

## Computational Options for Standard Errors and Test Statistics with Incomplete Normal and Nonnormal Data in SEM

Victoria Savalei & Yves Rosseel

To cite this article: Victoria Savalei & Yves Rosseel (2022) Computational Options for Standard Errors and Test Statistics with Incomplete Normal and Nonnormal Data in SEM, Structural Equation Modeling: A Multidisciplinary Journal, 29:2, 163-181, DOI: [10.1080/10705511.2021.1877548](https://doi.org/10.1080/10705511.2021.1877548)

To link to this article: <https://doi.org/10.1080/10705511.2021.1877548>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 08 Oct 2021.



Submit your article to this journal [↗](#)



Article views: 1090




View related articles [↗](#)



View Crossmark data [↗](#)

# Computational Options for Standard Errors and Test Statistics with Incomplete Normal and Nonnormal Data in SEM

Victoria Savalei<sup>a</sup> and Yves Rosseel<sup>b</sup> 

<sup>a</sup>University of British Columbia; <sup>b</sup>Ghent University

## ABSTRACT

This article provides an overview of different computational options for inference following normal theory maximum likelihood (ML) estimation in structural equation modeling (SEM) with incomplete normal and nonnormal data. Complete data are covered as a special case. These computational options include whether the information matrix is observed or expected, whether the observed information matrix is estimated numerically or using an analytic asymptotic approximation, and whether the information matrix and the outer product matrix of the score vector are evaluated at the saturated or at the structured estimates. A variety of different standard errors and robust test statistics become possible by varying these options. We review the asymptotic properties of these computational variations, and we show how to obtain them using *lavaan* in R. We hope that this article will encourage methodologists to study the impact of the available computational options on the performance of standard errors and test statistics in SEM.

## KEYWORDS

Incomplete data; nonnormal data; robust corrections; software implementation

## Introduction

The focus of this article is on the methods for drawing inferences from linear structural equation models (SEMs) with incomplete normal and nonnormal data. We cover complete data as a special case. We focus on the normal theory maximum likelihood (ML) estimator, which is the default for continuous data in all SEM software, including the *lavaan* package in R (R Core Team, 2020; Rosseel, 2012). With incomplete data, the ML estimator as implemented in SEM software is commonly referred to as the full information maximum likelihood (FIML) estimator (Arbuckle, 1996; Schafer & Graham, 2002). Because the FIML estimator is just the ML estimator, in this article we will refer to it as ML.

For incomplete data, we assume that the missing mechanism is either MCAR (missing completely at random) or MAR (missing at random) (Rubin, 1976). MCAR and MAR missing data mechanisms generally result in ignorable missing data (Little & Rubin, 2020), which means that inference based on the observed data log-likelihood remains valid. When data are normally distributed, the ML estimator is consistent with MCAR and MAR missing data, meaning that the ML parameter estimates approach the true values when the model is correct. When data are nonnormal but complete, the ML estimator is consistent as long as the data distribution has finite fourth order moments (Arminger & Schoenberg, 1989; Shapiro, 1983). When data are nonnormal and MCAR, the ML estimator remains consistent (Yuan & Bentler, 2000). The most complicated case is when data are nonnormal and MAR. Despite some sources claiming consistency in the general case of MAR data (Arminger & Sobel, 1990), counterexamples can be constructed to show this is not always the case (Yuan, 2009; Yuan & Savalei, 2014; Q. Zhang et al., 2019). Yuan (2009) showed that the ML estimator is consistent with nonnormal MAR data for several

common types of MAR mechanisms. Yuan and Bentler (2010) further showed that consistency is retained for all MAR mechanisms as long as the variables with missing values are linearly related to the observed variables in the model. In this article, we will assume that the ML estimator is consistent with MAR data when the model is correctly specified. When referring to the results that hold for MAR data, we assume either the type of missing mechanism or the model are such that the property of consistency is retained.

Once ML parameter estimates have been obtained, ML standard errors for each estimated parameter and the likelihood ratio (LR) chi-square to assess the overall model fit can be computed. In addition to the ML standard errors, so-called robust or sandwich standard errors can also be computed (Arminger & Schoenberg, 1989; Arminger & Sobel, 1990; Satorra & Bentler, 1994; Yuan & Bentler, 2000; Yuan & Hayashi, 2006), which can protect against nonnormality of the data, model misspecification, or both. In this article, we use the terms robust and sandwich interchangeably. In addition to the LR chi-square, two so-called robust test statistics are popular with nonnormal data: a scaled (mean-corrected) chi-square and an adjusted (mean-and-variance corrected) chi-square (Asparouhov & Muthén, 2010; Satorra & Bentler, 1994; Yuan & Bentler, 2000). These robust statistics, which are rescalings of the LR chi-square, can protect against distortion of Type I error rates due to nonnormality, and they have also been recommended for use even with normal data when the sample size is small (Maydeu-Olivares, 2017).

The goal of this article is to describe different computational versions of ML and robust standard errors, and of the scaled and adjusted chi-square statistics, with incomplete normal and nonnormal data in SEM. These computational options include whether the information matrix is observed or expected, whether the observed information matrix is obtained numerically or using

an analytic asymptotic (i.e., increasingly accurate in large samples) approximation, and whether the information matrix and the outer product matrix of the score vector are evaluated at the saturated or at the structured estimates. A variety of different computational versions of standard errors and test statistics can be obtained by varying these options. We list the properties of the resulting standard errors and test statistics under different missing data mechanisms, for normal and nonnormal data, and for correctly specified versus misspecified models.

While ML computations tend to be implemented similarly across software, different software packages have different computational defaults for robust computations, which are often hidden from the user. Further, methodologists who custom-program their own robust computations do not always clearly state which types of estimates were used (e.g., Lai, 2019; Yuan & Bentler, 2000). Yet, these choices may have a substantial effect on the performance of standard errors and test statistics in simulation studies. If the computational choices made in a particular simulation are not reported, knowledge aggregation across studies is difficult. There are very few simulation studies explicitly comparing the impact of the computational options described in this article on performance of standard errors and/or test statistics (Maydeu-Olivares, 2017; Xia et al., 2016), and only one that we know of with incomplete data (Savalei, 2010a). In this article, we illustrate how to obtain all computational options in *lavaan* 0.6–8 (Rosseel, 2012). This software package has the advantage of allowing a greater variety of types of standard errors and robust test statistics relative to any other SEM package; it is also free and open source. In addition, custom implementation of computational options that are not yet directly available in *lavaan* is straight-forward. In this article, we illustrate how to obtain all computational options for standard errors and test statistics using *lavaan*.

This article is organized as follows. We first review ML estimation with incomplete data and provide expressions for the relevant quantities based on the log-likelihood: the first and second derivatives of the log-likelihood, and the outer product matrix of the score vector. Next, we review the asymptotic distribution of the ML estimator under different assumptions and describe different computational options for ML and robust standard errors. We then review the asymptotic distribution of the LR test statistic and describe different computational options for the scaling corrections leading to robust statistics. These two sections also include an overview of the computational options available in *lavaan*, and how to obtain estimates that are not directly available. We conclude with a discussion, which includes an overview of existing limited literature studying the impact of these computational options and a summary of future research directions.

## Derivatives of the log-likelihood

### Overview of ML estimation

Let  $Y$  be a  $p$ -dimensional random variable with a continuous multivariate distribution with mean vector

$\mu^0$ , covariance matrix  $\Sigma^0$ , and finite 4th order moments. Under a general mean-and-covariance structure model,  $\mu^0 = \mu(\theta^0)$  and  $\Sigma^0 = \Sigma(\theta^0)$ , where  $\theta^0$  is a  $q \times 1$  vector of model parameters. Letting  $\beta^0 = ((\text{vech} \Sigma^0)', \mu^{0'})'$ , i.e., a  $(p^* + p) \times 1$  vector that contains the non-redundant elements of  $\Sigma^0$  (Magnus & Neudecker, 1999) and  $\mu^0$ , we can also represent the general mean-and-covariance structure model more succinctly as  $\beta^0 = \beta(\theta^0)$ . We will denote the hypothesized, or *structured*, model by  $H_0$ , and we will denote the saturated, or *unstructured*, model by  $H_1$ . The traditional saturated model of SEM has the parameters  $\theta^0 = \beta^0$ , i.e., it does not impose any restrictions on the means or the covariance matrix.

Let  $y_1, \dots, y_n$  be the observed values for a random sample of size  $n$  from the distribution of  $Y$ . Each  $y_i$  has the dimensions  $p_i \times 1$ , where  $1 \leq p_i \leq p$  for  $i = 1, \dots, n$ . The dimensions of the individual observations vary due to missing data. We estimate the model parameters by maximum likelihood under the assumption of a multivariate normal distribution for  $Y$ . The observed log-likelihood for an observation  $y_i$  is given by

$$l_i(\theta) = -\frac{1}{2}p_i \log(2\pi) - \frac{1}{2} \log |\Sigma_i(\theta)| - \frac{1}{2}(y_i - \mu_i(\theta))' \Sigma_i(\theta)^{-1} (y_i - \mu_i(\theta)), \quad (1)$$

where  $\mu_i$  is a  $p_i \times 1$  sub-vector of  $\mu$ , omitting elements corresponding to variables that are missing from  $y_i$ , and  $\Sigma_i$  is the  $p_i \times p_i$  sub-matrix of  $\Sigma$ , omitting rows and columns corresponding to the variables missing from  $y_i$ . To construct these matrices more formally, let  $\tau_i = \frac{\partial \mu_i}{\partial \mu}$ , a 0–1 matrix of size  $p_i \times p$  (Yuan & Bentler, 2000); then,  $\mu_i = \tau_i \mu$  and  $\Sigma_i = \tau_i \Sigma \tau_i'$ .

The observed log-likelihood for the entire sample is obtained by summing over the individual log-likelihoods:

$$l(\theta) = \sum_{i=1}^n l_i(\theta).$$

We denote the structured model parameter estimates by  $\hat{\theta}$  and the maximized log-likelihood by  $l(\hat{\theta})$ . We will write  $\hat{\mu} = \mu(\hat{\theta})$ ,  $\hat{\Sigma} = \Sigma(\hat{\theta})$ , as well as  $\hat{\beta} = \beta(\hat{\theta})$ . We denote the saturated model parameter estimates by  $\tilde{\beta}$  and the maximized log-likelihood by  $l(\tilde{\beta})$ . We will also write  $\tilde{\mu} = \mu(\tilde{\beta})$  and  $\tilde{\Sigma} = \Sigma(\tilde{\beta})$ . We assume that the missing data mechanism is such that  $\tilde{\beta}$  is consistent for  $\beta^0$ ; it follows that, when  $H_0$  is true,  $\hat{\theta}$  is consistent for  $\theta^0$  (Shapiro, 1985). When  $H_0$  is false, we will denote the limiting value of  $\hat{\theta}$  by  $\theta^*$ , which we will refer to as pseudo-parameters, i.e., the population values that the ML method consistently estimates when the model is wrong (Arminger & Schoenberg, 1989; Arminger & Sobel, 1990; Yuan & Hayashi, 2006).<sup>1</sup> We will write  $\mu^* = \mu(\theta^*)$ ,  $\Sigma^* = \Sigma(\theta^*)$ , and  $\beta^* = \beta(\theta^*)$ . When  $H_0$  is false,  $\beta^0 \neq \beta^*$ .

<sup>1</sup>It should be noted that  $\theta^*$  will be different depending on whether the data are complete or incomplete, and for different types of incomplete data; see X. Zhang and Savalei (2020). Thus, we cannot say that with incomplete data, the ML estimator retains consistency for the complete data pseudo-parameters when the model is wrong.

## The first derivative of the log-likelihood

### Saturated model ( $H_1$ )

We denote the first derivative, or the score vector, of Equation 1 with respect to  $\beta = ((\text{vech}(\Sigma))', \mu')'$  by

$$d'_{\beta,i}(\beta) = \frac{\partial l_i(\beta)}{\partial \beta'}. \quad (2)$$

An explicit expression is given in the Appendix (part B). The first derivative of the log-likelihood for the sample is the sum of the individual derivatives. It is more useful to define the average derivative and its limiting expression as the sample size goes to infinity:

$$d_{\beta,n} = \frac{1}{n} \sum_{i=1}^n d_{\beta,i}$$

$$d_{\beta} = \lim_{n \rightarrow \infty} d_{\beta,n}$$

Note that  $\tilde{d}_{\beta,n} = d_{\beta,n}(\tilde{\beta}) = 0$  by construction, and  $d_{\beta}^0 = d_{\beta}(\beta^0) = 0$ .

### Structured model ( $H_0$ )

By the chain rule, the first derivative of Equation 1 with respect to  $\theta$  is given by

$$d'_{\theta,i}(\theta) = \frac{\partial l_i(\theta)}{\partial \theta'} = \frac{\partial l_i(\beta)}{\partial \beta'} \frac{\partial \beta(\theta)}{\partial \theta'} = d'_{\beta,i} \Delta$$

where  $\Delta = \Delta(\theta) = \frac{\partial \beta(\theta)}{\partial \theta'}$  is a  $(p^* + p) \times q$  matrix of model derivatives. As well,  $d'_{\theta,n} = d'_{\beta,n} \Delta$  and  $d'_{\theta} = d'_{\beta} \Delta$ . Note that  $\hat{d}'_{\theta,n} = \hat{d}'_{\beta,n} \hat{\Delta} = 0$  by construction, where  $\hat{\Delta} = \Delta(\hat{\theta})$ . Under  $H_0$ ,  $d'_{\theta} = d'_{\theta}(\theta^0) = 0$ . However, when  $H_0$  is false, it is also the case that  $d'_{\theta} = d'_{\theta}(\theta^*) = 0$ ; the pseudo-parameters  $\theta^*$  are the values that make the first derivative vanish in the limit.

## The second derivative of the log-likelihood

### Saturated model ( $H_1$ )

The negative of the second derivative (or the Hessian) of the individual log-likelihood in Equation 1 with respect to  $\beta$  is

$$A_{\beta,i}(\beta) = -\frac{\partial^2 l_i(\beta)}{\partial \beta \partial \beta'}.$$

An explicit expression is given in Appendix (part A). The sample average and its limiting expression are as follows:

$$A_{\beta,n} = -\frac{1}{n} \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = \frac{1}{n} \sum_{i=1}^n A_{\beta,i} \quad (3)$$

$$A_{\beta} = \lim_{n \rightarrow \infty} A_{\beta,n} = \frac{1}{J} \sum_{i=1}^J q_j A_{\beta,i}, \quad (4)$$

where  $j = 1, \dots, J$  enumerates the missing data patterns,  $q_j$  is the probability of pattern  $j$  in the population,  $A_j = \lim_{n_j \rightarrow \infty} \frac{1}{n_j} \sum_{i=1}^{n_j} A_{\beta,i}$ , and  $n_j$  is the number of observations in the  $j$ th pattern. The matrices  $A_{\beta,n}$  and  $A_{\beta}$  can be described as the observed and the expected information matrices for the saturated model. In particular,  $A_j$  can be thought of as the conditional expected value of the second derivative given pattern  $j$ .

Under a general MAR mechanism, observations within each pattern are not randomly sampled from the population, and the limiting values of pattern-specific means and covariance matrices are not subcomponents of the overall population vector of means  $\mu^0$  and the population covariance matrix  $\Sigma^0$ . Denote the limiting values of these pattern-specific means and covariance matrices by  $\zeta_j$  and  $\Pi_j$ , respectively,  $j = 1, \dots, J$ . An explicit expression for  $A_j$  is given by:

$$A_j(\beta|\zeta_j, \Pi_j) = \begin{pmatrix} \kappa_j'(\Sigma_j^{-1} \otimes (\Sigma_j^{-1} \Pi_j \Sigma_j^{-1} + h_j h_j' - .5 \Sigma_j^{-1})) \kappa_j \kappa_j' (h_j \otimes \Sigma_j^{-1} \tau_j) \\ (h_j' \otimes \tau_j' \Sigma_j^{-1}) \kappa_j & \tau_j' \Sigma_j^{-1} \tau_j \end{pmatrix} \quad (5)$$

where  $\tau_j$  and  $\kappa_j = (\tau_j \otimes \tau_j) D_p$  add rows and columns of zeros corresponding to missing variables in pattern  $j$ ,  $D_p$  is the duplication matrix of order  $p$  (Magnus & Neudecker, 1999),  $\mu_j = \tau_j \mu$ ,  $\Sigma_j = \tau_j \Sigma \tau_j'$ ,  $h_j = \Sigma_j^{-1}(\zeta_j - \mu_j)$ . The selection matrices  $\tau_j$  and  $\kappa_j$  ensure the dimensions of  $A_j$  are  $p^* \times p^*$  for all  $j$ . In this expression,  $A_j$  is a function of  $\beta$  through  $\Sigma_j$  and  $\mu_j$ .

Because the pattern-specific population means and covariance matrices  $\zeta_j$  and  $\Pi_j$  are not known under a general MAR mechanism, in the sample we can estimate  $A_{j,n} = A_j(\beta|\bar{y}_j, S_j)$  for each  $j = 1, \dots, J$ , where  $\bar{y}_j$  and  $S_j$  are the sample mean and the sample covariance matrix in pattern  $j$ . These lead to the so-called observed information estimates, because  $A_{\beta,n} = \frac{1}{J} \sum_{i=1}^J \frac{n_j}{n} A_{j,n}$ .<sup>2</sup> We will use the following two observed information estimates of  $A_{\beta}$ :

$$\tilde{A}_{\beta} = \tilde{A}_{\beta,n} = \frac{1}{J} \sum_{i=1}^J q_j A_{j,n}(\tilde{\beta}|\bar{y}_j, S_j) \quad (6)$$

$$\hat{A}_{\beta} = \hat{A}_{\beta,n} = \frac{1}{J} \sum_{i=1}^J q_j A_{j,n}(\hat{\beta}|\bar{y}_j, S_j), \quad (7)$$

where the second estimate assumes a structured model has also been fit to data.

In the special case of MCAR data, observations falling within each pattern can be considered a random sample from the population, and the pattern-specific population means and covariances are subsets of the general vector of means and covariance matrix:  $\zeta_j = \mu_j^0$  and  $\Pi_j = \Sigma_j^0$  for  $j = 1, \dots, J$ . This consequence of MCAR is known as the homogeneity of means and covariance matrices (Kim & Bentler, 2002).<sup>3</sup> Note that

<sup>2</sup>To simplify notation, we assume that  $J$  enumerates all possible patterns that exist in the population (i.e., we implicitly set  $n_j = 0$  for those patterns missing from a particular sample).

<sup>3</sup>The converse is not true, however: it is possible to construct examples where homogeneity of means and covariance matrices holds while the mechanism is not MCAR (Yuan et al., 2018).

$$A_j(\beta|\mu_j, \Sigma_j) = \begin{pmatrix} .5\kappa'_j(\Sigma_j^{-1} \otimes \Sigma_j^{-1})\kappa_j & 0 \\ 0 & \tau'_j \Sigma_j^{-1} \tau_j \end{pmatrix} \equiv W_j(\beta), \quad (8)$$

Then, the general form of  $A_\beta$  in Equation 4 under MCAR is given by (Yuan & Bentler, 2000):

$$A_\beta = \frac{1}{J} \sum_{j=1}^J q_j W_j \equiv W, \quad (9)$$

and two additional estimates are possible:

$$\tilde{A}_{\beta,E} = W(\tilde{\beta}) = \tilde{W} \quad (10)$$

$$\hat{A}_{\beta,E} = W(\hat{\beta}) = \hat{W} \quad (11)$$

Because these estimates replace  $\bar{y}_j$  and  $S_j$  with their estimated expected values under a particular model, they are known as expected information estimates. They also necessarily assume the model under which the expectation is taken is true. Because complete data can be viewed as a special case of MCAR data, these estimates are valid for complete data as well. Table 1 provides parallel expressions to Equations 5 and 8 in the special case of complete data. Of note, observed and expected information estimates are actually the same with complete data when evaluated at the saturated model estimates:  $\tilde{A}_{\beta,E} = \tilde{A}_\beta = \tilde{W}_c$ , where  $W_c$  is the complete data simplification of  $W$ .

### Structured model ( $H_0$ )

The second derivative of Equation 1 with respect to  $\theta$  can be obtained from  $A_{\beta,i}$  via the multivariate version of the chain rule for the second derivative (e.g., Magnus & Neudecker, 1999, p. 110, Theorem 9)<sup>4</sup>

$$A_{\theta,i}(\theta) = -\frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta'} = \Delta' A_{\beta,i} \Delta - (d'_{\beta,i} \otimes I_q) H,$$

where  $I_q$  is a  $q \times q$  identity matrix and  $H = \frac{\partial \text{vec}(\Delta')}{\partial \theta'}$  is the  $q(p^* + p) \times q$  matrix of second derivatives of  $\beta(\theta)$ .<sup>5</sup> We define the sample average and its limiting expression as follows:

$$A_{\theta,n} = -\frac{1}{n} \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} = \frac{1}{n} \sum_{i=1}^n A_{\theta,i} \quad (12)$$

$$A_\theta = \lim_{n \rightarrow \infty} A_{\theta,n} = \Delta' A_\beta \Delta - (d'_\beta \otimes I_q) H \quad (13)$$

When  $H_0$  is true,  $d_\beta^0 = 0$ , and therefore

$$A_\theta^0 = \Delta^0 A_\beta^0 \Delta^0. \quad (14)$$

The following two estimates can be constructed based on the estimate  $\hat{A}_\beta$  from Equation 7:

$$\hat{A}_\theta = \hat{\Delta}' \hat{A}_\beta \hat{\Delta} - (\hat{d}'_\beta \otimes I_q) \hat{H} \quad (15)$$

$$\hat{A}_{\theta,h1} = \hat{\Delta}' \hat{A}_\beta \hat{\Delta} \quad (16)$$

where  $\hat{H} = H(\hat{\theta})$ .<sup>6</sup> The estimate in Equation 15 is typically obtained via numeric differentiation of the analytic first derivative of the log-likelihood (i.e.,  $\hat{d}_{\theta,n}$ ) because  $\hat{H}$  is too cumbersome to obtain analytically; for this reason,  $\hat{A}_\theta$  is commonly referred to as the numeric Hessian. The second term in Equation 15 will tend to be small, and Equation 16 constitutes a valid asymptotic approximation when  $H_0$  is true, by Equation 14. An additional estimate can be constructed based on the estimate  $\tilde{A}_\beta$  from Equation 6:

$$\tilde{A}_\theta = \hat{\Delta}' \tilde{A}_\beta \hat{\Delta} - (\tilde{d}'_\beta \otimes I_q) \hat{H} = \hat{\Delta}' \tilde{A}_\beta \hat{\Delta}, \quad (17)$$

where the second term drops out because  $\tilde{d}_\beta = 0$ . All three estimates in Equations 15–17 can be considered observed information estimates, although most commonly this term refers to the estimate in Equation 15.

In the special case of MCAR data, two additional estimates based on expected information are also possible:

$$\hat{A}_{\theta,E} = \hat{\Delta}' \hat{W} \hat{\Delta} \quad (18)$$

$$\tilde{A}_{\theta,E} = \hat{\Delta}' \tilde{W} \hat{\Delta}, \quad (19)$$

based on the Equations 11 and 10. With complete data, equations 17 and 19 yield identical estimates, i.e.,  $\tilde{A}_\theta = \tilde{A}_{\theta,E}$ . Other than this equivalence, complete data versions of these equations do not provide any additional insight and are omitted (see Browne & Arminger, 1995; Lee, 2007; Yuan & Hayashi, 2006).

### Outer product matrix of the score vector

#### Saturated model ( $H_1$ )

The outer product of the score vector in Equation 2 with itself will be denoted by

$$B_{\beta,i}(\beta) = \frac{\partial l_i(\beta)}{\partial \beta} \frac{\partial l_i(\beta)}{\partial \beta'} = d_{\beta,i} d'_{\beta,i}.$$

The average outer product and its limiting expression are given by:

$$B_{\beta,n} = \frac{1}{n} \sum_{i=1}^n B_{\beta,i} = \frac{1}{J} \sum_{j=1}^J \frac{n_j}{n} W_j \Gamma_{j,n} W_j \quad (20)$$

<sup>4</sup>Yuan and Bentler (2000) and Savalei (2010a) both inaccurately stated a simplified version of this equation that omits the second term, but this simplification only holds asymptotically and when  $H_0$  is true; see Equation 13.

<sup>5</sup>The matrix  $H = H(\theta)$  is formed by stacking vertically  $(p^* + p)$  number of  $q \times q$  matrices of the form:  $H_r = \frac{\partial^2 B_r(\theta)}{\partial \theta \partial \theta'} = \frac{\partial d_r}{\partial \theta'}$ , where  $d_r$  is the transpose of the  $r$ th row of  $\Delta$ ,  $r = 1, \dots, (p^* + p)$ . The following identity also holds:  $(d'_{\beta,i} \otimes I_q) H = \sum_{r=1}^{p^*+p} \frac{\partial l_i(\beta)}{\partial \beta'} H_r$ .

<sup>6</sup>The subscript “h1” in Equation 16 corresponds to the syntax used to call this estimate in *lavaan*. It is meant to capture that the difference between the two structured estimates of observed information,  $\hat{A}_\theta$  and  $\hat{A}_{\theta,h1}$ , is that the latter only uses the first term based on the information matrix for the  $H_1$  model.



**Table 1.** Complete data expressions for  $A_\beta$  and  $B_\beta$ .

Complete Data Expressions
$A_\beta = A(\beta \mu^0, \Sigma^0) = \begin{pmatrix} D'_p(\Sigma^{-1} \otimes (\Sigma^{-1}\Sigma^0\Sigma^{-1} + hh' - .5\Sigma^{-1}))D_p & D'_p(h \otimes \Sigma^{-1}) \\ (h' \otimes \Sigma^{-1})D_p & \Sigma^{-1} \end{pmatrix},$ <p style="text-align: center;">where <math>h = \Sigma^{-1}(\mu^0 - \mu)</math></p> $A_{\beta,n} = A(\beta \bar{y}, S), \text{ where } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, S = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})'$ $A_\beta^0 = A(\beta^0 \mu^0, \Sigma^0) = W_c^0, \text{ where } W_c = \begin{pmatrix} .5D'_p(\Sigma^{-1} \otimes \Sigma^{-1})D_p & 0 \\ 0 & \Sigma^{-1} \end{pmatrix}$
$B_{\beta,n} = W_c \Gamma_n W_c, \text{ where } \Gamma_n = \frac{1}{n} \sum_{i=1}^n t_i t_i', t_i = ((\text{vech}(y_i - \mu)(y_i - \mu)' - \text{vech}(\Sigma))', (y_i - \mu)')'$ $B_\beta = W_c \Gamma W_c, \text{ where } \Gamma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n t_i t_i'$

$$B_\beta = \lim_{n \rightarrow \infty} B_{\beta,n} = \frac{1}{J} \sum_{j=1}^J q_j W_j \Gamma_j W_j \quad (21)$$

where  $W_j = W_j(\beta)$  is given in Equation 8,  $\Gamma_{j,n} = \Gamma_{j,n}(\beta)$  depends on third- and fourth-order sample moments of the data within each pattern (see Appendix, part B), and  $\Gamma_j = \Gamma_j(\beta)$  is the population limit of  $\Gamma_j$ ,  $j = 1, \dots, J$ .

Sample estimates are as follows:

$$\hat{B}_\beta = \hat{B}_{\beta,n} = B_{\beta,n}(\hat{\beta}) \quad (22)$$

$$\tilde{B}_\beta = \tilde{B}_{\beta,n} = B_{\beta,n}(\tilde{\beta}). \quad (23)$$

In practice, these estimates are obtained by computing the average outer product of case-wise derivatives  $\hat{d}_{\beta,i}$  and  $\tilde{d}_{\beta,i}$ , respectively, rather than using estimates of  $W_j$  and  $\Gamma_{j,n}$ .

In the special case of complete data, Equation 21 simplifies to  $B_\beta = W_c \Gamma W_c$ , where  $W_c$  and  $\Gamma$  are defined in Table 1. The following estimate of  $\Gamma^0 = \Gamma(\beta^0)$  is famous in the SEM literature and is known as the asymptotic covariance matrix of sample covariances and sample means (Browne, 1984):

$$\tilde{\Gamma} = \frac{1}{n} \sum_{i=1}^n \tilde{t}_i \tilde{t}_i', \quad (24)$$

where  $\tilde{t}_i = ((\text{vech}(y_i - \bar{y})(y_i - \bar{y})' - \text{vech}(S))', (y_i - \bar{y})')'$ .

### Structured model ( $H_0$ )

By the chain rule,

$$B_{\theta,i}(\theta) = \frac{\partial l_i(\theta)}{\partial \theta} \frac{\partial l_i(\theta)}{\partial \theta'} = d_{\theta,i} d'_{\theta,i} = \Delta' B_{\beta,i} \Delta$$

$$B_{\theta,n} = \frac{1}{n} \sum_{i=1}^n \frac{\partial l_i \theta}{\partial \theta} \frac{\partial l_i \theta}{\partial \theta'} = \Delta' B_{\beta,n} \Delta$$

$$B_\theta = \lim_{n \rightarrow \infty} B_{\theta,n} = \Delta' B_\beta \Delta$$

The following estimates are based on estimates in Equations 22 and 23, respectively:

$$\hat{B}_\theta = \hat{\Delta}' \hat{B}_\beta \hat{\Delta} \quad (25)$$

$$\tilde{B}_\theta = \hat{\Delta}' \tilde{B}_\beta \hat{\Delta} \quad (26)$$

Note that Equation 26 does use the  $\hat{\theta}$  to estimate  $\hat{\Delta}$ , so the estimate  $\tilde{B}_\theta$  is a kind of mixed estimate.

In the special case of complete data,

$$B_\theta = \Delta' W_c \Gamma W_c \Delta, \quad (27)$$

where  $\Gamma$  is defined in Table 1. Based on this expression, several additional estimates of  $B_\theta$  that mix structured and saturated estimates can be constructed. Two such estimates are as follows:

$$\tilde{\tilde{B}}_\theta = \hat{\Delta}' \hat{W}_c \tilde{\Gamma} \hat{W}_c \hat{\Delta}, \text{ and } \hat{\hat{B}}_\theta = \hat{\Delta}' \hat{A}_\beta \tilde{\Gamma} \hat{A}_\beta \hat{\Delta}, \quad (28)$$

where  $\tilde{\Gamma}$  is computed using Equation 24. The estimate  $\hat{\hat{B}}_\theta$  makes use of the fact that with complete data  $W_c^0 = A_\beta^0$ , and observed information estimate  $\hat{A}_\beta$  is a consistent estimate of  $A_\beta^0$ . When the mean structure is saturated (i.e.,  $\hat{\mu} = \bar{y}$ ), it can be shown that  $\tilde{\tilde{B}}_\theta = \hat{\hat{B}}_\theta$  (see Appendix, part D, also Yuan & Hayashi, 2006).

### Standard errors for $\hat{\theta}$

#### Asymptotic distributions of estimators

##### Saturated model ( $H_1$ )

Under the assumption of multivariate normality of the data, and assuming the missing mechanism is such that  $\beta$  is consistent for  $\beta^0$ , the asymptotic distribution of  $\tilde{\beta}$  is given by (e.g., Little & Rubin, 2020):

$$\sqrt{n}(\tilde{\beta} - \beta^0) \rightarrow N(0, (A_\beta^0)^{-1}), \quad (29)$$

where  $A_\beta^0 = A_\beta(\beta^0)$  is defined in Equation 4. With MCAR data,  $A_\beta^0 = W^0 = W(\beta^0)$  where  $W$  is defined in Equation 9. With complete data,  $\tilde{\beta} = ((\text{vech}(S))', \bar{y}')'$  and  $A_\beta^0 = W_c^0$ , where  $W_c$  is defined in Table 1.

When multivariate normality cannot be assumed, the asymptotic distribution of  $\tilde{\beta}$  is given by (Arminger & Schoenberg, 1989; Arminger & Sobel, 1990; White, 1982; Yuan, 2009; Yuan & Bentler, 2000):

$$\sqrt{n}(\tilde{\beta} - \beta^0) \rightarrow N(0, \Omega^0) = N(0, (A_\beta^0)^{-1} B_\beta^0 (A_\beta^0)^{-1}), \quad (30)$$

where  $B_\beta^0 = B_\beta(\beta^0)$  is defined in Equation 21. The covariance matrix in Equation 30 is a type of sandwich covariance matrix, where the bread is the inverse of the information matrix and the meat is the outer product matrix of the score vector. It follows from Equations 29 and 30 that when data are multivariate normal,  $B_\beta^0 = A_\beta^0$ . For MCAR data, this result can be shown algebraically (see Appendix, part C). With complete data,  $\Omega^0 = (W_c^0)^{-1} B_\beta^0 (W_c^0)^{-1} = \Gamma^0$ , where  $\Gamma$  is defined in Table 1.

##### Structured model ( $H_0$ )

When  $H_0$  is true and data are multivariate normal, the asymptotic distribution of  $\hat{\theta}$  is given by:

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, (A_\theta^0)^{-1}), \quad (31)$$

where  $A_\theta^0$  is defined in Equation 14. It is also the case that  $A_\theta^0 = B_\theta^0$ . When  $H_0$  is true but normality cannot be assumed, the asymptotic distribution of  $\hat{\theta}$  is given by (Yuan & Bentler, 2000):

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N\left(0, (A_\theta^0)^{-1} B_\theta^0 (A_\theta^0)^{-1}\right). \quad (32)$$

When  $H_0$  is false, the asymptotic distribution of  $\hat{\theta}$  is given by:

$$\sqrt{n}(\hat{\theta} - \theta^*) \rightarrow N(0, (A_\theta^*)^{-1} B_\theta^* (A_\theta^*)^{-1}) \quad (33)$$

where  $A_\theta^* = A_\theta(\theta^*)$  is defined in Equation 13 and  $B_\theta^* = B_\theta(\theta^*)$  is defined in Equation 21. No simplification of the sandwich covariance matrix in Equation 33 occurs even for normal data because  $A_\theta^* \neq B_\theta^*$ , in general.

The expressions for the asymptotic covariance matrix of  $\hat{\theta}$  are summarized in Table 2, which also includes the special case of complete data. Of particular note, when data are complete and  $H_0$  is true, the asymptotic covariance matrix in Equation 32 is commonly written as (Satorra & Bentler, 1994):

$$(\Delta^{0'} W_c^0 \Delta^0)^{-1} \Delta^{0'} W_c^0 \Gamma^0 W_c^0 \Delta^0 (\Delta^{0'} W_c^0 \Delta^0)^{-1}, \quad (34)$$

which follows by Equations 9, 14, and 27, and the famous Satorra-Bentler standard errors are estimated from this equation.<sup>7</sup>

### Estimates of standard errors

Table 3 provides a summary of sample estimates of the asymptotic covariance matrix of  $\sqrt{n}\hat{\theta}$  in Table 2 that remain consistent under various conditions. Standard errors for  $\hat{\theta}$  are obtained by taking the square-root of the diagonal elements of the estimated asymptotic covariance matrix and dividing by the square-root of the sample size.

When the data are normal and  $H_0$  is true (first column of Table 2), there are five different types of standard errors that remain consistent with MAR data, three based on estimates of the information matrix  $A_\theta^0$  (Equations 15–17) and two based on estimates of the outer product matrix  $B_\theta^0$  (Equations 25–26). Estimates based on the outer product matrix are rarely used in practice. With incomplete data, most software packages implement standard errors based on the numeric Hessian by default, which is equivalent to  $\hat{A}_\theta$ . However, the approximation  $\hat{A}_{\theta,h1}$  is a viable analytic alternative that is also based on structured estimates. The saturated estimate  $\hat{A}_\theta$  is also an interesting alternative. It should be noted that while  $\hat{A}_\theta$  is always consistent for  $A_\theta^0$ , it is not

**Table 2.** Asymptotic covariance matrix of  $\sqrt{n}\hat{\theta}$ .

	Normal Data	Nonnormal Data
$H_0$ is true	$(A_\theta^0)^{-1} = (\Delta^{0'} A_\theta^0 \Delta^0)^{-1}$ Note: $A_\theta^0 = B_\theta^0$	$(A_\theta^0)^{-1} B_\theta^0 (A_\theta^0)^{-1} =$ $(\Delta^{0'} A_\theta^0 \Delta^0)^{-1} \Delta^{0'} B_\theta^0 \Delta^0 (\Delta^{0'} A_\theta^0 \Delta^0)^{-1}$
$H_0$ is false	$(A_\theta^*)^{-1} B_\theta^* (A_\theta^*)^{-1} = (A_\theta^*)^{-1} \Delta^{*'} B_\theta^* \Delta^* (A_\theta^*)^{-1}$	
Complete Data Special Case:		
$H_0$ is true	$(\Delta^{0'} W_c^0 \Delta^0)^{-1}$	$(\Delta^{0'} W_c^0 \Delta^0)^{-1} \Delta^{0'} W_c^0 \Gamma^0 W_c^0 \Delta^0 (\Delta^{0'} W_c^0 \Delta^0)^{-1}$
$H_0$ is false	$(A_\theta^*)^{-1} \Delta^{*'} W_c^* \Gamma^* W_c^* \Delta^* (A_\theta^*)^{-1}$	

**Table 3.** Some consistent estimates of the asymptotic covariance matrix of  $\sqrt{n}\hat{\theta}$ .

	Normal Data	Nonnormal Data
$H_0$ is true; MAR, MCAR, or complete data	$\hat{A}_\theta^{-1}$ (see Eq. 15) $\hat{A}_{\theta,h1}^{-1} = (\hat{\Delta}' \hat{A}_\theta \hat{\Delta})^{-1}$ $\tilde{A}_\theta^{-1} = (\hat{\Delta}' \tilde{A}_\theta \hat{\Delta})^{-1}$ $\hat{B}_\theta^{-1} = (\hat{\Delta}' \hat{B}_\theta \hat{\Delta})^{-1}$ $\tilde{B}_\theta^{-1} = (\hat{\Delta}' \tilde{B}_\theta \hat{\Delta})^{-1}$	$A^{-1} B A^{-1}$ , where $A \in \{\hat{A}_\theta, \hat{A}_{\theta,h1}, \tilde{A}_\theta\}$ and $B \in \{\hat{B}_\theta, \tilde{B}_\theta\}$
$H_0$ is true; MCAR or complete data only	$\hat{A}_{\theta,E}^{-1} = (\hat{\Delta}' \hat{W} \hat{\Delta})^{-1}$ $\tilde{A}_{\theta,E}^{-1} = (\hat{\Delta}' \tilde{W} \hat{\Delta})^{-1}$	$A^{-1} B A^{-1}$ , where $A \in \{\hat{A}_{\theta,E}, \tilde{A}_{\theta,E}\}$ and $B \in \{\hat{B}_{\theta,E}, \tilde{B}_{\theta,E}\}$
$H_0$ is true; complete data only	$\tilde{B}_\theta^{-1} = (\hat{\Delta}' \tilde{W}_c \tilde{W}_c \hat{\Delta})^{-1}$ $\tilde{B}_{\theta}^{-1} = (\hat{\Delta}' \hat{A}_\theta \tilde{A}_\theta \hat{\Delta})^{-1}$	$A^{-1} B A^{-1}$ , where $A \in \{\hat{A}_\theta, \hat{A}_{\theta,h1}, \tilde{A}_\theta, \hat{A}_{\theta,E}\}$ and $B \in \{\tilde{B}_\theta, \tilde{B}_{\theta}\}$ Note: $\tilde{A}_{\theta,E} = \tilde{A}_\theta$ with complete data
$H_0$ may be false	$\hat{A}_\theta^{-1} \hat{B}_\theta \hat{A}_\theta^{-1}$	

Note. Other mixed estimates are possible when  $H_0$  is true and data are complete, such as  $(\hat{\Delta}' \tilde{W}_c \tilde{W}_c \hat{\Delta})^{-1}$  ( $\hat{A}_\theta = \tilde{W}_c$  with complete data) and  $(\hat{\Delta}' \hat{A}_\theta \tilde{A}_\theta \hat{\Delta})^{-1}$ . Because such mixed estimates are somewhat odd, the table highlights only those that are used in software. With complete data and when the mean structure is saturated,  $\tilde{B}_\theta = \hat{B}_\theta$  (see part C of the Appendix), and thus the combination  $\hat{A}_\theta^{-1} \tilde{B}_\theta \hat{A}_\theta^{-1}$  would also provide consistent estimates regardless of whether  $H_0$  is true or false in this case.

actually consistent for the asymptotic covariance matrix of  $\hat{\theta}$  when  $H_0$  is false (given by Equation 33). Thus, simply using observed information does not protect against model misspecification, and there is no theoretical advantage of using  $\hat{A}_\theta$  over  $\hat{A}_{\theta,h1}$  or  $\tilde{A}_\theta$ . Table 3 also lists two additional types of standard errors that are consistent for MCAR or complete data only (Equations 18–19) as they are based on expected information. It also lists two mixed estimates (Equation 28), which are only used with complete data. When data are complete, most software packages default to standard errors based on a structured expected information estimate, i.e., those based on  $\hat{A}_{\theta,E}$  in Equation 18.

When the data are not normal but  $H_0$  is true, many different sandwich-type standard error estimates can be obtained by crossing different computational versions of the “bread” estimate with those of the “meat” estimate (see the second column of Table 3). The properties of the resulting standard errors are inherited from these components; for example, if a component of the sandwich assumes MCAR data, then the resulting standard errors are only consistent under MCAR data. The most likely combinations to be used in practice are those that use the same types of estimates (e.g., both structured or both saturated), though there is no theoretical reason why mixed estimates should not be studied or used. The combination  $\hat{A}_{\theta,E} \tilde{B}_\theta \hat{A}_{\theta,E}$  yields the classic Satorra-Bentler standard errors based on estimating Equation 34. These standard errors are based on  $\tilde{\Gamma}$  (via  $\tilde{B}_\theta$ , see Equation 28) and the estimate  $\hat{W}_c$ .

When  $H_0$  is false, only one sandwich combination provides consistent estimates of the variability of  $\hat{\theta}$  (see the last row of Table 3), regardless of whether the data are normal. This is because only  $\hat{A}_\theta$  is consistent for  $A_\theta^*$  and only  $\hat{B}_\theta$  is consistent

<sup>7</sup>The covariance matrix in Equation 32 can also be expressed as  $(\Delta^{0'} A_\theta^0 \Delta^0)^{-1} \Delta^{0'} A_\theta^0 \Omega^0 A_\theta^0 \Delta^0 (\Delta^{0'} A_\theta^0 \Delta^0)^{-1}$ , which parallels Equation 34. However, this expression is not used with incomplete data because there is no direct way to estimate  $\Omega^0$ . With complete data,  $\Omega^0 = \Gamma^0$ , and an estimate of  $\Gamma^0$  is available (see Equation 24).

for  $B_{\theta}^*$  in general. The resulting standard errors are often called Huber-White standard errors (Eicker, 1967; Huber, 1967; White, 1980). In the SEM literature, the default meaning of the term “sandwich standard errors” usually refers to these estimates. When the mean structure is saturated and the data are complete,  $\tilde{B}_{\theta} = \hat{B}_{\theta}$  (see Appendix, part D), and using  $\tilde{B}_{\theta}$  as the meat would produce standard errors identical to Huber-White standard errors.

### Lavaan options for standard errors

In *lavaan* 0.6–8, the default estimator is **estimator = “ML”**, and the default treatment of missing data is listwise deletion. To appropriately treat missing data while performing ML estimation (i.e., to do FIML estimation), the user must add **missing = “ML”** to the syntax.<sup>8</sup> The information options in *lavaan* are toggled via the following options: **information**, **observed.information**, and **h1.information**. Table 4 provides *lavaan* syntax for obtaining most of the standard error estimates in the first column of Table 3. These standard error estimates are available via **se = “standard”** (default with **estimator = “ML”**). To obtain estimates based on expected information, we set **information = “expected”**, and further specify whether the expected information matrix should be based on structured (**h1.information = “structured”**) or saturated estimates (**h1.information = “unstructured”**). To obtain estimates based on observed information, we set **information = “observed”**, which defaults to a secondary setting of **observed.information = “hessian”** to produce standard errors based on the numeric Hessian.<sup>9</sup> We can change this setting to **observed.information = “h1”** to obtain the analytic estimates of observed information based on either structured or saturated estimates. Finally, we can set **information = “first.order”** to obtain standard error estimates based on the outer product matrix, which can also be structured (**h1.information = “structured”**) or unstructured (**h1.information = “unstructured”**).

Table 5 provides syntax for obtaining sandwich standard errors, which produce most of the options listed in the second column of Table 3, as well as the Huber-White standard errors shown in the last row of Table 3. For historic reasons, there are two types of standard error commands available in *lavaan* to obtain sandwich standard errors. The first is **se = “robust.huber.white”**<sup>10</sup> (default with **estimator = “MLR”**, discussed shortly), which can be used with or without **missing = “ML”**. The second is **se = “robust.sem”** (default with **estimator = “MLM”**), which can only be used with complete data. The default computation of sandwich standard errors when **se = “robust.huber.white”** is to produce the Huber-White standard errors, i.e., shown in the last row of Table 3. As Table 5 shows, most other combinations

**Table 4.** Available options for ML standard errors in *lavaan*.

<i>lavaan</i> options with <b>se = “standard”</b>	Asymptotic Covariance Matrix Estimate
information = “expected” , h1.information = “structured”	$\hat{A}_{\theta,E}^{-1}$ (default without <b>missing = “ML”</b> )
information = “expected” , h1.information = “unstructured”	$\hat{A}_{\theta,E}^{-1}$ (equal to $\hat{A}_{\theta,E}^{-1}$ with complete data)
information = “observed” , observed.information = “hessian” , h1.information = “structured”	$\hat{A}_{\theta}^{-1}$ (default with <b>missing = “ML”</b> )
information = “observed” , observed.information = “h1” , h1.information = “structured”	$\hat{A}_{\theta,h1}^{-1}$
information = “observed” , observed.information = “h1” , h1.information = “unstructured”	$\tilde{A}_{\theta}^{-1}$ (equal to $\tilde{A}_{\theta,E}^{-1}$ with complete data)
information = “first.order” , h1.information = “structured”	$\hat{B}_{\theta}^{-1}$
information = “first.order” , h1.information = “unstructured”	$\tilde{B}_{\theta}^{-1}$

Current *lavaan* defaults with **missing = “ML”** are in italics. All estimates are consistent only when  $H_0$  is true. Options based on  $\tilde{A}_{\theta}^{-1}$  and  $\tilde{B}_{\theta}$  are unavailable.

in the second column of Table 3 can be obtained by toggling the same *lavaan* options as with normal data, plus an additional option, **h1.information.meat**, which permits mixed estimates by allowing the meat of the sandwich (i.e., the outer product matrix) to use different types of estimates (structured vs. unstructured) than what is specified for the bread on the sandwich in **h1.information**.

When **se = “robust.sem”**, the standard errors are computed by estimating Equation 34. By default, this command produces classic Satorra-Bentler standard errors, which use  $\tilde{\Gamma}$  and  $\hat{W}_c$ . Toggling options does not impact the estimate of  $\Gamma^0$ , which is always  $\tilde{\Gamma}$ , but only affects how  $W_c^0$  is estimated, leading to expressions that can be equivalently written in sandwich form with the mixed estimates  $\tilde{B}_{\theta}$  or  $\hat{B}_{\theta}$  (see Equation 28 or Table 3) as the meat of the sandwich. Only some combinations of these mixed meat estimates with various information matrix (bread) estimates are possible in *lavaan*, because *lavaan* generally uses the same estimate of  $W_c$  in all parts of Equation 34.<sup>11</sup>

As Table 5 shows, **se = “robust.huber.white”** and **se = “robust.sem”** will produce identical standard error estimates with complete data when **information = “expected”**, **h1.information = “unstructured”**. In this case, using saturated estimates  $\tilde{W}_c$  and  $\tilde{\Gamma}$  in Equation 34 is equivalent to estimating the sandwich using saturated estimates of both the bread and the meat. In addition, because  $\tilde{B}_{\theta} = \hat{B}_{\theta}$  when the mean structure is saturated, **se = “robust.huber.white”** and **se = “robust.sem”** will produce identical

<sup>8</sup>Aliases include **missing = “FIML”** and **missing = “direct”**.

<sup>9</sup>When **observed.information = “hessian”**, *lavaan* currently does not respond to the **h1.information** option if it is also included; the numerically obtained Hessian is always structured.

<sup>10</sup>**se = “sandwich”** is an alias.

<sup>11</sup>An exception is when **information = “observed”**, **observed.information = “hessian”**. In this case, *lavaan* uses the numeric Hessian to estimate  $\Delta^0 W^0 \Delta^0$  in Equation 34. This is *lavaan*’s best interpretation of what the user wants when **observed.information = “hessian”** is used alongside **se = “robust.sem”**.



**Table 5.** Available “bread” and “meat” options for sandwich-type standard errors in *lavaan*.

$\hat{B}_\theta$	$\hat{A}_{\theta E}^{-1}$	$\hat{A}_{\theta E}^{-1}$	$\hat{A}_\theta^{-1}$	$\hat{A}_\theta^{-1}$	$\hat{A}_{\theta H1}^{-1}$
se = “robust.huber.white”, information = “expected”, h1.information = “structured”	se = “robust.huber.white”, information = “expected”, h1.information = “unstructured”, h1.information.meat = “structured”	se = “robust.huber.white”, information = “observed”, observed.information = “h1”, h1.information = “unstructured”, h1.information.meat = “structured”	se = “robust.huber.white”, information = “observed”, observed.information = “hessian”, h1.information = “structured”	se = “robust.huber.white”, information = “observed”, observed.information = “h1”, h1.information = “structured”	se = “robust.huber.white”, information = “observed”, observed.information = “h1”, h1.information = “structured”
$\tilde{B}_\theta$	se = “robust.huber.white”, information = “expected”, h1.information.meat = “unstructured” or se = “robust.sem”, information = “expected”, h1.information = “unstructured”	se = “robust.huber.white”, information = “observed”, observed.information = “h1”, h1.information = “unstructured”	se = “robust.huber.white”, information = “observed”, observed.information = “hessian”, h1.information = “structured”, h1.information.meat = “unstructured”	se = “robust.huber.white”, information = “observed”, observed.information = “h1”, h1.information = “structured”, h1.information.meat = “unstructured”	
$\bar{B}_\theta$	se = “robust.sem”, information = “expected”, h1.information = “structured”	Not available	Complete data only:	Not available	Not available
$\hat{B}_\theta$	Not available	Not available	se = “robust.sem”, information = “observed”, observed.information = “hessian”, h1.information = “structured”	se = “robust.sem”, information = “observed”, observed.information = “hessian”, h1.information = “structured”	se = “robust.sem”, information = “observed”, observed.information = “h1”, h1.information = “structured”

Sandwich-type standard errors are obtained from the diagonals of the asymptotic covariance matrix of the form  $A^{-1}BA^{-1}$ , where the options for  $A^{-1}$  are given in columns and options for  $B$  are given in rows. Current *lavaan* defaults are in *italics*. With complete data, the second and third columns will produce identical estimates ( $\hat{A}_{\theta E} = \hat{A}_\theta$ ). Specifying **information = “observed”**, **observed.information = “hessian”**, **h1.information = “unstructured”** is another way to obtain  $\hat{A}_\theta^{-1}\hat{B}_\theta\hat{A}_\theta^{-1}$ .

standard error estimates when `information = "expected"`, `h1.information = "structured"` but only for models with no mean structure (see Appendix, part D).

Rather than manually specifying desired standard error options to accompany ML estimates, most end users opt for *lavaan*'s prepackaged options to address nonnormality or model misspecification, available via `estimator = "MLM"` (complete data only) and `estimator = "MLR"` (complete or incomplete data). Table 10 summarizes the defaults for these choices. Because these names are also used in the popular commercial package *Mplus* (Muthén & Muthén, 2012), but with different defaults, the options for `estimator = "MLM"` to match *Mplus* are also listed in the table. When `estimator = "MLR"`, *lavaan* and *Mplus* use the same defaults. We note that the specification of these prepackaged options under `estimator` is somewhat misleading, because the estimator remains the same (ML) but rather the specifications MLM and MLR trigger the computations of robust standard errors and test statistics. The default computations under MLM and MLR can be modified in the same manner (i.e., using `information`, `observed.information`, and `h1.information` options).

## Scaling corrections to the test statistic

### Asymptotic distribution of the test statistics

The likelihood ratio (LR) test statistic  $T$  and the two most popular robust test statistics, the scaled (or mean-corrected) chi-square  $T_M$  and the adjusted (or mean-and-variance corrected) chi-square  $T_{MV}$ , are given by:

$$T = 2(l(\tilde{\beta}) - l(\hat{\theta})) \quad (35)$$

$$T_M = \frac{1}{c} T; T_{MV} = \frac{1}{a} T + b, \quad (36)$$

where  $b = d(1 - c/a)$  and  $d = p^* + p - q$  is the degrees of freedom for the model. In these equations,  $c$  is an estimate of

$$c_0 = \text{Tr}(U^0 \Omega^0) / d, \quad (37)$$

$a$  is an estimate of

$$a_0 = \sqrt{\text{Tr}((U^0 \Omega^0)^2) / d}, \quad (38)$$

where  $\Omega^0$  is defined in Equation 30 and

$$U^0 = A_\beta^0 - A_\beta^0 \Delta^0 (\Delta'^0 A_\beta^0 \Delta^0)^{-1} \Delta'^0 A_\beta^0. \quad (39)$$

The  $(p^* + p) \times (p^* + p)$  matrix  $U^0$  is called the residual weight matrix, and it has reduced rank  $d$ .<sup>12</sup> The constants  $c_0$  and  $a_0$  are commonly referred to as scaling factors or scaling corrections.

The statistic  $T_M$  is the most popular type of robust test statistic for nonnormal data, also known as the Satorra-Bentler chi-square, or the Yuan-Bentler chi-square when extended to incomplete data (Satorra & Bentler, 1994; Yuan

& Bentler, 2000). The version of  $T_{MV}$  in Equation 36 was proposed by Asparouhov and Muthén (2010); it is an improvement on the older version of the mean-and-variance corrected chi-square that required an adjustment to the degrees of freedom (Satorra & Bentler, 1994).<sup>13</sup>

Under multivariate normality and when  $H_0$  is true,  $T$  is asymptotically distributed as a central chi-square with  $d$  degrees of freedom, and thus its approximate expected value is  $\mathbb{E}(T) \approx d$ . However, if the data are nonnormal,  $T$  is asymptotically distributed as a mixture of  $d$  independent chi-square variates, each with 1 degree of freedom; the weights of the mixture are given by the eigenvalues of the matrix product  $U^0 \Omega^0$ . The mean-corrected chi-square  $T_M$ , which corrects by the estimated average eigenvalue of  $U^0 \Omega^0$ , is such that  $\mathbb{E}(T_M) \approx d$ . The mean-and-variance adjusted chi-square  $T_{MV}$  is such that  $\mathbb{E}(T_{MV}) \approx d$  and  $\text{var}(T_{MV}) \approx 2d$ . A chi-square reference distribution with  $d$  degrees of freedom is used for both robust statistics, although neither is asymptotically chi-square distributed, in general. When the data are normal and  $H_0$  is true,  $a_0 = c_0 = 1$ , and  $b_0 = 0$ , so that asymptotically the scaling corrections do nothing.

Different computational versions of the robust test statistics can be obtained by using different estimates of  $a_0$  and  $c_0$ . Table 6 provides several estimates of  $U^0$  and  $\Omega^0$  that can be used to compute the scaling corrections. As described in the note to the table, some estimates of  $U^0$  will always have the property of reduced rank (equal to  $d$ ) by construction, whereas others will only have this property asymptotically (when  $H_0$  is true). Thus, not all versions of the estimate of  $c_0$  can be interpreted as correcting  $T$  by the average eigenvalue of the estimate of  $U^0 \Omega^0$ . It is unknown whether having this property in finite samples affects the performance of the robust statistics.

While different computational versions of  $U^0$  and  $\Omega^0$  in Table 6 are asymptotically equivalent when  $H_0$  is true, they will differ in finite samples, and the resulting scaling corrections may produce robust statistics that vary in their ability to control Type I error rates. The estimates of each matrix in Table 6 are also no longer asymptotically equivalent when  $H_0$  is false, and the resulting robust statistics may also vary in power. In fact, different computational versions of the robust test statistics may have different power even with normal data. This is because variants of  $a$  and  $c$  that depend on structured estimates of either  $U^0$  or  $\Omega^0$  will only approach 1 when  $H_0$  is true.

Asparouhov and Muthén (2005) offered an alternative expression for  $c_0$  in Equation 37:

$$c_0 = \left( \text{Tr}(B_\beta^0 (A_\beta^0)^{-1}) - \text{Tr}(B_\theta^0 (A_\theta^0)^{-1}) \right) / d, \quad (40)$$

which follows from Equations 14 and 30, the identity  $B_\theta^0 = \Delta'^0 B_\beta^0 \Delta^0$ , as well as properties of the trace operator. Additional estimates of  $c_0$  are possible based on this equation. Most of these estimates will be equivalent to some estimate of Equation 37, but some will not be expressible in this original form, most notably the version used by default with the MLR

<sup>12</sup> $U^0 = A_\beta^0 (I - P_{\Delta^0})$ , where  $I - P_{\Delta^0}$  is an oblique projection matrix onto the column space of  $(\Delta^0)^\perp$ , the orthogonal complement of  $\Delta^0$ , which has rank  $d$ . The meaning of

$U^0$  is that it is the inverse of the asymptotic covariance matrix of the model residuals, i.e., of  $\sqrt{n}(\tilde{\beta} - \hat{\theta})$ .

<sup>13</sup>This version is no longer in much use, but it is available in *lavaan* for complete data via `test = "mean.var.adjusted"`.

**Table 6.** Some estimates of  $U^0$  and  $\Omega^0$ .

	Estimates of $U^0$	Estimates of $\Omega^0$
MAR, MCAR, or complete data	$\hat{U}_{h1} = \hat{A}_\beta - \hat{A}_\beta \hat{\Delta} (\hat{\Delta}' \hat{A}_\beta \hat{\Delta})^{-1} \hat{\Delta}' \hat{A}_\beta$ $\tilde{U} = \tilde{A}_\beta - \tilde{A}_\beta \tilde{\Delta} (\tilde{\Delta}' \tilde{A}_\beta \tilde{\Delta})^{-1} \tilde{\Delta}' \tilde{A}_\beta$ $\hat{U} = \hat{A}_\beta - \hat{A}_\beta \hat{\Delta} \hat{A}_\beta^{-1} \hat{\Delta}' \hat{A}_\beta$ $\tilde{U}_{mix} = \tilde{A}_\beta - \tilde{A}_\beta \tilde{\Delta} \tilde{A}_\beta^{-1} \tilde{\Delta}' \tilde{A}_\beta$	$\hat{\Omega} = \hat{A}_\beta^{-1} \hat{B}_\beta \hat{A}_\beta^{-1}$ $\tilde{\Omega} = \tilde{A}_\beta^{-1} \tilde{B}_\beta \tilde{A}_\beta^{-1}$
MCAR or complete data only	$\hat{U}_E = \hat{W} - \hat{W} \hat{\Delta} (\hat{\Delta}' \hat{W} \hat{\Delta})^{-1} \hat{\Delta}' \hat{W}$ $\tilde{U}_E = \tilde{W} - \tilde{W} \tilde{\Delta} (\tilde{\Delta}' \tilde{W} \tilde{\Delta})^{-1} \tilde{\Delta}' \tilde{W}$ Note: $\tilde{U} = \tilde{U}_E$ with complete data	$\hat{\Omega}_E = \hat{W}^{-1} \hat{B}_\beta \hat{W}^{-1}$ $\tilde{\Omega}_E = \tilde{W}^{-1} \tilde{B}_\beta \tilde{W}^{-1}$ Note: $\tilde{\Omega} = \tilde{\Omega}_E = \tilde{\Gamma}$ with complete data

Not all estimates are listed; many mixed estimates are possible. All estimates of  $U^0$  but  $\tilde{U}$  and  $\tilde{U}_{mix}$  have rank equal to  $d$ . The estimates  $\hat{U}$  and  $\tilde{U}_{mix}$  approach this property only asymptotically (and only when  $H_0$  is true). The scaling corrections are meant to control Type I error rates, which are defined when  $H_0$  is true. All listed estimates are consistent for  $U^0$  and  $\Omega^0$ .

**Table 7.** Estimates of  $U^0 \Omega^0$  used in scaling corrections with complete data in *lavaan*.

<i>lavaan</i> options with <b>test</b> = "satorra.bentler" or <b>test</b> = "scaled.shifted"	Estimate
<i>information</i> = "expected", <i>h1.information</i> = "structured"	$\hat{U}_E \tilde{\Gamma}$
<i>information</i> = "expected", <i>h1.information</i> = "unstructured" or <i>information</i> = "observed", <i>observed.information</i> = "h1", <i>h1.information</i> = "unstructured"	$\tilde{U}_E \tilde{\Gamma} = \tilde{U} \tilde{\Gamma}$
<i>information</i> = "observed", <i>observed.information</i> = "h1", <i>h1.information</i> = "structured"	$\hat{U}_{h1} \tilde{\Gamma}$
<i>information</i> = "observed", <i>observed.information</i> = "hessian", <i>h1.information</i> = "structured"	$\hat{U} \tilde{\Gamma}$
<i>information</i> = "observed", <i>observed.information</i> = "hessian", <i>h1.information</i> = "unstructured"	$\tilde{U}_{mix} \tilde{\Gamma}$

Current *lavaan* defaults are in italics. Note that the inclusion of **se** = "robust.huber.white" will override the default of expected information. When **test** = "satorra.bentler" (default with **estimator** = "MLM"),  $T_M$  is computed; when **test** = "scaled.shifted" (default with **estimator** = "MLMV"),  $T_{MV}$  is computed (see Equation 36). These test options are not compatible with **missing** = "ML". With both tests, *lavaan* options only affect the estimate of  $U^0$ ; the saturated estimate  $\tilde{\Gamma} = \tilde{\Omega}$  of  $\Omega^0$  is always used. For different estimates of  $\Omega^0$ , use **test** = "yuan.bentler" (see Table 9).

estimator in *Mplus* and *lavaan* (see Table 10). The original rationale for Equation 40 was computational convenience.<sup>14</sup>

### Lavaan options for the scaling corrections

The original scaled chi-square  $T_M$  (Satorra & Bentler, 1994) was developed for complete data. In *lavaan*, this test is available via **test** = "satorra.bentler" (default with **estimator** = "MLM", see Table 10). Available computational variations are shown in Table 7. Notably, the estimate of  $\Omega^0$  is always unstructured, in accordance with the original references (Satorra & Bentler, 1994), and only the computation of  $U^0$  varies. The default estimate of  $U^0$  is based on expected information evaluated at the structured estimates.<sup>15</sup> The mean-and-variance adjusted chi-square  $T_{MV}$  is available in *lavaan* via **test** = "scaled.shifted" (default with **estimator** = "MLMV"); because it depends on the same two

matrices, the computational options in Table 7 also apply to this test. Neither **test** = "satorra.bentler" nor **test** = "scaled.shifted" are compatible with **missing** = "ML", so they must be applied to complete data (or listwise deletion will be performed by default).

For incomplete data, only  $T_M$  is directly available in *lavaan*, but  $T_{MV}$  can be obtained via custom computations, as we illustrate shortly. To obtain  $T_M$  with missing data, the user can specify either **test** = "yuan.bentler.mplus" (default with **estimator** = "MLR", see Table 10), which estimates  $c_0$  using Equation 40, or **test** = "yuan.bentler", which estimates  $c_0$  using Equation 37. Both these test options can be run with or without **missing** = "ML", so they can be used with complete data as well, albeit with different defaults (expected rather than observed information).

Computational variations that are available in *lavaan* for these tests are shown in Tables 8 and 9, respectively. With **test** = "yuan.bentler.mplus", *lavaan* computational options only affect the second trace term in Equation 40; the first trace term is always computed using saturated estimates. With **test** = "yuan.bentler", choices of estimates of  $U^0$  and  $\Omega^0$  can be completely crossed. As with **test** = "satorra.bentler", the default computation with **test** = "yuan.bentler" uses an unstructured estimate of  $\Omega^0$  (i.e.,  $\tilde{\Omega}$  when **missing** = "ML" is included) and a structured estimate of  $U^0$ , which also has the property of reduced rank (i.e.,  $\hat{U}_{h1}$  when **missing** = "ML" is included). Because the default type of information for **test** = "yuan.bentler" changes to expected information when **missing** = "ML" is not included, **test** = "satorra.bentler" and **test** = "yuan.bentler" will yield an identical value for  $T_M$  with complete data (but see footnote 15).

### Custom computations

We now illustrate how to verify computations available in *lavaan* using its internal functions and how to obtain custom computations that are not available in *lavaan* 0.6–8. We will use the **HolzingerSwineford1939** dataset available in *lavaan* for illustration.<sup>16</sup> Only the first nine variables were used in the analysis; incomplete data were created by randomly

<sup>14</sup>The estimate in Table 10 can be obtained without computing the matrix  $\hat{\Delta}$ , by obtaining numeric Hessians and by taking the outer product of the individual score vectors (i.e.,  $\tilde{a}_{\beta,i}, \tilde{a}_{\beta,i}$ ).

<sup>15</sup>This default will be overwritten if **se** = "robust.huber.white" is also included, because these types of standard errors trigger the use of observed information, and by default the type of information is set to be the same for both standard errors and tests.

<sup>16</sup>See <https://www.rdocumentation.org/packages/lavaan/versions/0.6-5/topics/HolzingerSwineford1939> for description.

**Table 8.** Scaling correction options based on Equation 40 in *lavaan*.

<i>lavaan</i> options with <b>test = "yuan.bentler.mplus"</b>	Trace Expression
information = "expected", h1.information = "structured" (default without missing = "ML")	$\text{Tr}(\tilde{B}_\beta \tilde{A}_{\beta,E}^{-1}) - \text{Tr}(\hat{B}_\theta \hat{A}_{\theta,E}^{-1})$
information = "expected", h1.information = "unstructured" [*]	$\text{Tr}(\tilde{B}_\beta \tilde{A}_{\beta,E}^{-1}) - \text{Tr}(\hat{B}_\theta \hat{A}_{\theta,E}^{-1}) = \text{Tr}(\tilde{U}_E \tilde{\Omega}_E)$
information = "observed", observed.information = "hessian", h1.information = "structured"	$\text{Tr}(\tilde{B}_\beta \tilde{A}_{\beta}^{-1}) - \text{Tr}(\hat{B}_\theta \hat{A}_{\theta}^{-1})$
information = "observed", observed.information = "h1", h1.information = "structured"	$\text{Tr}(\tilde{B}_\beta \tilde{A}_{\beta}^{-1}) - \text{Tr}(\hat{B}_\theta \hat{A}_{\theta,h1}^{-1})$
information = "observed", observed.information = "h1", h1.information = "unstructured" [*]	$\text{Tr}(\tilde{B}_\beta \tilde{A}_{\beta}^{-1}) - \text{Tr}(\hat{B}_\theta \hat{A}_{\theta}^{-1}) = \text{Tr}(\tilde{U} \tilde{\Omega})$
information = "observed", observed.information = "hessian", h1.information = "unstructured" [*]	$\text{Tr}(\tilde{B}_\beta \tilde{A}_{\beta}^{-1}) - \text{Tr}(\hat{B}_\theta \hat{A}_{\theta}^{-1}) = \text{Tr}(\tilde{U}_{\text{mix}} \tilde{\Omega})$

The scaling correction  $c$  is obtained by dividing the corresponding trace expression by  $d$ . The italics show defaults in the current version of *lavaan* with **missing = "ML"**. The defaults without **missing = "ML"** are as shown unless **se = "robust.huber.white"**. The **h1.information** option does not affect the first term of the trace expression, which is always computed using saturated estimates. Estimates in cells marked with [\*] can also be obtained via **test = "yuan.bentler"** (see Table 9) because the two mathematical forms of the scaling correction become equivalent. Not all possible options are shown; the **h1.information.meat** option can be used to specify a different type of estimate for  $B_\theta^0$  only.

**Table 9.** Scaling correction options based on Equation 37 in *lavaan*.

<i>lavaan</i> options with <b>test = "yuan.bentler"</b>	Estimate of
information = "expected", h1.information = "structured" (default without missing = "ML")	$U^0$
information = "expected", h1.information = "unstructured"	$\hat{U}_E$
information = "observed", observed.information = "h1", h1.information = "unstructured"	$\tilde{U}$
information = "observed", observed.information = "h1", h1.information = "structured"	$\hat{U}_{h1}$
information = "observed", observed.information = "hessian", h1.information = "structured"	$\hat{U}$
information = "observed", observed.information = "hessian", h1.information = "unstructured"	$\tilde{U}_{\text{mix}}$
Additional <i>lavaan</i> options with <b>test = "yuan.bentler"</b>	Estimate of
omega.information = "expected", omega.h1.information = "structured"	$\hat{\Omega}_E$
omega.information = "expected", omega.h1.information = "unstructured" (default without missing = "ML")	$\tilde{\Omega}_E$
omega.information = "observed", omega.h1.information = "unstructured"	$\tilde{\Omega}$
omega.information = "observed", omega.h1.information = "structured"	$\hat{\Omega}$

The defaults in the current version of *lavaan* with **missing = "ML"** are in italics. The defaults without **missing = "ML"** are as shown unless **se = "robust.huber.white"**. The options for estimates of  $U^0$  and  $\Omega^0$  can be completely crossed. If the additional options are omitted, the default computation of  $\Omega^0$  is performed. Not all possible options are shown; the **h1.information.meat** option can be used to specify mixed estimates of  $U^0$ , and the **omega.h1.information.meat** option can be used to specify mixed estimates of  $\Omega^0$ .

deleting 20 observations from these variables. The final dataset contains  $n = 301$  rows and  $p = 9$  columns. The model fit to data is a 3-factor model with three indicators per factor and correlated factors. This model has 24 degrees of freedom. The code below creates the dataset and the model:

```
library(lavaan)
set.seed(11)
HS.comp <- HolzingerSwineford1939[, paste0("x", 1:9)]
HS.missing <- as.data.frame(lapply(HS.comp, function(x) {
  x[ sample(1:length(x), 20) ] <- NA
  x })))
HS.model <- 'visual =~ x1 + x2 + x3
textual =~ x4 + x5 + x6
speed =~ x7 + x8 + x9'
```

The syntax below fits the model using MLR and shows how to verify the default settings for information options under this choice of estimator:

```
fit.mlr <- sem(HS.model, data = HS.missing, missing = "ml", estimator="MLR")
lavInspect(fit.mlr, "options")$test
## [1] "yuan.bentler.mplus"
lavInspect(fit.mlr, "options")$se
## [1] "robust.huber.white"
lavInspect(fit.mlr, "options")$information
## [1] "observed" "observed"
lavInspect(fit.mlr, "options")$observed.information
## [1] "hessian" "hessian"
lavInspect(fit.mlr, "options")$h1.information
## [1] "structured" "structured"
```

As this output illustrates, the information options are vectors with two entries, which correspond to the defaults for the standard errors and the test statistic, respectively. This output above verifies the current defaults for "MLR" as shown in Table 10. The default information options can be changed for standard errors and the test statistic together or separately. The run **fit.mlr** produces the following values of  $T$ ,  $T_M$ , and  $c$ :

```
fitMeasures(fit.mlr, c("df", "chisq", "chisq.scaled", "chisq.scaling.factor"))
##          df          chisq      chisq.scaled
##      24.000         89.043         89.654
## chisq.scaling.factor
##          0.993
```

The default scaling correction for **test = "yuan.bentler.mplus"** uses a mix of unstructured and structured estimates (see Table 8). The following code replicates this scaling correction using *lavaan*'s internal functions:

```
fit.mlr@options$h1.information <- "unstructured"
Abeta.unstr <- lavInspect(fit.mlr, "h1.information.observed")
Bbeta.unstr <- lavInspect(fit.mlr, "h1.information.firstorder")
tr1 <- sum(diag(solve(Abeta.unstr) %*% Bbeta.unstr))

fit.mlr@options$h1.information <- "structured"
Atheta.hessian <- lavInspect(fit.mlr, "information.observed")
Btheta.str <- lavInspect(fit.mlr, "first.order")
tr2 <- sum(diag(solve(Atheta.hessian) %*% Btheta.str))

df <- fitMeasures(fit.mlr, "df")
c <- (tr1 - tr2)/df; names(c) <- "c.yb.mplus"
c
## c.yb.mplus
##          0.993
```

This code illustrates that one can obtain the estimates of  $A_\beta$ ,  $B_\beta$ ,  $A_\theta$ ,  $B_\theta$  using the **lavInspect** function.<sup>17</sup> Options for  $A_\beta$  include **"h1.information.observed"** and **"h1.information.expected"**. Similarly, options for  $A_\theta$  include **"information.observed"** (which produces the

<sup>17</sup>For advanced users, we note that *lavaan* organizes  $H_1$  matrices corresponding to the order  $(\mu', (\text{vech}\Sigma)')$ ; e.g., see **row.names(Abeta.unstr)**.



**Table 10.** “MLM” and “MLR” definitions in *lavaan*.

	estimator = “MLM”	estimator = “MLM”, mimic = “Mplus”	estimator = “MLR”
Equivalent Specification	estimator = “ML”, se = “robust.sem”, test = “satorra.bentler”, information = “expected”, h1.information = “structured”	estimator = “ML”, se = “robust.sem”, test = “satorra.bentler”, information = “expected”, h1.information = “unstructured”	estimator = “ML”, [missing = “ML”], se = “robust.huber.white”, test = “yuan.bentler.mplus”, information = “observed”, observed.information = “hessian”, h1.information = “structured”
Standard Errors	$\hat{A}_{\theta,E}^{-1} \hat{B}_{\theta} \hat{A}_{\theta,E}^{-1}$	$\tilde{A}_{\theta,E}^{-1} \tilde{B}_{\theta} \tilde{A}_{\theta,E}^{-1}$	$\hat{A}_{\theta}^{-1} \hat{B}_{\theta} \hat{A}_{\theta}^{-1}$
Scaling Correction	$\text{Tr}(\hat{U}_E \tilde{\Gamma})/d$	$\text{Tr}(\tilde{U}_E \tilde{\Gamma})/d$	$(\text{Tr}(\tilde{B}_{\theta} \tilde{A}_{\theta}^{-1}) - \text{Tr}(\hat{B}_{\theta} \hat{A}_{\theta}^{-1}))/d$

Options triggered by default are in italics. Specifying `mimic = “Mplus”` leads *lavaan* to mimic the “MLM” method as implemented in *Mplus*, which uses unstructured estimates (in the “Equivalent Specification” syntax for this option, setting `information = “observed”` will produce identical results because  $\hat{A}_{\theta,E} = \hat{A}_{\theta}$  with complete data). The “MLM” option is not compatible with `missing = “ML”`. Not included in the table is the “MLMV” option, which differs from “MLM” only in that it sets `test = “scaled.shifted”`, so that  $T_{MV}$  instead of  $T_M$  is computed, with the same defaults. Under the default options for “MLM”, the resulting standard errors will correctly adjust for nonnormality, but not for model misspecification. Under the default options for “MLR” (i.e., using the numeric Hessian as the “bread”), the resulting standard errors will adjust for nonnormality and model misspecification.

numeric Hessian) and “`information.expected`”. Whether structured or saturated estimates are used in the computation of estimates can be controlled via `@Options $h1.information`. This option can be toggled before the computation of each matrix, making it useful for obtaining structured and unstructured estimates from the same model run. Note that the numeric Hessian does not respond to this option (i.e., the object `Atheta.hessian` will always have the same value regardless of what options precede it) because it is not obtained analytically. To get the analytic approximation in Equation 16, one would need to compute it according to that equation.

The next example verifies the default computation of the scaling correction  $c$  used in the statistic  $T_M$  corresponding to `test = “yuan.bentler”`, and also shows how to obtain  $T_{MV}$  from this run, which is not available in *lavaan* with incomplete data. The type of test and the type of standard errors cannot be changed with `estimator = “MLR”` because the choice of estimator will override any changes. Thus, to request `test = “yuan.bentler”` instead of `test = “yuan.bentler.mplus”`, the user needs to specify instead:

```
fit.mlr.yb <- sem(HS.model, data = HS.missing, missing = “ml”,
  estimator=“ML”,test=“yuan.bentler”,se=“robust.huber.white”)
```

The default scaling correction for this test is  $c = \text{Tr}(\hat{U}_{h1} \hat{\Omega})/d$ . Below is the code to reproduce it:

```
Omega <- solve(Abeta.unstr) %*% Bbeta.unstr %*% solve(Abeta.unstr)
Delta <- lavInspect(fit.mlr, “delta”)
fit.mlr@Options$h1.information <- “structured”
Abeta.str <- lavInspect(fit.mlr, “h1.information.observed”)
Uhat <- Abeta.str - Abeta.str %*% Delta %*% solve(t(Delta)) %*%
  Abeta.str %*% Delta %*% t(Delta) %*% Abeta.str
c <- (sum(diag(Uhat %*% Omega)))/df; names(c) <- “c”

c
## c
## 1.04
T.M <- fitMeasures(fit.mlr, “chisq”)/c; names(T.M) <- “T.M”
T.M
## T.M
## 85.638

fitMeasures(fit.mlr.yb, c(“chisq.scaled”, “chisq.scaling.factor”))
## chisq.scaled chisq.scaling.factor
## 85.638 1.040
```

The following code obtains  $T_{MV}$  based on the same estimates:

```
a <- sqrt(sum(diag(Uhat %*% Omega %*% Uhat %*% Omega))/df)
b <- df * (1 - c/a)
names(a) <- “a”; names(b) <- “b”
a; b
## a
## 1.276
## b
## 4.441
T.MV <- fitMeasures(fit.mlr.yb, “chisq”)/a + b; names(T.MV) <- “T.MV”
T.MV
## T.MV
## 74.233
```

The code in this section can be adapted to compute any other variation of the robust statistics or robust standard errors.

## Discussion

In this article, we have summarized the different computational versions of standard errors and test statistics that are used with incomplete normal and nonnormal data in SEM. Complete data equations were included as a special case. We have illustrated how to obtain the available options in the popular R package *lavaan*, and have shown how to obtain currently unavailable options via *lavaan*’s internal functions. We hope that this article encourages the extensive study of these computational options by methodologists. Scenarios of interest would be formed by the intersection of the following options: complete and different types of incomplete data, normal and different types of nonnormal data, and different types of correctly specified and misspecified models. We briefly review the existing literature, which is largely limited to complete data.

## Existing comparisons

With complete data, some research has been devoted to investigating whether saturated or structured estimates of the expected information matrix result in better performance of robust standard errors and test statistics. Noting that with complete data, EQS uses structured estimates in the computation of the expected information matrix whereas *Mplus* uses saturated estimates, Xia et al. (2016) compared the impact of this choice on the performance of robust standard errors and of  $T_M$ . They investigated data with varying degrees of nonnormality, and



included both correct and incorrect models. They found that structured estimates were preferred in both computations. Foldnes and Olsson (2019) replicated this finding with models of varying size and degrees of freedom and with additional types of nonnormal data; these authors only studied correctly specified models.

The impact of using observed versus expected information estimates with complete data has been investigated by several authors. Dolan and Molenaar (1991) found very minor differences between observed and expected information in ML standard errors with normal data (see also Maydeu-Olivares, 2017). Studying only correct models, Savalei (2010a) found that observed information outperformed expected information in both ML and robust standard errors, with both complete and incomplete data. Differences were more substantial with incomplete MAR data, as would be theoretically expected.

In the context of complete data, Maydeu-Olivares (2017) conducted a comprehensive investigation of ML and robust standard errors as implemented in *Mplus*. Models with varying degree of misspecification were studied. Notably, ML standard error estimates based on the outer product matrix were included, which exhibited significant bias even with normal data and under correct model specification. Three types of robust standard errors were investigated, which correspond to *Mplus*'s version of MLM (see Table 10), MLR, and MLR with expected information. When the model was correctly specified, all three types of standard errors performed well with normal and nonnormal data. When the model was misspecified, the default MLR standard errors outperformed the other two, as would be theoretically expected (see also Lai, 2018). Finally, robust test statistics under the same three specifications were also investigated. Scaling corrections based on Equation 37, which are used with MLR, performed the worst, regardless of type of information. The MLM statistic performed only slightly better, although this may be because it is based on saturated estimates in *Mplus* (Xia et al., 2016). The statistic that performed best was  $T_{MV}$  (estimator = "MLMV"), also based on the saturated estimates. Overall, it is hard to draw firm conclusions from this study because several options are confounded in the default implementation of options in *Mplus*.

### Future research

In this section, we outline some interesting future research directions. When it comes to standard errors, given the good relative performance of Huber-White standard errors (default with MLR) in existing studies with complete data (Lai, 2018; Maydeu-Olivares, 2017), further studies of how these standard errors perform with different types of incomplete data are needed. These standard errors are robust to both nonnormality and model misspecification. Often in statistics, the price of such broad robustness is poorer performance relative to methods that make more assumptions, as long as those assumptions are correct. However, good performance of these standard errors with complete data established in prior research is encouraging in this regard. When the model is correct or approximately

correct, many other comparisons become of interest. One is between the use of the numeric Hessian versus the analytic asymptotic approximations in both ML and robust standard errors; these two estimates have never been directly compared, though the latter version has been used in some studies of standard errors with incomplete data (Savalei & Bentler, 2009; Savalei & Falk, 2014). Another comparison of interest is between structured and saturated estimates in robust standard errors, to determine whether findings from complete data (e.g., Xia et al., 2016) generalize to incomplete data.

Another research direction would be to explore whether the property of consistency of Huber-White standard errors is meaningful when the model is misspecified. In existing studies with misspecified models, Huber-White standard errors have been evaluated based on whether they accurately approximate the actual empirical variability of parameter estimates across replications and whether they have good coverage of the pseudo-parameters. To understand the meaning of the pseudo-parameters, consider a population 1-factor model with five indicators, with unit variances, with loadings of .7, and a residual covariance of .3 between the first two indicators. Suppose a misspecified model with no residual covariance is fit to this population. The true loadings are  $\lambda^0 = (.7, .7, .7, .7, .7)'$ , but the loadings obtained by fitting the 1-factor model without a residual covariance to the population covariance matrix are  $\lambda^* = (.87, .87, .595, .595, .595)'$ . To complicate matters further, with incomplete data, these pseudo-parameters may vary with the characteristics of missing data (X. Zhang & Savalei, 2020). Consistency of Huber-White standard errors in this situation means that in large samples they will provide accurate estimates of the variability of  $\hat{\lambda}$  around the pseudo-parameters  $\lambda^*$ , not around the true loadings  $\lambda^0$ . Simulation studies with misspecified models should consider adding measures such as coverage defined around the true values (in our example,  $\lambda^0$ ) rather than around the pseudo-values ( $\lambda^*$ ). Huber-White standard errors are no longer guaranteed to have better performance for such measures.

When it comes to the test statistics, the impact of the computational options on the scaling corrections has been studied very little even with complete data, and with incomplete data, it is entirely unstudied. Studies with complete data have shown that structured estimates of  $U^0$  are preferred when computing  $T_M$  (Xia et al., 2016). Future research should explore whether this finding generalizes to incomplete data. With complete data, typically a saturated estimate of  $\Omega^0$  is always used regardless of how  $U^0$  is estimated. Future research should investigate the impact of using saturated versus structured estimates in the matrices involved in the scaling, with both complete and incomplete data. We should note that existing studies of robust with incomplete data statistics have not always made clear how these matrices were estimated.

Studies with complete data have also found that the variant of the scaling correction for  $T_M$  based on Equation 40 (default with MLR) performs worse (Maydeu-Olivares, 2017), but similar comparisons with incomplete data have not been conducted. While some authors have argued that the estimate of

$\text{Tr}(U^0\Omega^0)$  needs to have rank  $d$  in order for the scaling correction to make sense (Bentler & Yuan, 1999; Jiang & Yuan, 2017), in our view it is the finite sample performance that should be used to judge the quality of the various versions of the scaling corrections. Asparouhov and Muthén (2005) have also identified other benefits of estimates of Equation 37, such as ease of difference tests computations.

The statistic  $T_{MV}$  has been recommended with both complete and incomplete MCAR data based on its performance in limited simulation studies (Maydeu-Olivares, 2017; Savalei, 2010b). However, it is not often used with incomplete data because it is not readily available in software such as *Mplus* or *lavaan*. We have shown how to obtain it using custom computations in *lavaan*. The various computational options for it have never been studied. Aside from comparing the relative performance of the different computational versions of  $T_M$  and  $T_{MV}$  to find the best implementation of each, a thorough comparison, in terms of Type I error rates and power, of the best performing version of  $T_M$  and  $T_{MV}$  to each other would be needed with both complete and incomplete data.

In addition to the standard errors and test statistic, the options affecting the computation of the information matrix and the outer product matrix will impact other quantities, such as fit indices and estimates of fraction of missing information. When data are nonnormal, robust difference tests rely on the same scaling corrections, robust equations for RMSEA and CFI use the scaling correction  $c$  to improve finite sample performance (Brosseau-Liard & Savalei, 2014; Brosseau-Liard et al., 2012). Thus, they too will vary depending on how the scaling correction is computed. With complete data, Xia et al. (2016) found that robust RMSEA and CFI were affected very little by whether saturated or structured estimates were used in the information matrix to compute  $c$ .<sup>18</sup> Further, the *Mplus* variant of the scaling correction has not been studied in this context. The information matrix is also used to estimate the fraction of missing information for each parameter; we are currently investigating how the computational choices for the information matrix affect these estimates. Scaling corrections are also used when computing different versions of robust difference tests for nested models (Satorra & Bentler, 2001, 2010), and the impact of computational choices on the performance of these tests has not been studied. Finally, other estimation methods for incomplete nonnormal data, such as two-stage ML (Savalei & Falk, 2014; Yuan & Lu, 2008) are also affected by the same computational choices, and their impact needs to be studied.

Most simulation studies in the SEM literature generate nonnormal data using the Vale-Maurelli method (Vale & Maurelli, 1983), which is a generalization of the Fleishman method (Fleishman, 1978). This method permits the user to specify univariate skewness and kurtosis for each variable, and then subsequently rotates the set of variables to the pre-specified mean and covariance structure. This method has been criticized on several grounds (e.g., Astivia & Zumbo, 2019), but the most significant criticism is that it generates very limited kinds of multivariate nonnormal distributions, ones with multivariate properties very close to the Gaussian despite nonnormal marginals (Foldnes &

Grønneberg, 2015). While many existing studies of test statistics took care to demonstrate that the LR chi-square statistic  $T$  did break down in the generated nonnormal data, it is not known whether the findings that robust statistics performed well in these studies will hold up for nonnormal data generated in other ways. For robust standard errors, it has been shown that the type of nonnormality does influence performance (Falk, 2018). Many alternative methods of nonnormal data generation exist (e.g., Grønneberg & Foldnes, 2017; Mair et al., 2012; Qu et al., 2019), and the performance of the various computational versions of robust standard errors and test statistics should be evaluated on data with more varied types of multivariate nonnormality.

Alternative types of uncertainty estimates for nonnormal data exist. Lai (2019) proposed an improvement to MLR standard errors based on a second-order approximation, and Falk (2018) proposed robust likelihood-based confidence intervals. Extensions of these methods to incomplete data present interesting future research directions. Future research should also consider extending and studying variants of Huber-White standard errors for handling outlying observations, which have been developed in the regression context (e.g., MacKinnon & White, 1985), but extensions to SEM are limited (but see Yuan & Zhang, 2012). Finally, in this article, we have dealt with the most straight-forward type of mean and covariance structure model. More complicated models would involve conditioning on exogenous covariates, the presence of constraints, or the presence of interaction terms. The equations in this article do not cover these extensions, and these extensions should be subject of future research.

## Conclusion

We hope that the present article encourages methodologists to investigate the performance of the different computational versions of standard errors and robust test statistics in a systematic way and across a wide variety of types of incomplete data, model misspecification, and nonnormality. It would be important to conduct such studies without confounding the available computational variations by using software defaults. We have provided the *lavaan* code to disentangle the available options. We also hope that this article communicates the importance of reporting exactly how a particular method was implemented even in studies that do not make computational variations their focus; otherwise, future aggregation of results across studies will not be meaningful.

While we have outline types of standard error and scaling correction estimates that are consistent under various scenarios, we believe it is too early for applied researchers to use this information in order to select which options to use in practice. Theoretical expectations and practical performance do not always align. Consistency is not the only desirable property of an estimator, and small sample performance, measured by bias and efficiency, would be relevant. For example, while theoretically there is usually no good reason to use expected information in computations with incomplete data (exceptions would include a planned missingness design, Graham et al., 2006), in practice

<sup>18</sup>While Maydeu-Olivares (2017) included RMSEA in the study, the default *Mplus* implementation was studied, which does not correctly adjust for nonnormality; see Brosseau-Liard et al. (2012).

the expected information matrix may be more stable in small samples. We have found, for example, that its use in the scaling correction sometimes produces more interpretable results when the independence model is fit to data (e.g., for the purposes of computing CFI). Overall, the scarcity of research on the relative performance of standard errors, scaling corrections, and other statistics dependent on them, especially with incomplete data, means that it would be premature to make recommendations. The goal of the present article is to make the available options explicit and to encourage their study by methodologists.

## Funding

This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC) to Victoria Savalei.

## ORCID

Yves Rosseel  <http://orcid.org/0000-0002-4129-4477>

## References

- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Lawrence Erlbaum Associates.
- Arminger, G., & Schoenberg, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika*, 54, 409–425. <https://doi.org/10.1007/BF02294626>
- Arminger, G., & Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, 85, 195–203. <https://doi.org/10.1080/01621459.1990.10475326>
- Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) research conference*, Arlington, VA. <http://www.statmodel.com/download/2005FCSM.pdf>
- Asparouhov, T., & Muthén, B. (2010). *Simple second order chi-square correction*. [https://www.statmodel.com/download/WLSMV\\_new\\_chi21.pdf](https://www.statmodel.com/download/WLSMV_new_chi21.pdf)
- Astivia, O. L. O., & Zumbo, B. D. (2019). A note on the solution multiplicity of the Vale-Maurelli intermediate correlation equation. *Journal of Educational and Behavioral Statistics*, 44, 127–143. <https://doi.org/10.3102/1076998618803381>
- Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34, 181–197. <https://doi.org/10.1207/S15327906Mb340203>
- Brosseau-Liard, P., & Savalei, V. (2014). Adjusting relative fit indices for nonnormality. *Multivariate Behavioral Research*, 49, 460–470. <https://doi.org/10.1080/00273171.2014.933697>
- Brosseau-Liard, P., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two non-normality corrections for RMSEA. *Multivariate Behavioral Research*, 47, 904–930. <https://doi.org/10.1080/00273171.2012.715252>
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83. <https://doi.org/10.1111/j.2044-8317.1984.tb00789.x>
- Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance- structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). Springer. <https://www.springer.com/gp/book/9780306448058>
- Dolan, C. V., & Molenaar, P. C. (1991). A comparison of four methods of calculating standard errors of maximum-likelihood estimates in the analysis of covariance structure. *British Journal of Mathematical and Statistical Psychology*, 44, 359–368. <https://doi.org/10.1111/j.2044-8317.1991.tb00967.x>
- Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Berkeley, CA: University of California Press.
- Falk, C. F. (2018). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 244–266. <https://doi.org/10.1080/10705511.2017.1367254>
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521–532. <https://doi.org/10.1007/BF02293811>
- Foldnes, N., & Grønneberg, S. (2015). How general is the Vale-Maurelli simulation approach? *Psychometrika*, 80, 1066–1083. <https://doi.org/10.1007/s11336-014-9414-0>
- Foldnes, N., & Olsson, U. H. (2019). The choice of normal-theory weight matrix when computing robust standard errors in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 861–875. <https://doi.org/10.1080/10705511.2019.1600408>
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 11, 323. <https://doi.org/10.1037/1082-989X.11.4.323>
- Grønneberg, S., & Foldnes, N. (2017). Covariance model simulation using regular vines. *psychometrika*, 82, 1035–1051. <https://doi.org/10.1007/s11336-017-9569-6>
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Berkeley, CA: University of California Press.
- Jiang, G., & Yuan, K.-H. (2017). Four new corrected statistics for sem with small samples and nonnormally distributed data. *Structural Equation Modeling: A Multidisciplinary Journal*, 24, 479–494. <https://doi.org/10.1080/10705511.2016.1277726>
- Kim, K. H., & Bentler, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika*, 67, 609–623. <https://doi.org/10.1007/BF02295134>
- Lai, K. (2018). Estimating standardized sem parameters given nonnormal data and incorrect model: Methods and comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 600–620. <https://doi.org/10.1080/10705511.2017.1392248>
- Lai, K. (2019). More robust standard error and confidence interval for sem parameters given incorrect model and nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 260–279. <https://doi.org/10.1080/10705511.2018.1505522>
- Lee, S.-Y. (2007). *Structural equation modeling: A bayesian approach* (Vol. 711). John Wiley & Sons.
- Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119013563>
- MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 305–325. [https://doi.org/10.1016/0304-4076\(85\)90158-7](https://doi.org/10.1016/0304-4076(85)90158-7)
- Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley.
- Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, 47, 547–565. <https://doi.org/10.1080/00273171.2012.692629>
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 24, 383–394. <https://doi.org/10.1080/10705511.2016.1269606>
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.
- Qu, W., Liu, H., & Zhang, Z. (2019). A method of generating multivariate non-normal random numbers with desired multivariate skewness and kurtosis. *Behavior Research Methods*, 51, 1–8. <https://doi.org/10.3758/s13428-018-1072-1>



- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In C. C. C. von Eye Alexander (Ed.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514. <https://doi.org/10.1007/BF02296192>
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75, 243–248. <https://doi.org/10.1007/s11336-009-9135-y>
- Savalei, V. (2010a). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, 15, 352. <https://doi.org/10.1037/a0020143>
- Savalei, V. (2010b). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a monte carlo comparison. *Structural Equation Modeling*, 17, 241–264. <https://doi.org/10.1080/10705511003659375>
- Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 16, 477–497. <https://doi.org/10.1080/10705510903008238>
- Savalei, V., & Falk, C. F. (2014). Robust two-stage approach outperforms robust full information maximum likelihood with incomplete nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 280–302. <https://doi.org/10.1080/10705511.2014.882692>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures. *South African Statistical Journal*, 17, 33–81. [https://journals.co.za/doi/10.10520/AJA0038271X\\_800](https://journals.co.za/doi/10.10520/AJA0038271X_800)
- Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statistical Journal*, 19, 73–81. [https://journals.co.za/doi/abs/10.10520/AJA0038271X\\_810](https://journals.co.za/doi/abs/10.10520/AJA0038271X_810)
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48, 465–471. <https://doi.org/10.1007/BF02293687>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838. <https://doi.org/10.2307/1912934>
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 1, 1–25. <https://doi.org/10.2307/1912526>
- Xia, Y., Yung, Y.-F., & Zhang, W. (2016). Evaluating the selection of normal-theory weight matrices in the satorra-bentler correction of chi-square and standard errors. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 585–594. <https://doi.org/10.1080/10705511.2016.1141354>
- Yuan, K.-H. (2009). Normal distribution based pseudo ml for missing data: With applications to mean and covariance structure analysis. *Journal of Multivariate Analysis*, 100, 1900–1918. <https://doi.org/10.1016/j.jmva.2009.05.001>
- Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30, 165–200. <https://doi.org/10.1111/0081-1750.00078>
- Yuan, K.-H., & Bentler, P. M. (2010). Consistency of normal-distribution-based pseudo maximum likelihood estimates when data are missing at random. *The American Statistician*, 64, 263–267. <https://doi.org/10.1198/tast.2010.09203>
- Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, 59, 397–417. <https://doi.org/10.1348/000711005X85896>
- Yuan, K.-H., Jamshidian, M., & Kano, Y. (2018). Missing data mechanisms and homogeneity of means and variances-covariances. *Psychometrika*, 83, 425–442. <https://doi.org/10.1007/s11336-018-9609-x>
- Yuan, K.-H., & Lu, L. (2008). Sem with missing data and unknown population distributions using two-stage ml: Theory and its application. *Multivariate Behavioral Research*, 43, 621–652. <https://doi.org/10.1080/00273170802490699>
- Yuan, K.-H., & Savalei, V. (2014). Consistency, bias and efficiency of the normal-distribution-based mle: The role of auxiliary variables. *Journal of Multivariate Analysis*, 124, 353–370. <https://doi.org/10.1016/j.jmva.2013.11.006>
- Yuan, K.-H., & Zhang, Z. (2012). Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika*, 77(4), 803–826. <https://doi.org/10.1007/s11336-012-9282-4>
- Zhang, Q., Yuan, K.-H., & Wang, L. (2019). Asymptotic bias of normal-distribution-based maximum likelihood estimates of moderation effects with data missing at random. *British Journal of Mathematical and Statistical Psychology*, 72, 334–354. <https://doi.org/10.1111/bmsp.12151>
- Zhang, X., & Savalei, V. (2020). Examining the effect of missing data on RMSEA and CFI under normal theory full-information maximum likelihood. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 219–239. <https://doi.org/10.1080/10705511.2019.1642111>

## Appendix

### A. Derivation of $A_{\beta,i}$ and $A_{\beta,n}$

Let  $\beta'_i = ((\text{vech}\Sigma_i)', \mu'_i)'$ . The first differential of the log-likelihood in Equation 1 is given by (for how to work with differentials, see Magnus & Neudecker, 1999):

$$dl_i(\beta_i) = -\frac{1}{2}\text{Tr}(\Sigma_i^{-1}d\Sigma_i) + (y_i - \mu_i)'\Sigma_i^{-1}d\mu_i + \frac{1}{2}(y_i - \mu_i)'\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}(y_i - \mu_i)$$

The first differential of the first term of  $dl_i(\beta_i)$  is given by:

$$-\frac{1}{2}d\text{Tr}(\Sigma_i^{-1}d\Sigma_i) = -\frac{1}{2}\text{Tr}(d\Sigma_i^{-1}d\Sigma_i) = \frac{1}{2}\text{Tr}(\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}d\Sigma_i) = \frac{1}{2}(d\text{vec}\Sigma_i)'(\Sigma_i^{-1} \otimes \Sigma_i^{-1})d\text{vec}\Sigma_i$$

The first differential of the second term of  $dl_i(\beta_i)$  is given by:

$$d((y_i - \mu_i)'\Sigma_i^{-1}d\mu_i) = -d\mu_i'\Sigma_i^{-1}d\mu_i - (y_i - \mu_i)'\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}d\mu_i = -d\mu_i'\Sigma_i^{-1}d\mu_i - (d\text{vec}\Sigma_i)'(\Sigma_i^{-1}(y_i - \mu_i) \otimes \Sigma_i^{-1})d\mu_i$$

The first differential of the third term of  $dl_i(\beta_i)$  is given by:

$$\begin{aligned} \frac{1}{2}d((y_i - \mu_i)'\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}(y_i - \mu_i)) &= -.5(2)(y_i - \mu_i)'\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}d\mu_i - .5(2)(y_i - \mu_i)'\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}d\Sigma_i\Sigma_i^{-1}(y_i - \mu_i) \\ &= -(d\text{vec}\Sigma_i)'(\Sigma_i^{-1}(y_i - \mu_i) \otimes \Sigma_i^{-1})d\mu_i - (d\text{vec}\Sigma_i)'(\Sigma_i^{-1} \otimes \Sigma_i^{-1})(y_i - \mu_i)(y_i - \mu_i)'\Sigma_i^{-1}(d\text{vec}\Sigma_i) \end{aligned}$$

The second differential of  $l_i(\beta_i)$  is the sum of the three components above (components of the first differential of  $dl_i(\beta_i)$ ), and is given by:

$$\begin{aligned} d^2l_i(\beta_i) &= .5(d\text{vec}\Sigma_i)'(\Sigma_i^{-1} \otimes \Sigma_i^{-1})d\text{vec}\Sigma_i - d\mu_i'\Sigma_i^{-1}d\mu_i - \\ &2(d\text{vec}\Sigma_i)'(\Sigma_i^{-1}(y_i - \mu_i) \otimes \Sigma_i^{-1})d\mu_i - (d\text{vec}\Sigma_i)'(\Sigma_i^{-1} \otimes \Sigma_i^{-1})(y_i - \mu_i)(y_i - \mu_i)'\Sigma_i^{-1}(d\text{vec}\Sigma_i) \end{aligned}$$

The negative Hessian for the  $i$ th observation is (Magnus & Neudecker, 1999):

$$-\frac{\partial^2 l_i \beta_i}{\partial \beta_i \partial \beta_i'} = \begin{pmatrix} D'_{p_i}(\Sigma_i^{-1} \otimes (\Sigma_i^{-1}(y_i - \mu_i)(y_i - \mu_i)'\Sigma_i^{-1} - .5\Sigma_i^{-1}))D_{p_i} & D'_{p_i}(\Sigma_i^{-1}(y_i - \mu_i) \otimes \Sigma_i^{-1}) \\ ((y_i - \mu_i)'\Sigma_i^{-1} \otimes \Sigma_i^{-1})D_{p_i} & \Sigma_i^{-1} \end{pmatrix}$$

To obtain the Hessian with respect to  $\beta = (\text{vech}\Sigma', \mu')'$ , we write  $\beta'_i = ((\text{vech}\Sigma_i)', \mu'_i)' = ((D_{p_i}^+ \kappa_i \text{vech}\Sigma)', (\tau_i \mu')')$ , so that  $\frac{\partial \beta_i}{\partial \beta'} = \begin{pmatrix} D_{p_i}^+ \kappa_i & 0 \\ 0 & \tau_i \end{pmatrix}$ . Using the identity  $D_{p_i} D_{p_i}^+ \kappa_i = \kappa_i$ , we obtain

$$\begin{aligned} A_{\beta,i} &= -\frac{\partial^2 l_i(\beta)}{\partial \beta \partial \beta'} = -\frac{\partial \beta_i}{\partial \beta} \frac{\partial^2 l_i(\beta_i)}{\partial \beta_i \partial \beta_i'} \frac{\partial \beta_i}{\partial \beta'} \\ &= \begin{pmatrix} \kappa'_j \left( \Sigma_j^{-1} \otimes (\Sigma_j^{-1}(y_i - \mu_j)(y_i - \mu_j)'\Sigma_j^{-1} - .5\Sigma_j^{-1}) \right) \kappa_j & \kappa'_j (\Sigma_j^{-1}(y_i - \mu_j) \otimes \Sigma_j^{-1} \tau_j) \\ ((y_i - \mu_j)'\Sigma_j^{-1} \otimes \tau'_j \Sigma_j^{-1}) \kappa_j & \tau'_j \Sigma_j^{-1} \tau_j \end{pmatrix}. \end{aligned}$$

Sorting by missing data patterns and averaging across  $n$ , we obtain

$$A_{\beta,n} = \frac{1}{J} \sum_{j=1}^J \frac{n_j}{n} \begin{pmatrix} \kappa'_j \left( \Sigma_j^{-1} \otimes \left( (\Sigma_j^{-1} S_j \Sigma_j^{-1} + h_{j,n} h'_{j,n}) - .5\Sigma_j^{-1} \right) \right) \kappa_j & \kappa'_j (h_{j,n} \otimes \Sigma_j^{-1} \tau_j) \\ (h'_{j,n} \otimes \tau'_j \Sigma_j^{-1}) \kappa_j & \tau'_j \Sigma_j^{-1} \tau_j \end{pmatrix},$$

where  $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_i$ ,  $h_{j,n} = \Sigma_j^{-1}(\bar{y}_j - \mu_j)$ , and  $S_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_i - \bar{y}_j)(y_i - \bar{y}_j)'$  for  $j = 1, \dots, J$ . See also (Magnus & Neudecker, 1999), p. 319, Exercise 2.

### B. Derivation of $B_{\beta,n}$

Here we provide expressions for the components of  $B_{\beta,n}$  in Equation 20. We write



$$B_{\beta,n} = \begin{pmatrix} B_{11} & B'_{21} \\ B_{21} & B_{22} \end{pmatrix},$$

where the blocks correspond to the partitioning of  $\beta = (\text{vech}\Sigma', \mu')'$ . An explicit expression for  $d_{\beta,i}$  in Equation 2 is

$$\begin{aligned} d_{\beta,i} &= \left( \frac{\partial l_i(\beta)}{\partial (\text{vech}\Sigma)'} , \frac{\partial l_i(\beta)}{\partial \mu'} \right) \\ &= \left( .5\text{vec}\left(\Sigma_j^{-1}(y_i - \mu_j)(y_i - \mu_j)' \Sigma_j^{-1} - \Sigma_j^{-1}\right) \kappa_j, (y_i - \mu_j)' \Sigma_j^{-1} \tau_j \right)'. \end{aligned}$$

Let  $w_i = \text{vec}(y_i - \mu_j)(y_i - \mu_j)'$ . For the  $i$ th observation belonging to the  $j$ th missing data pattern, we have:

$$\begin{aligned} \frac{\partial l_i(\beta)}{\partial \mu} \frac{\partial l_i(\beta)}{\partial \mu'} &= \tau_j' \Sigma_j^{-1} (y_i - \mu_j)(y_i - \mu_j)' \Sigma_j^{-1} \tau_j \\ \frac{\partial l_i(\beta)}{\partial \mu} \frac{\partial l_i(\beta)}{\partial (\text{vech}\Sigma)'} &= .5 \tau_j' \Sigma_j^{-1} (y_i - \mu_j) \left( w_i' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) - (\text{vec} \Sigma_j^{-1})' \right) \kappa_j \\ &= .5 \tau_j' \Sigma_j^{-1} (y_i - \mu_j) (w_i - \text{vec} \Sigma_j) (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j \\ \frac{\partial l_i(\beta)}{\partial \text{vech}\Sigma} \frac{\partial l_i(\beta)}{\partial (\text{vech}\Sigma)'} &= \frac{1}{4} \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) (w_i - \text{vec} \Sigma_j) (w_i - \text{vec} \Sigma_j)' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j. \end{aligned}$$

Averaging across  $n$  leads to:

$$\begin{aligned} B_{22} &= \sum_{j=1}^J \frac{n_j}{n} \tau_j' \Sigma_j^{-1} M_j \Sigma_j^{-1} \tau_j, \\ B_{21} &= \sum_{j=1}^J \frac{n_j}{2n} \tau_j' \Sigma_j^{-1} T_j (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j, \\ B_{11} &= \sum_{j=1}^J \frac{n_j}{4n} \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) V_j (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j, \end{aligned}$$

where  $M_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_i - \mu_j)(y_i - \mu_j)' = S_j + (\bar{y}_j - \mu_j)(\bar{y}_j - \mu_j)'$ ,  $T_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_i - \mu_j)(w_i - \text{vec} \Sigma_j)'$ , and  $V_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (w_i - \text{vec} \Sigma_j)(w_i - \text{vec} \Sigma_j)'$ . Define the following matrix:

$$\Gamma_{j,n} = \frac{1}{n_j} \sum_{i=1}^{n_j} t_i t_i' = \begin{pmatrix} (D_p' D_p)^{-1} \kappa_j' V_j \kappa_j (D_p' D_p)^{-1} & \kappa_j' T_j' \tau_j \\ \tau_j' T_j \kappa_j (D_p' D_p)^{-1} & \tau_j' M_j \tau_j \end{pmatrix},$$

where  $t_i = \left( (w_i - \text{vec} \Sigma_j)' \kappa_j (D_p' D_p)^{-1}, (y_i - \mu_j)' \tau_j \right)'$ . The expression in Equation 20 follows by noting that  $d_{\beta,i} = W_j t_i$ . The following identity has been used:  $\kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) = \kappa_j' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \kappa_j (D_p' D_p)^{-1} \kappa_j'$ . In the special case of complete data,

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^n t_i t_i' = \begin{pmatrix} D_p^+ V D_p^{+'} & D_p^+ T' \\ T D_p^{+'} & S + (\bar{y} - \mu)(\bar{y} - \mu)' \end{pmatrix},$$

where  $t_i = ((w_i - \text{vec} \Sigma)' D_p^{+'}, (y_i - \mu)')' = ((\text{vech}(y_i - \mu)(y_i - \mu)' - \text{vech} \Sigma)', (y_i - \mu)')'$ ,  $w_i = \text{vec}((y_i - \mu)(y_i - \mu)'),$   $T = \frac{1}{n} \sum_{i=1}^n (y_i - \mu)(w_i - \text{vec} \Sigma)'$ , and  $V = \frac{1}{n} \sum_{i=1}^n (w_i - \text{vec} \Sigma)(w_i - \text{vec} \Sigma)'$ .

### C. Equivalence of $B_{\beta}^0$ and $A_{\beta}^0$ with MCAR

As shown in part B of this Appendix, components of  $B_{\beta,n}^0$  depend on third- and fourth-order sample moments of the data. In the limit these moments simplify with normally distributed MCAR data. Under MCAR, data within each pattern are a random sample from the same population, so that  $\bar{y}_j \rightarrow \mu_j^0$  and  $S_j \rightarrow \Sigma_j^0$  for  $j = 1, \dots, J$ . This implies that  $M_j^0 = S_j + (\bar{y}_j - \mu_j^0)(\bar{y}_j - \mu_j^0)' \rightarrow \Sigma_j^0$  for all  $j$ . Normality implies that the third-order moments about the true mean tend to zero, so that  $T_j^0 = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_i - \mu_j^0) \left( (\text{vec}(y_i - \mu_j^0)(y_i - \mu_j^0)' - \text{vec} \Sigma_j^0) \right)' \rightarrow 0$  for all  $j$ . Under normality, fourth-order moments are

functions of second-order moments,  $V_j^0 = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_i^1 - \text{vec}\Sigma_j^0)(y_i^0 - \text{vec}\Sigma_j^0)' \rightarrow 2D_{p_j}D_{p_j}^+(\Sigma_j^0 \otimes \Sigma_j^0)$  for all  $j$ . Then,  $B_{\beta,n}^0 \rightarrow W^0$ . Under MCAR,  $\zeta_j = \mu^0$  for all  $j$ , so that  $A_\beta^0 = W^0$ , and therefore  $B_\beta^0 = A_\beta^0$ . This result also holds for complete data by viewing it as a special case of MCAR data with  $J = 1$ .

#### D. Equivalence of $\tilde{B}_\theta$ and $\hat{B}_\theta$ with complete data when $\hat{\mu} = \bar{y}$

We show that with complete data and when the mean structure is saturated,  $\tilde{B}_\theta = \hat{B}_\theta$ , where  $\tilde{B}_\theta = \hat{\Delta}' \hat{W}_c \tilde{\Gamma} \hat{W}_c \hat{\Delta}$ ,  $\hat{B}_\theta = \hat{\Delta}' \hat{W}_c \hat{\Gamma} \hat{W}_c \hat{\Delta}$ , where  $\tilde{\Gamma}$  is defined in Equation 24 and  $\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n \hat{t}_i \hat{t}_i'$ , with  $\hat{t}_i = ((\text{vech}(y_i - \hat{\mu})(y_i - \hat{\mu})' - \text{vec}\hat{\Sigma})', (y_i - \hat{\mu})')'$ . When the mean structure is saturated,  $\hat{\mu} = \tilde{\mu} = \bar{y}$ ,  $\hat{w}_i = \tilde{w}_i = \text{vec}((y_i - \bar{y})(y_i - \bar{y})')$ ,  $\hat{T} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\tilde{w}_i - \text{vec}\hat{\Sigma})' = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})\tilde{w}_i' = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\tilde{w}_i - \text{vec}\hat{S})' = \tilde{T}$ , and  $\hat{V} = \frac{1}{n} \sum_{i=1}^n (\tilde{w}_i - \text{vec}\hat{\Sigma})'(\tilde{w}_i - \text{vec}\hat{\Sigma})' = \frac{1}{n} \sum_{i=1}^n (\tilde{w}_i - \text{vec}\hat{S})'(\tilde{w}_i - \text{vec}\hat{S})' = \tilde{V} + (\text{vec}\hat{S} - \text{vec}\hat{\Sigma})(\text{vec}\hat{S} - \text{vec}\hat{\Sigma})'$ . Then,  $\hat{\Gamma} = \tilde{\Gamma} + \begin{pmatrix} D_p^+(\text{vec}\hat{S} - \text{vec}\hat{\Sigma})(\text{vec}\hat{S} - \text{vec}\hat{\Sigma})' D_p^{+'} & 0 \\ 0 & 0 \end{pmatrix}$ , and  $\hat{B}_\theta = \tilde{B}_\theta + .25\hat{\Delta}'_1 D_p' (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})(\text{vec}\hat{S} - \text{vec}\hat{\Sigma})(\text{vec}\hat{S} - \text{vec}\hat{\Sigma})' (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) D_p \hat{\Delta}_1$ , where  $\hat{\Delta}_1 = \frac{\partial \text{vech}\Sigma(\hat{\theta})}{\partial \theta'}$  is a  $p^* \times q$  submatrix of  $\hat{\Delta}$ . For properties of duplication matrices, see Magnus and Neudecker (1999). With complete data,  $d_{\beta,i} = (.5\text{vec}(\Sigma^{-1}(y_i - \mu)(y_i - \mu)' \Sigma^{-1} - \Sigma^{-1}) D_p, (y_i - \mu)' \Sigma^{-1})'$ , and  $d_{\beta,n} = (.5\text{vec}(\Sigma^{-1} S \Sigma^{-1} + \Sigma^{-1}(\bar{y} - \mu)(\bar{y} - \mu)' \Sigma^{-1} - \Sigma^{-1})' D_p, (\bar{y} - \mu)' \Sigma^{-1})'$ . At the solution,  $\hat{d}_{\theta,n} = \hat{d}_{\beta,n} \hat{\Delta} = 0$ . When  $\bar{y} = \hat{\mu}$ , the expression for  $\hat{d}_{\beta,n}$  simplifies. Setting it to zero yields, for the covariance structure part,  $(\text{vec}(\hat{\Sigma}^{-1} S \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}))' D_p \hat{\Delta}_1 = 0$ , which implies that  $(\text{vec}\hat{S} - \text{vec}\hat{\Sigma})' (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) D_p \hat{\Delta}_1 = 0$ . Therefore,  $\hat{B}_\theta = \tilde{B}_\theta$  whenever  $\hat{\mu} = \bar{y}$ .