Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

The Impact of Misspecifying the Within-Subject Covariance Structure in Multiwave Longitudinal Multilevel Models: A Monte Carlo Study

Oi-man Kwok , Stephen G. West & Samuel B. Green

To cite this article: Oi-man Kwok , Stephen G. West & Samuel B. Green (2007) The Impact of Misspecifying the Within-Subject Covariance Structure in Multiwave Longitudinal Multilevel Models: A Monte Carlo Study, Multivariate Behavioral Research, 42:3, 557-592, DOI: 10.1080/00273170701540537

To link to this article: http://dx.doi.org/10.1080/00273170701540537

Published online: 05 Dec 2007.

MULTIVARIATE BEHAVIORAL RESEARCH, *42*(3), 557–592

The Impact of Misspecifying the Within-Subject Covariance Structure in Multiwave Longitudinal Multilevel Models: A Monte Carlo Study

Oi-man Kwok
Texas A&M University, College Station, Texas

Stephen G. West and Samuel B. Green
Arizona State University, Tempe, Arizona

This Monte Carlo study examined the impact of misspecifying the $\mathbf{\Sigma}$ matrix in longitudinal data analysis under both the multilevel model and mixed model frameworks. Under the multilevel model approach, under-specification and general-misspecification of the $\mathbf{\Sigma}$ matrix usually resulted in overestimation of the variances of the random effects (e.g., τ_{00}, $\tau_{\tau_{11}}$) and standard errors of the corresponding growth parameter estimates (e.g., $SE_{\beta 0}$, $SE_{\beta 1}$). Overestimates of the standard errors led to lower statistical power in tests of the growth parameters. An unstructured $\mathbf{\Sigma}$ matrix under the mixed model framework generally led to underestimates of standard errors of the growth parameter estimates. Underestimates of the standard errors led to inflation of the type I error rate in tests of the growth parameters. Implications of the compensatory relationship between the random effects of the growth parameters and the longitudinal error structure for model specification were discussed.

The first two authors contributed equally to the development of this manuscript. The authors are grateful to Leona Aiken and David MacKinnon for their comments on an earlier version of this manuscript.

Correspondence concerning this article should be addressed to Oi-man Kwok, Department of Educational Psychology, 4225 TAMU, College Station, TX 77843-4225. E-mail: omkwok@neo.tamu.edu

When researchers analyze multiwave longitudinal data (i.e., three or more measured waves) using the Multi-Level Modeling (MLM) framework (Bryk & Raudenbush, 1987; Laird & Ware, 1982; Raudenbush & Bryk, 2002), they typically propose an average growth trend (e.g., linear, or quadratic). The corresponding between-subject variation associated with the residuals from the intercepts and the individual growth trends is also modeled. However, researchers typically assume the *within-subject residuals* to be independently and identically distributed (i.i.d.) with mean zero and homogenous variance σ^2 for all participants (i.e., $\underline{e} \sim N(\underline{0}, \sigma^2\mathbf{I})$). There are two issues when researchers only focus on modeling the average growth trends and the corresponding between-subject variation. First, important aspects of longitudinal change of some phenomena may not be reflected in the average growth trends but rather in the form of the variance-covariance structure of the within-subject residuals (Biesanz, West, & Kwok, 2003; Cook & Weisberg, 1999; Hedeker & Mermelstein, 2007). For example, Hedeker and Mermelstein (2007) showed that mood change in smokers could be reflected by the changes in the within-subject covariance structure (i.e., the *variation* of the mood change in smokers decreased over time) rather than the average growth. Second, the simplification of the within-subject covariance structure (i.e., identity structure $= \sigma^2\mathbf{I}$, which assumes an homogeneous within-subject residual variance) may bias the estimation of the standard errors of the fixed effects, which, in turn, may lead to either Type I or Type II error when testing the fixed effects, and affect the construction of confidence intervals for the effects of interest (Davis, 2002; Diggle et al., 2002; Singer & Willett, 2003; Weiss, 2005). The major focus of this paper is to examine the effect of misspecifying the within-subject covariance structure on the estimation and testing of the fixed effects and the variances and covariance of the random effects. Paralleling with the conditions considered in Ferron, Dailey and Yi's (2002) study, we also examined the plausible influence of other factors including sample size, number of waves, magnitude of the fixed effects, and the magnitude of the elements of the covariance structure. Some guidelines for analyzing longitudinal data under the MLM framework will be provided.

Multi-Level Modeling (MLM)

MLM can be viewed as an extension of familiar linear models such as analysis of variance and multiple regression. For example, consider a longitudinal study with N participants and T different measurement occasions on the same instrument for each of the N participants. To simplify the illustration, we use a simple linear growth model which can be written in reduced (mixed model) form as follows:

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{T1} \\ \vdots \\ \vdots \\ \vdots \\ y_{1N} \\ \vdots \\ y_{TN} \end{bmatrix} = \begin{bmatrix} 1 & TIME_1 \\ \vdots & \vdots \\ 1 & TIME_T \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & TIME_1 \\ \vdots & \vdots \\ 1 & TIME_T \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & TIME_1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 1 & TIME_T & \vdots & \dots & \vdots & 0 & 0 \\ 0 & 0 & \vdots & \dots & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & \dots & \vdots & 0 & 0 \\ 0 & 0 & \vdots & \dots & \vdots & 1 & TIME_1 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & TIME_T \end{bmatrix} \begin{bmatrix} u_{01} \\ u_{11} \\ u_{02} \\ u_{12} \\ \vdots \\ \vdots \\ \vdots \\ u_{0N} \\ u_{1N} \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{T1} \\ \vdots \\ \vdots \\ \vdots \\ e_{1N} \\ \vdots \\ e_{TN} \end{bmatrix} \tag{1}$$

This equation can be rewritten in matrix form as:

$$\underline{\boldsymbol{y}} = \mathbf{X}\underline{\beta} + \mathbf{Z}\underline{\boldsymbol{u}} + \underline{\boldsymbol{e}} \tag{2}$$

In equation (2), $\underline{y}$ is a [TN] column vector containing the T repeated measures for all N participants. $\mathbf{X}$ is a [(TN) by 2] matrix containing the intercept (i.e., $\underline{1}$) and the predictor variable TIME. $\underline{\beta}$ is a [2] column vector containing the unknown growth parameters β_0 and β_1. $\mathbf{Z}$ is a [(TN) by (2N)] design matrix, and $\underline{u}$ is a [2N] column vector containing the random effects representing between subject variation (individual differences). $\underline{e}$ is a [TN] column vector containing the within subject random errors. Conceptually, the model shown in equation (2) can be divided into two components: fixed effects (i.e., $\mathbf{X}\underline{\beta}$), and random effects (i.e., $\mathbf{Z}\underline{u} + \underline{e}$). Fixed effect coefficients β_0 and β_1 represent the overall/average model intercept and slope, respectively. Random effects u_{0i} and u_{1i} (i.e., elements in $\underline{u}$) represent the deviation of the ith subject's intercept and slope, respectively, from the average intercept and slope (between subject variation). Random effect e_{ti} (i.e., elements in $\underline{e}$) represents the deviation of the observation at the t-th measurement occasion for the i-th participant from that participant's individual regression line (within subject random errors). The variance of the random effects in equation (2) is equal to:

$$VAR(Z\underline{u} + \underline{e}) = VAR(Z\underline{u}) + VAR(\underline{e})$$

$$= \underset{\mathbf{Z}}{\begin{bmatrix} Z_1 & 0 & . & . & 0 \\ 0 & Z_2 & . & . & 0 \\ \vdots & \vdots & . & \vdots & \vdots \\ \vdots & \vdots & \vdots & . & \vdots \\ 0 & . & . & . & Z_N \end{bmatrix}} \underset{\mathbf{G}}{\begin{bmatrix} T & \dots & \dots & 0 \\ 0 & T & \dots & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & . & . & T \end{bmatrix}} \underset{\mathbf{Z^T}}{\begin{bmatrix} Z_1 & 0 & . & . & 0 \\ 0 & Z^2 & . & . & 0 \\ \vdots & \vdots & . & \vdots & \vdots \\ \vdots & \vdots & \vdots & . & \vdots \\ 0 & . & . & . & Z_N \end{bmatrix}^T}$$

$$+ \underset{\mathbf{R}}{\begin{bmatrix} \Sigma & 0 & \dots & \dots & 0 \\ 0 & \Sigma & \dots & \dots & 0 \\ \vdots & \vdots & . & . & \vdots \\ \vdots & \vdots & . & . & \vdots \\ 0 & 0 & \dots & \dots & \Sigma \end{bmatrix}} \tag{3}$$

where the elements on the diagonals of the block diagonal matrices are defined as following:

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 & \mathrm{TIME}_1 \\ 1 & \mathrm{TIME}_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & \mathrm{TIME}_\mathrm{T} \end{bmatrix} \tag{4}$$

$$\mathbf{T} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \tag{5}$$

and

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \dots & \sigma_{2T} \\ \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \dots & \dots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \dots & \dots & \sigma_{TT}^2 \end{bmatrix} \tag{6}$$

T is a [2 by 2] variance-covariance matrix[1] containing the variances and covariances of the random effects related to between subject differences. $\mathbf{\Sigma}$ is a [T by T] variance-covariance matrix containing the variances and covariances of random errors related to within subject random errors. Equation (3) assumes that the covariance structures of $\underline{u}$ and $\underline{e}$ are independent and can be modeled separately. Indeed, equation (3) can be rewritten as:

$$VAR(\boldsymbol{Z}\underline{\boldsymbol{u}} + \underline{\boldsymbol{e}})$$

$$= \begin{bmatrix} \boldsymbol{Z}_1 \boldsymbol{T} \boldsymbol{Z}_1^{\boldsymbol{T}} + \Sigma & 0 & . & . & 0 \\ 0 & \boldsymbol{Z}_2 \boldsymbol{T} \boldsymbol{Z}_2^{\boldsymbol{T}} + \Sigma & . & . & 0 \\ \vdots & \vdots & . & \vdots & \vdots \\ \vdots & \vdots & \vdots & . & \vdots \\ 0 & . & . & . & \boldsymbol{Z}_N \boldsymbol{T} \boldsymbol{Z}_N^{\boldsymbol{T}} + \Sigma \end{bmatrix} \tag{7}$$

which is a block diagonal structure with $\mathbf{Z_i T Z_i^{\mathit{T}}} + \mathbf{\Sigma}$ on the main diagonal and zeros off the main diagonal given $\mathbf{i} = 1 \ldots N$. Another important assumption underlying equation (7) is that all N participants are *independent* from each other (no clustering of individuals). Given these assumptions, the covariance structure of the random effects for each individual (i.e., $\mathbf{Z_i T Z_i^{\mathit{T}}} + \mathbf{\Sigma}$) can be seen to be a combination of the covariance elements of the between individual random effects (i.e., $VAR(\underline{u}) = \mathbf{T}$) and within individual random errors (i.e., $VAR(\underline{e}) = \mathbf{\Sigma}$).

Another way to analyze longitudinal data is to use traditional approaches such as Univariate Analysis of Variance (UANOVA). Unlike equation (2) which contains two random effect components (i.e., $\underline{u}$ and $\underline{e}$), UANOVA contains only one random effect component (i.e., $\underline{e}$):

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e} \tag{8}$$

[1]The size of both Z and T depends on the number of random effects related to the between subject differences. For example, if quadratic growth is examined (see Study 2) and a (between-subject) random effect is associated with the quadratic term, T will be a [3 by 3] matrix containing the variance and covariance components of the three random effects associated with the three growth parameters, namely, intercept, linear term, and quadratic term. Z will be a [T by 3] design matrix containing the predictor values corresponding to the intercept, linear term and quadratic term.

with

$$VAR(\underline{e}) = \begin{bmatrix} \Sigma & 0 & \dots & \dots & 0 \\ 0 & \Sigma & \dots & \dots & 0 \\ \vdots & \vdots & . & . & \vdots \\ \vdots & \vdots & . & . & \vdots \\ 0 & 0 & \dots & \dots & \Sigma \end{bmatrix} \tag{9}$$

where $\mathbf{\Sigma}$, the within-subject covariance structure, is a [T by T] matrix with a *compound symmetry*[2] (CS) structure containing a constant covariance for all possible pairs of the T repeated measures (Kirk, 1995; Maxwell & Delaney, 2004; Rogan, Keselman, & Mendoza, 1979). The CS structure can be presented as follows:

$$\Sigma = \begin{bmatrix} \sigma_A^2 & \sigma_B^2 & \dots & \dots & \sigma_B^2 \\ \sigma_B^2 & \sigma_A^2 & \dots & \dots & \sigma_B^2 \\ \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \dots & \dots & \vdots \\ \sigma_B^2 & \sigma_B^2 & \dots & \dots & \sigma_A^2 \end{bmatrix} \tag{10}$$

which contains constant diagonal and constant off-diagonal elements (i.e., σ_A^2 & σ_B^2 respectively).

Indeed, MLM is a more general model framework and UANOVA can be viewed as a special case of MLM. The UANOVA model can be reproduced by fitting a random intercept model in MLM (i.e., all elements in equation (5) are constrained to zero except τ_{00} and Σ in equation (6) has an ID structure ($\sigma^2\mathbf{I}_{(T)}$)). The advantage of using UANOVA to analyzing longitudinal data is that it has higher statistical power if the sphericity assumption is met (Algina & Keselman, 1997; Keselman, Algina, & Kowalchuk, 2001; Rogan, et al., 1979; Wolfinger, 1996). However, this assumption is likely to be violated in longitudinal studies (Keselman et al., 2001; McCall & Appelbaum, 1973). Generally, correlations (or covariances) between y measures collected closer together in time (e.g., 1st time point and 2nd time point) should be higher than correlations (or covariances)

[2]Compound symmetry structure of $\mathbf{\Sigma}$ is a sufficient condition for fulfilling the sphericity assumption. A covariance structure which can reflect the sphericity assumption is the Huynh-Feldt (H-F) covariance structure. The H-F covariance structure is a more general form of the compound symmetry structure (Wolfinger, 1996).

between y measures collected further apart in time (e.g., 1st time point and 20th time point). Violation of the sphericity assumption results in the inflation of the nominal alpha level, which in turn, affects inferences based on the F-test statistics (Keselman et al., 2001).

Advantages of Using MLM over the Traditional ANOVA Approaches

A major difference between MLM and the traditional ANOVA approaches is the covariance structure of the random errors. The covariance structure of the random effects in the traditional ANOVA approaches is modeled as a single component (i.e., $\underline{e}$; see equations (8) and (9)), whereas the random effects in MLM are divided into two components: $\underline{u}$ and $\underline{e}$, which represent the between subject variation and within subject random errors respectively. Several advantages accrue from modeling the between-subject and within-subject random effects as two separate components. First, MLM permits the examination of new effects of interest such as *cross-level* interaction effects which represent the interaction between one or more between subject (individual level) predictors and the growth trend of the within subject repeated measures. For example, researchers can examine how personal characteristics such as age and gender (i.e., individual level predictors) influence the growth trajectories (i.e., within subject repeated measures) of mathematical achievement in young children. Second, the covariance matrices of both the between subject random effects and within subject random errors can be flexibly and simultaneously modeled in MLM (Chi & Reinsel, 1988; Diggle, 1988; Laird & Ware, 1982; Jones & Boadi-Boateng, 1991; Wolfinger, 1993). Heterogeneity (i.e., differences of the covariance elements of the random effects or errors in different populations under the *same* fixed-effect and random-effect structure) can also be easily addressed under the MLM framework (Littell, et al., 1996; Raudenbush, et al., 2000).

However, when researchers adopt the MLM approach to analyzing longitudinal data, they typically focus strongly on the examination of the covariance structure of the between-subject random effects while retaining the simple default covariance structure of the within-subject random errors (i.e., specifying the $\mathbf{\Sigma}$ matrix (see equations (3) and (6)) as having the default identity structure (ID)—$\sigma^2\mathbf{I}_{(T)}$; for examples, see Fredricks & Eccles, 2002; Jacobs et al., 2002). This oversimplified $\mathbf{\Sigma}$ matrix implies that there is no correlation/covariance between any pair of random errors after partialing out the between-subject covariances, which is very unlikely to reflect reasonable assumptions about the data (Goldstein, Healy, & Rasbash, 1994; Sivo, Fan & Witta, 2005). The existence of (auto)correlated residuals in longitudinal data is very common (Sivo, Fan & Witta, 2005) and has been shown in different studies such as the change of the household income over time (Bollen & Curran, 2004) and the relation between depression and antisocial behavior in children (Curran & Bollen, 2001). One possible reason of the occurrence of the correlated residuals may be due to the test-retest property in longitudinal data (Sivo, Fan & Witta, 2005). Such misspecification of the covariance structure can result in biased estimation of the standard errors, which in turn, will affect the tests of significance and the construction of confidence intervals for the effects of interest in MLM (Davis, 2002; Diggle et al., 2002; Singer & Willet, 2003; Weiss, 2005).

Types and Effects of Misspecification in the Covariance Structure of Errors within a Single Population

There are three major types of misspecification in the covariance structure of errors within a single population: under-specification, over-specification, and general misspecification. Figure 1 presents an illustration of the different types of misspecification using four different within-subject covariance (or $\mathbf{\Sigma}$) matrices,[3] namely, identity (ID), first-order autoregression (AR(1)), first-order autoregression and first-order moving average (ARMA(1,1)), and second-banded Toeplitz (TOEP(2)). As shown in Figure 1, ID contains a single parameter (σ^2) on the main diagonal of an identity matrix, whereas TOEP(2) contains two parameters with σ^2 on the main diagonal and σ_1 to represent constant correlated error terms only for the first lag of the matrix. AR(1) contains two parameters (σ^2 and the autocorrelation coefficient ρ), and ARMA(1,1) contains not only the same two parameters (i.e., σ^2 and ρ) as in AR(1) but also the moving average coefficient (γ). A *nested* relationship between different $\mathbf{\Sigma}$ matrices is defined by whether one can obtain a specific $\mathbf{\Sigma}$ matrix by imposing constraint(s) on another $\mathbf{\Sigma}$ matrix. For example, AR(1) can be reduced to ID if ρ is set to zero. Hence, ID is nested within AR(1). On the other hand, TOEP(2) cannot be reduced to AR(1) (or vice versa) by imposing constraint(s). Figure 1(A) shows that ID is nested within AR(1) and AR(1) is nested within ARMA(1,1), whereas ID is nested within TOEP(2) but TOEP(2) is not nested within either AR(1) or ARMA(1,1).

As shown in Figure 1(B), under-specification occurs within nested $\mathbf{\Sigma}$ matrices when the true $\mathbf{\Sigma}$ matrix is more complex than the $\mathbf{\Sigma}$ matrix chosen for

[3]There are more choices for modeling the within-subject covariance structure in some common statistical programs (e.g., there are more than 18 different covariance structures available in SAS PROC MIXED). Here we only chose four commonly used structures among a large variety of possible covariance structures that were sufficient to study the effects of the four different types of misspecification.

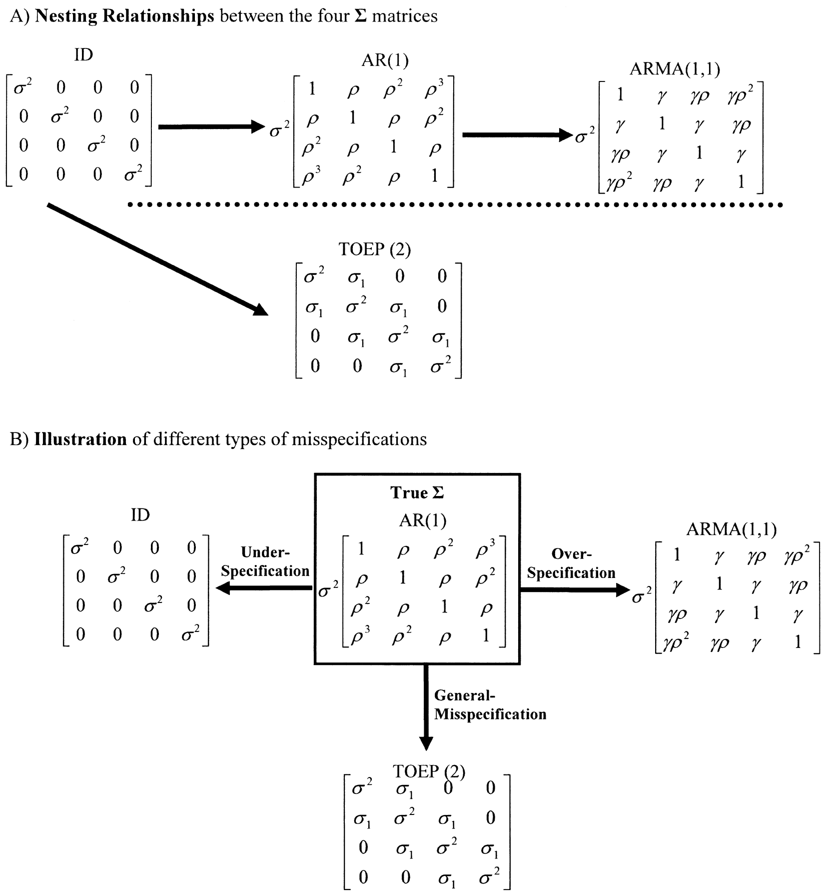

FIGURE 1 Different types of misspecifications of the within-subject covariance structure.

the analysis (e.g., the chosen $\mathbf{\Sigma}$ matrix is ID but the true $\mathbf{\Sigma}$ matrix is AR(1)). Over-specification occurs within nested $\mathbf{\Sigma}$ matrices when the true $\mathbf{\Sigma}$ matrix is more constrained than the chosen $\mathbf{\Sigma}$ matrix (e.g., the chosen $\mathbf{\Sigma}$ matrix is ARMA(1,1) but the true $\mathbf{\Sigma}$ matrix is AR(1)). Once again, these two types of misspecifications apply only to *nested* $\mathbf{\Sigma}$ matrices. General misspecification occurs when the true $\mathbf{\Sigma}$ matrix and the chosen $\mathbf{\Sigma}$ matrix are not nested. For example, general misspecification would occur if the chosen $\mathbf{\Sigma}$ matrix is TOEP(2) but the true $\mathbf{\Sigma}$ matrix is AR(1)).

The effect of under-specification on estimation and test of significance has been studied in the past (Ferron et al., 2002; Sivo, Fan & Witta, 2005; Sivo

& Willson, 2000). For example, Ferron et al (2002) studied the effect of an underspecified within-subject covariance structure in MLM framework. In their Monte Carlo study, the true $\boldsymbol{\Sigma}$ was AR(1) but was underspecified as ID. The estimates of the fixed effects were unbiased but the corresponding standard errors were positively biased (i.e., larger than the true standard errors), which in turn, affected the confidence intervals and reduced the statistical power of the target fixed effects. Sivo and colleagues (Sivo et al., 2005) have studied the same under-specification effect with broader conditions under the latent growth modeling framework (using structural equation models) and a similar pattern of results were found.[4] However, unlike under-specification, the effects of over-specification and general misspecification on the parameter estimation and the test of significance of the fixed effects *of interest* have not yet been intensively studied.

In another context, Kowalchuk and Keselman (2001) suggested using the overspecified structure, UN-H, for the within-subject covariance structure[5] with Satterthwaite estimation of degrees of freedom in the context of conducting *pairwise multiple comparisons* under the ANOVA framework. Nevertheless, in the context of the full multilevel modeling framework, Singer and Willet (2003) showed in an example analysis that the standard errors of the fixed effects became smaller when a more optimal and parsimonious covariance structure rather than UN was adopted. Wolfinger (1996) defined the meaning of optimal and parsimonious covariance structure as follows: "to obtain the most efficient inferences about the mean model (i.e., fixed effects or the average growth trend), one selects the most parsimonious covariance structure possible that still reasonably fits the data" (p. 208). We will use "optimal covariance structure" to stand for the "optimal and parsimonious covariance structure". In general, optimal covariance structure is believed to be able to maximize the statistical power of detecting the target fixed effects given a nominal alpha level (Keselman et al., 2001; Singer & Willet, 2003; Wolfinger, 1996). An overly specified

[4]Sivo et al (2005) not only found biases in the estimates of the variances and covariance of the random effects but also in the estimates of the fixed effects. One possible cause for the different findings between these two studies may be due to the use of the estimation method. Ferron et al (2002) analyzed their data under the MLM framework in which Restricted Maximum Likelihood (REML) was used. In constrast, Sivo et al (2005) analyzed their data under the *latent* growth model framework in which Full Information Maximum Likelihood (FIML) was used. The major difference between these two estimation methods is that both regression coefficients and variance components are included in the likelihood function when using FIML but only variance components are included when using REML.

[5]UN is the most general form of the covariance structure for a single population (Wolfinger, 1996). In the same vein, UN-H (i.e., Heterogeneous Unstructured) is the most general form of the covariance structure for multiple populations.

covariance structure can reduce the generalizability of the hypothesized model (Myung & Pitt, 2004). Hence, Kowlachuk and Keselman's suggestion of over-specifying the within-subject covariance structure as UN may potentially only be applicable to specific multiple comparison tests but not to more general conditions.

Purposes of This Study

As described previously, the effects of misspecification have not yet been intensively studied. Generally, researchers believe that misspecification of the covariance structure affects the estimation of the standard errors of the fixed effects, which in turn, affects the tests of significance, estimation of confidence intervals, and calculation of the pseudo R^2 for each fixed effect.

Only a small number of studies to date have examined the effect of misspecification of the covariance structure under the MLM framework. No systematic study has examined the effect of all three different types of misspecification of the covariance structure (i.e., under-specification, over-specification, and general misspecification) on the estimation of fixed effects and their corresponding standard errors in MLM. The major goal of this paper is to investigate the effect of different forms of misspecification of the covariance structure of the within-subject residuals for longitudinal models under the MLM framework. We examined the impact of misspecification in two Monte Carlo studies.

STUDY 1

In Study 1, we focused on a common two-level growth model with level-1 modeling the repeated measures within individuals and level-2 modeling the differences of individual growth models between individuals. We limited our focus to a simple *linear* growth model (see equation (1)) with correctly specified fixed effects collected in a balanced design.

Method for Study 1 Linear Growth Model

The simulation used a 2 (30 or 210 cases) $\times$ 2 (4 or 8 waves) $\times$ 3 (magnitude of growth parameter β_1: 0, .05 or .16) $\times$ 2 (T matrix: small or medium) $\times$ 4 (true $\boldsymbol{\Sigma}$ matrices for generating the data: ID, TOEP(2), AR(1), or ARMA(1,1)) factorial design to generate the data. A total of 500 replications were generated for each condition using the Mplus (V4.1) Monte Carlo procedure (Muthén & Muthén, 2006), yielding 48,000 total datasets. All data were generated under Mplus with a multivariate normal distribution (Muthén & Muthén, 2001).

TABLE 1
Conditions for Examining the Effects of Over-Specifications, Under-Specifications, and General-Misspecifications

		True *Covariance Structure*			
		IDENTITY	*AR(1)*	*TOEP(2)*	*ARMA(1,1)*
Hypothesized/Tested	IDENTITY	X	Under	Under	Under
Covariance Structure	AR(1)	Over	X	General	Under
	TOEP(2)	Over	General	X	General
	ARMA(1,1)	Over	Over	General	X
	UN	Over-Null **T**	Over-Null **T**	Over-Null **T**	Over-Null **T**

Note. X: Correct-specification; Over: Over-specification; Under: Under-specification; General: General misspecification; Over-Null **T**: Over-specification with null **T** matrix. There are two parts of analyses based on the specification conditions: (a) analyses within the complete MLM framework including under-specification, general misspecification, over-specification, and correct specification of the $\mathbf{\Sigma}$ matrix in which the **T** matrix is specified as unstructured permitting estimation of variance of the random intercept, the random slope, and their covariance (i.e., conditions above the double lines); (b) analyses within the general mixed model framework using an unstructured $\mathbf{\Sigma}$ matrix (i.e., UN) and null **T** matrix in which the intercept and the slope are fixed for all participants (i.e., conditions below the double lines).

Each dataset was then analyzed using *five* separate specifications of the Σ matrix[6] (ID, TOEP(2), AR(1), ARMA(1,1) and UN) using SAS PROC MIXED (Littell et al., 1996) yielding a total of 240,000 records (i.e., 48,000*5). The combination of the four $\mathbf{\Sigma}$ matrices for data generation and the five $\mathbf{\Sigma}$ matrices for analysis yielded five specification categories (see Table 1). The details of each design factor are described below together with a justification of the values selected for study.

Number of participants. The number of the participants (or cases) is based on the conditions used in past simulation studies (Keselman et al., 1998; Ferron

[6]When we generated the data based on any one of the four covariance structures (i.e., ID, TOEP(2), AR(1), and ARMA(1,1)), we could obtain the overall covariance matrix as shown in equation (7). Then, we could either analyze the data under the MLM framework and use these four covariance structures with freely estimated elements in **T** as shown in equation (5), or we could analyze the data under the traditional approach as shown in equations (8) and (9) (i.e., no $\underline{u}$), without imposing any specific structure for $\mathbf{\Sigma}$ (i.e., UN) and constraining all the elements in **T** to zeros (i.e., null **T**). Equation (7) can be reduced to equation (9) by constraining all the elements in **T** to zero. The number of parameters in the covariance matrix of the later approach (i.e., modeling $\mathbf{\Sigma}$ as UN along with null **T**) is equal to: $[T^*(T+1)]/2$ with T equal to the number of repeated measures, which also implies that the covariance matrix will always be over-specified under this later approach.

et al., 2002) and Khoo et al.'s (2006) review of the multiwave longitudinal studies published in *Developmental Psychology* in 2002. Thirty (individuals) was the smallest sample size considered in both Keselman et al (1998) and Ferron et al (2002) studies. The mean number of individuals of the multiwave longitudinal studies published in *Developmental Psychology* was 210 (*SD* = 180). Hence, we chose 30 as a "small" number of individuals and 210 as a "medium" number of individuals.[7]

Number of measurement waves. Among the multiwave longitudinal studies[8] published in *Developmental Psychology* in 2002, more than half (52%) of these studies collected data with three or four occasions. The mean number of waves of the other 48% of studies was 8. Hence we chose 4 waves as the small number of repeated measures and 8 waves as the medium number of measures. We centered the time variable so it had a mean of 0 and 1 unit between adjacent observations (i.e., $Time_{4waves'} = [-1.5\ \ -.5\ \ .5\ \ 1.5]$, and $Time_{8waves'} = [-3.5\ \ -2.5\ \ -1.5\ \ -.5\ \ .5\ \ 1.5\ \ 2.5\ \ 3.5]$).

Standardized effect size of the average growth trajectory. Two different magnitudes of the standardized effect size of the growth trajectory (i.e., β_1, see Equation (1)) were examined in this study, small effect size (i.e., $\beta_1 = .05$) and medium effect size (i.e., $\beta_1 = .16$). These values were obtained by the following effect size equation (Raudenbush & Liu, 2001):

$$\delta = \frac{\beta_1}{\sqrt{\tau_{11}}} \tag{11}$$

where δ is the standardized effect size, β_1 is the average linear growth trend (see Equation (1)), and τ_{11} is the variance of the random effect associated with the growth parameter (see Equation (5)), which indicates the differences between individual growth trends and the average growth trend. Cohen (1988) provided some standardized effect size guidelines in which small effect size (i.e., δ) is equal to .20 and medium effect size is equal to .50. Raudenbush and Liu (2001) proposed similar guidelines for the size of τ_{11}, where .05 was for small τ_{11} and .10 was for medium τ_{11}. Given the values of δ and τ_{11}, the corresponding β_1 could be easily computed. β_0 was fixed to a constant value in all conditions (i.e., $\beta_0 = .10$).

[7]We only considered small and medium numbers of participants (and repeated measures) because longitudinal studies with a large number of *both* individuals and observations are rare in most areas of psychology.

[8]One study (Lagattuta & Wellman, 2002) contained a very large number of repeated measures (i.e., more than 50). This study was treated as an outlier and excluded from the proportion calculation.

*Overall effect size of **T** matrix.* The parameters which have to be determined in the **G** matrix are the elements in the **T** matrices lying on the diagonal of the **G** matrix. As shown in equation (5), the elements in **T** matrices are τ_{00}, τ_{11}, and τ_{01} (which is identical to τ_{10}), where τ_{00} captures the variance of the intercepts of individual growth models and τ_{11} captures the variance of the linear growth trends of individual growth models. τ_{10} captures the covariance between the individual intercepts and linear growth trends. According to the criteria provided by Raudenbush and Liu (2001), a medium **T** and a small **T** could be specified as:

$$T_{Medium} = \begin{bmatrix} .200 & .050 \\ .050 & .100 \end{bmatrix} \text{ and } T_{Small} = \begin{bmatrix} .100 & .025 \\ .025 & .050 \end{bmatrix}$$

In $\mathbf{T}_{Medium}$, τ_{11} was set to .10, whereas τ_{00} was set to .20 (i.e. double of τ_{11}) because the variation of the intercepts was generally larger than the variation of the growth trends in longitudinal studies. The size of the covariance, τ_{01} (or τ_{10}), was half of the size of τ_{11} because the covariance between intercepts and growth trends was generally small in longitudinal studies. The parameters in $\mathbf{T}_{Small}$ were set by halving the corresponding parameter values in $\mathbf{T}_{Medium}$ (Raudenbush & Liu, 2001).

Specification of $\boldsymbol{\Sigma}$ matrices. The final parameters we had to determine were the elements in the $\boldsymbol{\Sigma}$ matrices, which were lying on the main diagonal of the **R** matrix (see equation (3)). Four covariance structures, namely, ID, AR(1), TOEP(2), and ARMA(1,1) were adopted. One key feature of these four error covariance structures was that the numbers of parameters of these four error covariance structures are invariant across different numbers of repeated observations[9] (here, 4 versus 8). Moreover, as shown in Figure 1, these four structures have both nested and non-nested relations, which can be used for studying the effects of under-specification, over-specification, and general misspecification (see Table 1). All of these structures are available in SAS PROC MIXED Version 8 and higher.

The four error covariance structures are commonly used when analyzing longitudinal data. As noted previously, researchers who use the MLM approach nearly always use the default ID covariance structure. AR(1) and ARMA(1,1) are commonly considered in time series analysis (Velicer & Fava, 2003; West & Hepworth, 1991). TOEP(2) is closely related to the moving average (1) structure which is also commonly used in time series.

[9]The number of parameters in some covariance structures varies as the number of repeated measures changes. For example, the number of parameters for an UN(1) (first-banded or main diagonal unstructured) with 4-wave measures is 4 whereas the number of parameters for an UN(1) with 8-wave measures is 8.

As shown in Figure 1, four parameters are necessary to specify the four chosen error covariance structures. These parameters are: σ^2 (variance of the within subject random errors), σ_1 (i.e., the parameter in TOEP(2)), γ (i.e., moving average coefficient), and ρ (i.e., autoregressive correlation coefficient). To reduce the number of unknown parameters, the within subject random errors were assumed to be normally distributed with variance equal to one (i.e., $\sigma^2 = 1$), which is the general practice when conducting power analyses under the MLM framework (Snijders & Bosker, 1993; Bosker, Snijders, & Guldemond, 2003). σ_1 was assumed to be the same as γ. Hence, we only needed to specify two parameters, ρ and γ, for the four within subject covariance structures. ρ was fixed to .80 and γ was fixed to .50, which are within the reasonable range of values used in past simulation studies (Hamaker, Dolan, & Molenaar, 2002; Sivo & Willson, 2000).

Evaluation criteria. There were several evaluation criteria used to study the effects of misspecifications of the within subject covariance structures along with other design factors (i.e., number of cases, number of waves, magnitude of growth parameter, and magnitude of **T** matrix). The criteria were: 1) convergence of the analyses, 2) relative bias[10] of the estimates of the fixed effects (i.e., β_0 and β_1) and their corresponding standard errors (i.e., $SE_{\beta 0}$ and $SE_{\beta 1}$), and the variances (i.e., τ_{00} and τ_{11}) and covariance (i.e., τ_{01}) of the random effects, and 3) Type I error rate and statistical power, as appropriate, of the tests of the fixed effects.

Results for Study 1 Linear Growth Model

Conceptually, the analyses can be partitioned into two broad parts based on the specification conditions. (a) The first part corresponds to analyses within the complete multi-level model (MLM) framework. Conditions corresponding to under-specification, general misspecification, over-specification, and correct specification of the $\mathbf{\Sigma}$ matrix were created (see Table 1). In all cases the **T** matrix was specified as unstructured permitting estimation of variance of the random intercept, the random slope, and their covariance. (b) The second part corresponds to analyses within the general mixed model framework by specifying an unstructured (UN) $\mathbf{\Sigma}$ matrix with null **T** matrix. This analysis assumes that the intercept and the slope are *fixed* for all participants. The results pertaining to the

[10]Relative bias is calculated by: $RB = \frac{\hat{\theta}-\theta}{\theta}$, where θ is the true parameter value, and $\hat{\theta}$ is the corresponding sample estimate. RB equal to zero indicates an unbiased estimate of the parameter. A negative RB indicates an underestimation of the parameter (i.e., the estimated value is smaller than the true parameter value), whereas a positive RB indicates an overestimation of the parameter (i.e., the estimated value is larger than the true parameter value).

four specification categories under the MLM framework are presented at the beginning of each section. These results are then followed by a comparison of the performance of the correctly specified condition with those of the unstructured (UN) condition with null **T** matrix. The corresponding SAS computer script for the data analyses are presented in Appendix A.

The results are presented in the following sequence: 1) convergence of the analyses when specifying different $\boldsymbol{\Sigma}$ matrices regardless of the true $\boldsymbol{\Sigma}$ matrices, 2) the impact of different design factors on the relative bias (RB) of the target parameters (i.e., β_0, β_1, $SE_{\beta 0}$, $SE_{\beta 1}$, τ_{00}, τ_{01}, & τ_{11}), and 3) the impact of different design factors on the type I error rate and the statistical power for testing the average growth parameters (i.e., β_0 and β_1). Of particular interest are the effects of the specification of the $\boldsymbol{\Sigma}$ matrix under both MLM and general mixed model frameworks.

The impact of the specification category and other design factors on the relative bias of a specific parameter was examined under the UANOVA framework[11] with η^2 (i.e., $\eta^2 = \frac{SS_{Effect}}{SS_{Total}}$) as the effect size indicator. The reason for using η^2 instead of the significance test (or p-value) was due to the very large number of records (i.e., 48,000 datasets * 5 different types of $\boldsymbol{\Sigma}$ matrix for analyses = 240,000 records), which substantially reduced the size of the sum of squared errors (SSE) and resulted in detecting a large number of effects (at $p < .05$) with tiny effect sizes. Hence, $\eta^2 \geq .005$ was adopted as the effect size indicator for filtering out the effects that are trivial in magnitude and evaluating the impact of the six design factors on the RB of different parameters.

Convergence. There were 48,000 datasets generated and each dataset was analyzed using one of the five different $\boldsymbol{\Sigma}$ matrices including ID, TOEP(2), AR(1), ARMA(1,1), and UN. All analyses converged when $\boldsymbol{\Sigma}$ was specified as either ID or AR(1), and 99.7% of the analyses converged when $\boldsymbol{\Sigma}$ was specified as TOEP(2). The convergence percentage decreased to 95.6% (i.e., 2109 out of 48,000 replications did not converge) when $\boldsymbol{\Sigma}$ was specified as ARMA(1,1). All analyses also converged when $\boldsymbol{\Sigma}$ was specified as UN. Only the converged results were adopted for further analyses.

[11] The structure of this dataset was a mixed structure with $\boldsymbol{\Sigma}$ specification as within-subject factor, and all other four design factors as between-subject factors. Analyzing the current dataset using a UANOVA model treating all five design factors as between-subject factors has the advantage of reducing the estimation difficulty, considering the impact of all five design factors on the target parameters simultaneously, providing a clear picture of the findings by reducing the number of trivial effects, and increasing the interpretability of the results because they are in a single metric. The major shortcoming is that the significance tests of the within-subject effects and the mixed interaction effects are too conservative. However, statistical significance is not the primary concern given the large sample size.

Relative bias of β_0 and β_1. Under the MLM framework, no η^2 of the five design factors and their interaction effects in the UANOVA model was larger than .005 when the relative bias (RB) of β_0 and β_1 were the dependent variables. The mean RB and simple bias[12] (SB) of both β_0 and β_1 were close to zero (i.e., $SB_{\beta0=0} = .000$, $RB_{\beta0>0} = .004$, $SB_{\beta1=0} = .000$, and $RB_{\beta1>0} = .000$). Similar results (i.e., no η^2 equal or larger than .005) were found when comparing UN with correct specification condition. The mean relative bias (RB) and simple bias (SB) of both β_0 and β_1 were close to zero under the UN specification (i.e., $SB_{\beta0=0} = .000$, $RB_{\beta0>0} = .004$, $SB_{\beta1=0} = -.001$, and $RB_{\beta1>0} = -.004$).

Relative bias of $SE_{\beta0}$ and $SE_{\beta1}$. For $SE_{\beta0}$, the specification category under the MLM framework had an impact on the estimation of $SE_{\beta0}$ ($\eta^2_{\mathbf{\Sigma}\,\text{Specification}} = .005$). As shown in Table 2, a larger overestimation of $SE_{\beta0}$ occurred when the $\mathbf{\Sigma}$ matrix was either under-specified or generally misspecified, whereas there was no significant difference on the relative bias of $SE_{\beta0}$ between over-specification and correct-specification of the $\mathbf{\Sigma}$ matrix. A similar pattern of results was also found for $SE_{\beta1}$ under the MLM framework ($\eta^2_{\mathbf{\Sigma}\,\text{Specification}} = .006$). That is, under-specification and general-misspecification led to larger estimated standard errors of β_1.

When comparing UN specification with correct specification under the mixed model framework, a three way interaction between number of cases, number of waves, and specification categories was found for $SE_{\beta0}$ ($\eta^2_{\text{Ncases}*\text{Nwaves}*\mathbf{\Sigma}\,\text{Specification}} = .007$; see Table 3). $SE_{\beta0}$ was generally underestimated (i.e., the estimate is smaller than the true parameter value) when $\mathbf{\Sigma}$ matrix was specified as UN. The range of the relative bias of $SE_{\beta0}$ was between $-.006$ and $-.119$ when $\mathbf{\Sigma}$ was specified as UN.

A similar three-way interaction effect for $SE_{\beta1}$ ($\eta^2_{\text{Ncases}*\text{Nwaves}*\mathbf{\Sigma}\,\text{Specification}} = .006$; see Table 3) was also found. In general, $SE_{\beta1}$ was also underestimated (i.e., smaller than the true parameter value) from .6% to 11.6% when $\mathbf{\Sigma}$ matrix was specified as UN.

Relative bias of τ_{01}, τ_{00}, and τ_{11}. The relative bias of τ_{01}, τ_{00}, and τ_{11} were only examined under the MLM framework because no element of the **T** matrix was estimated under the general mixed model framework (i.e., with null **T** matrix). No η^2 of the five design factors and the interaction terms in the UANOVA model was larger than .005 under the MLM framework when the RB of τ_{01} was the dependent variable. The mean relative bias of τ_{01} was close to zero ($RB\tau_{01} = .002$). On the other hand, a substantial interaction effect between the magnitude of **T** matrix and the specification of $\mathbf{\Sigma}$ matrix was found for the

[12]Instead of relative bias, simple bias (i.e., $\mathbf{SB} = \mathbf{E}(\hat{\theta}) - \theta$) was used as the dependent variable when the true value of both β_0 and β_1 was equal to .00.

TABLE 2
Impact of Different $\mathbf{\Sigma}$ Specifications on Relative Bias (RB) and Test of Significance for the Linear Growth Model under MLM framework

Parameter	*Effects with η^2 (in Parenthesis) Equal or Larger Than .005*	*Under-Specification*	*General-Misspecification*	*Over-Specification*	*Correct-Specification*
RB of $SE_{\beta 0}$	**Specification** (.005)	.007	.013	−.003	−.003
RB of $SE_{\beta 1}$	**Specification** (.006)	.014	.005	−.004	−.004
RB of τ_{00}	Teffect***Specification** (.033)	Small **T**			
		4.020	2.183	.393	.719
		Medium **T**			
		2.006	.956	.071	.254[a]
RB of τ_{11}	a) Nwaves***Specification** (.047)	Small **T**			
		1.114	.498	.070	.138
		Medium **T**			
		.303	.111	.011	.027
	b) Teffect***Specification** (.015)	Small **T**			
		.949	.426	.079	.120
		Medium **T**			
		.469	.182	.002	.045
Statistical power for detecting β_0	**Specification** ($\eta^2 = .490$, $p < .001$)	.280	.295	.382	.322[b]

[a]This is the mean relative bias value for correct-specification with medium **T** (averaging across different sample size conditions). Under the correct specification condition with $N = 210$, 8 waves and medium **T** matrix, the Relative Bias of τ_{00} and τ_{11} were .002 and .004, respectively.

[b]We set the intercept (β_0) to .10 and the effect of the linear growth parameter (β_1) to .05 (small effect) and .16 (medium effect). This value (.322) is the average statistical power across all conditions under the correct specification. For condition with $N = 210$, 8 waves and medium **T** matrix, the empirical power for detecting β_1 regardless the magnitude of β_1 is 1.000 for all four $\mathbf{\Sigma}$ specifications.

RB of τ_{00} ($\eta^2_{\text{magnitude of } \mathbf{T} \text{ matrix} * \mathbf{\Sigma} \text{ Specification}} = .033$; see Table 2). Under-specified and general-misspecified $\mathbf{\Sigma}$ matrices resulted in relatively larger overestimation of τ_{00}, and this overestimation became greater when the data contained a small **T** matrix, whereas over-specified $\mathbf{\Sigma}$ matrices always resulted in the smallest τ_{00} compared with other specification conditions regardless the magnitude of the **T** matrix.

Similarly, an interaction effect between the magnitude of **T** matrix and the $\mathbf{\Sigma}$ specification was also found for the RB of τ_{11} ($\eta^2_{\text{magnitude of } \mathbf{T} \text{ matrix} * \mathbf{\Sigma} \text{ Specification}} = .015$; see Table 2). Compared with other specification categories, the overestimation of τ_{11} was the largest when the $\mathbf{\Sigma}$ matrix was under-specified, and this overestimation increased under the small **T** matrix condition. A similar pattern of results was also found between number of waves and specification categories ($\eta^2_{\text{Nwaves} * \mathbf{\Sigma} \text{Specification}} = .047$; see Table 2). Under-specified and general-misspecified $\mathbf{\Sigma}$ matrices resulted in relatively larger overestimation of τ_{11}, and this overestimation became greater when the data contained a small number

TABLE 3
Impact of Unstructured $\mathbf{\Sigma}$ Specifications on Relative Bias (RB) and Test of Significance for the Linear Growth Model under the Mixed (Null **T**) Model Framework

Parameter	*Effects with η^2 (in Parenthesis) Equal or Larger Than .005*	*Null-**T** with UN $\mathbf{\Sigma}$ Matrix*	*Correct-Specification*
RB of $SE_{\beta 0}$	Ncases*Nwaves***Specification** (.007)	N = 30, 4 waves −.043	N = 30, 4 waves −.006[a]
		N = 30, 8 waves −.119	N = 30, 8 waves −.005
		N = 210, 4 waves −.006	N = 210, 4 waves −.001
		N = 210, 8 waves −.015	N = 210, 8 waves −.000
RB of $SE_{\beta 1}$	Ncases*Nwaves***Specification** (.006)	N = 30, 4 waves −.045	N = 30, 4 waves −.008
		N = 30, 8 waves −.116	N = 30, 8 waves −.006
		N = 210, 4 waves −.006	N = 210, 4 waves −.001
		N = 210, 8 waves −.016	N = 210, 8 waves −.001
Type I error rate for detecting β_0	**Specification** ($\eta^2 = .423$, $p < .05$)	.075	.051[b]
Type I error rate for detecting β_1	**Specification** ($\eta^2 = .357$, $p < .05$)	.074	.049

[a]This is the average relative bias of $SE_{\beta 0}$ under correct-specification with 30 observations and 4 waves per observation.

[b]This is the average empirical Type I error rate for detecting β_0 under correct-specification.

of waves. On the other hand, over-specified $\mathbf{\Sigma}$ matrices always resulted in the smallest τ_{11} compared with other specification conditions regardless the magnitude of the **T** matrix and the number of waves.

Type I error rate of detecting β_0 and β_1. The type I error rate was examined for those conditions in which the true parameter value of both β_0 and β_1 was equal to zero. The full design under MLM framework included a total of 32 different conditions (2 number of cases × 2 number of waves × 2 magnitude of **T** matrices × 4 $\mathbf{\Sigma}$ specification categories) under which the empirical type I error rates could be evaluated.[13] The impact of the $\mathbf{\Sigma}$ matrix misspecification on the type I error inflation for detecting both β_0 and β_1 was evaluated using

[13]The empirical type I error rate was the proportion of the significant effects within each condition.

UANOVA,[14] controlling for the other three design factors including number of cases, number of waves, and magnitude of the **T** matrix. The results showed that none of these factors had an appreciable impact on the type I error rate of testing either β_0 or β_1. The range of the mean type I error rate for testing β_0 over the four $\boldsymbol{\Sigma}$ specifications was between .048 and .051. The range of the mean type I error rate for testing β_1 over the same four $\boldsymbol{\Sigma}$ specifications was between .049 and .052.

The type I error rates between UN and correct specification were also compared using UANOVA, controlling for the other design factors. Significant differences between UN and correct specification on the type I error rates of both β_0 and β_1 were found ($p < .05$). For β_0, the mean type I error rate under the UN specification (mean $\alpha = .075$) was significantly larger than the correct specification condition (mean $\alpha = .051$). A similar result was also found for β_1. That is, the mean type I error rate of the UN specification (mean $\alpha = .074$) was significantly larger than the type I error rate when $\boldsymbol{\Sigma}$ was correctly specified (mean $\alpha = .049$).

Statistical power of detecting β_0 and β_1. Statistical power was only examined for those conditions in which the true parameter value of both β_0 and β_1 were each larger than zero.[15] The statistical power was represented by the proportion of the significant effects within each condition. The impact of different $\boldsymbol{\Sigma}$ specifications and other design factors on the statistical power for detecting β_0 and β_1 under the MLM framework[16] was examined by using a UANOVA model.[17] For the statistical power of testing β_0, different $\boldsymbol{\Sigma}$ specifications ($p < .001$) had appreciable impact over and above other design factors. On average, the over-specified $\boldsymbol{\Sigma}$ matrix (power = .38) resulted in higher statistical power for testing β_0 than the correctly specified $\boldsymbol{\Sigma}$ matrix (power = .32), whereas both under-specified $\boldsymbol{\Sigma}$ (power = .28) and general-misspecified $\boldsymbol{\Sigma}$ (power = .29) resulted in lower statistical power.

On the other hand, $\boldsymbol{\Sigma}$ specification categories had no significant impact on the statistical power for testing β_1 after adjusting for the effects of other design factors. The range of the mean statistical power of different $\boldsymbol{\Sigma}$ specifications

[14]Because of the small sample size (N = 32 conditions), only the main effects of the design factors were included in the model.

[15]The magnitude of β_0 was set to be constant (.10) for all conditions regardless the magnitude of β_1.

[16]We only examined the impact of the design factors on the statistical power of detecting β_0 and of β_1 under the MLM framework because there was no obvious type I error inflation under the MLM framework but substantial inflation under the mixed model framework (i.e., the UN-Null **T** condition).

[17]Only main effects of the design factors were included in the UANOVA model.

was between .546 and .556 with mean statistical power equal to .551. Because of the substantial inflated type I error rate, the empirical power for detecting β_0 and β_1 in the UN specification was not examined.

STUDY 2

In Study 1, we considered a linear growth model with homogeneous within-residual variances over time. In Study 2, we conducted additional simulations considering quadratic growth (Study 2A) and unequal within-residual variances over time (i.e., $\mathbf{\Sigma} = \text{UN}(1)$ structure; Study 2B). The details of these two additional simulation conditions are presented below.

Method for Study 2A: Quadratic Growth Model

Two of the five simulation conditions, including sample size (i.e., 30 vs. 210) and the $\mathbf{\Sigma}$ matrices (i.e., same 4 structures for generating data: ID, TOEP(2), AR(1) and ARMA(1,1) with same parameter values), were exactly the same as in Study 1. Because of the additional quadratic term (associated with a random effect) in the new model, there were more fixed parameters for the average growth trajectory (i.e., β_0: the coefficient of the intercept term; β_1: the coefficient associated with the linear growth term; β_2: the coefficient associated with the quadratic term) and more variance and covariance parameters of the random effects for the corresponding $\mathbf{T}$ matrix:

$$T = \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}$$

where τ_{00} is the variance of the random effect associated with the intercept term, τ_{11} is the variance of the random effect associated with the linear growth term and τ_{22} is the variance of the random effect associated with the quadratic growth term. τ_{01} (equal to τ_{10}) is the covariance between the intercept and the linear growth terms; τ_{02} (equal to τ_{20}) is the covariance between the linear and the quadratic growth terms; τ_{03} (equal to τ_{30}) is the covariance between the intercept and quadratic growth terms. Following Study 1, we set the intercept coefficient (β_0) equal to .10. For both linear (β_1) and quadratic (β_2) coefficients, we used .05 for small effect and .16 for medium effect. Similarly, the elements in $\mathbf{T}$ matrix were set to be:

$$T_{Medium} = \begin{bmatrix} .200 & .050 & .050 \\ .050 & .100 & .035 \\ .050 & .035 & .100 \end{bmatrix} \text{ and } T_{Small} = \begin{bmatrix} .100 & .025 & .025 \\ .025 & .050 & .018 \\ .025 & .018 & .050 \end{bmatrix}$$

where all the correlations (i.e., $\gamma = \frac{\tau_{xy}}{\sqrt{\tau_{xx}\tau_{yy}}}, x \neq y$) were equal to .35. Because of the increased number of parameters in the quadratic model,[18] we only considered the 8-wave condition in which we could obtain more stable estimates. The same evaluation criteria was used as in Study 1, including convergence, relative bias of the parameter estimates, type I error rate and statistical power, were also examined.

Results for Study 2A: Quadratic Growth Model

Convergence. Similar to the findings from Study 1, all analyses converged when $\mathbf{\Sigma}$ was specified as ID, TOEP(2), AR(1) or UN. The convergence rate dropped slightly (96.4%) when $\mathbf{\Sigma}$ was specified as ARMA(1,1).

Relative bias. Following the analysis procedures presented in Study 1, we also found a very similar pattern of results from the simulations based on the quadratic growth model. There was no η^2 of the five design factors and their interaction effects larger than .005 when examining the relative bias of the fixed parameter estimates (i.e., β_0, β_1, and β_2) under either the MLM or mixed model frameworks. On the other hand, under the MLM framework, $\mathbf{\Sigma}$ specifications had substantial impact on the estimation of the standard errors of the fixed parameters (i.e., $SE_{\beta 0}$, $SE_{\beta 1}$, and $SE_{\beta 2}$; see Table 4). On average, under-specified and general-misspecified $\mathbf{\Sigma}$ resulted in overestimated (or larger) standard errors (except for $SE_{\beta 0}$ at the underspecified $\mathbf{\Sigma}$ condition), whereas over-specified $\mathbf{\Sigma}$ always resulted in similar estimates to those of the correctly-specified $\mathbf{\Sigma}$. Under the mixed model framework with the UN $\mathbf{\Sigma}$ specification, the standard errors of the fixed effects were generally underestimated (i.e., smaller than the true parameter value; see Table 5), and the underestimation became larger when the size of the sample was small ($N = 30$).

For the relative bias of the random effect variances (i.e., τ_{00}, τ_{11}, and τ_{22}), a similar pattern of results as presented in Study 1 was also found in Study 2 ($\eta^2_{\mathbf{\Sigma}\,\text{Specification}*\mathbf{T}\,\text{Matrix}} > .005$). That is, compared with the over-specified and correct-specified $\mathbf{\Sigma}$ conditions, under-specified and general-misspecified $\mathbf{\Sigma}$ resulted in substantially overestimated variances, especially in the condition with a small $\mathbf{T}$ matrix. Additionally, $\mathbf{\Sigma}$ specifications had significant impact on the relative bias of the covariance between the intercept and the quadratic growth (i.e., τ_{02}). This covariance was substantially underestimated in both under-specified and general-misspecified $\mathbf{\Sigma}$ conditions, and this underestimation became larger under the small $\mathbf{T}$ condition.

[18]In the ARMA(1,1) condition, there are 9 parameters associated with the random effects (i.e., τ_{00}, τ_{11}, τ_{22}, τ_{01}, τ_{02}, τ_{12}, σ^2, ρ, and γ), whereas there are only 10 moments (or pieces of information) under the 4-wave condition.

TABLE 4
Impact of Different $\boldsymbol{\Sigma}$ Specifications on Relative Bias (RB) and Test of Significance for the Quadratic Growth Model under MLM Framework

Parameter	*Effects with η^2 (in Parenthesis) Equal or Larger Than .005*	*Under-Specification*	*General-Misspecification*	*Over-Specification*	*Correct-Specification*
RB of $SE_{\beta 0}$	**Specification** (.008)	.001	.020	−.001	.003[a]
RB of $SE_{\beta 1}$	**Specification** (.010)	.019	.010	−.003	−.004
RB of $SE_{\beta 2}$	**Specification** (.004)	.013	.004	−.002	−.004
RB of τ_{00}	Teffect*$\textbf{Specification}$	Small **T**			
	(.039)	5.026	2.724	.547	.942
		Medium **T**			
		2.498	1.204	.170	.345
RB of τ_{11}	Teffect***Specification**	Small **T**			
	(.038)	1.744	.739	.135	.241
		Medium **T**			
		.877	.316	.051	.103
RB of τ_{21}	Teffect***Specification**	Small **T**			
	(.029)	−3.222	−1.549	−.240	−.431
		Medium **T**			
		−1.575	−.600	−.060	−.147
RB of τ_{22}	Teffect***Specification**	Small **T**			
	(.018)	.747	.312	.050	.072
		Medium **T**			
		.366	.119	.019	.025
Type I error rate for detecting β_1	**Specification** ($\eta^2 = .676$, $p < .010$)	.058	.058	.052	.057[b]
Statistical Power for detecting β_0	**Specification** ($\eta^2 = .642$, $p < .001$)	.231	.228	.330	.266[c]

[a]This is the mean relative bias of $SE_{\beta 0}$ under correct-specification (averaging across different sample size conditions).

[b]This is the mean empirical type I error rate for detecting β_1 under correct-specification (averaging across different sample size conditions).

[c]This is the mean empirical power for detecting β_0 under correct-specification (averaging across different sample size conditions).

Type I error of detecting the three fixed effects. Under the MLM framework, $\boldsymbol{\Sigma}$ specification and other design factors had no significant effect on the type I error rate for tests of β_0 (grand mean $\alpha = .047$) or β_2 (grand mean $\alpha = .056$); whereas, the type I error rate for testing β_1 (range between .052 and .058) was found to differ between the different $\boldsymbol{\Sigma}$ specifications in which over-specified $\boldsymbol{\Sigma}$ had the smallest α (.052). Under the mixed model framework, the UN $\boldsymbol{\Sigma}$ specification resulted in a substantial inflation of the type I error rates for testing the three fixed effects (range of mean α between .083 and .085).

Statistical power. Similar to the findings from the linear growth model presented in Study 1, $\boldsymbol{\Sigma}$ specification under the MLM framework had significant impact on the statistical power for detecting β_0 ($\eta^2 = .642$, $p < .001$) but not the other two fixed effects (i.e., β_1 and β_2). Over-specified $\boldsymbol{\Sigma}$ resulted in higher

TABLE 5
Impact of Unstructured $\mathbf{\Sigma}$ Specifications on Relative Bias (RB) and Test of Significance for the Quadratic Growth Model under the Mixed (Null T) Model Framework

Parameter	*Effects with η^2 (in Parenthesis) Equal or Larger Than .005*	*Null-**T** with UN $\mathbf{\Sigma}$ Matrix*	*Correct-Specification*
RB of $SE_{\beta 0}$	Ncases*$\mathbf{Specification}$ (.045)	N = 30	N = 30
		−.098	−.001[a]
		N = 210	N = 210
		−0.13	.000
RB of $SE_{\beta 1}$	Ncases***Specification** (.041)	N = 30	N = 30
		−.099	−.007
		N = 210	N = 210
		−0.13	−.001
RB of $SE_{\beta 2}$	Ncases***Specification** (.040)	N = 30	N = 30
		−.099	−.006
		N = 210	N = 210
		−0.13	−.001
Type I error rate for detecting β_0	**Specification** ($\eta^2 = .645$, $p < .10$)	.083	.048[b]
Type I error rate for detecting β_1	**Specification** ($\eta^2 = .579$, $p < .10$)	.085	.057
Type I error rate for detecting β_2	**Specification** ($\eta^2 = .591$, $p < .10$)	.083	.056

[a]This is the average relative bias of $SE_{\beta 0}$ under correct-specification with 30 observations.
[b]This is the average empirical Type I error rate for detecting β_0 under correct-specification (averaging across different sample size conditions).

statistical power (.330) than both under-specified $\mathbf{\Sigma}$ (.231) and general-misspecified $\mathbf{\Sigma}$ (.228). The average statistical power for detecting β_1 and β_2 was .502 and .544, respectively.

Method for Study 2B: Model with Unequal Within-Residual Variances over Time

In this study we generated data using the linear growth model as shown in equations (1) and (2) with unequal within-residual variances and zeros off diagonal for the $\mathbf{\Sigma}$ matrix (i.e., unstructured(1) or UN(1) structure). Then, we analyzed the data using different $\mathbf{\Sigma}$ matrices as discussed before. Hence, we could only examine the impact of under-specification (i.e., specified $\mathbf{\Sigma}$ matrix as ID) and general-misspecification (specified $\mathbf{\Sigma}$ matrix as TOEP(2), or AR(1), or ARMA(1,1)) but not over-specification on the estimation of the target parameters (i.e., β_0, β_1, $SE_{\beta 0}$, $SE_{\beta 1}$, τ_{00}, τ_{01}, and τ_{11}), and the tests of significance (i.e., type I error rate and statistical power) of the fixed effect parameters (i.e., β_0 and β_1). Except for the number of waves and the parameter values of the elements in the $\mathbf{\Sigma}$ matrix for data generation, all other conditions (i.e., number of participants, magnitude of the growth parameters, and magnitude of the elements in the $\mathbf{T}$ matrix) were exactly the same as in Study 1. To reduce the number of parameters[19] for the $\mathbf{\Sigma}$ matrix specification, we only considered a 4-wave condition with the true $\mathbf{\Sigma}$ matrix equal to:

$$UN(1) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix} = \begin{bmatrix} 1.00 & 0 & 0 & 0 \\ 0 & .800 & 0 & 0 \\ 0 & 0 & .640 & 0 \\ 0 & 0 & 0 & .512 \end{bmatrix}$$

where the residual variance of the first time point (σ_1^2) was set to 1.00 and the size of each of the following residual variances was 80% of the previous residual variance. The same evaluation criteria as presented in Study 1 were also used here.

Results for Study 2B: Model with Unequal Within-Residual Variances over Time

Convergence. Similar to the findings presented in previous sections, all analyses converged when $\mathbf{\Sigma}$ was specified as ID, TOEP(2), AR(1) or UN. The convergence rate dropped to 88% when $\mathbf{\Sigma}$ was specified as ARMA(1,1).

Relative bias. We found a pattern of results for the condition with heterogeneous within-residual variances that was similar to the results presented in previous sections. There was no η^2 of the design factors and their interaction effects larger than .005 when examining the relative bias of the fixed parameter estimates (i.e., β_0 and β_1) under either MLM or mixed model framework. On the other hand, under the MLM framework, $\mathbf{\Sigma}$ specifications had appreciable impact (i.e., $\eta^2 > .005$) on the estimation of the standard errors of β_1 (i.e., $SE_{\beta 1}$, see Table 6). Under-specified or general-misspecified $\mathbf{\Sigma}$ resulted in larger standard errors than the correctly specified $\mathbf{\Sigma}$ condition. Under the mixed model framework with UN $\mathbf{\Sigma}$ specification, the standard errors of the fixed effects were generally underestimated (i.e., relative bias of: $SE_{\beta 0} = -.025$ and $SE_{\beta 1} = -.025$).

[19]The number of parameters in the $\mathbf{\Sigma}$ matrix with an UN(1) structure varies as the number of waves (or repeated measures) varies. For example, the number of parameters in the $\mathbf{\Sigma}$ matrix with an UN(1) structure is 4 for a balanced 4-wave study, whereas the number of parameters for the same $\mathbf{\Sigma}$ structure will become 8 for a balanced 8-wave study.

TABLE 6
Impact of Different $\boldsymbol{\Sigma}$ Specifications on Relative Bias (RB) and Test of Significance for the Linear Growth Model under MLM framework

Parameter	*Effects with η^2 (in Parenthesis) Equal or Larger Than .005*	*Under-Specification*	*General-Misspecification*	*Correct-Specification*
RB of $SE_{\beta 1}$	**Specification** (.007)	.011	.005	−.013[a]
RB of τ_{01}	Teffect****Specification** (.018)	τ −1.632	−1.625	−.026
		T −.827	−.820	−.025
RB of τ_{11}	**Specification** (.003)	.089	.141	.044

[a]This is the average relative bias of $SE_{\beta 0}$ under correct-specification (averaging across different sample size conditions).

For the relative bias of the random effect variances and covariance (i.e., τ_{00}, τ_{01}, and τ_{11}), we found that the $\boldsymbol{\Sigma}$ specification had substantial impact on the estimation of both τ_{01} ($\eta^2_{\boldsymbol{\Sigma}\,\text{Specification}*\mathbf{T}\,\text{Matrix}} = .018$) and τ_{11} ($\eta^2_{\boldsymbol{\Sigma}\,\text{Specification}} = .003$) but not τ_{00} (see Table 6). That is, compared with correctly specified $\boldsymbol{\Sigma}$, under-specified and general-misspecified $\boldsymbol{\Sigma}$ resulted in substantial overestimation in τ_{11} and underestimation in τ_{01}, and the underestimation of τ_{01} became larger under small **T** matrix condition.

Type I error of detecting the two fixed effects. Under the MLM framework, $\boldsymbol{\Sigma}$ specification and other design factors had no significant effect on the type I error rate for testing β_0 (grand mean $\alpha = .050$) and β_1 (grand mean $\alpha = .065$). Similarly, $\boldsymbol{\Sigma}$ specification and other design factors had no significant effect on the type I error rate for testing β_0 (grand mean $\alpha = .055$) and β_1 (grand mean $\alpha = .062$) under the mixed model framework.

Statistical power. $\boldsymbol{\Sigma}$ specification and other design factors had no significant effect on the statistical power of detecting β_0 and β_1 under either MLM framework (grand mean power for detecting: $\beta_0 = .433$ and $\beta_1 = .477$) or mixed model framework (grand mean power for detecting: $\beta_0 = .440$ and $\beta_1 = .485$).

DISCUSSION

The main purpose of this study was to examine the impact of misspecifying the within-subject covariance matrix (i.e., $\boldsymbol{\Sigma}$ matrix) on the estimation of the target parameters including fixed effects, their corresponding standard errors, the variances and covariances of the random effects, and the type I error rate and statistical power of testing the fixed effects. All models examined in the present study were population mean centered in time to minimize random error in the estimates of β_0. The simulation results implied a general pattern of the effects of misspecifying the $\mathbf{\Sigma}$ matrix on the estimation of the random effect variances and the standard errors of the growth parameters. Under-specification and general misspecification of the $\mathbf{\Sigma}$ matrices were more likely to result in overestimation in both random effect variances and the standard errors[20] of the growth parameters, which in turn, resulted in lower statistical power relative to correct-specification. On the other hand, compared with the correct-specification condition, over-specification of the $\mathbf{\Sigma}$ matrix was more likely to result in smaller estimates of the random effect variances, which in turn, resulted in similar or slightly smaller standard errors of the growth parameters. These results led to a possible gain in statistical power relative to other specification categories. The UN specification generally resulted in underestimation of the standard errors of the growth parameters, leading to a greater likelihood of type I error inflation when testing the growth parameters. Figure 2 summarizes these findings.

Bias in Estimation of the Random Effect Variances

The structure of the multilevel model provides a possible explanation of the bias in estimation of the random effect variances (e.g., τ_{00} and τ_{11}). As shown in Equation (2), repeated below, the general form of a two-level linear model can be written as following:

$$\underline{y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\underline{u} + \underline{e}$$

Based on Equation (2), the expected value of $\underline{y}$ is:

$$E(\underline{y}) = \mathbf{X}\underline{\beta} \tag{12}$$

The variance of $\underline{y}$ is:

$$V(\underline{y}) = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R} \tag{13}$$

[20]The specification of $\mathbf{\Sigma}$ also had significant impact on the covariance between the intercept and the quadratic growth terms (Study 2A) and the covariance between intercept and the linear growth terms (Study 2B). The under-specified and general-misspecified $\mathbf{\Sigma}$ generally resulted in substantially underestimed covariances which could potentially cancel out the impact of these two $\mathbf{\Sigma}$ specifications on the (over)estimation of the variances of the random effects and standard errors (SE) of the corresponding growth parameters (e.g., η^2 of $\mathbf{\Sigma}$ specifications on SE of the quadratic growth term was .004 in Study 2A and η^2 of $\mathbf{\Sigma}$ specifications on SE of the intercept term was .001 in Study 2B)

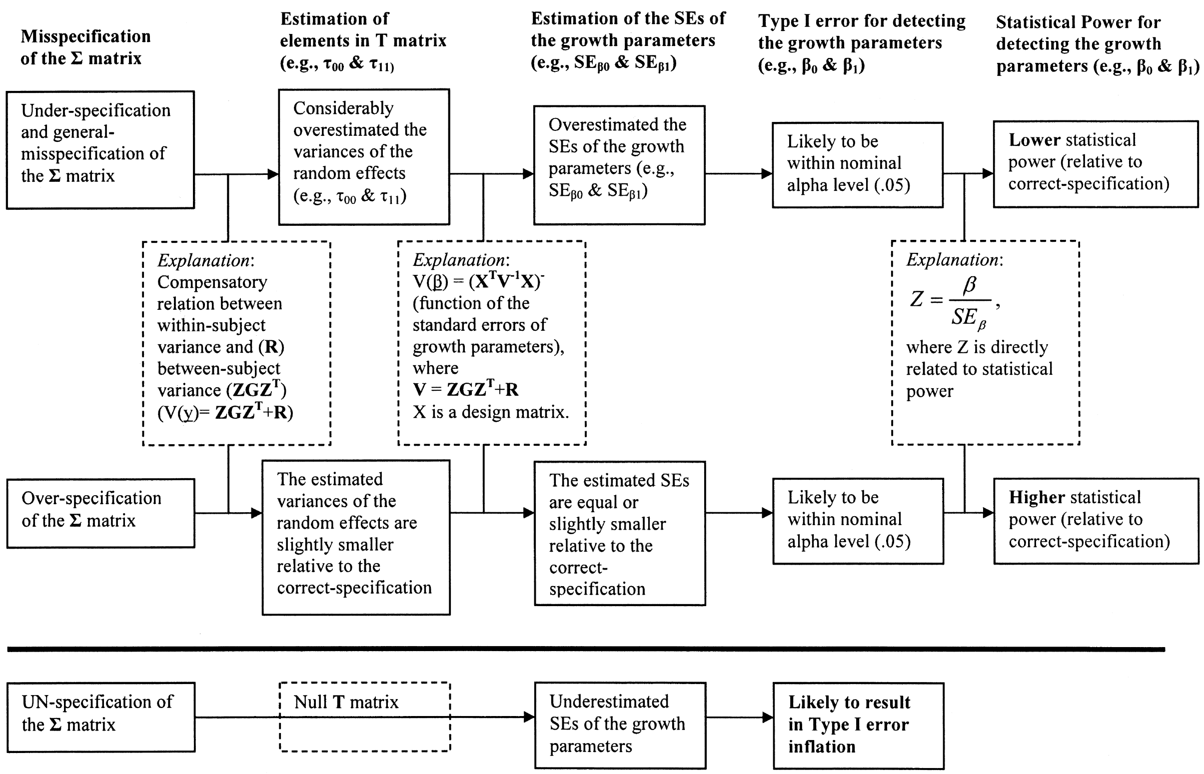

FIGURE 2 Summary of the impact of misspecifying the $\boldsymbol{\Sigma}$ matrix.

$E(\underline{y})$ and $V(\underline{y})$ are independent of each other given standard assumptions. In Equation (12), $\underline{\beta}$ contains the growth parameters, which are generally unbiased regardless of the effect of misspecifying the $\mathbf{\Sigma}$ matrix.[21] On the other hand, Equation (13) shows that the *total variance* of a two-level model (i.e., $V(\underline{y})$) is constructed from two separate sources of variance: variance from the between-subject random effects (i.e., the $\mathbf{ZGZ^T}$ part shown in Equation 13), and variance from the within-subject random errors (i.e., the $\mathbf{R}$ matrix shown in Equation 13). After controlling for the fixed effect part (i.e., $\mathbf{X}\underline{\beta}$), the total variance becomes a fixed quantity for a given dataset regardless of the structure of the sources of the variance. This result implies that the between-subject variance and the within-subject variance have a compensatory relationship.[22] For example, when one of these two sources of variance is smaller than the true value, the other source will be overestimated to compensate. This compensatory relationship tends to keep the total variance constant.

As shown in Equation (3), the core element of the $\mathbf{G}$ matrix is the $\mathbf{T}$ matrix, which contains the variances (i.e., τ_{00} and τ_{11}) and covariance (i.e., τ_{01}) of the random effects in the linear growth model. On the other hand, the core element of $\mathbf{R}$ matrix is the $\mathbf{\Sigma}$ matrix, which contains the variance of the residuals at each time point and the covariance of the residuals between each pair of time points. Generally, no specific structure is imposed on the $\mathbf{T}$ matrix when analyzing longitudinal data using the MLM approach. In contrast, the default structure of the common statistical packages is that the $\mathbf{\Sigma}$ matrix is assumed to be identity (ID) under the MLM framework. However, ID is a very restrictive assumption for longitudinal data, because it assumes no relation between the residuals of any pair of time points once the random effects have been modeled. The presumption of the $\mathbf{\Sigma}$ matrix as ID is unlikely to capture the true relation between the residuals of different time points (Goldstein, Healy, & Rasbash, 1994). This result means that the $\mathbf{\Sigma}$ matrix will typically be underspecified in longitudinal analyses. To compensate for the reduction in the contribution of the within-subject variance (i.e., the under-specified $\mathbf{\Sigma}$ matrix), the elements in the $\mathbf{T}$ matrix will tend to become larger to capture this missing variance as part of the between-subject variance. This result, in turn, can lead to the overestimation of the random effect variance terms (e.g., τ_{00} and τ_{11}). On the other hand, over-specification results in similar or slightly smaller estimates of the random effect variances relative to the correct-specification.

[21] Similar results (i.e., unbiased fixed effect estimates) were also found in other studies on the misspecification of the variance structure (i.e., $V(\underline{y})$) including: under-specifying the $\mathbf{\Sigma}$ matrix (Ferron et al., 2002), ignoring a nested level (Moerbeek, 2004) or a crossed level (Meyers & Beretvas, 2006) in multilevel models.

[22] This compensatory relationship has also been shown in other MLM related simulation studies (Moerbeek, 2004; Meyers & Beretvas, 2006).

Bias in Estimation of the Standard Errors of the Growth Parameters

Following the simple linear growth model presented in Equations (1) and (2), the variances and covariance of the growth parameters are equal to:[23]

$$V(\underline{\beta}) = (X^T V^{-1} X)^- \tag{14}$$

where

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \tag{15}$$

and

$$V = ZGZ^T + R \tag{16}$$

and X is the design matrix containing the intercept and the time predictor. The main diagonal elements of the square root of $V(\underline{\beta})$ are the standard errors of the two growth parameters. Hence, the standard errors of the growth parameters are a direct function of the **T** (i.e., core elements of **G**) and **Σ** (i.e., core elements of **R**) matrices. Recall that under-specification and general-misspecification of the **Σ** matrix generally led to overestimates of the elements of the **T** matrix (especially τ_{00} and τ_{11}), which in turn, likely led to overestimation of the standard errors of the growth parameters. On the other hand, the estimated standard errors of the growth parameters given over-specification were not appreciably different from the standard error estimates under correct-specification. This finding may be due to the slightly underestimated variances and covariance of the random effects in the over-specification relative to the correct-specification conditions. UN, can be viewed as the extreme case of the over-specification of the **Σ** matrix with all random effect variances in **G** matrix specified as zeros. UN resulted in substantial underestimation of the standard errors of the growth parameters.

Type I Error Rate and Statistical Power for Detecting the Growth Parameters

The Wald Z-value of a growth parameter is equal to:

$$Z = \frac{\beta}{SE_\beta} \tag{17}$$

[23] $(X^T V^{-1} X)^-$ is a generalized inverse. Suppose A^- is a generalized inverse, then $AA^-A = A$.

This Z-value is directly related to the type I error rate and statistical power. As shown in equation (17), when the standard error of a growth parameter decreases, the corresponding Z-value will increase. Given a specified alpha level, the over-estimated standard errors of the growth parameters in the under-specification and general-misspecification conditions are more likely to result in lower statistical power relative to the correct-specification condition. On the other hand, over-specifying the $\mathbf{\Sigma}$ matrix to a structure within which the true $\mathbf{\Sigma}$ matrix is *nested* may result in higher statistical power relative to the correct-specification without increasing the type I error rate beyond the nominal level. As indicated previously, the variances of the random effects were generally overestimated for all conditions in which a $\mathbf{T}$ matrix was estimated. Compared with other conditions, over-specification of the $\mathbf{\Sigma}$ matrix resulted in better estimates of the random effect variances, which in turn, reduced the standard errors toward the theoretically correct values. In contrast, leaving the $\mathbf{\Sigma}$ matrix as unstructured and the $\mathbf{T}$ matrix as a null matrix will result in underestimation of the standard errors, which in turn, likely leads to the type I error inflation when testing the growth parameters.

The simulation results showed that the four specification categories of the MLM framework only had an impact on the statistical power of testing β_0 but not other growth parameters (e.g., β_1 in linear growth model). A possible explanation of this result is the magnitude of the random effect variances specified in the simulation. In longitudinal data, the variation between intercepts is typically larger than the variation between slopes, so we set τ_{00} (i.e., variance of the intercept term) to be twice of the variance of other random effect (e.g., τ_{11}, variance of the linear growth term) for the simulation. In this simulation, the overall estimation bias was generally larger in τ_{00} than in τ_{11}. The same pattern was also found for $SE_{\beta 0}$ and $SE_{\beta 1}$. The estimation bias of $SE_{\beta 0}$ was generally larger than the estimation bias of $SE_{\beta 1}$, which in turn, had direct impact on the statistical power of detecting the growth parameters.

Recommendation

Misspecification of the $\mathbf{\Sigma}$ matrix has a substantial impact on the estimation of the random effect variances, which in turn, affects the estimation of the standard errors and the test of significance of the growth parameters. Leaving the $\mathbf{\Sigma}$ matrix as unstructured is very likely to underestimate the standard errors of the growth parameters and produce type I error inflation when testing the growth parameters. Moreover, using an unstructured $\mathbf{\Sigma}$ matrix results in the loss of information concerning the magnitude of the variation of the random intercepts and slopes, which can be used for calculating the pseudo R^2 (Singer & Willett, 2003). We recommend that researchers who believe that there may be random effects associated with growth parameters should attempt to find an optimal within-subject covariance matrix when analyzing longitudinal data.[24]

Ideally, researchers should be able to draw on substantive or statistical theory in specifying the within-subject covariance structure. Occasionally, substantive theory will specify that the variance is expected to change over time as in the Hedeker and Mermelstein (in press) study cited earlier. However, psychological theory almost never provides a clear basis for expecting a specified structure of within subject error covariances over time. Generalizing from the time series literature (e.g., Velicer & Fava, 2003; West & Hepworth, 1991), the error structure appears to be a function of the lag between adjacent measurements, whether there are periods in which measurements are not collected (e.g., periods of sleep), and the nature of the phenomenon under investigation. Particular problems arise in the study of phenomena in which there are within day (e.g., activity level), weekly (e.g., alcohol consumption), or monthly cycles (e.g., consumer behavior) in the data. Such cycles should ideally be included as part of the level 1 model of the data (see Armeli et al., 2000). Given that cycles have been properly modeled, researchers should consider carefully the nature of the phenomenon under study and the lag between adjacent measurements and choose a within subject error structure that appears to match their investigation. If they are uncertain, we encourage researchers to err in the direction of adopting a slightly over-specified $\boldsymbol{\Sigma}$ matrix[25] (e.g., TOEP(2) or AR(1) which are commonly used in longitudinal data analysis under SEM framework; Bollen & Curran 2006) if they have a *balanced* design.

Some Limitations of the Present Studies

As with any simulation study, one of the major potential limitations of this study is the generalizability of the findings. Further examination of the applicability of the current findings to the misspecification of the $\boldsymbol{\Sigma}$ matrix across a broader range of models (e.g., three or higher-level model with multiple predictors from different levels, and unbalanced data) and conditions is needed. A preliminary examination of the impact of different $\boldsymbol{\Sigma}$ specifications on a linear growth model with unbalanced data showed a pattern of results that was similar to the balanced condition under the mixed model framework (i.e., UN $\boldsymbol{\Sigma}$ specification with

[24]Several studies have examined the accuracy of the traditional search indices such as AIC and BIC in searching for the optimal covariance structure. The results have shown that these traditional indices provided only marginally to moderately acceptable accuracy (Ferron, et al., 2002; Keselman et al., 1998; Wolfinger, 1993). The development of more accurate search indices is needed.

[25]AR(1) is one of the commonly used covariance structures when analyzing longitudinal data (J. Hilbe & J. Hardin, personal communication, August 18, 2006), which has also been applied for the latent growth models under the SEM framework (Bollen & Curran, 2004; 2006; Curran & Bollen, 2001).

null **T** matrix). Under the MLM framework, the pattern of results was also similar between the balanced and unbalanced conditions with one exception: the standard errors of the growth parameters were less affected with unbalanced data. To date, little simulation work has addressed the effects of non-normality of the within-subject error structure on the performance of multilevel models and further study on this issue is needed.

The impact of the bias estimation of the random effect variances on the calculation of the explained variance (or pseudo R^2) has not yet been examined. There are two issues related to the biased estimation of the random effect variances: 1) the rationale for adding the level-2 predictors; 2) the calculation of explained variance. Typically, researchers add level-2 predictor when appreciable random effect variance exists in the data (Snijders & Bosker, 1999). The bias in estimation of the random effect variances, especially in the under-specification conditions, implies that the occurrence of the random effect variance may be due to the misspecification of the $\mathbf{\Sigma}$ matrix rather than the true existence of the variation between intercepts and between slopes within the data. Indeed, the impact of the added level-2 predictors is evaluated by the change of the random effect variances after including the level-2 predictors in the model. Hence, the calculation of the explained variance may not be valid because of the bias in the estimation of the random effect variance.

REFERENCES

Algina, J., & Keselman, H. J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods, 2(2)*, 208–218.

Armeli, S., Carney, M. A., Tennen, H., Affleck, G., & O'Neil, T. P. (2000). Stress and alcohol use: A daily process examination of the stressor-vulnerability model. *Journal of Personality and Social Psychology*, *78*, 979–994.

Biesanz, J. C., West, S. G., & Kwok, O. (2003). Personality over time: Methodological approaches to the study of short-term and long-term development and change. *Journal of Personality, 71,* 905–941.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research, 32*, 336–383.

Bollen, K. A., & Curran, P. J. (2006). Latent curve models: A structural equation perspective. Hoboken, NJ: John Wiley & Sons.

Bosker, R. J., Snijders, T. A. B., & Guldemond, H. (2003). *PINT: Power IN Two-level Designs: User's Manual (V2.1)*. Retrieved from http://stat.gamma.rug.nl/

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.

Chi, E. M., & Reinsel, G. C. (1988). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association, 84,* 452–459.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Mahwah, NJ: Erlbaum.

Cook, D. R., & Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York, NY: John Wiley & Sons.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New Methods for the Analysis of Change* (pp. 107–135). Washington, DC: American Psychological Association.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. San Diego, CA: Springer.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics, 44*, 959–971.

Diggle, P. J., Heagerty, P., Liang, K., & Zeger, S. L. (2002). *Analysis of Longitudinal Data (2nd Ed.)*. New York, NY: Oxford.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research, 37*, 379–403.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from children through adolescence: Growth trajectories in two male-sex typed domains. *Developmental Psychology, 38*, 519–533.

Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13*, 1643–1655.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling, 9*, 347–368.

Hedeker, D., & Mermelstein, R. (2007). Mixed-effects regression models with heterogeneous variance: Analyzing ecological momentary assessment (EMA) data of smoking. In T. D. Little, J. A. Bovaird, and Noel A. Card (Eds.), *Modeling contextual effects in longitudinal studies*. Mahwah, NJ: LEA.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: gender and domain differences across grades one through twelve. *Child Development, 73*, 509–527.

Jones, R. H., & Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics, 47*, 161–175.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology, 54*, 1–20.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics-Simulation and Computation, 27*, 591–604.

Khoo, S., West, S. G., Wu, W., & Kwok, O. (2006). Longitudinal Methods. In M. Eid & E. Diener (Eds.), *Handbook of Psychological Measurement: A Multimethod Perspective* (pp. 301–317). Washington, DC: APA.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd Ed.). Pacific Grove, CA: Brooks/Cole.

Kowalchuk, R. K., & Keselman, H. J. (2001). Mixed-model pairwise multiple comparisons of repeated measures means. *Psychological Methods, 6*, 282–296.

Laird, N., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963–974.

Lagattuta, K. H., & Wellman, H. M. (2002). Differences in early parent-child conversations about negative versus positive emotions: Implications for the development of psychological understanding. *Developmental Psychology, 38*, 564–580.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Linear Mixed Models*. NC: SAS Institute.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd Ed.). Mahwah, NJ: LEA.

McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development, 44*, 401–415.

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*(4), 473–497.

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*, 129–149.

Muthén, B., & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*, 371–402.

Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide* (V2.0). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide* (V4.1). Los Angeles, CA: Muthén & Muthén.

Myung, J. I., & Pitt, A. A. (2004). Model comparison methods. *Methods in Enzymology, 383*, 351–366.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd Ed.).* Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. (2000). *HLM5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: SSI.

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6,* 387–401.

Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 32*, 269–286.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford University Press.

Sivo, S. A., & Willson, V. L. (2000). Modeling causal error structures in longitudinal panel data: A Monte Carlo study. *Structural Equation Modeling, 7*, 174–205.

Sivo, S. A., Fan, X., & Witta, L. (2005). The biasing effects of unmodeled ARMA time series processes on latent growth curve model estimates. *Structural Equation Modeling, 12*, 215–231.

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*, 237–259.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* CA: Sage Publications.

Velicer, W. F., & Fava, J. L. (2003). Time series analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology*. Vol. 2: *Research methods in psychology* (pp. 581–606). Hoboken, NJ: Wiley.

Weiss, R. E. (2005). *Modeling Longitudinal Data*. New York, NY: Springer.

West, S. G., & Hepworth, J. T. (1991). Data analytic strategies for temporal data and daily events. *Journal of Personality, 59,* 60-9-662.

Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics—Simulation and Computation, 22*, 1079–1106.

Wolfinger, R. D. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1*, 205–230.

APPENDIX A

SAS script for Analyzing the Linear Growth Model in Study 1 with Different $\mathbf{\Sigma}$ Specifications

a) Linear growth model with **Σ** specified as ID structure

```
proc mixed data=all;
class Subj_id;
model y=time /solution;
random intercept time / type=un subject=Subj_id;
repeated / type=ID subject=Subj_id;
```

b) Linear growth model with **Σ** specified as TOEP(2) structure

```
proc mixed data=all;
class Subj_id;
model y=time /solution;
random intercept time / type=un subject=Subj_id;
repeated / type=TOEP(2) subject=Subj_id;
```

c) Linear growth model with **Σ** specified as AR(1) structure

```
proc mixed data=all;
class Subj_id;
model y=time /solution;
random intercept time / type=un subject=Subj_id;
repeated / type=AR(1) subject=Subj_id;
```

d) Linear growth model with **Σ** specified as ARMA(1,1) structure

```
proc mixed data=all;
class Subj_id;
model y=time /solution;
random intercept time / type=un subject=Subj_id;
repeated / type=ARMA(1,1) subject=Subj_id;
```

e) Linear growth model with **Σ** specified as UN structure (and null T matrix)

```
proc mixed data=all;
class Subj_id;
model y=time /solution;
repeated / type=UN subject=Subj_id;
```

Note.* ***class*: specifies categorical variable; Subj_id: Subject ID (i.e., level-2 cluster ID); ***model***: specify the average growth (or fixed-effect) model; ***solution***: request for growth parameter estimates (i.e., β_0 & β_1) and their corresponding standard errors (i.e., $SE_{\beta 0}$ & $SE_{\beta 1}$); ***random***: specify the random effects (or request for estimating the **T** matrix); ***type=UN*** (next to the "random" command): specify the structure of the **T** matrix as unstructured (UN) structure; ***subject***: specify the level-2 cluster ID; ***repeated***: request for estimating the **Σ** matrix; ***type=ID*** (next to the "repeated" command): specify the structure of the **Σ** matrix as ID structure.