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

Daniel Oberski

Universitat Pompeu Fabra, Spain

June 2, 2011

Abstract

In structural equation models (SEM), the standard errors not only of parameters of the model, but also of standardized parameters may be of interest to the researcher. Examples are the comparison of reliability coefficients across different groups, and meta-analysis of standardized quantities.

We provide the analytical asymptotic variance matrix of standardized parameters in SEM, expressed only in terms of the model parameters. The expression is straightforward to implement in standard SEM software. The asymptotic variance matrix of Fisher z-transformed standardized parameters is also provided, allowing for the construction of confidence intervals. We demonstrate the use of the derived expression on an example 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 (Muthén and Satorra, 1995; Muthén, 2002).

Researchers' interest often focuses on the so-called 'standardized' parameters of the model (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 applied here to derive standard errors and confidence intervals for standardized parameters is well known (e.g. Oehlert, 1992), to our knowledge the literature does not provide the asymptotic standard errors of standardized coefficients in SEM directly as a function of the parameter estimates. As a consequence, standard errors and confidence intervals for these coefficients are usually not provided by the standard software[1]. We

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

remedy this situation by deriving an expression for the asymptotic variance-covariance matrix of the standardized coefficients in terms of the unstandardized parameter estimates. The solution requires only the parameter estimates of the model and can be readily implemented in SEM software.

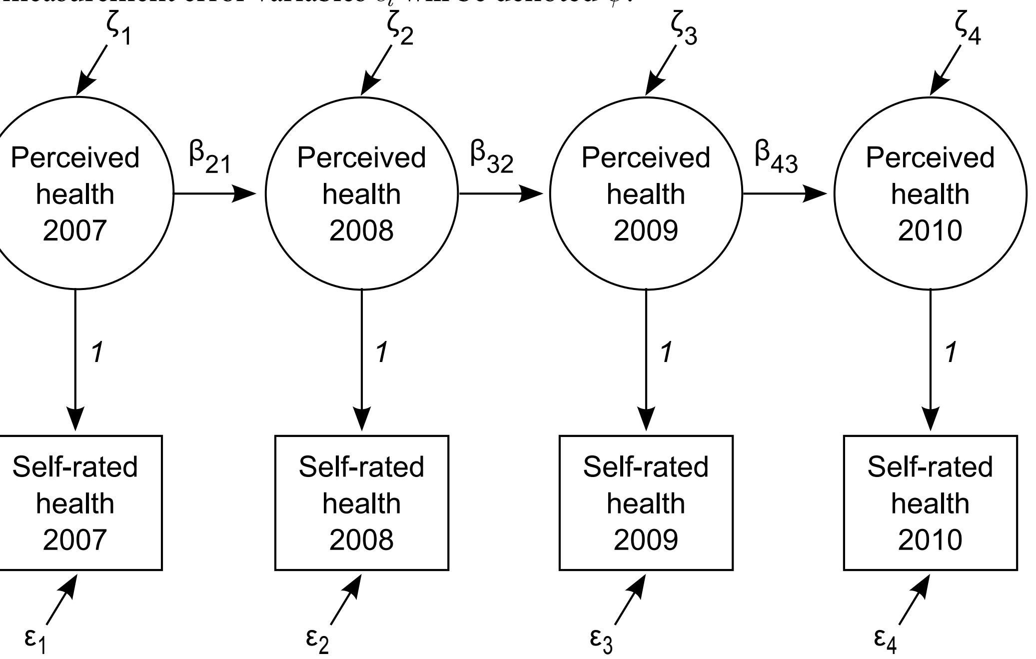
Section 2 provides a motivating example, using a SEM model with latent variables where standardized parameters and their standard errors and confidence intervals are of interest to the researcher. Section 3 derives the asymptotic variance-covariance matrix of the standardized estimates. Section 4 applies this expression to the example. A Monte Carlo study summarized in section 5 then evaluates the performance of the analytical standard errors and confidence intervals given here by simulation. The last section concludes, discusses the scope and limitations of this proposal, and suggests future research.

2 Example SEM with interest in standardized parameters

Several researchers in public health have studied differences in self-rated health across societal groups. 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. Steenkamp and Baumgartner, 1998; Saris and Gallhofer, 2007a). For example, for comparisons of correlations between self-rated health and other measures across groups with different levels of educations to be meaningful, the reliability of self-rated health should be equal for people with different educational attainments. Therefore not only between-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 suitably long interval: once a year, for, instance. 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

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.



overview of other designs for the estimation of reliability of single questions 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.

We estimate this model using the LISS panel study, a probability sample of 8735 Dutch citizens, who regularly 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?

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

		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 (2007–2010), in groups with different levels of education.

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. Table 1 shows the mean and standard deviation of self-rated health in groups with six different levels of education. We estimate the quasi-simplex model shown in figure 1 as a multiple group SEM to yield four standardized loadings for each of the six educational groups, 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.

The last row of table 2 shows that the error variance parameter ψ differs somewhat across groups with different levels of education. The differences are statistically significant ($p < 0.01$), as revealed by a test that compared the estimated model with a model containing equality constraints on ψ across the groups. However, it is not clear whether the reliability coefficients differ across the groups also. Since the reliability is a function of both the error variances and the total variances of the latent variables (which involve the β_{ji} and ψ

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

	Group: education level					
Par.	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	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

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 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.

		Year			
	n	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.

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; in order to evaluate this last possibility the variance matrix of the standardized parameters is needed.

Another type of analysis that would require the variances of these estimates are meta-analyses of the reliabilities; to combine the reliability coefficients across years or educational groups for the purpose of secondary analysis, their variance-covariance matrix would be required (Cooper et al., 2009, 261, 271-2). Such secondary analyses may, for instance, be used to assess the effect of the choice of survey mode, sampling characteristics, or other question characteristics on the reliability – examples are Andrews (1984); Scherpenzeel and Saris (1997); Saris and Gallhofer (2007b); Alwin (2007).

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. Structural equation models can be formulated as:

$$y = \Lambda\eta + \epsilon \tag{1}$$

$$\eta = B_0\eta + \zeta, \tag{2}$$

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.

The model described corresponds to the so-called LISREL "all-y" framework (Jöreskog and Sörbom, 1996). It also encompasses a variety of other models due to the fact the LISREL all-y model is equivalent to other SEM frameworks in use by standard software such as the Bentler-Weeks and RAM models (Bentler and Weeks, 1980; McArdle and McDonald, 1984).

The parameters of the model are collected in a parameter vector

$$\theta \equiv [\text{vec}\,\Lambda, \text{vec}\,B_0, \text{vech}\,\Phi, \text{vech}\,\Psi] \equiv [\lambda, \beta, \phi, \psi].$$

In general model 4 is not identified without additional restrictions. Typically these restrictions set certain elements to a constant or impose equality constraints. Assume, therefore, that θ is a continuously differentiable function $\theta = \theta(\delta)$, where δ is a t-dimensional vector of free parameters of the model, which may then be estimated by minimization of a discrepancy function (Shapiro, 1985). We also define $A \equiv \partial\theta(\delta)/\partial\delta$.

Interest focuses not only on the parameter vector θ, but also on the so-called "standardized" parameter vector, denoted $\tilde{\theta} \equiv [\text{vec}\,\tilde{\Lambda}, \text{vec}\,\tilde{B}_0, \text{vech}\,\tilde{\Phi}, \text{vech}\,\tilde{\Psi}] \equiv [\tilde{\lambda}, \tilde{\beta}, \tilde{\phi}, \tilde{\psi}]$., where:

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

$$\tilde{B}_0 \equiv D_\eta^{-1} B_0 D_\eta, \tag{6}$$

$$\tilde{\Phi} \equiv D_\eta^{-1} \Phi D_\eta^{-1}, \tag{7}$$

$$\tilde{\Psi} \equiv D_y^{-1} \Psi D_y^{-1}, \tag{8}$$

and $D_y \equiv \sqrt{I \circ \Sigma_y}$, and $D_\eta \equiv \sqrt{I \circ \Sigma_\eta}$. Given these definitions, the explained variances in

η and y, respectively R^2_η and R^2_y, which may also be of interest, equal the diagonal elements of the matrices $I - \tilde{\Phi}$ and $I - \tilde{\Psi}$.

By application of the delta method, the asymptotic variance of $\tilde{\theta}$ can be shown to equal

$$\text{var}(\tilde{\theta}) = \left(\frac{\partial\tilde{\theta}}{\partial\theta}\right) A\text{var}(\delta) A' \left(\frac{\partial\tilde{\theta}}{\partial\theta}\right)'. \tag{9}$$

Here $\text{var}(\delta)$ is the appropriate asymptotic variance matrix of the free model parameters δ (e.g. Satorra, 1989, 143-4). The choice of the appropriate $\text{var}(\delta)$ matrix allows for incorporation of non-normal and complex sample data with possibly missing observations (Muthén and Satorra, 1995).

The variances of $\tilde{\psi}$ and $\tilde{\phi}$ from equation 9 also provide the variances of R^2_y and R^2_η. Therefore we will not derive the variance of the explained variances separately, with the understanding that the derivations for $\tilde{\psi}$ and $\tilde{\phi}$ already provide these[4].

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

$$\text{var}[\text{arctanh}(\tilde{\theta})] = T\,\text{var}(\tilde{\theta})\,T', \tag{10}$$

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

3.1 Derivatives of the standardized parameters

The derivatives $\partial\tilde{\theta}/\partial\theta$ in equation 9 are not available in the literature and are derived here.

From definition 5,

$$\begin{aligned} \text{d}\,\text{vec}\,\tilde{\Lambda} = (D_\eta \Lambda' \otimes I_p)\,\text{d}\,\text{vec}\,D_y^{-1} + (D_\eta \otimes D_y^{-1})\,\text{d}\,\text{vec}\,\Lambda + \\ (I_q \otimes D_y^{-1}\Lambda)\,\text{d}\,\text{vec}\,D_\eta, \end{aligned} \tag{11}$$

[4] Let $R^2_y = Q\text{vech}\,(I - \tilde{\Psi})$, where Q is a selection matrix selecting diagonal elements. Then $\text{var}(R^2_y) = Q(\partial\tilde{\psi}/\partial\theta)\text{var}(\theta)(\partial\tilde{\psi}/\partial\theta)'Q'$. The same argument applies to $\text{var}(R^2_\eta)$.

and from definition 6,

$$\mathrm{d}\,\mathrm{vec}\,\tilde{B}_0 = (D_\eta B_0' \otimes I_p)\,\mathrm{d}\,\mathrm{vec}\,D_\eta^{-1} + (D_\eta \otimes D_\eta^{-1})\,\mathrm{d}\,\mathrm{vec}\,B_0 + (I_q \otimes D_\eta^{-1} B_0)\,\mathrm{d}\,\mathrm{vec}\,D_\eta. \quad (12)$$

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:

$$\begin{aligned}\mathrm{d}\,\mathrm{vec}\,\Sigma_y = (I+K)(\Lambda B^{-1}\Phi B^{-T} \otimes I)\,\mathrm{d}\,\mathrm{vec}\,\Lambda + \\ (I+K)(\Lambda B^{-1}\Phi B^{-T} \otimes \Lambda B^{-1})\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ (\Lambda B^{-1} \otimes \Lambda B^{-1})P\,\mathrm{d}\,\mathrm{vech}\,\Phi + \\ \mathrm{d}\,\mathrm{vech}\,\Psi,\end{aligned} \quad (13)$$

where the commutation matrix K, the duplication matrix P, and the operators vec and vech are defined in Magnus and Neudecker (1988) Let $\Delta_{y\lambda}$ equal the coefficient matrix of $\mathrm{d}\,\mathrm{vec}\,\Lambda$ in equation 13, $\Delta_{y\beta}$ the coefficient of $\mathrm{d}\,\mathrm{vec}\,B_0$, and $\Delta_{y\phi}$ the coefficient of $\mathrm{d}\,\mathrm{vech}\,\Phi$.

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

$$\begin{aligned}\mathrm{d}\,\mathrm{vec}\,\Sigma_\eta = (I+K)(B^{-1}\Phi B^{-T} \otimes B^{-1})\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ (B^{-1} \otimes B^{-1})P\,\mathrm{d}\,\mathrm{vech}\,\Phi.\end{aligned} \quad (14)$$

Let $\Delta_{\eta\beta}$ and $\Delta_{\eta\phi}$ be the coefficients of $\mathrm{d}\,\mathrm{vec}\,B_0$ and $\mathrm{d}\,\mathrm{vech}\,\Phi$ respectively in equation 14.

Also, let

$$G^{(\sigma_y^{-1})} \equiv -(I_p \otimes \frac{1}{2(I_p \circ \Sigma_y)^{\frac{3}{2}}})\mathrm{diag}[\mathrm{vec}\,(I_p)], \quad (15)$$

$$G^{(\sigma_\eta)} \equiv (I_q \otimes \frac{1}{2(I_q \circ \Sigma_\eta)^{\frac{1}{2}}})\mathrm{diag}[\mathrm{vec}\,(I_q)], \quad (16)$$

$$G^{(\sigma_\eta^{-1})} \equiv -(I_q \otimes \frac{1}{2(I_q \circ \Sigma_\eta)^{\frac{3}{2}}})\mathrm{diag}[\mathrm{vec}\,(I_q)]. \quad (17)$$

The differential of the standardized $\tilde{\Lambda}$ matrix may then be written

$$\begin{aligned}\mathrm{d}\,\mathrm{vec}\,\tilde{\Lambda} =& [(D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} \Delta_{y\lambda} + (D_\eta \otimes D_y^{-1})]\,\mathrm{d}\,\mathrm{vec}\,\Lambda + \\ & [(D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} \Delta_{y\beta} + (I_q \otimes D_y^{-1}\Lambda) G^{(\sigma_\eta)} \Delta_{\eta\beta}]\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ & [(D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} \Delta_{y\phi} + (I_q \otimes D_y^{-1}\Lambda) G^{(\sigma_\eta)} \Delta_{\eta\phi}] P\,\mathrm{d}\,\mathrm{vech}\,\Phi + \\ & (D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} P\,\mathrm{d}\,\mathrm{vech}\,\Psi. \end{aligned} \tag{18}$$

And similarly, the differential of the standardized $\tilde{B}_0$ matrix from equation 12 is

$$\begin{aligned}\mathrm{d}\,\mathrm{vec}\,\tilde{B}_0 =& [\{(D_\eta B_0' \otimes I_q) G^{(\sigma_\eta^{-1})} + (I_q \otimes D_\eta^{-1} B_0) G^{(\sigma_\eta)}\}\Delta_{\eta\beta} + (D_\eta \otimes D_\eta^{-1})]\,\mathrm{d}\,\mathrm{vec}\,B_0 + \\ & [(D_\eta B_0' \otimes I_q) G^{(\sigma_\eta^{-1})} + (I_q \otimes D_\eta^{-1} B_0) G^{(\sigma_\eta)}]\Delta_{\eta\phi} P\,\mathrm{d}\,\mathrm{vech}\,\Phi. \end{aligned} \tag{19}$$

From the differentials in equations 18 and 19, we conclude that the derivative matrices of the standardized parameters vec $\tilde{\Lambda}$ and vec $\tilde{B}$ with respect to the parameters of the model θ will be the partitioned matrices

$$\begin{aligned}\frac{\partial \tilde{\lambda}}{\partial \theta} = [& (D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} \Delta_{y\lambda} + (D_\eta \otimes D_y^{-1}), \\ & (D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} \Delta_{y\beta} + (I_q \otimes D_y^{-1}\Lambda) G^{(\sigma_\eta)} \Delta_{\eta\beta}, \\ & (D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} \Delta_{y\phi} P + (I_q \otimes D_y^{-1}\Lambda) G^{(\sigma_\eta)} \Delta_{\eta\phi} P, \\ & (D_\eta \Lambda' \otimes I_p) G^{(\sigma_y^{-1})} P] \end{aligned} \tag{20}$$

$$\begin{aligned}\frac{\partial \tilde{\beta}_0}{\partial \theta} = [& \mathbf{0}, \{(D_\eta B_0' \otimes I_q) G^{(\sigma_\eta^{-1})} + (I_q \otimes D_\eta^{-1} B_0) G^{(\sigma_\eta)}\}\Delta_{\eta\beta} + (D_\eta \otimes D_\eta^{-1}), \\ & \{(D_\eta B_0' \otimes I_q) G^{(\sigma_\eta^{-1})} + (I_q \otimes D_\eta^{-1} B_0) G^{(\sigma_\eta)}\}\Delta_{\eta\phi} P, \mathbf{0}]. \end{aligned} \tag{21}$$

By a similar derivation (not shown here), we also obtain the derivatives of vech $\tilde{\Phi}$, and vech $\tilde{\Psi}$ with respect to the parameters.

$$\frac{\partial \tilde{\phi}}{\partial \theta} = [\mathbf{0}, \{(D_\eta^{-1}\Phi \otimes I_q) + (I_q \otimes D_\eta^{-1}\Phi)\}G^{(\sigma_\eta^{-1})}\Delta_{\eta\beta},$$
$$\{(D_\eta^{-1}\Phi \otimes I_q) + (I_q \otimes D_\eta^{-1}\Phi)\}G^{(\sigma_\eta^{-1})}\Delta_{\eta\phi}P + (D_\eta^{-1} \otimes D_\eta^{-1})P, \mathbf{0}]. \quad (22)$$

$$\frac{\partial \tilde{\psi}}{\partial \theta} = [\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}G^{(\sigma_y^{-1})}\Delta_{y\lambda},$$
$$\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}G^{(\sigma_y^{-1})}\Delta_{y\beta},$$
$$\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}G^{(\sigma_y^{-1})}\Delta_{y\phi}P,$$
$$\{(D_y^{-1}\Psi \otimes I_p) + (I_p \otimes D_y^{-1}\Psi)\}G^{(\sigma_y^{-1})}P + D_y^{-1} \otimes D_y^{-1}P] \quad (23)$$

By stacking the derivatives in equations 20–23 rowwise, we obtain the matrix $d\tilde{\theta}/d\theta$:

$$\frac{\partial \tilde{\theta}}{\partial \theta} = [\left(\frac{\partial \tilde{\lambda}}{\partial \theta}\right)', \left(\frac{\partial \tilde{\beta}_0}{\partial \theta}\right)', \left(\frac{\partial \tilde{\phi}}{\partial \theta}\right)', \left(\frac{\partial \tilde{\psi}}{\partial \theta}\right)']'. \quad (24)$$

This equation can be applied to equation 9 to to obtain the variance matrix of the standardized parameters, while equation 10 provides the variance matrix of z-transformed standardized parameters. Consistent estimates of both may be obtained by replacing the parameters by their sample estimates (Satorra, 1989).

4 Application: standard errors of standardized parameters

Table 3 showed the standardized estimates for our example application. We can now obtain a consistent estimate of the asymptotic variance matrix of those estimates by applying equation 9, together with the derivatives derived in equation 24, replacing the parameter matrices by their sample estimates. The resulting standard errors for the different groups are shown in table 4.

Table 4 shows, as is to be expected, that educational subgroups with more observations

	n	*Year* 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.

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 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 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.

The researcher may also wish to examine 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, $z = \text{arctanh}(\tilde{\theta})$. 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 10 to the variance matrix obtained for the example and replacing the standardized parameters in that equation with their sample estimates, 95% confidence intervals on the reliability coefficients were constructed. The resulting confidence intervals for the reliabilities are shown for different years and educational groups as a coefficient plot 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.

5 Monte Carlo evaluation of the approach

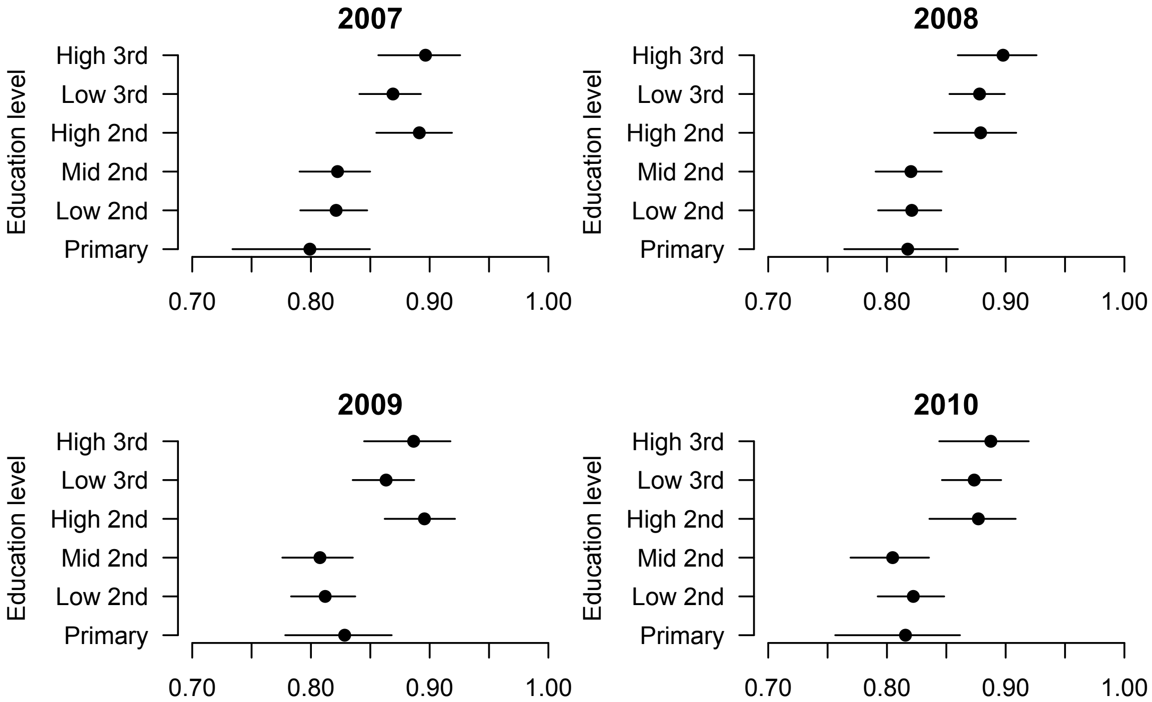
To evaluate the performance of the asymptotic standard errors and confidence intervals in the example, a Monte Carlo study was performed. The multiple group model estimated in section 2 is taken as the true population model, and 512 samples from this population model were simulated[5]. Thus each sample contains six groups of different education levels, and the

[5]All the analyses were performed in R 2.13.0 64-bit (R Development Core Team, 2011) .

	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 (rows versus columns). The cells show t-values for the test of no difference between the groups in reliability. Comparisons marked with "*" have an adjusted $p < 0.05$.

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.



four observed variables within each group are drawn from a multivariate normal distribution with population covariance matrix and sample size equal to the implied covariance matrix and sample size in that group from section 2 .

In each of the 512 simulated samples, the model shown in figure 1 was estimated, and the standardized loadings[6] (reliability coefficients) and a consistent estimate of their asymptotic variance matrix was calculated. Note that the analysis is carried out separately for each of the groups, some of which have a small sample size (see table 4). We then calculated the bias in the estimates by subtracting the average standardized loading over simulations from the corresponding true population value given in table 4. The asymptotic standard errors from table 4 were subtracted from the corresponding standard deviations of the standardized estimates over simulated samples. The bias in the standardized estimates and their asymptotic standard errors over the 512 simulations for the 24 standardized loadings is shown in table 6.

[6]Since the interest in this example focuses on the standardized loadings and for the sake of brevity, we will focus only on these standardized loadings here.

	2007			2008			2009			2010		
	Bias			Bias			Bias			Bias		
Grp	$\hat{\tilde{\lambda}}$	s.e	cvg.	$\hat{\tilde{\lambda}}$	s.e	cvg.	$\hat{\tilde{\lambda}}$	s.e	cvg.	$\hat{\tilde{\lambda}}$	s.e	cvg.
1	0.003	-0.003	95%	0.003	-0.002	94%	0.004	-0.002	93%	0.003	-0.002	94%
2	0.001	0.001	97%	0.001	0.001	96%	0.000	0.000	96%	0.001	0.000	94%
3	-0.001	0.000	95%	-0.001	0.000	96%	-0.001	-0.000	96%	-0.000	-0.000	96%
4	-0.000	-0.000	94%	0.000	-0.000	94%	-0.000	0.001	95%	0.000	-0.000	94%
5	-0.000	0.000	95%	0.000	-0.000	94%	-0.000	-0.001	93%	-0.000	-0.000	93%
6	0.000	-0.001	96%	-0.000	-0.000	95%	0.000	-0.001	95%	0.001	-0.000	95%

Table 6: Monte Carlo results of 512 random samples taking the example model in section 2 as the true model. Shown are the bias in the standardized estimates and their asymptotic standard errors as discussed above. Also shown is the coverage of 95% confidence intervals via the Fisher z-transformation.

It can be seen in table 6 that the standardized loadings are unbiased estimates of the true population values. The absolute relative bias (not shown here) was smaller than 0.5% in all cases. The asymptotic standard errors are very close to the standard deviations observed over the simulations, although there are some deviations. The absolute relative bias of the standard errors as compared to the observed standard deviations had a maximum of 10% (for the reliability coefficient in 2007 in the lowest educational group).

Confidence intervals were obtained in each sample as follows:

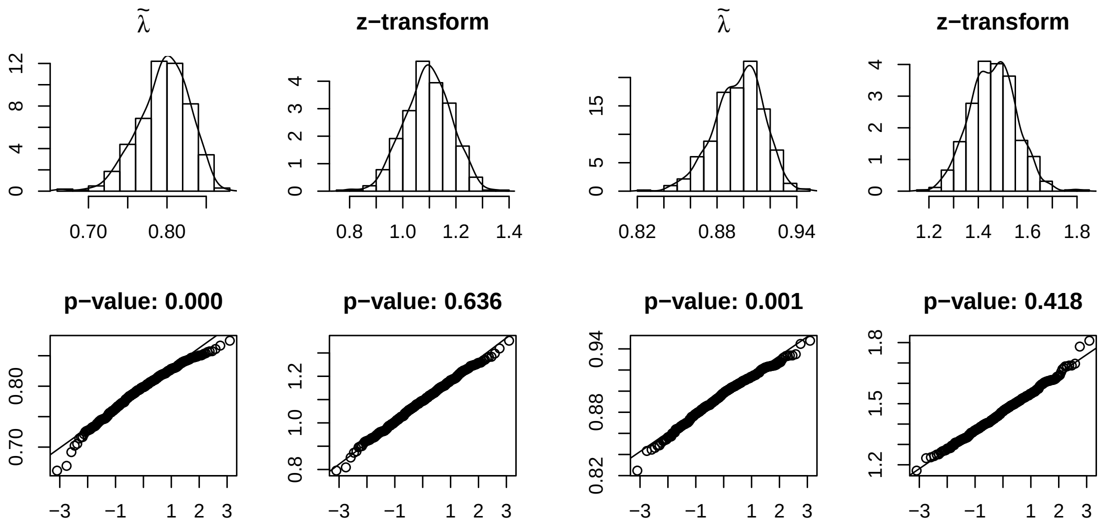
1. The variance of z-transformed standardized loadings was calculated from the parameter values;

2. Standard errors of the z-transformed standardized loadings were used to construct a 95% confidence interval for $z(\tilde{\lambda})$;

3. The upper and lower limits of the 95% confidence interval for $z(\tilde{\lambda})$ were back-transformed to the original scale as $z^{-1}(x) = \tanh(x)$.

It was then evaluated in each sample whether the nominal 95% confidence interval thus obtained contained the true population value shown in table 4 or not. The percentage of samples in which this was the case is shown in table 6 in the columns marked "cvg.".

Table 6 shows that in all cases the coverage of the 95% confidence confidence intervals using the Fisher z-transform is close to 95%, with the lowest observed coverage 93% and the



(a) Primary education

(b) University education

Figure 3: Distribution and normal Q-Q plots of the reliability coefficients $\tilde{\lambda}$ and their Fisher z-transforms in 2007 in the lowest and highest educational group. The p-value for a Shapiro-Wilks test of normality is also shown.

highest coverage 97%. The mean and median coverage of the 95% coverage intervals across the 24 standardized parameters are 94.7% and 95.0% respectively. The confidence intervals via the Fisher z-transform therefore appear to perform well.

Some authors have expressed concern about the use of the normal distribution as an approximation for the sampling distribution of standardized parameters. We therefore examined the quality of this approximation for all standardized loadings. Figure 3 presents the results for one of the standardized loadings in the lowest and highest educational groups. It can be seen in figure 3 that the distribution of the raw standardized loading in both cases is somewhat skewed, and indeed a Shapiro-Wilks test of normality is rejected in both cases. The effect appears to be more pronounced for the Primary group, which has the highest variance (see table 4). The Fisher z-transformed quantities do appear to follow a normal distribution in these two cases.

More generally, 9 out of 24 Shapiro-Wilks tests for the raw standardized estimates are rejected (when using adjusted p-values), while none of the tests for transformed estimates are. Additionally, in exactly two-thirds of the cases the fit to a normal distribution is better for the transformed than for the raw standardized values. When the normal distribution does not provide an accurate approximation to the sampling distribution of the standardized estimates, confidence intervals via the z-transformation may still be expected to provide accurate coverage.

This result is of course in line with the classic results for correlations (e.g. Johnson et al., 1995, chapter 32), and indeed in the example model, the standardized loadings are correlations: the difference in this case is that the classical theory deals with correlations between observed variables, while the standardized loadings examined here are correlations between an observed variable and a latent variable. In addition, it should be clear that the presentation given here is not limited to models where the standardized parameters represent correlations.

In summary, this short simulation study shows that for the example the suggested procedure performs well. The standard errors for standardized parameters are very close to the observed standard deviation across the simulations. Additionally, the derived asymptotic confidence intervals via the Fisher z-transformation show good coverage properties, while the sampling distribution of the transformed standardized parameters is the better approximated by the normal distribution. Note that the sample sizes in some of the groups are relatively small, with n ranging between 256 and 799 observations (see first column of table 4). In spite of these smaller sample sizes the approach taken here of providing asymptotic standard errors and confidence intervals for the standardized estimates works well in this example.

6 Discussion and conclusion

In this paper we have given analytic solutions for the asymptotic variance matrices of standardized estimates in SEM and their z-transformations, which were not yet available in the literature.

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, providing the derivative matrix of the standardized parameters with respect to the unstandardized ones. In additional, the variance-covariance matrix of Fisher z-transformed standardized parameters was given to allow the construction of confidence intervals.

Section 4 then obtained standard errors for the reliability coefficients (standardized loadings) in the example analysis of self-rated health. Confidence intervals were calculated via the Fisher z-transformation. 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). We were able to conclude that the self-rated health question has acceptable reliability across groups of education, although there are significant differences in reliability between the groups (differential measurement error).

Subsequently section 5 evaluated the asymptotic standard errors and confidence intervals for the example model by simulation. It was found that the confidence intervals for standardized parameters indeed provided the nominal coverage rates.

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 can also be accommodated (Muthén and Satorra, 1995). Another application is meta-analysis of reliability coefficients as in Andrews (1984); Scherpenzeel and Saris (1997); Saris and Gallhofer (2007b); Alwin (2007).

Our discussion of providing standard errors and confidence intervals for standardized estimates has parallels with the classical work on correlations between observed variables (e.g. Johnson et al., 1995, chapter 32); it is more general than that theory, however, in the sense that standardized parameters in SEM may comprise not only correlations between observed variables, but also between latent variables, as well as multiple regression coefficients of latent and observed variables, factor loadings, error correlations, and R^2 measures for both observed and latent variables.

The analytic solution to the problem of standard errors and confidence intervals for standardized parameters is not, of course, the only one possible. Other approaches include direct estimation by imposing model constraints (Chan and Kwan, 2009), analysis of correlation structures with constraints (Bentler and Lee, 1983), 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. Our Monte Carlo study suggests that the approach works well in the example given, and with moderately small samples, but a systematic investigation of the conditions under which the approach may be expected to perform well remains to be done. 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.

Bentler, P. and Lee, S. (1983). Covariance structures under polynomial constraints: Applications to correlation and alpha-type structural models. *Journal of Educational and Behavioral Statistics*, 8(3):207.

Bentler, P. M. and Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45(3):289–308.

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.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Distributions in statistics: Continuous univariate distributions*, volume 2. John Wiley & Sons, 2nd edition.

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8 user's reference guide*. Scientific Software.

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.

McArdle, J. and McDonald, R. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37(2):234–251.

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

Muthén, 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.

Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statistical Journal*, 19:73–81.

Steenkamp, J. and Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research*, pages 78–90.

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.