Robust standard errors and chi-squares

Contents

1. Introduction

This document describes how we implemented the computation of robust standard errors and chi-squares in LISREL. Several authors have contributed to the statistical inference theory for covariance structures in single and multiple groups, notably Browne (1977, 1984, 1987), Jöreskog (1981), and Satorra (1987, 1993). The purpose of this document is not to review this literature, but rather to give the formulas we use in LISREL to compute standard errors and chi-squares.

It should be noted that these formulas are for continuous variables. The theory for ordinal variables is another matter.

1. Single group: Covariance structures
2. Single group: Mean and covariance structures
3. Single group: Augmented moment matrices
4. Multiple groups: covariance structures
5. Multiple groups: Mean and covariance structures
6. Multiple groups: Augmented moment matrices

2. Single group: covariance structures

2.1 Definitions

k = number of observed variables (1)

$$s = \frac{1}{2}k(k+1) \quad (2)$$

t = number of independent parameters $< s$ (3)

$$d = s - t \quad (4)$$

$\mathbf{K} : k^2 \times s = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}$, where $\mathbf{D} : k^2 \times s$ is the duplication matrix (5)

$$\mathbf{s} : s \times 1 = \mathbf{K}'\text{vec}(\mathbf{S}), \quad vec(\mathbf{S}) = \mathbf{Ds} \quad (6)$$

$$\mathbf{\Omega} : s \times s = n\text{ACov}(\mathbf{s}) \quad n = N - 1 \quad (7)$$

$$\mathbf{W} : s \times s = n\text{Est}\left[\text{ACov}(\mathbf{s})\right] \quad (8)$$

$\mathbf{W}_{\text{NT}} = 2\mathbf{K}'\left(\hat{\mathbf{\Sigma}} \otimes \hat{\mathbf{\Sigma}}\right)\mathbf{K}$ under NT, *i.e.*, if AC is not read or computed (9)

$\mathbf{W}_{\text{NNT}} = \left[w_{gh,ij}\right]$, where $w_{gh,ij}$ is defined as in (14) (10)

$\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$ with $\boldsymbol{\sigma} : s \times 1, \boldsymbol{\theta} : t \times 1$ (11)

$\mathbf{\Delta} : s \times t = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\theta}'}$ evaluated at $\hat{\boldsymbol{\theta}}$ and assumed to have rank t (12)

$\mathbf{\Delta}_c : s \times d$ = orthogonal complement to $\mathbf{\Delta}$ (13)

$\mathbf{\Delta}_c'\mathbf{\Delta} = \mathbf{0}$ $[\mathbf{\Delta} \,|\, \mathbf{\Delta}_c]$ non-singular

Comments:

- In LISREL notation $k = p + q$ where p and q are the numbers of y- and x-variables, respectively.
- s is the number of independent elements of the sample covariance matrix **S**.
- The assumption $t < s$ implies that we exclude the case of saturated models where $t = s$. For such models all residuals and chi-squares are zero. The asymptotic covariance matrix of parameter estimates can still be obtained by (24) but some of the other formulas break down. For example, $\mathbf{\Delta}_c$ does not exist.
- We assume that the model is identified so that d is the degrees of freedom of the model.
- $\text{vec}(\mathbf{S})$ is a vector of order $k^2 \times 1$ consisting of the columns of **S** stacked under each other.

- $\mathbf{s} = (s_{11}, s_{21}, s_{22}, s_{31}, \ldots, s_{kk})'$ is a vector or order $s \times 1$ consisting of the nonduplicated elements of **S**. **D** is the duplication matrix (see Magnus & Neudecker, 1988) which transforms **s** to $\text{vec}(\mathbf{S})$. $\mathbf{K}'$ is the generalized inverse of **D**, which transforms $\text{vec}(\mathbf{S})$ to **s**.
- $\mathbf{\Omega}$ is *n* times the asymptotic covariance matrix of **s**. This is unknown but can be estimated by **W**. The estimated asymptotic covariance matrix **W** is abbreviated AC.
- Under normal theory NT, *i.e.*, if the observations come from a multivariate normal population or if **S** has a Wishart distribution, **W** can be estimated by $\mathbf{W}_{\mathbf{NT}}$ in (9). Here $\otimes$ is the symbol for a Kronecker product. LISREL uses this formula to estimate **W** if the AC matrix is not read or computed. In (9), $\hat{\mathbf{\Sigma}}$ is the fitted covariance matrix obtained after the model has been estimated.
- In the non-normal case NNT, *i.e.*, if the observations come from a distribution which is not multivariate normal, the elements of **W** can be estimated by the formula

$$w_{gh,ij} = n\text{Est}[\text{ACov}(s_{gh}, s_{ij})] = m_{ghij} - s_{gh}s_{ij} \tag{14}$$

where

$$s_{gh} = (1/N)\sum_{a=1}^{N}(z_{ag} - \bar{z}_g)(z_{ah} - \bar{z}_h)$$

and

$$m_{ghij} = (1/N)\sum_{a=1}^{N}(z_{ag} - \bar{z}_g)(z_{ah} - \bar{z}_h)(z_{ai} - \bar{z}_i)(z_{aj} - \bar{z}_j) \tag{15}$$

is a fourth-order central sample moment. This defines $\mathbf{W}_{\text{NNT}}$. PRELIS computes $\mathbf{W}_{\text{NNT}}$ for continuous variables and LISREL will use this matrix if it is read or the LISREL RO command or the SIMPLIS Robust Estimation command is used with an LSF.
- The covariance structure model is $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{\theta})$, *i.e.*, the elements of $\mathbf{\Sigma}$ are functions of the parameters $\mathbf{\theta}$. In terms of the nonduplicated elements of $\mathbf{\Sigma}$, this is expressed as in (11).
- $\mathbf{\Delta}$ is a matrix of derivatives. It can be evaluated at any point in the parameter space. However, its operational use is when it is evaluated at the parameter estimates, and this is assumed in what follows.
- Since $\mathbf{\Delta}$ is of rank $t < s$, there exists $s - t = d$ columns orthogonal to the columns in $\mathbf{\Delta}$. These *d* columns form the orthogonal complement $\mathbf{\Delta}_c$. This is not unique but any orthogonal complement will do.

2.2 Fit functions

With these definitions, all fit functions in LISREL are of the same form:

$$\text{F} = (\mathbf{s} - \mathbf{\sigma})'\mathbf{V}(\mathbf{s} - \mathbf{\sigma}), \quad \mathbf{V} : s \times s$$

where the weight matrix **V** is defined differently for different estimation methods:

$$\text{ULS:} \quad \mathbf{V} = \mathbf{I}^* = \text{Diag}(1, 2, 1, 2, 2, 1, \ldots) \tag{16}$$

$$\text{GLS:} \quad \mathbf{V} = \mathbf{D}'(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\mathbf{D} \tag{17}$$

ML: $$\mathbf{V} = \mathbf{D}'\left(\hat{\mathbf{\Sigma}}^{-1} \otimes \hat{\mathbf{\Sigma}}^{-1}\right)\mathbf{D} \tag{18}$$

WLS: $$\mathbf{V} = \mathbf{W}_{\text{NNT}}^{-1} \text{ or } \mathbf{W}_{\text{NNT}}^{-} \text{ if } \mathbf{W}_{\text{NNT}} \text{ is singular} \tag{19}$$

$$\mathbf{V} = \mathbf{D}_{\text{W}}^{-1} = [\text{Diag}(\mathbf{W})]^{-1} \tag{20}$$

DWLS: with $\mathbf{W} = \mathbf{W}_{\text{NT}}$ if AC not read or computed

and $\mathbf{W} = \mathbf{W}_{\text{NNT}}$ if AC read or computed

Comments:

- The fit function for ML is usually written

$$\text{F} = \log\|\mathbf{\Sigma}\| + \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \log\|\mathbf{S}\| - k \tag{21}$$

but an asymptotically equivalent form is

$$\text{F} = (\mathbf{s} - \boldsymbol{\sigma})'\mathbf{D}'\left(\hat{\mathbf{\Sigma}}^{-1} \otimes \hat{\mathbf{\Sigma}}^{-1}\right)\mathbf{D}(\mathbf{s} - \boldsymbol{\sigma}) \tag{22}$$

- Equation (22) can be interpreted as ML estimated by means of iteratively reweighted least squares in which $\hat{\mathbf{\Sigma}}$ is updated in each iteration. Both of these fit functions have a minimum at the same point in the parameter space, namely at the ML estimates. However, the minimum value of the functions is not the same.

2.3 Results

$$\mathbf{E} : q \times q \quad = \mathbf{\Delta}'\mathbf{V}\mathbf{\Delta} \tag{23}$$

$$n\text{ACov}(\hat{\boldsymbol{\theta}}) \quad = \mathbf{E}^{-1}\mathbf{\Delta}'\mathbf{V}\mathbf{W}\,\mathbf{V}\,\mathbf{\Delta}\mathbf{E}^{-1} \tag{24}$$

$$n\text{ACov}(\mathbf{s} - \hat{\boldsymbol{\sigma}}) \quad = \mathbf{W} - \mathbf{\Delta}\,\mathbf{E}^{-1}\mathbf{\Delta}' \tag{25}$$

$$\text{with } \mathbf{W} \quad = \mathbf{W}_{\text{NT}} \text{ if AC not read or computed} \tag{26}$$

$$\mathbf{W} \quad = \mathbf{W}_{\text{NNT}} \text{ if AC read or computed} \tag{27}$$

$$c_1 \quad = n \min \text{F} \text{ for ULS, GLS, WLS, DWLS} \tag{28}$$

$$= n(\log\left\|\hat{\mathbf{\Sigma}}\right\| + \text{tr}(\mathbf{S}\,\hat{\mathbf{\Sigma}}^{-1}) - \log\|\mathbf{S}\| - k) \text{ for ML}$$

$$c_{2\text{NT}} \quad = n(\mathbf{s} - \hat{\boldsymbol{\sigma}})'\mathbf{\Delta}_c(\mathbf{\Delta}_c'\mathbf{W}_{\text{NT}}\mathbf{\Delta}_c)^{-1}\mathbf{\Delta}_c'(\mathbf{s} - \hat{\boldsymbol{\sigma}}) \tag{29}$$

$$c_{2\mathrm{NNT}} = n(\mathbf{s}-\hat{\boldsymbol{\sigma}})'\boldsymbol{\Delta}_c(\boldsymbol{\Delta}_c'\mathbf{W}_{NNT}\boldsymbol{\Delta}_c)^{-1}\boldsymbol{\Delta}_c'(\mathbf{s}-\hat{\boldsymbol{\sigma}}) \tag{30}$$

$$h_1 = \mathrm{tr}[(\boldsymbol{\Delta}_c'\mathbf{W}_{\mathrm{NT}}\boldsymbol{\Delta}_c)^{-1}(\boldsymbol{\Delta}_c'\mathbf{W}_{\mathrm{NT}}\boldsymbol{\Delta}_c)] \tag{31}$$

$$c_3 = \frac{d}{h_1}c_2 \tag{32}$$

Comments:

- **E** is the information matrix.
- (24) gives the estimated asymptotic covariance matrix of the parameter estimates. The standard errors of the parameter estimates are obtained from the diagonal elements of this matrix.
- Note that if AC is not read or computed and ML is used then $\mathbf{W} = \mathbf{V}^{-1}$ and (24) reduces to

$$n\mathrm{ACov}(\hat{\boldsymbol{\theta}}) = \mathbf{E}^{-1} \tag{33}$$

- Similarly, if AC is read or computed and WLS is used, then $\mathbf{VWV} = \mathbf{V}$ and then also (24) reduces to (33).
- The standard errors of the residuals are obtained from the diagonal elements of (25) with $\mathbf{W} = \mathbf{W}_{\mathrm{NT}}$ or $\mathbf{W} = \mathbf{W}_{\mathrm{NNT}}$. From these standardized residuals are obtained.
- $c_1, c_{2\mathrm{NT}}, c_{2\mathrm{NNT}}$ and c_3 define four different chi-squares:
  - c_1 is *n* times the minimum value of the fit function.
  - $c_{2\mathrm{NT}}$ is *n* times the minimum value of the WLS fit function using a weight matrix estimated under multivariate normality.
  - $c_{2\mathrm{NNT}}$ is equation (2.20a) in Browne (1984) using the asymptotic covariance matrix provided.
  - c_3 is the Satorra-Bentler scaled chi-square statistic (Satorra & Bentler, 1988, equation 4.1).
  - Under multivariate normality of the observed variables, c_1 and $c_{2\mathrm{NT}}$ are asymptotically equivalent and have an asymptotic chi-square distribution if the model holds exactly and an asymptotic non-central chi-square distribution if the model holds approximately. The same holds for $c_{2\mathrm{NNT}}$ under the more general assumption that the observed variables have a multivariate distribution with finite moments up to order four. c_3 is a correction to $c_{2\mathrm{NT}}$ which makes c_3 have the correct asymptotic mean even under non-normality. This correction is applied to $c_{2\mathrm{NT}}$, but not to c_1.
- Note that the inverse of **W** is not required in any of these formulas, except if WLS is used. This is important because the inverse of **W** is often more unstable than **W** itself.

3. Single group: mean and covariance structures

There are two possible approaches to analyze the covariance matrix and the mean vector jointly. One is described in this section and the other in the next section.

In principle, the approach of the previous section can be applied to any vector of moments. So we just extend the vector **s** with the mean vector $\bar{\mathbf{z}}$ and then use the same definitions as before:

$$s = \frac{1}{2}k(k+1)+k \quad (34)$$

$$t = \text{number of independent parameters} \quad (35)$$

$$\boldsymbol{\beta}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\sigma}(\boldsymbol{\theta}) \\ \boldsymbol{\mu}(\boldsymbol{\theta}) \end{bmatrix} \quad (36)$$

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}'} \quad (37)$$

$$\boldsymbol{\Delta}_c = \text{the orthogonal complement of } \boldsymbol{\Delta} \quad (38)$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{s} \\ \bar{\mathbf{z}} \end{bmatrix} \quad (39)$$

¡ $\boldsymbol{\Omega}$ is now n times the asymptotic covariance matrix of **b** and **W** a consistent estimate of it. We assumed that **s** and $\bar{\mathbf{z}}$ are asymptotically independent so that $\boldsymbol{\Omega}$ and **W** are block diagonal.

Under normal theory NT, **W** is estimated as

$$\mathbf{W}_{\text{NTE}} = \begin{bmatrix} \mathbf{W}_{\text{NT}} & \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}} \end{bmatrix} \quad (40)$$

and under non-normal theory NNT, **W** is estimated as

$$\mathbf{W}_{\text{NNTE}} = \begin{bmatrix} \mathbf{W}_{\text{NNT}} & \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}} \end{bmatrix} \quad (41)$$

where $\mathbf{W}_{\text{NT}}$ and $\mathbf{W}_{\text{NNT}}$ are defined as in the previous section.

3.1 Fit functions

$$\text{F} = (\mathbf{b}-\boldsymbol{\beta})'\mathbf{V}(\mathbf{b}-\boldsymbol{\beta})$$

ULS: $$\mathbf{V} = \begin{bmatrix} \mathbf{I}^* & \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}^{-1} \end{bmatrix} \quad (42)$$

GLS: $$\mathbf{V} = \begin{bmatrix} \mathbf{D}'\left(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1}\right)\mathbf{D} & \\ \mathbf{0} & \mathbf{S}^{-1} \end{bmatrix} \quad (43)$$

$$\text{ML:} \quad \mathbf{V} = \begin{bmatrix} \mathbf{D}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{D} & \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}^{-1} \end{bmatrix} \tag{44}$$

$$\text{WLS:} \quad \mathbf{V} = \begin{bmatrix} \mathbf{W}_{\text{NNT}}^{-} & \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}^{-1} \end{bmatrix} \tag{45}$$

$$\text{DWLS:} \quad \mathbf{V} = \begin{bmatrix} \mathbf{D}_{W}^{-1} & \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}^{-1} \end{bmatrix} \tag{46}$$

The same results as in Section 2 apply but with $\mathbf{W}_{\text{NT}}$ replaced with $\mathbf{W}_{\text{NTE}}$ and $\mathbf{W}_{\text{NNT}}$ with $\mathbf{W}_{\text{NNTE}}$.

4. Single group: augmented moment matrix

In the previous section it was assumed that $\mathbf{s}$ and $\bar{\mathbf{z}}$ are asymptotically independent. Under non-normality this may not hold. A way to avoid this assumption is to use the augmented moment matrix. This is the matrix of sample moments about zero for the vector $\mathbf{z}$ augmented with a variable which is constant equal to 1 for every case. The population augmented moment matrix and the sample augmented moment matrix are defined in equations (49) and (50) that follow. $\mathbf{a}$ is a vector of the non-duplicated elements in $\mathbf{A}$. Because the last element of $\mathbf{a}$ is constant equal to 1, its covariance matrix $\mathbf{W}_a$ is singular. However, the inverse of $\mathbf{W}_a$ is only used with WLS in which case it is replaced by its generalized inverse. Under non-normal theory $\mathbf{W}_a$ is estimated as $\mathbf{W}_{a\text{NNT}}$ whose elements are

$$w_{gh,ij} = n\text{Est}[\text{ACov}(a_{gh}, a_{ij})] = n_{ghij} - a_{gh}a_{ij}, \tag{47}$$

where

$$a_{gh} = (1/N)\sum_{a=1}^{N} z_{ag}z_{ah}$$

$$n_{ghij} = (1/N)\sum_{a=1}^{N} z_{ag}z_{ah}z_{ai}z_{aj} \tag{48}$$

is a fourth-order sample moment about zero.

4.1 Definitions

$$\boldsymbol{\Upsilon} = E\left[\begin{pmatrix}\mathbf{z}\\1\end{pmatrix}\right]\left[\mathbf{z}' \quad 1\right] = \begin{bmatrix}\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}' & \\ \boldsymbol{\mu}' & 1\end{bmatrix} \tag{49}$$

$$\mathbf{A} = \frac{1}{N}\sum_{c=1}^{N}\left[\begin{pmatrix}\mathbf{z}_c\\1\end{pmatrix}\right]\left[\mathbf{z}_c' \quad 1\right] = \begin{bmatrix}\mathbf{S}+\bar{\mathbf{z}}\bar{\mathbf{z}}' & \\ \bar{\mathbf{z}}' & 1\end{bmatrix} \tag{50}$$

$$\mathbf{a} = \mathbf{K}'vec(\mathbf{A}) \quad vec(\mathbf{A}) = \mathbf{D}\mathbf{a} \tag{51}$$

$$\boldsymbol{\alpha} = \mathbf{K}'vec(\boldsymbol{\Upsilon}) \tag{52}$$

$$\mathbf{W}_a = n\mathrm{Est}[n\mathrm{ACov}(\mathbf{a})] \text{ singular} \tag{53}$$

$$\mathbf{W}_{aNT} = 2\mathbf{K}'(\hat{\boldsymbol{\Upsilon}}\otimes\hat{\boldsymbol{\Upsilon}})\mathbf{K} \tag{54}$$

4.2 Fit Functions

$$\mathrm{F} = (\mathbf{a}-\boldsymbol{\alpha})'\,\mathbf{V}(\mathbf{a}-\boldsymbol{\alpha}) \tag{55}$$

with **V** defined in Section 2 but with

A instead of **S**

$\boldsymbol{\Upsilon}$ instead of $\boldsymbol{\Sigma}$

The same results as in Section 2 apply but with $\mathbf{W}_{\mathrm{NT}}$ replaced with $\mathbf{W}_{a\mathrm{NT}}$ and $\mathbf{W}_{\mathrm{NNT}}$ replaced with $\mathbf{W}_{a\mathrm{NNT}}$.

5. Multiple group: covariance structures

To generalize the results in Section 2 to multiple groups we need only consider that the samples from different groups are supposed to be independent. The presentation here follows Satorra (1993).

5.1 Definitions

$$\mathbf{s}' = (\mathbf{s}_1', \mathbf{s}_2', ..., \mathbf{s}_G') \tag{57}$$

$$\boldsymbol{\sigma}' = (\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2', ..., \boldsymbol{\sigma}_G') \tag{58}$$

$$n_g = N_g - 1, \quad n = \sum n_g$$

$$\mathrm{F}(\mathbf{s},\boldsymbol{\sigma}) = \sum \frac{n_g}{n} \mathrm{F}_g(\mathbf{s}_g, \boldsymbol{\sigma}_g) \tag{59}$$

$$\mathbf{W}_g = n_g \mathrm{Est}[\mathrm{ACov}(\mathbf{s}_g)] \tag{60}$$

$$\mathbf{W} = \begin{bmatrix} W_1 & & & \mathbf{0} \\ & W_2 & & \\ & & \ddots & \\ \mathbf{0} & & & W_G \end{bmatrix} \text{ block diagonal} \tag{61}$$

$$\mathbf{V}_g = \frac{n_g}{n} \times \mathbf{V} \text{ defined in (16) to (20)} \tag{62}$$

different for different methods

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & & & \mathbf{0} \\ & \mathbf{V}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{V}_G \end{bmatrix} \text{ block diagonal} \tag{63}$$

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial \boldsymbol{\sigma}_1}{\partial \boldsymbol{\theta}'} \\ \frac{\partial \boldsymbol{\sigma}_2}{\partial \boldsymbol{\theta}'} \\ \vdots \\ \frac{\partial \boldsymbol{\sigma}_G}{\partial \boldsymbol{\theta}'} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \\ \vdots \\ \boldsymbol{\Delta}_G \end{bmatrix} \tag{64}$$

$$\boldsymbol{\Delta}_c = \begin{bmatrix} \boldsymbol{\Delta}_{1c} \\ \boldsymbol{\Delta}_{2c} \\ \vdots \\ \boldsymbol{\Delta}_{Gc} \end{bmatrix} \tag{65}$$

The fit function in (59) is a weighted sum of the fit functions for each group. The asymptotic covariance matrix in (60) can be estimated for each group separately under NT and NNT as in Section 2. The total asymptotic covariance matrix is defined in (61) and the **V**-matrix is defined in (63) where each $\mathbf{V}_g$ is n_g / n times the $\mathbf{V}$ in Section 2. $\boldsymbol{\Delta}$ is defined in (64). $\boldsymbol{\Delta}_c$ is an orthogonal complement to $\boldsymbol{\Delta}$ partitioned as in (65).

5.2 Results

With these definitions, the same results as in Section 2 apply, namely

$$\mathbf{E} = \mathbf{\Delta}'\mathbf{V}\mathbf{\Delta} \tag{66}$$

$$n\text{ACov}(\hat{\boldsymbol{\theta}}) = \mathbf{E}^{-1}\mathbf{\Delta}'\mathbf{V}\mathbf{W}\mathbf{V}\mathbf{\Delta}\mathbf{E}^{-1} \tag{67}$$

$$n\text{ACov}(\mathbf{s}-\hat{\boldsymbol{\sigma}}) = \mathbf{W}-\mathbf{\Delta}\mathbf{E}^{-1}\mathbf{\Delta}' \tag{68}$$

$$c_1 = n\min \mathrm{F} \tag{69}$$

$$c_{2\mathrm{NT}} = n(\mathbf{s}-\hat{\boldsymbol{\sigma}})'\mathbf{\Delta}_c(\mathbf{\Delta}_c'\mathbf{W}_{NT}\mathbf{\Delta}_c)^{-1}\mathbf{\Delta}_c'(\mathbf{s}-\hat{\boldsymbol{\sigma}}) \tag{70}$$

$$c_{2\mathrm{NNT}} = n(\mathbf{s}-\hat{\boldsymbol{\sigma}})'\mathbf{\Delta}_c(\mathbf{\Delta}_c'\mathbf{W}_{\mathrm{NNT}}\mathbf{\Delta}_c)^{-1}\mathbf{\Delta}_c'(\mathbf{s}-\hat{\boldsymbol{\sigma}}) \tag{71}$$

$$h_1 = \mathrm{tr}[(\mathbf{\Delta}_c'\mathbf{W}_{\mathrm{NT}}\mathbf{\Delta}_c)^{-1}(\mathbf{\Delta}_c'\mathbf{W}_{\mathrm{NT}}\mathbf{\Delta}_c)] \tag{72}$$

$$c_3 = \frac{d}{h_1}c_2 \tag{73}$$

but note that

$$\mathbf{E} = \mathbf{\Delta}'\mathbf{V}\mathbf{\Delta} = \sum_g \mathbf{\Delta}_g'\mathbf{V}_g\mathbf{\Delta}_g \tag{74}$$

$$\mathbf{\Delta}'\mathbf{V}\mathbf{W}\mathbf{V}\mathbf{\Delta} = \sum_g \mathbf{\Delta}_g'\mathbf{V}_g\mathbf{W}_g\mathbf{V}_g\mathbf{\Delta}_g \tag{75}$$

$$n\text{ACOV}(\mathbf{s}-\hat{\boldsymbol{\sigma}}) = \sum_g\left(\mathbf{W}_g-\mathbf{\Delta}_g\mathbf{E}^{-1}\mathbf{\Delta}_g'\right) \tag{76}$$

$$\mathbf{\Delta}_c'\mathbf{W}\mathbf{\Delta}_c = \sum_g \mathbf{\Delta}_{gc}'\mathbf{W}_g\mathbf{\Delta}_{gc} \tag{77}$$

$$\mathbf{\Delta}_c'(\mathbf{s}-\hat{\boldsymbol{\sigma}}) = \sum_g \mathbf{\Delta}_{gc}'(\mathbf{s}_g-\hat{\boldsymbol{\sigma}}_g) \tag{78}$$

6. Multiple group: mean and covariance structures

The generalization of the results in Section 3 to multiple groups follows by using a **b** vector instead of **s** in (59) and definitions in analog to Sections 3 and 4.

7. Multiple group: augmented moment matrices

The generalization of the results in Section 4 to multiple groups follows by using an **a** vector instead of **s** in (59) and definitions in analog to Sections 4 and 5.

8. References

Browne, M.W. (1977) Generalized least squares estimators in the analysis of covariance structures. Pp. 205-226 in D.J.Aigner and A.S. Goldberger (Eds): *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Browne, M.W. (1984). Asymptotically Distribution-free Methods in the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Browne, M.W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika*, **74**, 375-384.

Jöreskog, K.G. (1981) Analysis of covariance structures. *Scandinavian Journal of Statistics*, **8**, 65-92.

Magnus, J.R. & Neudecker, H. (1988) *Matrix differential calculus with applications in statistics and econometrics*. New York, Wiley.

Satorra, A. (1987) Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, **54**, 131-151.

Satorra, A. (1993). Multi-sample analysis of moment structures: Asymptotic validity of inferences based on second-order moments. Pp. 283-298 in K. Haagen, D.J. Bartholomew, & M. Deistler (Eds): *Statistical Modeling and Latent variables*. Elsevier Science Publishers.

Satorra, A. & Bentler, P.M. (1988) Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 1988, 308-313.