Effect Comparison in Multilevel Structural Equation Models with Non-Metric Outcomes

Christoph Kern* Petra Stein†

Abstract
This study discusses difficulties of effect comparisons in multilevel structural equation models with non-metric outcomes, such as nonlinear dyadic mixed-effects regression. In these models, the fixation of the level-1 error variances induces substantial drawbacks in the context of effect comparisons which align with the well-known problems of standard single- and multilevel nonlinear models. Specifically, the level-1 and level-2 coefficients as well as the level-2 variance components are implicitly rescaled by the amount of unobserved level-1 residual variation and thus may apparently differ across (and within) equations despite of true effect equality. Against this background, the present study discusses a multilevel extension of the method proposed by Sobel and Arminger (1992) with which potential differences in level-1 residual variation can be taken into account through the specification of non-linear parameter constraints. The problems of effect comparisons in multilevel probit SEM's and the proposed correction method are exemplified with a simulation study.

Key Words: multilevel probit SEM, effect comparison, non-linear constraints

1. Introduction

Nonlinear regression methods – such as logit and probit models – involve a number of distinctive features, resulting in substantial difficulties in the context of effect comparisons between different model specifications and/ or groups. Specifically, the fixed error variance in nonlinear models (logit $V(\varepsilon) = \frac{\pi^2}{3}$, probit $V(\varepsilon) = 1$) leads to implicitly rescaled coefficients, which additionally depend on the extent of unobserved residual variation such that naive effect comparisons can lead to false conclusions. In a multilevel setup with categorical y variables, the same holds true for the respective level-1 residual variation. Against this background, various correction techniques have been introduced for effect comparisons in single- as well as multilevel regression models with non-metric outcomes. In the case of standard logit and probit models, e.g. the usage of y^*-standardized coefficients, average marginal effects (AME's) or the KHB method have been proposed for effect comparisons across different model specifications (Karlson et al. 2012, Mood 2010), whereas correction methods in the context of group comparisons are discussed by Allison (1999) and Williams (2009, 2010, cf. also Long 2009). Concerning multilevel regression with non-metric dependent variables, effect comparisons between different nested models can be carried out through the usage of the rescaling procedures proposed by Hox (2010) and Bauer (2009, cf. also Fielding 2004).

Furthermore, it can be shown that problems of effect comparisons also arise in nonlinear models with multiple outcomes, such as dyadic logit and probit regression in a structural equation modeling framework (Stein / Pavetic 2013, Kern / Stein 2015). In this case, effect comparisons between and within equations are likewise complicated by the implicit fixation of error variances, leading to potential drawbacks when the assumption of equal residual variation is not met. A potential solution in the case of single-level probit SEM's

*University of Duisburg-Essen, Lotharstr. 65, 47057 Duisburg
†University of Duisburg-Essen, Lotharstr. 65, 47057 Duisburg

has been proposed by Sobel and Arminger (1992) through the specification of non-linear parameter constraints.

Combining the multilevel and multivariate perspective, the present study discusses difficulties of effect comparisons in multilevel structural equation models with non-metric outcomes. In this context, it can be demonstrated that the level-1 and level-2 coefficients as well as the level-2 variance components are implicitly rescaled by the amount of unobserved level-1 residual variation and thus may apparently differ across (and within) equations despite of true effect equality. Consequently, a correction method is also needed in this case. Against this background, the present study discusses a multilevel extension of the Sobel and Arminger (1992) approach, with which potential differences in level-1 residual variation can be taken into account within the specification of equality constraints. This approach thereby enables the researcher to impose "robust" equality restrictions for effect comparisons between and within level-1 and level-2 equations. However, it is important to note that this approach rests on the initial assumption of true effect equality between equations and therefore involves the same limitations as the corresponding procedure in the single-level case (Kern / Stein 2015).

In the following, section 2 outlines the issue of implicit rescaling in mixed-effects models with multiple non-metric outcomes and introduces non-linear parameter constraints in the multilevel SEM context. In section 3, the problems of effect comparisons in multilevel probit SEM's and the proposed correction method are exemplified with a simulation study. Finally, limitations of the proposed approach are discussed in the last section (4) of this paper.

2. Effect comparison in mixed-effects probit SEM's

2.1 Implicit rescaling

Denoting y^*_{pij} the p–th latent response variable ($p = 1, ..., P$) for individual $i = 1, ..., N_j$ in cluster $j = 1, ..., C$, one can set up a two-level structural equation model by decomposing y^*_{pij} into a within (y^*_{wpij}) and a between (y^*_{bpj}) component (Asparouhov / Muthén 2007):

$$y^*_{pij} = y^*_{wpij} + y^*_{bpj} \tag{1}$$

On this basis, structural equations and measurement equations can be specified on both levels. On level-1, these equations are given by:

$$\mathbf{y}^*_{wij} = \boldsymbol{\Lambda}_w \boldsymbol{\eta}_{wij} + \boldsymbol{\epsilon}_{wij} \tag{2}$$

$$\boldsymbol{\eta}_{wij} = \mathbf{B}_w \boldsymbol{\eta}_{wij} + \boldsymbol{\Gamma}_w \mathbf{x}_{wij} + \boldsymbol{\zeta}_{wij} \tag{3}$$

Likewise, the level-2 equations are:

$$\mathbf{y}^*_{bj} = \mathbf{v}_b + \boldsymbol{\Lambda}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\epsilon}_{bj} \tag{4}$$

$$\boldsymbol{\eta}_{bj} = \boldsymbol{\alpha}_b + \mathbf{B}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\Gamma}_b \mathbf{x}_{bj} + \boldsymbol{\zeta}_{bj} \tag{5}$$

The following sections focus solely on the structural part, thus it is assumed that each of the latent η_{ij} variables is perfectly measured by one y^*_{ij} variable ($\boldsymbol{\Lambda}_w = \mathbf{I}$, $\boldsymbol{\Lambda}_b = \mathbf{I}$, $\mathbf{v}_b = \mathbf{0}$, $\boldsymbol{\epsilon}_{wij} = \mathbf{0}$ and $\boldsymbol{\epsilon}_{bj} = \mathbf{0}$.).

On this basis, $\boldsymbol{\eta}_{wij}$ represents a $(M \times 1)$ vector of latent dependent variables on level-1 (within-part of $\boldsymbol{\eta}_{ij}$), $\mathbf{B}_w$ is a $(M \times M)$ matrix of regression coefficients covering relationships between the η_{wij} variables, $\boldsymbol{\Gamma}_w$ contains a $(M \times Q_1)$ matrix of within regression slopes, $\mathbf{x}_{wij}$ is a $(Q_1 \times 1)$ vector of (observed) independent variables on level-1 and $\boldsymbol{\zeta}_{wij}$

is a $(M \times 1)$ vector of level-1 residuals with $\boldsymbol{\zeta}_{wij} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_w)$. Correspondingly, $\boldsymbol{\eta}_{bj}$ represents a $(M \times 1)$ vector of latent dependent variables on level-2 (between-part of $\boldsymbol{\eta}_{ij}$), $\boldsymbol{\alpha}_b$ is a $(M \times 1)$ vector of level-2 intercepts, $\mathbf{B}_b$ is a $(M \times M)$ matrix of coefficients covering relationships between the η_{bj} variables, $\boldsymbol{\Gamma}_b$ contains a $(M \times Q_2)$ matrix of between regression slopes, $\mathbf{x}_{bj}$ is a $(Q_2 \times 1)$ vector of (observed) independent variables on level-2 and $\boldsymbol{\zeta}_{bj}$ is a $(M \times 1)$ vector of level-2 residuals with $\boldsymbol{\zeta}_{bj} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_b)$. With this setup, the diagonal elements of $\boldsymbol{\Psi}_b$ allow the specification of random intercepts for $\boldsymbol{\eta}_{bj}$.

Without a measurement model, the reduced form equations for the y^* variables are

$$\mathbf{y}^*_{wij} = (\mathbf{I} - \mathbf{B}_w)^{-1}\boldsymbol{\Gamma}_w\mathbf{x}_{wij} + \boldsymbol{\varepsilon}_{wij} \tag{6}$$

$$\mathbf{y}^*_{bj} = (\mathbf{I} - \mathbf{B}_b)^{-1}\boldsymbol{\alpha}_b + (\mathbf{I} - \mathbf{B}_b)^{-1}\boldsymbol{\Gamma}_b\mathbf{x}_{bj} + \boldsymbol{\varepsilon}_{bj} \tag{7}$$

with $\boldsymbol{\varepsilon}_{wij} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_w)$ and $\boldsymbol{\varepsilon}_{bj} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_b)$. In this case, the covariance matrices $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_b$ are given by:

$$\boldsymbol{\Sigma}_w = (\mathbf{I} - \mathbf{B}_w)^{-1}\boldsymbol{\Psi}_w(\mathbf{I} - \mathbf{B}_w)'^{-1} \tag{8}$$

$$\boldsymbol{\Sigma}_b = (\mathbf{I} - \mathbf{B}_b)^{-1}\boldsymbol{\Psi}_b(\mathbf{I} - \mathbf{B}_b)'^{-1} \tag{9}$$

With categorical observed y variables, restrictions have to be introduced into the model structure. Relating y and y^* with a threshold model, $\boldsymbol{\alpha}_b = \mathbf{0}$ has to be imposed since a simultaneous specification of all intercepts and thresholds leads to identification issues.

Most importantly, additional standardizations have to be made with respect to $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Psi}_w$. First, a diagonal matrix $\boldsymbol{\Delta}_w$ with elements $1/\sqrt{\Sigma_{wmm}}$ is introduced in order to normalize $\boldsymbol{\Sigma}_w$ so that $\text{diag}(\boldsymbol{\Sigma}^*_w) = \mathbf{I}$ (resulting in a multivariate probit model for $\mathbf{y}^*_{ijs}$). In this case, $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}^*_w$ as well as $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}^*_b$ are related as follows:

$$\boldsymbol{\Sigma}^*_w = \boldsymbol{\Delta}_w\boldsymbol{\Sigma}_w\boldsymbol{\Delta}_w \tag{10}$$

$$\boldsymbol{\Sigma}^*_b = \boldsymbol{\Delta}_w\boldsymbol{\Sigma}_b\boldsymbol{\Delta}_w \tag{11}$$

Secondly, identifying assumptions have to be made concerning the unobservable error variances of $\boldsymbol{\Psi}_w$ within $\boldsymbol{\Sigma}_w$, whereas in the following the standardization $\text{diag}(\boldsymbol{\Psi}^*_w) = \mathbf{I}$ is imposed (Theta parameterization; Muthén / Asparouhov 2002). More explicitly, the unrestricted $\boldsymbol{\Psi}_w$ matrix

$$\boldsymbol{\Psi}_w = \begin{bmatrix} \psi_{w11} & \psi_{w12} & \cdots & \psi_{w1m} \\ \psi_{w21} & \psi_{w22} & \cdots & \psi_{w2m} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{wm1} & \psi_{wm2} & \cdots & \psi_{wmm} \end{bmatrix} \tag{12}$$

is replaced by

$$\boldsymbol{\Psi}^*_w = \begin{bmatrix} 1 & \psi^*_{w12} & \cdots & \psi^*_{w1m} \\ \psi^*_{w21} & 1 & \cdots & \psi^*_{w2m} \\ \vdots & \vdots & \ddots & \vdots \\ \psi^*_{wm1} & \psi^*_{wm2} & \cdots & 1 \end{bmatrix} \tag{13}$$

through the (implicit) introduction of

$$\boldsymbol{\Delta}^*_w = \begin{bmatrix} \sigma_{w1} & 0 & \cdots & 0 \\ 0 & \sigma_{w2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{wm} \end{bmatrix}^{-1} . \tag{14}$$

In this context, $\mathbf{\Delta}_w^*$ contains the inverted standard deviations of the unobserved level-1 residuals with $\sigma_{wm} = \sqrt{\psi_{wmm}}$. The relation between $\mathbf{\Psi}_w$ and $\mathbf{\Psi}_w^*$ is given by:

$$\mathbf{\Psi}_w^* = \mathbf{\Delta}_w^* \mathbf{\Psi}_w \mathbf{\Delta}_w^* \tag{15}$$

Applying the first standardization through the introduction of $\mathbf{\Delta}_w$, the two-level model for the standardized y^* variables follows

$$\mathbf{y}_{wijs}^* = \mathbf{\Delta}_w (\mathbf{I} - \mathbf{B}_w)^{-1} \mathbf{\Gamma}_w \mathbf{x}_{wij} + \varepsilon_{wij}^* \tag{16}$$

$$\mathbf{y}_{bjs}^* = \mathbf{\Delta}_w (\mathbf{I} - \mathbf{B}_b)^{-1} \mathbf{\Gamma}_b \mathbf{x}_{bj} + \varepsilon_{bj}^* \tag{17}$$

with $\varepsilon_{wij}^* \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_w^*)$ and $\varepsilon_{bj}^* \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_b^*)$.

Furthermore – building on the previous differentiation between $\mathbf{\Psi}_w$ and $\mathbf{\Psi}_w^*$ – it becomes clear that (16) and (17) are subject to a second standardization. With the fixation of the level-1 error variances of $\mathbf{\Psi}_w$ within $\mathbf{\Sigma}_w^*$, the elements of (16) and (17) are implicitly rescaled through the additional introduction of $\mathbf{\Delta}_w^*$. More specific, the relationships between the standardized and fully standardized matrices are given by (cf. Sobel / Arminger 1992, Stein / Pavetic 2013):

$$\mathbf{B}_w^* = \mathbf{\Delta}_w^* \mathbf{B}_w \mathbf{\Delta}_w^{*-1} \tag{18}$$

$$\mathbf{B}_b^* = \mathbf{\Delta}_w^* \mathbf{B}_b \mathbf{\Delta}_w^{*-1} \tag{19}$$

$$\mathbf{\Gamma}_w^* = \mathbf{\Delta}_w^* \mathbf{\Gamma}_w \tag{20}$$

$$\mathbf{\Gamma}_b^* = \mathbf{\Delta}_w^* \mathbf{\Gamma}_b \tag{21}$$

$$\mathbf{\Psi}_b^* = \mathbf{\Delta}_w^* \mathbf{\Psi}_b \mathbf{\Delta}_w^* \tag{22}$$

Thus, as a result of $\text{diag}(\mathbf{\Psi}_w^*) = \mathbf{I}$, the estimable level-1 and level-2 coefficients additionally depend on the amount of unobserved level-1 error variation.

2.2 Non-linear constraints

Using scalar notation, equation (18) implies that e.g. $\beta_{w31}^* = \frac{\sigma_{w1}\beta_{w31}}{\sigma_{w3}}$ and $\beta_{w32}^* = \frac{\sigma_{w2}\beta_{w32}}{\sigma_{w3}}$. Likewise, equation (20) corresponds to $\gamma_{w1q}^* = \frac{\gamma_{w1q}}{\sigma_{w1}}$ and $\gamma_{w2q}^* = \frac{\gamma_{w2q}}{\sigma_{w2}}$, while equation (21) implies that $\gamma_{b1q}^* = \frac{\gamma_{b1q}}{\sigma_{w1}}$ and $\gamma_{b2q}^* = \frac{\gamma_{b2q}}{\sigma_{w2}}$. On this basis, effect comparisons between ($\mathbf{\Gamma}_w^*$, $\mathbf{\Gamma}_b^*$) and within ($\mathbf{B}_w^*$, $\mathbf{B}_b^*$) equations can lead to false conclusions in models with differences in unobserved error variances across equations.

Building on Sobel / Arminger (1992), equality restrictions within the specified model must therefore be formulated in terms of the rescaled coefficients. For the hypothesis that e.g. $\gamma_{w1q} = \gamma_{w2q}$, it follows:

$$\begin{aligned} \gamma_{w1q}^* \sigma_{w1} &= \gamma_{w2q}^* \sigma_{w2} \\ \gamma_{w1q}^* &= \frac{\sigma_{w2}}{\sigma_{w1}} \gamma_{w2q}^* \\ \gamma_{w1q}^* &= \lambda \gamma_{w2q}^* \end{aligned} \tag{23}$$

Defining $\lambda = \frac{\sigma_{w2}}{\sigma_{w1}}$, the relation of the unobserved level-1 error variances of the first two equations is taken into account within the imposed equality restriction through the introduction of λ. Given λ, the hypothesis that e.g. $\beta_{w32} = \beta_{w31}$ can be replaced by:[1]

$$\begin{aligned} \beta_{w32}^* \frac{\sigma_{w3}}{\sigma_{w2}} &= \beta_{w31}^* \frac{\sigma_{w3}}{\sigma_{w1}} \\ \beta_{w32}^* &= \beta_{w31}^* \frac{\sigma_{w3}}{\sigma_{w1}} \frac{\sigma_{w2}}{\sigma_{w3}} \\ \beta_{w32}^* &= \lambda \beta_{w31}^* \end{aligned} \tag{24}$$

[1] Likewise for corresponding hypotheses concerning $\mathbf{B}_b^*$.

Finally, the hypothesis $\gamma_{b1q} = \gamma_{b2q}$ can be formulated in terms of:

$$\begin{aligned} \gamma^*_{b1q}\sigma_{w1} &= \gamma^*_{b2q}\sigma_{w2} \\ \gamma^*_{b1q} &= \frac{\sigma_{w2}}{\sigma_{w1}}\gamma^*_{b2q} \\ \gamma^*_{b1q} &= \lambda\gamma^*_{b2q} \end{aligned} \tag{25}$$

The empirical implementation of (23), (24) and (25) results in the specification of non-linear parameter constraints, which can be utilized in advanced SEM-software like Mplus (Muthén / Muthén 1998–2012). In this context, a constrained model specification can be compared with a less restrictive model in order to draw conclusions concerning the postulated hypothesis of effect equality for specific coefficients. Likewise, different sets of parameter restrictions can be tested in a stepwise manner.

3. Simulation study

3.1 Simulation setup

The problems of effect comparisons in multilevel probit SEM's and the proposed correction method can be exemplified with a simulation study. As a starting point, a simple multivariate mixed-effects model has been specified from which data for a hypothetical population has been generated. This population consists of 1000 level-2 units with 1000 level-1 cases, respectively. In line with the previous derivations, the data-generating model only includes a structural part, which in this case involves three dependent variables. Here, the first two η variables are each specified as a function of six level-1 and one level-2 predictors (with $x_{w1} \sim$ unif$(1, 10)$, ..., $x_{w6} \sim$ unif$(1, 10)$ and $x_b \sim$ unif$(1, 10)$), whereas the third η variable is in turn dependent on the first two η variables:

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{w31} & \beta_{w32} & 0 \end{bmatrix} \boldsymbol{\eta}_{wij} + \begin{bmatrix} \gamma_{w11} & \gamma_{w12} & \gamma_{w13} & \gamma_{w14} & \gamma_{w15} & \gamma_{w16} \\ \gamma_{w21} & \gamma_{w22} & \gamma_{w23} & \gamma_{w24} & \gamma_{w25} & \gamma_{w26} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}_{wij} + \begin{bmatrix} \gamma_{b11} \\ \gamma_{b21} \\ 0 \end{bmatrix} x_{bj} + \boldsymbol{\zeta}_{wij} + \boldsymbol{\zeta}_{bj}$$

On this basis, a effect structure has been specified which involves equal γ_w and γ_b coefficients between the first two equations and equal β_w coefficients within the third equation. More specific, the true parameter values are:

$$\begin{aligned} \gamma_{w11} = \gamma_{w24} &= -0.5 \\ \gamma_{w12} = \gamma_{w25} &= 0.3 \\ \gamma_{w13} = \gamma_{w26} &= 0.2 \\ \gamma_{w14} = \gamma_{w21} &= -0.4 \\ \gamma_{w15} = \gamma_{w22} &= 0.2 \\ \gamma_{w16} = \gamma_{w23} &= 0.1 \\ \gamma_{b11} = \gamma_{b21} &= -0.2 \\ \beta_{w31} = \beta_{w32} &= 0.4 \end{aligned}$$

Furthermore, error components have been added on both levels. The residual variables have been drawn from a multivariate normal distribution, such that $\boldsymbol{\zeta}_{wij} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^+_w)$ with

$$\boldsymbol{\Psi}^+_w = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$$

and $\boldsymbol{\zeta}_{bj} \sim N(\mathbf{0}, \boldsymbol{\Psi}_b)$ with

$$\boldsymbol{\Psi}_b = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}.$$

With this setup, the generated data set is based on a model structure with equal effects between and within level-1 and level-2 equations as well as equal level-2 error variances, but differences in level-1 residual variation across equations. Finally, three binary ("observed") y variables have been drawn from a Bernoulli distribution, whereas $P(y = 1)$ is determined by the cumulative distribution function (cdf) of the standard normal distribution at η:[2]

$$y_{ij} \sim B(1, \Phi(\eta_{ij}))$$

The simulation then consists of 1000 random draws from the population, whereas each sample contains 100 level-2 units with 50 level-1 cases, respectively. With each sample, a multilevel probit SEM (with respect to the binary y variables) has been estimated via WLSM using Mplus (Version 7; Muthén / Muthén 1998–2012).[3]

3.2 Results

The simulation process produced 1000 Mplus results, which constitute a distribution of estimates for each model component. The kernel density plots of these distributions are illustrated in Figure A.1, A.2 and A.3.

Concerning the effects of the x_w variables, the γ^*_w estimates of the first equation correspond to their predefined values (black lines in Figure A.1 and A.2). Specifically, the mean values of the respective distributions each approximate the true effects, whereas $\bar{\gamma}^*_{w11} = -0.512$, $\bar{\gamma}^*_{w12} = 0.306$, $\bar{\gamma}^*_{w13} = 0.203$, $\bar{\gamma}^*_{w14} = -0.412$, $\bar{\gamma}^*_{w15} = 0.206$ and $\bar{\gamma}^*_{w16} = 0.103$.[4] In contrast, the distributions of the γ^*_w coefficients of the second equation indicate substantial differences between the estimated coefficients and their predefined values in the data-generating model (red lines in Figure A.1 and A.2). Here, $\bar{\gamma}^*_{w24} = -0.387$, $\bar{\gamma}^*_{w25} = 0.233$, $\bar{\gamma}^*_{w26} = 0.154$, $\bar{\gamma}^*_{w21} = -0.309$, $\bar{\gamma}^*_{w22} = 0.155$ and $\bar{\gamma}^*_{w23} = 0.078$. These deviations can be ascribed to the elevated level-1 error variation in the second equation, which in the multivariate probit case results in rescaled γ^*_{w2q} coefficients. Thus, the Mplus estimates falsely imply effect differences across $\boldsymbol{\Gamma}_w$ equations as a result of differences in level-1 residual variation.

Similar mechanisms can be observed in Figure A.3. Here, the distribution of the γ^*_{b11} coefficients indicate a nearly unbiased x_b effect in the first equation ($\bar{\gamma}^*_{b11} = -0.206$), whereas the mean of the γ^*_{b21} distribution does not reflect the true parameter value ($\bar{\gamma}^*_{b21} = -0.144$; Figure A.3a). Furthermore, it becomes clear that the level-2 variance components are also affected by the specified level-1 error structure, which in this case results in an underestimated level-2 error variance in the second equation ($\bar{\psi}^*_{b11} = 0.250$, $\bar{\psi}^*_{b22} = 0.159$; Figure A.3b). Finally, Figure A.3c illustrates that the predefined differences in level-1 error variation between the first two equations also lead to apparent effect differences between the β^*_w estimates within the third equation ($\bar{\beta}^*_{w31} = 0.380$, $\bar{\beta}^*_{w32} = 0.471$). These coefficients

[2]Consequently, the complete level-1 error structure is given by $\boldsymbol{\Psi}_w = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1.75 & 0 \\ 0 & 0 & 1.25 \end{bmatrix}$.

[3]The simulation process has been implemented using R (Version 3.1.1; R Core Team 2014) and the R package "MplusAutomation" (Hallquist / Wiley 2014).

[4]Deviations from the true parameter values are in this case a result of the defined model structure with $\mathbf{B}_w \neq \mathbf{0}$ and the specification of $\boldsymbol{\Psi}^+_w$ as a diagonal matrix.

are subject to two (potential) standardizations: On the one hand, both estimates are scaled downwards as a result of the (modestly) elevated level-1 residual variation of the third equation, whereas β^*_{w32} is additionally scaled upwards through the elevated level-1 error variance of equation two ($\beta^*_{w31} = \frac{\sigma_{w1}\beta_{w31}}{\sigma_{w3}}$, $\beta^*_{w32} = \frac{\sigma_{w2}\beta_{w32}}{\sigma_{w3}}$; cf. section 2). As a result, the β^*_w coefficients cannot be compared directly within the third equation.

The results of a naive and a "λ-corrected" approach for effect comparisons between the β^*_{w31} and β^*_{w32} coefficients are illustrated in Table 1. First, 1000 fully restricted Mplus models have been estimated, in which all γ^*_w and γ^*_b coefficients have been constraint to be equal across equations and the naive restriction $\beta^*_{w32} = \beta^*_{w31}$ within the third equation has been imposed (Model 1 in Table 1a). Subsequently, the latter restriction has been relaxed in Model 2 such that both model specifications can be compared through (1000) SB-corrected χ^2 difference tests (Satorra / Bentler 2001). It becomes clear that this approach results in a mean SB-corrected χ^2 difference of 6.381, which in 59.0% of the conducted χ^2 difference tests leads to the false conclusion of significant effect differences between β_{w31} and β_{w32}.

In contrast, Table 1b displays the results of (1000) SB-corrected χ^2 difference tests between Model 3 and Model 4, in which the former restrictions have now been implemented through the specification of non-linear parameter constraints (cf. section 2.2.). On this basis, it can be seen that the relaxation of $\beta^*_{w32} = \lambda\beta^*_{w31}$ leads to a mean SB-corrected χ^2 difference of 2.148, such that in this case only 18.2% of the χ^2 difference tests suggest significant differences between the β_w coefficients. Thus, in the (ideal-typical) scenario at hand, the specification of non-linear constraints protects against the false rejection of the hypothesis of equal effects.

Table 1: Scaled χ^2 difference tests†

(a) Naive restrictions

Model	Restriction	$\bar{\chi}^2$	df	$\bar{\chi}^2_{sc}$ Diff.	Median p	Rate $p <= 0.05$
1	$\gamma^*_{w11} = \gamma^*_{w24} \cdots \gamma^*_{w13} = \gamma^*_{w26}$, $\gamma^*_{w14} = \gamma^*_{w21} \cdots \gamma^*_{w16} = \gamma^*_{w23}$, $\gamma^*_{b11} = \gamma^*_{b21}$, $\beta^*_{w32} = \beta^*_{w31}$	213.5	18			
2	$\gamma^*_{w11} = \gamma^*_{w24} \cdots \gamma^*_{w13} = \gamma^*_{w26}$, $\gamma^*_{w14} = \gamma^*_{w21} \cdots \gamma^*_{w16} = \gamma^*_{w23}$, $\gamma^*_{b11} = \gamma^*_{b21}$	207.4	17	6.38	0.027	0.590

(b) Non-linear constraints

Model	Restriction	$\bar{\chi}^2$	df	$\bar{\chi}^2_{sc}$ Diff.	Median p	Rate $p <= 0.05$	$\bar{\lambda}$
3	$\gamma^*_{w11} = \lambda\gamma^*_{w24} \cdots \gamma^*_{w13} = \lambda\gamma^*_{w26}$, $\gamma^*_{w14} = \lambda\gamma^*_{w21} \cdots \gamma^*_{w16} = \lambda\gamma^*_{w23}$, $\gamma^*_{b11} = \lambda\gamma^*_{b21}$, $\beta^*_{w32} = \lambda\beta^*_{w31}$	120.2	17				1.32
4	$\gamma^*_{w11} = \lambda\gamma^*_{w24} \cdots \gamma^*_{w13} = \lambda\gamma^*_{w26}$, $\gamma^*_{w14} = \lambda\gamma^*_{w21} \cdots \gamma^*_{w16} = \lambda\gamma^*_{w23}$, $\gamma^*_{b11} = \lambda\gamma^*_{b21}$	119.0	16	2.15	0.339	0.182	1.33

† Satorra / Bentler (2001)

4. Discussion

In this study, problems of effect comparisons have been discussed in the context of multilevel structural equation models with non-metric outcomes. It has been shown that the level-1 and level-2 coefficients as well as level-2 variance components are subject to implicit rescaling due to the fixation of residual variances on level-1. Thus, problems which arise within standard single- and multilevel nonlinear models extend – in correspondence with Kern / Stein (2015) – to the multivariate case, resulting in substantial difficulties of effect comparisons between equations, which may be of particular interest e.g. in a dyadic modeling framework (Kenny et al. 2006). Furthermore, comparisons of coefficients within equations are complicated concerning coefficients of the $\mathbf{B}_w$ and $\mathbf{B}_b$ matrices. Against this background, a multilevel extension of the method proposed by Sobel and Arminger (1992) has been discussed, with which potential differences in level-1 residual variation can be taken into account within the imposed equality restrictions through the specification of non-linear parameter constraints. Using simulated data, it has been shown that this approach enables the researcher to impose "robust" equality restrictions which are less sensitive to apparent effect differences of e.g. β_w^* coefficients.

However, it is important to note that this technique involves some limitations. Whereas in the previous derivations λ was considered to represent differences in level-1 residual variation between equations, it may likewise contain "true" effect differences in models with homogeneously different effects across equations. Since the present technique rests on the initial assumption of true effect equality (which may then be relaxed for specific coefficients in a stepwise manner), the hypothesis of equal effects may be falsely maintained in models with specific – i.e. homogeneous – patterns of effect differences, because these differences can be absorbed into λ besides differences in residual variation.[5] On the other hand, the implementation of non-linear constraints protects against the false rejection of the hypothesis of equal effects in nonlinear models with apparent effect differences across and within equations due to differences in level-1 residual variation. Therefore, the outlined procedure may be viewed as a flexible, supplementary tool in the context of multilevel structural equation modeling with categorical dependent variables. In any case, implicit rescaling should be considered as a substantial drawback in mixed-effects (as well as single-level) probit SEM's, demanding particular caution when coefficients of such models are compared.

REFERENCES

Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research*, 28(2):186–208.

Asparouhov, T. and Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. Technical report, Proceedings of the 2007 Joint Statistical Meetings. Alexandria, VA: American Statistical Association.

Bauer, D. J. (2009). A note on comparing the estimates of models for cluster-correlated or longitudinal data with binary or ordinal outcomes. *Psychometrika*, 74(1):97–105.

Fielding, A. (2004). Scaling for residual variance components of ordered category responses in generalized linear mixed multilevel models. *Quality and Quantity*, 38:425–433.

Hallquist, M. and Wiley, J. (2014). Mplusautomation: Automating Mplus model estimation and interpretation. Technical report, http://CRAN.R-project.org/package=MplusAutomation.

Hox, J. J. (2010). *Multilevel analysis – techniques and applications*. New York, NY: Routledge.

[5] For similar problems in the case of group comparisons with single-equation logit models, see Williams (2010).

Karlson, K. B., Holm, A., and Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology*, 42(1):286–313.

Kenny, D. A., Kashy, D. A., and Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Kern, C. and Stein, P. (2015). Comparing coefficients of nonlinear multivariate regression models between equations. *Survey Research Methods*, 9(3):159–167.

Long, J. S. (2009). Group comparisons in logit and probit using predicted probabilities. Technical report, Unpublished manuscript. http://www.indiana.edu/~jslsoc/files_research/groupdif/groupwithprobabilities/groups-with-prob-2009-06-25.pdf.

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1):67–82.

Muthén, B. O. and Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Technical report, Mplus Web Notes: No. 4.

Muthén, L. K. and Muthén, B. O. (1998–2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/.

Satorra, A. and Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4):507–514.

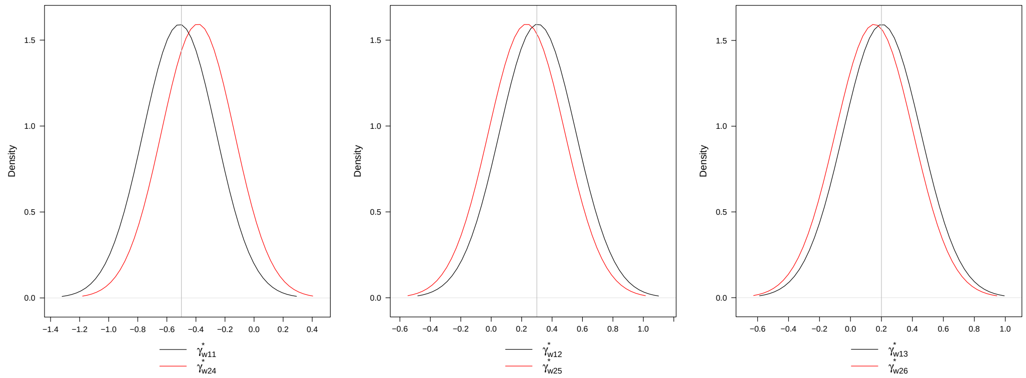
Sobel, M. E. and Arminger, G. (1992). Modeling household fertility decisions: A nonlinear simultaneous probit model. *Journal of the American Statistical Association*, 87(417):38–47.

Stein, P. and Pavetic, M. (2013). A nonlinear simultaneous probit-model for the investigation of decision-making processes: Modelling the process of setting up a family in partnerships. *Quality and Quantity*, 47(3):1717–1732.

Williams, R. (2009). Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research*, 37(4):531–559.

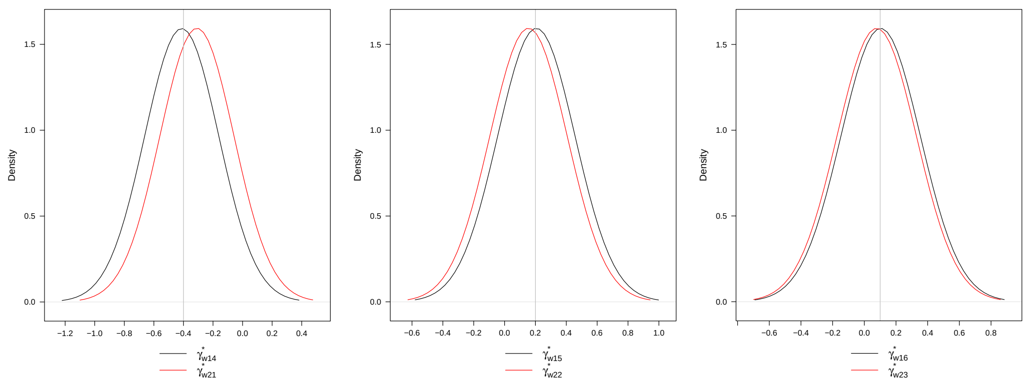
Williams, R. (2010). Fitting heterogeneous choice models with oglm. *The Stata Journal*, 10(4):540–567.

Appendix



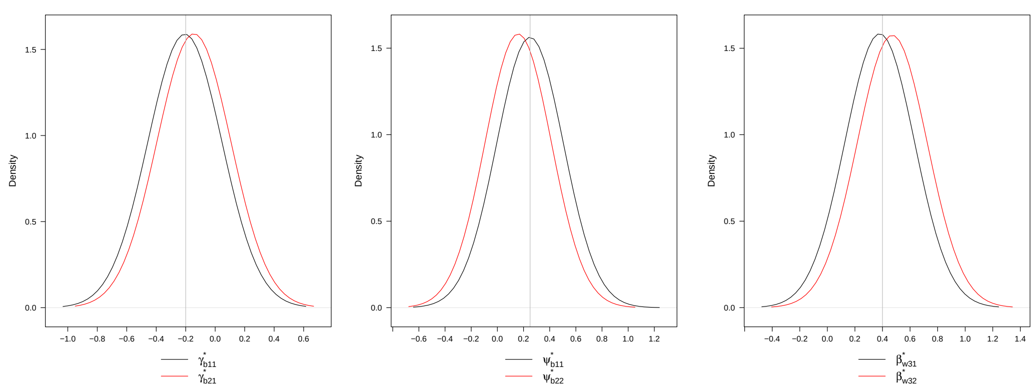
(a) $\gamma_{w11} = \gamma_{w24} = -0.5$ (b) $\gamma_{w12} = \gamma_{w25} = 0.3$ (c) $\gamma_{w13} = \gamma_{w26} = 0.2$

Figure A.1: Level-1 γ^* coefficients 1



(a) $\gamma_{w14} = \gamma_{w21} = -0.4$ (b) $\gamma_{w15} = \gamma_{w22} = 0.2$ (c) $\gamma_{w16} = \gamma_{w23} = 0.1$

Figure A.2: Level-1 γ^* coefficients 2



(a) $\gamma_{b11} = \gamma_{b21} = -0.2$ (b) $\psi_{b11} = \psi_{b22} = 0.25$ (c) $\beta_{w31} = \beta_{w32} = 0.4$

Figure A.3: Level-2 γ^* coefficients, ψ^*'s & β^*'s