PSYCHOMETRIKA—VOL. 84, NO. 3, 781–801
SEPTEMBER 2019
https://doi.org/10.1007/s11336-018-09655-0

CREATING MISSPECIFIED MODELS IN MOMENT STRUCTURE ANALYSIS

KEKE LAI
UNIVERSITY OF CALIFORNIA

To understand how SEM methods perform in practice where models always have misfit, simulation studies often involve incorrect models. To create a wrong model, traditionally one specifies a perfect model first and then removes some paths. This approach becomes difficult or even impossible to implement in moment structure analysis and fails to control the amounts of misfit separately and precisely for the mean and covariance parts. Most importantly, this approach assumes a perfect model exists and wrong models can eventually be made perfect, whereas in practice models are all implausible if taken literally and at best provide approximations of the real world. To improve the traditional approach, we propose a more realistic and flexible way to create model misfit for multiple group moment structure analysis. Given (a) the model $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\Sigma}(\cdot)$, (b) population model parameters $\boldsymbol{\theta}_0$, and (c) F_1 and F_2 specified by the researcher, our method creates $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ to simultaneously satisfy (a) $\boldsymbol{\theta}_0 = \arg\min F[\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \boldsymbol{\mu}(\cdot), \boldsymbol{\Sigma}(\cdot)]$, (b) the mean structure's misfit equals F_1, and (c) the covariance structure's misfit equals F_2.

Key words: Monte Carlo experiments, model misspecification, moment structure analysis, multiple group analysis.

Many methods in structural equation modeling (SEM) are developed by assuming the model is correct, but this assumption never holds in reality. The consequences of incorrect models are often unknown and need examinations on a case-by-case basis. Methods robust to model misspecifications exist (e.g., Arminger & Schoenberg, 1989; Gourieroux et al., 1984; Vuong, 1989; White, 1982), but their robustness is usually asymptotic. It is thus important to ask to what extent those methods can remain robust to model misspecifications given sample sizes typical of the behavioral sciences. Therefore, no matter whether a method requires the correct model assumption or not, to better understand the method's performance in practice, it is necessary to evaluate the method in simulation studies. Such simulation studies in the literature often proceed as follows. First, the researcher specifies the model $\boldsymbol{\Sigma}(\cdot)$ and its parameter values $\boldsymbol{\theta}_0$, obtaining the model-implied covariance matrix $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. Second, the researcher removes some parameters from the correct model $\boldsymbol{\Sigma}(\cdot)$ and creates an incorrect model $\boldsymbol{\Sigma}^*(\cdot)$. Third, random data are generated from the population with $\boldsymbol{\Sigma}_0$, whereas data analyses are based on the wrong model $\boldsymbol{\Sigma}^*(\cdot)$. Hereafter, we refer to this type of methods to create model misfit as the Type I approach.

The Type I perspective is easy to implement if the simulation scenario is simple, but becomes unwieldly as the problem becomes more complex, especially when the model involves mean structure. First, the form of $\boldsymbol{\Sigma}(\cdot)$ partly determines the range of options in the model parameters one can remove. For example, consider Model 1 depicted in Fig. 1. It is impossible to remove the factor loadings, and the only parameters available for removal is the factor covariances. To misspecify the mean structure, the only possibility is to remove the latent means (keep fixing the intercepts of the indicators to 0 at the same time). Second, because the removable model parameters are limited, the analysis model $\boldsymbol{\Sigma}^*(\cdot)$ one obtains is sometimes unrealistic. For example, using the Type I method, one might remove c_{F1F2} from Model 1, but in practice a researcher almost

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11336-018-09655-0) contains supplementary material, which is available to authorized users.

Correspondence should be made to Keke Lai, Psychological Sciences, University of California, Merced, CA 95343, USA. Email: KLai25@UCmerced.edu

© 2019 The Psychometric Society

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

never leaves F_1 and F_2 independent. Similarly, the model without a_{F1} is also unlikely to occur in practice. These two limitations become salient in some important applications of mean and covariance structure analysis, including growth curve models and multiple group comparisons. For growth curve models, cross-loadings do not exist, the values of (many or all) factor loadings are often fixed, and the means of latent factors need to be free parameters. All these characteristics make it difficult to remove any of the model parameters. Although there are clever ways to implement the Type I misspecification in this case, they are ad hoc and hold only for peculiar model forms or special parameter values. For multigroup analysis, it is even more difficult to find paths in a model to remove. The Type I approach usually comes in the form of incorrect equality constraints between groups, namely $\theta_j = \theta_k$ instead of $\theta_j = 0$.

The third limitation of the Type I method is the difficulty in controlling the amount of misfit and the location of misfit. For example, removing c_{F1F3} from Model 1 (see Fig. 1) leads to $F_{\mathrm{ML}} = .166$ (RMSEA = .073; CFI = .950), and this is the smallest F_{ML} one can obtain by removing a path. Without changing $\boldsymbol{\theta}_0$ values, it is impossible to create a wrong model with smaller F_{ML} values such as .15 or .10. Regarding the lack of control on the location of misfit, removing c_{F1F2} from Model 1 will not introduce any misfit on the covariances among X_1 to X_3, on the covariances among X_4 to X_6, or on the means of X_1 to X_9. If one removes a_{F1} from Model 1, then the model-implied means of X_1 to X_3 will always be zero. The lack of control on the misfit location not only restricts the model forms and parameter values available for simulation design, it also impairs the verisimilitude of a simulation study, because in practice the discrepancy between model and data usually permeates all the elements of $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ rather than takes place on few elements only. Without a realistic design, a simulation study is not helpful for guiding research in practice. The fourth limitation is the difficulty in separating the misfit in mean structure from that in the covariance structure. The factor loadings, as well as structural coefficients if any, play a role in both $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and thus any changes in the

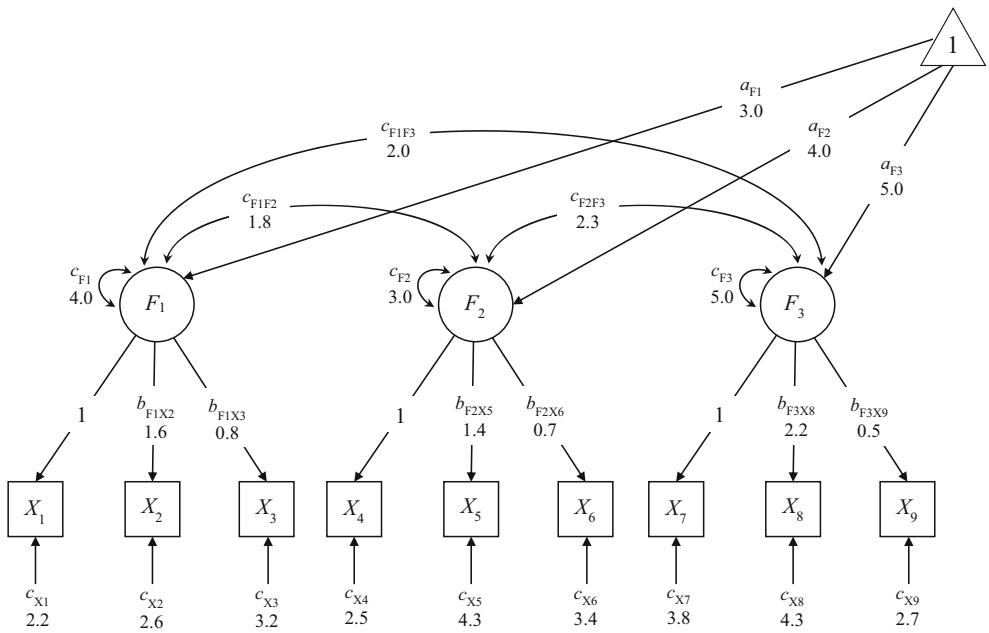


FIGURE 1.
Path diagram of Model 1 used in introduction and Demonstration 1. Values are the population model parameters specified by the researcher. Parameters originating from the triangle "1" have the label "a." Single-headed arrows from one variable to another have the label "b." Double-headed arrows have the label "c".

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

factor loadings affect both the mean and covariance parts. On the other hand, it is important to understand the different roles of mean structure and covariance structure in model misfit, and methods are desirable that can change the misfit in one part while maintain the misfit in the other part constant. For the Type I misspecification, unless designing the simulation in an unrealistic and peculiar way, it is impossible to separately study the misfit in $\boldsymbol{\mu}(\theta)$ and $\boldsymbol{\Sigma}(\theta)$.

The fifth and most fundamental limitation of the Type I method is how it conceptualizes mistakes in models. Statistical models are at best over-simplifications of reality. They are literally implausible because they never describe the real process that gives rise to the phenomenon, but are merely artificial and convenient approximations of the real process. The notion "all models are wrong" (Box, 1979a) has been reiterated over the history of psychometrics by influential scholars such as Thurstone (1930), Tukey (1961), Box (1979b), Meehl (1990), Cudeck and Henly (1991), Thissen (2001), and MacCallum (2003). Because all models are wrong, in practice there never exists a θ such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$. The literal implausibility of models implies that there are always inherently non-fixable mistakes in modeling; no matter how hard one improves the model by correcting the fixable mistakes, there will always be discrepancy between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\theta)$ because $\boldsymbol{\Sigma}(\cdot)$ is not the real process that gives rise to $\boldsymbol{\Sigma}$. However, the Type I approach begins by assuming there is a perfect model, and in this framework, a wrong model can eventually become perfect if one keeps adding the missing paths back to the model. This perspective imitates only the fixable mistakes, but ignores the non-fixable mistakes in modeling. A more realistic approach is thus to imitate both the fixable and non-fixable mistakes, and we refer to this approach as the Type II perspective. Given a $\boldsymbol{\Sigma}$ created from the Type II perspective, even when a model has all the parameters it should have, there is still discrepancy between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\theta)$.

Important applications of the Type II perspective include MacCallum and Tucker (1991) and MacCallum and colleagues (1999, 2001; see also MacCallum 2003; Tucker et al., 1969) in the context of factor analysis. In particular, data are generated based on $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\text{model}} + \boldsymbol{\Sigma}_{\text{minor}}$, where $\boldsymbol{\Sigma}_{\text{minor}}$ is the covariance matrix of a large number (50 in MacCallum & Tucker, 1991) of minor latent factors. The rationale is that manifest variables X_j and X_k are correlated because they are affected by numerous common but unknown sources in reality, but a model explains σ_{jk} with only few common (major) factors. Misfit comes in because the model omits many other sources of influence, referred to as the minor factors. The best effort in modeling will lead to $\boldsymbol{\Sigma}_{\text{model}}$, but the misfit due to $\boldsymbol{\Sigma}_{\text{minor}}$ is not fixable because the minor factors are too many to be included in the model. This application of the Type II perspective is reasonable but difficult to extend outside of factor analysis. Another application of the Type II perspective is Cudeck and Browne (1992), where the researcher specifies (a) the model $\boldsymbol{\Sigma}(\cdot)$, (b) population model parameter values θ_0, and (c) desired fit function value c ($c > 0$) based on F_{ML} or F_{OLS}. Then, the data covariance matrix $\boldsymbol{\Sigma}^*$ is created such that (a) $\theta_0 = \arg\min F[\boldsymbol{\Sigma}^*, \boldsymbol{\Sigma}(\cdot)]$ and (b) $F[\boldsymbol{\Sigma}^*, \boldsymbol{\Sigma}(\theta_0)] = c$. Therefore, the covariance structure holds only approximately in the population. In simulations, random data are generated from $\boldsymbol{\Sigma}^*$ and the analysis model remains $\boldsymbol{\Sigma}(\cdot)$. Compared to MacCallum and Tucker's method, Cudeck and Browne's method is applicable to covariance structure analysis in general and can control the amount of misfit more easily and precisely. Both methods are free from the five limitations of the Type I perspective discussed above. However, both methods are limited to single-group covariance structure analysis.

In this paper, we extend Cudeck and Browne's (1992) method to multiple group mean and covariance structure analysis. We propose a method to create $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ so that, given $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\Sigma}(\cdot)$, θ_0, F_{mean}, and F_{cov} specified by the researcher, it simultaneously satisfies (a) $\theta_0 = \arg\min F[\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \boldsymbol{\mu}(\cdot), \boldsymbol{\Sigma}(\cdot)]$, (b) the mean structure's misfit equals F_{mean}, and (c) the covariance structure's misfit equals F_{cov}. Our method is applicable to any simulations involving moment structure analysis, such as growth curve modeling, measurement invariance, and mixture modeling. In the rest of the paper, we first describe the basic procedure in the context of single-

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

group mean and covariance structure analysis, followed by an empirical demonstration. Then, we further extend our method to multigroup moment analysis, followed by a second demonstration.

1. The Basic Procedure

In this section, we focus on single-group mean and structure analysis and will extend to multigroup analysis later. Let the $p \times 1$ vector $\mathbf{x}$ denote the manifest variables, $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\Sigma}(\cdot)$ the model of interest, and the $q \times 1$ vector $\boldsymbol{\theta}_0$ the population model parameter values specified by the researcher. The goal is to find some noise $\mathbf{t}$ and $\mathbf{E}$ to imitate the lack of fit in reality and construct $\boldsymbol{\mu}^* = \boldsymbol{\mu}(\boldsymbol{\theta}_0) + \mathbf{t}$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) + \mathbf{E}$. Then, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ are considered the population moments of $\mathbf{x}$ and used to generate random data for simulation studies. We hope when fitting the model to $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$, the discrepancy function achieves its minimum at the specified parameter values $\boldsymbol{\theta}_0$, and the amount of misfit equals some desired values. Here, we select the normal theory maximum likelihood (ML) fit function as the measure of misfit:

$$F_{\text{ML}} = [\boldsymbol{\mu}^* - \boldsymbol{\mu}(\boldsymbol{\theta})]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\boldsymbol{\mu}^* - \boldsymbol{\mu}(\boldsymbol{\theta})] + \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\boldsymbol{\Sigma}^*| + \text{tr}[\boldsymbol{\Sigma}^* \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] - p. \quad (1)$$

In constructing $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ (or equivalently $\mathbf{t}$ and $\mathbf{E}$), we seek to satisfy the following three requirements simultaneously:

$$\boldsymbol{\theta}_0 = \arg\min F[\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})]; \quad (2)$$

$$[\boldsymbol{\mu}^* - \boldsymbol{\mu}(\boldsymbol{\theta}_0)]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)[\boldsymbol{\mu}^* - \boldsymbol{\mu}(\boldsymbol{\theta}_0)] = F_{\text{mean}}; \quad (3)$$

$$\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)| - \ln|\boldsymbol{\Sigma}^*| + \text{tr}[\boldsymbol{\Sigma}^* \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)] - p = F_{\text{cov}}; \quad (4)$$

where $\boldsymbol{\theta}_0$, F_{mean}, and F_{cov} are constants specified by the researcher. The desired F_{ML} value for the whole model is then $F_{\text{mean}} + F_{\text{cov}}$. The necessary condition for Eq. (2) is (see "Appendix"):

$$\dot{F}(\boldsymbol{\theta}_0) = 2\dot{\boldsymbol{\mu}}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \cdot \mathbf{t} + \dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)] \cdot \text{vec}(\mathbf{t}\mathbf{t}' + \mathbf{E}) = \mathbf{0}, \quad (5)$$

where $\dot{F}(\boldsymbol{\theta}_0) = \partial F / \partial \boldsymbol{\theta} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ (a $q \times 1$ vector), $\dot{\boldsymbol{\mu}}(\boldsymbol{\theta}_0) = \partial \boldsymbol{\mu}(\boldsymbol{\theta})' / \partial \boldsymbol{\theta} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ (a $q \times p$ matrix), $\dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) = \partial \text{vec}'[\boldsymbol{\Sigma}(\boldsymbol{\theta})] / \partial \boldsymbol{\theta} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ (a $q \times p^2$ matrix), $\text{vec}(\cdot)$ stack the columns of a matrix into a vector, and $\otimes$ is the Kronecker product. We define the derivatives in this way so that the j-th row in $\dot{F}(\boldsymbol{\theta})$ is the derivative of $F(\boldsymbol{\theta})$ with respect to θ_j. That is, we define $\dot{F}(\boldsymbol{\theta})$ as the transpose of the Jacobian matrix of $F(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, because doing so facilitates the following exposition of our method. We define $\dot{\boldsymbol{\mu}}(\boldsymbol{\theta})$ and $\dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$ in the same manner.

Note some rows in $\dot{\boldsymbol{\mu}}(\boldsymbol{\theta}_0)$ and $\dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ are zeros because some model parameters do not play a role in the mean or the covariance structure. More specifically, parameters for the mean or intercept of a variable (i.e., coefficients originating from the triangle "1" in a path diagram) affect $\boldsymbol{\mu}(\boldsymbol{\theta})$ but not $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and we refer to them as $\boldsymbol{\theta}_a$ or Type-a parameters. Regression coefficients from one variable to another (e.g., factor loadings, structural coefficients) affect both $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and we refer to them as $\boldsymbol{\theta}_b$ or Type-b parameters. Variances and covariances of variables affect $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ only but not $\boldsymbol{\mu}(\boldsymbol{\theta})$, and we refer to them as $\boldsymbol{\theta}_c$ or Type-c parameters. An example for such notations is in Fig. 1. Throughout the paper, we use the subscripts "a," "b," and "c" to denote the subsets corresponding to the Type-a, b, and c parameters. For example, q_a, q_b, and q_c denote the number of elements in $\boldsymbol{\theta}_a$, $\boldsymbol{\theta}_b$, and $\boldsymbol{\theta}_c$. Next, we study the specific forms of $\dot{\boldsymbol{\mu}}(\boldsymbol{\theta}_0)$ and $\dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$. Without loss of generality, let us assume the model parameters are arranged in such an order that $\boldsymbol{\theta} = (\boldsymbol{\theta}_a', \boldsymbol{\theta}_b', \boldsymbol{\theta}_c')'$. For $\dot{\boldsymbol{\mu}}(\boldsymbol{\theta}_0)$, the first $(q_a + q_b)$ rows are nonzero, whereas the last q_c rows are all zeros. For $\dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$, the first q_a rows are all zeros, but the last $(q_b + q_c)$ rows are nonzero. In this paper, we focus on the case where $\text{rank}[\dot{\boldsymbol{\mu}}(\boldsymbol{\theta}_0)_a] = q_a$ and $\text{rank}[\dot{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)_{b,c}] = q_b + q_c$, meaning

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

the first q_a rows in $\dot{\boldsymbol{\mu}}(\theta_0)$ are linearly independent, and so are the last $(q_b + q_c)$ rows in $\dot{\boldsymbol{\Sigma}}(\theta_0)$. This is generally true for SEM analyses in practice.

Using the shorthand $\boldsymbol{\Omega} = 2\dot{\boldsymbol{\mu}}(\theta_0)\boldsymbol{\Sigma}_0^{-1}$ and $\boldsymbol{\Delta} = \dot{\boldsymbol{\Sigma}}(\theta_0)(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})$ (recall $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\theta_0)$), we rewrite Eq. (5) as

$$\boldsymbol{\Omega} \cdot \mathbf{t} + \boldsymbol{\Delta} \cdot \text{vec}(\mathbf{t}\mathbf{t}') + \boldsymbol{\Delta} \cdot \text{vec}\mathbf{E} = \mathbf{0}. \tag{6}$$

Because the last q_c rows in $\dot{\boldsymbol{\mu}}(\theta_0)$ are zeros, the last q_c rows in $\boldsymbol{\Omega}$ are also zeros. Similarly, the first q_a rows in $\boldsymbol{\Delta}$ are also zeros. Accordingly, Eq. (6) has the form

$$\begin{bmatrix} \boldsymbol{\Omega}_a \\ \boldsymbol{\Omega}_b \\ \mathbf{O} \end{bmatrix} \cdot \mathbf{t} + \begin{bmatrix} \mathbf{O} \\ \boldsymbol{\Delta}_b \\ \boldsymbol{\Delta}_c \end{bmatrix} \cdot \text{vec}(\mathbf{t}\mathbf{t}') + \begin{bmatrix} \mathbf{O} \\ \boldsymbol{\Delta}_b \\ \boldsymbol{\Delta}_c \end{bmatrix} \cdot \text{vec}\mathbf{E} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{7}$$

where $\mathbf{O}$ is a matrix of zeros. The strategy for finding $\mathbf{t}$ and $\mathbf{E}$ is as follows. First, we find an initial solution to

$$\boldsymbol{\Omega}_a \cdot \mathbf{t} = \mathbf{0}, \tag{8}$$

and denote this solution as $\tilde{\mathbf{t}}$. Second, we find a scalar π for $\tilde{\mathbf{t}}$ such that $(\pi\tilde{\mathbf{t}})'\boldsymbol{\Sigma}_0^{-1}(\pi\tilde{\mathbf{t}}) = F_{\text{mean}}$, and thus, $\mathbf{t} = \pi\tilde{\mathbf{t}}$ satisfies Eqs. (3) and (8). Third, given $\mathbf{t}$, an identity about $\mathbf{E}$ can be obtained based on Eqs. (6) and (7):

$$\boldsymbol{\Delta} \cdot \text{vec}\mathbf{E} = -\boldsymbol{\Omega} \cdot \mathbf{t} - \boldsymbol{\Delta} \cdot \text{vec}(\mathbf{t}\mathbf{t}') \equiv \boldsymbol{\eta}. \tag{9}$$

Fourth, we substitute Eq. (9) into Eq. (4) and solve for $\mathbf{E}$.

Now let us consider how to implement these four steps. We do not consider the trivial case where the intercept of an endogenous manifest variable is freely estimated. This is because, say if X_j is such a variable and its intercept a_j is free, then the model can always fit the mean of X_j perfectly. Thus, there is no nonzero solution for t_j in $\mathbf{t}$. In a later section, when we extend the procedure to multigroup analyses, due to between-group constraints, it becomes possible to create misfit on the intercepts of manifest variables. Returning to Eq. (6), given that $\text{rank}[\dot{\boldsymbol{\mu}}(\theta_0)_a] = q_a$ and $\boldsymbol{\Sigma}_0^{-1}$ has full rank p, it follows $\text{rank}(\boldsymbol{\Omega}_a) = q_a$. Because $\text{rank}([\boldsymbol{\Omega}_a \mid \mathbf{0}]) = \text{rank}(\boldsymbol{\Omega}_a)$, $\boldsymbol{\Omega}_a$ is a $q_a \times p$ matrix, and $\text{rank}(\boldsymbol{\Omega}_a) < p$, it follows that Eq. (8) has infinitely many solutions. The solutions are of the form

$$\mathbf{t} = (\mathbf{I} - \boldsymbol{\Omega}_a^{-}\boldsymbol{\Omega}_a)\mathbf{y}_t, \tag{10}$$

where $\boldsymbol{\Omega}_a^{-}$ is a generalized inverse of $\boldsymbol{\Omega}_a$, and $\mathbf{y}_t$ can be any $p \times 1$ vector. For convenience, one can just use the Moore–Penrose inverse for the generalized inverse. Given a randomly generated $\tilde{\mathbf{y}}_t$, an initial solution to Eq. (8) is thus $\tilde{\mathbf{t}} = (\mathbf{I} - \boldsymbol{\Omega}_a^{-}\boldsymbol{\Omega}_a)\tilde{\mathbf{y}}_t$. Next, we calculate $\tilde{\mathbf{t}}'\boldsymbol{\Sigma}_0^{-1}\tilde{\mathbf{t}} \equiv \tilde{F}_{\text{mean}}$. Defining $\mathbf{t}$ as $\mathbf{t} = \tilde{\mathbf{t}}\sqrt{F_{\text{mean}}/\tilde{F}_{\text{mean}}}$ will satisfy Eqs. (8) and (3) simultaneously. Given $\mathbf{t}$, the value of $\text{vec}(\mathbf{t}\mathbf{t}')$ is also determined, and thus, the only unknown in Eq. (7) is $\mathbf{E}$.

Next, we construct $\mathbf{E}$. The first q_a rows in Eq. (7) always hold regardless of the choice of $\mathbf{E}$, and thus, we just need to focus on the nonzero rows concerning θ_b and θ_c. To ensure the solution $\mathbf{E}$ is symmetric, we rewrite Eq. (9) as

$$\boldsymbol{\Delta}\mathbf{D} \cdot \text{vec}h(\mathbf{E}) = \mathbf{B} \cdot \text{vec}h(\mathbf{E}) = \boldsymbol{\eta}, \tag{11}$$

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

where vec$h(\cdot)$ stacks the lower triangular elements of a matrix into a vector, $\mathbf{B} = \boldsymbol{\Delta}\mathbf{D}$, and $\mathbf{D}$ is the duplication matrix such that $\text{vec}(\mathbf{E}) = \mathbf{D} \cdot \text{vec}h(\mathbf{E})$. Given that $\text{rank}[\dot{\boldsymbol{\Sigma}}(\theta_0)] = q_b + q_c$, $(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})$ has full rank p^2, and $\mathbf{D}$ has full column rank p^* (where $p^* = p(p+1)/2$), we have $\text{rank}(\boldsymbol{\Delta}) = q_b + q_c$ and $\text{rank}(\mathbf{B}) \leq q_b + q_c$. The first q_a rows of $\mathbf{B}$ are all zeros, and if $\text{rank}(\mathbf{B}) = q_b + q_c$, then the rest $(q_b + q_c)$ rows are linearly independent. Because the first q_a elements of $\boldsymbol{\eta}$ are all zeros, it follows that $\text{rank}([\mathbf{B} \mid \boldsymbol{\eta}]) = \text{rank}(\mathbf{B})$ and Eq. (11) is a consistent linear system. Given that $\mathbf{B}$ is a $q \times p^2$ matrix and $\text{rank}(\mathbf{B}) < p^2$, Eq. (11) has infinitely many solutions. If $\text{rank}(\mathbf{B}) < q_b + q_c$, $\text{rank}([\mathbf{B} \mid \boldsymbol{\eta}])$ may not equal $\text{rank}(\mathbf{B})$ and thus sometimes $\mathbf{B} \cdot \text{vec}h(\mathbf{E}) = \boldsymbol{\eta}$ is not a consistent system. In that case, one can simply find a different $\boldsymbol{\eta}$ to satisfy $\text{rank}([\mathbf{B} \mid \boldsymbol{\eta}]) = \text{rank}(\mathbf{B})$ by generating a new $\mathbf{t}$ from Eq. (10). Therefore, Eq. (11) can be guaranteed to have infinitely many solutions, and their general form is

$$\text{vec}h(\mathbf{E}) = \mathbf{B}^{-}\boldsymbol{\eta} + (\mathbf{I} - \mathbf{B}^{-}\mathbf{B})\mathbf{y}_E, \tag{12}$$

where $\mathbf{y}_E$ can be any $p^* \times 1$ vector. Next, we rewrite $\mathbf{E}$ in Eq. (4) in terms of $\mathbf{y}_E$ and solve the equation for $\mathbf{y}_E$. More specifically, we define a new function:

$$\phi = \ln|\boldsymbol{\Sigma}_0| - \ln|\boldsymbol{\Sigma}_0 + \mathbf{E}| + \text{tr}(\mathbf{I} + \mathbf{E}\boldsymbol{\Sigma}_0^{-1}) - p - F_{\text{cov}}, \tag{13}$$

and then find a root for $\phi = 0$ in terms of $\mathbf{y}_E$ using the Newton method. In particular, the Jacobian of ϕ with respect to $\mathbf{y}_E$ is $\mathbf{J}_\phi(\mathbf{y}_E) = \partial\phi/\partial\mathbf{y}_E' = (\partial\phi/\partial\mathbf{E})(\partial\mathbf{E}/\partial\mathbf{y}_E')$, where

$$\frac{\partial\phi}{\partial\mathbf{E}} = \text{vec}'[\boldsymbol{\Sigma}_0^{-1} - (\boldsymbol{\Sigma}_0 + \mathbf{E})^{-1}] \cdot \frac{\partial\mathbf{E}}{\partial\mathbf{E}}; \tag{14}$$

$$\frac{\partial\mathbf{E}}{\partial\mathbf{y}_E} = \frac{\partial[\mathbf{D}\text{vec}h(\mathbf{E})]}{\partial\mathbf{y}_E} = \mathbf{D}(\mathbf{I} - \mathbf{B}^{-}\mathbf{B}). \tag{15}$$

Note $\partial\mathbf{E}/\partial\mathbf{E}$ does not equal $\mathbf{I}$ but another constant matrix instead, because each off-diagonal element of $\mathbf{E}$ appears twice in $\mathbf{E}$, leading to $\partial e_{ij}/\partial e_{ji} = 1$ even if $i \neq j$. Then we can find a root for $\phi(\mathbf{y}_E) = 0$ using an iterative process. The update from step k to step $(k+1)$ is:

$$\mathbf{y}_E^{(k+1)} = \mathbf{y}_E^{(k)} - \mathbf{J}_\phi^{-1}[\mathbf{y}_E^{(k)}] \cdot \phi[\mathbf{y}_E^{(k)}]. \tag{16}$$

At convergence, the $\mathbf{y}_E^{(k+1)}$ value is a root to $\phi(\mathbf{y}_E) = 0$. Then, we can construct $\mathbf{E}$ based on Eq. (12). Now that $\mathbf{t}$ and $\mathbf{E}$ are available, it is straightforward to compute $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$.

The $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ obtained with the above procedure will satisfy Eqs. (3) to (5), but any θ satisfying Eq. (5) is only a stationary point of $F_{\text{ML}}[\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta)]$. Equation (2) requires θ_0 to be the global minimizer of $F_{\text{ML}}[\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta)]$, not just a stationary point. Therefore, it needs to continue to show that θ_0 is indeed the global minimizer. To prove this, we borrow from the arguments that established the consistency of the ML point estimator (Kano, 1986; Shapiro, 1984; see also Chun & Shapiro, 2010). More specifically, let $\mathbf{m}(\theta) = [\boldsymbol{\mu}(\theta)', \text{vec}'[\boldsymbol{\Sigma}(\theta)]]'$ denote the model-implied moments, $\tilde{\mathbf{m}} = [\tilde{\boldsymbol{\mu}}', \text{vec}'(\tilde{\boldsymbol{\Sigma}})]'$ the moments of the data, and $\mathbf{m}_0 = [\boldsymbol{\mu}_0', \text{vec}'(\boldsymbol{\Sigma}_0)]'$. Because $\mathbf{m}_0 = \mathbf{m}(\theta_0)$, θ_0 is always a global minimizer of $F_{\text{ML}}[\mathbf{m}_0, \mathbf{m}(\theta)]$. Moreover, if the model is identified at θ_0, then θ_0 is the unique minimizer (Kano, 1986; Shapiro, 1984). Let $\tilde{\theta} = \underset{\theta \in \boldsymbol{\Theta}}{\text{argmin}} F_{\text{ML}}[\tilde{\mathbf{m}}, \mathbf{m}(\theta)]$, and $\tilde{\theta}$ is consistent if, for all $\tilde{\mathbf{m}}$ sufficiently close to $\mathbf{m}_0$, $\tilde{\theta} \to \theta_0$ as $\tilde{\mathbf{m}} \to \mathbf{m}_0$. Under mild regularity conditions, the consistency of $\tilde{\mathbf{m}}$ holds if the model is identified at θ_0 and the set $\boldsymbol{\Theta}$ is compact (Kano, 1986; Shapiro, 1984).

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

Proposition 1. *Suppose the consistency of* $\tilde{\Theta}$ *holds for the researcher's model, and the Hessian* $\partial^2 F[\mathbf{m}, \mathbf{m}(\Theta)]/\partial\Theta\partial\Theta$ *evaluated at* $\mathbf{m} = \mathbf{m}_0$ *and* $\Theta = \Theta_0$ *is positive definite (denoted as* $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\Theta_0)]$*). Then, there exists a neighborhood* $\mathcal{U}$ *of* $\mathbf{m}_0$*, such that for any* $\mathbf{m}_k \in \mathcal{U}$ *satisfying* $\dot{F}[\mathbf{m}_k, \mathbf{m}(\Theta_0)] = \mathbf{0}$ *(i.e.,* Θ_0 *is a stationary point when fitting the model to* $\mathbf{m}_k$*), it follows that* Θ_0 *is the* unique *minimizer of* $F[\mathbf{m}_k, \mathbf{m}(\Theta)]$ *over* $\Theta \in \boldsymbol{\Theta}$.

Proof. We argue by a contradiction. First, suppose the assertion is false. Suppose there is a sequence $\{\mathbf{m}_k\} \to \mathbf{m}_0$, such that $\dot{F}[\mathbf{m}_k, \mathbf{m}(\Theta_0)] = \mathbf{0}$ but $\Theta_0 \neq \underset{\Theta\in\boldsymbol{\Theta}}{\operatorname{argmin}} F[\mathbf{m}_k, \mathbf{m}(\Theta)]$. Let this minimizer be denoted as $\tilde{\Theta}_k = \underset{\Theta\in\boldsymbol{\Theta}}{\operatorname{argmin}} F[\mathbf{m}_k, \mathbf{m}(\Theta)]$. Because $\mathbf{m}_k \to \mathbf{m}_0$ as $k \to \infty$, the consistency of $\tilde{\Theta}_{\mathrm{ML}}$ implies that $\tilde{\Theta}_k \to \Theta_0$ as $k \to \infty$. In addition, because $\tilde{\Theta}_k$ is the minimizer of $F[\mathbf{m}_k, \mathbf{m}(\Theta)]$ but Θ_0 is not, it follows that $F[\mathbf{m}_k, \mathbf{m}(\Theta_0)] > F[\mathbf{m}_k, \mathbf{m}(\tilde{\Theta}_k)]$.

On the other hand, because the Hessian $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\Theta_0)]$ is positive definite, by continuity arguments the Hessian $\ddot{F}[\mathbf{m}_k, \mathbf{m}(\Theta_0)]$ is also positive definite if k is large enough (i.e., $\mathbf{m}_k$ is close enough to $\mathbf{m}_0$; see "Appendix" for details). Because $\dot{F}[\mathbf{m}_k, \mathbf{m}(\Theta_0)] = \mathbf{0}$ and $\ddot{F}[\mathbf{m}_k, \mathbf{m}(\Theta_0)]$ is positive definite, Θ_0 is a strict local minimizer of $F[\mathbf{m}_k, \mathbf{m}(\Theta)]$. Accordingly, there is a neighborhood $\mathcal{V}$ of Θ_0, such that for all the $\Theta_v \in \mathcal{V}$, $F[\mathbf{m}_k, \mathbf{m}(\Theta_0)] < F[\mathbf{m}_k, \mathbf{m}(\Theta_v)]$. Note the neighborhood $\mathcal{V}$ can be taken independent of k. Moreover, because $\tilde{\Theta}_k \to \Theta_0$, if k is large enough, we have $\tilde{\Theta}_k \in \mathcal{V}$ and thus $F[\mathbf{m}_k, \mathbf{m}(\Theta_0)] < F[\mathbf{m}_k, \mathbf{m}(\tilde{\Theta}_k)]$. But this result contradicts with $F[\mathbf{m}_k, \mathbf{m}(\Theta_0)] > F[\mathbf{m}_k, \mathbf{m}(\tilde{\Theta}_k)]$, which is obtained by assuming $\Theta_0 \neq \underset{\Theta\in\boldsymbol{\Theta}}{\operatorname{argmin}} F[\mathbf{m}_k, \mathbf{m}(\Theta)]$. Hence, the proof is complete. □

Proposition 1 implies that if $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\Theta_0)]$ is positive definite and the misfit is not too large (i.e., $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ do not depart from $\boldsymbol{\mu}(\Theta_0)$ and $\boldsymbol{\Sigma}(\Theta_0)$ by too much), Eq. (2) will hold as long as Eq. (5) holds. In the context of covariance structure analysis (i.e., fixing $\mathbf{t} = \mathbf{0}$ and $F_{\mathrm{mean}} = 0$), Cudeck and Browne (1992) and Chun and Shapiro (2010) proved that the stationary point Θ_0 is the global minimizer if $\mathbf{E}$ is not too large. Chun and Shapiro (2010) also showed that the misfit can be in fact quite large before Θ_0 stops being the global minimizer. Based on Proposition 1, therefore, the $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ constructed using our proposed procedure can satisfy Eqs. (2) through (4) simultaneously. This concludes the basic procedure.

2. Demonstration 1

In this section we use the proposed method to create several sets of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ for Model 1 in Fig. 1. The model form and Θ_0 values are presented in Fig. 1. For the amount of misfit, we choose values that render the misfit quite serious and greater than values of interest to a researcher when designing simulation studies or analyzing real data. We use huge F_{ML} values because, based on the above discussion of Proposition 1, for the proposed method to work well, it needs the misfit being not too large. Therefore, if our method works well given huge F_{mean} and F_{cov} in this demonstration, it should perform even better given smaller F_{mean} and F_{cov} values in practice. In choosing F_{mean} and F_{cov} for this demonstration, we draw analogy with the definition of RMSEA $\varepsilon = \sqrt{F/df}$, and set $F_{\mathrm{mean}} = (.15)^2 df_{\mathrm{mean}} = .135$ and $F_{\mathrm{cov}} = (.20)^2 df_{\mathrm{cov}} = .96$. Accordingly, $F_{all} = 1.095$ and the model's RMSEA is .191, indicating serious misfit. Next, we used the proposed method to create four sets of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ randomly (as $\mathbf{y}_t$ and $\mathbf{y}_E$ can be chosen randomly, see Eqs. (10) and (12)). We then fitted Model 1 to $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ using the "lavaan" package (Rosseel, 2012) in R (R Core Team, 2017) and recorded Θ_{fit}, $F_{\mathrm{mean}}^{(\mathrm{fit})}$, and $F_{\mathrm{cov}}^{(\mathrm{fit})}$, where Θ_{fit} is the fitted model parameter.

Table 1 presents the four sets of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ and model estimation results. Comparing $\boldsymbol{\mu}(\Theta_0)$ and $\boldsymbol{\Sigma}(\Theta_0)$ to $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$, we see the model can largely reproduce the data and the misfit spreads

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

TABLE 1.
Population model-implied moments and four generated moments with specified misfit and parameter values in Demonstration 1.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Mean	Fit results
Model-implied moments based on θ_0											
X_1	**6.200**	.717	.535	.308	.317	.229	.271	.331	.202	3.00	F_{Cov} Diff
X_2	6.400	**12.840**	.595	.343	.353	.255	.301	.368	.225	4.80	F_{Mean} Diff
X_3	3.200	5.120	**5.760**	.256	.263	.190	.225	.275	.168	2.40	RMSEA
X_4	1.800	2.880	1.440	**5.500**	.561	.406	.331	.404	.247	4.00	CFI
X_5	2.520	4.032	2.016	4.200	**10.180**	.418	.340	.416	.254	5.60	SRMR
X_6	1.260	2.016	1.008	2.100	2.940	**4.870**	.246	.301	.184	2.80	
X_7	2.000	3.200	1.600	2.300	3.220	1.610	**8.800**	.695	.424	5.00	
X_8	4.400	7.040	3.520	5.060	7.084	3.542	11.000	**28.500**	.518	11.00	
X_9	1.000	1.600	0.800	1.150	1.610	0.805	2.500	5.500	**3.950**	2.50	
Generated moments set 1											
X_1	**8.050**	.801	.687	.272	.426	.433	.318	.384	.329	2.67	8.356E−09
X_2	7.362	**10.496**	.508	.187	.339	.351	.269	.330	.283	5.04	−5.551E−17
X_3	4.668	3.936	**5.730**	.213	.324	.331	.243	.291	.249	2.41	.191
X_4	1.508	1.184	0.995	**3.819**	.432	.531	.266	.311	.265	4.20	.759
X_5	3.979	3.623	2.556	2.781	**10.862**	.715	.371	.427	.361	5.54	.155
X_6	3.125	2.894	2.012	2.640	5.996	**6.466**	.369	.426	.360	2.50	
X_7	2.523	2.432	1.624	1.451	3.422	2.623	**7.817**	.644	.554	5.10	
X_8	5.756	5.634	3.676	3.204	7.422	5.717	9.508	**27.854**	.694	11.03	
X_9	2.084	2.051	1.335	1.159	2.663	2.049	3.463	8.186	**4.998**	2.28	
Generated moments set 2											
X_1	**4.897**	.666	.560	.136	.381	.348	.293	.324	.282	3.21	8.738E−09
X_2	5.384	**13.336**	.724	.155	.430	.395	.324	.348	.304	4.75	5.551E−17
X_3	3.258	6.955	**6.913**	.128	.356	.327	.275	.294	.257	2.15	.191
X_4	0.535	1.010	0.598	**3.169**	.461	.438	.241	.286	.251	4.28	.750
X_5	2.930	5.460	3.253	2.852	**12.090**	.748	.406	.459	.394	5.43	.187
X_6	1.935	3.628	2.162	1.959	6.544	**6.324**	.348	.384	.332	2.53	
X_7	1.775	3.241	1.982	1.174	3.861	2.392	**7.491**	.627	.584	5.13	
X_8	3.777	6.689	4.076	2.678	8.405	5.088	9.033	**27.750**	.743	11.03	
X_9	1.431	2.550	1.551	1.027	3.146	1.918	3.663	8.975	**5.262**	2.22	
Generated moments set 3											
X_1	**4.927**	.622	.714	.287	.391	.338	.302	.366	.306	3.21	8.641E-09
X_2	4.843	**12.295**	.818	.279	.351	.285	.332	.344	.318	4.86	−2.776E−17
X_3	4.469	8.094	**7.955**	.248	.315	.269	.245	.274	.244	1.89	.191
X_4	1.327	2.037	1.459	**4.336**	.532	.276	.339	.378	.368	4.14	.764
X_5	2.940	4.164	3.005	3.748	**11.467**	.632	.361	.383	.404	5.48	.131
X_6	1.743	2.320	1.762	1.336	4.965	**5.388**	.300	.351	.324	2.71	
X_7	2.043	3.547	2.105	2.152	3.726	2.125	**9.302**	.695	.641	4.95	
X_8	4.160	6.174	3.968	4.029	6.654	4.172	10.862	**26.274**	.655	11.10	
X_9	1.510	2.478	1.530	1.704	3.044	1.673	4.343	7.461	**4.940**	2.29	
Generated moments set 4											
X_1	**5.789**	.643	.775	.342	.328	.286	.310	.355	.275	3.07	8.617E−09
X_2	5.155	**11.109**	.815	.305	.308	.248	.287	.350	.287	4.98	1.110E−16
X_3	5.316	7.746	**8.130**	.363	.350	.293	.246	.291	.289	1.84	.191
X_4	2.056	2.535	2.583	**6.230**	.641	.391	.342	.392	.305	3.91	.749

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

TABLE 1.
continued

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Mean	Fit results
X_5	2.467	3.204	3.119	5.001	**9.773**	.186	.362	.392	.235	5.64	.121
X_6	1.404	1.688	1.709	1.994	1.186	**4.171**	.301	.299	.276	2.92	
X_7	2.378	3.048	2.239	2.722	3.610	1.960	**10.158**	.756	.326	4.86	
X_8	4.546	6.200	4.417	5.201	6.507	3.244	12.812	**28.258**	.349	11.01	
X_9	1.165	1.680	1.450	1.338	1.290	0.991	1.830	3.262	**3.096**	2.67	

The lower triangle contains covariances and upper triangle correlations. F_{Cov} Diff & F_{Mean} Diff = Fitted F_{ML} value for the covariance (or mean) part − Desired F_{ML} value for the covariance (or mean) part.

somewhat evenly over all the elements of $\boldsymbol{\mu}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, representing a realistic modeling scenario. For all the four sets of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$, the $\boldsymbol{\theta}_{fit}$ obtained is exactly equal to the specified value $\boldsymbol{\theta}_0$, indicating the goal in Eq. (2) is satisfied. Regarding the amount of misfit, the difference in F_{cov} between fitted and desired values is around 10^{-9}, and the difference in F_{mean} between fitted and desired values is around 10^{-17} (see Table 1). Such a small difference between the desired and fitted F_{ML} values can be well explained by rounding, and the goals in Eqs. (3) and (4) are also satisfied.[1] Therefore, even when such a large amount of model misfit is specified, the proposed method still performs extremely well and is able to yield $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ as desired.

2.1. Contrast to the Type I Approach

For Model 1, if a researcher wants to create a misspecified model from the Type I perspective, there are only six model parameters possible for removal, namely the three latent means and the three latent covariances. To better illustrate the limitations of the Type I perspective, we fit the six misspecified models (only one parameter is missing in each model) to $\boldsymbol{\mu}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and record the fit indices and F_{ML} values in Table 2. First, it is almost impossible for the Type I approach to control the amount of model misfit. Among the six wrong models, the best-fitting model is the one without c_{F1F3}, which yields RMSEA = .073; CFI = .950; SRMR = .122. Unless changing the $\boldsymbol{\theta}_0$ values, it is impossible to have a wrong model with, say an RMSEA of .05 or .06. It is also impossible to achieve a larger RMSEA values, say .08 or .10 in the current example, and the only available RMSEA values are those six values in Table 2. However, the method we proposed allows the researcher to specify the desired F_{ML} values on a continuous scale. Second, if one removes a latent covariance from Model 1, the model is wrong in the covariance part only and its $F_{\text{mean}}^{(\text{fit})}$ is always 0 (see Table 2). To introduce misfit on the mean part, one has to remove a latent mean, but in that case the model will be too wrong to be useful in a simulation study (see Table 2 for the fit indices). By contrast, our method is able to control both $F_{\text{mean}}^{(\text{fit})}$ and $F_{\text{cov}}^{(\text{fit})}$ at a reasonable level. Third, the Type I approach is unable to control the location of model misfit. If one removes a latent covariance, say c_{F1F2}, then only some elements in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ will have misfit and the other elements will remain in perfect fit, showing an unrealistic pattern in the residual matrix (see Table 3). In addition, if one removes a latent mean (e.g., a_{F1}), the model-implied means will all be zeros for the indicators that load on this latent factor (e.g., X_1 to X_3; see Table 3). Again this is not a problem for our method, as we saw earlier in Table 1 that for the four sets of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$, all the elements in $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ have some lack of fit, depicting a more realistic scenario.

[1] In R, by default two quantities are considered functionally equal if their difference is less than 1.5×10^{-8}.

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

TABLE 2.
Population fit indices and F_{ML} values after removing a parameter from the model in Demonstration 1.

Remove path	RMSEA	CFI	SRMR	F_{Cov}	F_{Mean}
c_{F1F2}	.079	.942	.128	.193	0
c_{F1F3}	.073	.950	.122	.166	0
c_{F2F3}	.093	.920	.138	.265	0
a_{F1}	.187	.674	.664	.421	.661
a_{F2}	.228	.515	1.235	.809	.800
a_{F3}	.233	.492	1.364	.872	.815

TABLE 3.
Residuals or model-implied moments after removing a parameter from Model 1 in Demonstration 1.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Residuals after removing c_{F1F2}									
X_1	**0**								
X_2	**0**	**0**							
X_3	**0**	**0**	**0**						
X_4	1.80	2.88	1.44	**0**					
X_5	2.52	4.03	2.02	**0**	**0**				
X_6	1.26	2.02	1.01	**0**	**0**	**0**			
X_7	.83	1.33	.66	.46	.64	.32	.42		
X_8	1.82	2.92	1.46	1.01	1.41	.70	.93	2.04	
X_9	.41	.66	.33	.23	.32	.16	.21	.46	.11
Residuals for mean	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Model-implied means after removing a_{F1}									
	0	**0**	**0**	2.93	3.96	1.98	3.70	8.14	1.85

3. Multiple Group Analysis

In this section, we first introduce some new notations for the multigroup context. We illustrate the notations with Model 2 in Fig. 2. We focus on the cases where the model form is the same across all the G groups. We use "(g)" flexibly in the subscript or superscript (e.g., $\theta_a^{(g)}$, $\theta'_{(g)}$) to denote the elements within group g (where $g = 1, 2, \cdots, G$) and use vectors or matrices without "(g)" to denote those that contain the corresponding elements of all the G groups. For example, for the model in Fig. 2, we have $\theta = [\theta'_{(1)}, \theta'_{(2)}]'$ and $\boldsymbol{\Sigma} = Diag[\boldsymbol{\Sigma}_{(1)}, \boldsymbol{\Sigma}_{(2)}]$, where $Diag[\cdot]$ denotes a block diagonal matrix. Suppose Model 2 has the constraints $a_F^{(1)} = 0$, $\mathbf{a}_x^{(1)} = \mathbf{a}_x^{(2)}$, and $\mathbf{b}^{(1)} = \mathbf{b}^{(2)}$, where $\mathbf{a}_x = [a_1, a_2, a_3]'$ and $\mathbf{b} = [b_2, b_3]'$. Then, the Type-a parameters in the two groups are written as $\theta_a^{(1)} = [0, a_1^{(1)}, a_2^{(1)}, a_3^{(1)}]'$ and $\theta_a^{(2)} = [a_F^{(2)}, a_1^{(2)}, a_2^{(2)}, a_3^{(2)}]'$. That is, the number of elements in $\theta_a^{(g)}$ remains the same across all groups, and we define $\theta_b^{(g)}$ and $\theta_c^{(g)}$ in the same manner. If a parameter is set to a special value in a certain group g (e.g., $a_F^{(1)} = 0$), we still include the fixed value in $\theta_{(g)}$. (Thus, the first element of $\theta_a^{(1)}$ is 0.) In addition, for the parameters constrained to be equal over some groups, we consider them as the same parameter but give them different labels in $\theta_{(g)}$ (e.g., $\theta_b^{(1)} = [b_2^{(1)}, b_3^{(1)}]'$, $\theta_b^{(2)} = [b_2^{(2)}, b_3^{(2)}]'$). Regarding the type of constraints, we focus on between-group equality constraints only, because other constraint types seldom appear in practice. Let q denote the number of elements in $\theta_{(g)}$,

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

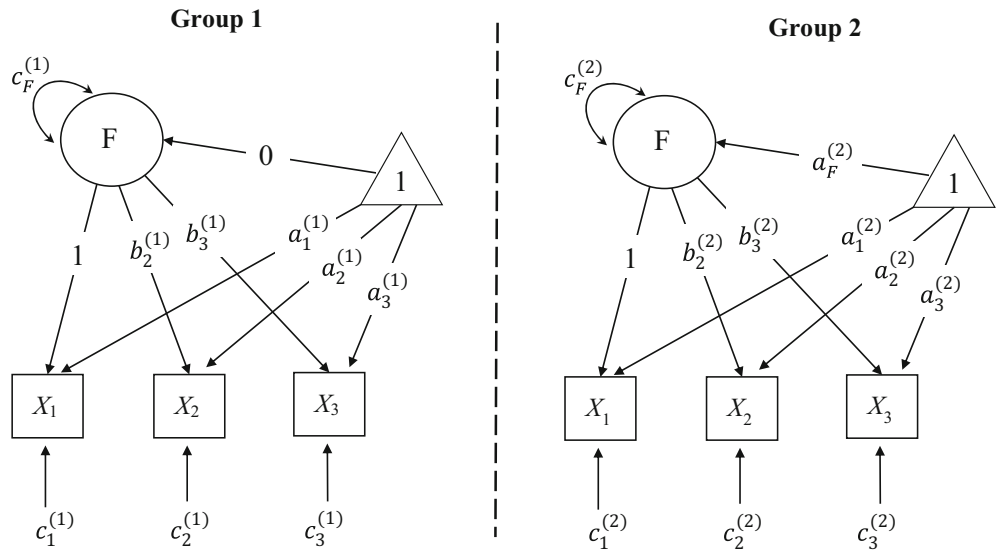


FIGURE 2.
Path diagram of Model 2 used to introduce new notations for our proposed method in the multiple group context. The latent mean is fixed at 0 in Group 1, but freely estimated in Group 2. All the factor loadings and intercepts are constrained equal across the two groups. The parameters constrained to be equal across groups are considered as the same parameter, but have different labels (e.g., $a_1^{(1)}$ and $a_1^{(2)}$ are the same parameters, but have two different labels).

$q_{(g)}$ the number of free parameters in $\theta_{(g)}$, and q_{all} the total number of free parameters in θ. The subsets corresponding to the Type-a, b, and c parameters are defined in the same way. For example, in Fig. 2, $q_a^{(1)} = 3$ and $q_a = q_a^{(2)} = 4$. Vector θ_a has $q_a G = 8$ elements, but $q_a^{(\text{all})} = 4$. We consider the model-implied moments in group g as functions of θ (not $\theta_{(g)}$) and denote them as $\boldsymbol{\mu}_{(g)}(\theta)$ and $\boldsymbol{\Sigma}_{(g)}(\theta)$. The model-implied moments of the whole multigroup analysis are $\boldsymbol{\mu}(\theta) = [\boldsymbol{\mu}_{(1)}(\theta)', \boldsymbol{\mu}_{(2)}(\theta)', \cdots, \boldsymbol{\mu}_{(G)}(\theta)']'$ and $\boldsymbol{\Sigma}(\theta) = Diag[\boldsymbol{\Sigma}_{(1)}(\theta), \boldsymbol{\Sigma}_{(2)}(\theta), \cdots, \boldsymbol{\Sigma}_{(G)}(\theta)]$.

Given θ_0, the parameter values of all the G groups, we continue to describe the misfit as $\boldsymbol{\mu}^* = \boldsymbol{\mu}(\theta_0) + \mathbf{t}$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}(\theta_0) + \mathbf{E}$, where $\mathbf{t} = [\mathbf{t}'_{(1)}, \mathbf{t}'_{(2)}, \cdots, \mathbf{t}'_{(G)}]'$ and $\mathbf{E} = Diag[\mathbf{E}_{(1)}, \mathbf{E}_{(2)}, \cdots, \mathbf{E}_{(G)}]$. The population F_{ML} value in multigroup analysis is $F_{\text{ML}} = \sum_{g=1}^{G} F_{(g)}/G$, where

$$\begin{aligned} F_{(g)} = {} & [\boldsymbol{\mu}^*_{(g)} - \boldsymbol{\mu}_{(g)}(\theta)]' \cdot \boldsymbol{\Sigma}^{-1}_{(g)}(\theta) \cdot [\boldsymbol{\mu}^*_{(g)} - \boldsymbol{\mu}_{(g)}(\theta)] \\ & + \ln|\boldsymbol{\Sigma}_{(g)}(\theta)| - \ln|\boldsymbol{\Sigma}^*_{(g)}| + \text{tr}[\boldsymbol{\Sigma}^*_{(g)}\boldsymbol{\Sigma}^{-1}_{(g)}(\theta)] - p. \end{aligned} \tag{17}$$

is the ML fit function value in group g. To construct misfit for multigroup moment structures, we try to find $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ (or equivalently $\mathbf{t}$ and $\mathbf{E}$) that simultaneously satisfy

$$\begin{aligned} & \theta_0 = \arg\min F_{\text{ML}}[\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta)] \\ & \text{subject to } \gamma_a(\theta_a) = \mathbf{0}; \gamma_b(\theta_b) = \mathbf{0}; \gamma_c(\theta_c) = \mathbf{0}; \end{aligned} \tag{18}$$

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

$$\sum_{g=1}^{G}[\boldsymbol{\mu}^*_{(g)} - \boldsymbol{\mu}_{(g)}(\boldsymbol{\theta}_0)]' \cdot \boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0) \cdot [\boldsymbol{\mu}^*_{(g)} - \boldsymbol{\mu}_{(g)}(\boldsymbol{\theta}_0)] = F_{\text{mean}}; \tag{19}$$

$$\sum_{g=1}^{G}\left\{\ln|\boldsymbol{\Sigma}_{(g)}(\boldsymbol{\theta}_0)| - \ln|\boldsymbol{\Sigma}^*_{(g)}| + \text{tr}[\boldsymbol{\Sigma}^*_{(g)}\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)] - p\right\} = F_{\text{cov}}, \tag{20}$$

where $\gamma_a(\cdot)$, $\gamma_b(\cdot)$, and $\gamma_c(\cdot)$ are between-group equality constraints on the Type-a, Type-b, and Type-c model parameters. Note the $\boldsymbol{\theta}_0$ specified by the researcher must already satisfy those constraints; otherwise, $\boldsymbol{\theta}_0$ is not even a feasible point in the optimization.

Similar to the basic procedure, we first calculate the derivative of F_{ML} with respect to $\boldsymbol{\theta}$ and set to zero the derivative evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Note that $F_{(g)}$ in Eq. (17) is defined as a function of $\boldsymbol{\theta}$ (not $\boldsymbol{\theta}_{(g)}$), and the derivative of F_{ML} is easily calculated as $\dot{F}(\boldsymbol{\theta}) = \sum_{g=1}^{G}\dot{F}_{(g)}(\boldsymbol{\theta})$. Accordingly, we have

$$\dot{F}(\boldsymbol{\theta}_0) = \sum_{g=1}^{G}\left[\boldsymbol{\Omega}_{(g)} \cdot \mathbf{t}_{(g)} + \boldsymbol{\Delta}_{(g)} \cdot \text{vec}(\mathbf{t}_{(g)}\mathbf{t}'_{(g)}) + \boldsymbol{\Delta}_{(g)} \cdot \text{vec}\mathbf{E}_{(g)}\right] = \mathbf{0}, \tag{21}$$

where $\boldsymbol{\Omega}_{(g)} = 2\dot{\boldsymbol{\mu}}_{(g)}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Delta}_{(g)} = \dot{\boldsymbol{\Sigma}}_{(g)}(\boldsymbol{\theta}_0)[\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)]$. To proceed, first let us consider the forms of $\dot{\boldsymbol{\mu}}_{(g)}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{(g)}$ more closely. In particular, $\dot{\boldsymbol{\mu}}_{(g)}(\boldsymbol{\theta})$ is a $qG \times p$ block vector:

$$\dot{\boldsymbol{\mu}}_{(g)}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{(1)}} \\ \frac{\partial \boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{(2)}} \\ \vdots \\ \frac{\partial \boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{(G)}} \end{bmatrix}. \tag{22}$$

The derivative $\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_{(g)}$ is calculated in the same way as that in single-group analysis. If $\theta^{(g)}_j$ is a fixed value, $\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})/\partial\theta^{(g)}_j = \mathbf{0}'$. Similar to the single-group context, we require the nonzero rows of $\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{(g)}$ concerning the Type-a parameters are linearly independent, and thus, $\text{rank}(\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^{(g)}_a) = \min\{q^{(g)}_a, p\}$. For the derivative of $\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})$ with respect to parameters in another group k, if there is a constraint $\theta^{(k)}_j = \theta^{(g)}_j$, then we consider $\theta^{(k)}_j$ and $\theta^{(g)}_j$ as the same parameter and write $\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})/\partial\theta^{(k)}_j = \partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})/\partial\theta^{(g)}_j$. If there is no constraint between $\theta^{(k)}_j$ and $\theta^{(g)}_j$, then $\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})/\partial\theta^{(k)}_j$ is simply $\mathbf{0}'$. Therefore, $\partial\boldsymbol{\mu}_{(g)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_{(g)}$ contains all the unique rows of $\dot{\boldsymbol{\mu}}_{(g)}(\boldsymbol{\theta})$. Because the j-th row in $\boldsymbol{\Omega}_{(g)}$ is the j-th row of $\dot{\boldsymbol{\mu}}_{(g)}(\boldsymbol{\theta}_0)$ post-multiplied by $2\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)$ has full rank p, it follows $\text{rank}(\boldsymbol{\Omega}^{(g)}_a) = \min\{q^{(g)}_a, p\}$. Similarly to Eq. (22), $\dot{\boldsymbol{\Sigma}}_{(g)}(\boldsymbol{\theta})$ is also a block vector, where the k-th block of rows is $\partial\text{vec}'[\boldsymbol{\Sigma}_{(g)}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}_{(k)}$. We also require the $[q^{(g)}_b + q^{(g)}_c]$ nonzero rows of $\partial\boldsymbol{\Sigma}_{(g)}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{(g)}$ being linearly independent. Applying the same argument, it can be shown that $\text{rank}(\dot{\boldsymbol{\Sigma}}_{(g)}(\boldsymbol{\theta}_0)) = \text{rank}(\partial\boldsymbol{\Sigma}_{(g)}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_{(g)}) = \min\{q^{(g)}_b + q^{(g)}_c, p^2\}$. Because $[\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)]$ has full rank p^2 and $q^{(g)}_b + q^{(g)}_c < p^2$, it follows $\text{rank}(\boldsymbol{\Delta}_{(g)}) = q^{(g)}_b + q^{(g)}_c$.

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

Similar to Eq. (7), we rearrange the rows of $\boldsymbol{\Omega}_{(g)}$ and $\boldsymbol{\Delta}_{(g)}$ in terms of Type-a, b, and c parameters (the rows were in group order before) and rewrite Eq. (21) as

$$\sum_{g=1}^{G}\begin{bmatrix}\boldsymbol{\Omega}_a^{(g)}\\ \boldsymbol{\Omega}_b^{(g)}\\ \mathbf{O}\end{bmatrix}\cdot\mathbf{t}_{(g)}+\sum_{g=1}^{G}\begin{bmatrix}\mathbf{O}\\ \boldsymbol{\Delta}_b^{(g)}\\ \boldsymbol{\Delta}_c^{(g)}\end{bmatrix}\cdot\text{vec}[\mathbf{t}_{(g)}\mathbf{t}'_{(g)}]+\sum_{g=1}^{G}\begin{bmatrix}\mathbf{O}\\ \boldsymbol{\Delta}_b^{(g)}\\ \boldsymbol{\Delta}_c^{(g)}\end{bmatrix}\cdot\text{vec}\mathbf{E}_{(g)}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}. \tag{23}$$

First, let us consider $\sum_{g=1}^{G}\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}=\mathbf{0}$. If $\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}=\mathbf{0}$ holds for any g in $1, 2, \cdots, G$, certainly $\sum_{g=1}^{G}\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}$ will be zero, but this case is trivial and often unrealistic in multigroup analyses. This is because $\text{rank}(\boldsymbol{\Omega}_a^{(g)})=\min\{q_a^{(g)}, p\}$, and $\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}=\mathbf{0}$ has nonzero solutions only if $q_a^{(g)}<p$, namely the free intercepts and means together in any given group are fewer than p. In the special case where $q_a^{(g)}<p$ holds for all the groups, one can simply solve $\sum_{g=1}^{G}\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}=\mathbf{0}$ by solving $\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}=\mathbf{0}$ individually for the G groups. In that case, it is even possible to control the misfit $\mathbf{t}'_{(g)}\boldsymbol{\Sigma}_{(g)}^{-1}(\theta_0)\mathbf{t}_{(g)}=F_{\text{mean}}^{(g)}$ separately for the G groups. This special case amounts to conducting the single-group procedure G times separately, and thus, we do not elaborate on it. However, multigroup analyses often involve Type-a parameters for all the manifest and latent variables, causing $q_a^{(g)}>p$ (for example, Model 2 has $q_a^{(1)}=3$ and $q_a^{(2)}=4$, but $p=3$). The model is not identified within group g, but with proper constraints $q_a^{(\text{all})}<pG$ becomes possible, and the multigroup analysis can be identified on the mean part. Accordingly, it often needs to solve $\sum_{g=1}^{G}\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}=\mathbf{0}$ simultaneously for all the G groups.

We recognize that

$$\sum_{g=1}^{G}\boldsymbol{\Omega}_a^{(g)}\cdot\mathbf{t}_{(g)}=\boldsymbol{\Omega}_a\cdot\mathbf{t}=\mathbf{0}, \tag{24}$$

where $\boldsymbol{\Omega}_a=[\boldsymbol{\Omega}_a^{(1)}|\boldsymbol{\Omega}_a^{(2)}|\cdots|\boldsymbol{\Omega}_a^{(G)}]$ is a $q_aG\times pG$ block row vector (recall q_a is the length of $\theta_a^{(g)}$) and $\text{rank}(\boldsymbol{\Omega}_a)=q_a^{(\text{all})}<pG$. Accordingly, Eq. (24) has infinitely many solutions, and a generic solution for $\mathbf{t}$ is $\mathbf{t}=(\mathbf{I}-\boldsymbol{\Omega}_a^{-}\boldsymbol{\Omega}_a)\mathbf{y}_t$, in the same form as $\mathbf{t}$ in the single-group context. Next, we find a specific $\mathbf{t}$ to satisfy Eq. (19). We recognize that

$$\begin{aligned}&\sum_{g=1}^{G}[\boldsymbol{\mu}_{(g)}^{*}-\boldsymbol{\mu}_{(g)}(\theta_0)]'\cdot\boldsymbol{\Sigma}_{(g)}^{-1}(\theta_0)\cdot[\boldsymbol{\mu}_{(g)}^{*}-\boldsymbol{\mu}_{(g)}(\theta_0)]\\ &=[\mathbf{t}'_{(1)},\mathbf{t}'_{(2)},\cdots,\mathbf{t}'_{(G)}]\cdot Diag[\boldsymbol{\Sigma}_{(1)}^{-1}(\theta_0),\boldsymbol{\Sigma}_{(2)}^{-1}(\theta_0),\cdots,\boldsymbol{\Sigma}_{(G)}^{-1}(\theta_0)]\cdot[\mathbf{t}'_{(1)},\mathbf{t}'_{(2)},\cdots,\mathbf{t}'_{(G)}]'\\ &=\mathbf{t}'\cdot\boldsymbol{\Sigma}^{-1}(\theta_0)\cdot\mathbf{t}\end{aligned} \tag{25}$$

which is also in the same form as Eq. (3) in the single-group context. Therefore, we can directly apply the procedure in the single-group context to the current problem and find a solution for $\mathbf{t}$. Given $\mathbf{t}$, its subsets $\mathbf{t}_{(1)}, \mathbf{t}_{(2)}, \cdots, \mathbf{t}_{(G)}$ can be easily obtained.

Next, we return to Eq. (23) and consider $\mathbf{E}_{(g)}$. For group g, given the values of $\mathbf{t}_{(g)}$ and $\text{vec}[\mathbf{t}_{(g)}\mathbf{t}'_{(g)}]$, Eq. (23) leads to

$$\sum_{g=1}^{G}\boldsymbol{\Delta}_{(g)}\text{vec}\mathbf{E}_{(g)}=-\sum_{g=1}^{G}\boldsymbol{\Omega}_{(g)}\mathbf{t}_{(g)}-\sum_{g=1}^{G}\boldsymbol{\Delta}_{(g)}\text{vec}[\mathbf{t}_{(g)}\mathbf{t}'_{(g)}]\equiv\sum_{g=1}^{G}\boldsymbol{\eta}_{(g)}, \tag{26}$$

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

where

$$\boldsymbol{\eta}_{(g)} = -\begin{bmatrix} \boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)} \\ \boldsymbol{\Omega}_b^{(g)}\mathbf{t}_{(g)} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Delta}_b^{(g)}\text{vec}[\mathbf{t}_{(g)}\mathbf{t}'_{(g)}] \\ \boldsymbol{\Delta}_c^{(g)}\text{vec}[\mathbf{t}_{(g)}\mathbf{t}'_{(g)}] \end{bmatrix}. \tag{27}$$

Note $\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)}$ is often not $\mathbf{0}$ based on the discussion above, but the first q_aG rows of $\boldsymbol{\Delta}_{(g)}$ are all zeros, and thus it is usually not possible to solve $\boldsymbol{\Delta}_{(g)}\text{vec}\mathbf{E}_{(g)} = \boldsymbol{\eta}_{(g)}$ separately for group g. Instead, we construct the $Gq \times Gp^*$ block row vector $\mathbf{B} = [\boldsymbol{\Delta}_{(1)}\mathbf{D}|\boldsymbol{\Delta}_{(2)}\mathbf{D}|\cdots|\boldsymbol{\Delta}_{(G)}\mathbf{D}]$ and solve

$$\mathbf{B}\cdot\mathbf{v} = \boldsymbol{\eta}_{\text{sum}}, \tag{28}$$

where $\mathbf{v} = [\text{vec}h'(\mathbf{E}_{(1)}), \text{vec}h'(\mathbf{E}_{(2)}), \cdots, \text{vec}h'(\mathbf{E}_{(G)})]'$ and $\boldsymbol{\eta}_{\text{sum}} = \sum_{g=1}^{G}\boldsymbol{\eta}_{(g)}$. Note $\mathbf{v} \neq \text{vec}h(\mathbf{E})$. Because $\sum_{g=1}^{G}\boldsymbol{\Omega}_a^{(g)}\mathbf{t}_{(g)} = \mathbf{0}$, the first q_aG elements of $\boldsymbol{\eta}_{\text{sum}}$ are all zeros, and so are the first q_aG rows of $\mathbf{B}\cdot\mathbf{v}$. Using the same argument as that in the single-group context, $\text{rank}(\mathbf{B}) \leq q_b^{(\text{all})} + q_c^{(\text{all})}$ and it is possible to have $\text{rank}([\mathbf{B}|\boldsymbol{\eta}_{\text{sum}}]) = \text{rank}(\mathbf{B})$. Therefore, Eq. (28) can be a consistent system with infinitely many solutions. We also require the solution $\mathbf{v}$ to satisfy Eq. (20). Given that

$$\sum_{g=1}^{G}\ln|\boldsymbol{\Sigma}_{(g)}(\boldsymbol{\theta}_0)| = \ln\left\{\prod_{g=1}^{G}|\boldsymbol{\Sigma}_{(g)}(\boldsymbol{\theta}_0)|\right\} = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)|; \tag{29}$$

$$\sum_{g=1}^{G}\ln|\boldsymbol{\Sigma}^*_{(g)}| = \ln|\boldsymbol{\Sigma}^*|; \tag{30}$$

$$\sum_{g=1}^{G}\text{tr}[\boldsymbol{\Sigma}^*_{(g)}\boldsymbol{\Sigma}^{-1}_{(g)}(\boldsymbol{\theta}_0)] = \text{tr}[\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)]; \tag{31}$$

Equation (20) is equivalent to

$$\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)| - \ln|\boldsymbol{\Sigma}^*| + \text{tr}[\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)] - Gp = F_{\text{cov}}, \tag{32}$$

which has the same form as Eq. (4). Therefore, the steps for finding $\mathbf{v}$ to satisfy Eq. (20) are the same as the method described in the single-group context. In particular, we write the general solution to Eq. (28) as $\mathbf{v} = \mathbf{B}^{-}\boldsymbol{\eta}_{\text{sum}} + (\mathbf{I} - \mathbf{B}^{-}\mathbf{B})\mathbf{y}_v$ and define a new function $\phi = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)| - \ln|\boldsymbol{\Sigma}^*| + \text{tr}[\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)] - Gp - F_{\text{cov}}$. The Jacobian of ϕ with respect to $\mathbf{y}_v$ is $\mathbf{J}_\phi(\mathbf{y}_v) = (\partial\phi/\partial\mathbf{E})(\partial\mathbf{E}/\partial\mathbf{v})(\partial\mathbf{v}/\partial\mathbf{y}_v)$, where $\partial\phi/\partial\mathbf{E}$ has the same form as Eq. (14). Also, $\partial\mathbf{E}/\partial\mathbf{v} = \partial[\mathbf{D}\text{vec}h(\mathbf{E})]/\partial\mathbf{v}$ is a constant matrix with 0's and 1's, and $\partial\mathbf{v}/\partial\mathbf{y}_v = \mathbf{I} - \mathbf{B}^{-}\mathbf{B}$. The solution $\mathbf{y}_v$ to $\phi = 0$ can be obtained with the Newton method in the same way as the single-group case.

Based on the discussion so far, the $\mathbf{t}$ and $\mathbf{E}$ obtained will satisfy Eqs. (19), (20), and (21), and the only task left is to show that such $\mathbf{t}$ and $\mathbf{E}$ will also satisfy Eq. (18). This is easily achieved with Proposition 1, with the adaptation that the Hessian $\partial^2 F[\mathbf{m}_0, \mathbf{m}(\boldsymbol{\theta}_0)]/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}$ of the whole multigroup analysis is positive definite. In fact, Proposition 1 implies that $\boldsymbol{\theta}_0$ is the global minimizer of F_{ML} even in unconstrained optimization, and obviously, $\boldsymbol{\theta}_0$ will also satisfy the constrained optimization in Eq. (18).

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

TABLE 4.
Population model parameter values for the simulation study in Demonstration 2.

	Unstd loading		Unstd error var		Intercept/mean		Std loading	
	G1	G2	G1	G2	G1	G2	G1	G2
X_1	**1***		2.2	2.4	2.5		.803	.822
X_2	1.6		2.6	2.8	2.0		.893	.906
X_3	0.8		3.2	3.5	1.5		.667	.691
X_4	**1***		2.5	2.7	2.5		.739	.773
X_5	1.4		4.3	4.4	2.0		.760	.800
X_6	0.7		3.4	3.6	1.5		.549	.594
X_7	**1***		3.8	3.5	2.5		.754	.795
X_8	2.2		4.3	4.0	2.0		.921	.938
X_9	0.5		2.7	3.0	1.5		.563	.577
F_1					**0***	7		
F_2					**0***	8		
F_3					**0***	9		

Factor covariance (lower) & correlation (upper)

	G1			G2	
4	.520	.447	5	.626	.475
1.8	3	.594	2.8	4	.490
2.0	2.3	5	2.6	2.4	6

The unstandardized factor loadings and intercepts of X_1 to X_9 are the same in Groups 1 and 2.
Bold values with a star indicate the model parameter is fixed at that particular value for model identification.
G1 & G2 = Groups 1 & 2.

4. Demonstration 2

In this section, we carry out an example simulation study in the multigroup analysis context. Simulation studies in the literature of measurement invariance or latent mean comparisons have predominantly used the Type I perspective to misspecify models, where equality constraints are imposed on groups whose parameters actually differ. Therefore, if the incorrect constraints are removed the model will have perfect fit. A more realistic situation is, even when the model has only the correct constraints, the model still fails to perfectly reproduce the data but only holds approximately. Accordingly, in this demonstration, we conduct simulations to study the quality of parameter estimations in this more realistic situation. The model is a 3-factor CFA model, where X_1 to X_3 load on F_1, X_4 to X_6 load on F_2, and X_7 to X_9 load on F_3. No cross-loadings are present, the three factors are correlated, and the measurement errors are independent of each other and of the factors. The nine intercepts and three latent means are all estimated. The analysis involves two groups, where all the factor loadings and intercepts are constrained equal over the two Groups. The loadings of X_1, X_4, and X_7 are fixed at 1, and the three latent means are fixed at 0 in Group 1 but free in Group 2. The specified model parameter values θ_0 are given in Table 4. It is clear that all the between-group constraints are correct.

Giving the model and θ_0, we used the proposed method to create population moments based on various desired misfit levels. We chose values of .075, .192, .300, and .432 to represent four conditions of misfit, corresponding to RMSEA values of .05, .08, .10, and .12. Note $F_{ML} = (F_{mean} + F_{cov})/2$ as $G = 2$. Given F_{ML}, the desired values of F_{mean} and F_{cov} are calculated using an 15/85 ratio, so as to keep F_{mean}/F_{cov} consistent with $df_{mean}/df_{cov} = 1/9$ while slightly worsen the misfit on the mean part. Given $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ in a condition, we generated random data

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

TABLE 5.
Relative bias of point estimates and of standard errors for selected model parameters in Demonstration 2.

F_{ML}	.075		.192		.300		.432	
RMSEA	.050		.080		.100		.120	
CFI	.979		.951		.923		.893	
SRMR	.043		.072		.082		.098	
	Pt Est	SE	Pt Est	SE	Pt Est	SE	Pt Est	SE
$a_{F1}^{(2)}$	.001	.006	.000	.006	.001	− .042	.000	.000
$a_{F2}^{(2)}$	.001	.016	− .001	− .033	.000	.010	.001	− .008
$a_{F3}^{(2)}$	.000	.005	− .001	.010	.000	.027	.000	.037
a_{X7}	− .001	.053	.002	− .028	.000	.099	.002	.072
a_{X8}	− .003	−.007	.002	.001	.008	.018	.000	.013
a_{X9}	.000	.017	.004	−.059	.004	−.079	.002	−.**136**
b_{F3X8}	.000	− .002	.001	− .024	.000	− .**100**	.000	− .**145**
b_{F3X9}	.001	− .038	.000	.067	− .001	− .**109**	− .001	− .073

Pt Est = relative bias of point estimates, SE = relative bias of standard errors. Bold values = Relative bias exceeding 10% for SE or 5% for point estimates.

from a multivariate normal distribution using $n = 150$ per group, and replicated 2,000 times in each condition. Of interest are point estimates and standard errors. All the replications converged and representative results are in Table 5. Results suggest that latent means have largely unbiased SE even when the model is relatively poor (RMSEA = .120; CFI = .893; SRMR = .098). Intercepts are slightly more difficult to estimate, but the SE becomes problematic only in the condition with the most serious misfit. Factor loadings begin to have unacceptable SEs if the misfit is somewhat large (RMSEA = .100; CFI = .923; SRMR = .082). Therefore, for the situations we have explored, it seems it is still safe to compare latent means when a model is quite poor, but investigations of factor loading invariance require a much better model. Moreover, it is dangerous to solely rely on fit indices to judge whether a constraint is valid, as we see in this demonstration the constraints are always correct, but the misfit can range from somewhat small to fairly large.

5. Implications of Many Possible Population Moments

Given the model form, θ_0 and the desired F_{mean} and F_{cov} values, there are infinite sets of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ that can satisfy the goals in Eqs. (2) to (4). Accordingly, it is natural to ask whether some $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ are better than others and how should one choose the population moments for a simulation study. Before we discuss this, it is important to note the possibility of generating data from many populations is not a limitation of our method, but rather the normal state of affairs in statistical simulation studies. MacCallum and Tucker's (1991) method and Cudeck and Browne's (1992) method are two examples of the Type II approach, and they both can yield infinite $\boldsymbol{\Sigma}^*$ matrices that satisfy their research goals. Similarly, in the context of data analysis, different studies do not sample from exactly the same population (e.g., Tucker et al., 1969). Wu and Browne (2015) conceptualized this problem as one where study j collects a sample $\mathbf{S}_j$ from its population $\boldsymbol{\Sigma}_j$, while all the $\boldsymbol{\Sigma}_j$'s come from a hyper-population with mean $\boldsymbol{\Sigma}_{\text{hyper}}$. From Wu and Browne's perspective, finding a $\boldsymbol{\Sigma}^*$ for simulations amounts to choosing a $\boldsymbol{\Sigma}_j$ randomly and then generating random samples from it. However, this paper concerns how to construct $\boldsymbol{\Sigma}_j$'s , whereas Wu and Browne's method pertains to estimating $\boldsymbol{\Sigma}_{\text{hyper}}$ given an $\mathbf{S}_j$.

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

TABLE 6.
Residuals for fitting Model 1 to population moments created with fixed initial values.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Mean
X_1	−0.097									0.016
X_2	0.217	0.531								−0.056
X_3	−0.701	−0.357	−0.528							0.108
X_4	−0.032	0.064	0.204	0.363						−0.046
X_5	−0.187	−0.178	0.176	0.548	−0.041					0.004
X_6	−0.007	0.085	0.167	−0.280	−0.972	−0.471				0.083
X_7	−0.136	−0.204	0.105	−0.309	−0.492	−0.185	−0.681			0.068
X_8	0.002	−0.098	0.475	0.374	0.073	0.192	−0.156	1.655		−0.076
X_9	0.009	0.061	0.095	−0.103	−0.217	−0.094	−1.566	−0.907	−0.575	0.113

Regardless of the method for creating model misfit, one could always ask why a simulation study is based on one model (say three-factor CFA) rather than another (say two-factor CFA). Once the model form is chosen, one could ask why a population model parameter equals a certain value instead of being 2% larger or 3% smaller. All such questions pertain to the $\boldsymbol{\Sigma}$ that gives rise to the random data in simulations, and slight changes in $\boldsymbol{\Sigma}(\cdot)$ or θ_0 can lead to a different $\boldsymbol{\Sigma}$. Thus, the $\boldsymbol{\Sigma}$ matrices available for a simulation study are always infinite. If the researcher uses the Type I approach to misspecify the model, one could also ask why a model parameter is removed instead of another parameter being removed. Although on surface the Type I approach yields a definite $\boldsymbol{\Sigma}$ and model, implicitly such uniqueness results from an arbitrary selection from the otherwise infinite suitable $\boldsymbol{\Sigma}$ matrices and models. In essence, asking how to choose $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ in applying our method is similar to asking questions like "Is 0.72 or 0.75 better for the population factor loading of X_1?" All the $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ satisfying Eqs. (2) to (4) are reasonable, and one can just randomly pick a set for their simulation study. This strategy has been widely used in other simulation studies that applied MacCallum and Tucker's (1991) method or Cudeck and Browne's (1992) method (see, e.g., MacCallum et al., 1999, 2001; Lai & Green, 2016; Lai, 2018).

Nevertheless, is it possible to obtain a unique $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$? We discuss several possible directions. First, to obtain $\boldsymbol{\mu}^*$, we need $\mathbf{y}_t$ for Eq. (10), where $\mathbf{y}_t$ can be *any* $p \times 1$ vector. Therefore, to ensure $\boldsymbol{\mu}^*$ is unique, one can simply assign specific values to $\mathbf{y}_t$ instead of getting a $\mathbf{y}_t$ randomly. Even vectors like $[1, 1, \cdots, 1]$ and $[1, 0, \cdots, 0]$ for $\mathbf{y}_t$ will satisfy Eq. (10). Moreover, such a special $\mathbf{y}_t$ will not undermine the verisimilitude of $\boldsymbol{\mu}^*$, because the "perfect" values in $\mathbf{y}_t$ are later multiplied by other ordinary matrices in constructing $\boldsymbol{\mu}^*$ (see Eq. (10)). Similarly, one can assign convenient values to $\mathbf{y}_E$ in constructing $\boldsymbol{\Sigma}^*$. To illustrate, we revisit Model 1 in Demonstration 1 and create $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ using $F_{\text{mean}} = (.05)^2 df_{\text{mean}} = .015$ and $F_{\text{cov}} = (.10)^2 df_{\text{cov}} = .24$ as the desired misfit amount. We set all the elements in $\mathbf{y}_t$ and $\mathbf{y}_E$ to 1, fit Model 1 to the resulting (unique) $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ and report the residuals on the mean and covariance parts in Table 6. Clearly, the residuals do not exhibit any systematic pattern.

A second method to obtain unique $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ is to impose additional constraints on these data moments. For example, one could choose the moments with the smallest or largest $\|\mathbf{m}_0 - \mathbf{m}^*\|$ (recall $\mathbf{m} = [\boldsymbol{\mu}', \boldsymbol{\sigma}']'$). Geometrically, this means to find $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ that are closest to, or farthest from, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. To achieve this goal, the current problem becomes to minimize (or maximize) $\|\mathbf{m}_0 - \mathbf{m}^*\|$ subject to the constraints in Eqs. (2) to (4). How to perform such an optimization is beyond the scope of this paper and requires future research. Third, in the context of covariance structure analysis, Chun and Shapiro (2010) studied a similar optimization problem with the goal to obtain a unique $\boldsymbol{\Sigma}^*$ for Cudeck and Browne's (1992) method. In particular, Chun and Shapiro sought $\mathbf{E}$ to maximize $F[\boldsymbol{\Sigma}_0 + \mathbf{E}, \boldsymbol{\Sigma}_0]$ subject to $\theta_0 = \text{argmin} F[\boldsymbol{\Sigma}_0 + \mathbf{E}, \boldsymbol{\Sigma}(\cdot)]$. Given the model

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

and θ_0, $\mathbf{E}$ is unique. Once $\mathbf{E}$ is found, one can shrink the length of $\mathbf{e}$ (where $\mathbf{e} = \text{vec}\mathbf{E}$) along its direction and obtain a new vector $\tilde{\mathbf{e}} = \tau \cdot \mathbf{e}$ ($0 \leq \tau \leq 1$). Then, $F[\boldsymbol{\Sigma}_0 + \tau\mathbf{E}, \boldsymbol{\Sigma}_0]$ is a strictly decreasing function of τ, and θ_0 remains the minimizer of $F[\boldsymbol{\Sigma}_0 + \tau\mathbf{E}, \boldsymbol{\Sigma}(\cdot)]$ (Chun & Shapiro, 2010). Choosing a proper τ value will ensure $F[\boldsymbol{\Sigma}_0 + \tau\mathbf{E}, \boldsymbol{\Sigma}(\theta_0)]$ equals the desired F_{ML} value. Constructing $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_0 + \tau\mathbf{E}$ in this way ensures $\boldsymbol{\Sigma}^*$ is unique. However, one cannot directly apply Chun & Shapiro's approach to mean and covariance structure analysis. That is, for the $\boldsymbol{\beta}$ (where $\boldsymbol{\beta} = [\mathbf{t}', \mathbf{e}']'$) that maximizes $F[\mathbf{m}_0 + \boldsymbol{\beta}, \mathbf{m}_0]$ subject to $\theta_0 = \text{argmin} F[\mathbf{m}_0 + \boldsymbol{\beta}, \mathbf{m}(\cdot)]$, the minimizer of $F[\mathbf{m}_0 + \tau\boldsymbol{\beta}, \mathbf{m}(\cdot)]$ is no longer θ_0.

6. Discussion

Our method requires the following conditions: (a) The model $\boldsymbol{\Sigma}(\theta)$ is identified, and θ_0 is an interior point in the parameter space; (b) all partial derivatives of the first three orders of $\boldsymbol{\Sigma}(\theta)$ and $\boldsymbol{\mu}(\theta)$ with respect to θ are continuous and bounded in a neighborhood of θ_0; (c) $\boldsymbol{\Sigma}^*$ is positive definite; (d) $\dot{\boldsymbol{\mu}}(\theta_0)_a$ and $\dot{\boldsymbol{\Sigma}}(\theta_0)_{b,c}$ both have full rank (or $\text{rank}(\boldsymbol{\Omega}_a) = q_a^{(\text{all})}$ and $\text{rank}(\mathbf{B}) = q_b^{(\text{all})} + q_c^{(\text{all})}$ in the multigroup context); (e) $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\theta_0)]$ is positive definite. Conditions (a) to (d) are simply the mild regularity conditions commonly assumed in SEM analysis. We need them to guarantee the consistency of $\hat{\theta}_{\text{ML}}$. Condition (e) is necessary for deriving Proposition 1. Because Conditions (c) to (e) have special implications for our method, we study them more closely. Affecting Condition (c) are the values of F_{mean}, F_{cov}, $\mathbf{y}_t$, and $\mathbf{y}_E$. Although F_{mean} and F_{cov} are specified independently by the researcher, indiscreet choices of their values can sometimes cause the $\boldsymbol{\Sigma}^*$ created to be non-positive definite. If the desired misfit is somewhat small, then it does not matter how to choose F_{mean} and F_{cov}. Otherwise, the proposed method performs best when $F_{\text{mean}}/F_{\text{cov}}$ is proportional to or less than $df_{\text{mean}}/df_{\text{cov}}$. If $F_{\text{mean}}/F_{\text{cov}}$ exceeds $df_{\text{mean}}/df_{\text{cov}}$ by too much (i.e., overly large weight is given to misfit in the mean part), non-positive definite $\boldsymbol{\Sigma}^*$ tends to appear more often. In addition, initial values $\mathbf{y}_t$ and $\mathbf{y}_E$ (or $\mathbf{y}_v$ in multigroup context), especially $\mathbf{y}_t$, play a crucial role in whether the $\boldsymbol{\Sigma}^*$ obtained is positive definite. Based on our experience, we recommend randomly generating negative values for $\mathbf{y}_t$ (e.g., $Unif[-1, 0]$, $Unif[-2, 0]$) and positive values for $\mathbf{y}_E$. If the $\boldsymbol{\Sigma}^*$ created is non-positive definite, changing the initial values can easily solve the problem.

To verify Condition (d) and for constructing $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$, it needs to compute $\dot{\boldsymbol{\mu}}(\theta_0)$ and $\dot{\boldsymbol{\Sigma}}(\theta_0)$. In practice, $\dot{\boldsymbol{\mu}}(\theta_0)$ and $\dot{\boldsymbol{\Sigma}}(\theta_0)$ are usually obtained with numerical differentiation instead of analytic methods. The numerical accuracy of these derivatives plays a more important role in our method than in SEM data analysis (e.g., calculating the standard error of $\hat{\theta}_{\text{ML}}$). Although traditional numerical methods such as the central difference formula and Richardson's extrapolation (see, e.g., Linfield & Penny, 1989) are acceptable, to achieve a higher level of accuracy, we recommend the complex variable method (see Squire & Trapp, 1998). In R, numerical differentiation with complex variables is available in the "numDeriv" package (GilBert & Varadhan, 2016). Regardless of the numerical method, note software usually returns the Jacobian of $\boldsymbol{\mu}(\theta)$ or $\boldsymbol{\Sigma}(\theta)$ with respect to θ, and one needs to transpose the Jacobians to obtain the $\dot{\boldsymbol{\mu}}(\theta_0)$ and $\dot{\boldsymbol{\Sigma}}(\theta_0)$ defined in this paper. Regarding Condition (e), some SEM software (e.g., lavaan in R) has built-in functions to calculate $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\theta_0)]$. If the Hessian is not a standard output in the software, given the $\dot{\boldsymbol{\mu}}(\theta_0)$ and $\dot{\boldsymbol{\Sigma}}(\theta_0)$ obtained in Condition (d), it is also easy to calculate the Hessian manually (see "Appendix"):

$$\ddot{F}[\mathbf{m}_0, \mathbf{m}(\theta_0)] = 2\dot{\boldsymbol{\mu}}(\theta_0)\boldsymbol{\Sigma}^{-1}(\theta_0)\dot{\boldsymbol{\mu}}(\theta_0)' + \dot{\boldsymbol{\Sigma}}(\theta_0)[\boldsymbol{\Sigma}^{-1}(\theta_0) \otimes \boldsymbol{\Sigma}^{-1}(\theta_0)]\dot{\boldsymbol{\Sigma}}(\theta_0)'. \quad (33)$$

Conditions (d) and (e) generally hold in practice and are both easy to verify. In the rare occasions where a condition fails to hold, specifying a slightly different θ_0 often solves the problem.

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

Many applications of SEM include the mean structure, such as growth curve models, mixture models, and measurement invariance, but the statistical theories for these methods often assume a correct model. It is thus important to conduct simulations to study the consequences when a model is imperfect, a case always true in practice. As a simulation problem becomes more complex, it becomes increasingly difficult or even impossible to create model misfit by removing paths from a model. Currently, simulation studies on moment structure analysis often create incorrect models from the Type I perspective, and their design often requires a peculiar model form or special parameter values. Most importantly, the Type I approach fails to reflect how people define and use statistical models to understand the real world. The framework we proposed is both conceptually and mathematically more refined than the Type I approach and can help design SEM simulations in a manner not only more realistic but also more flexible.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

7. Appendix

This Appendix derives some results used in Eq. (5) and Proposition 1. In particular, Eq. (5) pertains to the first derivative of $F_{\mathrm{ML}}[\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta)]$ with respect to θ, and Proposition 1 pertains to the second derivative. Note the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are to be realized, and in general, $\boldsymbol{\mu} \neq \boldsymbol{\mu}(\theta)$ and $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}(\theta)$. The first derivative is

$$\begin{aligned}
\frac{\partial F}{\partial \theta} &= \frac{\partial \mathbf{t}'\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t} + \ln|\boldsymbol{\Sigma}(\theta)| + \mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}(\theta)]}{\partial \theta} \\
&= \frac{\partial \mathbf{t}}{\partial \theta}\{\mathbf{I}_1 \otimes [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]\} + \frac{\partial [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]}{\partial \theta}[\mathbf{t} \otimes \mathbf{I}_1] + \frac{\partial \boldsymbol{\Sigma}(\theta)}{\partial \theta}\mathrm{vec}[\boldsymbol{\Sigma}^{-1}(\theta)] + \frac{\partial \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}(\theta)}{\partial \theta}\mathrm{vec}\mathbf{I}_p \\
&= -\dot{\boldsymbol{\mu}}(\theta)\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t} - \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W}[\mathbf{I}_p \otimes \mathbf{t}]\mathrm{vec}(\mathbf{t}) - \dot{\boldsymbol{\mu}}(\theta)\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t} \\
&\quad + \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W} \cdot \mathrm{vec}\boldsymbol{\Sigma}(\theta) - \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W} \cdot \mathrm{vec}\boldsymbol{\Sigma} \\
&= -2\dot{\boldsymbol{\mu}}(\theta)\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t} - \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W} \cdot \mathrm{vec}(\mathbf{t}\mathbf{t}') - \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W} \cdot \mathrm{vec}\mathbf{E}, \qquad \text{(A1)}
\end{aligned}$$

where $\mathbf{W} = \boldsymbol{\Sigma}^{-1}(\theta) \otimes \boldsymbol{\Sigma}^{-1}(\theta)$. Accordingly, Eq. (5) in the main text follows.

To derive a result needed in Proposition 1, we continue to calculate the second derivative. In particular, the derivative of the first term in Eq. (A1) with respect to θ is as follows.

$$\begin{aligned}
&\partial[-2\dot{\boldsymbol{\mu}}(\theta)\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]/\partial\theta \\
&= \frac{\partial - 2\dot{\boldsymbol{\mu}}(\theta)}{\partial \theta}\{\mathbf{I}_q \otimes [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]\} + \frac{\partial [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]}{\partial \theta}[-2\dot{\boldsymbol{\mu}}(\theta)' \otimes \mathbf{I}_1] \\
&= -2\ddot{\boldsymbol{\mu}}(\theta)\{\mathbf{I}_q \otimes [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]\} + \left\{\frac{\partial \boldsymbol{\Sigma}^{-1}(\theta)}{\partial \theta}(\mathbf{I}_p \otimes \mathbf{t}) + \frac{\partial \mathbf{t}}{\partial \theta}[\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{I}_1]\right\}[-2\dot{\boldsymbol{\mu}}(\theta)'] \\
&= -2\ddot{\boldsymbol{\mu}}(\theta)\{\mathbf{I}_q \otimes [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]\} + 2\dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W}(\mathbf{I}_p \otimes \mathbf{t})\dot{\boldsymbol{\mu}}(\theta)' + 2\dot{\boldsymbol{\mu}}(\theta)\boldsymbol{\Sigma}^{-1}(\theta)\dot{\boldsymbol{\mu}}(\theta)' \qquad \text{(A2)}
\end{aligned}$$

The derivative of the second term in Eq. (A1) respect to θ is as follows.

$$\begin{aligned}
&\partial\{-\dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W} \cdot \mathrm{vec}(\mathbf{t}\mathbf{t}')\}/\partial\theta \\
&= \partial\{-\dot{\boldsymbol{\Sigma}}(\theta) \cdot \mathrm{vec}[\boldsymbol{\Sigma}^{-1}(\theta)(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta)]\}/\partial\theta
\end{aligned}$$

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

$$= -\ddot{\boldsymbol{\Sigma}}(\theta)\{\mathbf{I}_q \otimes \text{vec}[\boldsymbol{\Sigma}^{-1}(\theta)(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta)]\} - \frac{\partial \text{vec}[\boldsymbol{\Sigma}^{-1}(\theta)(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta)]}{\partial \theta}[\dot{\boldsymbol{\Sigma}}(\theta)' \otimes \mathbf{I}_1]. \tag{A3}$$

To proceed, we define the shorthand $\mathbf{Q}_1 = \boldsymbol{\Sigma}^{-1}(\theta)(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta)$. Then, we have

$$\begin{aligned}
&-\frac{\partial \text{vec}\mathbf{Q}_1}{\partial \theta}[\dot{\boldsymbol{\Sigma}}(\theta)' \otimes \mathbf{I}_1] \\
&= -\frac{\partial \boldsymbol{\Sigma}^{-1}(\theta)}{\partial \theta}\Big\{\mathbf{I}_p \otimes [(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta)]\Big\}\dot{\boldsymbol{\Sigma}}(\theta)' - \frac{\partial [(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta)]}{\partial \theta}[\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{I}_p]\dot{\boldsymbol{\Sigma}}(\theta)' \\
&= \dot{\boldsymbol{\Sigma}}(\theta)[\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{Q}_1]\dot{\boldsymbol{\Sigma}}(\theta)' - \frac{\partial (\mathbf{t}\mathbf{t}')}{\partial \theta}\mathbf{W}\dot{\boldsymbol{\Sigma}}(\theta)' \\
&\quad + \dot{\boldsymbol{\Sigma}}(\theta) \cdot \mathbf{W}[(\mathbf{t}\mathbf{t}')\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{I}_p] \cdot \dot{\boldsymbol{\Sigma}}(\theta)' \\
&= \dot{\boldsymbol{\Sigma}}(\theta)[\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{Q}_1]\dot{\boldsymbol{\Sigma}}(\theta)' - [\frac{\partial \mathbf{t}}{\partial \theta}(\mathbf{I}_p \otimes \mathbf{t}') + \frac{\partial \mathbf{t}}{\partial \theta}(\mathbf{t}' \otimes \mathbf{I}_p)] \cdot \mathbf{W}\dot{\boldsymbol{\Sigma}}(\theta)' \\
&\quad + \dot{\boldsymbol{\Sigma}}(\theta)[\mathbf{Q}_1 \otimes \boldsymbol{\Sigma}^{-1}(\theta)]\dot{\boldsymbol{\Sigma}}(\theta)' \\
&= \dot{\boldsymbol{\mu}}(\theta)\Big\{\boldsymbol{\Sigma}^{-1}(\theta) \otimes [\mathbf{t}'\boldsymbol{\Sigma}^{-1}(\theta)]\Big\}\dot{\boldsymbol{\Sigma}}(\theta)' + \dot{\boldsymbol{\mu}}(\theta)\Big\{[\mathbf{t}'\boldsymbol{\Sigma}^{-1}(\theta)] \otimes \boldsymbol{\Sigma}^{-1}(\theta)\Big\}\dot{\boldsymbol{\Sigma}}(\theta)' \\
&\quad + \dot{\boldsymbol{\Sigma}}(\theta)[\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{Q}_1]\dot{\boldsymbol{\Sigma}}(\theta)' + \dot{\boldsymbol{\Sigma}}(\theta)[\mathbf{Q}_1 \otimes \boldsymbol{\Sigma}^{-1}(\theta)]\dot{\boldsymbol{\Sigma}}(\theta)'.
\end{aligned} \tag{A4}$$

Similarly, the derivative of the third term in Eq. (A1) with respect to θ is as follows.

$$\begin{aligned}
&\partial\{-\dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W} \cdot \text{vec}\mathbf{E}\}/\partial\theta \\
&\quad = -\ddot{\boldsymbol{\Sigma}}(\theta)(\mathbf{I}_q \otimes \text{vec}\mathbf{Q}_2) + \dot{\boldsymbol{\Sigma}}(\theta)[\boldsymbol{\Sigma}^{-1}(\theta) \otimes \mathbf{Q}_2]\dot{\boldsymbol{\Sigma}}(\theta)' \\
&\qquad + \dot{\boldsymbol{\Sigma}}(\theta)[\mathbf{Q}_2 \otimes \boldsymbol{\Sigma}^{-1}(\theta)]\dot{\boldsymbol{\Sigma}}(\theta)' + \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W}\dot{\boldsymbol{\Sigma}}(\theta)',
\end{aligned} \tag{A5}$$

where $\mathbf{Q}_2 = \boldsymbol{\Sigma}^{-1}(\theta)\mathbf{E}\boldsymbol{\Sigma}^{-1}(\theta)$. Combining Eqs. (A2) to (A5) and rearranging the terms, we have:

$$\frac{\partial^2 F[\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta)]}{\partial\theta\partial\theta} \equiv \frac{\partial^2 F[\mathbf{m}, \mathbf{m}(\theta)]}{\partial\theta\partial\theta} = \mathbf{H}_1(\theta) + \mathbf{H}_2(\mathbf{t}, \mathbf{E}, \theta), \tag{A6}$$

where

$$\mathbf{H}_1(\theta) = 2\dot{\boldsymbol{\mu}}(\theta)\boldsymbol{\Sigma}^{-1}(\theta)\dot{\boldsymbol{\mu}}(\theta)' + \dot{\boldsymbol{\Sigma}}(\theta)\mathbf{W}\dot{\boldsymbol{\Sigma}}(\theta)', \tag{A7}$$

and

$$\begin{aligned}
\mathbf{H}_2(\mathbf{t}, \mathbf{E}, \theta) = &-\ddot{\boldsymbol{\Sigma}}(\theta)[\mathbf{I}_q \otimes \text{vec}\mathbf{Q}_1] - \ddot{\boldsymbol{\Sigma}}(\theta)[\mathbf{I}_q \otimes \text{vec}\mathbf{Q}_2] - 2\ddot{\boldsymbol{\mu}}(\theta)\{\mathbf{I}_q \otimes [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]\} \\
&+ 2\dot{\boldsymbol{\Sigma}}(\theta)\Big\{\boldsymbol{\Sigma}^{-1}(\theta) \otimes [\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{t}]\Big\}\dot{\boldsymbol{\mu}}(\theta)' \\
&+ \dot{\boldsymbol{\Sigma}}(\theta)\Big\{\boldsymbol{\Sigma}^{-1}(\theta) \otimes (\mathbf{Q}_1 + \mathbf{Q}_2)\Big\}\dot{\boldsymbol{\Sigma}}(\theta)' + \dot{\boldsymbol{\Sigma}}(\theta)\Big\{(\mathbf{Q}_1 + \mathbf{Q}_2) \otimes \boldsymbol{\Sigma}^{-1}(\theta)\Big\}\dot{\boldsymbol{\Sigma}}(\theta)' \\
&+ \dot{\boldsymbol{\mu}}(\theta)\Big\{\boldsymbol{\Sigma}^{-1}(\theta) \otimes [\mathbf{t}'\boldsymbol{\Sigma}^{-1}(\theta)]\Big\}\dot{\boldsymbol{\Sigma}}(\theta)' + \dot{\boldsymbol{\mu}}(\theta)\Big\{[\mathbf{t}'\boldsymbol{\Sigma}^{-1}(\theta)] \otimes \boldsymbol{\Sigma}^{-1}(\theta)\Big\}\dot{\boldsymbol{\Sigma}}(\theta)'.
\end{aligned} \tag{A8}$$

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press

Therefore, evaluating $\partial^2 F[\mathbf{m}, \mathbf{m}(\theta)]/\partial\theta\partial\theta$ at $\mathbf{m} = \mathbf{m}_0$ and $\theta = \theta_0$, we have $\mathbf{H}_2(\mathbf{t}_0, \mathbf{E}_0, \theta_0) = \mathbf{O}$ and $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\theta_0)] = \mathbf{H}_1(\theta_0)$, as $\mathbf{t}_0 = \boldsymbol{\mu}_0 - \boldsymbol{\mu}(\theta_0) = \mathbf{0}$ and $\mathbf{E}_0 = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}(\theta_0) = \mathbf{O}$. The assumption in Proposition 1 that $\ddot{F}[\mathbf{m}_0, \mathbf{m}(\theta_0)]$ is positive definite leads to that $\mathbf{H}_1(\theta_0)$ is positive definite. When $\mathbf{m}_k$ is sufficiently close to $\mathbf{m}_0$, by continuity arguments $\mathbf{t}_k = \boldsymbol{\mu}_k - \boldsymbol{\mu}(\theta_0)$ and $\mathbf{E}_k = \boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}(\theta_0)$ will be sufficiently close to $\mathbf{t}_0$ and $\mathbf{E}_0$, and thus, $\mathbf{H}_2(\mathbf{t}_k, \mathbf{E}_k, \theta_0)$ will also be sufficiently close to $\mathbf{H}_2(\mathbf{t}_0, \mathbf{E}_0, \theta_0)$, which is $\mathbf{O}$. Evaluated at $\mathbf{m} = \mathbf{m}_k$ and $\theta = \theta_0$, the Hessian is $\ddot{F}[\mathbf{m}_k, \mathbf{m}(\theta_0)] = \mathbf{H}_1(\theta_0) + \mathbf{H}_2(\mathbf{t}_k, \mathbf{E}_k, \theta_0)$. For any vector $\mathbf{z}$, we have $\mathbf{z}' \cdot \ddot{F}[\mathbf{m}_k, \mathbf{m}(\theta_0)] \cdot \mathbf{z} = \mathbf{z}'\mathbf{H}_1(\theta_0)\mathbf{z} + \mathbf{z}'\mathbf{H}_2(\mathbf{t}_k, \mathbf{E}_k, \theta_0)\mathbf{z} > 0$, and thus $\ddot{F}[\mathbf{m}_k, \mathbf{m}(\theta_0)]$ is also positive definite.

References

Arminger, G., & Schoenberg, R. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika*, *54*, 409–426.

Box, G. E. P. (1979a). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201–236). New York: Academic.

Box, G. E. P. (1979b). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, *74*, 1–4.

Chun, S. Y., & Shapiro, A. (2010). Construction of covariance matrices with a specified discrepancy function minimizer, with application to factor analysis. *SIAM Journal on Matrix Analysis and Applications*, *31*, 1570–1583.

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, *57*, 357–369.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. *Psychological Bulletin*, *109*, 512–519.

GilBert, P., & Varadhan, R. (2016). numDeriv: Accurate numerical derivatives [software and manual]. Retrieved January 10, 2018 from https://CRAN.R-project.org/package=numDeriv.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, *52*, 681–700.

Kano, Y. (1986). Conditions on consistency of estimators in covariance structure model. *Journal of the Japan Statistical Society*, *16*, 75–80.

Lai, K. (2018). Estimating standardized SEM parameters given nonnormal data and incorrect model: Methods and comparison. *Structural Equation Modeling*, *25*, 600–620.

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, *51*, 220–239.

Linfield, G. R., & Penny, J. E. T. (1989). *Microcomputers in numerical analysis*. New York: Halsted Press.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research*, *38*, 113–139.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502–511.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611–637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84–99.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*, 108–141.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved January 10, 2018 from http://www.R-project.org/.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 84–88.

Squire, W., & Trapp, G. (1998). Using complex variables to estimate derivatives of real functions. *SIAM Review*, *40*, 110–112.

Thissen, D. (2001). Psychometric engineering as art. *Psychometrika*, *66*, 473–486.

Thurstone, L. L. (1930). The learning function. *Journal of General Psychology*, *3*, 469–478.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421–459.

Tukey, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics*, *3*, 191–219.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica*, *57*, 307–333.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25.

Wu, H., & Browne, M. W. (2015). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*, 571–600.

Manuscript Received: 10 JAN 2018
Published Online Date: 9 JAN 2019

https://doi.org/10.1007/s11336-018-09655-0 Published online by Cambridge University Press