Comparing coefficients of nonlinear multivariate regression models between equations

Christoph Kern*

Petra Stein*

07/2015

Abstract: The present study discusses the usage of non-linear constraints in regression models with multiple categorical outcomes. With this approach, effect differences between equations are made accessible to statistical tests while potential differences in residual variation are explicitly taken into account. In this context, it can be shown that the techniques reviewed by Williams (2010) are conjointly equivalent to the specification of non-linear constraints in multivariate regression models. However, the application of non-linear constraints extends these approaches into a structural equation modeling framework, which allows the researcher to address a broader range of research questions.

Keywords: non-linear constraints, structural equation modeling, effect comparison

1 Introduction

In comparison with standard linear models, the fixation of the unobserved error variance imposes several pitfalls in the application of nonlinear regression methods. Since the coefficients of these models are inevitably rescaled so that the respective residual variance equals 1 (probit) or $\pi^2/3$ (logit), a naive comparison of coefficients of nonlinear models between different groups can lead to false conclusions. While a variety of studies evaluate the difficulties of effect comparisons within the limits of single dependent variables, the present study focuses on nonlinear models with multiple outcomes, e.g. dyadic probit models in a structural equation modeling framework. In this context, it can be shown that comparisons of coefficients between different equations are invalid when

* University of Duisburg-Essen, Institute of Sociology (IfS).

the assumption of equal residual variances is not met. Thus, in this case even effect comparisons *within* the specified model are an error-prone task.

Against this background, the aim of this study is twofold: First, the usage of non-linear constraints in multivariate regression models with categorical outcomes is discussed as a means for testing effect differences across equations. In this context, it is demonstrated that the specification of non-linear constraints enables the researcher to impose a variety of equality constraints while taking potential differences in residual variation explicitly into account. Second, the outlined approach proposed in this paper is related to previous techniques that have been developed in the context of group comparisons with nonlinear models. Here, building on Williams (2010), it is shown that the techniques proposed by Allison (1999), Hauser and Andrew (2006) and Williams (2009) are conjointly equivalent to the specification of non-linear constraints in multivariate regression models. However, the latter extends these methods into a structural equation modeling framework, enabling the researcher the specification of more elaborated model structures.

This paper is organized as follows: The next section (2) provides a short review of previously proposed methods concerning effect comparisons in standard (i.e. single equation) logit and probit models. In the following section (3), difficulties of effect comparisons are discussed in the context of structural equation models with multiple categorical outcomes, resulting in the introduction of non-linear constraints. The empirical application of this technique is exemplified in section 4, whereas in subsection 4.1 non-linear constraints are related to the methods outlined in section 2 within the application of a dyadic logit model, followed by an example using an extended SEM-structure (subsection 4.2). The paper closes with a summarizing discussion concerning the advantages and limitations in the application of non-linear constraints (section 5).

2 Effect comparison in logit & probit models

The application of logit and probit models is characterized by a number of substantial, distinctive features. Due to implicit assumptions in the context of model identification, particularly the comparison of coefficients across different models and/ or groups involves some difficulties. These problems can be demonstrated referring to a latent response variable y^*, which is considered causal for the observed value of y within a threshold model (Allison 1999):

$$y_i = \begin{cases} 0 \;\; if \;\; -\infty \leq y_i^* < \tau \\ 1 \;\; if \quad\;\; \tau \leq y_i^* < \infty \end{cases} \tag{2.1}$$

Consequently, y_i equals 1 if y_i^* exceeds a threshold value τ, whereas τ is typically restricted a priori to zero ($\tau = 0$). The following linear model can be specified for y_i^*:

$$y_i^* = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_J x_{iJ} + \sigma\varepsilon_i \tag{2.2}$$

Here, ε_i is an error term with constant variance and σ is a non-fixed scale parameter. Thus, $\sigma\varepsilon_i$ allows for a variable error variance. However, since the scale of y_i^* is unknown and $\sigma\varepsilon_i$ therefore not determinable, the following model is based on implicit assumptions about ε_i:

$$g[P(y_i = 1)] = \beta_0^* + \beta_1^* x_{i1} + \ldots + \beta_J^* x_{iJ} \tag{2.3}$$

Assuming ε_i has a logistic distribution with $E(\varepsilon) = 0$ and $V(\varepsilon) = \frac{\pi^2}{3}$, g corresponds to the "logit-link" so that $P(y_i = 1) = \Lambda(\mathbf{x}_i'\boldsymbol{\beta}^*)$.[1] The assumption of $\varepsilon \sim N(0, 1)$ leads to $P(y_i = 1) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}^*)$, so that g corresponds to the "probit-link" function.[2] The relation between the "true" coefficients β and β^* is given by:

$$\beta_j^* = \frac{\beta_j}{\sigma} \tag{2.4}$$

Thus, the β^*- coefficients of logit and probit models are implicitly rescaled by σ due to the fixed error variance. Compared to β they therefore additionally depend on the extent of unobserved residual variation. Consequently, the identification assumptions of nonlinear models ((1) $\tau = 0$, (2) $E(\varepsilon) = 0$, (3) $V(\varepsilon) = \frac{\pi^2}{3}$ respectively $V(\varepsilon) = 1$; Long 1997) do not allow for naive comparisons of coefficients between different model specifications, groups, points in time or samples if the value of σ differs between the corresponding models (see also Mood 2010).

The difficulties concerning effect comparisons between different (nested) model specifications (case 1) result from the fixation of the error variance component in the context of variance decomposition. Here, the fixed error variance does not allow for a decrease in

[1] Λ = cumulative distribution function (cdf) of the standard-logistic distribution.
[2] Φ = cumulative distribution function (cdf) of the standard normal distribution.

residual variation as compensation for an elevated explained variance in models with additional explanatory variables. In consequence, the total variance of y^* must be adjusted and thus differs between the models. There have been different attempts at solving this problem, for example the usage of (fully) standardized β- coefficients, average marginal effects (AME's) or the KHB method (Karson et al. 2012; see also the overview in Best / Wolf 2012).[3] Correspondingly, comparisons of coefficients between different groups (case 2) can be complicated by differences in explanatory power (of the same model specification) between these groups. In this case, the β^*- coefficients refer to y^*- variables with different (group specific) scales. Furthermore, it is possible that the ("true") error variances differ between groups in spite of equal explanatory power, so that the rescaled β^*- coefficients cannot be compared either (Hoetker 2004, Williams 2010). Due to this circumstance, specific approaches have been developed in the context of case 2.

Given the outlined difficulties, Allison's (1999) solution for the inter-group comparison of coefficients in nonlinear regression models is based on the expansion of the standard-logistic model by the additional parameter δ:[4]

$$P(y_i = 1) = \Lambda(\beta_0^* + \sum \beta_j^* x_{ij}) \times (1 + \delta G_i) \tag{2.5}$$

$$= \Lambda\left(\frac{\beta_0^* + \sum \beta_j^* x_{ij}}{1/(1 + \delta G_i)}\right)$$

$$= \Lambda\left(\frac{\beta_0^* + \sum \beta_j^* x_{ij}}{\sigma_i}\right)$$

with

$$\sigma_i = \frac{1}{1 + \delta G_i}$$

Here, G_i is a dummy variable which indicates the affiliation to the respective group and δ is an adjustment factor (with $\delta > -1$) that allows the group $G_i = 1$ to deviate from the fixed residual variance with $\sigma_i = 1/(1 + \delta G_i)$ (for $G_i = 0$, σ_i equals 1). Furthermore the model contains the grouping variable in $\sum \beta_j^* x_{ij}$, enabling group-specific intercepts. Based on the assumption that several (at least one) variables exert the same "true" influence in both groups, the corresponding coefficients are rescaled by the same factor in groups with unequal error variances, whereas the latter is integrated into the model

[3] A correction procedure concerning multilevel logit and probit models can be found in Hox (2010).

[4] Analogue to Allison (1999), Hauser / Andrew (2006) and Williams (2009) focus on logit models, subsequent examinations are based on the usage of logit-link functions. Furthermore, comparisons between two groups are considered.

(2.5) as a function of δ. If group specific error variances are discovered (e.g. through a χ^2- difference test between models with and without δ) the model can be extended with (at max $J-1$) interaction effects between x_{ij} and G_i to detect differences in effects between the groups while taking differences in residual variances into account.

Hauser and Andrew (2006) propose a similar model, yet in a completely different context. Their "logistic response model with proportionality constraints" (LRPC) includes the additional parameter λ_k:

$$P(y_{ik}=1) = \Lambda(\beta^*_{k0} + \lambda_k \sum \beta^*_j x_{ikj}) \tag{2.6}$$

$$= \Lambda\left(\beta^*_{k0} + \frac{\sum \beta^*_j x_{ikj}}{\sigma_k}\right)$$

with

$\sigma_k = \dfrac{1}{\lambda_k}$

Even though the model has originally been conceptualized to fit $k = 1, ..., K$ educational transitions, it can be applied to analyze $K = 2$ groups. In that case, β^*_{k0} contains two group specific intercepts, thus the grouping variable is not included in $\sum \beta^*_j x_{ikj}$ as additional explanatory factor. As a key feature of (2.6), the coefficients of group 2 can deviate from the analogue coefficients of group 1 due to λ_k, which is restricted to 1 in group 1 ($\lambda_1 = 1$). A comparison of (2.5) and (2.6) immediately illustrates the analogy between Allison's approach and the LRPC in the case of two groups, whereas $\lambda_2 = 1 + \delta$. A relaxation of the proportionality assumption of the LRPC can be made with the introduction of interaction terms between x_{ij} and the corresponding grouping variable ("logistic response model with partial proportionality constraints"; LRPPC).

Finally, a third approach contains the specification of a "heteroscedastic logit model" (as a special case of "heterogeneous choice" models; Williams 2009):

$$P(y_i = 1) = \Lambda\left(\frac{\sum \beta^*_j x_{ij} - \tau}{\exp(G_i\gamma)}\right) \tag{2.7}$$

$$= \Lambda\left(\frac{\sum \beta^*_j x_{ij} - \tau}{\sigma_i}\right)$$

with

$\sigma_i = \exp(G_i\gamma)$

Compared to the former models, (2.7) imposes the restriction $\beta_0^* = 0$, thus the threshold τ is estimated instead of the intercept. In the context of inter-group comparisons, a grouping dummy G_i can be added to the explanatory variables in $\sum \beta_j^* x_{ij}$. Within the "heteroscedastic logit model", the additional parameter γ allows a deviation of the fixed error variance within the group $G_i = 1$ through $\exp(G_i\gamma)$ (given the assumption of equal effects between the groups).[5] A comparison of (2.5), (2.6) and (2.7) illustrates the equivalence of the three modeling strategies, whereas $\gamma = \ln(1/(1+\delta))$ and $\gamma = \ln(1/\lambda_2)$ (Williams 2010). With the inclusion of interaction effects between x_{ij} and G_i, the "heteroscedastic logit model" additionally enables the specification of an extended model with group specific effects of β^* while taking potential differences in error variation into account.

Even though all three outlined techniques lead to empirically equivalent models, both Allison's approach and the "heteroscedastic logit model" are based on entirely different rationales of the existing group differences compared to the LRPC. While the first two approaches initially assume that the difference between the groups lies within the error variances (which results in apparently different effects across groups), the LRPC attributes (real) effect differences instead of group specific error variances to the very same mechanism (Williams 2010). Since these interpretations are not empirically distinguishable from each other, the underlying assumptions of each model need to be considered carefully when these methods are applied.[6]

3 SEM-Extension

So far, difficulties of effect comparisons in logit and probit models have been discussed for cases with a single dependent variable. However, it can easily be seen that these problems extend to models with multiple outcomes. Consider the structural equation model

$$\boldsymbol{\eta} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\varepsilon} \tag{3.1}$$

where $\boldsymbol{\eta}$ represents a $(K \times 1)$ vector of latent dependent variables, $\boldsymbol{\mu}$ is a $(K \times 1)$ vector of intercepts, $\boldsymbol{\Gamma}$ contains a $(K \times J)$ matrix of regression slopes, $\mathbf{x}$ is a $(J \times 1)$ vector of (ob-

[5] In addition, it is possible to specify an extended variance equation with several variance conditioning variables ($\exp(\sum z_{ij}\gamma_j)$; Williams 2009).

[6] Alternative approaches to this set of problems can further be found in Long (2009) and Breen et al. (2014).

served) independent variables and $\boldsymbol{\varepsilon}$ is a $(K \times 1)$ vector of residuals with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega})$.[7] For simplicity, the following derivations focus on the limiting case of two dependent variables ($K = 2$), thus equation (3.1) contains:

$$\begin{aligned} \eta_1 &= \mu_1 + \mathbf{x}'\boldsymbol{\gamma}_1 + \varepsilon_1 \\ \eta_2 &= \mu_2 + \mathbf{x}'\boldsymbol{\gamma}_2 + \varepsilon_2 \end{aligned} \tag{3.2}$$

Here, $\boldsymbol{\gamma}_1$ is a $(J \times 1)$ vector of regression coefficients which relates $\mathbf{x}$ to η_1 and $\boldsymbol{\gamma}_2$ is a $(J \times 1)$ coefficient vector relating $\mathbf{x}$ to η_2, respectively. With categorical y- variables, the η- variables are related to their observable counterparts with a threshold model (cf. y^* in section 2). In the following, two binary y- variables are considered, where $y_1 = 1$ if $\eta_1 \geq \tau_1$, $y_1 = 0$ if $\eta_1 < \tau_1$, $y_2 = 1$ if $\eta_2 \geq \tau_2$ and $y_2 = 0$ if $\eta_2 < \tau_2$.

Since a simultaneous specification of all intercepts and thresholds leads to identification issues, restrictions have to be imposed on these parameters. In this derivation, the restriction $\boldsymbol{\mu}^* = \mathbf{0}$ is specified. Furthermore, identifying assumptions have to be made concerning the unobservable error variances in $\boldsymbol{\Omega}$, whereas in the following the standardization $\boldsymbol{\varepsilon}^* \sim N(\mathbf{0}, \boldsymbol{\Psi}^*)$ with $diag(\boldsymbol{\Psi}^*) = \mathbf{I}$ is imposed. Given these restrictions, equation (3.1) results in a multivariate probit model:[8]

$$\boldsymbol{\eta}^* = \boldsymbol{\Gamma}^*\mathbf{x} + \boldsymbol{\varepsilon}^* \tag{3.3}$$

The error covariance and coefficient matrices of equation (3.1) and (3.3) are related as follows:

$$\boldsymbol{\Psi}^* = \boldsymbol{\Lambda}\boldsymbol{\Omega}\boldsymbol{\Lambda} \tag{3.4}$$

$$\boldsymbol{\Gamma}^* = \boldsymbol{\Lambda}\boldsymbol{\Gamma} \tag{3.5}$$

Where $\boldsymbol{\Lambda}$ contains the inverted standard deviations of the unobserved residuals $\boldsymbol{\varepsilon}$:

[7] For a related discussion of model structures which cover relationships between the η- variables (via $\mathbf{B}\boldsymbol{\eta}$) cf. Stein and Pavetic (2013).

[8] In this case, the probability of e.g. $y_1 = 1$ is given by $P(y_1 = 1) = \Phi(\tau_1 - E(\eta_1^*))$.

$$\mathbf{\Lambda} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}^{-1}$$

Using scalar notation, equation (3.5) implies that $\gamma^*_{1j} = \frac{\gamma_{1j}}{\sigma_1}$ and $\gamma^*_{2j} = \frac{\gamma_{2j}}{\sigma_2}$, which corresponds with the well-known problem outlined in section 2. However, in a SEM-framework the fixation of the unobserved error variances poses problems concerning effect comparisons *within* the specified model (i.e. effect comparisons between equations). Consider the hypothesis $\gamma_{1j} = \gamma_{2j}$, assuming that the effect of a given predictor is equal in both equations. Such effect comparisons may be of particular interest in e.g. dyadic models, where the dependent variables of two partners are related to the same set of actor-specific independent variables (Kenny et al. 2006). With categorical outcomes, such hypotheses must be formulated in terms of $\mathbf{\Gamma}^*$ (Stein / Pavetic 2013, Pavetic 2009, Sobel / Arminger 1992):

$$\begin{aligned} \gamma^*_{1j}\sigma_1 &= \gamma^*_{2j}\sigma_2 \\ \gamma^*_{1j} &= \frac{\sigma_2}{\sigma_1}\gamma^*_{2j} \\ \gamma^*_{1j} &= \lambda\gamma^*_{2j} \end{aligned} \tag{3.6}$$

Thus, it can be seen that with the specification of $\gamma^*_{1j} = \lambda\gamma^*_{2j}$, the relation of the unobserved error variances is taken into account within the imposed equality restriction. The empirical implementation of equation (3.6) results in the specification of non-linear parameter constraints, imposing proportionality restrictions on the coefficients of interest through the introduction of λ. Subsequently, the constrained model can be compared with an unrestricted model specification in order to draw conclusions concerning the postulated hypothesis of equal effects across equations. In this context, different sets of parameter restrictions (i.e. multiple hypotheses) can be tested in a stepwise manner.

The outlined procedure of (sequenced) effect comparisons in nonlinear multivariate regression models requires SEM-software which allows the specification of non-linear constraints. Examples of statistical software packages with corresponding capabilities include MECOSA (Arminger et al. 1996) and Mplus (Muthén 1998-2004), whereas the latter is used for the empirical applications of this study.[9]

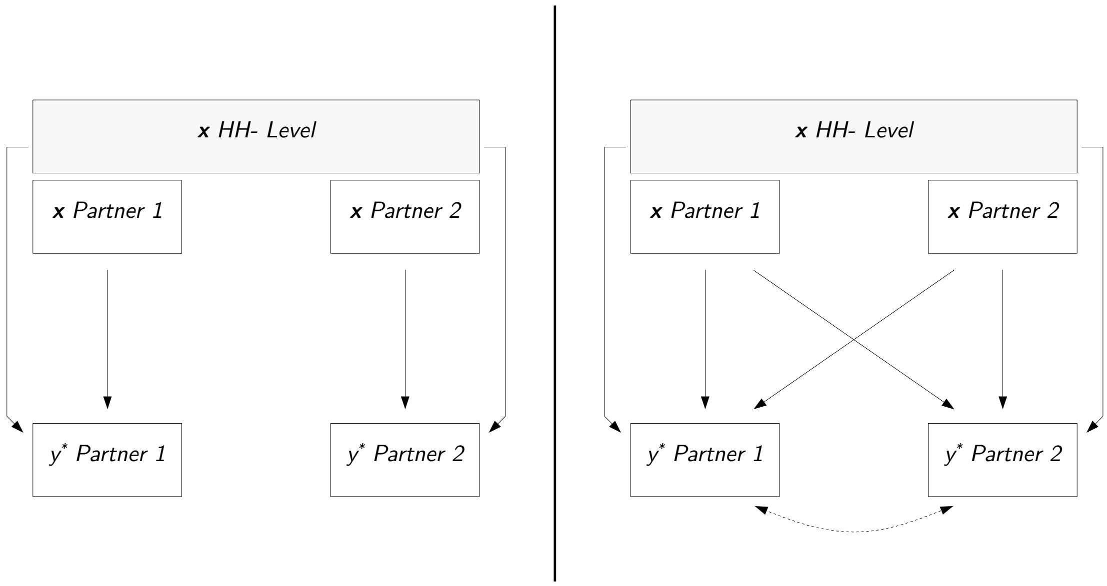
[9] In this case, non-linear constraints can be easily specified within Mplus' MODEL CONSTRAINT subsection (Muthén / Muthén 1998-2012).

4 Application

In this section, the previously outlined technique will be exemplified within some empirical applications. In this context, it is useful to distinguish between two types of models, which are illustrated in Figure 1. *Model type 1* encompasses a simple dyadic model structure, where the non-metric dependent variables of two partners (male & female) are related to a set of individual-level and household-level predictors. Without the specification of additional partner-related effect structures (and given the availability of a distinct partner identification in the data set), this model can be fitted either in a structural equation modeling framework (data in "wide format") or within a simple group comparison between gender using standard logit models (data in "long format"). Thus, any previously outlined technique is applicable for model structures of type 1. However, dyadic research typically involves the specification of partner effects and the consideration of error covariances, which leads to an elaborated model structure (*model type 2*). Since models of this type inherently imply the usage of specific SEM estimation methods, the specification of non-linear constraints is the remaining method of choice concerning effect comparisons between gender in this context.

Figure 1: Dyadic model structures with non-metric outcomes

(a) Model type 1 (b) Model type 2



The following empirical application consists of two steps, covering both model types of

Figure 1. At first, the analogy of the previously outlined approaches will be illustrated within a simple dyadic model structure of type 1 (example 1, section 4.1). Subsequently, the usage of non-linear constraints will be exemplified based on an extended model specification of type 2 (example 2, section 4.2).[10] In both cases, the (binary) dependent variables represent the mobility disposition of partner 1 (male; y_1) and partner 2 (female; y_2), which are assumed to be a function of the exogenous variables described in Table 1.[11] Whereas in both applications the same set of independent variables is used, the mobility dispositions of example 2 are assumed to be additionally dependent on individual characteristics of the respective partner ("partner effects" in Table 1). The empirical investigations are based on data from wave z (2009) of the German Socio-Economic Panel Study (GSOEP; Wagner et al. 2007), which allows a distinct partner matching through the usage of partner identification codes.

Table 1: Description of exogenous variables (Example 1 & 2)

		actor effects		partner effects	
	Variables	male	female	male	female
x_1, x_{16}	Age	γ_{11}	γ_{216}		
x_2, x_{17}	Education (in years)	γ_{12}	γ_{217}	γ_{117}	γ_{22}
x_3, x_{18}	marginal Emp. (Ref: full / part-time Emp.)	γ_{13}	γ_{218}	γ_{118}	γ_{23}
x_4, x_{19}	Non-Working (Ref: full / part-time Emp.)	γ_{14}	γ_{219}	γ_{119}	γ_{24}
x_5, x_{20}	Life Satisfaction	γ_{15}	γ_{220}		
x_6, x_{21}	Risk Tolerance	γ_{16}	γ_{221}		
x_7	Household Income	γ_{17}	γ_{27}		
x_8	Household Size	γ_{18}	γ_{28}		
x_9	Household Size^2	γ_{19}	γ_{29}		
x_{10}	Owner (Ref: Renter)	γ_{110}	γ_{210}		
x_{11}	Tenure (in years)	γ_{111}	γ_{211}		
x_{12}	Local ties	γ_{112}	γ_{212}		
x_{13}	Number of Children $<$ 6 y.	γ_{113}	γ_{213}		
x_{14}	Number of Children 6 - 16 y.	γ_{114}	γ_{214}		
x_{15}	Mover 2008 (Ref: Stayer)	γ_{115}	γ_{215}		

4.1 Example 1: Comparison of correction methods

In order to compare the techniques reviewed in section 2 with the proposed approach of section 3, the same model specification has been fitted using Allison's (1999), Hauser

[10] The corresponding Stata and Mplus Codes are available from the authors upon request.

[11] The y- variables are based on the question "Could you imagine moving away from here because of family or career reasons?" with the dichotomized response categories 0="No / It depends" and 1="Yes".

/ Andrews' (2006) and Williams' (2009) methods. In addition, an equivalent model has been implemented in Mplus, using non-linear constraints as proposed in this paper. In all cases, the mobility disposition of both partners (male & female) are assumed to be solely dependent on actor-related individual features and household characteristics, so that the model can either be implemented within a dyadic (two equation) SEM-framework or with a simple group comparison between gender using standard (single equation) logit regression.[12] Selected results of the four model specifications are presented in Table 2. It can be seen that apparently not only the first three approaches produce empirically equivalent results (cf. Williams 2010) but also that the model fitted with Mplus using non-linear constraints (and ML estimation with a logit link) induces the same LogLikelihood as the former models. Thus, in this case (i.e. applying models of type 1) all four techniques generate the same empirical results. However, following Williams (2010) it should be noted that these results may be interpreted quite differently: From the perspective of Allison's (1999) and Williams' (2009) methods, one would conclude that the standard deviation of the error term is $5,1\%$ (δ) lower for the male partner in comparison with the respective error variation of the female partner, or – equivalently – that σ_ε is $\exp(\gamma) = 1.054$ times larger for women than for men. On the other hand, the LRPC implies that the gender-specific *effects* differ by .949 ($\beta^*_{female} = \lambda_{LRPC}\beta^*_{male}$) or 1.054 ($\beta^*_{male} = \frac{1}{\lambda_{LRPC}}\beta^*_{female}$), respectively. Finally, in line with the derivations of section 3, the Mplus model with non-linear constraints constitutes that the ratio $\frac{\sigma_{female}}{\sigma_{male}}$ of the (unobserved) error standard deviations equals $\lambda_{Mplus} = 1.053$, which corresponds with the perspective of Allison's (1999) and Williams' (2009) techniques.

Table 2: Comparison of correction methods

	Allison's approach (1999)	LRPC (Hauser / Andrew 2006)	Heteroscedastic logit model (Williams 2009)	Mplus model with non-linear constraints
δ	-.051 (.094)			
λ (LRPC)		.949 (.094)		
γ			.052 (.099)	
λ (Mplus)				1.053 (0.110)
	$\frac{1}{1+\delta G_i} = 1.054$	$\frac{1}{\lambda} = 1.054$	$\exp(G_i\gamma) =$ 1.054	
LL	-3721.903	-3721.903	-3721.903	-3721.903

[12] In this context, the same effect structure is implied for both gender, thus a fully restricted Mplus model is estimated and the single equation models are fitted without any gender interactions.

4.2 Example 2: Non-linear constraints in dyadic probit models

As with model structures of type 1, non-linear constraints can be easily implemented in extended dyadic models with partner effects and error covariances (type 2 models). Furthermore, for both model types, series of variously restricted models can be specified in order to test effect differences of specific sets of parameters while taking potential differences in residual variation into account. This procedure has its analogy in the inclusion of interaction terms with the grouping variable within the single equation approaches discussed in section 2.[13] In the context of SEM-structures, this practice requires a fully constraint model to be fitted first, followed by a set of less restrictive models. Subsequently, χ^2- difference tests can be carried out in order to test which specification (i.e. hypothesis) should be preferred.

Turning to the previous example from mobility research, the outlined testing procedure with non-linear constraints is illustrated on the basis of an extended dyadic model of type 2 (including partner effects and a specified error covariance).[14] In this context, five model versions have been specified: Starting from the fully constraint model 1, model 2 relaxes the assumption that the partner effects γ_{117} and γ_{22} (partner effects of education) are proportionally equivalent. The next set of partner effect restrictions (concerning the effects of employment status) is relaxed in model 3, thus in this case only the actor-related individual- and household-level effects are constraint to be (proportionally) equal across equations. Finally, model 4 solely poses restrictions on the individual-level actor effects, whereas in model 5 no constraints are specified.

The results of the corresponding SB-corrected χ^2- difference tests (Satorra / Bentler 1999) are summarized in Table 3. It can be seen that the relaxation of the first restriction induces a significant improvement in model fit, thus from the perspective developed here, it can be argued that substantial differences between the gender-specific partner effects of education can be observed which cannot be solely attributed to differences in residual variation (since these differences are already accounted for by λ). Given the insignificant test result of the next χ^2- difference test (model 2 vs. model 3), it becomes clear that such differences cannot be observed concerning the partner effects of employment status. Furthermore, also the relaxation of the next two sets of parameter restrictions (actor effects of individual- and household-level predictors) do not induce any substantial improvement in model fit, thus on the basis of the conducted χ^2- difference

[13] Therefore, the present strategy bears the same restrictions and assumptions as outlined by Williams (2009).

[14] The following results are based on Mplus' WLSM estimation procedure, which uses the probit link (e.g. Muthén 1998-2004).

tests model 2 provides the best balance between model fit and parsimony.

Table 3: Scaled χ^2- difference tests

Model	Restriction	χ^2	df	χ^2_{sc} *Diff.*	p	r^2_{y1}	r^2_{y2}
1	$\gamma^*_{11} = \lambda\gamma^*_{216}$ ... $\gamma^*_{16} = \lambda\gamma^*_{221}$,† $\gamma^*_{17} = \lambda\gamma^*_{27}$ ... $\gamma^*_{115} = \lambda\gamma^*_{215}$,†† $\gamma^*_{118} = \lambda\gamma^*_{23}$, $\gamma^*_{119} = \lambda\gamma^*_{24}$,††† $\gamma^*_{117} = \lambda\gamma^*_{22}$†††	23.426	23			.127	.123
2	$\gamma^*_{11} = \lambda\gamma^*_{216}$ ... $\gamma^*_{16} = \lambda\gamma^*_{221}$, $\gamma^*_{17} = \lambda\gamma^*_{27}$ ... $\gamma^*_{115} = \lambda\gamma^*_{215}$, $\gamma^*_{118} = \lambda\gamma^*_{23}$, $\gamma^*_{119} = \lambda\gamma^*_{24}$,	19.852	22	2.839	0.092	.124	.128
3	$\gamma^*_{11} = \lambda\gamma^*_{216}$ ... $\gamma^*_{16} = \lambda\gamma^*_{221}$, $\gamma^*_{17} = \lambda\gamma^*_{27}$ ... $\gamma^*_{115} = \lambda\gamma^*_{215}$,	18.545	20	1.388	0.500	.122	.131
4	$\gamma^*_{11} = \lambda\gamma^*_{216}$ ... $\gamma^*_{16} = \lambda\gamma^*_{221}$,	9.487	11	9.521	0.391	.127	.131
5		6.050	6	3.509	0.622	.124	.134

†actor effects, ††HH-effects, †††partner effects

The preferred model 2 of the previously outlined model series is illustrated in Table 4 in more detail. Here, all coefficients except for the partner effects of education (γ_{117} and γ_{22}) are constrained to be proportionally equal across equations, with $\gamma^*_{male} = \lambda\gamma^*_{female}$. As in the previous example, λ accounts for potential differences in residual variation, whereas in this case $\frac{\sigma_{female}}{\sigma_{male}} = 1.071$. As a result of the imposed restrictions, symmetric partner effects can be observed concerning the coefficients of the employment status dummys, which show a positive effect of a non-working spouse on the mobility disposition for both gender. In contrast, the unconstraint partner effects of education exhibit a different effect pattern: Here, higher levels of education of the male partner are related to an increase in the willingness to move of the female partner, whereas a corresponding effect cannot be observed in the male's equation. Thus, as indicated in the previous section, substantial effect differences can be observed in this case. Additionally, it can be seen that the inclusion of ψ_{21} (error covariance) accounts for substantial interdependencies between both partners, underlining the utility of an extended model specification in applications with dyadic data structures.

Table 4: Dyadic probit model with non-linear constraints

	Partner 1 (Male)			Partner 2 (Female)		
	γ^*	*se*	$\gamma^*_{S_{xy}}$	γ^*	*se*	$\gamma^*_{S_{xy}}$
actor effects						
Age	$-.009^*$	(.004)	−.081	$-.008^*$	(.003)	−.076
Education	$.037^{***}$	(.008)	.097	$.034^{***}$	(.007)	.085
marginal Emp.†	.033	(.078)	.005	.031	(.073)	.009
Non-Working†	$.107^*$	(.048)	.036	$.100^*$	(.045)	.042
Life Satisfaction	$-.049^{***}$	(.011)	−.078	$-.046^{***}$	(.010)	−.073
Risk Tolerance	$.061^{***}$	(.009)	.122	$.057^{***}$	(.008)	.106
HH-effects						
HH-Income	$.000^{***}$	(.000)	.123	$.000^{***}$	(.000)	.114
HH-Size	$-.100^{**}$	(.032)	−.102	$-.093^{***}$	(.029)	−.095
HH-Size^2	$.047^{***}$	(.014)	.112	$.044^{***}$	(.013)	.104
Owner††	$-.311^{***}$	(.050)	−.141	$-.290^{***}$	(.048)	−.132
Tenure	$-.008^{***}$	(.003)	−.087	$-.008^{***}$	(.002)	−.081
Local ties	$-.119^{***}$	(.025)	−.093	$-.111^{***}$	(.024)	−.086
Children $<$ 6 y.	.030	(.052)	.014	.028	(.049)	.013
Children 6 - 16 y.	.017	(.039)	.012	.015	(.037)	.012
Mover 2008†††	$-.156^+$	(.087)	−.034	$-.145^+$	(.081)	−.032
partner effects						
Education	.015	(.011)	.037	$.042^{***}$	(.011)	.111
marginal Emp.†	.105	(.074)	.029	.098	(.069)	.013
Non-Working†	$.105^*$	(.047)	.043	$.098^*$	(.044)	.033
λ	1.071	(.122)				
ψ_{21}	$.558^{***}$	(.024)				
τ	.713	(.058)		.700	(.059)	
r^2_{MZ}	.124			.128		
χ^2 (22)	19.852					
RMSEA	.000					
CFI	1.000					
TLI	1.005					
n	3631					

†Ref.: full / part-time Emp., ††Ref.: Renter, †††Ref.: Stayer 2008

$^+$: p $\leq$ 0.1; *: p $\leq$ 0.05; **: p $\leq$ 0.01; ***: p $\leq$ 0.001

5 Discussion

In the present study, effect comparisons in nonlinear models have been discussed from a structural equation modeling perspective. In this case, it has been shown that the well-known problems which arise in the context of standard (single equation) logit and probit regression extend to models with multiple non-metric outcomes. More specific, the fixation of the unobserved error variances poses substantial problems concerning effect comparisons between equations when the assumption of equal residual variances is not met. As a result, a naive (direct) comparison of coefficients across equations can lead to false conclusions. A potential solution for these problems was discussed through the implementation of non-linear constraints, which enable the specification of equality restrictions while taking potential differences in residual variation explicitly into account. Since non-linear constraints can be easily specified in advanced SEM software (e.g. Mplus), this method provides a handy tool when categorical variables are analyzed in a multivariate framework.

Furthermore, the technique outlined in this paper has been related to previously proposed methods which have been (mainly) developed in the context of group comparisons with single non-metric outcomes (Allison 1999, Hauser / Andrew 2006, Williams 2009). It has been shown that these approaches induce the same empirical results concerning model specifications which can be fitted either in a nonlinear SEM- framework or within simple group comparisons using standard logit regression. Thus, the application of non-linear constraints involves the same advantages as well as limitations as the considered single equation techniques (cf. Williams 2009, 2010): On the one hand, markedly different interpretations of the same (constrained) model result are possible.[15] Correspondingly, the hypothesis of unequal effects may be falsely rejected in models with homogenously different effect patterns between equations, because these effect differences can be absorbed into λ. On the other hand, the specification of non-linear constraints protects against the false rejection of the hypothesis of equal effects in cases where apparent effect differences are induced by differences in residual variance. Given the latter capability, the procedure proposed in this paper provides a flexible technique concerning effect comparisons in nonlinear multivariate regression models. However – as with the corresponding single equation approaches – the underlying assumptions should be kept in mind when non-linear constraints are empirically applied.

[15] Specifically, the coefficients may not be interpreted to be proportionally equivalent across equations because of differences in residual variation but to be actually different in effect size (LRPC perspective; Hauser / Andrew 2006).

References

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

Arminger, G., Wittenberg, J. and Schepers, A. (1996): MECOSA 3: A program for the analysis of general mean- and covariance structures with nonmetric variables, User Guide. *Frauenfeld: SLI-AG.*

Best, H. and Wolf, C. (2012): Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64, 2, pp. 377-395.

Breen, R., Holm, A. and Karlson, K. B. (2014): Correlations and Nonlinear Probability Models. In: *Sociological Methods & Research* 43, 4, pp. 571-605.

Hauser, R. M. and Andrew, M. (2006): Another look at the stratification of educational transitions: The logistic response model with partial proportionality constraints. In: *Sociological Methodology* 36, 1, pp. 1-26.

Hoetker, G. (2004): Confounded coefficients: Extending recent advances in the accurate comparison of logit and probit coefficients across groups. *Working Paper.*

Hox, J. J. (2010): *Multilevel Analysis. Techniques and Applications.* New York, NY: Routledge.

Karlson, K. B., Holm, A. and Breen, R. (2012): Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method. In: *Sociological Methodology* 42, 1, pp. 286-313.

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

Long, J. S. (1997): *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage.

Long, J. S. (2009): Group comparisons in logit and probit using predicted probabilities. *Working Paper.*

Mood, C. (2010): Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. In: *European Sociological Review* 26, 1, pp. 67-82.

Muthén, B. O. (1998-2004): Mplus Technical Appendices. *Los Angeles, CA: Muthén & Muthén.*

Muthén, L. K. and Muthén, B. O. (1998-2012): Mplus Users Guide. Seventh Edition. *Los Angeles, CA: Muthén & Muthén.*

Pavetic, M. (2009): *Familiengründung und -erweiterung in Partnerschaften: Statistische Modellierung von Entscheidungsprozessen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Satorra, A. and Bentler, P. (1999): A Scaled Difference Chi-square Test Statistic for Moment Structure Analysis. *UCLA Statistics Series No. 260.*

Sobel, M. E. and Arminger, G. (1992): Modeling Household Fertility Decisions: A Nonlinear Simultaneous Probit Model. In: *Journal of the American Statistical Association* 87, 417, pp. 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. In: *Quality and Quantity* 47, 3, pp. 1717-1732.

Wagner, G. G., Frick, J. R. and Schupp, J. (2007): The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements. In: *Schmollers Jahrbuch* 127, 1, pp. 139-169.

Williams, R. (2009): Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups. In: *Sociological Methods & Research* 37, 4, pp. 531-559.

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