ON CHI-SQUARE DIFFERENCE AND z TESTS IN MEAN AND COVARIANCE STRUCTURE ANALYSIS WHEN THE BASE MODEL IS MISSPECIFIED

KE-HAI YUAN
University of Notre Dame

PETER M. BENTLER
University of California, Los Angeles

In mean and covariance structure analysis, the chi-square difference test is often applied to evaluate the number of factors, cross-group constraints, and other nested model comparisons. Let model M_a be the base model within which model M_b is nested. In practice, this test is commonly used to justify M_b even when M_a is misspecified. The authors study the behavior of the chi-square difference test in such a circumstance. Monte Carlo results indicate that a nonsignificant chi-square difference cannot be used to justify the constraints in M_b. They also show that when the base model is misspecified, the z test for the statistical significance of a parameter estimate can also be misleading. For specific models, the analysis further shows that the intercept and slope parameters in growth curve models can be estimated consistently even when the covariance structure is misspecified, but only in linear growth models. Similarly, with misspecified covariance structures, the mean parameters in multiple group models can be estimated consistently under null conditions.

Keywords: *chi-square difference; nested models; model misspecification; parameter bias; mean comparison; growth curves*

Measurements in the social and behavioral sciences are typically subject to errors. By separating measurement errors from latent constructs, structural

---

This research was supported by Grant DA01070 from the National Institute on Drug Abuse and by a Faculty Research Grant from the Graduate School at the University of Notre Dame. Correspondence concerning this article should be addressed to Ke-Hai Yuan, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556; e-mail: kyuan@nd.edu.

Educational and Psychological Measurement, Vol. 64 No. 5, October 2004 737-757
DOI: 10.1177/0013164404264853

equation modeling (SEM) provides means of modeling the latent variables directly (e.g., Bollen, 2002; MacCallum & Austin, 2000). Compared to models that do not take measurement errors into account, SEM can provide more accurate conclusions regarding the relationship between interesting attributes. To achieve such an objective, the methodology of SEM has to be appropriately used. In practice, researchers often elaborate on the substantive side of a structural model even when it barely fits the data. We will show that such a practice most likely leads to biased or misleading conclusions. Specifically, we will discuss the misuse of the chi-square difference test and the z test. For the discussion of the misuse of the chi-square difference test, we will focus on using this test in deciding the number of factors and for adding cross-group constraints. For the discussion of the misuse of the z test, we will focus on its use in evaluating the statistical significance of mean parameter estimates in the growth curve models and latent mean comparisons.

There are many indices for evaluating the adequacy of a model. Among these, only a chi-square statistic judges the model using probability as characterized by Type I and Type II errors. Although the chi-square test is limited due to its reliance on sample sizes, it is still commonly reported in applications. In practice, many reported chi-square statistics are significant even when sample sizes are not large, and in the context of nested models, the chi-square difference test is often not significant; this is used to justify model modifications or constraints across groups (e.g., Larose, Guay, & Boivin, 2002). The practice for relying on difference tests has a long history in psychometrics. For example, in the context of exploratory studies, Jöreskog (1978) stated,

> If the drop in χ^2 is large compared to the difference in degrees of freedom, this is an indication that the change made in the model is a real improvement. If, on the other hand, the drop in χ^2 is close to the difference in number of degrees of freedom, this is an indication that the improvement in fit is obtained by "capitalization on chance" and the added parameters may not have any real significance or meaning. (p. 448)

This statement may give encouragement for using the chi-square difference test to guide model modifications or adding constraints even when the less constrained model is highly significant. We will show that the difference test cannot be used reliably in this manner.

We will mainly study the misuse of statistical significance tests. In the context of multiple groups, even when a model may barely fit an individual sample, further constraints may be added across the groups. Let T_a be the statistic corresponding to the models without the constraints and T_b be the statistic corresponding to the models with constraints. Even when both T_a and T_b are statistically significant, implying rejection of both models, the difference $\Delta T = T_b - T_a$ can still be nonsignificant. This is often used to justify the cross-group constraints in practice. See Drasgow and Kanfer (1985), Brouwers and Tomic (2001), and Vispoel, Boo, and Bleiler (2001) for such applications. Similarly, a model with two factors may correspond to a significant statistic T_a whereas the substantive theory may support only a one-factor model. The one-factor model may have a significant statistic T_b. In such a context, many researchers regard the one-factor model as "attainable" if $\Delta T = T_b - T_a$ is not statistically significant at the .05 level. In the context of latent growth curves and latent mean comparisons, there are mean structures in addition to covariance structures. These models are nested within the covariance structure models with saturated means. The statistic T_a corresponding to only the covariance structure may be already highly statistically significant. Adding a mean structure generally makes the overall model even more statistically significant, that is, less fitting. Nonetheless, researchers still elaborate on the significance of the intercept or slope estimates or significant mean differences as evaluated by z tests.

Let the model M_a be the base model within which model M_b is nested. When M_a is an adequate model as reflected by a nonsignificant T_a and supported by other model fit indices, one may want to test the further restricted model M_b. If $\Delta T = T_b - T_a$ is not statistically significant, M_b is generally preferred because it is more parsimonious. When M_a is not adequate as indicated by a significant T_a, can we still justify M_b by a nonsignificant ΔT? Although there exist statistical theories (Steiger, Shapiro, & Browne, 1985) in this context, and wide applications (e.g., Brouwers & Tomic, 2001; Drasgow & Kanfer, 1985; Vispoel et al., 2001) justify M_b using nonsignificant ΔTs, in our view, the effect of such a practice on the substantive aspect of SEM is not clear. A related question is when the overall model is misspecified, can a test be used to indicate the statistical significance of a parameter estimate? Examples in this direction include whether the intercept and slope parameters in a latent growth curve model are zeros, whether the means are different in latent mean comparisons, and whether a parameter should be freed or fixed as in model modifications. The interest here is to study the effect of misspecified models on ΔT and the z tests. By simulation, the second section studies the behavior of ΔT when M_a is misspecified. The third section explores the reason why ΔT does not perform properly when M_a is misspecified. Detailed results show that a misspecified model leads to biased parameters, which explains why model inferences based on ΔT and parameter inference based on the z test actually can be quite misleading.

Chi-Square Difference Test When the Base Model Is Misspecified

Jöreskog (1971) and Lee and Leung (1982) recommended using the chi-square difference test for cross-group constraints in analyzing multiple samples. Under some standard regularity conditions, Steiger et al. (1985) proved that the chi-square difference statistic asymptotically follows a noncentral chi-square distribution (see also Satorra & Saris, 1985). Chou and Bentler (1990) studied the chi-square difference test when M_a is correctly specified and found that it performs the best compared to the Lagrange multiplier test and the Wald test in identifying omitted parameters. The chi-square difference test has been widely used in SEM, essentially in every application of SEM with multiple groups. However, how to appropriately apply the chi-square difference test in practice is not clear at all. Paradoxes readily occur, for example, a nonsignificant T_a and a nonsignificant $\Delta T = T_b - T_a$ cannot guarantee a statistically nonsignificant T_b. Although $T_a = 3.84 \sim \chi_1^2$ is statistically nonsignificant at the .05 level and $\Delta T_b = 3.84 \sim \chi_1^2$ is statistically nonsignificant at the .05 level, $T_b = 7.68 \sim \chi_2^2$ is statistically significant at .05 level. Another paradox occurs when sequential application of nonsignificant ΔT may lead to a highly significant final model. The general point is that when ΔT is not statistically significant, one may derive the conclusion that M_b is less misspecified than M_a. However, we will show that this is not necessarily the case. In this section, we will show the effect of a misspecified base model M_a on the significance of ΔT through three simulation studies. Because the normal theory-based likelihood ratio statistic T_{ML} is commonly used in practice, we study only the performance of ΔT based on this statistic for simulated normal data. When data are not normal or when another statistic is used in practice, one cannot expect ΔT to perform better.

Type II Error of ΔT in Deciding the Number of Factors

We first study using ΔT to judge the number of factors in a confirmatory factor model. Using ΔT to decide the number of factors in the exploratory factor model was recommended by Lawley and Maxwell (1971). It is also commonly applied when confirmatory factor analysis is used for scale development.

Let us consider a confirmatory factor model with five manifest variables and two latent factors. The population is generated by

$$\mathbf{x} = \boldsymbol{\mu}_0 + \boldsymbol{\Lambda}_0\mathbf{f} + \mathbf{e}$$

with

$$E(\mathbf{x}) = \boldsymbol{\mu}_0,\ \mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}_0 = \boldsymbol{\Lambda}_0\boldsymbol{\Phi}_0\boldsymbol{\Lambda}_0' + \boldsymbol{\Psi}_o, \tag{1}$$

where

$$\boldsymbol{\Lambda}_0 = \begin{pmatrix} .700 & .790 & 0 & 0 & 0 \\ 0 & 0 & .926 & .774 & .725 \end{pmatrix}',\ \boldsymbol{\Phi}_0 = \begin{pmatrix} 1.0 & .818 \\ .818 & 1.0 \end{pmatrix},$$

$\psi_{150} = \psi_{510} = 0.285$, and the diagonal elements of $\boldsymbol{\Psi}_0$ are adjusted so that $\boldsymbol{\Sigma}_0$ is a correlation matrix. Note that the subscript 0 is used to denote the population value of a parameter. The corresponding model parameter without the subscript 0 is subject to estimation before its value can be obtained. Except for ψ_{150}, the population parameter values for the model defined in Equation 1 are obtained from fitting the two-factor model to the open-closed book data set in Table 1.2.1 of Mardia, Kent, and Bibby (1979). The purposes of choosing this set of population values are (a) they are represented by real data and thus realistic, (b) $\phi_{120} = 0.818$ is large enough so that ΔT will not be able to judge the correct number of factors when M_a is misspecified.

Let the covariance structure model be

$$\mathbf{M}(\boldsymbol{\theta}) = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}\ ,$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{21} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{32} & \lambda_{42} & \lambda_{52} \end{pmatrix}',\ \boldsymbol{\Phi} = \begin{pmatrix} 1.0 & \phi_{12} \\ \phi_{21} & 1.0 \end{pmatrix}$$

and $\boldsymbol{\Psi}$ is a diagonal matrix. Because we ignore the covariance ψ_{15}, the above two-factor model is no longer correct for the population covariance matrix in Equation 1. Of course, the one-factor model excluding ψ_{15} is not correct either. In such a circumstance, however, a researcher may be tempted in practice to justify the one-factor model by a nonsignificant ΔT. We next evaluate the effect of ignoring ψ_{15} on ΔT for such a purpose.

Without a mean structure, there is only one degree of freedom difference between M_a (the one-factor model) and M_b (the two-factor model). We refer ΔT to the 95th percentile of χ_1^2 for statistical significance. With 500 replications, Table 1 contains the number of replications with nonsignificant ΔT. For comparison purposes, we also include the performance of ΔT when ψ_{15} is explicitly included in both M_a and M_b. When ψ_{15} is excluded, although the one-factor model is inadequate due to a misspecification, ΔT cannot reject the one-factor model more than 50% at sample size $n = 100$. With correct model specification in M_a, ΔT has a much greater power to reject the wrong model M_b.

Type II Error of ΔT in Testing Invariance in Factor Pattern Coefficients

With a misspecified base model M_a, the statistic ΔT not only loses its power with smaller sample sizes but also may have a weak power even with very large sample sizes. We will illustrate this through a two-group comparison.

Table 1
Number of Nonsignificant $\Delta T = T_b - T_a$ *(Type II Error) Out of 500 Replications: One-Factor Model* (M_a) *Versus Two-Factor Model* (M_b)

Sample Size	Misspecified M_a and M_b	Correct M_a and Misspecified M_b
50	363	176
100	276	48
200	148	1
300	60	0
400	34	0
500	8	0

Consider two groups, each with four manifest variables that are generated by a one-factor model. The population covariance matrix $\mathbf{\Sigma}_{10}$ of the first group is generated by

$$\mathbf{x}_1 = \boldsymbol{\mu}_{10} + \boldsymbol{\lambda}_{10} f_1 + \mathbf{e}_1,$$

where

$$\boldsymbol{\lambda}_{10} = (1, 0.80, 0.50, 0.40)', \text{ Var}(f_1) = \phi_0^{(1)} = 1.0, \text{ Cov}(\mathbf{e}_1) = \boldsymbol{\Psi}_{10} = (\psi_{ij0}^{(1)}),$$

with $\psi_{110}^{(1)} = \psi_{220}^{(1)} = 1.0, \psi_{330}^{(1)} = 1.24, \psi_{440}^{(1)} = 1.09, \psi_{140}^{(1)} = .32,$ and $\psi_{240}^{(1)} = .25.$ The population covariance matrix $\mathbf{\Sigma}_{20}$ of the second group is generated by

$$\mathbf{x}_2 = \boldsymbol{\mu}_{20} + \boldsymbol{\lambda}_{20} f_2 + \mathbf{e}_2,$$

where

$$\boldsymbol{\lambda}_{20} = (1, .80, .70, .80)', \text{ Var}(f_2) = \phi_0^{(2)} = 1.0, \text{ Cov}(\mathbf{e}_2) = \boldsymbol{\Psi}_{20} = (\psi_{ij0}^{(2)}),$$

with $\psi_{110}^{(2)} = \psi_{220}^{(2)} = \psi_{330}^{(2)} = \psi_{440}^{(2)} = 1.0,$ and $\psi_{340}^{(2)} = -.559.$ It is obvious that the two groups do not have invariant factor pattern coefficients. In model M_a, the one-factor model $\mathbf{M}(\boldsymbol{\theta}) = \boldsymbol{\lambda}\phi\boldsymbol{\lambda}' + \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix, is fitted to a normal sample from each of the populations corresponding to $\mathbf{\Sigma}_{10}$ and $\mathbf{\Sigma}_{20}$ and the statistic T_a is the sum of the two T_{ML}s. The first factor pattern coefficient was set at 1.0 for identification purposes. In model M_b, the three-factor pattern coefficients as well as the factor variances were set equal across the two groups, which results in the statistic T_b. Referring $\Delta T = T_b - T_a$ to the 95th percentile of χ_1^2, the number of nonsignificant replications are given in the middle column of Table 2. For the purpose of comparison, a parallel study in which the three error covariances are included in $\boldsymbol{\Psi}$ in both M_a and M_b was also performed, and the corresponding results are in the last column of Table 2.

Table 2
Number of Nonsignificant $\Delta T = T_b - T_a$ (Type II Error) Out of 500 Replications: Incorrect Equality Constraints Across Two-Group Factor Pattern Coefficients

Sample Size ($n_1 = n_2$)	Misspecified M_a and M_b	Correct M_a and Misspecified M_b
100	447/494[a]	342/496
200	444/497	239
300	417	131
400	402	66
500	387	23
1,000	329	1
3,000	66	0

a. Converged solutions out of 500 replications.

When ignoring the error covariances, only 494 replications out of the 500 converged when $n_1 = n_2 = 100$, and 497 replications converged when $n_1 = n_2 = 200$. When error covariances were accounted for, 496 replications converged when $n_1 = n_2 = 100$. The number of nonsignificant replications is based on the converged replications only. When the base model is misspecified, although the power for ΔT to reject the incorrect constraints increases as sample sizes increase, the speed is extremely slow. Even when $n_1 = n_2 = 1{,}000$, more than 60% of the replications could not reject the incorrect constraints. When M_a is correctly specified, the statistic ΔT has a power greater than 0.95 in rejecting the incorrect constraints at sample size $n_1 = n_2 = 500$.

Type I Error of ΔT in Testing Invariance in Factor Pattern Coefficients

A misspecified M_a not only leads to attenuated power for the chi-square difference test, it can also lead to inflated Type I errors, as illustrated in the following two-group comparison.

Again, consider two groups, each with four manifest variables that are generated by a one-factor model. The first group, $\mathbf{\Sigma}_{10} = \text{Cov}(\mathbf{x}_1)$, is generated by

$$\mathbf{x}_1 = \boldsymbol{\mu}_{10} + \boldsymbol{\lambda}_{10} f_1 + \mathbf{e}_1,$$

where

$$\boldsymbol{\lambda}_{10} = (1, .80, .70, .50)', \text{Var}(f_1) = \phi_0^{(1)} = 1.0, \text{Cov}(\mathbf{e}_1) = \mathbf{\Psi}_{10} = (\psi_{ij0}^{(1)})$$

with $\psi_{110}^{(1)} = \psi_{220}^{(1)} = \psi_{330}^{(1)} = \psi_{440}^{(1)} = 1.0$, $\psi_{140}^{(1)} = .70$ and $\psi_{240}^{(1)} = .30$. The second group, $\mathbf{\Sigma}_{20} = \text{Cov}(\mathbf{x}_2)$, is generated by

$$\mathbf{x}_2 = \boldsymbol{\mu}_{20} + \boldsymbol{\lambda}_{20} f_2 + \mathbf{e}_2,$$

where

$$\boldsymbol{\lambda}_{20} = (1, .80, .70, .50)', \text{Var}(f_2) = \phi_0^{(2)} = 1.0, \text{Cov}(\mathbf{e}_2) = \boldsymbol{\Psi}_{20} = (\psi_{ij0}^{(2)}),$$

with $\psi_{110}^{(2)} = \psi_{220}^{(2)} = \psi_{330}^{(2)} = \psi_{440}^{(2)} = 1.0$, and $\psi_{340}^{(2)} = -.25$. Now, the two groups have invariant factor pattern coefficients and factor variances. We want to know whether ΔT can endorse the invariance when M_a is misspecified. Let the three error covariances be ignored in M_a when fitting the one-factor model to both samples and M_b be the model in which the factor pattern coefficients and factor variances are constrained equal. Instead of reporting the nonsignificant replications of ΔT, we report the significant ones in Table 3. When M_a is misspecified, ΔT is not able to justify the cross-group constraints. As indicated in Table 3, even when $n_1 = n_2 = 100$, more than 70% of the equal factor pattern coefficients and factor variances are rejected. When the error covariances were accounted for in M_a and M_b, Type I errors are around the nominal level of 5% for all the sample sizes in Table 3.

In summary, when the base model M_a is misspecified, the chi-square difference test cannot control either the Type I errors or the Type II errors for realistic sample sizes. Conclusions based on ΔT are misleading. For the simulation results in Tables 1 to 3, we did not distinguish the significant T_as from those that are not significant. Some of the nonsignificant ΔTs in Tables 1 and 2 have nonsignificant T_as, and some of the significant ΔTs in Table 3 also correspond to nonsignificant T_as. As was discussed at the beginning of this section, even when both T_a and ΔT are not significant at the .05 level, we are unable to control the errors of inference regarding model M_b. When constraints across groups hold partially, Kaplan (1989) studied the performance of T_{ML}, which is essentially the T_b here. The results in Tables 1 to 3 are not in conflict with Kaplan's results, which indicate that T_b has a nice power in detecting misspecifications. Actually, both T_a and T_b can also be regarded as chi-square difference tests due to M_a and M_b being nested within the saturated model M_s. Because M_s is always correctly specified, T_a and T_b do not possess the problems discussed above.

Because ΔT, the Lagrange multiplier, and the Wald tests are asymptotically equivalent (Buse, 1982; Lee, 1985; Satorra, 1989), the results in Tables 1 to 3 may also imply that the two other tests cannot perform well in similar circumstances. All of these tests are used in model modification, and our results may explain some of the poor performance of empirically based model modification methods (e.g., MacCallum, 1986).

Steiger et al. (1985) showed that chi-square differences in sequential tests are asymptotically independent, and each difference follows a noncentral chi-square even when M_a is misspecified. The results in this section imply

Table 3
Number of Significant $\Delta T = T_b - T_a$ *(Type I Error) Out of 500 Replications: Correct Equality Constraints Across Two-Group Factor Pattern Coefficients*

Sample Size ($n_1 = n_2$)	Misspecified M_a and M_b	Correct M_a and M_b
100	362/497[a]	25
200	481	28
300	498	23
400	500	25

a. Converged solutions out of 500 replications.

that (a) when the base model M_a is wrong and the constraints that differentiate M_a and M_b are substantially incorrect, the noncentrality parameter of the chi-square difference can be tiny so that ΔT loses its power and (b) when the base model M_a is wrong and the constraints that differentiate M_a and M_b are correct, the noncentrality parameter of the chi-square difference can be substantial so that ΔT always rejects the correct hypothesis. The next section explains why the noncentrality parameter is tiny or substantial due to misspecifications.

The Effect of a Misspecified Model on Parameters

In this section, we explain why the chi-square difference test is misleading when the base model is misspecified. Specifically, when a model is misspecified, parameter estimates converge to different values from those of a correctly specified model. Thus, equal parameters in a correctly specified model become unequal in a misspecified model. Consequently, ΔT for testing constraints will be misleading. In the context of mean structures, rather than using a chi-square statistic to evaluate the overall model, researchers often use z tests to evaluate the statistical significance of mean parameter estimates (see Hong, Malik, & Lee, 2003; Whiteside-Mansell & Corwyn, 2003). We will also show the effect of a misspecified model in evaluating the mean parameters. We use results in Yuan and Bentler (2004) for this purpose.

Let $\bar{\mathbf{x}}$ and $\mathbf{S}$ be the sample mean vector and sample covariance matrix from a p-variate normal distribution $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Let $\boldsymbol{\nu}^*(\boldsymbol{\gamma})$ and $\mathbf{M}^*(\boldsymbol{\gamma})$ be the correct mean and covariance structure; thus, there exists a vector $\boldsymbol{\gamma}_0$ such that $\boldsymbol{\mu}_0 = \boldsymbol{\nu}^*(\boldsymbol{\gamma}_0)$ and $\boldsymbol{\Sigma}_0 = \mathbf{M}^*(\boldsymbol{\gamma}_0)$. Let the misspecified model be $\boldsymbol{\nu}(\boldsymbol{\theta})$ and $\mathbf{M}(\boldsymbol{\theta})$. We assume that the misspecification is due to model $\boldsymbol{\nu}(\boldsymbol{\theta})$ and $\mathbf{M}(\boldsymbol{\theta})$ missing parameters $\boldsymbol{\delta}$ of $\boldsymbol{\gamma} = (\boldsymbol{\theta}', \boldsymbol{\delta}')'$. In the context of mean and covariance structure analysis, one obtains the normal theory-based maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ by minimizing (see, e.g., Browne & Arminger, 1995)

$$F_{ML}(\boldsymbol{\theta}, \bar{\mathbf{x}}, \mathbf{S}) = [\bar{\mathbf{x}} - \boldsymbol{\nu}(\boldsymbol{\theta})]'\mathbf{M}^{-1}(\boldsymbol{\theta})[\bar{\mathbf{x}} - \boldsymbol{\nu}(\boldsymbol{\theta})] + \mathrm{tr}[\mathbf{S}\mathbf{M}^{-1}(\boldsymbol{\theta})] - \log|\mathbf{S}\mathbf{M}^{-1}(\boldsymbol{\theta})| - p.$$

Under some standard regularity conditions (e.g., Kano, 1986; Shapiro, 1984), $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}^*$, which minimizes $F_{ML}(\boldsymbol{\theta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Note that in general, $\boldsymbol{\theta}^*$ does not equal its counterpart $\boldsymbol{\theta}_0$ in $\boldsymbol{\gamma}_0 = (\boldsymbol{\theta}_0', \boldsymbol{\delta}_0')'$, which is the population value of the correctly specified model. We will call $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}^* - \boldsymbol{\theta}_0$ the bias in $\boldsymbol{\theta}^*$, which is also the asymptotic bias in $\hat{\boldsymbol{\theta}}$. It is obvious that if the sample is generated by $\boldsymbol{\mu}^0 = \boldsymbol{\nu}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}^0 = \mathbf{M}(\boldsymbol{\theta}_0)$, then $\boldsymbol{\theta}^*$ will have no bias. We may regard the true population $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ as a perturbation to $(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$. Because of the perturbation, $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, although some parameters in $\boldsymbol{\theta}^*$ can still equal the corresponding ones in $\boldsymbol{\theta}_0$ (see Yuan, Marshall, & Bentler, 2003). Yuan et al. (2003) studied the effect of the misspecified model on parameter estimates in covariance structure analysis. Extending their result to mean and covariance structure models, Yuan and Bentler (2004) characterized $\boldsymbol{\theta}$ as a function of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in a neighborhood of $(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$. Denote this function as $\boldsymbol{\theta} = \mathbf{g}(\boldsymbol{\mu}, \boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ is the vector containing the nonduplicated elements of $\boldsymbol{\Sigma}$. Then there approximately exists

$$\Delta\boldsymbol{\theta} \approx \dot{\mathbf{g}}_1(\boldsymbol{\mu}^0, \boldsymbol{\sigma}^0)\Delta\boldsymbol{\mu} + \dot{\mathbf{g}}_2(\boldsymbol{\mu}^0, \boldsymbol{\sigma}^0)\Delta\boldsymbol{\sigma}, \tag{2}$$

where $\dot{\mathbf{g}}_1$ is the partial derivative of $\mathbf{g}$ with respect to $\boldsymbol{\mu}$ and $\dot{\mathbf{g}}_2$ is the partial derivative of $\mathbf{g}$ with respect to $\boldsymbol{\sigma}$; $\Delta\boldsymbol{\mu}_0 = \boldsymbol{\mu}_0 - \boldsymbol{\mu}^0$ and $\Delta\boldsymbol{\sigma} = \boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^0$. Explicit expressions of $\dot{\mathbf{g}}_1$ and $\dot{\mathbf{g}}_2$ are given in Yuan and Bentler (2004). Equation 2 implies that the bias in $\boldsymbol{\theta}^*$ caused by $\Delta\boldsymbol{\mu}$ and $\Delta\boldsymbol{\sigma}$ are approximately additive. Let q be the number of free parameters in $\boldsymbol{\theta}$, and then $\dot{\mathbf{g}}_1(\boldsymbol{\mu}^0, \boldsymbol{\sigma}^0)$ is a $q \times p$ matrix and $\dot{\mathbf{g}}_2(\boldsymbol{\mu}^0, \boldsymbol{\sigma}^0)$ is a $q \times p^*$ matrix, where $p^* = p(p+1)/2$. For the lth parameter θ_l, we can rewrite Equation 2 as

$$\Delta\theta_l \approx \sum_{i=1}^{p} c_{li}\Delta\mu_i + \sum_{i=1}^{p}\sum_{j=1}^{p} c_{lij}\Delta\sigma_{ij}. \tag{3}$$

When the parameter is clear, we will omit the subscript in reporting the coefficients in examples.

Now we can use the result in Equation 2 or Equation 3 to explain the misleading behavior of ΔT when M_a is misspecified. Because of the misspecification, $\boldsymbol{\theta}^*$ may not equal $\boldsymbol{\theta}$. Most nested models can be formulated by imposing constraints $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$. When $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$, $\mathbf{h}(\boldsymbol{\theta}^*)$ may not equal zero. With a misspecified M_a, it is the constraints $\mathbf{h}(\boldsymbol{\theta}^*) = \mathbf{0}$ that are being tested by ΔT. Because $\mathbf{h}(\boldsymbol{\theta}^*) \neq \mathbf{0}$, T_b will be significantly greater than T_a, and thus ΔT tends to be statistically significant as reflected in Table 3. Similarly, when $\mathbf{h}(\boldsymbol{\theta}_0)$ does not equal zero, $\mathbf{h}(\boldsymbol{\theta}^*)$ may approximately equal zero. Consequently, the power for ΔT to reject $\mathbf{h}(\boldsymbol{\theta}^*) = \mathbf{0}$ is low, as reflected in Tables 1 and 2. However, researchers in practice treat $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$ as plausible.

In general, it is difficult to control the two types of errors by ΔT when M_a is misspecified. If treating ΔT as if M_a were correctly specified when it is actually not, the conclusion regarding $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$ will be misleading. For example, the ΔT that produced the results in Table 1 tests whether $\phi_{120} = 1$. When ignoring ψ_{15} in $\mathbf{M}(\boldsymbol{\theta})$, using Equation 3 and the population parameter values in Table 1, we have $\Delta\phi_{12} \approx 0.166\ \Delta\sigma_{15} = 0.166 \times 0.285 = 0.047$. This leads to $\phi_{12}^* \approx 0.865$, which is closer to 1.0 than $\phi_{120} = 0.818$. Actually, any positive perturbation on σ_{ij}, $i = 1,2$; $j = 3,4,5$ will cause a positive bias in ϕ_{12}^*, as illustrated in the following example.

Example 1

Let $\boldsymbol{\theta}_0$ be the population parameter values of the model in Equation 1 excluding ψ_{150}; evaluating Equation 3 at $\boldsymbol{\theta}_0$, we obtain the coefficients c_{ij} for the approximate bias $c_{ij}\Delta\sigma_{ij}$ of ϕ_{12}^* in Table 4. For purposes of comparison, the exact biases when $\Delta\sigma_{ij} = 0.05$, 0.10, and 0.20 were also computed by minimizing $F_{ML}(\boldsymbol{\theta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ directly. The approximate biases $c_{ij}\Delta\sigma_{ij}$ are very close to the exact ones when $\Delta\sigma_{ij} = 0.05$. The accuracy of the approximation decreases as the amount of perturbation $\Delta\sigma_{ij}$ increases. This is because Equation 2 is based on a local linearization. The smallest c_{ij} is with σ_{45}, implying that the function $\phi_{12} = \phi_{12}(\boldsymbol{\Sigma})$ is quite flat in the direction of σ_{45}. The direction obtained at this point is usually not stable. Actually, the $c_{45}\Delta\sigma_{45}$ predicts a small positive bias in ϕ_{12}^* when $\Delta\sigma_{45} = 0.10$ or 0.20, but the actual biases are negative. Except for this element, the predicted biases and the actual biases agree reasonably well for perturbations on all the other covariances σ_{ij}. Notice that positive perturbations on the covariances between indicators for different factors (σ_{ij}, $i = 1,2$; $j = 3,4,5$) lead to an inflated ϕ_{12}^*. Perturbations in the opposite direction will lead to an attenuated ϕ_{12}^*. So the estimate $\hat{\phi}_{12}$ and any testing for $\phi_{120} = 0$ or 1 based on $\hat{\phi}_{12}$ are not trustworthy when model M_a is misspecified, especially when ϕ_{120} is near 0 or 1.0.

Similarly, due to the changes in parameters, the chi-square difference test for the equivalent constraints across groups is misleading when either of the models does not fit the data within a group. Instead of providing more examples about the bias on factor pattern coefficients when σ_{ij} are perturbed, we illustrate the effect of a misspecified model on the mean parameters in simultaneously modeling mean and covariance structures.

Let $\mathbf{y} = (y_1, y_2, \ldots, y_p)'$ be repeated measures at p time points. Then a latent growth curve model can be expressed as (Curran, 2000; Duncan, Duncan, Strycker, Li, & Alpert, 1999; McArdle & Epstein, 1987; Meredith & Tisak, 1990)

$$\mathbf{y} = \boldsymbol{\Lambda}\mathbf{f} + \mathbf{e}, \tag{4}$$

Table 4
The Effect of a Perturbation $\Delta\sigma_{ij}$ *on Factor Correlation* ϕ^*_{12}

		$\Delta\sigma_{ij} = 0.05$		$\Delta\sigma_{ij} = 0.10$		$\Delta\sigma_{ij} = 0.20$	
σ_{ij}	c_{ij}	$c_{ij} \times \Delta\sigma_{ij}$	$\Delta\phi_{12}$	$c_{ij} \times \Delta\sigma_{ij}$	$\Delta\phi_{12}$	$c_{ij} \times \Delta\sigma_{ij}$	$\Delta\phi_{12}$
σ_{12}	–0.740	–0.037	–0.035	–0.074	–0.065	–0.148	–0.117
σ_{13}	0.464	0.023	0.023	0.046	0.044	0.093	0.068
σ_{14}	0.212	0.011	0.011	0.021	0.024	0.042	0.056
σ_{15}	0.166	0.008	0.009	0.017	0.018	0.033	0.041
σ_{23}	0.363	0.018	0.017	0.036	0.031	0.073	0.032
σ_{24}	0.221	0.011	0.012	0.022	0.025	0.044	0.058
σ_{25}	0.173	0.009	0.009	0.017	0.019	0.035	0.044
σ_{34}	–0.362	–0.018	–0.020	–0.036	–0.042	–0.072	–0.090
σ_{35}	–0.305	–0.015	–0.018	–0.030	–0.040	–0.061	–0.092
σ_{45}	0.097	0.005	0.002	0.010	–0.002	0.019	–0.046

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1.0 & 1.0 & 1.0 & \ldots & 1.0 \\ 0 & 1.0 & \lambda_1 & \ldots & \lambda_{p-2} \end{pmatrix}',$$

$\mathbf{f} = (f_1, f_2)'$, with f_1 being the latent intercept and f_2 being the latent slope, $\boldsymbol{\mu}_f = E(\mathbf{f}) = (\alpha, \beta)'$,

$$\boldsymbol{\Phi} = \text{Cov}(\mathbf{f}) = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

and $\text{Cov}(\mathbf{e}) = \boldsymbol{\Psi} = \text{diag}(\psi_{11}, \psi_{22}, \ldots, \psi_{pp})$. This setup leads to the following mean and covariance structures:

$$\boldsymbol{\nu}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\mu}_f,\ \mathbf{M}(\boldsymbol{\theta}) = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}.$$

In fitting such a model in practice, researchers often need to elaborate on the significance of the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$, although the overall model fit is typically significant as judged by a chi-square statistic. If the misspecification affects the mean structure to such a degree that the significances of $\hat{\alpha}$ and $\hat{\beta}$ are due to only a systematic bias, then caution is needed to specify the model before meaningful $\hat{\alpha}$ and $\hat{\beta}$ can be obtained. We will consider the models for both linear growth and nonlinear growth.

Example 2

When letting $\lambda_1 = 2, \lambda_2 = 3, \ldots, \lambda_{p-2} = p-1$, Equation 4 describes the linear growth model. The unknown parameters in this model are

$$\boldsymbol{\theta} = (\alpha, \beta, \phi_{11}, \phi_{21}, \phi_{22}, \psi_{11}, \ldots, \psi_{pp})'.$$

Detailed calculation (see Yuan & Bentter, 2004) shows that all the c_{1ij}s and c_{2ij}s in Equation 3 are zero. So there is no effect of misspecification in $\mathbf{M}(\boldsymbol{\theta})$ on α^* and β^*. This implies that we can still get consistent parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ when $\boldsymbol{\nu}(\boldsymbol{\theta})$ is correctly specified even if $\mathbf{M}(\boldsymbol{\theta})$ is misspecified.

However, the misspecification in $\boldsymbol{\nu}(\boldsymbol{\theta})$ does have an effect on α^* and β^* as presented in Table 5 using $p = 4$, where Equation 3 was evaluated at

$$\alpha_0 = 1, \beta_0 = 1, \phi_{110} = \phi_{220} = 1.0, \phi_{120} = 0.5, \psi_{110} = \ldots = \psi_{pp0} = 1.0,$$

and the perturbation was set at $\Delta\mu_i = 0.2$. The positive perturbations $\Delta\mu_1$ and $\Delta\mu_2$ cause positive biases on α^* but negative biases on β^*. The positive perturbation $\Delta\mu_4 = 0.2$ causes a negative bias on α^* but a positive bias on β^*. Because $\boldsymbol{\nu}(\boldsymbol{\theta})$ is a linear model, the approximate biases given by Equation 2 or Equation 3 are identical to the corresponding exact ones.

When the trend in $\boldsymbol{\mu}_0 = E(\mathbf{y})$ cannot be described by a linear model, a nonlinear model may be more appropriate. However, any misspecification in $\mathbf{M}(\boldsymbol{\theta})$ will affect the α^* and β^* as illustrated in the following example.

Example 3

When $\lambda_1, \lambda_2, \ldots, \lambda_{p-2}$ are free parameters, Equation 4 subjects the shape of growth to estimation. The unknown parameters in this model are

$$\boldsymbol{\theta} = (\alpha, \beta, \lambda_1, \ldots, \lambda_{p-2}, \phi_{11}, \phi_{21}, \phi_{22}, \psi_{11}, \ldots, \psi_{pp})'.$$

Because the λ_is are in both $\boldsymbol{\nu}(\boldsymbol{\theta})$ and $\mathbf{M}(\boldsymbol{\theta})$, misspecification in $\mathbf{M}(\boldsymbol{\theta})$ will cause biases in α^* and β^*. To illustrate this, let us consider a population that is generated by Equation 4 with

$$\alpha_0 = 1, \beta_0 = 1, \lambda_{j0} = j + 1, \phi_{110} = \phi_{220} = 1.0, \phi_{120} = 0.5, \psi_{110} = \ldots = \psi_{pp0} = 1.0$$

and $p = 4$. Table 6 gives the approximate biases in α^* and β^* as described in Equation 2 and Equation 3 when $\Delta\mu_i = 0.2$ or $\Delta\sigma_{ij} = 0.2$ whereas the remaining elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are fixed at $\boldsymbol{\theta}_0$ as specified above. When μ_i is perturbed, the changes in α^* and β^* are no longer linear functions of $\Delta\mu_i$, and the approximate biases given in Equations 2 or 3 are no longer identical to the

Table 5
The Effect of a Perturbation $\Delta\mu_i = 0.2$ on the Intercept α^* and Slope β^* for the Linear Growth Curve Model

	α			β		
μ_i	c_i	$c_i \times \Delta\mu_i$	$\Delta\alpha$	c_i	$c_i \times \Delta\mu_i$	$\Delta\beta$
μ_1	0.700	0.140	0.140	–0.300	–0.060	–0.060
μ_2	0.400	0.080	0.080	–0.100	–0.020	–0.020
μ_3	0.100	0.020	0.020	0.100	0.020	0.020
μ_4	–0.200	–0.040	–0.040	0.300	0.060	0.060

Table 6
The Effect of a Perturbation $\Delta\mu_i = 0.2$ or $\Delta\sigma_{ij} = 0.2$ on the Intercept α^* and Slope β^* for the Nonlinear Growth Curve Model

	α			β		
μ_i	c_i	$c_i \times \Delta\mu_i$	$\Delta\alpha$	c_i	$c_i \times \Delta\mu_i$	$\Delta\beta$
μ_1	0.869	0.174	0.171	–0.684	–0.137	–0.131
μ_2	0.185	0.037	0.035	0.487	0.097	0.097
μ_3	0.021	0.004	0.002	0.078	0.016	0.018
μ_4	–0.076	–0.015	–0.015	0.119	0.024	0.023
σ_{ij}	c_{ij}	$c_{ij} \times \Delta\sigma_{ij}$	$\Delta\alpha$	c_{ij}	$c_{ij} \times \Delta\sigma_{ij}$	$\Delta\beta$
σ_{12}	0.020	0.004	0.003	–0.032	–0.006	–0.007
σ_{13}	0.013	0.003	0.002	–0.017	–0.003	–0.003
σ_{14}	0.020	0.004	0.005	–0.064	–0.013	–0.013
σ_{23}	–0.079	–0.016	–0.016	0.123	0.025	0.023
σ_{24}	0.001	0.000	0.000	0.063	0.013	0.011
σ_{34}	0.026	0.005	0.005	–0.073	–0.015	–0.015

exact biases $\Delta\alpha$ or $\Delta\beta$. This occurs even though the population mean vector and covariance matrix are identical to those in Example 2. The $\Delta\alpha$ or $\Delta\beta$ in Table 6 do not equal the corresponding ones in Table 5 due to the nonlinear nature of the model.

We need to notice that when $\boldsymbol{\mu}_0 = E(\mathbf{y}) = \mathbf{0}$ in the population, there is no effect of a misspecified $\mathbf{M}(\boldsymbol{\theta})$ on α^* or β^*. This can be see from the form of $F_{ML}(\boldsymbol{\theta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. With any given $\mathbf{M}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_0$, when $\boldsymbol{\mu}_0 = \mathbf{0}$, the minimum of $F_{ML}(\boldsymbol{\theta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ is at $\alpha^* = \beta = 0$.

We next consider comparing factor means across groups. For convenience, we will give details for only two groups. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ represent random vectors from the two groups that are generated by

$$\mathbf{y}_1 = \boldsymbol{\iota}_1 + \boldsymbol{\Lambda}_1\mathbf{f}_1 + \mathbf{e}_1 \text{ and } \mathbf{y}_2 = \boldsymbol{\iota}_2 + \boldsymbol{\Lambda}_2\mathbf{f}_2 + \mathbf{e}_2, \tag{5}$$

whose first two moment structures are

$$\boldsymbol{\nu}_1(\boldsymbol{\theta}) = \boldsymbol{\iota}_1 + \boldsymbol{\Lambda}_1\boldsymbol{\tau}_1,\ \mathbf{M}_1(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_1\boldsymbol{\Phi}_1\boldsymbol{\Lambda}_1' + \boldsymbol{\Psi}_1,$$

$$\boldsymbol{\nu}_2(\boldsymbol{\theta}) = \boldsymbol{\iota}_2 + \boldsymbol{\Lambda}_2\boldsymbol{\tau}_2,\ \mathbf{M}_2(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_2\boldsymbol{\Phi}_2\boldsymbol{\Lambda}_2' + \boldsymbol{\Psi}_2.$$

It is typical to assume $\boldsymbol{\iota}_1 = \boldsymbol{\iota}_2 = \boldsymbol{\iota}$ and $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}$ in studying the mean difference $\boldsymbol{\tau}_2 - \boldsymbol{\tau}_1$ (Sörbom, 1974). But there can be exceptions (Byrne, Shavelson, & Muthen, 1989). For the purpose of identification, one typically fixes $\boldsymbol{\tau}_1 = \mathbf{0}$, and consequently the interesting null hypothesis is $H_0 : \boldsymbol{\tau}_{20} = \mathbf{0}$. The free parameters in Equation 5 are

$$\boldsymbol{\theta} = (\boldsymbol{\iota}', \boldsymbol{\tau}_2', \boldsymbol{\lambda}', \boldsymbol{\phi}_1', \boldsymbol{\psi}_1', \boldsymbol{\phi}_2', \boldsymbol{\psi}_2')',$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\phi}_1$, $\boldsymbol{\psi}_1$, $\boldsymbol{\phi}_2$, and $\boldsymbol{\Psi}_2$ are vectors containing the free parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}_1$, $\boldsymbol{\Psi}_1$, $\boldsymbol{\Phi}_2$, and $\boldsymbol{\Psi}_2$. With the sample moments $\bar{\mathbf{y}}_1$, $\mathbf{S}_1$, and $\bar{\mathbf{y}}_2$, $\mathbf{S}_2$, the normal theory-based MLE $\hat{\boldsymbol{\theta}}$ is obtained by minimizing

$$F_{ML}(\boldsymbol{\theta}, \bar{\mathbf{y}}_1, \mathbf{S}_1, \bar{\mathbf{y}}_2, \mathbf{S}_2) = n^{-1}n_1F_{ML}(\boldsymbol{\theta}, \bar{\mathbf{y}}_1, \mathbf{S}_1) + n^{-1}n_2F_{ML}(\boldsymbol{\theta}, \bar{\mathbf{y}}_2, \mathbf{S}_2),$$

where n_1 and n_2 are the sample sizes for the two groups with $n = n_1 + n_2$. Under standard regularity conditions, $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}^*$, which minimizes $F_{ML}(\boldsymbol{\theta}, \boldsymbol{\mu}_{10}, \boldsymbol{\Sigma}_{10}, \boldsymbol{\mu}_{20}, \boldsymbol{\Sigma}_{20})$, where $\boldsymbol{\mu}_{10} = E(\mathbf{y}_1)$, $\boldsymbol{\Sigma}_{10} = \text{Cov}(\mathbf{y}_1)$, $\boldsymbol{\mu}_{20} = E(\mathbf{y}_2)$, and $\boldsymbol{\Sigma}_{20} = \text{Cov}(\mathbf{y}_2)$.

Notice that when the population parameter values satisfy $\boldsymbol{\mu}_{10} = \boldsymbol{\mu}_{20} = \boldsymbol{\mu}_0$, whether $\mathbf{M}_1(\boldsymbol{\theta})$ and $\mathbf{M}_2(\boldsymbol{\theta})$ are misspecified or not, the $\boldsymbol{\iota}^*$ has to take the value $\boldsymbol{\mu}_0$ and $\boldsymbol{\tau}_2^*$ has to be zero in order for $F_{ML}(\boldsymbol{\theta}, \boldsymbol{\mu}_{10}, \boldsymbol{\Sigma}_{10}, \boldsymbol{\mu}_{20}, \boldsymbol{\Sigma}_{20})$ to reach its minimum. So when $\boldsymbol{\mu}_{10} = \boldsymbol{\mu}_{20} = \boldsymbol{\mu}_0$, there will be no bias in $\boldsymbol{\tau}_2^*$ even when $\mathbf{M}_1(\boldsymbol{\theta})$ and $\mathbf{M}_2(\boldsymbol{\theta})$ are misspecified. The converse is also partially true. That is, when $\boldsymbol{\tau}_2^* \neq \mathbf{0}$, $\boldsymbol{\mu}_{10}$ will not equal $\boldsymbol{\mu}_{20}$ regardless of whether $\mathbf{M}_1(\boldsymbol{\theta})$ or $\mathbf{M}_2(\boldsymbol{\theta})$ are correctly specified. This partially explains the results of Kaplan and George (1995) and Hancock, Lawrence, and Nevitt (2000) regarding the performance of T_{ML} in testing factor mean differences when factor pattern coefficients are partially invariant. They found that T_{ML} performs well in controlling Type I and Type II errors when $n_1 = n_2$, and it is preferable to other types of analysis.

However, any misspecification will cause an asymptotic bias in $\hat{\boldsymbol{\tau}}_2$ when H_0 is not true or when $\boldsymbol{\mu}_{10} \neq \boldsymbol{\mu}_{20}$. We illustrate how misspecified $(\boldsymbol{\nu}_1(\boldsymbol{\theta}), \mathbf{M}_1(\boldsymbol{\theta}))$ and $(\boldsymbol{\nu}_2(\boldsymbol{\theta}), \mathbf{M}_2(\boldsymbol{\theta}))$ interfere with the estimate $\hat{\boldsymbol{\tau}}_2$ and with testing the null hypothesis $\boldsymbol{\tau}_{20} = \mathbf{0}$. Let $\boldsymbol{\theta}_0$ be the population value of $\boldsymbol{\theta}$ corresponding to correctly specified models and $\boldsymbol{\nu}_1^0 = \boldsymbol{\nu}_1(\boldsymbol{\theta}_0)$, $\boldsymbol{\nu}_2^0 = \boldsymbol{\nu}_2(\boldsymbol{\theta}_0)$, $\boldsymbol{\Sigma}_1^0 = \mathbf{M}_1(\boldsymbol{\theta}_0)$, $\boldsymbol{\Sigma}_2^0 = \mathbf{M}_2(\boldsymbol{\theta}_0)$. Similar to the one-group situation, $\boldsymbol{\theta}$ is a function of $(\boldsymbol{\nu}_1, \boldsymbol{\sigma}_1, \boldsymbol{\nu}_2,$

$\boldsymbol{\sigma}_2$) in a neighborhood of ($\boldsymbol{\nu}_1^0, \boldsymbol{\sigma}_1^0, \boldsymbol{\nu}_2^0, \boldsymbol{\sigma}_2^0$). For the $\Delta\boldsymbol{\theta} = (\Delta\theta_1, \ldots, \Delta\theta_q)' = \boldsymbol{\theta}^* - \boldsymbol{\theta}_0$, we have

$$\Delta\theta_l \approx \sum_{i=1}^{p} c_{li}^{(1)} \Delta\mu_i^{(1)} + \sum_{i=1}^{p}\sum_{j=1}^{p} c_{lij}^{(1)} \Delta\sigma_{ij}^{(1)} + \sum_{i=1}^{p} c_{li}^{(2)} \Delta\mu_i^{(2)} + \sum_{i=1}^{p}\sum_{j=1}^{p} c_{lij}^{(2)} \Delta\sigma_{ij}^{(2)}. \quad (6)$$

Explicit expressions for $c_{li}^{(1)}, c_{li}^{(2)}, c_{lij}^{(1)}$, and $c_{lij}^{(2)}$ are provided in Yuan and Bentler (2004). Equation (6) can be used to evaluate the effect of any misspecifications of ($\boldsymbol{\nu}_1(\boldsymbol{\theta}), \mathbf{M}_1(\boldsymbol{\theta})$) and/or ($\boldsymbol{\nu}_2(\boldsymbol{\theta}), \mathbf{M}_2(\boldsymbol{\theta})$) on $\boldsymbol{\theta}^*$, as illustrated in the following example.

Example 4

Let the population means and covariances be generated by Equation 5 with four variables measuring one factor. We will use $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ to denote the vectors of factor pattern coefficients instead of their matrix versions $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$. Set the population values

$$\boldsymbol{\iota}_{10} = \boldsymbol{\iota}_{20} = (1.0, 1.0, 1.0, 1.0)', \tau_{10} = 0, \tau_{20} = 0.5, \boldsymbol{\lambda}_{10} = \boldsymbol{\lambda}_{20} = (1.0, 1.0, 1.0, 1.0)',$$
$$\phi_0^{(1)} = 1.0, \phi_0^{(2)} = 1.0, \psi_{110}^{(1)} = \ldots = \psi_{440}^{(1)} = 1.0 \text{ and } \psi_{110}^{(2)} = \ldots = \psi_{440}^{(2)} = 1.0.$$

So the model in Equation 5 is correct for the population if there are no perturbations. Fix the first factor pattern coefficient at 1.0 for the purpose of identification and let

$$\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = (1, \lambda_1, \lambda_2, \lambda_3)'$$

and $\tau_1 = 0$ in the model; the free parameters are

$$\boldsymbol{\theta} = (\boldsymbol{\iota}', \tau_2, \lambda_1, \lambda_2, \lambda_3, \phi^{(1)}, \psi_{11}^{(1)}, \psi_{22}^{(1)}, \psi_{33}^{(1)}, \psi_{44}^{(1)}, \phi^{(2)}, \psi_{11}^{(2)}, \psi_{22}^{(2)}, \psi_{33}^{(2)}, \psi_{44}^{(2)})'.$$

Using Equation 6, with equal sample size in the two groups, we get the coefficients c_i and c_{ij} in the first column of Table 7 for the biases in τ_2^*. With $\Delta\mu_i^{(1)} = 0.2, \Delta\mu_i^{(2)} = 0.2, \Delta\sigma_{ij}^{(1)} = 0.2$ and $\Delta\sigma_{ij}^{(2)} = 0.2$, the approximate biases using Equation 6 as well as the exact ones in τ_2^* are given in the second and third columns of Table 7, in which the approximate biases closely match the corresponding exact ones.

According to the coefficients in Table 7, any positive perturbation on $\mu_i^{(1)}$ will cause a negative bias on τ_2^*, and the opposite is true when $\mu_i^{(2)}$ is positively perturbed. Similarly, τ_2^* will change in the direction specified by c_{ij} when σ_{ij} is perturbed. The results in Table 7 imply that one has to be cautious when using a z test for $\tau_{20} = 0$. When $\boldsymbol{\iota}_{10}$ and $\boldsymbol{\iota}_{20}$ are not equal, or the factor pat-

Table 7
The Effect of a Perturbation $\Delta\mu_i = 0.2$ or $\Delta\sigma_{ij} = 0.2$ *on the Difference* τ_2^* *of Factor Means in Latent Mean Comparison*

μ_i	c_i	$c_i \times \Delta\mu_i$	$\Delta\tau_2$
$\mu_1^{(1)}$	-0.314	-0.063	-0.059
$\mu_2^{(1)}$	–0.229	–0.046	–0.044
$\mu_3^{(1)}$	–0.229	–0.046	–0.044
$\mu_4^{(1)}$	–0.229	–0.046	–0.044
$\mu_1^{(2)}$	0.314	0.063	0.066
$\mu_2^{(2)}$	0.229	0.046	0.047
$\mu_3^{(2)}$	0.229	0.046	0.047
$\mu_4^{(2)}$	0.229	0.046	0.047
σ_{ij}	c_{ij}	$c_{ij} \times \Delta\mu_i$	$\Delta\tau_2$
$\sigma_{12}^{(1)}$	0.057	0.011	0.012
$\sigma_{13}^{(1)}$	0.057	0.011	0.012
$\sigma_{14}^{(1)}$	0.057	0.011	0.012
$\sigma_{23}^{(1)}$	–0.057	–0.011	–0.015
$\sigma_{24}^{(1)}$	–0.057	–0.011	–0.015
$\sigma_{34}^{(1)}$	–0.057	–0.011	–0.015
$\sigma_{12}^{(2)}$	0.057	0.011	0.012
$\sigma_{13}^{(2)}$	0.057	0.011	0.012
$\sigma_{14}^{(2)}$	0.057	0.011	0.012
$\sigma_{23}^{(2)}$	–0.057	–0.011	–0.015
$\sigma_{24}^{(2)}$	–0.057	–0.011	–0.015
$\sigma_{34}^{(2)}$	–0.057	–0.011	–0.015

tern coefficients $\boldsymbol{\lambda}_{10}$ and $\boldsymbol{\lambda}_{20}$ are not invariant, or the structural models $\mathbf{M}_1(\boldsymbol{\theta})$ and $\mathbf{M}_2(\boldsymbol{\theta})$ are misspecified, the estimate $\hat{\tau}_2$ cannot be regarded as the estimate of the latent mean difference τ_{20}. The bias $\Delta\tau_2$ can be substantial. Just like a nonzero parameter, the bias in $\hat{\tau}_2$ will be statistically significant when sample sizes are relatively large.

For the four examples in this section, we studied only $\Delta\theta_l$ for a few interesting parameters when the mean μ_i or covariance[1] σ_{ij} are perturbed individually. Equations 3 or 6 can also be used to obtain an approximate bias on any parameter in a model with simultaneous perturbations on elements of means and covariances. For example, when $\mu_1^{(1)}$ and $\sigma_{34}^{(2)}$ are perturbed by $\Delta\mu_1^{(1)} = 0.2$ and $\Delta\mu_{34}^{(2)} = 0.2$ simultaneously, the approximate bias on τ_2^* is about $\Delta\tau_2 = -0.314 \times 2 - 0.057 \times 2 = -0.074$.

Discussion and Conclusion

When variables contain measurement errors, correlation or regression analysis might lead to biased parameter estimates. SEM supposedly removes the biases in regression or correlation coefficients. However, if a model is misspecified, the correlation or regression coefficients among latent variables are also biased. Because the measurement errors are partialled out, SEM also has merits over the traditional MANOVA in comparing mean differences, as discussed in Cole, Maxwell, Arvey, and Salas (1993) and Kano (2001). However, this methodology can also be easily misused. In such a case, the estimated latent mean differences may not truly reflect the mean differences of the latent variables.

There are many model fit indices in the literature of SEM. For example, SAS CALIS provides about 20 fit indices in its default output. Consequently, there is no unique criterion for judging whether a model fits the data. Conceivably, these different criteria might provide good resources because each fit index may provide additional information for looking at the discrepancy between data and model. Actually, Hu and Bentler (1999) recommended using multiple indices in judging the fit of a model. However, people in practice often pick the most favorable index to sell a model. Particularly, with a given fit index, the cutoff value between a good and a bad model is not clear; the commonly used terms *adequate*, *plausible*, or *tenable* for models have never been defined clearly. For example, for the comparative fit index (CFI), the criterion CFI > 0.95 has been recommended for an acceptable model (Bentler, 1990; Hu & Bentler, 1999), but CFI > 0.90 is also commonly used for indicating adequate, plausible, or tenable models. It is interesting to observe that fit indices are often used when judging a covariance structure because of the need to accept the model, whereas chi-squares or z tests are generally used when judging a mean difference because of the need to find significance (see Hong et al., 2003; Whiteside-Mansell & Corwyn, 2003). Such a practice most likely leads to misleading conclusions.

We agree that any model is an approximation to the real world and that there is some need to quantify the degree of approximation. But there are good approximations and bad ones. As we have shown, if a significance or a substantive conclusion following an SEM model is due to systematic biases, caution is needed in elaborating on the findings from the model. To minimize the misuse of ΔT and z tests, one should use multiple criteria to make sure the base model M_a is correctly specified. When M_a is not good enough, one may need to find a different model structure that better fits the data before adding extra constraints or performing a z test. An alternative is to further explore the structure of the data to better understand the substantive theory.

Our study leads to two humble but definite conclusions with regard to the specific types of models. In the latent growth curve models as represented by Equation 4, when $\hat{\alpha}$ or $\hat{\beta}$ is statistically significant at the .05 level, then with 95% confidence one can claim that $E(\mathbf{y})$ is different from zero. In comparing factor means as represented in the model in Equation 5, if $\hat{\boldsymbol{\tau}}_2$ is statistically significant at the .05 level, then one can be 95% confident that $E(\mathbf{y}_1) \neq E(\mathbf{y}_2)$. But the significance in $\hat{\alpha}$ or $\hat{\beta}$ may not be due to nonzero $E(f_1)$ or $E(f_2)$, and the significance of $\hat{\boldsymbol{\tau}}_2$ may not be due to a nonzero $E(\mathbf{f}_2 - \mathbf{f}_1)$.

Note

1. Tables 4, 6 and 7 do not contain σ_{ii} because its perturbation does not cause any biases on the reported paramenters.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605-634.

Brouwers, A., & Tomic, W. (2001). The factorial validity of scores on the teacher interpersonal self-efficiency scale. *Educational and Psychological Measurement*, *61*, 433-445.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185-249). New York: Plenum.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *American Statistician*, *36*, 153-157.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factorial covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research*, *25*, 115-136.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*, 174-184.

Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J. R. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 1-42). Mahwah, NJ: Lawrence Erlbaum.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, *70*, 662-680.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum.

Hancock, G. R., Lawrence, F. R., & Nevitt, J. (2000). Type I error and power of latent mean methods and MANOVA in factorially invariant and noninvariant latent variable systems. *Structural Equation Modeling*, *7*, 534-556.

Hong, S. H., Malik, M. L., & Lee, M.-K. (2003). Testing configural, metric, scalar, and latent mean invariance across genders in sociotropy and autonomy using a non-Western sample. *Educational and Psychological Measurement*, *63*, 636-654.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*, 443-477.

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

Kano, Y. (2001). Structural equation modeling for experimental data. In R. Cudeck, S. H. C. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 381-402). Chicago: Scientific Software.

Kaplan, D. (1989). Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance. *Educational and Psychological Measurement*, *49*, 579-586.

Kaplan, D., & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling*, *2*, 101-118.

Larose, S., Guay, F., & Boivin, M. (2002). Attachment, social support, and loneliness in young adulthood: A test of two models. *Personality and Social Psychology Bulletin*, *28*, 684-693.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd ed.). New York: American Elsevier.

Lee, S. Y. (1985). On testing functional constraints in structural equation models. *Biometrika*, *72*, 125-131.

Lee, S. Y., & Leung, T. K. (1982). Covariance structure analysis in several populations. *Psychometrika*, *47*, 297-308.

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107-120.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201-226.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves with developmental structure equation models. *Child Development*, *58*, 110-133.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

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

Satorra, A., & Saris, W. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83-90.

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.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229-239.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, *50*, 253-264.

Vispoel, W., Boo, J., & Bleiler, T. (2001). Computerized and paper-and-pencil versions of the Rosenberg Self-Esteem Scale: A comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement*, *61*, 461-474.

Whiteside-Mansell, L., & Corwyn, R. F. (2003). Mean and covariance structure analysis: An examination of the Rosenberg Self-Esteem Scale among adolescents and adults. *Educational and Psychological Measurement*, *63*, 163-173.

Yuan, K.-H., & Bentler, P. M., (2004). *Mean comparison: Manifest variable versus latent variable.* Under review.

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology*, *33*, 241-265.