the british psychological society
promoting excellence in psychology

British Journal of Mathematical and Statistical Psychology (2021)
© 2021 The Authors. *British Journal of Mathematical and Statistical Psychology* published by
John Wiley Sons Ltd on behalf of British Psychological Society

www.wileyonlinelibrary.com

Pairwise likelihood estimation for confirmatory factor analysis models with categorical variables and data that are missing at random

Myrsini Katsikatsou[1], Irini Moustaki[2]* and Haziq Jamil[3]
[1]Horrothia FX, Falmouth, UK
[2]London School of Economics and Political Science, UK
[3]Universiti Brunei Darussalam, Brunei

Methods for the treatment of item non-response in attitudinal scales and in large-scale assessments under the pairwise likelihood (PL) estimation framework and under a missing at random (MAR) mechanism are proposed. Under a full information likelihood estimation framework and MAR, ignorability of the missing data mechanism does not lead to biased estimates. However, this is not the case for pseudo-likelihood approaches such as the PL. We develop and study the performance of three strategies for incorporating missing values into confirmatory factor analysis under the PL framework, the complete-pairs (CP), the available-cases (AC) and the doubly robust (DR) approaches. The CP and AC require only a model for the observed data and standard errors are easy to compute. Doubly-robust versions of the PL estimation require a predictive model for the missing responses given the observed ones and are computationally more demanding than the AC and CP. A simulation study is used to compare the proposed methods. The proposed methods are employed to analyze the UK data on numeracy and literacy collected as part of the OECD Survey of Adult Skills.

No survey ever attains 100% response. Item non-response occurs when the individual fails to respond to some of the items and/or due to the design of a survey. An example of the latter is adaptive testing where only a part of the test items is administered to a respondent, the choice of which is determined by respondent's answers to previous questions in the test. Multivariate outcomes are often analysed using an exploratory or confirmatory factor analysis type model. In factor analysis for categorical variables, full-information maximum likelihood (FIML) estimation is not computationally feasible with a large number of observed ordinal variables (Lee, Poon, & Bentler, 1990; Poon & Lee, 1987). Instead, limited-information estimation methods have been developed. The most widely used is a three-stage weighted least squares method, with diagonally weighted least squares (DWLS) and unweighted least squares (ULS) as special cases (Jöreskog, 1994; Muthén, 1984). An alternative limited-information estimation method, recently proposed, is pairwise likelihood (PL) estimation, which is the focus of this paper. PL has been found to

*Correspondence should be addressed to Irini Moustaki, Department of Statistics, Houghton Street, London WC2A 2AE, UK (email: i.moustaki@lse.ac.uk).

DOI:10.1111/bmsp.12243

be a competitive alternative to DWLS and ULS for fitting confirmatory factor analysis (CFA) models with binary/ordinal variables (De Leon, 2005; Jöreskog & Moustaki, 2001; Katsikatsou, 2013; Katsikatsou, Moustaki, Yang-Wallentin, & Jöreskog, 2012; Liu, 2007; Xi, 2011).

In the context of item non-response the missing at random (MAR) assumption implies that the probability of missingness depends only on observed data (Rubin, 1976). For models with latent variables, it is also important to distinguish between covariate-dependent MAR (CD-MAR) (Asparouhov & Muthén, 2010), in which the probability of missingness depends on observed covariates, and indicator-dependent MAR (ID-MAR), in which the probability of missingness depends on the observed variables that are used as the indicators for measuring the latent variable(s). In Rubin's (1976) terminology, when the parameters of the missingness mechanism and those of the substantive model for the observed data are distinct, MAR missingness is ignorable for likelihood-based estimation, meaning that the estimation of the substantive model can be done without specifying a model for the probability of missingness. The result of MAR ignorability, however, applies only to FIML and cannot be extended, in general, to limited-information estimation methods such as three-stage weighted least squares and pairwise likelihood which is a pseudo-likelihood method (Molenberghs, Kenward, Verbeke, & Birhanu, 2011).

Item non-response is typically dealt with by either pairwise deletion (for each computation involving a pair of variables, individuals missing either item in the pair are excluded) or by listwise deletion, also known as complete-case analysis (each individual with any missing data is excluded). These approaches deal adequately only with cases where the data are missing completely at random. Otherwise, there remains the concern that excluding the non-response in this way may lead to bias. For factor analysis models with ordinal variables, Asparouhov and Muthén (2010) noted that, applying listwise or pairwise deletion, DWLS and ULS provide unbiased estimates for the model parameters only under CD-MAR and not under ID-MAR. Thus, in the paper, we focus on ID-MAR data missingness under the PL estimation framework.

Within the pseudo-likelihood estimation framework, Molenberghs et al. (2011) study the ignorability issue under MAR for incomplete clustered data and longitudinal data with monotone missingness. They found that pairwise and listwise estimators yield biased estimates in general under MAR. They propose instead a method that is based on inverse probability weighting (IPW) and another method that combines both IPW and ideas from doubly robust (DR) methods (Bang & Robins, 2005; Robins, Rotnitzky, & Zhao, 1995). In the IPW approach, the contribution of an individual to the likelihood is weighted by the inverse probability of their response pattern being observed (Robins & Rotnitzky, 1992; Robins, Rotnitzky, & Zhao, 1994; Robins et al., 1995; Rotnitzky, 2009). In the DR case, weighting is complemented by incorporating a predictive model for the missing values given the observed values (Bang & Robins, 2005; Rotnitzky, 2009; Scharfstein, Rotnitzky, & Robins, 1999). DR is preferred to the IPW approach because IPW is sensitive to misspecification of the missing-data mechanism model, while DR estimators remain consistent when either the model for the missingness mechanism or the predictive model is correctly specified (Bang & Robins, 2005). For this reason, in this paper we do not consider any IPW formulations of PL for CFA models. Molenberghs et al., 2011 apply the DR approach to complete pairs (CP) and available cases (AC) separately to arrive at the same log-likelihood function where the terms referring to the missing-data mechanism model cancel out. Thus, the DR formulation of PL includes only the predictive terms. Molenberghs et al., 2011 provide details of the DR approach for the special case of longitudinal data with drop-out and acknowledge that the implementation of DR in the case of general missingness patterns is more complicated. Birhanu (2012) conducts a simulation study with longitudinal binary data and reports that DR performs better than CP, AC and IPW estimators with respect to bias and MSE. Also, AC and IPW estimators exhibit efficiency comparable to that of FIML, while DR is found to be more efficient than FIML. The naive CP estimator has also been studied by (He & Yi, 2011) and (Yi, Zeng, & Cook, 2011) for models with longitudinal binary clustered data and correlated binary data respectively, and by Fonseca and Grassetti (2010) for vector autoregressive models. In these specific cases, the proposed CP versions are found to have acceptable performance and are recommended.

Our goal is to develop methods for handling item non-response within the PL pairwise likelihood (PL) estimation framework in the analysis of attitudinal scales and large-scale assessment data under an MAR missing at random mechanism. More specifically, following the work done by Molenberghs et al. (2011), we develop the CP, AC and DR versions of PL for CFA models with ordinal variables in the case of item non-response with an ID-MAR missing-data mechanism and any pattern of missingness. Our main research question is whether the general result, that CP and AC yield biased estimators, applies to the specific framework of CFA models.

We examine the performance of CP, AC and DR in a simulation study. For completeness we also show results from multiple imputation under the diagonally weighted least squares method (MI-DWLS), which is the standard estimation framework for CFA models with ordinal variables and data that are MAR. The results generalize to exploratory factor analysis and structural equation modelling.

The rest of the paper is structured as follows. Section 2 presents the notation and the model framework. Section 3, after briefly discussing PL estimation for CFA models with ordinal variables and completely observed data, details the proposed methodology for handling item non-response under ID-MAR. The aim of Section 5 though is the performance of the proposed estimators using a simulation study, while Section 4 analyses Programme for the International Assessment of Adult Competencies (PIAAC) data using the proposed PL methods. Section 6 discusses areas for future research and concludes.

2. Notation and model framework

Let $\mathbf{y} = (y_1, \cdots, y_p)'$ be a p-dimensional vector of categorical (binary, ordinal) variables, and $\boldsymbol{\eta}$ a q-dimensional vector of continuous latent variables. y_i is assumed to be the manifestation of an underlying continuous variable y_i^* where

$$y_i = a \Leftrightarrow \tau_{i,a-1} < y_i^* < \tau_{i,a}, \tag{1}$$

a is the *ath* response category of variable $y_i, a = 1, \ldots, C_i$, $\tau_{i,a}$ is the *ath* threshold of variable y_i, and $-\infty = \tau_{i,0} < \tau_{i,1} < \ldots < \tau_{i,C_i-1} < \tau_{i,C_i} = +\infty$. Let $\boldsymbol{\tau}$ be the vector of all thresholds, and $\mathbf{y}^*$ be the p-dimensional vector of the underlying continuous variables. The factor analysis model for $\mathbf{y}^*$ is

$$\mathbf{y}^* = \Lambda\boldsymbol{\eta} + \varepsilon, \tag{2}$$

where Λ is a $p \times q$ matrix of factor loadings, ε is the vector of unique error terms with $\varepsilon \sim \mathcal{N}_p(\mathbf{0}, \Theta_\varepsilon)$, $\boldsymbol{\eta} \sim \mathcal{N}_q(0, \Phi)$, and $\text{Cov}(\boldsymbol{\eta}, \varepsilon) = \mathbf{0}$. The parameter vector of the model, denoted by $\boldsymbol{\theta}$, includes the free parameters in Λ, Φ, Θ_ε and τ. Based on the model, $\mathbf{y}^* \sim \mathcal{N}_p(\boldsymbol{\mu}, P)$ where $P = \Lambda\Phi\Lambda' + \Theta_\varepsilon$. A structural equation model (SEM) with categorical variables consists of equations (1) and (2), and equation (3) below. The latter defines the relationships among the latent variables, including covariates:

$$\boldsymbol{\eta} = B\boldsymbol{\eta} + \Gamma\mathbf{x} + \zeta, \tag{3}$$

where $\mathbf{x}$ is the vector of covariates, $\mathbf{B}$ and Γ are parameter matrices, $\mathbf{I} - \mathbf{B}$ is a non-singular matrix with I being the identity matrix, and ζ is the vector of error terms for which it is typically assumed $\zeta \sim \mathcal{N}_q(\mathbf{0}, \Psi)$ and $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\zeta}) = \text{Cov}(\varepsilon, \boldsymbol{\zeta}) = \mathbf{0}$. The main difference between an SEM and a CFA model is that the former, through equation (3), imposes a parametric structure on the factor covariance matrix Φ.

3. Pairwise likelihood estimation

3.1. PL for CFA with ordinal variables and completely observed data

Pairwise likelihood, a member of the family of composite likelihood methods, has been proposed as an alternative to the standard DWLS approach for estimating CFA and SEM with ordinal variables (De Leon, 2005; Jöreskog & Moustaki, 2001; Katsikatsou, 2013; Katsikatsou et al., 2012; Liu, 2007).

The pairwise log-likelihood (*pl*) function is defined as the sum of the bivariate log-likelihood functions. For a single observation denoted by n, it takes the form

$$pl_n(\boldsymbol{\theta}; \mathbf{y}_n) = \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \log f\left(y_{ni}, y_{nj}; \boldsymbol{\theta}\right). \tag{4}$$

For ordinal variables, a bivariate log-likelihood function is a multinomial one, that is,

$$\log f\left(y_{ni}, y_{nj}; \boldsymbol{\theta}\right) = \sum_{a=1}^{c_i} \sum_{b=1}^{c_j} I\left(y_{ni} = a, y_{nj} = b\right) \ln \pi\left(y_{ni} = a, y_{nj} = b; \boldsymbol{\theta}\right), \tag{5}$$

where $I\left(y_{ni} = a, y_{nj} = b\right)$ is an indicator variable indicating whether y_{ni} and y_{nj} fall into categories a and b, respectively, and $\pi\left(y_{ni} = a, y_{nj} = b; \boldsymbol{\theta}\right)$ is the corresponding probability, which, under the CFA model defined by equations (1) and (2), is

$$\begin{aligned}\pi\left(y_{ni} = a, y_{nj} = b; \boldsymbol{\theta}\right) &= \int_{\tau_{i,a-1}}^{\tau_{i,a}} \int_{\tau_{j,b-1}}^{\tau_{j,b}} f\left(y^*_{ni}, y^*_{nj}; \boldsymbol{\theta}\right) dy^*_{ni} dy^*_{nj} \\ &= \Phi_2\left(\tau_{i,a}, \tau_{j,b}; \rho_{ij}\right) - \Phi_2\left(\tau_{i,a-1}, \tau_{j,b}; \rho_{ij}\right) \\ &\quad - \Phi_2\left(\tau_{i,a}, \tau_{j,b-1}; \rho_{ij}\right) + \Phi_2\left(\tau_{i,a-1}, \tau_{j,b-1}; \rho_{ij}\right),\end{aligned} \tag{6}$$

where ρ_{ij} is the polychoric correlation between y_i^* and y_j^*, and $\Phi_2(\tau_1, \tau_2; \rho)$ is the bivariate cumulative normal distribution with correlation ρ evaluated at the point (τ_1, τ_2). For a random sample of N observations, the pairwise log-likelihood function is $pl(\boldsymbol{\theta};\mathbf{y}) = pl(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_N)) = \sum_{n=1}^{N} pl_n(\boldsymbol{\theta};\mathbf{y}_n)$. Maximizing it over $\boldsymbol{\theta}$, we obtain the PL estimator,$\hat{\boldsymbol{\theta}}_{\mathrm{PL}}$.

Based on composite likelihood theory results (Lindsay, 1988), $\hat{\boldsymbol{\theta}}_{\mathrm{PL}}$ is asymptotically consistent and normally distributed. In particular, we have $\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\mathrm{PL}} - \boldsymbol{\theta}\right) \rightarrow^{d} \mathcal{N}(0, G^{-1}(\boldsymbol{\theta}))$, where $G(\boldsymbol{\theta})$ is the Godambe information matrix (also known as the sandwich information matrix),

$$G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J^{-1}(\boldsymbol{\theta})H(\boldsymbol{\theta}),$$

with $H(\boldsymbol{\theta})$ and$J(\boldsymbol{\theta})$ estimated respectively by

$$\hat{H}(\hat{\boldsymbol{\theta}}_{\mathrm{PL}}) = -\frac{1}{N}\left(\frac{\partial^2}{\partial\theta'\partial\boldsymbol{\theta}} pl(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_N))\right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{PL}}} \tag{7}$$

and

$$\hat{J}(\hat{\boldsymbol{\theta}}_{\mathrm{PL}}) = \frac{1}{N}\sum_{n=1}^{N}\left(\left(\frac{\partial}{\partial\theta'} pl_n(\boldsymbol{\theta};\mathbf{y}_n)\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{PL}}}\right)\left(\left(\frac{\partial}{\partial\theta'} pl_n(\boldsymbol{\theta};\mathbf{y}_n)\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{PL}}}\right)'. \tag{8}$$

In finite samples, simulation studies indicate that the PL estimates and standard errors have close to zero bias and mean square error, both decreasing with increasing sample size (Katsikatsou et al., 2012).

3.2. Proposed PL methods for CFA with ordinal variables under MAR

Let$pl^{\mathrm{CP}}(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_N))$,$pl^{\mathrm{AC}}(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_N))$ and$pl^{\mathrm{DR}}(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_N))$ denote the complete-pairs, available-cases and doubly robust pairwise log-likelihood functions respectively, for a sample of$N$observations. Maximizing the functions over $\boldsymbol{\theta}$, we obtain the CP estimator, $\hat{\boldsymbol{\theta}}_{\mathrm{CP}}$, the AC estimator, $\hat{\boldsymbol{\theta}}_{\mathrm{AC}}$, and the DR estimator, $\hat{\boldsymbol{\theta}}_{\mathrm{DR}}$, respectively. For a random sample of observations, each log-likelihood function is equal to the sum of the N individual contributions, the exact form of which is given below. Let $\tilde{p}_n$ and m_n be the number of items with observed values and missing values respectively, for sample unit n, where $\tilde{p}_n + m_n = p$, and$\tilde{p}_n > 0$ in the case of item non-response. Also, let $\mathbf{y}_n^{\circ}$ and $\mathbf{y}_n^{\mathrm{m}}$ denote the$\tilde{p}_n$ -dimensional vector of observed variables and the

m_n-dimensional vector of missing variables, respectively, for that unit. The contribution of observation n to$pl^{\mathrm{CP}}(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_N))$ is

$$pl_n^{\mathrm{CP}}(\boldsymbol{\theta};\mathbf{y}_n) = \sum_{i=1}^{\tilde{p}_n-1}\sum_{j=i+1}^{\tilde{p}_n} \log f\left(y_{ni}^{\circ}, y_{nj}^{\circ};\boldsymbol{\theta}\right), \tag{9}$$

where $\log f\left(y_{ni}^{\circ}, y_{nj}^{\circ};\boldsymbol{\theta}\right)$ is defined in (5). In the case of AC, the contribution of observation n is

$$pl_n^{\mathrm{AC}}(\boldsymbol{\theta};\mathbf{y}_n) = pl_n^{\mathrm{CP}}(\boldsymbol{\theta};\mathbf{y}_n) + m_n \sum_{i=1}^{\tilde{p}_n} \log f\left(y_{ni}^{\circ};\boldsymbol{\theta}\right), \tag{10}$$

where, based on the CFA model defined in Section 2,

$$\log f\left(y_{ni}^{\mathrm{o}};\boldsymbol{\theta}\right) = \sum_{a=1}^{c_i} I\left(y_{ni}^{\circ} = a\right) \ln \pi\left(y_{ni}^{\circ} = a;\boldsymbol{\theta}\right), \tag{11}$$

$I\left(y_{ni}^{\circ} = a\right)$ is an indicator for whether y_{ni}° falls into category a, and the corresponding probability $\pi\left(y_{ni}^{\circ} = a;\boldsymbol{\theta}\right)$ is.

$$\pi\left(y_{ni}^{\circ} = a;\boldsymbol{\theta}\right) = \int_{\tau_{i,a-1}}^{\tau_{i,a}} f\left(y_{ni}^{*};\boldsymbol{\theta}\right) dy_{ni}^{*} = \Phi_1(\tau_{i,a}) - \Phi_1(\tau_{i,a-1}). \tag{12}$$

with $\Phi_1(\tau)$ being the univariate cumulative standard normal distribution evaluated at point τ. Note that in equation (10) the number of missing variables for sample unit n appears as a weight of the univariate log-likelihood functions of the observed variables. This is derived by taking into account those pairs of variables of which one is observed and one is missing, written as $\log f\left(y_{ni}^{\circ}, y_{nj}^{\mathrm{m}};\boldsymbol{\theta}\right) = \log f\left(y_{ni}^{\circ}\right) + \log f\left(y_{nj}^{\mathrm{m}} | y_{ni}^{\circ};\boldsymbol{\theta}\right)$. In AC, by definition, we keep only the information that is observed.

The standard errors of $\hat{\boldsymbol{\theta}}_{\mathrm{CP}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{AC}}$ are obtained from the Godambe information matrix using the expressions given in equations (7) and (8), where pl is replaced by pl^{CP} and pl^{AC}, and $\hat{\boldsymbol{\theta}}_{\mathrm{PL}}$ is replaced by $\hat{\boldsymbol{\theta}}_{\mathrm{CP}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{AC}}$, respectively. Note that for likelihood-based inference, when data are MAR, Kenward and Molenberghs (1998) have shown that the classical expected information matrix is biased and recommend the use of the observed information matrix.

DR requires a predictive model of the missing responses given the observed responses. The contribution of observation n is defined as

$$\begin{aligned} pl_n^{\mathrm{DR}}(\boldsymbol{\theta};\mathbf{y}_n) &= pl_n^{\mathrm{AC}}(\boldsymbol{\theta};\mathbf{y}_n) + \sum_{i=1}^{m_n-1} \sum_{j=i+1}^{m_n} E_{(y_{ni}^{\mathrm{m}}, y_{nj}^{\mathrm{m}})|\mathbf{y}_n^{\circ}}\left[\log f\left(y_{ni}^{\mathrm{m}}, y_{nj}^{\mathrm{m}};\boldsymbol{\theta}\right)\right] \\ &\quad + \sum_{i=1}^{\tilde{p}_n} \sum_{j=1}^{m_n} E_{y_{nj}^{\mathrm{m}}|\mathbf{y}_n^{\circ}}\left[\log f\left(y_{nj}^{\mathrm{m}} | y_{ni}^{\circ};\boldsymbol{\theta}\right)\right], \end{aligned} \tag{13}$$

where

$$\begin{aligned} E_{(y_{ni}^{\mathrm{m}}, y_{nj}^{\mathrm{m}})|\mathbf{y}_n^{\circ}}\left[\log f\left(y_{ni}^{\mathrm{m}}, y_{nj}^{\mathrm{m}};\boldsymbol{\theta}\right)\right] &= \sum_{a=1}^{c_i} \sum_{b=1}^{c_j} \left\{E_{(y_{ni}^{\mathrm{m}}, y_{nj}^{\mathrm{m}})|\mathbf{y}_n^{\circ}}\left[I\left(y_{ni}^{\mathrm{m}} = a, y_{nj}^{\mathrm{m}} = b\right)\right]\right\} \ln \pi\left(y_{ni}^{\mathrm{m}} = a, y_{nj}^{\mathrm{m}} = b;\boldsymbol{\theta}\right) \\ &= \sum_{a=1}^{c_i} \sum_{b=1}^{c_j} \Pr\left(y_{ni}^{\mathrm{m}} = a, y_{nj}^{\mathrm{m}} = b | \mathbf{y}_n^{\circ}\right) \ln \pi\left(y_{ni}^{\mathrm{m}} = a, y_{nj}^{\mathrm{m}} = b;\boldsymbol{\theta}\right) \end{aligned}$$

and

$$\begin{aligned} E_{y^{\mathrm{m}}_{nj}|\mathbf{y}^{\circ}_{n}}\left[\log f\left(y^{\mathrm{m}}_{nj}|y^{\circ}_{ni};\boldsymbol{\theta}\right)\right] &= \sum_{b=1}^{c_j}\left\{E_{y^{\mathrm{m}}_{nj}|\mathbf{y}^{\circ}_{n}}\left[I\left(y^{\mathrm{m}}_{nj}=b\right)\right]\right\}\ln\pi\left(y^{\mathrm{m}}_{nj}=b|y^{\circ}_{ni}=a;\boldsymbol{\theta}\right) \\ &= \sum_{b}\Pr\left(y^{\mathrm{m}}_{nj}=b|\mathbf{y}^{\circ}_{n}\right)\ln\pi\left(y^{\mathrm{m}}_{nj}=b|y^{\circ}_{ni}=a;\boldsymbol{\theta}\right). \end{aligned}$$

Furthermore, $\pi\left(y^{\mathrm{m}}_{ni}=a,y^{\mathrm{m}}_{nj}=b;\boldsymbol{\theta}\right)$ is given in (6), and

$$\pi\left(y^{\mathrm{m}}_{nj}=b|y^{\circ}_{ni}=a;\boldsymbol{\theta}\right)=\frac{\pi\left(y^{\mathrm{m}}_{nj}=b,y^{\circ}_{ni}=a;\boldsymbol{\theta}\right)}{\pi(y^{\circ}_{ni}=a;\boldsymbol{\theta})}.$$

with $\pi(y^{\circ}_{ni}=a;\boldsymbol{\theta})$ defined in (12).

The probabilities $\Pr\left(y^{\mathrm{m}}_{ni}=a,y^{\mathrm{m}}_{nj}=b|\mathbf{y}^{\circ}_{n}\right)$ and $\Pr\left(y^{\mathrm{m}}_{nj}=b|\mathbf{y}^{\circ}_{n}\right)$ may or may not depend on the model parameter vector $\boldsymbol{\theta}$. If not, they need to be computed for each of the N sample units and then plugged into their corresponding $pl^{\mathrm{DR}}_{n}(\boldsymbol{\theta};\mathbf{y}_n)$. So, in this case, the DR estimation involves two steps.

We consider three alternative models for $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b|\mathbf{y}^{\circ}\right)$ (the subscript n is dropped since the same model applies to all sample units). The first model is the unconstrained model, $\mathbf{y}^{*}\sim\mathcal{N}_p(\mathbf{0},P)$, where the polychoric correlation matrix P is unconstrained. After expressing $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b|\mathbf{y}^{\circ}\right)$ as $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b,\mathbf{y}^{\circ}\right)/\Pr(\mathbf{y}^{\circ})$, the probabilities in the numerator and denominator are computed in the same fashion as in (6) with the difference that the dimensions of the integrals are equal to $\tilde{p}+2$ and $\tilde{p}$, respectively. We will refer to this approach as the unconstrained model exact probability (UMEP) approach. Since the unconstrained model requires enough data for all pairs of the p variables, so that their polychoric correlations can be estimated, the UMEP cannot be employed in designs with planned missingness, where certain pairs of items are never administered together.

The second model we consider is the hypothesized CFA model, $\mathbf{y}^{*}\sim\mathcal{N}_p(\mathbf{0},P)$ with $P=\Lambda\Phi\Lambda'+\Theta_{\varepsilon}$. As in UMEP, after expressing $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b|\mathbf{y}^{\circ}\right)$ as $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b,\mathbf{y}^{\circ}\right)/\Pr(\mathbf{y}^{\circ})$, the latter two are computed in the same fashion as in equation (6). We will refer to this approach as the hypothesized model exact probability (HMEP) approach. When the hypothesized model is true, the UMEP and HMEP are expected to give very similar results as the hypothesized model is nested in the unconstrained model. The advantage of the HMEP is that it can be applied to data collected from designs with planned missingness.

Both the UMEP and the HMEP are computationally demanding since one integration of dimension $\tilde{p}+2$ and one of dimension $\tilde{p}$ need to be computed. A less computationally intensive approach is to approximate $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b|\mathbf{y}^{\circ}\right)$ with $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b|\boldsymbol{\eta}\right)$, where $\boldsymbol{\eta}$ is replaced by the regression factor scores, $\tilde{\boldsymbol{\eta}}$.

Following the model described in Section 2, we have that $\Pr\left(y^{\mathrm{m}}_{i}=a,y^{\mathrm{m}}_{j}=b|\boldsymbol{\eta}\right)=\int_{\tau_{i,a-1}}^{\tau_{i,a}}\int_{\tau_{i,b-1}}^{\tau_{i,b}}f\left(y^{*}_{i},y^{*}_{j}|\boldsymbol{\eta}\right)dy^{*}_{j}dy^{*}_{i}$, where

$$\left(y^{*}_{i}\ \ y^{*}_{j}\right)'|\boldsymbol{\eta}\sim\mathcal{N}_2\left([\boldsymbol{\Lambda}]_{ij,.}\boldsymbol{\eta},[\Theta_{\varepsilon}]_{ij,ij}\right),$$

$[\Lambda]_{ij,.}$ denotes the $2\times q$ sub-matrix of Λ that includes only the *i*th and *j*th rows and all columns, and $[\Theta_{\varepsilon}]_{ij,ij}$ is the 2×2 sub-matrix of Θ_{ε} that includes the elements that are simultaneously in the *i*th and *j*th rows and *i*th and *j*th columns. This approach will be referred to as the hypothesized model approximate probability (HMAP) approach. It requires the computation of only one two-dimensional integral regardless of the sizes of p and m. However, it relies on the assumption that the hypothesized model is true and that $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \boldsymbol{\eta}\right)$ is a good approximation of $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$. Apart from the UMEP, HMEP, and HMAP, any other model for $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ motivated by a specific application and/or the data at hand could be employed. Here, we just consider the obvious candidate models within the context of CFA with ordinal variables.

The univariate conditional probability $\Pr\left(y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ involved in DR can be calculated from the bivariate conditional probabilities $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$, summing over the response categories of y_i^{m} when at least two variables have missing values. If there is only one variable missing, the approach adopted for $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ is applied. In particular, after expressing $\Pr\left(y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ as $\Pr\left(y_j^{\mathrm{m}}=b, \mathbf{y}^{\circ}\right)/\Pr(\mathbf{y}^{\circ})$, we employ the unconstrained model $\mathbf{y}^* \sim \mathcal{N}(\mathbf{0}, P)$ in the case of UMEP or the hypothesized model in the case of HMEP to compute $\Pr\left(y_j^{\mathrm{m}}=b, \mathbf{y}^{\mathrm{o}}\right)$ and $\Pr(\mathbf{y}^{\mathrm{o}})$ in the same fashion as in (6). For the HMAP, we approximate $\Pr\left(y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\mathrm{o}}\right)$ with $\Pr\left(y_j^{\mathrm{m}}=b \mid \boldsymbol{\eta}\right)$ using the hypothesized model and the regression factor scores.

In the UMEP and the HMEP approaches, although the estimation of $\boldsymbol{\theta}$ could theoretically be done in one step, since $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ and $\Pr\left(y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ are functions of $\boldsymbol{\theta}$, it defeats the purpose of PL. PL is suggested as a computationally feasible alternative to maximum likelihood estimation because it requires the computation of up to two-dimensional integrals (written in closed form in (6) and (12)) regardless of the size of p. Thus, in practice, all three proposed versions of DR involve two steps. At the first step, the selected model, unconstrained or hypothesized, is fitted to the data at hand. For fitting the model, we recommend AC because our simulation results (reported in the next section) show that, although AC and CP provide very similar results for loadings and factor correlations, AC exhibits smaller average standardized bias for thresholds than CP. Using $\hat{\boldsymbol{\theta}}_{\mathrm{AC}}$, we can estimate the probabilities $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$ and $\Pr\left(y_j^{\mathrm{m}}=b \mid \mathbf{y}^{\circ}\right)$. In the HMAP approach, $\hat{\boldsymbol{\theta}}_{\mathrm{AC}}$ is used to first estimate the regression factor scores and then $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b \mid \boldsymbol{\eta}\right)$ and $\Pr\left(y_j^{\mathrm{m}}=b \mid \boldsymbol{\eta}\right)$ can be estimated.

4. Simulation study

4.1. Set-up

We use a simulation study to examine the finite-sample performance of the proposed pairwise likelihood estimators (CP, AC, UMEP, HMEP and HMAP), when applied to CFA with ordinal variables and data that are ID-MAR. For completeness, we also present the results for MI-DWLS as implemented in Mplus. Ten imputed data sets are produced (in the literature, five to ten imputed data sets are usually considered enough) and the imputation model is the variance–covariance model (i.e., $\mathbf{y}^* \sim \mathcal{N}_p(\mathbf{0}, P)$ with unconstrained P). For the imputation, Mplus uses a Markov chain Monte Carlo (MCMC) simulation procedure (Rubin, 1978; Schafer, 1997) and the imputation model is estimated with the Bayesian estimation method.

To compute the proposed PL estimators, we wrote our R routines which are incorporated in the R package *lavaan* (version 0.5-23.1043 and beyond). The specific commands with detailed explanations are given in the Supporting Information. Our simulation study consists of two parts; the results of part I partly inform the design of part II.

4.2. Performance criteria

We report percentages of convergence and proper solutions for each simulation condition, average and individual parameter relative bias, raw bias, and root mean square error (RMSE) of factor loadings, thresholds, factor correlations and their standard errors as well as coverage rates. When average statistics are reported for each parameter type, the absolute values of the relative or raw bias for each individual parameter are used.

Let us denote by $\hat{\theta}_{l,k,r}$ the estimated parameter value of the lth parameter in the rth replication for the kth estimation method, where $r = 1,\ldots,R$, $l = 1,\ldots,L$ and $k = \text{CP, AC, UMEP, HMEP, HMAP}$ (R is the total number of replications, L is the total number of parameters). To study the performance of the different estimation methods, we compute first the raw bias and relative bias of parameter estimates $\hat{\theta}_{l,k}$, given by

$$RawB_{(l,k)} = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_{l,k,r} - \theta_l\right), \tag{14}$$

and

$$RelB_{(l,k)} = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_{l,k,r} - \theta_l\right)/\theta_l \times 100. \tag{15}$$

respectively, where θ_l is the true parameter value. Values of relative bias less than 10% are considered acceptable, values of 10–20% indicate substantial bias, and values greater than 20% indicate unacceptable bias ((Forero & Maydeu-Olivares, 2009). The RMSE assesses the combined effect of parameter bias and parameter variance and is given by

$$RMSE_{(l,k)} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_{l,k,r} - \theta_l\right)^2} \tag{16}$$

Since we are not interested in the performance of the individual parameters, we average the absolute relative bias, the absolute raw bias and the root mean square error across all parameters of the same type such as loadings, thresholds and factor correlations. We also provide either the $RawB_{(l,k)}$ or $RelB_{(l,k)}$ for each individual parameter estimate in the Supporting Information.

The 95% confidence interval (CI) coverage rate for the parameters of interest is computed as the proportion of the 95% CIs that include the true parameter value across replications, given by $\sum_{r=1}^{R} I\left[\theta \in \hat{\theta}_{l,k,r} \pm 1.96\hat{se}\left(\hat{\theta}_{l,k,r}\right)\right]/R$, where $I(\cdot)$ is the indicator function, and $\hat{se}\left(\hat{\theta}_{l,k,r}\right)$ is the estimated standard error of $\hat{\theta}_{l,k,r}$ for the lth parameter at the rth replication for method k, k = CP, AC, MI-DWLS.

Equations (14), (15) and (16) can be also applied to the estimated standard errors by replacing θ_l with the standard deviation of the parameter estimates obtained from the R replications. For UMEP, HMEP and HMAP, we first consider only the relative or raw bias and RMSE of parameter estimates.

The simulation results for these three criteria will inform us whether we need to undertake the complex task of deriving the standard errors of their estimated parameters.

Table 1. True parameter values of factor loadings and factor correlations for the binary and ordinal data generating models considered in part I of the simulation study

	η_1						η_2						
	y_1^*	y_2	y_3	y_4	y_5	y_6	y_7^*	y_8	y_9	y_{10}	y_{11}	y_{12}	
Model	$\lambda_{1,1}$	$\lambda_{2,1}$	$\lambda_{3,1}$	$\lambda_{4,1}$	$\lambda_{5,1}$	$\lambda_{6,1}$	$\lambda_{7,1}$	$\lambda_{8,1}$	$\lambda_{9,1}$	$\lambda_{10,1}$	$\lambda_{11,1}$	$\lambda_{12,1}$	ϕ_{12}
1	.8	.8	.8	.8	.8	.8	–	–	–	–	–	–	–
2	.4	.8	.8	.8	.8	.8	–	–	–	–	–	–	–
3	.6	.6	.6	.6	.6	.6	–	–	–	–	–	–	–
4	.4	.5	.6	.7	.8	.9	.4	.5	.6	.7	.8	.9	.3
5	.4	.5	.6	.7	.8	.9	.4	.5	.6	.7	.8	.9	0.6

*Always observed.

Note that the probabilities $\Pr\left(y_i^{\mathrm{m}}=a, y_j^{\mathrm{m}}=b|\mathbf{y}^{\circ}\right)$ and $\Pr\left(y_j^{\mathrm{m}}=b|\mathbf{y}^{\circ}\right)$ included in the objective function for the DR estimators are estimated at step 1, preceding the function's optimization, and the standard errors need to reflect the sampling variability of their estimates.

4.3. Part I of the simulation study

We consider ten experimental conditions derived from five models and a small and medium sample size of 300 and 1,000 respectively. The same ten conditions are used for binary and ordinal variables with four categories. The five data-generating models are summarized in Table 1 for the binary and ordinal variables, respectively. The first three models are one-factor models with six binary or six ordinal variables each, where the first variable, y_1, is observed for all sample members. In model 1 all loadings are equal to .8 (strong discrimination); in model 2 the loading of y_1, which determines the missingness, is .4 and the rest of the loadings remain .8; and in model 3 the loadings of all variables are .6 (medium discrimination). In all three models, the factor variance is fixed to 1 to define the unit of the factor scale. Models 4 and 5 are two-factor models where each factor is measured by six binary or ordinal variables. The first variable of each set (i.e., y_1 for factor 1 and y_7 for factor 2) is always observed and determines the probability of missingness for the remaining variables as defined in equations (17) and (18) for binary and ordinal variables, respectively. The factor loadings for models 4 and 5 range from .4 to .9 for each factor, with .4 being given to the loadings of y_1 and y_7, which determine the missingness. In models 4 and 5, the factor variances are fixed to 1 (i.e.$\phi_{11}=\phi_{22}=1$, to define the units of the factor scales), and the factor correlation, ϕ_{12}, is .3 for model 4 and .6 for model 5. In all five models, Θ_ε is a diagonal matrix with $\Theta_\varepsilon = I - \mathrm{diag}(\Lambda\Phi\Lambda')$, where **I** is the identity matrix. In each of the ten experimental conditions, we conduct 1,000 replications.

All binary variables have thresholds equal to .5. Missing data are generated for variables $y_2,\ldots,y_6$ using the following mechanism:

$$\begin{aligned}\Pr(y_i\text{mis.}|y_1=0)&=\exp(-2)/(1+\exp(-2))=.119,\\ \Pr(y_i\text{mis.}|y_1=1)&=\exp(+1)/(1+\exp(+1))=.731,\end{aligned}\tag{17}$$

where $i=2,\ldots,6$ which gives about 29% missing data in each of $y_2,\ldots,y_6$. All ordinal variables have four response categories and the thresholds are set to −1.5, .5 and 1.5. Missing data are generated for variables $y_2,\ldots,y_6$ using the following mechanism:

$$\begin{aligned}\Pr(y_i\text{mis.}|y_1=1)&=\exp(-2)/(1+\exp(-2))=.119,\\ \Pr(y_i\text{mis.}|y_1=2)&=\exp(-1)/(1+\exp(-1))=.269,\\ \Pr(y_i\text{mis.}|y_1=3)&=\exp(-1)/(1+\exp(-1))=.269,\\ \Pr(y_i\text{mis.}|y_1=4)&=\exp(+1)/(1+\exp(+1))-.731,\end{aligned}\tag{18}$$

where $i=2,\ldots,6$ which gives about 31% missing data in each of $y_2,\ldots,y_6$. This procedure generates MAR missingness (Asparouhov & Muthén, 2010). The missing-data mechanism for variables $y_8,\ldots,y_{12}$ is exactly the same, except that the probabilities in (17) and (18) are conditional on the value of y_7, the variable that is always observed for this set.

A total of $R\ =\ 1{,}000$ replications were run for part I of the simulation study. However, not all of these replications yielded a solution. We define convergence/completion rates as the percentage of replications for each condition that converged, excluding improper solutions (i.e., estimation was completed). A solution was also defined as improper when at least one estimated parameter was outside of expected range (i.e., error variances were non-negative and factor correlations were $<\ 1$ in absolute value). The convergence and proper solution percentages as well as the number of replications for which all methods provide a proper solution denoted by R' are given in Tables S1–S4. The overall convergence rates for all methods studied in the paper are between 99.6 and 100% except for models 2, 4 and 5 for binary data and model 4 for ordinal data and for sample size 300,

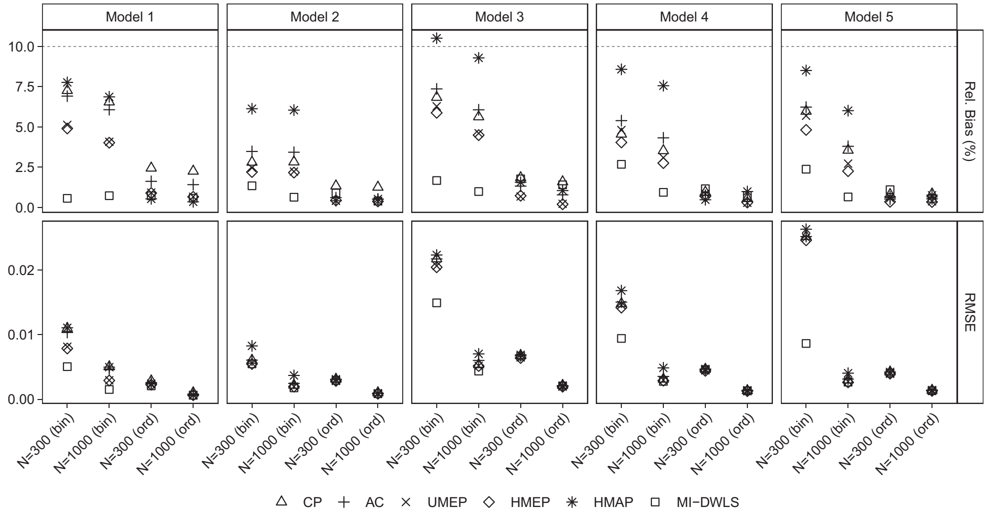

Figure 1. Absolute relative bias (top panel) and RMSE (bottom panel) of estimated factor loadings averaged over all variables and all factors when applicable, for all experimental conditions, N denotes the sample size.

which are 96.4, 98.8, 98.0 and 98.8%, respectively. Models 1–3 and all methods gave 100% proper solutions when the sample size was 1,000 for binary data and 300 or 1,000 for ordinal data. The same is true for models 4 and 5 for ordinal data and sample size 1,000. In most other cases the percentage of proper solutions ranges from 80.4 to 99.6%, where 100% is achieved only for MI-DWLS under model 5, ordinal data and sample size 300. In summary, binary data and smaller sample size exhibited larger percentages of improper solutions. We conclude that convergence rates and proper solutions obtained by the proposed method are satisfactory and do not raise any concerns. Non-convergent and improper solutions were removed from the analysis.

Figure 1 shows that all methods exhibit acceptable relative bias (<10%) in the estimated factor loadings in all conditions except for HMAP under model 3 for binary data and $N = 300$. Among the proposed methods and for the binary case, HMAP exhibits the largest bias for all models and MI-DWLS the smallest. In the ordinal case, all methods perform similarly and show low bias. For all proposed PL methods, the relative bias for the loading estimates tends to decrease as the loading value of the indicator which determines the missingness mechanism decreases (model 2 compared to models 1 and 3). Overall, larger but still acceptable relative biases are found in the binary models. The value of the factor correlation (model 5 compared to model 4) seems not to have much of an effect on the bias of the factor loadings. The bottom panel of Figure 1 shows that the average RMSE is smaller for the MI-DWLS for the binary case and $N = 300$ but similar across methods in all other conditions and decreases with the sample size. A higher factor correlation seems to have no impact on the average RMSE of estimated loadings.

The results from Figure 1 indicate that UMEP and HMEP do not perform clearly better than the CP and AC estimators in any of the experimental conditions. Therefore, we do not proceed with computing their standard errors and do not report results for them in Figure 2.

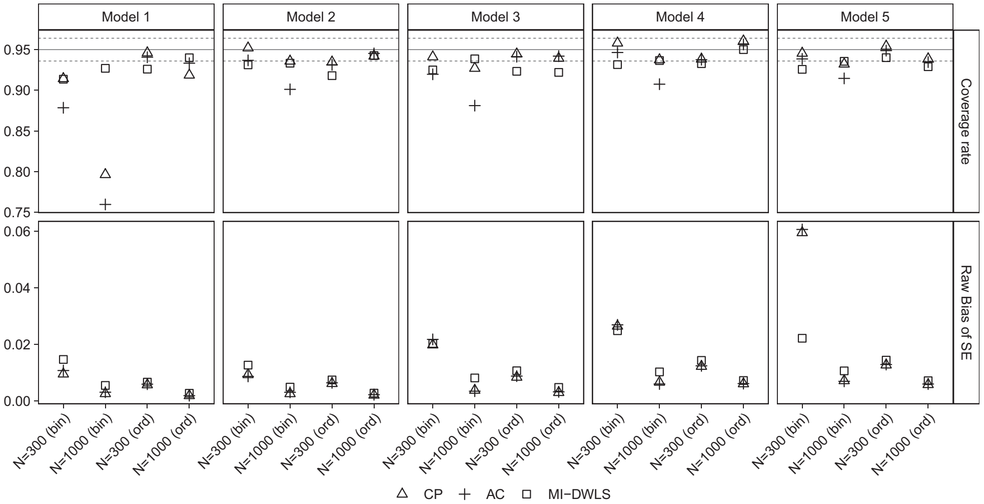

Figure 2. Coverage rate of 95% confidence intervals (top panel) and absolute raw bias of standard errors (bottom panel) for estimated factor loadings averaged over all variables and all factors when applicable, all experimental conditions, N denotes the sample size; a point, in the top panel graph, lying within the grey horizontal dashed lines, drawn at values 0.964 and 0.936, yields a 95% CI that includes the value 0.95.

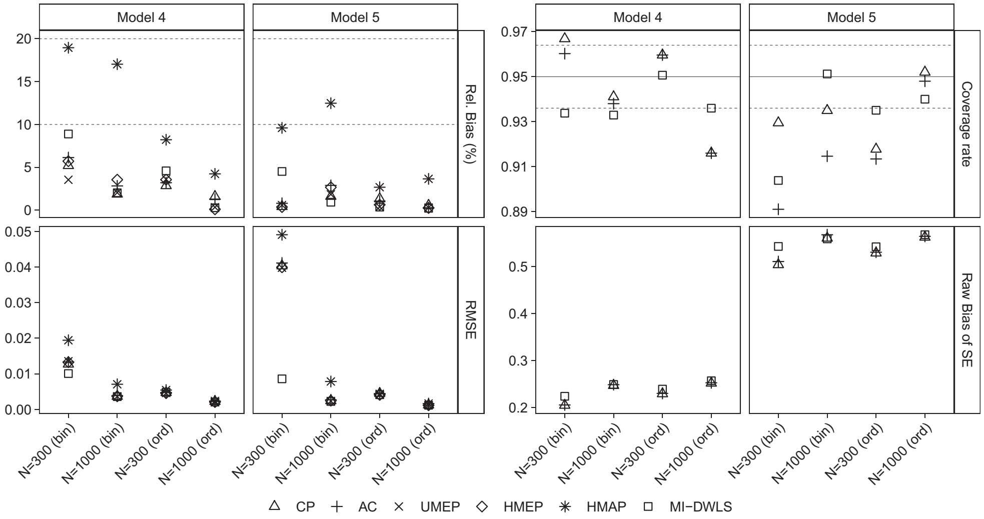

Figure 3. Absolute relative bias (top left panel), RMSE (bottom left panel), coverage rate of 95% confidence intervals (top right panel), and absolute raw bias of standard errors (bottom right panel) for estimated factor correlations for all experimental conditions (*N* denotes the sample size); a point, in the top panel graph, lying within the grey horizontal dashed lines, drawn at values 0.964 and 0.936, yields a 95% CI that includes the value 0.95.

The top panel of Figure 2 displays the average coverage rate of 95% CIs for factor loadings. The bold horizontal line is drawn at .95 and the thin horizontal lines are drawn at .964 and .936, respectively. A rate lying within the thin horizontal lines yields a 95% CI that includes .95.

The proposed estimators for model 1, binary data and $N = 1{,}000$ exhibit unexpected and poor coverage rate performance. The AC estimator also exhibits lower than expected coverage in the binary models 1–3 for $N = 300$ and 1,000. For the ordinal case, all models exhibit acceptable coverage rates in both sample sizes.

The bottom panel of Figure 2 shows that CP and AC have very similar average absolute raw bias for the estimated standard errors of the loadings. The average bias decreases with the sample size increase.

The results for the factor correlation in models 4 and 5 are presented in Figure 3. The top left panel shows that the proposed PL estimators (except HMAP) all have acceptable relative bias in both models, for binary and ordinal data, which decreases with a sample size increase. The RMSE, depicted in the bottom left panel, is very similar for all methods except for model 5, binary data and $N = 300$, for which all proposed methods have larger RMSE than MI-DWLS. It is again the case that UMEP and HMEP do not perform significantly better than CP and AC in terms of relative bias and RMSE, and thus we omit them in the comparisons of coverage rate and bias of standard error. The coverage rates are better for model 4 (lower factor correlation) and better in model 5 for the larger sample size but quite unsatisfactory for $N = 300$ (top right panel). Finally, the average absolute raw bias for the standard errors is similar across the estimators under each condition (bottom right panel).

Figure 4 displays the results for the thresholds. The CP estimator exhibits unacceptable relative bias and has been excluded from the top panel of the figure so as not to distort

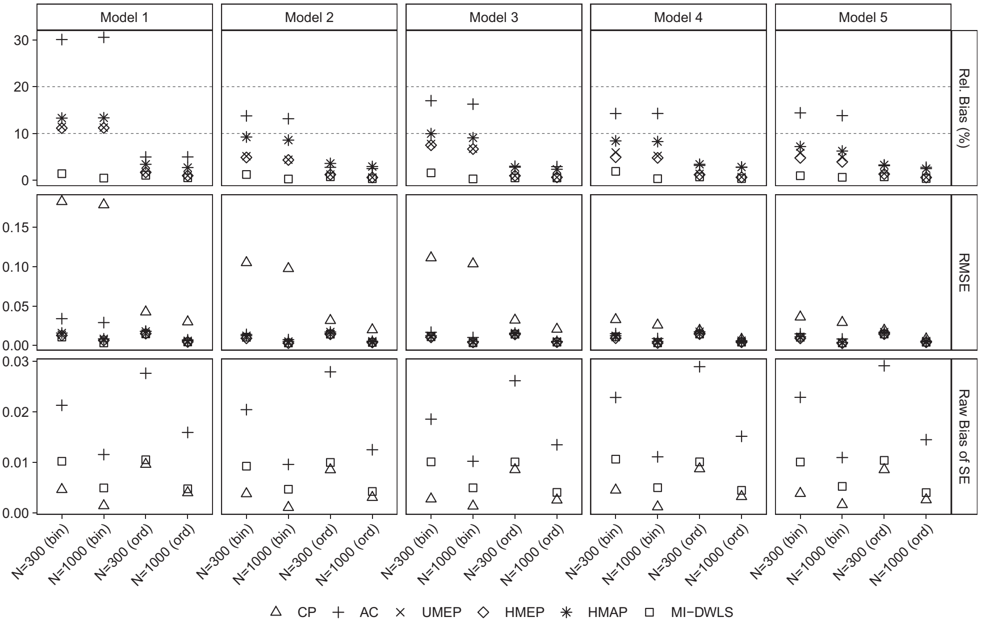

Figure 4. Absolute relative bias (top panel), RMSE (middle panel), and absolute raw bias of standard errors (bottom panel) of estimated thresholds averaged over all variables and all factors when applicable, for all experimental conditions, where N denotes the sample size.

the remaining results. However, we provide the figure with CP included in the Supporting Information (Figure S5). As seen in the top panel of Figure 4, AC exhibits substantial relative bias in the binary case but acceptable levels in the ordinal case for $N = 300$ and 1,000. UMEP and HMEP also exhibit acceptable and lower relative bias than AC in all cases except for model 1 and for both sample sizes. The average RMSE, in the middle panel of Figure 4, is nearly identical for all proposed PL methods and MI-DWLS except for AC, which in the cases of binary data and both sample sizes is systematically larger. In both the binary and ordinal case for $N = 300$, UMEP and HMEP clearly outperform CP and AC, but we are not convinced that this is a strong argument for preferring the more complicated DR methods even when it comes to binary data and the estimation of factor loadings and factor correlations. The simulation results indicate that the quality of threshold estimation does not seem to affect the quality of the estimation of loadings and factor correlation, which are typically the parameters of interest. Therefore, we do not compute the threshold standard errors for UMEP and HMEP and no results for them are presented in the bottom panel of Figure 4. AC has systematically larger average absolute raw biases in the standard errors in all conditions. An explanation could be that the information in the univariate likelihood functions is repeated in AC's objective function. The CP bias is not much larger than zero in absolute value, but that of MI-DWLS is smaller in all conditions. For all methods, the bias decreases as the sample size increases. Finally, Figures S1–S4 provide the raw bias for each model parameter (factor loadings, factor correlations and thresholds) under all simulation conditions.

Table 2. Results averaged over the same type of parameters (loadings, factor correlations, thresholds) for all performance criteria (absolute raw bias, RMSE, coverage rate of 95% CIs) for CP, AC, and MI-DWLS estimation methods for the four-factor model with 20 variables (sample size 1,000)

	Absolute raw bias	RMSE	Coverage rate	Absolute raw bias of *SE*
Loadings				
CP	.0034	.0019	.9506	.0030
AC	.0032	.0019	.9471	.0030
MI-DWLS	.0019	.0020	.9373	.0053
Factor correlations				
CP	.0022	.0024	.9453	.4041
AC	.0026	.0024	.9437	.4045
MI-DWLS	.0035	.0025	.9363	.4053
Thresholds				
CP	.0334	.0052	.9158	.0025
AC	.0188	.0044	.8569	.0132
MI-DWLS	.0029	.0041	.9468	.0050

4.4. Part II of the simulation study

The results of part I indicate that, for loadings and factor correlation estimates, CP and AC perform nearly the same as UMEP and HMEP with respect to the average absolute relative bias and RMSE, and exhibit acceptable performance in average coverage rate of 95% CIs and in average bias of standard errors. As long as thresholds are not parameters of interest, these findings, along with the merits of CP and AC (they require only a model for the observed data and the estimation is done in one step), render them preferable to UMEP and HMEP. The latter require the computation of the probabilities $\Pr\left(y_i^{\mathrm{m}} = a, y_j^{\mathrm{m}} = b|\mathbf{y}^{\circ}\right)$ and $\Pr\left(y_j^{\mathrm{m}} = b|\mathbf{y}^{\circ}\right)$, which is infeasible when p is large. For example, the R package *mnormt*, which is used in *lavaan* internally for the calculations, computes up to 30-dimensional normal probabilities. Thus, part II aims to study the performance of CP and AC in a larger model that could be encountered in practice.

The model includes 20 ordinal indicators, $y_1, \ldots, y_{20}$, and four factors, $\eta_1, \ldots, \eta_4$, each measured by a distinct set of five variables: $(y_1, \ldots, y_5)$ measure η_1, $(y_6, \ldots, y_{10})$ measure η_2, and so on. The loadings of each set of five variables are .6, .6, .6, .6, .6. The factor correlations are set to .3, .3, .3, .6, .6, .6. All factor variances are fixed to 1. All 20 indicators have four response categories and their threshold values are −1.25, .5 and 1.25. The variables y_1, y_6, y_{11}, y_{16} are always observed, and their values determine the probability of missingness of the remaining variables measuring the same factor (i.e., y_1 determines the probability of missingness for $y_2, \ldots, y_5, y_6$ determines the probability of missingness for $y_7, \ldots, y_{10}$, etc.) following the model in (18) with one exception: the probability of missingness conditional on the always-observed item scoring a ‘4’ is now set to .5 and the rest remain the same as in (18). The reasoning here is to induce slightly less severe missingness for the data compared to simulation part I. Indeed, the missing proportion is now 27% on average for each variable (as compared to 31% previously). The sample size is taken to be 1,000 and 50,000, and 250 replications are conducted within each sample size. For sample size 1,000, we give the average absolute raw biases, RMSE, and coverage rates for all parameter estimates. For sample size 50,000, we only study the raw bias of the parameter estimates in order to get an idea of the asymptotic behaviour of CP and AC.

Table 2 reports the results for sample size 1,000 which are similar to those of part I. CP and AC exhibit low raw biases for the estimates of factor loadings and factor correlations. Compared to MI-DWLS, the CP and AC average absolute raw bias is larger for the loadings and the thresholds but smaller for the factor correlations. The biases for the thresholds are substantially larger. The average RMSE of loadings and factor correlations is nearly the same for CP, AC and MI-DWLS. The average coverage rate of the 95% CIs for the loadings and factor correlations is very close to .95 for all methods but low for the thresholds under CP and AC. CP, AC and MI-DWLS exhibit very similar and low average absolute raw bias for the estimated standard errors of the loadings. The estimated standard errors for the factor correlations exhibit quite large biases with all methods.

For sample size 50,000, we compute the raw bias for each parameter estimate, $\bar{\hat{\theta}}_k - \theta$, for k = CP and AC, and compare it with the bias when the sample size is 1,000. The results are displayed in the Figure S6 for the loadings and the factor correlations, and in Figure S7 for the thresholds. For all three methods the bias of all loading and factor correlation estimates goes closer to zero when the sample size increases to 50,000. However, both CP and AC overestimate the thresholds, with AC exhibiting systematically smaller bias than CP.

5. A study on adult numeracy and literary in the UK

In large-scale assessments, adaptive testing has been implemented in the Organisation for Economic Co-operation and Development (OECD) Survey of Adult Skills developed by the PIAAC. The PIAAC data and the related documentation are publicly available on the OECD website. Here, we analyse a part of the UK data (collected in Round 1, from 2011 to 2012). The grouping variable we use, native versus non-native speakers, is included in the PIAAC data (labelled as 'NATIVESPEAKER') and defined as 'the respondent was considered a native speaker if his or her first language was one of the assessment languages' (OECD, 2016). For meaningful results and a fair comparison between native and non-native speakers, we restrict our analysis to those respondents who took the computer-based assessment after having successfully passed a short test on information and communication technology and reported a 'high' education level according to the background demographic questions (the levels of education are low, medium and high). A respondent who passed the core stage 1 test was administered a core stage 2 test (containing six cognitive items). We select those with scores above 3 (scores range from 0 to 6) who were routed to numeracy and literacy tests. The code to replicate the analysis is given in the Supporting Information.

The latent variables numeracy and literacy are each measured by 18 binary variables, but, due to the adaptive testing design, only nine are administered to a respondent. More specifically, the 18 variables are divided into three testlets of nine questions each. The testlets vary in difficulty, with testlet 1 being the easiest and testlet 3 the most difficult. Testlets 1 and 2 have five common questions and testlets 2 and 3 have four common questions, while testlets 1 and 3 have no common questions.

All three testlets have a positive probability of being administered which depends on the respondent's core stage 2 test score. The higher the score, the higher the probability that the more difficult testlet will be administered (see OECD, 2016, Chapter 1). Thus, there are missing values for test items not having been administered and the missingness is at random by design.

To ensure that the missingness is at random, we excluded the respondents who actively skipped administered questions or did not reach some of the questions because they ran out of time. The rate of this kind of item non-response ranges from .07% to 3.7% and the cases with at least one item skipped or not reached represent 18.5% of the initial sample. Tables S5 and S6 provide information on all types of missingness in the data (planned missingness, no response, not reached/not attempted) as well as the percentages of incorrect and correct response for each item. Keeping only the respondents for which both numeracy and literacy are measured and also eliminating the respondents who had either no responses or not reached/not attempted entries, the size of the analysed data is 1,170.

To estimate the correlation between numeracy and literacy, we fitted a two-factor model as defined in (1) and (2), where, $\boldsymbol{\eta}$ is a two-dimensional vector and Λ is of dimension 36×2. In Λ, the first 18 loadings of the first column and the last 18 loadings of the second column are free to be estimated, while the remaining loadings are fixed to 0. Φ is a 2×2 matrix with 1s on the main diagonal to define the scale of the latent variables, while ϕ_{12} is free to be estimated. Θ_ε satisfies the equation $\Theta_\varepsilon = I - \mathrm{diag}(\Lambda\Phi\Lambda\prime)$. The vector $\boldsymbol{\tau}$, in this example, includes 36 thresholds free to be estimated.

Among the proposed PL methods, we consider only the CP and the AC, and compare them with the MI-DWLS. Although the HMEP was attempted, it was found to be computationally infeasible. This is because for each respondent, 36 eleven-dimensional integrals over an eleven-variate normal distribution need to be computed and this needs to be repeated for 1,170 respondents. The use of the UMEP is not possible for PIAAC data as there are no data at all for some pairs of indicators (testlets 1 and 3 have no common questions). For the same reason, the variance–covariance model cannot be used as an imputation model in MI-DWLS, and instead we use the two-factor model assumed for the observed data. Ten imputed data sets are analysed. All three methods are quite fast. It took half a minute to produce the output for a single core an Intel Core i7 processor running at 2.20 GHz. Figure S8 displays the estimated parameters and their corresponding estimated standard errors. All three methods suggest that there is a fairly high correlation between literacy and numeracy for the subset of respondents analysed. CP and AC estimate the factor correlation to be .82 and MI-DWLS to be .77. The MI-DWLS estimated standard error for the factor correlation is only slightly larger than that of CP and AC, both of which yield almost identical standard errors. Regarding the loadings and the thresholds, CP and AC produce almost identical estimates, while the MI-DWLS ones are slightly larger. The CP and AC standard errors for the loadings are also nearly identical. For the thresholds, the AC standard errors are smaller than the CP ones. The MI-DWLS standard errors are larger than those of CP and AC for both the loadings and thresholds.

To test whether native and non-native English-speakers differ in numeracy and literacy, we carry out a two-group factor analysis. The CFA model defined in (1) and (2) can be extended to a two-group factor analysis model by adding a superscript g to all variables and parameters, with g denoting the group membership. Here, let $g = 1$ for the native speakers and $g = 2$ for the non-native speakers. Since the test items are meant to measure the factors in exactly the same way for both groups, we assume measurement equivalence, that is, $\Lambda^{(1)} = \Lambda^{(2)}$ and $\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$. The matrices $\Theta_\varepsilon^{(g)}$ should satisfy the equation $\Theta_\varepsilon^{(g)} = \mathrm{diag}\left(P^{(g)} - \Lambda\Phi^{(g)}\Lambda\prime\right)$, $g = 1,\ 2$. The model for the latent variables is modified to:

Table 3. CP and AC estimates and standard errors for factor means, variances and covariances as well as *p*-values for factor means for the two-group, two-factor model fitted to the UK data. Subscripts 1 and 2 denote literacy and numeracy, respectively; superscripts 1 and 2 in parentheses denote the native speakers group and the non-native speakers group, respectively

	Estimate		Standard error		*p*-value	
Parameter	CP	AC	CP	AC	CP	AC
$\phi_{21}^{(1)}$	.799	.797	.045	.045		
$\alpha_1^{(2)}$	.805	.834	.455	.182	.077	<.001
$\alpha_2^{(2)}$	.428	.447	.807	.445	.596	.316
$\phi_{11}^{(2)}$	.447	.365	.999	.526		
$\phi_{22}^{(2)}$	.872	.802	1.475	1.033		
$\phi_{21}^{(2)}$	.609	.532	.889	.541		

$$\begin{pmatrix}\eta_1^{(1)}\\ \eta_2^{(1)}\end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \\ \phi_{21}^{(1)} & 1\end{pmatrix}\right).$$

and

$$\begin{pmatrix}\eta_1^{(2)}\\ \eta_2^{(2)}\end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix}\alpha_1^{(2)}\\ \alpha_2^{(2)}\end{pmatrix},\begin{pmatrix}\phi_{11}^{(2)} & \\ \phi_{21}^{(2)} & \phi_{22}^{(2)}\end{pmatrix}\right),$$

where $\eta_1^{(g)}$ and $\eta_2^{(g)}$ denote literacy and numeracy in group g respectively, and $\alpha_1^{(2)}$ and $\alpha_2^{(2)}$ are the means of the factors in group 2. The means and variances of the factors in group 1 are fixed to 0 and 1, respectively, to define their scales.

For the two-group model, the CP and AC log-likelihood functions are

$$pl^{\text{CP}}\left(\boldsymbol{\theta};\left(\mathbf{y}_1^{(1)},\ldots\mathbf{y}_{N_1}^{(1)},\mathbf{y}_1^{(2)},\ldots\mathbf{y}_{N_2}^{(2)}\right)\right)=\sum_{g=1}^{2}\sum_{n=1}^{N_g} pl_n^{\text{CP}}\left(\boldsymbol{\theta};\mathbf{y}_n^{(g)}\right),$$

$$pl^{\text{AC}}\left(\boldsymbol{\theta};\left(\mathbf{y}_1^{(1)},\ldots\mathbf{y}_{N_1}^{(1)},\mathbf{y}_1^{(2)},\ldots\mathbf{y}_{N_2}^{(2)}\right)\right)=\sum_{g=1}^{2}\sum_{n=1}^{N_g} pl_n^{\text{AC}}\left(\boldsymbol{\theta};\mathbf{y}_n^{(g)}\right),$$

where $pl_n^{\text{CP}}\left(\boldsymbol{\theta};\mathbf{y}_n^{(g)}\right)$ and $pl_n^{\text{AC}}\left(\boldsymbol{\theta};\mathbf{y}_n^{(g)}\right)$ are defined in (9) and (10) respectively, with the difference of adding a superscript g to the y. The sample size of native speakers is $N_1 = 1{,}078$ and that of non-native speakers is $N_2 = 92$.

MI-DWLS in multi-group analysis is complicated. Different strategies have been suggested in the literature (e.g., product term imputation, separate group imputation) so that possible interactive effects between the grouping variable and the remaining variables will be preserved during the imputation, and in this way bias in the parameter estimates is avoided. Mplus version 7.11 does not offer the option of multiple imputation in the case of multi-group analysis within the CFA framework with ordinal indicators. For this, we fitted the model using only CP and AC.

Both CP and AC are fast to implement; it takes each of them $2\frac{1}{2}$ minutes to fit the two-group two-factor model. The estimated loadings and thresholds (which are nearly identical between the methods) and their standard errors (which are very similar for the loadings between the methods but for the thresholds; the AC ones are systematically smaller) are displayed in Figure S9. The estimates and standard errors for the factor means, variances and covariances, which are the parameters of interest, are reported in Table 3. The qualitative results are similar for both methods and need to be interpreted with caution because the size of the non-native speakers group is very small, 92; it is less than one-tenth of the size of the native speakers group, which is 1,078. The estimates of the factor means indicate that non-native speakers have, on average, lower levels of literacy and numeracy than native speakers, but the difference is rather small, less than one standard deviation. (Note that, for all binary indicators, the correct answer is coded with 1 and an incorrect answer with 2; since the loading estimates are positive, higher values of the factors denote lower level of the skills they represent.) Only the mean of literacy, $\alpha_1^{(2)}$, is found to be statistically significant in the case of AC only (p-value $<.001$). In both CP and AC, non-native speakers exhibit smaller variances for both literacy and numeracy and higher correlation between literacy and numeracy (approximately .98) than native speakers.

6. Conclusions and discussion

We develop and study the performance of complete-pairs (CP), available-cases (AC) and three variants of the doubly robust (DR) pairwise likelihood (PL) estimators for confirmatory factor analysis (CFA) models with binary and ordinal variables and missing at random (MAR) data. Our simulation results indicate that the general result, that CP and AC yield biased estimators because they ignore the missing-data mechanism, does not necessarily apply to CFA. CP and AC are found to have acceptable relative bias and close to zero raw bias for loading and factor correlation estimates, decreasing with a sample size increase. Overall, larger but still acceptable relative biases were found in the binary models compared to the ordinal ones. The CP and AC coverage rates of 95% CIs for loadings and factor correlations are satisfactory and improve with a sample size increase. The only systematic difference between CP and AC in their performance lies in the estimation of thresholds. AC yields threshold estimates that have unacceptable relative bias in the case of binary data but acceptable for ordinal data, and CP has unacceptable bias in all cases. We also study three variants of the DR PL estimator which we term the unconstrained model exact probability (UMEP), hypothesized model exact probability (HMEP), and hypothesized model approximate probability (HMAP) approaches. In the binary model HMAP often exhibits an unacceptable performance, while the performance of UMEP and HMEP in relative and raw bias for loadings and factor correlations is nearly identical to that of CP and AC. UMEP and HMEP outperform CP and AC only in the estimation of thresholds. However, as long as thresholds are not parameters of interest, CP and AC are preferred because they require only a model for the observed data, the estimation is done in one step, and the estimation of standard errors is straightforward. UMEP and HMEP are computationally intensive and become infeasible for a large number of variables (e.g., beyond 30).

Both CP and AC exhibit competitive performance compared to MI-DWLS except for the estimation of thresholds. An advantage of CP and AC over MI-DWLS is that they only require a model for the observed data. Moreover, the extension of CP and AC to the analysis of multi-group data is straightforward, while multiple imputation should be conducted with caution so that the data imputation will not distort possible interaction effects between the grouping variable and the remaining variables.

The results of the paper are applicable to exploratory factor analysis and structural equation modelling (SEM) with categorical variables. As the general formulation of the model is the same, the form of the objective function of the proposed methods remains unchanged. The CP and AC estimators can also be extended to factor analysis and SEM with mixed variables (categorical and continuous). What is needed in addition is to specify the bivariate log-likelihood functions for the pairs of continuous variables and the pairs of one categorical and one continuous variable. Finally, some first simulation results indicate that CP and AC may perform satisfactorily within CFA models with missing not at random data, which could be a topic for future research.

Finally, we should note that the estimators proposed here assume that the model is correctly specified. Lindsay et al. (2011) and Yi and Reid (2010) mention that composite likelihood estimators might exhibit more robustness compared to traditional likelihood methods since they require correct model specification in the lower order margins than in the full pattern. However, there is also more recent work that indicates that this might not be the case in all models and, in particular, Ogden (2016) studied the case of misspecifying the random effect distribution in generalized mixed effect models. Factor analysis models also depend on correctly specifying the distribution of the latent variables and therefore further investigation is needed to address the robustness of the pairwise likelihood estimator under model misspecification and missing values.

Acknowledgements

The first author was funded by the Economic and Social Research Council under the Future Research Leaders Scheme grant ES/L009838/1.

Conflicts of interest

All authors declare no conflict of interest.

Data Availability Statement

The data are publicly available on the OECD's website.

References

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, *2010*, 1–10.

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, *61*, 962–972.

Birhanu, T. (2012). *Pseudo-likelihood and estimating equation methodology for incomplete data*. Ph. D. thesis, Universiteit Hasselt, Belgium.

De Leon, A. R. (2005). Pairwise likelihood approach to grouped continuous model and its extension. *Statistics and Probability Letters*, *75*, 49–57. https://doi.org/10.1016/j.spl.2005.05.017

Fonseca, G., & Grassetti, L. (2010). Pairwise likelihood for missing data treatment in VAR models. Technical report, Department of Statistics, University of Udine, Italy.

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, *14*, 275–299. https://doi.org/10.1037/a0015825

He, W., & Yi, G. (2011). A pairwise likelihood method for correlated binary data with/without missing observations under generalized partially linear single-index models. *Statistica Sinica*, *21*, 207–229.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389. https://doi.org/10.1037/a0015825

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347–387. https://doi.org/10.1207/S15327906347-387

Katsikatsou, M. (2013). *Composite likelihood estimation for latent variable models with ordinal and continuous or ranking variables*. Ph. D. thesis, Uppsala Universitet.

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics and Data Analysis*, *56*, 4243–4258. https://doi.org/10.1207/S15327906347-387

Kenward, M. G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, *13*(3), 236–247. https://doi.org/10.1214/ss/1028905886

Lee, S., Poon, W., & Bentler, P. (1990). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, *9*, 91–97. https://doi.org/10.1016/0167-7152(90)90100-L

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.

Lindsay, B. G., Yi, G. Y., & Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, *21*, 71–105.

Liu, J. (2007). *Multivariate ordinal data analysis with pairwise likelihood and its extension to SEM*. Ph. D. thesis, University of California, Los Angeles.

Molenberghs, G., Kenward, M., Verbeke, G., & Birhanu, T. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica*, *21*, 187–206.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variables indicators. *Psychometrika*, *49*, 115–132.

OECD. (2016). *Technical report of the Survey of Adult Skills (PIAAC)*. Technical report, Organisation for Economic Co-operation and Development.

Ogden, H. E. (2016). A caveat on the robustness of composite likelihood estimators: The case of a misspecified random effect distribution. *Statistica Sinica*, *26*, 639–651.

Poon, W. Y., & Lee, S. Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika*, *52*, 409–430.

Robins, J. M., & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In N. P. Jewell, K. Dietz & V. Farewell (Eds.), *AIDS epidemiology: Methodological issues*. Boston, MA: Birkhauser.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some of the regressors are not always observed. *Journal of the American Statistical Association*, *89*, 846–866.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semi-parametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, *90*, 106–121. https://doi.org/10.1080/01621459.1995.10476493

Rotnitzky, A. (2009). Inverse probability weighted methods. In G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 453–476). Boca Raton, FL: CRC/Chapman & Hall.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. In I*mputation and editing of faulty or missing survey data* (pp. 1–23). Washington, DC: US Department of Commerce.

Schafer, J. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall/CRC.

Scharfstein, D. O., Rotnitzky, A., & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, *94*, 1096–1120.with Rejoinder, 1135-1146. https://doi.org/10.1080/01621459.1999.10473862

Xi, N. (2011). *A Composite likelihood approach for factor analyzing ordinal data*. Ph. D. thesis, The Ohio State University.

Yi, G. Y., & Reid, N. (2010). A note on mis-specified estimating functions. *Statistica Sinica*, *20*, 1749–1769.

Yi, G., Zeng, L., & Cook, R. (2011). A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *The Canadian Journal of Statistics*, *39*, 34–51. https://doi.org/10.1002/cjs.10089

Received 30 July 2019; revised version received 10 February 2021

Supporting Information

The following supporting information may be found in the online edition of the article:

Figure S1. Raw bias of individual parameter estimates (factor loadings, factor correlations and thresholds) for binary data and sample size $N = 300$ shown for all proposed estimation methods across the 5 data generating models.
Figure S2. Raw bias of individual parameter estimates (factor loadings, factor correlations and thresholds) for binary data and sample size $N = 1{,}000$ shown for all proposed estimation methods across the 5 data generating models.
Figure S3. Raw bias of individual parameter estimates (factor loadings, factor correlations and thresholds) for ordinal data and sample size $N = 300$ shown for all proposed estimation methods across the 5 data generating models.
Figure S4. Raw bias of individual parameter estimates (factor loadings, factor correlations and thresholds) for ordinal data and sample size $N = 1{,}000$ shown for all proposed estimation methods across the 5 data generating models.
Figure S5. Relative absolute bias (top panel), RMSE (middle panel), and bias of standard errors (bottom panel) of estimated thresholds with CP method included averaged over all variables and all factors when applicable, for all experimental conditions, where N denotes the sample size.
Figure S6. Raw bias of factor loading (parameter index 1-20) and factor correlation estimates (parameter index 21-26) for CP and AC for the four-factor model with 20 variables and sample sizes 1,000 and 50,000.
Figure S7. Raw bias of threshold estimates (parameter index 27-86) for CP and AC for the four-factor model with 20 variables and sample sizes 1,000 and 50,000.
Figure S8. CP, AC, and MI-DWLS parameter estimates and standard errors for the single-group two-factor model fitted to the UK data; the vertical lines separate parameters of different types; from left to right: loadings, factor correlation, thresholds.
Figure S9. CP and AC loading and threshold estimates and standard errors for the two-group two-factor model fitted to the UK data; the vertical line separates loadings in the left panel from thresholds in the right panel
Table S1. Percentage of overall completions by simulation and proper solutions by method and simulation, Binary data, sample size 300.
Table S2. Percentage of overall completions by simulation and proper solutions by method and simulation, Binary data, sample size 1,000.
Table S3. Percentage of overall completions by simulation and proper solutions by method and simulation, Ordinal data, sample size 300.
Table S4. Table Percentage of overall completions by simulation and proper solutions by method and simulation, Ordinal data, sample size 1,000.
Table S5. Literacy items: Percentages of planned missing data, ‘no response’, ‘not reached /not attempted’, incorrect and correct responses. Subjects with missing data to either all literacy or all numeracy or both sets of items have been excluded.
Table S6. Numeracy items: Percentages of planned missing data, ‘no response’, ‘not reached/not attempted’, incorrect and correct responses. Subjects with missing data to either all literacy or all numeracy or both sets of items have been excluded.