−1.63
C2	C	0.86	0.64	0.62	0.64	0.61	0.61	−0.74	−0.67	−0.66	−0.67	−0.66	−0.66
C3	C	0.98	0.40	0.40	0.42	0.37	0.35	−1.59	−1.22	−1.22	−1.23	−1.21	−1.20
C4	C	0.64	0.54	0.38	0.38	0.38	0.38	−0.75	−0.75	−0.67	−0.67	−0.67	−0.67

Note. MML = marginal maximum likelihood estimation ignoring local independence; PML = pairwise likelihood estimation; DWLS = diagonally least squares estimation; ULS = unweighted least squares estimation; WNH = weighted normal-ogive harmonic analysis robust method (WNOHARM) estimation; NH = normal-ogive harmonic analysis robust method (NOHARM) estimation.

4.2. Dataset `data.pisaMath`

The second dataset `data.pisaMath` contains item responses of $N = 565$ Austrian students on $I = 11$ mathematics items as a subset of students and items from a PISA study [65,66]. The eleven items are arranged into four testlets with two items each. Moreover, the dataset has three single items without a testlet structure.

Table 3 displays estimated item discriminations and item intercepts of the 2PNO model for this dataset. Substantial deviations for item parameters between MML estimation and the limited information estimation methods are observed for items in the testlets M406 and M496. The average absolute deviation of estimated item discriminations a_i of MML estimation with the five limited information methods is always 0.07. However, there are no substantial differences in item discriminations between the limited information methods, and they differ, on average, by no more than 0.01. For item intercepts b_i, the average absolute differences between MML and all limited information methods are smaller and at 0.02. Furthermore, there are no notable differences among the limited information estimation methods, with a maximum absolute difference of 0.002.

Table 3. Dataset `data.pisaMath`: Estimated item discriminations a_i and item intercepts b_i in the two-parameter normal-ogive (2PNO) model for different estimation methods.

		a_i						b_i					
Item	**Testlet**	**MML**	**PML**	**DWLS**	**ULS**	**WNH**	**NH**	**MML**	**PML**	**DWLS**	**ULS**	**WNH**	**NH**
M192Q01	—	0.75	0.79	0.81	0.80	0.78	0.78	0.14	0.14	0.14	0.14	0.14	0.14
M406Q01	M406	1.03	0.86	0.87	0.88	0.87	0.87	0.21	0.20	0.20	0.20	0.20	0.20
M406Q02	M406	1.28	1.04	1.04	1.04	1.07	1.06	0.94	0.84	0.83	0.83	0.85	0.84
M423Q01	—	0.31	0.31	0.31	0.31	0.31	0.31	−0.67	−0.67	−0.67	−0.67	−0.67	−0.67
M496Q01	M496	0.82	0.74	0.74	0.75	0.74	0.74	−0.17	−0.17	−0.17	−0.17	−0.17	−0.17
M496Q02	M496	0.65	0.55	0.55	0.55	0.55	0.55	−0.67	−0.65	−0.65	−0.65	−0.65	−0.65
M564Q01	M564	0.51	0.52	0.52	0.52	0.52	0.52	−0.04	−0.04	−0.04	−0.04	−0.04	−0.04
M564Q02	M564	0.49	0.50	0.50	0.50	0.50	0.50	−0.07	−0.07	−0.07	−0.07	−0.07	−0.07
M571Q01	—	0.85	0.91	0.93	0.89	0.92	0.93	−0.15	−0.16	−0.16	−0.16	−0.16	−0.16
M603Q01	M603	0.70	0.74	0.75	0.73	0.73	0.74	−0.17	−0.17	−0.17	−0.17	−0.17	−0.17
M603Q02	M603	0.85	0.91	0.93	0.89	0.89	0.91	0.09	0.09	0.09	0.09	0.09	0.09

Note. MML = marginal maximum likelihood estimation ignoring local independence; PML = pairwise likelihood estimation; DWLS = diagonally least squares estimation; ULS = unweighted least squares estimation; WNH = weighted normal-ogive harmonic analysis robust method (WNOHARM) estimation; NH = normal-ogive harmonic analysis robust method (NOHARM) estimation.

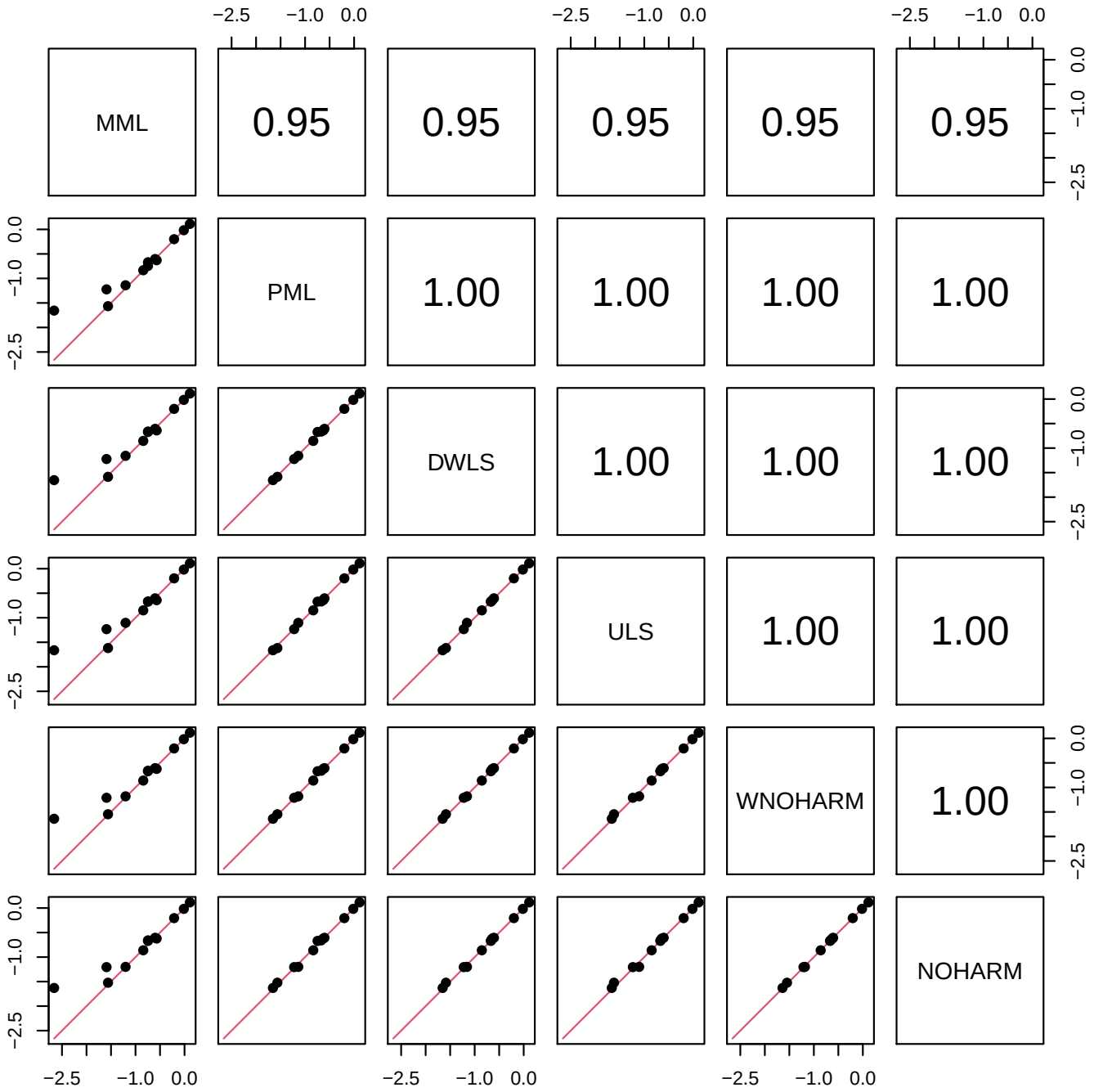


Figure 2. Dataset `data.read`: Pairwise scatter plots and correlations of estimated item intercepts b_i for between different estimation methods.

5. Discussion

5.1. Merits

In this article, we compared MML estimation with limited information estimation methods in the 2PNO model with locally dependent items. It turned out that limited information methods PML, DWLS, ULS, WNOHARM, and NOHARM were effective in handling local dependence and produced approximately unbiased item parameters in the 2PNO model. The employed limited information methods operate on statistical information (i.e., bivariate frequencies) of item pairs. This property allows the exclusion of item pairs prone to local dependence from estimation. By doing so, item parameters from the 2PNO model can be marginally interpreted. Overall, there were only minor differences in the quality of parameter estimates between the different limited information methods. From

the perspective of an applied researcher, the different estimation methods can be considered exchangeable, and the availability of software can determine the choice of a particular estimation method. Notably, MML estimation that ignores local dependence will typically provide biased item parameter estimates.

5.2. Limitations

As with any simulation study, our study had several limitations. First, we did not simulate item responses with larger item discriminations. It could be that the performance of the limited information methods more substantially differs with more extreme item parameters. Second, we did not investigate smaller sample sizes such as $N = 250$. In this case, it is more crucial how to handle estimation issues. Third, we did not treat polytomous data. However, the NOHARM procedure must be adapted from dichotomous to polytomous item responses [56]. Fourth, our analyses were restricted to unidimensional IRT models with local dependence. In a similar vein, multidimensional IRT models with locally dependent items can also be handled and investigated in future research. Fifth, the item discriminations used in the simulation study had moderate deviations. We carried out preliminary additional simulation studies in which there was more variation in item discriminations. It turned out that NOHARM (and WNOHARM) estimates performed worse than PML or DWLS (and ULS) in these situations. Sixth, we only considered a normally distributed θ variable and the probit link function in the 2PNO model. PML estimation has the advantage that it can be generalized to other link functions (such as in the 2PL model) or non-normal θ distributions [39]. However, DWLS/ULS and NOHARM rely on the probit link function and the normal distribution and cannot be generalized straightforwardly.

5.3. Meaning of Local Dependence

We would like to emphasize that any empirical dataset will necessarily have some deviations from local independence [67]. If a unidimensional IRT model is fitted to a dataset of item responses, positive and negative local dependencies will average out in some way. It is up to the researcher to define whether positive local dependence should only exist. Such a decision would imply that positive local dependence is treated as a nuisance that should be removed from model estimation. However, there might be situations in which local dependence is an in-built feature of a test, and it is not advised to control for such factors in statistical analysis [68].

6. Conclusions

Our study demonstrated that ignoring local dependence in the estimation of the 2PNO model results in biased item parameters, particularly in biases for item discriminations. We have shown in a simulation study and empirical examples that limited information methods can easily be adapted to item responses containing local dependence by only including item pairs in the estimation that are not prone to local dependence. Utilizing this modification in the estimation methods results in unbiased item parameter estimation under local dependence. Importantly, there is no need to model the local dependence structure in the modified limited information estimation methods. In this respect, limited information methods are more robust than the full information maximum likelihood estimation method, which requires a correctly specified distribution to model local dependence.

Author Contributions: Conceptualization, A.R.; methodology, A.R.; software, A.R.; validation, A.R.; formal analysis, A.R.; investigation, A.R.; resources, A.R.; data curation, A.R.; writing—original draft preparation, A.R.; writing—review and editing, A.R.; visualization, A.R.; supervision, A.R.; project administration, A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The empirical datasets used in Section 4 can be extracted from the R package sirt (https://cran.r-project.org/web/packages/sirt; accessed on 27 April 2024).

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

Abbreviations

The following abbreviations are used in this manuscript:

2PL	two-parameter logistic
2PNO	two-parameter normal-ogive
DWLS	diagonally weighted least squares
IRF	item response function
IRT	item response theory
MML	marginal maximum likelihood
NOHARM	normal-ogive harmonic analysis robust method
PML	pairwise maximum likelihood
ULS	unweighted least squares
WLS	weighted least squares
WNOHARM	weighted normal-ogive harmonic analysis robust method

References

1. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Item Response Theory—A Statistical Framework for Educational and Psychological Measurement. 2023, *Epub Ahead of Print*. Available online: https://imstat.org/journals-and-publications/statistical-science/statistical-science-future-papers/ (accessed on 27 April 2024).
2. van der Linden, W.J. Unidimensional logistic response models. In *Handbook of Item Response Theory, Volume 1: Models*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 11–30.
3. Yen, W.M.; Fitzpatrick, A.R. Item response theory. In *Educational Measurement*; Brennan, R.L., Ed.; Praeger Publishers: Westport, CT, USA, 2006; pp. 111–154.
4. Bock, R.D.; Moustaki, I. Item response theory in a general framework. In *Handbook of Statistics, Vol. 26: Psychometrics*; Rao, C.R., Sinharay, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 469–513. [CrossRef]
5. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. [CrossRef]
6. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*; Lord, F.M., Novick, M.R., Eds.; MIT Press: Reading, MA, USA, 1968; pp. 397–479.
7. Lord, F.M.; Novick, R. *Statistical Theories of Mental Test Scores*; Addison-Wesley: Reading, MA, USA, 1968.
8. Camilli, G. Origin of the scaling constant $d = 1.7$ in item response theory. *J. Educ. Behav. Stat.* **1994**, *19*, 293–295. [CrossRef]
9. Savalei, V. Logistic approximation to the normal: The KL rationale. *Psychometrika* **2006**, *71*, 763–767. [CrossRef]
10. Camilli, G. The scaling constant D in item response theory. *Open J. Stat.* **2017**, *7*, 780–785. [CrossRef]
11. Cho, E. Interchangeability between factor analysis, logistic IRT, and normal ogive IRT. *Front. Psychol.* **2023**, *14*, 1267219. [CrossRef] [PubMed]
12. Bradlow, E.T.; Wainer, H.; Wang, X. A Bayesian random effects model for testlets. *Psychometrika* **1999**, *64*, 153–168. [CrossRef]
13. Tuerlinckx, F.; De Boeck, P. Non-modeled item interactions lead to distorted discrimination parameters: A case study. *Methods Psychol. Res.* **2001**, *6*, 2. Available online: http://tinyurl.com/mrydstwz (accessed on 27 April 2024).
14. Braeken, J.; Tuerlinckx, F.; De Boeck, P. Copula functions for residual dependency. *Psychometrika* **2007**, *72*, 393–411. [CrossRef]
15. Ip, E.H. Testing for local dependency in dichotomous and polytomous item response models. *Psychometrika* **2001**, *66*, 109–132. [CrossRef]
16. Wang, W.C.; Wilson, M. The Rasch testlet model. *Appl. Psychol. Meas.* **2005**, *29*, 126–149. [CrossRef]
17. Hoskens, M.; De Boeck, P. A parametric model for local dependence among test items. *Psychol. Methods* **1997**, *2*, 261–277. [CrossRef]
18. Marais, I.; Andrich, D. Effects of varying magnitude and patterns of response dependence. *J. Appl. Meas.* **2008**, *9*, 105–124. Available online: http://tinyurl.com/yc7mhmkw (accessed on 27 April 2024). [PubMed]
19. Noventa, S.; Spoto, A.; Heller, J.; Kelava, A. On a generalization of local independence in item response theory based on knowledge space theory. *Psychometrika* **2019**, *84*, 395–421. [CrossRef] [PubMed]
20. Ye, S.; Kelava, A.; Noventa, S. Parameter estimation of KST-IRT model under local dependence. *Psych* **2023**, *5*, 908–927. [CrossRef]
21. Wilson, M.; Adams, R.J. Rasch models for item bundles. *Psychometrika* **1995**, *60*, 181–198. [CrossRef]
22. Eckes, T. Item banking for C-tests: A polytomous Rasch modeling approach. *Psychol. Test Assess. Model.* **2011**, *53*, 414–439. Available online: http://tinyurl.com/bap2z4nh (accessed on 27 April 2024).

23. Ip, E.H. Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *Br. J. Math. Stat. Psychol.* **2010**, *63*, 395–416. [CrossRef] [PubMed]
24. Braeken, J. A boundary mixture approach to violations of conditional independence. *Psychometrika* **2011**, *76*, 57–76. [CrossRef]
25. Braeken, J.; Kuppens, P.; De Boeck, P.; Tuerlinckx, F. Contextualized personality questionnaires: A case for copulas in structural equation models for categorical data. *Multivar. Behav. Res.* **2013**, *48*, 845–870. [CrossRef]
26. Joe, H. *Dependence Modeling with Copulas*; CRC Press: Boca Raton, FL, USA, 2014.
27. Schroeders, U.; Robitzsch, A.; Schipolowski, S. A comparison of different psychometric approaches to modeling testlet structures: An example with C-tests. *J. Educ. Meas.* **2014**, *51*, 400–418. [CrossRef]
28. Nikoloulopoulos, A.K.; Joe, H. Factor copula models for item response data. *Psychometrika* **2015**, *80*, 126–150. [CrossRef]
29. Bellio, R.; Varin, C. A pairwise likelihood approach to generalized linear models with crossed random effects. *Stat. Model.* **2005**, *5*, 217–227. [CrossRef]
30. Chen, Y.; Moustaki, I.; Zhang, S. On the estimation of structural equation models with latent variables. In *Handbook of Structural Equation Modeling*; Hoyle, R.H., Ed.; Guilford Press: New York, NY, USA, 2023; pp. 145–162.
31. Katsikatsou, M.; Moustaki, I.; Yang-Wallentin, F.; Jöreskog, K.G. Pairwise likelihood estimation for factor analysis models with ordinal data. *Comput. Stat. Data Anal.* **2012**, *56*, 4243–4258. [CrossRef]
32. Renard, D.; Molenberghs, G.; Geys, H. A pairwise likelihood approach to estimation in multilevel probit models. *Comput. Stat. Data Anal.* **2004**, *44*, 649–667. [CrossRef]
33. Varin, C.; Reid, N.; Firth, D. An overview of composite likelihood methods. *Stat. Sin.* **2011**, *21*, 5–42. Available online: https://bit.ly/38lbhom (accessed on 27 April 2024).
34. Vasdekis, V.G.S.; Rizopoulos, D.; Moustaki, I. Weighted pairwise likelihood estimation for a general class of random effects models. *Biostatistics* **2014**, *15*, 677–689. [CrossRef]
35. Fieuws, S.; Verbeke, G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* **2006**, *62*, 424–431. [CrossRef] [PubMed]
36. Fieuws, S.; Verbeke, G.; Boen, F.; Delecluse, C. High dimensional multivariate mixed models for binary questionnaire data. *J. R. Stat. Soc. Ser. Appl. Stat.* **2006**, *55*, 449–460. [CrossRef]
37. Mauff, K.; Erler, N.S.; Kardys, I.; Rizopoulos, D. Pairwise estimation of multivariate longitudinal outcomes in a Bayesian setting with extensions to the joint model. *Modelling* **2021**, *21*, 115–136. [CrossRef]
38. Fu, Z.H.; Tao, J.; Shi, N.Z.; Zhang, M.; Lin, N. Analyzing longitudinal item response data via the pairwise fitting method. *Multivar. Behav. Res.* **2011**, *46*, 669–690. [CrossRef]
39. Robitzsch, A. Pairwise likelihood estimation of the 2PL model with locally dependent item responses. *Appl. Sci.* **2024**, *14*, 2652. [CrossRef]
40. Christoffersson, A. Factor analysis of dichotomized variables. *Psychometrika* **1975**, *40*, 5–32. [CrossRef]
41. Christoffersson, A. Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika* **1977**, *42*, 433–438. [CrossRef]
42. Muthén, B. Contributions to factor analysis of dichotomous variables. *Psychometrika* **1978**, *43*, 551–560. [CrossRef]
43. Muthén, B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **1984**, *49*, 115–132. [CrossRef]
44. Muthén, B.O.; Satorra, A. Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika* **1995**, *60*, 489–503. [CrossRef]
45. Huang, P.H. Penalized least squares for structural equation modeling with ordinal responses. *Multivar. Behav. Res.* **2022**, *57*, 279–297. [CrossRef]
46. Takane, Y.; de Leeuw, J. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* **1987**, *52*, 393–408. [CrossRef]
47. Kamata, A.; Bauer, D.J. A note on the relation between factor analytic and item response theory models. *Struct. Equ. Model. Multidiscip. J.* **2008**, *15*, 136–153. [CrossRef]
48. Paek, I.; Cui, M.; Öztürk Gübeş, N.; Yang, Y. Estimation of an IRT model by Mplus for dichotomously scored responses under different estimation methods. *Educ. Psychol. Meas.* **2018**, *78*, 569–588. [CrossRef]
49. Forero, C.G.; Maydeu-Olivares, A.; Gallardo-Pujol, D. Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Struct. Equ. Model.* **2009**, *16*, 625–641. [CrossRef]
50. Kappenburg-ten Holt, J. A Comparison between Factor Analysis and Item Response Theory Modeling in Scale Analysis. Unpublished Dissertation, University of Groningen, Groningen, The Netherlands, 2014. Available online: https://tinyurl.com/52yewx34 (accessed on 27 April 2024).
51. McDonald, R.P. Linear versus non-linear models in item response theory. *Appl. Psychol. Meas.* **1982**, *6*, 379–396. [CrossRef]
52. McDonald, R.P. *Test Theory: A Unified Treatment*; Lawrence Erlbaum: Mahwah, NJ, USA, 1999. [CrossRef]
53. Fraser, C.; McDonald, R.P. NOHARM: Least squares item factor analysis. *Multivar. Behav. Res.* **1988**, *23*, 267–269. [CrossRef] [PubMed]
54. McDonald, R.P. Normal-ogive multidimensional model. In *Handbook of Modern Item Response Theory*; van der Linden, W., Hambleton, R., Eds.; Springer: New York, NY, USA, 1997; pp. 257–269. [CrossRef]

55. Maydeu-Olivares, A. Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates. *J. Educ. Behav. Stat.* **2001**, *26*, 51–71. [CrossRef]
56. Swaminathan, H.; Rogers, H.J. Normal-ogive multidimensional models. In *Handbook of Item Response Theory, Volume 1: Models*; van der Linden, W.J., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 167–187.
57. Finch, H. Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis—Based models. *Appl. Psychol. Meas.* **2010**, *34*, 10–26. [CrossRef]
58. Finch, H. Multidimensional item response theory parameter estimation with nonsimple structure items. *Appl. Psychol. Meas.* **2011**, *35*, 67–82. [CrossRef]
59. Sass, D.A. Factor loading estimation error and stability using exploratory factor analysis. *Educ. Psychol. Meas.* **2010**, *70*, 557–577. [CrossRef]
60. Svetina, D.; Levy, R. An overview of software for conducting dimensionality assessment in multidimensional models. *Appl. Psychol. Meas.* **2012**, *36*, 659–669. [CrossRef]
61. Svetina, D.; Levy, R. Dimensionality in compensatory MIRT when complex structure exists: Evaluation of DETECT and NOHARM. *J. Exp. Educ.* **2016**, *84*, 398–420. [CrossRef]
62. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2023. Available online: https://www.R-project.org/ (accessed on 15 March 2023).
63. Robitzsch, A. sirt: Supplementary Item Response Theory Models. R Package Version 4.2-57. 2024. Available online: https://github.com/alexanderrobitzsch/sirt (accessed on 20 April 2024).
64. Rosseel, Y. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [CrossRef]
65. Lietz, P.; Cresswell, J.C.; Rust, K.F.; Adams, R.J. (Eds.) *Implementation of Large-scale Education Assessments*; Wiley: New York, NY, USA, 2017. [CrossRef]
66. Rutkowski, L.; von Davier, M.; Rutkowski, D. (Eds.) *A Handbook of International Large-scale Assessment: Background, Technical Issues, and Methods of Data Analysis*; Chapman Hall; CRC Press: London, UK, 2013. [CrossRef]
67. Habing, B.; Roussos, L.A. On the need for negative local item dependence. *Psychometrika* **2003**, *68*, 435–451. [CrossRef]
68. Robitzsch, A.; Lüdtke, O. Some thoughts on analytical choices in the scaling model for test scores in international large-scale assessment studies. *Meas. Instruments Soc. Sci.* **2022**, *4*, 9. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.