Standard errors and confidence intervals of standardized parameters in structural equation models

Daniel Oberski
RECSM research paper, Universitat Pompeu Fabra

May 29, 2011

Abstract

In structural equation models (SEM), often not only the standard errors of parameters of the model, but also of the standardized coefficients are of interest to the researcher. Examples are the comparison of reliability coefficients across different groups, and meta-analysis of standardized quantities. The literature does not, however, provide explicit analytical standard errors for standardized parameters.

We derive an explicit expression in terms of the model parameters for derivatives and the asymptotic variance-covariance matrix of standardized parameters in structural equation models, that is straightforward to implement in standard SEM software. The construction of confidence intervals via the Fisher z-transformation is also discussed. The solution is applied to an analysis of the reliability of self-rated health in groups with different levels of education.

1 Introduction

Linear structural equation models (SEM) with latent variables have become a popular tool in the behavioral sciences. Such models encompass as special cases a diverse range of common models of interest such as factor analysis, multivariate regression, errors-in-variables models, growth, and multilevel and multigroup models (Bollen, 1989). Extensions are available for categorical, count, and censored dependent variables as well as complex sampling (Muthen and Satorra, 1995; Muthén, 2002).

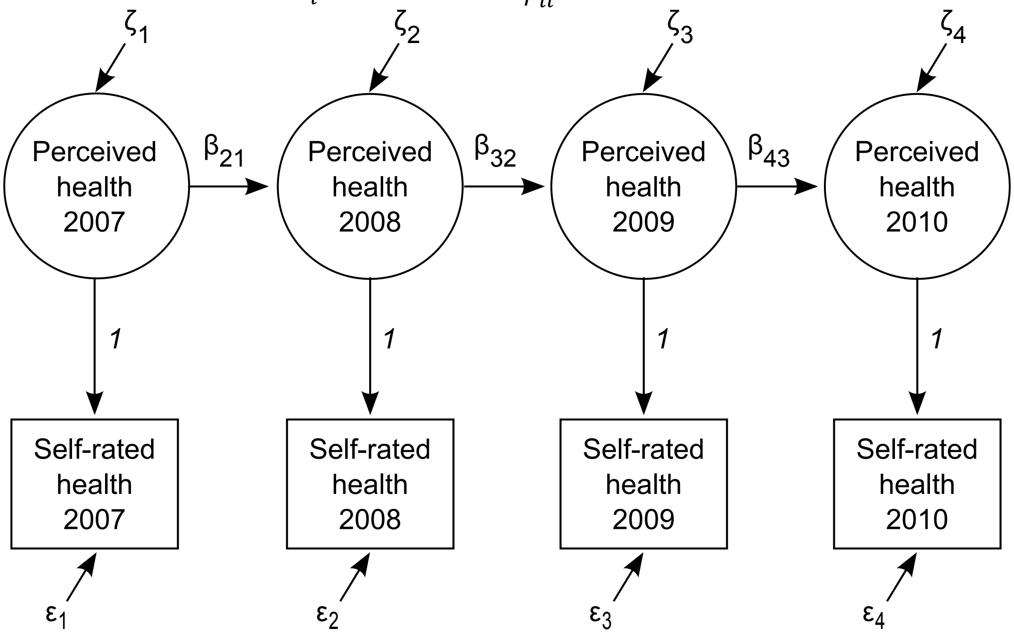
Although the main interest in such models lies in the parameters of the model, researchers' interest often also focuses on the so-called 'standardized' parameters (Bollen, 1989). Typical applications include examination of factor loadings and correlations in factor analysis, as well as the evaluation of the relative size of regression coefficients, possibly of latent variables.

Although the general principle of deriving standard errors and confidence intervals for standardized parameters is well known, to our knowledge the literature does not provide any explicit expression for the asymptotic standard errors of standardized coefficients in SEM. As a consequence, standard errors and confidence intervals for these coefficients are usually not provided by the standard software[1]. We remedy this situation by deriving an explicit expression, in terms of the unstandardized parameter estimates, for the asymptotic variance-covariance matrix of the standardized coefficients. The solution requires only the parameter estimates of the model and can be readily implemented in SEM software.

Section 2 provides a motivating example, using a SEM model with latent variables where standardized parameters and their standard errors and confidence intervals are of interest to

[1] At the time of writing, the exception is Mplus version 5.2 and above (Muthén and Muthén, 1998).

Figure 1: Quasi-simplex model for four repeated measures of self-rated health in the LISS panel 2007–2010. Parameter names for the so-called 'stability coefficients' are shown as β_{ji} in the picture. The variances of the disturbance terms ζ_i are denoted φ_{ii}, while the variances of the measurement error variables ε_i will be denoted ψ_{ii}.



the researcher. Section 3 derives the explicit expression for the asymptotic variance-covariance matrix of the standardized estimates. Section 4 applies this expression to the example. The last section discusses the scope and limitations of this proposal and suggests future research.

2 Example SEM with interest in standardized parameters

The study of differences across societal groups in self-rated health is of interest to researchers in public health. For example, von dem Knesebeck et al. (2006) compare people with different incomes, age groups, and level of education. It is well-known, however, that in order to be able to compare correlations across groups it is necessary for the reliability of the measures to be the same (e.g. Saris and Gallhofer, 2007a). Therefore not only across-group comparison of the levels of self-rated health are of interest, but also the evaluation of differences between the groups in reliability (Lundberg and Manderbacka, 1996).

Different types of designs exist to estimate the reliability of survey measures. One such design is the repeated measures design, wherein the same question is asked at least three times with a certain interval. One can then apply the so-called 'quasi-simplex' model in different groups to the estimate reliability, and compare reliabilities across groups (Heise and Bohrnstedt, 1970; Wiley and Wiley, 1970). For an overview of other designs we refer to Alwin (2007).

The quasi-simplex model can be formulated and estimated as a multiple-group structural equation model for groups with different levels of highest completed education. This model is shown for four repetitions in figure 1. Following Wiley and Wiley (1970), the unstandardized error variances are restricted to be equal across the four repetitions. The reliability coefficients of interest are then the standardized loadings. This yields the structural equation model shown as a path diagram in figure 1.

		Year							
		2007		2009		2009		2010	
	n	Mean	sd	Mean	sd	Mean	sd	Mean	sd
Education level									
Primary	279	2.83	(0.69)	2.91	(0.73)	2.84	(0.74)	2.82	(0.72)
Lower secondary	940	2.98	(0.73)	3.08	(0.73)	3.01	(0.71)	2.97	(0.73)
Middle secondary	782	3.21	(0.80)	3.24	(0.76)	3.27	(0.81)	3.20	(0.75)
Upper secondary	369	3.17	(0.75)	3.18	(0.74)	3.16	(0.72)	3.10	(0.72)
Lower tertiary	799	3.28	(0.73)	3.28	(0.76)	3.22	(0.72)	3.22	(0.74)
Upper tertiary	256	3.33	(0.86)	3.35	(0.86)	3.32	(0.82)	3.29	(0.83)

Table 1: Means and standard deviations (in brackets) of the self-rated health question across the four repetitions, in groups with different levels of education.

We estimate this model using the LISS panel study, a probability sample of 8000 Dutch citizens. The respondents answer questionnaires over the web. For more information about the LISS we refer to Scherpenzeel (2011). The panel contains a study that included the commonly used self-rated health question[2]. The question was asked as follows:

> How would you describe your health, generally speaking?
>
> 1. Poor
> 2. Moderate
> 3. Good
> 4. Very good
> 5. Excellent

This question was asked of 3425 LISS respondents in the years 2007, 2008, 2009, and 2010. Splitting the sample into groups with different levels of education, the means and standard deviations shown in table 1 are obtained. Using the data from all four years, we estimate the quasi-simplex model shown in figure 1 as a SEM to yield four standardized loadings for each educational group, which can be interpreted as the reliability coefficients for each group[3]. The unstandardized parameter estimates and model fit measures are shown in table 2, while standardized loadings (reliability coefficients) are shown in table 3.

Table 3 shows that there appear to be quite some differences across the groups in reliability. For the lowest educational level in 2007, the reliability coefficient of self-rated health is 0.799, while for the highest level it is 0.896. There also appear to be some differences between the years, although the differences across years are much smaller than those across different educational groups. Finally, the values of all reliability coefficients in table 3 are quite a bit higher than those reported by Lundberg and Manderbacka (1996) for self-rated health. This may be due to differences in the population, in the question or in the mode of interviewing, but it may also be due to sampling.

Since the reliability is a function of both the error variances and the total variances of the latent variables (which involve the β and ψ parameters), a test for invariance of these unstandardized parameters will not suffice to test the desired hypotheses. If the error variances are equal across groups, this does not mean that the reliabilities will be so, since they depend also on

[2] The original question was asked in Dutch. See `http://www.lissdata.nl/dataarchive/question_constructs/view/600`

[3] For the estimation we used `R` 2.13.0 64-bit (R Development Core Team, 2011) and `lavaan` 0.4.8 (Rosseel, 2011)

Par.	Group: education level					
	Primary	Lower secondary	Middle secondary	Upper secondary	Lower tertiary	Upper tertiary
φ_{11}	0.50 (0.049)	0.59 (0.068)	0.35 (0.026)	0.38 (0.030)	0.31 (0.044)	0.40 (0.029)
φ_{22}	0.13 (0.027)	0.17 (0.040)	0.10 (0.018)	0.03 (0.019)	0.05 (0.031)	0.12 (0.021)
φ_{33}	0.10 (0.026)	0.08 (0.027)	0.05 (0.012)	0.04 (0.013)	0.02 (0.024)	0.08 (0.015)
φ_{44}	0.06 (0.027)	0.06 (0.035)	-0.02 (0.017)	0.02 (0.019)	0.02 (0.030)	0.08 (0.020)
β_{21}	0.78 (0.051)	0.85 (0.062)	0.84 (0.045)	0.96 (0.050)	0.99 (0.098)	0.89 (0.043)
β_{32}	0.99 (0.055)	0.87 (0.052)	0.89 (0.039)	0.89 (0.041)	1.01 (0.074)	0.83 (0.034)
β_{43}	0.84 (0.044)	0.95 (0.057)	1.07 (0.041)	0.96 (0.044)	0.92 (0.065)	0.94 (0.039)
ψ_{ii}	0.13 (0.015)	0.14 (0.019)	0.17 (0.010)	0.18 (0.010)	0.17 (0.017)	0.13 (0.010)

Table 2: Unstandardized parameter estimates and standard errors for the multiple group quasi-simplex model shown in figure 1. Chi-square: 6.923 on 12 degrees of freedom ($p = 0.863$), SRMR: 0.0041, RMSEA: 0.000

	n	*Year*			
		2007	2008	2009	2010
Education level					
Primary	279	0.799	0.818	0.828	0.816
Lower secondary	940	0.821	0.821	0.812	0.822
Middle secondary	782	0.891	0.879	0.896	0.877
Upper secondary	369	0.822	0.820	0.808	0.805
Lower tertiary	799	0.869	0.878	0.863	0.874
Upper tertiary	256	0.896	0.898	0.886	0.887

Table 3: Reliability (standardized Λ) of self-rated health in the Netherlands 2007-2010 for groups with different levels of education.

the variances of the latent variables. Conversely, if all parameters are tested for invariance one simultaneously imposes equality of the stability parameters, which is not desired. One would require standard errors or confidence intervals of the standardized parameters to properly study the differences between people with different levels of attained education in reliability.

Another type of analysis that would require the variances of these estimates are meta-analyses of the reliabilities (Cooper et al., 2009, 261, 271-2). Examples of such meta-analyses are Andrews (1984); Scherpenzeel and Saris (1997); Saris and Gallhofer (2007b); Alwin (2007). In the reliability study done here, one may for example wish to combine the results with those of Lundberg and Manderbacka (1996), to see whether differences in reliability across educational groups hold across the studies.

It is clear, therefore, that to answer the questions that are of interest to researchers, standard errors and confidence intervals for the reliability coefficients would be useful. The next section derives these standard errors for general structural equation models.

3 Standard errors of standardized parameters

Let y be a p-vector of observed variables, from which a sample is obtained. The following SEM for y is specified:

$$y = \Lambda\eta + \varepsilon \tag{1}$$
$$\eta = B_0\eta + \zeta, \tag{2}$$

where η is a vector of unobserved variables, ζ is a vector of disturbance terms and ε is a vector of measurement errors. Model 2 implies the following model $\Sigma_\eta(\theta)$ for the variance-covariance matrix of the unobserved variables as a function of a parameter vector θ:

$$\Sigma_\eta(\theta) = B^{-1}\Phi B^{-T}, \tag{3}$$

where $B \equiv I - B_0$ is positive definite, and Φ is the variance-covariance matrix of ζ. Model 1 can then be seen to imply the following model $\Sigma_y(\theta)$ for the variance-covariance matrix of the observed variables:

$$\Sigma_y(\theta) = \Lambda B^{-1}\Phi B^{-T}\Lambda' + \Psi, \tag{4}$$

where Ψ is the variance-covariance matrix of ε. We assume throughout that both Σ_y and Σ_η are positive definite. In what follows we will write $\Sigma_.$ for $\Sigma_.(\theta)$ in the interest of clarity.

The parameters of the model are collected in a parameter vector

$$\theta \equiv [\operatorname{vec}\Lambda, \operatorname{vech}\Phi, \operatorname{vec}B_0, \operatorname{vech}\Psi].$$

Interest focuses not only on the parameter vector θ, but also on the so-called "standardized" parameter vector, denoted $\widetilde{\theta} \equiv [\widetilde{\Lambda}, \widetilde{B_0}]$, where:

$$\widetilde{\Lambda} \equiv D_y^{-1}\Lambda D_\eta \tag{5}$$

$$\widetilde{B_0} \equiv D_\eta^{-1} B_0 D_\eta, \tag{6}$$

and $D_y \equiv \sqrt{I \circ \Sigma_y}$, and $D_\eta \equiv \sqrt{I \circ \Sigma_\eta}$.

By standard application of the Delta method (e.g. Oehlert, 1992), the asymptotic variance of $\widetilde{\theta}$ can be shown to equal

$$\operatorname{var}(\widetilde{\theta}) = \left(\frac{d\widetilde{\theta}}{d\theta}\right)\operatorname{var}(\theta)\left(\frac{d\widetilde{\theta}}{d\theta}\right)'. \tag{7}$$

Here $\operatorname{var}(\theta)$ is the appropriate asymptotic variance matrix of the free model parameters θ (e.g. Satorra, 1989, 143-4). The choice of the appropriate $\operatorname{var}(\theta)$ matrix allows for incorporation of non-normal and complex sample data with possibly missing observations.

In order to construct confidence intervals, it may be advantageous to apply the z-transform, defined as $z = \operatorname{arctanh}(\widetilde{\theta})$ (Fisher, 1925, section 35), to standardized parameters. In this case, again following the delta method, the variance of the transformed parameters will equal

$$\operatorname{var}[\operatorname{arctanh}(\widetilde{\theta})] = T \operatorname{var}(\widetilde{\theta})\, T', \tag{8}$$

where T is the diagonal matrix with elements $T_{(i,i)} = (1 - \widetilde{\theta}_i^2)^{-1}$.

3.1 Derivatives of the standardized parameters

The derivatives $d\widetilde{\theta}/d\theta$ in equation 7 are not available in the literature and are derived here.

From definition 5,

$$\mathrm{d}\operatorname{vec}\widetilde{\Lambda} = (D_\eta\Lambda' \otimes I_p)\,\mathrm{d}\operatorname{vec}D_y^{-1} + (D_\eta \otimes D_y^{-1})\,\mathrm{d}\operatorname{vec}\Lambda + (I_q \otimes D_y^{-1}\Lambda)\,\mathrm{d}\operatorname{vec}D_\eta, \tag{9}$$

and from definition 6,

$$\mathrm{d}\operatorname{vec}\widetilde{B_0} = (D_\eta B_0' \otimes I_p)\,\mathrm{d}\operatorname{vec}D_\eta^{-1} + (D_\eta \otimes D_\eta^{-1})\,\mathrm{d}\operatorname{vec}B_0 + (I_q \otimes D_\eta^{-1}B_0)\,\mathrm{d}\operatorname{vec}D_\eta. \tag{10}$$

The differentials of the standardized parameter matrices are, thus, functions of the differentials of the covariance structure models Σ_y and Σ_η. Neudecker and Satorra (1991) derived the differential of the implied variance matrix Σ_y of the observed variables. To ensure completeness of the treatment, we repeat it here:

$$\begin{aligned} \mathrm{d}\,\mathrm{vec}\,\Sigma_y = (I+K)(\Lambda B^{-1}\Phi B^{-T} \otimes I)\,\mathrm{d}\,\mathrm{vec}\,\Lambda + \\ (I+K)(\Lambda B^{-1}\Phi B^{-T} \otimes \Lambda B^{-1})\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ (\Lambda B^{-1} \otimes \Lambda B^{-1})P\,\mathrm{d}\,\mathrm{vech}\,\Phi + \\ \mathrm{d}\,\mathrm{vech}\,\Psi, \end{aligned} \tag{11}$$

where the commutation matrix K, the duplication matrix P, and the operators vec and vech are defined in Magnus and Neudecker (1988) Let Δ_λ equal the coefficient matrix of $\mathrm{d}\,\mathrm{vec}\,\Lambda$ in equation 11, Δ_β the coefficient of $\mathrm{d}\,\mathrm{vec}\,B_0$, and Δ_φ the coefficient of $\mathrm{d}\,\mathrm{vech}\,\Phi$.

The differential of the variance matrix of η can be obtained as:

$$\begin{aligned} \mathrm{d}\,\mathrm{vec}\,\Sigma_\eta = (I+K)(B^{-1}\Phi B^{-T} \otimes B^{-1})\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ (B^{-1} \otimes B^{-1})P\,\mathrm{d}\,\mathrm{vech}\,\Phi. \end{aligned} \tag{12}$$

Let Δ^*_β and Δ^*_φ be the coefficients of $\mathrm{d}\,\mathrm{vec}\,B_0$ and $\mathrm{d}\,\mathrm{vech}\,\Phi$ respectively in equation 12.

Also, let

$$C_{y2} \equiv -(I_p \otimes \frac{1}{2(I_p \circ \Sigma_y)^{\frac{3}{2}}})\mathrm{diag}[\mathrm{vec}\,(I_p)], \tag{13}$$

$$C_{\eta 1} \equiv (I_q \otimes \frac{1}{2(I_q \circ \Sigma_\eta)^{\frac{1}{2}}})\mathrm{diag}[\mathrm{vec}\,(I_q)], \tag{14}$$

$$C_{\eta 2} \equiv -(I_q \otimes \frac{1}{2(I_q \circ \Sigma_\eta)^{\frac{3}{2}}})\mathrm{diag}[\mathrm{vec}\,(I_q)]. \tag{15}$$

Then, applying standard operations on 9 and rearranging terms, the differential of the standardized $\widetilde{\Lambda}$ matrix is

$$\begin{aligned} \mathrm{d}\,\mathrm{vec}\,\widetilde{\Lambda} = & [(D_\eta\Lambda' \otimes I_p)C_{y2}\Delta_\lambda + (D_\eta \otimes D_y^{-1})]\,\mathrm{d}\,\mathrm{vec}\,\Lambda + \\ & [(D_\eta\Lambda' \otimes I_p)C_{y2}\Delta_\beta + (I_q \otimes D_y^{-1}\Lambda)C_{\eta 1}\Delta^*_\beta]\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ & [(D_\eta\Lambda' \otimes I_p)C_{y2}\Delta_\varphi + (I_q \otimes D_y^{-1}\Lambda)C_{\eta 1}\Delta^*_\varphi]P\,\mathrm{d}\,\mathrm{vech}\,\Phi + \\ & (D_\eta\Lambda' \otimes I_p)C_{y2}P\,\mathrm{d}\,\mathrm{vech}\,\Psi. \end{aligned} \tag{16}$$

And similarly, the differential of the standardized $\widetilde{B}_0$ matrix from equation 10 is

$$\begin{aligned} \mathrm{d}\,\mathrm{vec}\widetilde{B}_0 = & [\{(D_\eta B_0' \otimes I_q)C_{\eta 2} + (I_q \otimes D_\eta^{-1}B_0)C_{\eta 1}\}\Delta^*_\beta + (D_\eta \otimes D_\eta^{-1})]\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ & [(D_\eta B_0' \otimes I_q)C_{\eta 2} + (I_q \otimes D_\eta^{-1}B_0)C_{\eta 1}]\Delta^*_\varphi P\,\mathrm{d}\,\mathrm{vech}\,\Phi. \end{aligned} \tag{17}$$

From the differentials in equations 16 and 17, we conclude that the derivative matrices of the standardized parameters $\mathrm{vec}\,\widetilde{\Lambda}$, $\mathrm{vec}\,\widetilde{B}$, $\mathrm{vech}\,\widetilde{\Phi}$, and $\mathrm{vech}\,\widetilde{\Psi}$ with respect to the free parameters of the model θ will be the partitioned matrices

$$\begin{aligned} \frac{d\tilde{\lambda}}{d\theta} = [& (D_\eta\Lambda' \otimes I_p)C_{y2}\Delta_\lambda + (D_\eta \otimes D_y^{-1}), \\ & (D_\eta\Lambda' \otimes I_p)C_{y2}\Delta_\varphi P + (I_q \otimes D_y^{-1}\Lambda)C_{\eta 1}\Delta^*_\varphi P, \\ & (D_\eta\Lambda' \otimes I_p)C_{y2}\Delta_\beta + (I_q \otimes D_y^{-1}\Lambda)C_{\eta 1}\Delta^*_\beta, \\ & (D_\eta\Lambda' \otimes I_p)C_{y2}P] \end{aligned} \tag{18}$$

$$\frac{d\widetilde{\beta}_0}{d\theta} = [0, \{(D_\eta B_0' \otimes I_q)C_{\eta 2} + (I_q \otimes D_\eta^{-1} B_0)C_{\eta 1}\}\Delta_\varphi^* P,$$
$$\{(D_\eta B_0' \otimes I_q)C_{\eta 2} + (I_q \otimes D_\eta^{-1} B_0)C_{\eta 1}\}\Delta_\beta^* + (D_\eta \otimes D_\eta^{-1}), 0]. \quad (19)$$

$$\frac{d\widetilde{\varphi}}{d\theta} = [0, \{(D_\eta^{-1}\Phi \otimes I_q) + (I_q \otimes D_\eta^{-1}\Phi)\}C_{\eta 2}\Delta_\beta^*,$$
$$\{(D_\eta^{-1}\Phi \otimes I_q) + (I_q \otimes D_\eta^{-1}\Phi)\}C_{\eta 2}\Delta_\varphi^* P + (D_\eta^{-1} \otimes D_\eta^{-1})P, 0]. \quad (20)$$

$$\frac{d\widetilde{\psi}}{d\theta} = [\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}C_{y2}\Delta_\lambda,$$
$$\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}C_{y2}\Delta_\beta,$$
$$\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}C_{y2}\Delta_\varphi P,$$
$$\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}C_{y2}P + D_y^{-1} \otimes D_y^{-1}P] \quad (21)$$

By stacking the derivatives in equations 18–21 rowwise, we obtain the matrix $d\widetilde{\theta}/d\theta$:

$$\frac{d\widetilde{\theta}}{d\theta} = [\left(\frac{d\widetilde{\lambda}}{d\theta}\right)', \left(\frac{d\widetilde{\beta}_0}{d\theta}\right)', \left(\frac{d\widetilde{\varphi}}{d\theta}\right)', \left(\frac{d\widetilde{\psi}}{d\theta}\right)']'. \quad (22)$$

This equation can be applied to equation 7 to to obtain the variance matrix of the standardized parameters, or to equation 8 if the variance matrix of z-transformed standardized parameters are desired.

4 Application: standard errors of standardized parameters

Table 3 showed the standardized estimates for our example application. We can now obtain the asymptotic variance-covariance matrix of those standardized estimates by applying equation 7, together with the derivatives derived in the previous section (equations 19 and 18). This was done for the multiple group SEM discussed earlier. The results are shown in table 4.

		Year			
	n	2007	2008	2009	2010
Education level					
Primary	279	0.799 (0.029)	0.818 (0.024)	0.828 (0.023)	0.816 (0.027)
Lower secondary	940	0.821 (0.014)	0.821 (0.013)	0.812 (0.014)	0.822 (0.014)
Middle secondary	782	0.891 (0.016)	0.879 (0.017)	0.896 (0.015)	0.877 (0.018)
Upper secondary	369	0.822 (0.015)	0.820 (0.014)	0.808 (0.015)	0.805 (0.017)
Lower tertiary	799	0.869 (0.013)	0.878 (0.012)	0.863 (0.013)	0.874 (0.013)
Upper tertiary	256	0.896 (0.017)	0.898 (0.017)	0.886 (0.018)	0.887 (0.019)

Table 4: Reliability of self-rated health in the Netherlands 2007-2010 for different educational groups. Asymptotic standard errors based on the method described in this paper are also given. For direct tests of differences between the groups in reliability, see table 5.

Table 4 shows, as is to be expected, that educational subgroups with more observations such as the Lower Secondary group have lower standard errors for the reliability coefficients. However, the sample size is not the only factor that determines the standard error of reliabilities: the value of the reliability itself exerts an influence as well, as can be confirmed by comparing the standard errors in the relatively small-sample Upper Tertiary group with those of the Primary group. The Primary group has the larger sample size (279 observations rather than 256), but the standard errors of the reliabilities of self-rated health in that group are larger. This is due to the difference in the reliability estimates, as the Upper Tertiary group has the higher reliability in all cases.

Using the obtained standard errors, one may construct pairwise hypothesis test of the hypothesis that the reliability of self-rated health is equal across groups with different levels of attained education. Table 5 shows the t-values and statistical significance of the resulting multiple comparisons – using corrected p-values (Holm, 1979) – for the reliability of self-rated health in 2007. It can be seen that this rather simplistic analysis suggests that the groups Upper Secondary and Upper Tertiary are not statistically significantly different from each other, but do differ significantly from the lower three groups.

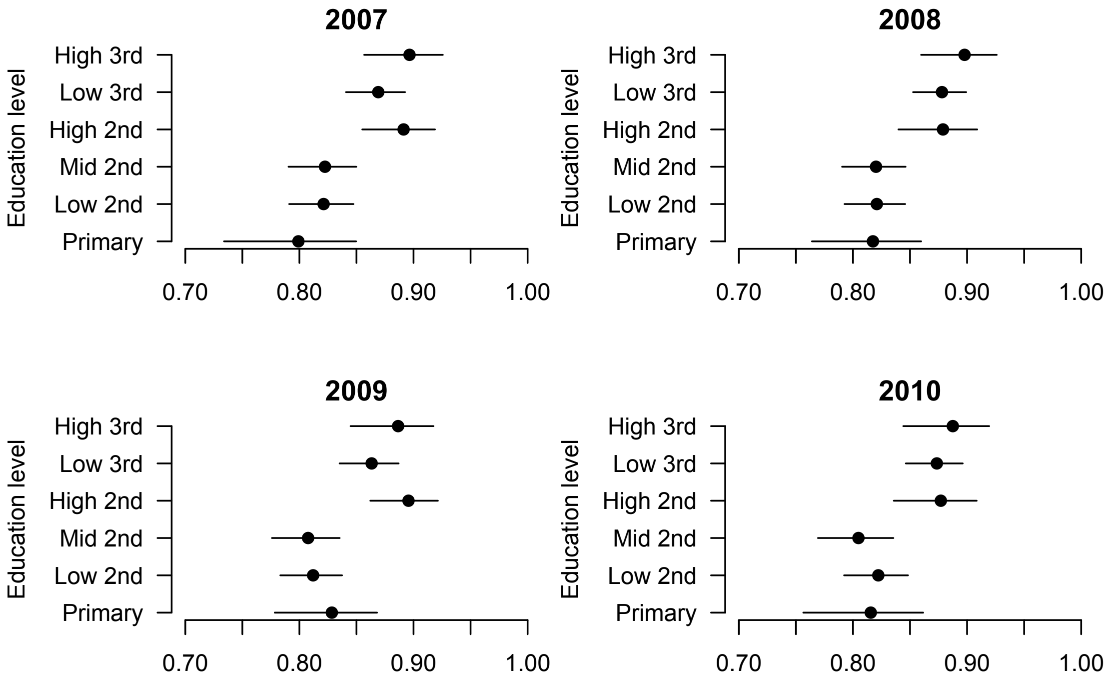
	Primary	Low 2nd	Mid 2nd	High 2nd	Low 3rd
Low 2nd	-0.68				
Mid 2nd	-0.71	-0.06			
High 2nd	-2.76*	-3.25*	3.12*		
Low 3rd	-2.18	-2.46	-2.33	1.07	
High 3rd	-2.86*	-3.35*	-3.22*	-0.22	-1.26

Table 5: t-values for comparisons of the reliability of self-rated health in 2007 between groups with different levels of education. The comparisons made can be read by examining the corresponding row and column. The cells show t-values for the test of no difference between the groups in reliability. Comparisons marked with a "*" have an adjusted p-value with $p_{\text{adj}} < 0.05$.

A different method of examining such hypotheses is the examination of 95% confidence intervals of the reliability coefficient of self-rated health in the different groups. In order to improve the normal approximation to the confidence interval for correlations, Fisher (1925) suggested the z-transformation, $z = \operatorname{arctanh}(\widetilde{\theta})$. This method can also be applied to the standardized estimates to obtain confidence intervals that cannot exceed the natural bounds of -1 and +1. By applying equation 8 to the variance matrix obtained for the example, 95% confidence intervals on the reliability coefficients were constructed. The resulting confidence intervals for the reliabilities are shown for different years and educational groups in figure 2.

Figure 2 confirms the observations that the upper three educational groups significantly differ from the lower three groups, but the 95% confidence intervals within the upper and the lower three groups overlap. Due to the use of the Fisher transformation, the confidence intervals are not symmetric around the point estimates. Another observation of interest is that none of the confidence intervals contain 0.70, the traditional cutoff point for acceptable reliability coefficients. It can be concluded that there are significant differences between people with higher and lower levels of education, but in all groups the reliability of self-rated health is acceptable.

Figure 2: Confidence intervals for the reliability coefficients (standardized λ's) constructed via the Fisher z-transformation. For each group of education level and year, the 95% C.I. is shown.



5 Discussion and conclusion

In section 2, we discussed an example analysis of the reliability of self-rated health in the Netherlands. Since this measure is widely used in public health research, it is vital that its reliability should be acceptable. Furthermore, if correlations between self-rated health and other variables are to be compared across groups with different levels of education, the reliability of self-rated health should be the same across the groups. A model allowing for the estimation of reliability was therefore estimated in six groups with different levels of educational attainment. The reliability coefficients were obtained in the different groups, but without their standard errors the questions of interest could not be answered. The example analysis of the first section therefore showed that, when standardized parameters such as reliability coefficients are of interest, it is necessary for the researcher to obtain the variance-covariance matrix of those standardized parameters in order to answer certain questions.

Although the principles applied here to obtain asymptotic standard errors and confidence intervals for standardized parameters are well-known, an explicit formula in terms only of the model's unstandardized parameters was not yet available. The subsequent section remedied this situation by providing this formula, deriving the Jacobian of the standardized parameters with respect to the unstandardized ones. In additional, the variance-covariance matrix of Fisher z-transformed standardized parameters was given for the construction of confidence intervals.

Section 4 then applied the derived equations to the unstandardized parameters obtained in the example analysis of self-rated health, shown in table 2 – thus obtaining standard errors for the reliability coefficients (standardized loadings). Confidence intervals obtained via the Fisher z-transformation were also calculated. These standard errors and confidence intervals for the standardized parameters allowed for investigation of the hypotheses that reliability does not differ across levels of education, and that the reliability is acceptable in different populations (i.e. higher than 0.70). Using the derived equations, we were able to conclude that the self-rated health measure is of acceptable reliability across groups of education, although there are significant differences in reliability between the educational groups (differential measurement error).

Due to the generality of SEM, the solution provided here encompasses many commonly used models as special cases. Standardized coefficients in (multivariate) regression (discussed in Bollen and Stine, 1990, 121), errors-in-variables models, factor analysis, and SUR models are special cases, for example. Complex sampling and non-normally distributed data are also accommodated (Muthen and Satorra, 1995). Another interesting application is meta-analysis of reliability coefficients as in Andrews (1984); Scherpenzeel and Saris (1997); Saris and Gallhofer (2007b); Alwin (2007).

In this paper we have given analytic solutions for the asymptotic variance matrices of standardized estimates in SEM and their z-transformations. These explicit solutions were not yet available in the literature. The analytic solution to this problem is not, of course, the only one possible. Other approaches include direct estimation by imposing model constraints (Chan and Kwan, 2009), bootstrapping, likelihood-based methods, and MCMC sampling of the standardized parameters. The advantage of the method presented here is that it is uses only the unstandardized parameter estimates which result from the estimation, is straightforward to implement in standard SEM software, and requires no further programming from the researcher. The question of which method provides the most adequate approximation to the true sampling distribution of the standardized parameters remains a topic for future study.

References

Alwin, D. F. (2007). *Margins of error: a study of reliability in survey measurement*. Wiley-Interscience.

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *The Public Opinion Quarterly*, 48:409–442.

Bollen, K. (1989). *Structural equations with latent variables*. Wiley New York.

Bollen, K. and Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological methodology*, 20(1):15–140.

Chan, W. and Kwan, J. L.-Y. (2009). Testing standardized effects in structural equation modeling: A model reparameterization approach. In *The 74th Meeting of the Pychometric Society*. The Psychometric Centre, Cambridge.

Cooper, H., Hedges, L., and Valentine, J. (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation Publications.

Fisher, S. (1925). *Statistical methods for research workers*. Number 5. Genesis Publishing Pvt Ltd.

Heise, D. and Bohrnstedt, G. (1970). Validity, invalidity, and reliability. *Sociological methodology*, 2:104–129.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Lundberg, O. and Manderbacka, K. (1996). Assessing reliability of a measure of self-rated health. *Scandinavian Journal of Public Health*, 24(3):218.

Magnus, J. and Neudecker, H. (1988). Matrix differential calculus with applications in statistics and econometrics.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29(1; ISSU 51):81–118.

Muthen, B. and Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological methodology*, 25:267–316.

Muthén, L. and Muthén, B. (1998). Mplus user's guide.

Neudecker, H. and Satorra, A. (1991). Linear structural relations: Gradient and Hessian of the fitting function. *Statistics & Probability Letters*, 11(1):57–61.

Oehlert, G. (1992). A note on the delta method. *The American Statistician*, 46(1):27–29.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rosseel, Y. (2011). *lavaan: Latent Variable Analysis*. R package version 0.4-8.

Saris, W. E. and Gallhofer, I. N. (2007a). *Design, evaluation, and analysis of questionnaires for survey research*. Wiley-Interscience.

Saris, W. E. and Gallhofer, I. N. (2007b). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54(1):131–151.

Scherpenzeel, A. (2011). Data collection in a probability-based internet panel: How the LISS panel was built and how it can be used. *Bulletin of Sociological Methodology*, 109(1):56.

Scherpenzeel, A. C. and Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods & Research*, 25:341.

von dem Knesebeck, O., Verde, P., and Dragano, N. (2006). Education and health in 22 European countries. *Social Science & Medicine*, 63(5):1344–1351.

Wiley, D. and Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, 35(1):112–117.