

## Article

# Estimation Methods of the Multiple-Group One-Dimensional Factor Model: Implied Identification Constraints in the Violation of Measurement Invariance

Alexander Robitzsch <sup>1,2</sup> <sup>1</sup> Department of Educational Measurement, IPN, Leibniz Institute for Science and Mathematics Education, 24118 Kiel, Germany; robitzsch@leibniz-ipn.de<sup>2</sup> Centre for International Student Assessment (ZIB), 24118 Kiel, Germany

**Abstract:** Factor analysis is one of the most important statistical tools for analyzing multivariate data (i.e., items) in the social sciences. An essential case is the comparison of multiple groups on a one-dimensional factor variable that can be interpreted as a summary of the items. The assumption of measurement invariance is a frequently employed assumption that enables the comparison of the factor variable across groups. This article discusses different estimation methods of the multiple-group one-dimensional factor model under violations of measurement invariance (i.e., measurement noninvariance). In detail, joint estimation, linking methods, and regularized estimation approaches are treated. It is argued that linking approaches and regularization approaches can be equivalent to joint estimation approaches if appropriate (robust) loss functions are employed. Each of the estimation approaches defines identification constraints of parameters that quantify violations of measurement invariance. We argue in the discussion section that the fitted multiple-group one-dimensional factor analysis will likely be misspecified due to the violation of measurement invariance. Hence, because there is always indeterminacy in determining group comparisons of the factor variable under noninvariance, the preference of particular fitting strategies such as partial invariance over alternatives is unjustified. In contrast, researchers purposely define fitting functions that minimize the extent of model misspecification due to the choice of a particular (robust) loss function.



**Citation:** Robitzsch, A. Estimation Methods of the Multiple-Group One-Dimensional Factor Model: Implied Identification Constraints in the Violation of Measurement Invariance. *Axioms* **2022**, *11*, 119. <https://doi.org/10.3390/axioms11030119>

Academic Editor: Hari Mohan Srivastava

Received: 20 February 2022

Accepted: 7 March 2022

Published: 9 March 2022

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



**Copyright:** © 2022 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

Factor analysis is one of the most important statistical tools for analyzing multivariate data (i.e., items) in the social sciences [1,2]. An important case is the comparison of multiple groups on a one-dimensional factor variable that can be interpreted as a summary of the multivariate input data. To enable a comparison on the factor variable, identification constraints for model estimation must be posed [3–5].

A popular and heavily discussed identification is the assumption of measurement invariance (MI, [6,7]) that assumes the existence of invariant (i.e., equal) item parameters across groups. Noninvariant item parameters occur if not all parameters are equal across groups. Practitioners and applied methodologists frequently claim that MI or only weak violations of MI (i.e., partial invariance) are necessary to enable group comparisons on the factor variable [8–10]. In this article, we discuss different estimation methods of the one-dimensional factor model and their implied identification constraints in the violation of MI. In more detail, we focus on joint estimation (maximum likelihood estimation), linking approaches (Haberman linking, invariance alignment), and regularized estimation (lasso-type regularization, fused regularization, Bayesian approximate invariance). We derive identification constraints on parameters that quantify violations of MI under different estimation methods. By doing so, it turns out that joint estimation, linking, and regularization can

be interpreted quite similarly under certain specifications. Therefore, this work discusses competing approaches for handling violations of measurement invariance in a unified framework and provides conditions for which these approaches can provide similar results. We also try to convince the reader that there cannot be a right choice of an approach on how to handle measurement invariance because an overidentified minimization problem is made identifiable to selecting a fitting function. We derive implied identification constraints by the different fitting functions and discuss the similarity in the sequel of this article.

The article is structured as follows: In Section 2, we discuss two important one-dimensional factor models and their estimation: the tau-equivalent and the tau-congeneric measurement model. In Section 3, the tau-equivalent model with noninvariant item intercepts. Section 4 treats the tau-congeneric model with noninvariant item intercepts and invariant item loadings, while Section 5 also allows noninvariant item loadings. Finally, Section 6 closes with a discussion.

## 2. One-Dimensional Factor Model

Assume that there are  $I$  random variables  $X_1, \dots, X_I$ . These variables are also referred to as items. Denote by  $\mathbf{X} = (X_1, \dots, X_I)$  the vector of all items. Denote by  $\boldsymbol{\mu} = E(\mathbf{X})$  the vector of means containing the entries  $E(X_i)$  and by  $\boldsymbol{\Sigma} = \text{Var}(\mathbf{X})$  the covariance matrix containing entries  $\sigma_{ij} = \text{Cov}(X_i, X_j)$  for  $i \neq j$ . In one-dimensional factor analysis, we represent the  $I$  items by a one-dimensional factor variable  $F$ . Hence, the covariances among items are presented by a rank-one matrix. In the following, we discuss two main measurement models of one-dimensional factor analysis: the tau-equivalent and the tau-congeneric models [11–14].

### 2.1. Tau-Equivalent Model

We now assume a one-dimensional factor  $F$  in the tau-equivalent model [14]:

$$X_i = \nu_i + F + \varepsilon_i, \quad \text{Var}(\varepsilon_i) = \phi,$$

where the residuals  $\varepsilon_i$  are uncorrelated. Note that it is assumed that there are equal loadings  $\lambda$  and equal residual variances  $\theta$ , while item intercepts  $\nu_i$  are item-specific parameters. For identification, we assume  $E(F) = 0$  and  $\psi = \text{Var}(F)$  is estimated. Denote by  $\mathbf{I}$  the  $I \times I$  identity matrix and by  $\mathbf{1}$  an  $I \times 1$  vector of ones. Then, the covariance matrix  $\boldsymbol{\Sigma}$  of the items  $\mathbf{X}$  is represented by a model-implied covariance matrix  $\boldsymbol{\Sigma}_0$

$$\boldsymbol{\Sigma}_0 = \psi \mathbf{1} \mathbf{1}^\top + \phi \mathbf{I}.$$

Note that the covariance matrix is parsimoniously represented by only two parameters. The mean vector  $\boldsymbol{\mu} = E(\mathbf{X}) = \boldsymbol{\nu}$  is estimated without constraints.

### 2.2. Tau-Congeneric Model

In the tau-congeneric measurement model [11], item-specific loadings and item-specific residual variances are allowed:

$$X_i = \nu_i + \lambda_i F + \varepsilon_i, \quad \text{Var}(\varepsilon_i) = \phi_i,$$

where residuals  $\varepsilon_i$  are uncorrelated across items. Denote by  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_I)$  the vector of item loadings  $\lambda_i$  and  $\boldsymbol{\Phi} = \text{diag}(\phi_1, \dots, \phi_I)$  the diagonal matrix containing residual variances  $\phi_i$ . For reasons of identification, we set  $E(F) = 0$  and  $\text{Var}(F) = 1$ . The covariance matrix  $\boldsymbol{\Sigma}$  is modelled as

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\lambda} \boldsymbol{\lambda}^\top + \boldsymbol{\Phi}.$$

### 2.3. Overview of Estimation Methods

The tau-equivalent and the tau-congeneric model are special cases of structural equation models that impose restrictions on the mean vector and the covariance matrix [15].

In maximum likelihood (ML) estimation assuming multivariate normality of  $X$ , the empirical mean vector  $\bar{x}$  and empirical covariance matrix  $S$  are sufficient statistics. We denote by  $\Theta$  the vector of all estimated parameters and define the fitting function

$$F_{\text{ML}}(\Theta; \bar{x}, S) = -\frac{N}{2} \left( -I \log(2\pi) + \log|\Sigma_0(\Theta)| + \text{tr}(S\Sigma_0(\Theta)^{-1}) + (\bar{x} - \mu_0(\Theta))^{\top} \Sigma_0(\Theta)^{-1} (\bar{x} - \mu_0(\Theta)) \right). \quad (1)$$

In this article, we are only concerned with the statistical behavior of parameter estimates in the population (i.e., infinite large sample sizes). Then, the sample quantities  $\bar{x}$  and  $S$  are replaced by population parameters  $\mu$  and  $\Sigma$ . The fitting function in Equation (1) can then be rewritten as

$$F_{\text{ML}}(\Theta; \mu, \Sigma) = -\log|\Sigma_0(\Theta)| - \text{tr}(\Sigma\Sigma_0(\Theta)^{-1}) - (\mu - \mu_0(\Theta))^{\top} \Sigma_0(\Theta)^{-1} (\mu - \mu_0(\Theta)).$$

In practice, the model-implied covariance matrix will be misspecified [16], and  $\Theta$  is a pseudo-true parameter that is defined as the maximizer of the fitting function  $F_{\text{ML}}$ .

A more general class of fitting functions is weighted least squares (WLS) estimation [15]. The parameter vector  $\Theta$  is determined as the minimizer of

$$F_{\text{WLS}}(\Theta; \mu, \sigma) = (\mu - \mu_0(\Theta))^{\top} W_1 (\mu - \mu_0(\Theta)) + (\sigma - \sigma_0(\Theta))^{\top} W_2 (\sigma - \sigma_0(\Theta)) \quad (2)$$

with known weight matrices  $W_1$  and  $W_2$ . The vectors  $\sigma$  and  $\sigma_0$  contain the nonduplicated elements from covariance matrices  $\Sigma$  and  $\Sigma_0(\Theta)$ . Diagonally weighted least squares (DWLS) estimation results by choosing diagonal weight matrices  $W_1$  and  $W_2$ . If these matrices are identity matrices, unweighted least squares (ULS) estimation is obtained. Interestingly, the minimization in (2) can be interpreted as a nonlinear least squares estimation problem with sufficient statistics  $\mu$  and  $\Sigma$  as input data [17].

It has been shown that ML estimation can be approximately written as DWLS estimation [18] with particular weight matrices. DWLS can be generally written as

$$F_{\text{DWLS}}(\Theta; \mu, \sigma) = \sum_{i=1}^I w_{1i} (\mu_i - \mu_{0,i}(\Theta))^2 + \sum_{i=1}^I \sum_{j=i}^I w_{2ij} (\sigma_{ij} - \sigma_{0,ij}(\Theta))^2, \quad (3)$$

where  $\mu_i$  etc. indicate the corresponding elements of vectors defined in (3). In ML estimation, the weights are approximately determined by  $w_{1i} = 1/u_i^2$  and  $w_{2ij} = 1/(u_i^2 u_j^2)$ , where  $u_i^2$  are sample unique standardized variances with  $u_i^2 = \phi_i/\sigma_{ii}$ . With smaller residual variances  $\phi_i$ , more trust is put on a mean  $\mu_i$  or a covariance  $\sigma_{ij}$  in the fitting function. This kind of weighting seems questionable in the case of misspecified models [18].

The model deviations  $\mu_i - \mu_{0,i}(\Theta)$  and  $\sigma_{ij} - \sigma_{0,ij}(\Theta)$  can be differently weighted by replacing the least squares functions with robust fitting functions  $\rho$  [19,20]:

$$F_{\text{rob}}(\Theta; \mu, \sigma) = \sum_{i=1}^I w_{1i} \rho(\mu_i - \mu_{0,i}(\Theta)) + \sum_{i=1}^I \sum_{j=i}^I w_{2ij} \rho(\sigma_{ij} - \sigma_{0,ij}(\Theta)). \quad (4)$$

Siemsen and Bollen [19] proposed the absolute value function  $\rho(x) = |x|$  for fitting the factor analysis model. This fitting function is robust to a few model violations such as unmodelled correlations of residuals  $\varepsilon_i$ . Alternative robust loss functions such as  $\rho(x) = |x|^p$  with  $0 < p < 1$  can ensure even more model-robust estimates [21].

#### 2.4. Estimation in the Presence of Slight Model Misspecifications

We now study the behavior of the estimate  $\Theta$  as the minimizer of  $F_{\text{rob}}$  in (4) (see [16,22]). A nondifferentiable loss function  $\rho$  is substituted by a differentiable approximation (e.g.,  $\rho(x) = |x|$  is replaced by  $\rho(x) = (x^2 + \varepsilon)^{1/2}$  for a small  $\varepsilon > 0$ ; see [21]) in the following derivation.

We investigate slight model misspecifications for the mean and the covariance structures. For the mean structure, we define residuals  $\gamma_i(\Theta) = \mu_i - \mu_{0,i}(\Theta)$  for  $i = 1, \dots, I$ .

Furthermore, we define  $\delta_{ij}(\theta) = \sigma_{ij} - \sigma_{0,ij}(\theta) \neq 0$  for model deviations in the covariance structure. The estimate  $\theta$  is obtained by taking the derivative of  $F_{\text{rob}}$  with respect to  $\theta$  and setting it to zero:

$$\mathbf{H}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \theta) = \sum_{i=1}^I w_{1i} \rho'(\mu_i - \mu_{0,i}(\theta)) \mathbf{a}_i(\theta) + \sum_{i=1}^I \sum_{j=i}^I w_{2ij} \rho'(\sigma_{ij} - \sigma_{0,ij}(\theta)) \mathbf{b}_{ij}(\theta) = \mathbf{0}, \quad (5)$$

where  $\mathbf{a}_i(\theta) = \frac{\partial \mu_{0,i}}{\partial \theta}$  and  $\mathbf{b}_{ij}(\theta) = \frac{\partial \sigma_{0,ij}}{\partial \theta}$ . A parameter estimate  $\theta$  is obtained by computing the root of  $\mathbf{H}$  in Equation (5). Note that  $\theta$  is a function of the mean vector  $\mu$  and the stacked covariance matrix  $\sigma$ .

Assume that there is a true parameter  $\theta_0$  if all model deviations  $\gamma_i$  and  $\delta_{ij}$  would be zero. That is, we assume

$$\mathbf{H}(\mu_0(\theta_0), \sigma_0(\theta_0), \theta_0) = \mathbf{0}. \quad (6)$$

Now, use the notation  $\boldsymbol{\gamma} = \boldsymbol{\mu} - \boldsymbol{\mu}_0(\theta_0)$  and  $\boldsymbol{\delta} = \boldsymbol{\sigma} - \boldsymbol{\sigma}_0(\theta_0)$ . A first-order Taylor expansion of  $\mathbf{H}$  using (6) provides

$$\mathbf{H}(\mu_0(\theta_0) + \boldsymbol{\gamma}, \sigma_0(\theta_0) + \boldsymbol{\delta}, \theta) \simeq \mathbf{H}_\mu(\mu_0(\theta_0))\boldsymbol{\gamma} + \mathbf{H}_\sigma(\sigma_0(\theta_0))\boldsymbol{\delta} + \mathbf{H}_\theta(\theta_0)(\theta - \theta_0) = \mathbf{0}, \quad (7)$$

where  $\mathbf{H}_\mu(\mu_0(\theta_0)) = \left( \frac{\partial \mathbf{H}}{\partial \boldsymbol{\mu}} \right) \Big|_{\boldsymbol{\mu}=\mu_0(\theta_0)}$ ,  $\mathbf{H}_\sigma(\sigma_0(\theta_0)) = \left( \frac{\partial \mathbf{H}}{\partial \boldsymbol{\sigma}} \right) \Big|_{\boldsymbol{\sigma}=\sigma_0(\theta_0)}$  and  $\mathbf{H}_\theta(\theta_0) = \left( \frac{\partial \mathbf{H}}{\partial \theta} \right) \Big|_{\theta=\theta_0}$ . Note that we suppress arguments in  $\mathbf{H}_\mu$ ,  $\mathbf{H}_\sigma$  and  $\mathbf{H}_\theta$  in our abbreviated notation. Using the approximation (7), we obtain

$$\theta = \theta_0 - \mathbf{H}_\theta^{-1}(\theta_0) \{ \mathbf{H}_\mu(\mu_0(\theta_0))\boldsymbol{\gamma} + \mathbf{H}_\sigma(\sigma_0(\theta_0))\boldsymbol{\delta} \}. \quad (8)$$

Model deviations  $\boldsymbol{\gamma}$  and  $\boldsymbol{\delta}$  enter the computation according to

$$\mathbf{H}_\mu(\mu_0(\theta_0))\boldsymbol{\gamma} + \mathbf{H}_\sigma(\sigma_0(\theta_0))\boldsymbol{\delta} = \sum_{i=1}^I w_{1i} \rho''(\gamma_i) \mathbf{a}_i(\theta_0) \gamma_i + \sum_{i=1}^I \sum_{j=i}^I w_{2ij} \rho''(\delta_{ij}) \mathbf{b}_{ij}(\theta_0) \delta_{ij}. \quad (9)$$

When using the square loss function  $\rho(x) = x^2$  in ULS estimation,  $\rho''(x) = 2$  and all model deviations contribute equally in the adapted parameter estimate  $\theta$ . In contrast, when using a robust loss function  $\rho(x) = |x|^p$  for  $p \leq 1$ , model deviations  $\gamma_i$  and  $\delta_{ij}$  are differentially weighted according to  $\rho''(\gamma_i)$  and  $\rho''(\delta_{ij})$  in Equation (9), respectively [21].

The result in Equation (8) highlights that model deviations  $\boldsymbol{\gamma}$  and  $\boldsymbol{\delta}$  enter the computation of the model parameter  $\theta$ . With suitable loss function  $\rho$ , the influence of model deviations can be reduced if the second derivative  $\rho''$  is sufficiently small for gross model misspecifications.

### 3. Group Comparisons in the Tau-Equivalent Model with Noninvariant Item Intercepts

Suppose that there are a fixed number of  $G$  groups. In each of the  $G$  groups, there exists a mean vector  $\boldsymbol{\mu}_g = (\mu_{1g}, \dots, \mu_{Ig})$  and a covariance matrix  $\boldsymbol{\Sigma}_g = (\sigma_{ijg})_{ij}$  for items  $\mathbf{X}_g = (X_{1g}, \dots, X_{Ig})$ . A latent variable  $F_g$  in a one-dimensional factor model summarizes the distribution of items in each group  $g = 1, \dots, G$ . Define  $\alpha_g = E(F_g)$  and  $\psi_g = \text{Var}(F_g)$ . In the sequel, we discuss the identification of  $\alpha_g$  and  $\psi_g$  for various fitting functions (i.e., estimation methods).

We now model the mean structure  $\boldsymbol{\mu}_g$  and the covariance structure  $\boldsymbol{\Sigma}_g$  with the tau-equivalent one-dimensional factor model in each group  $g$ . By assuming the identification constraint  $E(F_g) = 0$  and estimating  $\psi_g = \text{Var}(F_g)$ , this poses restrictions

$$\boldsymbol{\mu}_g = \boldsymbol{\nu}_g \quad \text{and} \quad \boldsymbol{\Sigma}_g = \psi_g \mathbf{1}\mathbf{1}^\top + \phi_g \mathbf{I}. \quad (10)$$

It can be seen from (10) that the mean structure is estimated without constraints, while severe constraints on the covariance structure are imposed.

If the mean  $\alpha_g$  and the variance  $\psi_g$  of the factor variable  $F_g$  should also be determined in each group for enabling a comparison of the factor variable  $F_g$  across groups, additional

identification constraints have to be posed. This can be seen by including these parameters in Equation (10)

$$\boldsymbol{\mu}_g = \alpha_g \mathbf{1} + \boldsymbol{\nu}_g \quad \text{and} \quad \boldsymbol{\Sigma}_g = \psi_g \mathbf{1}\mathbf{1}^\top + \phi_g \mathbf{I}. \quad (11)$$

A popular identification constraint is to assume invariant item intercepts  $\boldsymbol{\nu}_0$  across groups. In this case of MI, (11) simplifies to

$$\boldsymbol{\mu}_g = \alpha_g \mathbf{1} + \boldsymbol{\nu}_0 \quad \text{and} \quad \boldsymbol{\Sigma}_g = \psi_g \mathbf{1}\mathbf{1}^\top + \phi_g \mathbf{I}. \quad (12)$$

The group means and group variances can be identified by assuming  $\alpha_1 = 0$ . The condition (12) can be characterized as scalar invariance [7]. The MI assumption (12) can be statistically tested [7]. If the MI hypothesis (12) is not rejected,  $\alpha_g$  and  $\psi_g$  can be uniquely determined. In the violation of MI (i.e., measurement noninvariance; MNI), there is indeterminacy in defining group means and group variances. Identification constraints are implicitly posed by assuming particular fitting functions. We discuss several alternative fitting functions and draw relations among the different approaches below. In the following treatment, we allow group-specific item intercepts in the data-generating model:

$$\boldsymbol{\mu}_g = \alpha_g \mathbf{1} + \boldsymbol{\nu}_0 + \boldsymbol{\nu}_g^* \quad \text{and} \quad \boldsymbol{\Sigma}_g = \psi_g \mathbf{1}\mathbf{1}^\top + \phi_g \mathbf{I}. \quad (13)$$

The group-specific item intercepts  $\boldsymbol{\nu}_g^*$  are residuals that describe differences from the common item intercepts  $\boldsymbol{\nu}_0$ . Hence, violation of MI (i.e., MNI) is represented in  $\boldsymbol{\nu}_g^*$ . Condition (13) is also characterized as metric invariance [7].

### 3.1. Joint Estimation

In joint estimation, group means, group variances and common item parameters are estimated. However, MNI effects are not explicitly modeled as additional parameters. Group-specific means  $\boldsymbol{\mu}_g$  and covariances  $\boldsymbol{\Sigma}_g$  are used for determining the vector of model parameters  $\boldsymbol{\theta} = (\boldsymbol{\nu}_0, \alpha_2, \dots, \alpha_G, \psi_1, \dots, \psi_G, \phi_1, \dots, \phi_G)$ . In Section 2.3, it was argued that many estimation methods like ML estimation could be (approximately) characterized as DWLS estimation. DWLS uses the square loss function, but one can stick to the more general case of a loss function  $\rho$  (see Equation (4)). Using a set of known weights  $w_{1ig}$  and  $w_{2ijg}$  for  $i, j = 1, \dots, I$  and  $g = 1, \dots, G$ , the following fitting function is minimized:

$$F(\boldsymbol{\theta}) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\mu_{ig} - \mu_{0,ig}(\boldsymbol{\theta})) + \sum_{g=1}^G \sum_{i=1}^I \sum_{j=i}^I w_{2ijg} \rho(\sigma_{ijg} - \sigma_{0,ijg}(\boldsymbol{\theta})). \quad (14)$$

Note that the order in the summation in (14) across groups (index  $g$ ) and items (indices  $i$  and  $j$ ) can be swapped. The model assumption (13) for  $\mu_{0,ig}(\boldsymbol{\theta})$  and  $\sigma_{0,ijg}(\boldsymbol{\theta})$  can be included in (14), and we obtain

$$F(\boldsymbol{\theta}) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\mu_{ig} - \alpha_g - \nu_{0i}) + \sum_{g=1}^G \sum_{i=1}^I \sum_{j=i}^I w_{2ijg} \rho(\sigma_{ijg} - \psi_g - \phi_g \mathbf{1}_{\{i=j\}}), \quad (15)$$

where  $\mathbf{1}_A$  denotes the indicator function for a set  $A$ . Because of the invariance assumption for item loadings in (13), the second term in (15) will be exactly zero and  $\psi_1, \dots, \psi_G, \phi_1, \dots, \phi_G$  can be uniquely determined by data. One can choose  $\psi_g = \sigma_{ig}$  for any  $i \neq j$ . Then,  $\phi_g = \sigma_{ig} - \psi_g$  for any  $i$ . The group means  $\alpha_g$  and common item intercepts  $\nu_{0,i}$  can be estimated by minimizing

$$F_1(\boldsymbol{\theta}) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\mu_{ig} - \alpha_g - \nu_{i0}). \quad (16)$$

It can be seen that (16) corresponds to an analysis of variance model in which the two-way data  $\mu_{ig}$  for items  $i = 1, \dots, I$  and groups  $g = 1, \dots, G$  is represented by two sets of main effects  $\alpha_g$  and  $\nu_{i0}$  [23]. For item intercepts  $\nu_{i0}$ , we obtain estimating equations

$$\frac{\partial F_1}{\partial \nu_{i0}} = - \sum_{g=1}^G w_{1ig} \rho'(\mu_{ig} - \alpha_g - \nu_{i0}) = 0 \quad \text{for } i = 1, \dots, I. \quad (17)$$

For group means  $\alpha_g$ , we similarly obtain

$$\frac{\partial F_1}{\partial \alpha_g} = - \sum_{i=1}^I w_{1ig} \rho'(\mu_{ig} - \alpha_g - \nu_{i0}) = 0 \quad \text{for } g = 2, \dots, G. \quad (18)$$

Due to the assumption (13), we have  $\mu_{ig} - \alpha_g - \nu_{i0} = \nu_{ig}^*$  for the group-specific item intercepts. Hence, it follows from (18) that

$$\sum_{i=1}^I w_{1ig} \rho'(\nu_{ig}^*) = 0 \quad \text{for all } g = 2, \dots, G. \quad (19)$$

From (17), we get

$$\sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho'(\mu_{ig} - \alpha_g - \nu_{i0}) = 0. \quad (20)$$

From (19) and (20), we finally obtain

$$\sum_{i=1}^I w_{1ig} \rho'(\nu_{ig}^*) = 0 \quad \text{for all } g = 1, \dots, G. \quad (21)$$

The finding in Equation (21) demonstrates that there is an assumption of group-specific residual item intercepts  $\nu_{ig}^*$  when fitting a multiple group factor model under violation of MI. Hence, group means  $\psi_g$  depend on choosing sets of weights  $w_{1ig}$  and a loss function  $\rho$ .

When choosing  $w_{1ig} \equiv 1$ , the condition of partial invariance (PI; [24]) is received for the loss function  $\rho(x) = |x|^p$  for  $p \rightarrow 0$  which takes the value of 0 iff  $x = 0$  and 1 for  $x \neq 0$ . In PI, it is typically assumed there is a subset of items for which  $\nu_{ig0}^* \neq 0$  (i.e.,  $\rho(\nu_{ig0}^*) = 1$ ). For the majority of items, it holds that  $\nu_{ig0}^* = 0$  (i.e.,  $\rho(\nu_{ig0}^*) = 0$ ). The loss function in (16) then minimizes the number of group-specific residual item intercepts that differ from zero [25].

### 3.2. Linking

In linking methods [26], the one-dimensional factor model is firstly estimated in each of the groups. In the tau-equivalent model, the group variance  $\psi_g$  can be identified, but the group-specific estimation only provides identified item intercepts  $\nu_{ig}$  that are given as (see (13))

$$\nu_{ig} = \alpha_g + \nu_{i0} + \nu_{ig}^*.$$

Note that item intercepts  $\nu_{ig}$  coincide with group-specific item means  $\mu_{ig} = E(X_{ig})$ .

In a second step in the linking approach, the intercepts are used to determine group means  $\alpha_g$  and common item intercepts  $\nu_{i0}$  [21]. By defining  $\Theta = (\alpha_2, \dots, \alpha_G, \nu_{10}, \dots, \nu_{I0})$ , a linking function  $H$  defined by

$$H(\Theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\nu_{ig} - \alpha_g - \nu_{i0}) \quad (22)$$

using some set of weights  $w_{1ig}$  that are chosen equal to one in many applications. Again, the order in the summation in (22) across groups (index  $g$ ) and items (indices  $i$  and  $j$ )

can be swapped. The linking function in (22) can be considered as Haberman linking (HL; [21,27,28]). In [28], the loss function  $\rho(x) = x^2$  was used while [21] treated the robust loss function  $\rho(x) = |x|^p$  for  $p \in [0, 2]$ . Note that for the tau-equivalent model, the minimization problem (22) is exactly the same as the minimization problem (16) in joint estimation. This is trivial because the item intercepts coincide with the observed group-specific item means and if the covariance structure is correctly specified. Hence, the same condition as in (21) for group-specific residual item intercepts  $v_{ig}^*$ .

An alternative linking approach has been proposed that avoids estimating common item intercepts  $v_{i0}$ . In invariance alignment (IA; [29]), the following function  $G$  is minimized for determining  $\Theta = (\alpha_2, \dots, \alpha_G)$  while setting  $\alpha_1 = 0$ :

$$G(\Theta) = \sum_{g=1}^G \sum_{h=1}^H \sum_{i=1}^I w_{1ig} w_{1ih} \rho(v_{ig} - v_{ih} - \alpha_g + \alpha_h), \quad (23)$$

where the loss function  $\rho(x) = |x|^p$  for  $p \geq 0$  is utilized [21,30]. The power  $p = 0.5$  is most frequently chosen because it is the default in the software package Mplus [30]. It was empirically found that the IA minimization in (23) provides very similar group mean estimates as the minimization in (22) that also involves the estimation of common item intercepts [21]. Indeed, the loss function  $\rho(x) = |x|^p$  is a subadditive function for  $p \leq 1$  [31] which means that

$$\rho(x+y) \leq \rho(x) + \rho(y) \quad \text{for all } x, y \in \mathbb{R}. \quad (24)$$

By defining  $x = v_{ig} - \alpha_g - v_{i0}$  and  $y = -(v_{ih} - \alpha_h - v_{i0})$ , we get from (23) by using (24)

$$G(\Theta) \leq \sum_{g=1}^G \sum_{i=1}^I \tilde{w}_{1ig} \rho(v_{ig} - \alpha_g - v_{i0}) = \tilde{H}(\Theta), \quad (25)$$

where the weights  $\tilde{w}_{1ig}$  are defined as

$$\tilde{w}_{1ig} = 2w_{1ig} \sum_{h=1}^H w_{1ih}. \quad (26)$$

Hence, the majorizing function  $\tilde{H}$  in (25) is exactly given by the minimization function  $H$  in (22) when using properly defined weights  $\tilde{w}_{1ig}$  but the same loss function  $\rho$ . As a conclusion, joint estimation and linking methods can be regarded as exchangeable in the tau-equivalent model. They pose the same identification constraints on group-specific residual item intercepts.

Note that the two-step linking approach can be rewritten as a one-step estimation approach with overidentified parameters. Identification is ensured by posing side constraints implied by the linking function [32]. A joint optimization problem can be formulated by using Lagrange multipliers. Assume that there are associated weights  $w_{1ig}^*$  with group-wise ML estimation and different weights  $w_{1ig}$  in the linking method. Suppose that we parametrize  $v_{ig} = v_{i0} + \alpha_g + \tilde{v}_{ig}$ . The reformulated one-step fitting function  $F_{\text{Lagrange}}$  using Lagrange multipliers  $\ell_{1g}$  and  $\ell_{2i}$  of the two-step linking approaches is given by (see [32])

$$F_{\text{Lagrange}}(\Theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig}^* (\mu_{ig} - \alpha_g - v_{i0} - \tilde{v}_{ig})^2 + \sum_{g=2}^G \ell_{1g} \sum_{i=1}^I w_{1ig} \rho'(\tilde{v}_{ig}) + \sum_{i=1}^I \ell_{2i} \sum_{g=1}^G w_{1ig} \rho'(\tilde{v}_{ig}), \quad (27)$$

where  $\Theta$  now also includes the  $G - 1 + I$  Lagrange multipliers  $\ell_{1g}$  and  $\ell_{2i}$ . Note that the second and third term in (27) corresponds to estimating equations obtained from the linking method by determining group means  $\alpha_g$  and common item intercepts  $v_{i0}$ .

### 3.3. Regularization

MNI has also been tackled by regularization methods [33–35]. The main idea is to introduce nonidentified group-specific residual item intercepts  $\tilde{v}_{ig}$  in ML estimation. The estimation problem becomes identifiable by adding a penalty function  $\mathcal{P}$  to the negative likelihood function [36]. Again, the covariance structure is assumed to be correctly specified. The estimated model parameters are collected in the vector  $\theta = (\nu_{10}, \dots, \nu_{1I}, \alpha_2, \dots, \alpha_G, \dots, \tilde{v}_{ig}, \dots)$ . Using the result (3), the fitting function can be approximately written

$$F_{\text{reg}}(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} (\mu_{ig} - \alpha_g - \nu_{i0} - \tilde{v}_{ig})^2 + \sum_{g=1}^G \sum_{i=1}^I \mathcal{P}(\kappa, \tilde{v}_{ig}), \quad (28)$$

where  $\kappa > 0$  is a tuning parameter. A popular penalty function is the lasso penalty  $\mathcal{P}(\kappa, x) = \kappa|x|$ , but alternative lasso-type penalty functions with similar behavior to the former one around  $x = 0$  but more desirable statistical properties have been proposed [36]. Alternatively, the ridge penalty  $\mathcal{P}(\kappa, x) = \kappa x^2$  can be used that controls the variability of effects od MNI.

When using lasso-type penalty functions, the residual intercepts  $\tilde{v}_{ig}$  can be interpreted as outliers. Indeed, for the lasso penalty, it has been shown that the minimization of  $F_{\text{reg}}$  in (28) using regularization is equivalent to robust regression with outlier detection [37,38]. By defining  $\tilde{\theta} = (\nu_{10}, \dots, \nu_{1I}, \alpha_2, \dots, \alpha_G)$  and an appropriate loss function  $\tilde{\rho}$ , the minimization problem (28) can be rewritten as

$$F_{\text{rob}}(\tilde{\theta}) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \tilde{\rho}(\mu_{ig} - \alpha_g - \nu_{i0}),$$

Hence, regularized ML estimation can be equivalently recognized as joint estimation using a particular loss function that enables the efficient detection of outliers.

We now characterize the solution of (28) in more detail. From the behavior of the lasso penalty it is reasonable to assume that estimated residual item intercepts  $\tilde{v}_{ig} = 0$  equal zero iff  $|\nu_{ig}^*| < \kappa$  holds for true residual item intercepts  $\nu_{ig}^*$ . On the other hand, we can assume that  $\tilde{v}_{ig} = \nu_{ig}^*$  iff  $|\nu_{ig}^*| > \kappa$ . For determining the group mean  $\alpha_g$ , we get the estimating equation

$$\frac{\partial F_{\text{reg}}}{\partial \alpha_g} = -2 \sum_{i=1}^I w_{1ig} (\mu_{ig} - \alpha_g - \nu_{i0} - \tilde{v}_{ig}) = 0. \quad (29)$$

By relying on the just mentioned properties for estimated residual item intercepts  $\tilde{v}_{ig}$ , we obtain the condition

$$\sum_{i=1}^I w_{1ig} \nu_{ig}^* \mathbf{1}_{\{|\nu_{ig}^*| < \kappa\}} = 0 \quad \text{for all } g = 1, \dots, G. \quad (30)$$

The result in (30) indicates the MNI cancels out on average for small effects  $\nu_{ig}^*$  that fulfill  $|\nu_{ig}^*| < \kappa$ . Note that this set of effects is implicitly estimated in regularized ML.

In the case of a general penalty function  $\mathcal{P}$ , define  $\mathcal{P}_1 = \frac{\partial \mathcal{P}}{\partial x}$ . Note that we replace a nondifferentiable penalty function with a differentiable approximation  $\mathcal{P}$  [33,39]. For determining the group mean  $\alpha_g$ , the condition (29) does not change. For determining  $\tilde{v}_{ig}$ , we get the estimating equation

$$-w_{1ig} (\mu_{ig} - \alpha_g - \nu_{i0} - \tilde{v}_{ig}) + \mathcal{P}_1(\kappa, \tilde{v}_{ig}) = -w_{1ig} (\nu_{ig}^* - \tilde{v}_{ig}) + \mathcal{P}_1(\kappa, \tilde{v}_{ig}) = 0. \quad (31)$$

By using (31), there exists a function  $\mathcal{Q}$  such that  $\tilde{\nu}_{ig} = \mathcal{Q}(\nu_{ig}^*)$ . Moreover, by summing (31) across items  $i = 1, \dots, I$ , we receive

$$\sum_{i=1}^I \left\{ -w_{1ig}(\nu_{ig}^* - \tilde{\nu}_{ig}) + \mathcal{P}_1(\kappa, \tilde{\nu}_{ig}) \right\} = 0.$$

Because (29) holds, we get

$$\sum_{i=1}^I w_{1ig}(\nu_{ig}^* - \tilde{\nu}_{ig}) = 0 \text{ and } \sum_{i=1}^I \mathcal{P}_1(\kappa, \tilde{\nu}_{ig}) = 0.$$

This means that estimated effects  $\tilde{\nu}_{ig} = \mathcal{Q}(\nu_{ig}^*)$  somehow vanish on average according to their contribution in  $\mathcal{P}_1$ .

Instead of introducing group-specific residual item intercepts  $\nu_{ig}^*$  in regularized ML estimation, one can employ fused regularized ML estimation [40] that relies on overidentified group-specific item intercepts  $\check{\nu}_{ig}$ . The fitting function is fused regularized ML is defined as

$$F_{\text{fusedreg}}(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig}(\mu_{ig} - \alpha_g - \check{\nu}_{ig})^2 + \sum_{g=1}^{G-1} \sum_{h=g+1}^G \sum_{i=1}^I \mathcal{P}(\kappa, \check{\nu}_{ig} - \check{\nu}_{ih}). \quad (32)$$

In this case, the parameter vector  $\theta$  does not include common item intercepts  $\nu_0$ . The nonidentification issue of ML estimation is solved by defining a penalty function  $\mathcal{P}$  on deviations  $\check{\nu}_{ig} - \check{\nu}_{ih}$ . By using lasso-type penalty functions in (32), clusters of  $\check{\nu}_{ig}$  coefficients will be obtained. If there are only a few outlying parameters for each item, estimated group means  $\alpha_g$  from fused regularized ML using the fitting function in (32) will often be similar to those in regularized ML estimation using the fitting function in (28).

Another popular approach to handling MNI is the Bayesian approximate measurement invariance model (BAMI; [41–43]). The tau-equivalent model is estimated with an overidentified parameter vector that includes all item intercepts  $\nu_{ig}$ . To ensure the identification of the model, a normal prior with known variance on all pairwise deviations  $\nu_{ig} - \nu_{ih}$  is posed. The normal prior distribution on  $\nu_{ig} - \nu_{ih}$  can be regarded as a ridge penalty function of the form  $\tilde{\kappa}(\nu_{ig} - \nu_{ih})^2$  (see, e.g., [44]). Hence, BAMI can be recognized as fused regularized ML estimation with a particular penalty function in (32).

Interestingly, Battauz [39] showed for regularized estimation in the four-parameter item response model that the ridge penalty on differences  $\nu_{ig} - \nu_{ih}$  can be rewritten as a penalty for residual item intercepts  $\tilde{\nu}_{ig} = \nu_{ig} - \nu_{i0}$ :

$$\tilde{\kappa} \sum_{g=1}^G \sum_{h=1}^H \sum_{i=1}^I (\nu_{ig} - \nu_{ih})^2 = \kappa \sum_{g=1}^G \sum_{i=1}^I \tilde{\nu}_{ig}^2$$

using an appropriate tuning parameter  $\kappa$ . By replacing the Markov chain Monte Carlo estimation method of the BAMI model with regularized ML estimation, we obtain the fitting function

$$F_{\text{BAMI}}(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig}(\mu_{ig} - \alpha_g - \nu_{i0} - \tilde{\nu}_{ig})^2 + \kappa \sum_{g=1}^G \sum_{i=1}^I \tilde{\nu}_{ig}^2,$$

which is regularized ML estimation using a ridge penalty function. For determining group means  $\alpha_g$ , we get the estimating equation

$$\sum_{i=1}^I w_{1ig}(\mu_{ig} - \alpha_g - \nu_{i0} - \tilde{\nu}_{ig}) = 0. \quad (33)$$

For determining  $\tilde{\nu}_{ig}$ , we get

$$-w_{1ig}(\mu_{ig} - \alpha_g - \nu_{i0} - \tilde{\nu}_{ig}) + \kappa \tilde{\nu}_{ig} = 0$$

Hence, the estimated group-specific residual item intercepts are shrunken estimates of  $\nu_{ig}^*$

$$\tilde{\nu}_{ig} = \frac{w_{1ig}}{w_{1ig} + \kappa}(\mu_{ig} - \alpha_g - \nu_{i0}) = \frac{w_{1ig}}{w_{1ig} + \kappa}\nu_{ig}^* \quad (34)$$

that fulfill  $\sum_{i=1}^I \tilde{\nu}_{ig} = 0$ . By using (34), we get from (33)

$$\sum_{i=1}^I w_{1ig}(\nu_{ig}^* - \tilde{\nu}_{ig}) = \sum_{i=1}^I \frac{\kappa w_{1ig}}{w_{1ig} + \kappa}\nu_{ig}^* = 0.$$

With equal weights  $w_{1ig}$  for all items within a group  $g$ , this shows that BAMI and ML pose the same identification constraints on  $\nu_{ig}^*$ ; that is,  $\sum_{i=1}^I \nu_{ig}^* = 0$ .

A variant of IA has been proposed that uses the output of BAMI for determining the alignment solution [45–47]. BAMI produces adjusted group-specific item means  $\tilde{\mu}_{ig} = \alpha_g + \nu_{i0} + \tilde{\nu}_{ig}$ , which are subsequently used as input data for Bayesian IA. Note that

$$\tilde{\mu}_{ig} = \alpha_g + \nu_{i0} + \tilde{\nu}_{ig} = \mu_{ig} + \tilde{\nu}_{ig} - \nu_{ig}^* = \mu_{ig} - \frac{\kappa}{w_{1ig} + \kappa}\nu_{ig}^*. \quad (35)$$

We now substitute the quantity  $\tilde{\mu}_{ig}$  obtained in (35) in the IA fitting function (see (25) and weights  $\tilde{w}_{1ig}$  defined in (26)):

$$\tilde{H}(\Theta) = \sum_{g=1}^G \sum_{i=1}^I \tilde{w}_{1ig} \rho(\tilde{\mu}_{ig} - \alpha_g - \nu_{i0}) = \sum_{g=1}^G \sum_{i=1}^I \tilde{w}_{1ig} \rho\left(\mu_{ig} - \frac{\kappa}{w_{1ig} + \kappa}\nu_{ig}^* - \alpha_g - \nu_{i0}\right).$$

The estimating equation for  $\alpha_g$  is then given by

$$\sum_{i=1}^I \tilde{w}_{1ig} \rho' \left( \mu_{ig} - \frac{\kappa}{w_{1ig} + \kappa}\nu_{ig}^* - \alpha_g - \nu_{i0} \right) = 0.$$

Using the definition  $\mu_{ig} = \alpha_g + \nu_{i0} + \nu_{ig}^*$ , we get the identification constraint

$$\sum_{i=1}^I \tilde{w}_{1ig} \rho' \left( \frac{w_{1ig}}{w_{1ig} + \kappa}\nu_{ig}^* \right) = \sum_{i=1}^I \tilde{w}_{1ig} \rho'(\tilde{\nu}_{ig}) = 0. \quad (36)$$

When equal weights  $w_{1ig}$  (and  $\tilde{w}_{1ig}$ ) are used in (36), the identification constraints  $\sum_{i=1}^I \rho'(\omega \nu_{ig}^*) = 0$  with a scaling factor  $\omega > 0$  are obtained. For  $\rho(x) = x^2$ , it holds that  $\rho'(x) = 2x$ , and we receive the same identification constraints as those obtained with ULS estimation or DWLS estimation with equal weights.

#### 4. Group Comparisons in the Tau-Congeneric Model with Noninvariant Item Intercepts

In the following, we discuss group comparisons in the tau-congeneric model. We investigate the consequences of MNI in item intercepts but assume MI in item loadings; that is, item loadings are invariant, and metric invariance holds [6,7]. It will be shown that the derivations of the tau-equivalent model only have slightly to be modified.

For the following examinations, we assume common loadings  $\lambda_0$  where the first loading  $\lambda_{10}$  is fixed to 1 for reasons of identification in the data-generating model

$$\mu_g = \alpha_g \lambda_0 + \nu_0 + \nu_g^* \quad \text{and} \quad \Sigma_g = \Psi_g \lambda_0 \lambda_0^\top + \Phi_g,$$

where  $\Phi_g$  is a diagonal matrix of group-specific residual variances. Violation of MI only pertains to the mean structure due to the presence of group-specific residual item intercepts  $\nu_g^*$ .

#### 4.1. Joint Estimation

In the joint estimation of the tau-congeneric measurement model, the parameter of interest is given as  $\theta = (\nu_0, \lambda_0, \alpha_2, \dots, \alpha_G, \psi_1, \dots, \psi_G, \Phi_1, \dots, \Phi_G)$ , where  $\Phi_g$  include the diagonal entries of  $\Phi_g$ . The general fitting function is defined by

$$F(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\mu_{ig} - \alpha_g \lambda_{i0} - \nu_{0i}) + \sum_{g=1}^G \sum_{i=1}^I \sum_{j=i}^I w_{2ijg} \rho(\sigma_{ijg} - \psi_g \lambda_{i0} \lambda_{j0} - \phi_{ig} \mathbf{1}_{\{i=j\}}). \quad (37)$$

Because the covariance structure is correctly specified, common item loadings  $\lambda_0$  and group-specific residual variance matrices  $\Phi_g$  can be uniquely determined by minimizing the second term in (37). Hence, for determining group means  $\alpha_g$ , only the first term in (37) must be considered, which results in a fitting function

$$F_1(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\mu_{ig} - \alpha_g \lambda_{i0} - \nu_{0i}). \quad (38)$$

Similar to Equation (21) in Section 3.1, we get the identification constraints by taking the same steps as in Section 3.1

$$\sum_{i=1}^I w_{1ig} \lambda_{i0} \rho'(\nu_{ig}^*) = 0 \quad \text{for all } g = 1, \dots, G. \quad (39)$$

Comparing (39) with (21), it is vital that the common item loadings  $\lambda_{i0}$  now also enter the identification constraint (see also [48]).

#### 4.2. Linking

We now investigate HL in the tau-equivalent model [21] (see [48] for a similar approach). Again, identified parameters (i.e., item intercepts  $\nu_{ig}$  and item loadings  $\lambda_{ig}$ ) are firstly obtained in group-wise estimations

$$\nu_{ig} = \alpha_g \lambda_{i0} + \nu_{i0} + \nu_{ig}^* \quad \text{and} \quad \lambda_{ig} = \psi_g^{1/2} \lambda_{i0}. \quad (40)$$

From (40) and assuming  $\lambda_{10} = 1$  for the first item for reasons of identification, it can be seen that all group variances  $\psi_g$  can be uniquely identified. Obviously, we get  $\psi_g = \lambda_{1g}^2$ . For determining group means  $\alpha_g$ , we define a linking function  $H$  for the parameter  $\theta = (\nu_{i1}, \dots, \nu_{i0}, \psi_2, \dots, \psi_G)$  of interest:

$$H(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\nu_{ig} - \alpha_g \lambda_{i0} - \nu_{i0}) \quad (41)$$

Like in Section 3.2, it can be seen that the minimization problem in (41) corresponds to the problem in (38). Hence, the same identification constraints as in (39) are obtained.

As also discussed in Section 3.2, one can also argue that IA appears very similar to the linking problem in (41) and therefore resembles joint estimation by using an appropriate loss function  $\rho$ .

#### 4.3. Regularization

The fitting function for the mean structure of regularized ML estimation for the tau-congeneric model with invariant item loadings can be written as

$$F_{\text{reg}}(\theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} (\mu_{ig} - \alpha_g \lambda_{i0} - \nu_{i0} - \tilde{\nu}_{ig})^2 + \sum_{g=1}^G \sum_{i=1}^I \mathcal{P}(\kappa, \tilde{\nu}_{ig}). \quad (42)$$

Similar arguments like in Section 3.3 can be made for demonstrating the equivalence of regularized ML estimation using (42) and joint estimation using a robust loss function  $\rho$  (see (38)).

Similar findings as in Section 3.3 regarding the equivalence of fused regularized ML estimation and BAMI can be made. The similarity of linking methods and regularized ML estimation was also pointed out by [48] in item response models for dichotomous items.

### 5. Group Comparisons in the Tau-Congeneric Model with Noninvariant Item Intercepts and Noninvariant Item Loadings

Finally, we consider the estimation of the tau-congeneric measurement model. The data-generating model allows for noninvariant item intercepts and noninvariant item loadings.

Let  $x \circ y$  be the Hadamard product (i.e., element-wise multiplication of vectors  $x$  and  $y$ ). We assume

$$\mu_g = \alpha_g \lambda_0 \circ \lambda_g^* + \nu_0 + \nu_g^* \quad \text{and} \quad \Sigma_g = \psi_g (\lambda_0 \circ \lambda_g^*) (\lambda_0 \circ \lambda_g^*)^\top + \Phi_g. \quad (43)$$

For entries in  $\mu_g$  and  $\Sigma_g$ , we get from (43)

$$\mu_{ig} = \alpha_g \lambda_{i0} \lambda_{ig}^* + \nu_{i0} + \nu_{ig}^* \quad \text{and} \quad \sigma_{ijg} = \psi_g \lambda_{i0} \lambda_{j0} \lambda_{ig}^* \lambda_{jg}^* + \phi_{ig} \mathbf{1}_{\{i=j\}}. \quad (44)$$

Then, MNI is represented in  $\lambda_g^*$  and  $\nu_g^*$ . Note that values of  $\lambda_{ig}^*$  equal to 1 indicate MI of a particular item parameter, while MNI is represented by values different from 1.

It is important to emphasize that deviations from MI in item loadings are modeled as multiplicative effects. This assumption means that there is an additive representation for logarithmized identified group-specific item loadings  $l_{ig} = \log \lambda_{ig}$

$$l_{ig} = f_g + l_{i0} + l_{ig}^*,$$

where  $l_{i0} = \log \lambda_{i0}$ ,  $l_{ig}^* = \log \lambda_{ig}^*$ , and  $f_g = \frac{1}{2} \log \psi_g$ . Instead of treating deviations as multiplicative errors, one could also assume additive errors of deviations ([49], see also [50]) such that  $\lambda_{ig} = \psi_g^{1/2} (\lambda_{i0} + \lambda_{ig}^*)$ . In this case, the group-specific mean and covariances can be written as

$$\mu_{ig} = \alpha_g (\lambda_{i0} + \lambda_{ig}^*) + \nu_{i0} + \nu_{ig}^* \quad \text{and} \quad \sigma_{ijg} = \psi_g (\lambda_{i0} + \lambda_{ig}^*) (\lambda_{j0} + \lambda_{jg}^*) + \phi_{ig} \mathbf{1}_{\{i=j\}}. \quad (45)$$

Note the difference to the parameterization in (44).

We now study the effects of MNI in intercepts and loadings in different estimation approaches for determining group means and group variances in the tau-congeneric model.

#### 5.1. Joint Estimation

In joint estimation, the parameter of interest is given by  $\Theta = (\nu_0, \lambda_0, \alpha_2, \dots, \alpha_G, \psi_2, \dots, \psi_G, \Phi_1, \dots, \Phi_G)$ , while we set  $\alpha_1 = 0$  and  $\psi_1 = 1$  in the estimation. We consider the general fitting function

$$F(\Theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho(\mu_{ig} - \alpha_g \lambda_{i0} - \nu_{0i}) + \sum_{g=1}^G \sum_{i=1}^I \sum_{j=i}^I w_{2ijg} \rho(\sigma_{ijg} - \psi_g \lambda_{i0} \lambda_{j0} - \phi_{ig} \mathbf{1}_{\{i=j\}}).$$

We first investigate the determination of group variances  $\psi_g$ . Assume that common item loadings  $\lambda_0$  have already been determined. Then, we get the estimating equation by taking  $\frac{\partial F}{\partial \psi_g}$

$$\sum_{i=1}^{I-1} \sum_{j=i+1}^I w_{2ijg} \lambda_{i0} \lambda_{j0} \rho'(\sigma_{ijg} - \psi_g \lambda_{i0} \lambda_{j0}) = 0.$$

Note that  $\phi_{ig}$  can be uniquely determined and there is a vanishing contribution of terms for  $i = j$ . Using (44), this implies the identification constraint

$$\sum_{i=1}^{I-1} \sum_{j=i+1}^I w_{2ijg} \lambda_{i0} \lambda_{j0} \rho' \left( \psi_g \lambda_{i0} \lambda_{j0} (\lambda_{ig}^* \lambda_{jg}^* - 1) \right) = 0. \quad (46)$$

For the loss function  $\rho(x) = |x|^p$ , we get due to  $\rho'(\lambda x) = \rho'(\lambda) \rho'(x)$  for any  $\lambda \geq 0$  from (46)

$$\sum_{i=1}^{I-1} \sum_{j=i+1}^I w_{2ijg} \lambda_{i0} \lambda_{j0} \rho' (\psi_g \lambda_{i0} \lambda_{j0}) \rho' (\lambda_{ig}^* \lambda_{jg}^* - 1) = 0. \quad (47)$$

Group-specific residual item loadings  $\lambda_{ig}^*$  vanish on average according to the identification constraint (47). We can further specialize (47) for DWLS estimation (which can also approximate ML estimation), which employs the square loss function  $\rho(x) = x^2$ :

$$\sum_{i=1}^{I-1} \sum_{j=i+1}^I w_{2ijg} \lambda_{i0}^2 \lambda_{j0}^2 \psi_g (\lambda_{ig}^* \lambda_{jg}^* - 1) = 0$$

Now, we determine group means  $\alpha_g$ . Assume that common item loadings  $\lambda_0$  and  $\nu_0$  have already been determined. By taking  $\frac{\partial F}{\partial \alpha_g}$ , we get the identification constraints

$$\sum_{i=1}^I w_{1ig} \lambda_{i0} \rho' (\alpha_g \lambda_{i0} (\lambda_{ig}^* - 1) + \nu_{ig}^*) = 0. \quad (48)$$

For DWLS estimation, we get the identification constraint

$$\sum_{i=1}^I w_{1ig} \lambda_{i0} (\alpha_g \lambda_{i0} (\lambda_{ig}^* - 1) + \nu_{ig}^*) = 0. \quad (49)$$

It can be seen that MNI in loadings due to  $\lambda_{ig}^*$  as well as MNI in intercepts due to  $\nu_{ig}^*$  determine the group mean  $\alpha_g$ .

For additive deviations from MI that follow (45), the condition (46) is replaced by

$$\sum_{i=1}^{I-1} \sum_{j=i+1}^I w_{2ijg} \lambda_{i0} \lambda_{j0} \rho' (\psi_g (\lambda_{i0} \lambda_{jg}^* + \lambda_{ig}^* \lambda_{j0} + \lambda_{ig}^* \lambda_{jg}^*)) = 0.$$

The condition (49) is replaced by

$$\sum_{i=1}^I w_{1ig} \lambda_{i0} \rho' (\alpha_g \lambda_{ig}^* + \nu_{ig}^*) = 0.$$

Due to the arbitrariness of using either the multiplicative or additive representation of MNI effects  $\lambda_{ig}^*$ , the different obtained identification conditions should not be interpreted as conflicting findings but rather as different ways of representing the same identification condition.

## 5.2. Linking

This subsection discusses the estimation of group means and group variances for linking approaches. We assume that item loadings follow the multiplicative representation of MNI in (44). This corresponds to an additivity assumption for logarithmized item loadings.

At first, the tau-congeneric measurement model is estimated separately in each group. By assuming  $E(F_g) = 0$  and  $\text{Var}(F_g) = 1$ , identified group-specific item intercepts  $\nu_{ig}$  and group-specific item loadings  $\lambda_{ig}$  are given as

$$\nu_{ig} = \alpha_g \lambda_{i0} \lambda_{ig}^* + \nu_{i0} + \nu_{ig}^* \quad \text{and} \quad \lambda_{ig} = \psi_g^{1/2} \lambda_{i0} \lambda_{ig}^*. \quad (50)$$

Note that (50) can be equivalently written as

$$\nu_{ig} = \alpha_g \lambda_{i0} \lambda_{ig}^* + \nu_{i0} + \nu_{ig}^* \quad \text{and} \quad \log \lambda_{ig} = \frac{1}{2} \log \psi_g + \log \lambda_{i0} + \log \lambda_{ig}^*. \quad (51)$$

In HL [21,28], common logarithmized item loadings  $\mathbf{l}_0 = (l_{10}, \dots, l_{I0}) = \log \lambda_0$  and logarithmized group variances  $f_g = \frac{1}{2} \psi_g$  are computed in the first step. Note that  $\psi_g = \exp(2f_g)$ . The linking function  $H_2$  using  $\Theta = (f_2, \dots, f_G, \mathbf{l}_0)$  using the identification constraint  $f_1 = 0$  (i.e.,  $\psi_1 = 1$ ) is defined as

$$H_2(\Theta) = \sum_{g=1}^G \sum_{i=1}^I w_{2ig} \rho(l_{ig} - f_g - l_{i0}),$$

where  $l_{ig} = \log \lambda_{ig}$ . For determining  $f_g$  (and, hence, the group variance  $\psi_g$ ), applying  $\frac{\partial H_2}{\partial f_g}$  and considering (51) provides the identification constraint

$$\sum_{i=1}^I w_{2ig} \rho'(\log \lambda_{ig}^*) = 0.$$

In the second step in HL, group means  $\alpha_g$  are determined based on identified item intercepts  $\nu_{ig}$  and identified item loadings  $\lambda_{ig}$  (see Equation (50)) and group variances  $\psi_g$  that are determined in a first step. We discuss a variant of HL (see Equation (25) in [21])

$$H_1(\Theta) = \sum_{g=1}^G \sum_{i=1}^I w_{1ig} \rho\left(\nu_{ig} - \nu_{i0} - \psi_g^{-1/2} \lambda_{ig} \alpha_g\right) \quad (52)$$

for  $\Theta = (\alpha_2, \dots, \alpha_G, \mathbf{\nu}_0)$  and employing the identification constraint  $\alpha_1 = 0$ . For determining group means  $\alpha_g$ , we consider  $\frac{\partial H_1}{\partial \alpha_g}$  and get from (52) the identification constraints

$$\sum_{i=1}^I w_{1ig} \psi_g^{-1/2} \lambda_{ig} \rho'\left(\nu_{ig} - \nu_{i0} - \psi_g^{-1/2} \lambda_{ig} \alpha_g\right) = \sum_{i=1}^I w_{1ig} \lambda_{i0} \lambda_{ig}^* \rho'(\nu_{ig}^*) = 0.$$

Another popular linking approach is IA [29]. Originally it was discussed as a simultaneous linking method for determining group means and group variances, it has been shown in [21] that alignment is equivalent to a two-step linking approach.

In the first step of IA, group standard deviations  $p_g$  defined as  $\psi_g = p_g^2$  are determined by minimizing the linking function

$$G_2(\Theta) = \sum_{g=1}^G \sum_{h=1}^H \sum_{i=1}^I w_{2igh} \rho(p_g^{-1} \lambda_{ig} - p_h^{-1} \lambda_{ih})$$

where  $\Theta = (p_2, \dots, p_G)$  using the identification constraint  $p_1 = 1$  (and, hence,  $\psi_1 = 1$ ). For determining  $p_g$  (and subsequently  $\psi_g$ ), we receive the following identification condition by taking  $\frac{\partial G_2}{\partial p_g}$

$$\sum_{h=1}^H \sum_{i=1}^I w_{2igh} \lambda_{i0} \lambda_{ig}^* \rho'(\lambda_{i0}(\lambda_{ig}^* - \lambda_{ih}^*)) = 0.$$

In the second step of IA, group means  $\alpha_g$  using the identification constraint  $\alpha_1 = 0$  are obtained by minimizing the linking function using  $\Theta = (\alpha_2, \dots, \alpha_G)$

$$G_1(\Theta) = \sum_{g=1}^G \sum_{h=1}^H \sum_{i=1}^I w_{1igh} \rho\left(\nu_{ig} - \nu_{ih} - \psi_g^{-1/2} \lambda_{ig} \alpha_g + \psi_h^{-1/2} \lambda_{ih} \alpha_h\right). \quad (53)$$

Then, we get the following identification condition by taking  $\frac{\partial G_1}{\partial \alpha_g}$

$$\sum_{h=1}^H \sum_{i=1}^I w_{1igh} \psi_g^{-1/2} \lambda_{ig} \rho'(\nu_{ig} - \nu_{ih} - \psi_g^{-1/2} \lambda_{ig} \alpha_g + \psi_h^{-1/2} \lambda_{ih} \alpha_h) = \sum_{h=1}^H \sum_{i=1}^I w_{1igh} \lambda_{i0} \lambda_{ig}^* \rho'(\nu_{ig}^* - \nu_{ih}^*) = 0.$$

It has been shown in [21] that HL and IA provide very similar results for the tau-congeneric model if the same loss function  $\rho(x) = |x|^p$  for  $p \in [0, 1]$  is utilized. We can rely on the subadditivity property (24) for finding a majorizing function  $\tilde{H}_1$  in (53)

$$G_1(\theta) \leq \sum_{g=1}^G \sum_{i=1}^I \tilde{w}_{1ig} \rho(\nu_{ig} - \nu_{i0} - \psi_g^{-1/2} \lambda_{ig} \alpha_g) = \tilde{H}_1(\tilde{\theta}),$$

where  $\tilde{\theta}$  now also includes common item intercepts  $\nu_0$  and  $\tilde{w}_{1ig}$  are appropriately defined weights. By minimizing  $\tilde{H}_1$  for determining the group mean  $\alpha_g$ , we receive the identification constraint

$$\sum_{i=1}^I \tilde{w}_{1ig} \lambda_{i0} \lambda_{ig}^* \rho(\nu_{ig}^*) = 0.$$

### 5.3. Regularization

We now discuss identification constraints in regularized ML estimation of the multiple-group tau-congeneric measurement model. Deviations of MI in item loadings can be either modeled in a multiplicative (see (44)) or an additive (see (45)) form. In the following, we assume additive effects because this specification appears more frequently in practical applications.

In regularized ML estimation, we collect in the parameter vector  $\theta$  the parameters  $\alpha_g$ ,  $\nu_{i0}$ ,  $\tilde{\nu}_{ig}$ ,  $\psi_g$ ,  $\lambda_{i0}$ , and  $\tilde{\lambda}_{ig}$ . Moreover, we set  $\alpha_1 = 0$  and  $\psi_1 = 1$  for reason of identification and define the fitting function

$$\begin{aligned} F_{\text{reg}}(\theta) = & \sum_{g=1}^G \sum_{i=1}^I w_{1ig} (\mu_{ig} - \alpha_g (\lambda_{i0} + \tilde{\lambda}_{ig}) - \nu_{i0} - \tilde{\nu}_{ig})^2 \\ & + \sum_{g=1}^G \sum_{i=1}^I \sum_{j=i}^I w_{2ig} (\sigma_{ijg} - \psi_g (\lambda_{i0} + \tilde{\lambda}_{ig}) (\lambda_{j0} + \tilde{\lambda}_{jg}) - \Phi_{ig} \mathbf{1}_{\{i=j\}})^2 \\ & + \sum_{g=1}^G \sum_{i=1}^I \mathcal{P}(\kappa_1, \tilde{\nu}_{ig}) + \sum_{g=1}^G \sum_{i=1}^I \mathcal{P}(\kappa_2, \tilde{\lambda}_{ig}), \end{aligned} \quad (54)$$

where  $\kappa_1$  and  $\kappa_2$  are regularization parameters for group-specific residual intercepts  $\tilde{\nu}_{ig}$  and residual loadings  $\tilde{\lambda}_{ig}$ , respectively. Using the additive representation of MNI (45), we get the following identification constraint for determining the group mean  $\alpha_g$

$$\sum_{g=1}^G \sum_{i=1}^I w_{1ig} (\nu_{ig}^* - \tilde{\nu}_{ig} + \alpha_g (\lambda_{ig}^* - \tilde{\lambda}_{ig})) = 0. \quad (55)$$

The condition (55) means that MNI effects  $\nu_{ig}^*$  and  $\lambda_{ig}^*$  cancel out on average, where the average is computed mainly on those effects that are set to zero in regularized ML.

As an alternative approach, fused regularized ML estimation can be employed. In this approach, all item intercepts and item loadings are estimated group-wise, and identification is ensured using a fused penalty function. The fitting function is defined as

$$\begin{aligned}
F_{\text{fusedreg}}(\Theta) = & \sum_{g=1}^G \sum_{i=1}^I w_{1ig} (\mu_{ig} - \alpha_g \check{\lambda}_{ig} - \check{v}_{ig})^2 \\
& + \sum_{g=1}^G \sum_{i=1}^I \sum_{j=i}^I w_{2ig} (\sigma_{ijg} - \psi_g \check{\lambda}_{ig} \check{\lambda}_{jg} - \phi_{ig} \mathbf{1}_{\{i=j\}})^2 \\
& + \sum_{g=1}^G \sum_{h=g+1}^G \sum_{i=1}^I \mathcal{P}(\kappa_1, \check{v}_{ig} - \check{v}_{ih}) + \sum_{g=1}^G \sum_{h=g+1}^G \sum_{i=1}^I \mathcal{P}(\kappa_2, \check{\lambda}_{ig} - \check{\lambda}_{ih}).
\end{aligned}$$

With additive MNI effects (see (45)), the identification constraint for the group mean is given by

$$\sum_{i=1}^I w_{1ig} (\alpha_g (\lambda_{i0} + \lambda_{ig}^* - \check{\lambda}_{ig}) + v_{i0} + v_{ig}^* - \check{v}_{ig}) = 0.$$

In the tau-congeneric measurement model, BAMI [41,47] uses normal prior distributions for intercept differences  $v_{ig} - v_{ih}$  and loading differences  $\lambda_{ig} - \lambda_{ih}$  [51]. Therefore, BAMI can be viewed as fused regularized ML estimation with a ridge penalty. Using the finding of Battauz [39] and the same reasoning as in Section 3.3, it is evident that BAMI as fused regularization with a ridge penalty function is equivalent to regularized ML estimation in (54) that involves the estimation of common item intercepts  $\nu_0$  and common item loadings  $\Lambda_0$ . As argued in Section 3.3, regularized ML estimation with a ridge penalty can be quite close to DWLS with appropriate weights (and, hence, ML) in terms of estimated group means and estimated variances because only shrinkage for otherwise nonidentified residual MNI effects for intercepts and loadings is introduced. Consequently, BAMI will provide similar results compared to ML estimation.

As also discussed in Section 3.3 for the tau-equivalent model, the output of BAMI can be used in a subsequent IA estimation [45]. Using the same steps as in Section 3.3 for the derivations, an identification constraint similar to (36) can be obtained.

## 6. Discussion

In this article, we have argued that joint estimation, linking, and regularized ML estimation in the tau-equivalent and the tau-congeneric model can provide similar if not identical estimates in the violation of MI if an appropriate loss function  $\rho$  in joint estimation or linking is used. In the violation of MI, it is important to emphasize that researchers can use arbitrary identification constraints to determine group means. Resulting estimates depend on the chosen weights and loss function or the used penalty functions in regularized ML estimation. Therefore, researchers use implicit definitions for identification constraints on effects that quantify MNI by choosing a particular fitting function. The wisdom under applied researchers that partial invariance is necessary for determining group-mean comparisons [24] is unsound because it would imply that a particular loss function should always be preferred in practice.

It is important to emphasize that choosing of a particular loss function weighs discrepancies between sample input data (i.e., means and covariances in factor analysis) and assumed population parameters. Two error types must be distinguished: sampling error and model error. The former sampling error can be reduced in large samples while the latter model error (i.e., MNI in multiple-group factor analysis) does not vanish in large samples. Certain types of misspecifications can be downweighted by utilizing a model-robust loss function  $\rho$ . Consequently, estimated model parameters are not influenced by some misspecifications. We think that the preference of ML estimation factor analysis (and in structural equation modeling in general) is misguided because ML can only be the most efficient estimation method in (very) large samples and if the model of interest is correctly specified. Because MI is rare as unicorns in applications of multiple-group factor analysis, we cannot imagine many situations in which a robust loss function should not be preferred. Note that we are discussing robustness in the sense of model misspecification in

modelled means and covariances. In contrast, robust factor analysis is mainly devoted to the misspecification of the multivariate normal distribution [52].

Joint estimation in the case of measurement invariance still seems to be the most frequent approach. In contrast to regularization approaches, joint estimation does not include additional parameters for model deviations. Hence, joint estimation is less computationally demanding than regularization. Linking approaches are typically implemented in a two-step method. Because the one-dimensional factor model is estimated separately for each group in a first step, the linking approach might be less prone to convergence issues or ill-defined parameter estimates. However, group-wise estimation of the one-dimensional factor model might require a sufficiently large sample size. Hence, linking methods could result in less stable estimates than joint estimation or regularization.

Our arguments in this paper are based on fitting vectors of means and covariances that are computed for factor analysis of continuous items (i.e., assuming a multivariate normal distribution). The arguments likely generalize to the fitting of vectors of thresholds and polychoric correlation matrices for factor analysis of ordinal data [53]. Future research could investigate violations of measurement invariance in the one-dimensional factor model for a continuous covariate instead of a finite number of groups  $g = 1, \dots, G$  [54].

We limited our discussion to the most popular estimation methods of the multiple-group one-dimensional factor model (i.e., joint estimation, linking, and regularized estimation) in the violation of MI. More flexible handling of MNI has recently been proposed using deep learning methods [55]. These fitting functions also imply identification constraints for MNI effects. Because it is our conviction that all fitted models are grossly misspecified (and not only misspecified to a certain degree), there is no reason to believe that more complex models will provide more valid estimates for group comparisons. In contrast, researchers purposely choose fitting function that describe a complex dataset and define a pseudo-true parameter through the choice of this fitting function. It is likely not reasonable to talk about true parameters (i.e., true group means and true group variances) without explicitly mentioning identification constraints.

## 7. Conclusions

This article presented a formal analysis of different estimation methods in the violation of measurement invariance. We have shown how different fitting functions result in implied identification constraints on parameters that characterize the extent of measurement invariance. In our view, the choice of fitting functions should be mainly made regarding the weighing of model deviations because it is unlikely in practical applications that the doctrine of measurement invariance exactly holds.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

BAMI	Bayesian approximate measurement invariance
DWLS	diagonally weighted least squares
HL	Haberman linking
IA	invariance alignment
ML	maximum likelihood
MI	measurement invariance

MNI	measurement noninvariance
PI	partial invariance
ULS	unweighted least squares
WLS	weighted least squares

## References

1. Bartholomew, D.J. The foundations of factor analysis. *Biometrika* **1984**, *71*, 221–232. [[CrossRef](#)]
2. Jöreskog, K.G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **1969**, *34*, 183–202. [[CrossRef](#)]
3. Bechger, T.M.; Maris, G. A statistical test for differential item pair functioning. *Psychometrika* **2015**, *80*, 317–340. [[CrossRef](#)] [[PubMed](#)]
4. Schulze, D.; Pohl, S. Finding clusters of measurement invariant items for continuous covariates. *Struct. Equ. Model. A Multidiscip. J.* **2021**, *28*, 219–228. [[CrossRef](#)]
5. Robitzsch, A. Robust and nonrobust linking of two groups for the Rasch model with balanced and unbalanced random DIF: A comparative simulation study and the simultaneous assessment of standard errors and linking errors with resampling techniques. *Symmetry* **2021**, *13*, 2198. [[CrossRef](#)]
6. Meredith, W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **1993**, *58*, 525–543. [[CrossRef](#)]
7. Millsap, R.E. *Statistical Approaches to Measurement Invariance*; Routledge: New York, NY, USA, 2011. [[CrossRef](#)]
8. Davidov, E.; Meuleman, B. Measurement invariance analysis using multiple group confirmatory factor analysis and alignment optimisation. In *Invariance Analyses in Large-Scale Studies*; van de Vijver, F.J.R., Ed.; OECD: Paris, France, 2019; pp. 13–20.
9. Vandenberg, R.J.; Lance, C.E. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* **2000**, *3*, 4–70. [[CrossRef](#)]
10. Wicherts, J.M.; Dolan, C.V. Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educ. Meas. Issues Pract.* **2010**, *29*, 39–47. [[CrossRef](#)]
11. Jöreskog, K.G. Statistical analysis of sets of congeneric tests. *Psychometrika* **1971**, *36*, 109–133. [[CrossRef](#)]
12. Lewis, C. Selected topics in classical test theory. In *Handbook of Statistics*; Rao, C.R., Sinharay, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 26, pp. 29–43. [[CrossRef](#)]
13. Mellenbergh, G.J. A unidimensional latent trait model for continuous item responses. *Multivariate Behav. Res.* **1994**, *29*, 223–236. [[CrossRef](#)]
14. Steyer, R. Models of classical psychometric test theory as stochastic measurement models: Representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika* **1989**, *3*, 25–60. Available online: <https://bit.ly/3Js7N3S> (accessed on 20 February 2022).
15. Jöreskog, K.G.; Olsson, U.H.; Wallentin, F.Y. *Multivariate Analysis with LISREL*; Springer: Basel, Switzerland, 2016 [[CrossRef](#)]
16. Kolenikov, S. Biases of parameter estimates in misspecified structural equation models. *Sociol. Methodol.* **2011**, *41*, 119–157. [[CrossRef](#)]
17. Savalei, V. Understanding robust corrections in structural equation modeling. *Struct. Equ. Model. A Multidiscip. J.* **2014**, *21*, 149–160. [[CrossRef](#)]
18. MacCallum, R.C.; Browne, M.W.; Cai, L. Factor analysis models as approximations. In *Factor Analysis at 100*; Cudeck, R., MacCallum, R.C., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2007; pp. 153–175. [[CrossRef](#)]
19. Siemsen, E.; Bollen, K.A. Least absolute deviation estimation in structural equation modeling. *Sociol. Methods Res.* **2007**, *36*, 227–265. [[CrossRef](#)]
20. Van Kesteren, E.J.; Oberski, D.L. Flexible extensions to structural equation models using computation graphs. *Struct. Equ. Model. A Multidiscip. J.* **2021**. [[CrossRef](#)]
21. Robitzsch, A.  $L_p$  loss functions in invariance alignment and Haberman linking with few or many groups. *Stats* **2020**, *3*, 246–283. [[CrossRef](#)]
22. Yuan, K.H.; Marshall, L.L.; Bentler, P.M. Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociol. Methodol.* **2003**, *33*, 241–265. [[CrossRef](#)]
23. Davies, P.L. *Data Analysis and Approximate Models*; CRC Press: Boca Raton, FL, USA, 2014. [[CrossRef](#)]
24. Byrne, B.M.; Shavelson, R.J.; Muthén, B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol. Bull.* **1989**, *105*, 456–466. [[CrossRef](#)]
25. Davies, P.L.; Terbeck, W. Interactions and outliers in the two-way analysis of variance. *Ann. Statist.* **1998**, *26*, 1279–1305. [[CrossRef](#)]
26. Kolen, M.J.; Brennan, R.L. *Test Equating, Scaling, and Linking*; Springer: New York, NY, USA, 2014. [[CrossRef](#)]
27. Battauz, M. Multiple equating of separate IRT calibrations. *Psychometrika* **2017**, *82*, 610–636. [[CrossRef](#)]
28. Haberman, S.J. *Linking Parameter Estimates Derived from An Item Response Model through Separate Calibrations*; Research Report No. RR-09-40; Educational Testing Service: Princeton, NJ, USA, 2009. [[CrossRef](#)]
29. Asparouhov, T.; Muthén, B. Multiple-group factor analysis alignment. *Struct. Equ. Model. A Multidiscip. J.* **2014**, *21*, 495–508. [[CrossRef](#)]

30. Pokropek, A.; Lüdtke, O.; Robitzsch, A. An extension of the invariance alignment method for scale linking. *Psychol. Test Assess. Model.* **2020**, *62*, 303–334. Available online: <https://bit.ly/2UEp9GH> (accessed on 20 February 2020).
31. Schechter, E. *Handbook of Analysis and Its Foundations*; Academic Press: San Diego, CA, USA, 1996. [CrossRef]
32. Von Davier, M.; von Davier, A.A. A unified approach to IRT scale linking and scale transformations. *Methodology* **2007**, *3*, 115–124. [CrossRef]
33. Geminiani, E.; Marra, G.; Moustaki, I. Single- and multiple-group penalized factor analysis: A trust-region algorithm approach with integrated automatic multiple tuning parameter selection. *Psychometrika* **2021**, *86*, 65–95. [CrossRef]
34. Huang, P.H. A penalized likelihood method for multi-group structural equation modelling. *Br. J. Math. Stat. Psychol.* **2018**, *71*, 499–522. [CrossRef]
35. Li, X.; Jacobucci, R.; Ammerman, B.A. Tutorial on the use of the regsem package in R. *Psych* **2021**, *3*, 579–592. [CrossRef]
36. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press: Boca Raton, FL, USA, 2015. [CrossRef]
37. She, Y.; Owen, A.B. Outlier detection using nonconvex penalized regression. *J. Am. Stat. Assoc.* **2011**, *106*, 626–639. [CrossRef]
38. Yu, C.; Yao, W. Robust linear regression: A review and comparison. *Commun. Stat. Simul. Comput.* **2017**, *46*, 6261–6282. [CrossRef]
39. Battauz, M. Regularized estimation of the four-parameter logistic model. *Psych* **2020**, *2*, 269–278. [CrossRef]
40. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 91–108. [CrossRef]
41. Muthén, B.; Asparouhov, T. Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychol. Methods* **2012**, *17*, 313–335. [CrossRef]
42. Pokropek, A.; Schmidt, P.; Davidov, E. Choosing priors in Bayesian measurement invariance modeling: A Monte Carlo simulation study. *Struct. Equ. Model.* **2020**, *27*, 750–764. [CrossRef]
43. Van de Schoot, R.; Kluytmans, A.; Tummers, L.; Lugtig, P.; Hox, J.; Muthén, B. Facing off with scylla and charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Front. Psychol.* **2013**, *4*, 770. [CrossRef]
44. Van Erp, S.; Oberski, D.L.; Mulder, J. Shrinkage priors for Bayesian penalized regression. *J. Math. Psychol.* **2019**, *89*, 31–50. [CrossRef]
45. Arts, I.; Fang, Q.; Meitinger, K.; van de Schoot, R. Approximate measurement invariance of willingness to sacrifice for the environment across 30 countries: The importance of prior distributions and their visualization. *Front. Psychol.* **2021**, *12*, 624032. [CrossRef]
46. De Bondt, N.; Van Petegem, P. Psychometric evaluation of the overexcitability questionnaire-two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Front. Psychol.* **2015**, *6*, 1963. [CrossRef]
47. Muthén, B.; Asparouhov, T. Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociol. Methods Res.* **2018**, *47*, 637–664. [CrossRef]
48. Chen, Y.; Li, C.; Xu, G. DIF statistical inference and detection without knowing anchoring items. *arXiv* **2021**, arXiv:2110.11112.
49. Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models: A Modern Perspective*; Chapman and Hall: New York, NY, USA; CRC: Boca Raton, FL, USA, 2006. [CrossRef]
50. Robitzsch, A. A comparison of linking methods for two groups for the two-parameter logistic item response model in the presence and absence of random differential item functioning. *Foundations* **2021**, *1*, 116–144. [CrossRef]
51. Lek, K.; van de Schoot, R. Bayesian approximate measurement invariance. In *Invariance Analyses in Large-Scale Studies*; van de Vijver, F.J.R., Ed.; OECD: Paris, France, 2019; pp. 21–35.
52. Yuan, K.H.; Bentler, P.M. Robust procedures in structural equation modeling. In *Handbook of Latent Variable and Related Models*; Lee, S.Y., Ed.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 367–397. [CrossRef]
53. Cai, L.; Moustaki, I. Estimation methods in latent variable models for categorical outcome variables. In *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test*; Irving, P., Booth, T., Hughes, D.J., Eds.; Wiley: New York, NY, USA, 2018; pp. 253–277. [CrossRef]
54. Hildebrandt, A.; Wilhelm, O.; Robitzsch, A. Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Sociol. Methods Res.* **2009**, *16*, 87–102.
55. Pokropek, A.; Pokropek, E. Deep neural networks for detecting statistical model misspecifications. The case of measurement invariance. *arXiv* **2022**, arXiv:2107.12757.