7-12-2014

# Ignoring individual differences in times of assessment in growth curve modeling

Patrick Coulombe

Follow this and additional works at: https://digitalrepository.unm.edu/psy_etds

Patrick Coulombe
*Candidate*

Psychology
*Department*

This thesis is approved, and it is acceptable in quality and form for publication:

*Approved by the Thesis Committee:*

Harold Delaney , Chairperson

James Selig , Co-chair

Timothy Goldsmith

**IGNORING INDIVIDUAL DIFFERENCES IN TIMES OF ASSESSMENT IN GROWTH CURVE MODELING**

**by**

**PATRICK COULOMBE**

**B.S., PSYCHOLOGY, UNIVERSITÉ DU QUÉBEC À MONTRÉAL, 2010**

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Master of Science
Psychology**

The University of New Mexico
Albuquerque, New Mexico

**May 2014**

# DEDICATION

To my brother, Philippe Coulombe, who is 2100 miles (3400 km) away.

# ACKNOWLEDGMENTS

# IGNORING INDIVIDUAL DIFFERENCES IN TIMES OF ASSESSMENT IN GROWTH CURVE MODELING

by

**Patrick Coulombe**

**B.S., Psychology, Université du Québec à Montréal, 2010**

**M.S., Psychology, University of New Mexico, 2014**

## ABSTRACT

Researchers often collect longitudinal data so as to model change over time in a phenomenon and for a population of interest. Inevitably, there will be some variation across individuals in specific time intervals between assessments. By necessity or by choice, a researcher can decide to ignore these individual differences in times of assessments. In this simulation study of growth curve modeling, I investigate how ignoring individual differences in time points when modeling change over time relates to convergence and admissibility of solutions, bias in estimates of parameters, power to detect change over time, and, when there is no change over time, Type I error rate. The simulation factors that I manipulate in this study are magnitude of the individual differences in assessment times that are ignored, magnitude of change over time, number of time points, and sample size. Results show that, in contrast to the correct analysis, ignoring individual differences in time points frequently led to inadmissible solutions, especially with few time points and small samples, regardless of the specific magnitude

of individual differences that were ignored. Mean intercept and slope were generally estimated without bias. With few time points and small samples, ignoring individual differences in time points yielded overestimated intercept and slope variances and underestimated intercept-slope covariance and residual variance, more so than when using the correct analysis. When there were more than 3 time points, or when there were 3 time points and sample size was 500, ignoring individual differences in time points yielded overestimated residual variance, but only if individual differences were large. Power and Type I error rate for the linear slope were unaffected by the type of analysis. Overall, this study suggests that it is advisable to account for individual differences in time points whenever possible.

**TABLE OF CONTENTS**

# LIST OF FIGURES

# LIST OF TABLES

**Chapter 1**

**Introduction**

In longitudinal research, a dataset takes one of two forms (Singer & Willett, 2003). It can be *time-structured*, meaning that all individuals are assessed at the same time points, whether the time points are equally spaced or not. Otherwise, the data are *time-unstructured*, meaning that there are differences in times of assessment across individuals. Sometimes, a researcher ignores individual differences in times of assessment in time-unstructured datasets by necessity. Perhaps the specific timing of assessments was not recorded at the time of data collection, either because such recording was not embedded into the research design, or because those in charge of data collection failed to record this information. Other times, a researcher ignores individual differences in time points merely for simplicity of analysis or exposition. It may or may not be possible to measure such individual differences (e.g., by using the dates of data collection), but either way, the researcher chooses to simplify a time-unstructured dataset into a time-structured dataset.

The current study seeks to determine the effects of ignoring individual differences in times of assessment in the context of growth curve models. Specifically, through Monte Carlo simulation, I seek to answer the question: what are the consequences of ignoring individual differences in times of assessment when modeling change over time in terms of convergence, Type I error rate, power, and accuracy of the parameter estimates? In doing so, I aim to fill a gap in the methodological literature, since no study has systematically investigated the consequences of treating time-unstructured datasets as time-structured in growth curve models before. Further, the findings presented in this

paper should help consumers of research critically evaluate how the incorrect analysis of a longitudinal dataset may have affected published results.

The introduction is separated into the following sections. I start by describing two frameworks used for growth curve modeling—multilevel modeling and structural equation modeling—and I discuss how each relates to time-unstructured data. Then, I discuss the issue of the *temporal design*, or the sampling of time points, in longitudinal research. Following that, I mention previous studies that have manipulated the temporal design in some fashion in the context of growth curve modeling. In doing so, I also expose several other factors that have been found to affect the outcome of a growth curve analysis. Finally, I present a few studies that have explicitly addressed the issue of ignoring individual differences in times of assessment, and I describe the method used in the current study.

**Multilevel Modeling**

One statistical technique often used for modeling change over time is multilevel modeling (MLM). In this framework, repeated observations are treated as nested within each individual. The basic growth curve is represented using a set of equations at the observation level (level 1) and the individual level (level 2) which allow for both random slopes and intercepts:

$$\text{Level 1: } y_{ti} = \beta_{0i} + \beta_{1i}T_{ti} + \epsilon_{ti} \tag{1}$$

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + u_{0i} \tag{2}$$

$$\beta_{1i} = \gamma_{10} + u_{1i}$$

These equations can be combined into a reduced-form expression:

$$y_{ti} = \gamma_{00} + \gamma_{10}T_{ti} + u_{0i} + u_{1i}T_{ti} + \epsilon_{ti} \tag{3}$$

In these equations, $y_{ti}$ is the value of the dependent variable for individual *i* at assessment wave *t*, $\gamma_{00}$ is the mean intercept, and $\gamma_{10}$ is the mean linear slope. These two parameters do not vary from person to person, and represent fixed effects. Variable $T_{ti}$ is the time at wave *t*, and the subscript *i* on variable *T* indicates that times of measurement are allowed to vary across persons. The other terms in the model represent error terms, or deviation from mean effects: $u_{0i}$ is the difference between each person's intercept and the mean intercept; $u_{1i}$ is the difference between each person's time slope and the mean time slope; and $\epsilon_{ti}$ is the difference between each person's predicted score at wave *t* and the person's actual score at that time. The two *u* error terms reside at level 2 and vary only between persons, while the $\epsilon$ error term resides at level 1 and varies across time within each person. Together, these error terms represent random effects. The error terms are assumed to be normally distributed with a mean of 0. Further, level-1 residuals are assumed to be independent from level-2 residuals:

$$\epsilon_{ti} \sim N(0, \sigma^2) \; ; \; (u_{0i}, u_{1i}) \sim MVN([0 \quad 0], \begin{bmatrix} \tau_{00} & \tau_{10} \\ \tau_{10} & \tau_{11} \end{bmatrix}) \qquad (4)$$

Change over time is allowed to vary in magnitude for each person, as is the predicted score at time 0. Importantly, nothing in this model constrains the time points to be the same for everyone. In this way, MLM is well suited for the analysis of time-unstructured longitudinal data.

**Structural Equation Modeling**

Another statistical technique used in modeling change over time is latent growth curve modeling in the structural equation modeling (SEM) framework. Meredith and Tisak (1990) have shown that it is possible to model change over time using a special case of confirmatory factor analysis. Using the notation employed by Preacher,

Wichman, MacCallum, and Briggs (2008), we can provide the following equations (with added $i$ subscripts) that define the scores for an individual:

$$y_i = \tau_y + \Lambda_y \eta_i + \varepsilon_i \tag{5}$$

where

$$\eta_i = \alpha + \zeta_i \tag{6}$$

$$\varepsilon_i \sim N(\mathbf{0}, \sigma^2); \ \zeta_i \sim MVN(\mathbf{0}, \mathbf{\Psi}) \tag{7}$$

In this model, with $t$ assessments per person and an intercept and slope factor, $y_i$ is the $(t \times 1)$ vector of scores; $\tau_y$ is a $(t \times 1)$ vector of intercepts that are typically fixed to 0 (and is different from the $\tau$ matrix in Equation 4 above); $\Lambda_y$ is a $(t \times 2)$ matrix of factor loadings; $\eta_i$ is a $(2 \times 1)$ vector of factor scores made of $\alpha$, a $(2 \times 1)$ vector of factor means, and $\zeta_i$, a $(2 \times 1)$ vector of factor-level residuals; and $\varepsilon_i$ is a $(t \times 1)$ vector of occasion-specific residuals.

In this case, and contrary to the multilevel model described in the previous section, the specific times at which the data are collected (represented in the matrix of factor loadings $\Lambda_y$) are not allowed to vary from person to person. In this way, this specific model cannot handle time-unstructured data.

Yet, there are reasons why a researcher might opt for conducting his or her analysis in this framework rather than through multilevel modeling. One reason is missing data: While software implementations of both frameworks handle missing data in the dependent variables similarly, SEM software often allows easy handling of missing data on *predictor* variables as well, whereas this is typically not the case in MLM software. Another reason is measurement error: SEM allows a researcher to model change over time in latent variables using several measures at each time point as

indicators, thereby controlling for measurement error, while MLM is restricted to the analysis of observed variables. A third reason is that SEM allows a researcher to model growth in multiple dependent variables simultaneously much more easily than MLM.

**Analyzing Time-Unstructured Data via SEM**

When the data are time-structured, Curran (2003) showed that the MLM and SEM specifications are fully equivalent. However, when the data are time-unstructured, a researcher does not necessarily have to turn away from SEM in order to analyze the data correctly. In this section, I present three possible solutions to this problem.

One solution is to use a missing-data approach (Curran, 2003), where a variable exists for each observed time point in the dataset. Then, the observed scores are stored in the appropriate variables for each individual. This method yields the exact same solution as the one that would be obtained were the analysis run in MLM software. However, this method is less than adequate. First, it can be difficult and time-consuming to implement: The researcher needs to create as many variables as there are different times of observations. Second, the method does not always work: If a specific time point is observed only once (i.e., at least one variable is left empty for all participants but one), then the model cannot be estimated.

Another, more adequate solution is to use *definition variables* (Mehta & West, 2000), which allow the factor loadings from matrix $\mathbf{\Lambda_y}$ to vary from person to person. In this case, the SEM specification becomes once again equivalent to the MLM specification, even in the presence of time-unstructured data. This method does not suffer from the shortcomings of the missing-data approach, and it is the method that I use later in this paper.

Another recent solution is to use *multilevel structural equation modeling*, which combines features of both MLM and SEM. I do not use this technique here, but the interested reader is referred to Mehta and Neale (2005), and Kaplan, Kim, and Kim (2009).

**The Sampling of Time in Longitudinal Research**

A longitudinal design is intrinsically tied to its temporal design, which refers to the way time points are sampled (Collins & Graham, 2002). In other words, the temporal design refers to two components: 1) the number of measurements contributed by each individual (i.e., the number of time points); and 2) the location of the measurements on the time scale (i.e., the *absolute* location of the time points on the time scale, and their *relative* location to one another; see Collins, 2006). In this way, researchers sample not only individuals, but also time points for each individual. This makes sense, given change over time (or lack thereof) is continuous instead of discrete, and could be observed at any point in time, but in most cases it would be impractical to monitor research participants continuously. Therefore, for each longitudinal study, a researcher has to choose a temporal design. However, despite calls for more explicit justification of the choice of the temporal design and description of how it may have affected the statistical analysis (e.g., Collins, 2006; Collins & Graham, 2002), such discussion has yet to become standard practice in published reports.

The first component of the temporal design, the number of time points, can affect what a researcher can extract from his or her data. For instance, a researcher's ability to detect either linear or nonlinear change over time depends on the number of time points sampled. In particular, barring the cost of additional assessments and the increased risk of

attrition, one has much to gain by sampling more frequently (Adolph & Robinson, 2011; Cole & Maxwell, 2009; Collins & Graham, 2002; but see Vouloumanos, 2011, for a discussion on why nonlinear trajectories might be artifacts). For example, Adolph and colleagues (2008) showed using real data that the number of time points affects conclusions regarding stage transitions, particularly when the dependent variable is dichotomous. Other researchers have developed techniques to make use of frequent time points to separate true long-term change over time from short-term noise. For example, Shiyko and Ram (2011) show how multilevel modeling can be used to model change over time in two different time scales, one shorter (e.g., within-day variation) and one longer (e.g. across-day variation; see also Ram & Gerstorf, 2009). Cole and Maxwell (2009) present a model that distinguishes between a state (a temporary condition) and a trait (a stable condition) using as few as four time points, while McArdle and Woodcock (1997) present a similar model using only two assessments per individual. Similarly, specific longitudinal designs, such as the measurement burst design (Nesselroade, 1991), the intensive longitudinal design (see, e.g., Collins, 2006; Ellis-Davies, Sakkalou, Fowler, Hilbrink & Gattis, 2012; Tan, Shiyko, Li, Li & Dierker, 2012), and the accelerated longitudinal design (see, e.g., Singer & Willett, 2003), have been developed to maximize the number of time points sampled.

The second component of the temporal design, the location of the time points on the time scale, also deserves attention. Just as characteristics of sampled individuals can restrict generalizability of results, location of sampled time points can restrict conclusions to the very specific time points that are sampled. The *absolute* location of time points is of concern when variables that are of interest change over time, for example when

variables follow a daily or seasonal cycle (Selig, Hoy, & Little, in press). In those cases, findings might have been different had the study taken place at a different time in the day or year. Cole and Maxwell (2009) also remind us that a study's findings are influenced by the timing of the study, and mention that even for a simple linear growth model, the average linear trend depends on when the study started (assuming change over time is not perfectly linear for every individual in the study). The other aspect of the location of time points is the *relative* location of the time points to one another. Cole and Maxwell (2009) show that relationships between variables vary as a function of the interval between assessments, which can lead to misleading conclusions when the specific interval is left ignored. For this reason, researchers should choose the location of time points so as to test the hypotheses that are of interest; when they fail to do so, they can face the possibility of hypothesizing an effect that corresponds to a specific time point, but testing the effect at a different time point. Selig, Preacher, and Little (2012) and Tan et al. (2012) have developed techniques to not only account for, but to model the time-varying relationships between variables.

**Time and Other Factors in Growth Curve Modeling**

In this section, I discuss how the temporal design influences the outcome of a growth curve analysis. In doing so, I also mention other factors which affect estimation of growth curve models.

Some simulation studies have investigated how the number of time points, along with other factors such as sample size and effect size, impact the results of a growth curve analysis. Hertzog, Lindenberger, Ghisletta, and van Oertzen (2006) were interested in the power to detect the correlation between two slopes in a multivariate growth curve.

They varied number of time points, along with sample size, effect size (magnitude of the correlation of interest), and growth curve reliability (i.e., how little residual variance was left unaccounted for at each time point). As expected, power to detect the covariance between slopes increased as number of time points, effect size, and sample size increased. However, power was generally low unless the model fit the data particularly well (high growth curve reliability).

In a follow-up study, Hertzog, van Oertzen, Ghiselli, and Lindenberger (2008) were interested in the power to detect the variance of a linear slope in a linear growth curve. They again varied number of time points, sample size, effect size, and growth curve reliability, and found similar results. In particular, growth curve reliability had a high impact on power, such that poor model fit was associated with low power to detect a nonzero slope variance. Power was again found to monotonically increase as number of time points, sample size, and effect size increased. In a related study investigating power to detect a slope variance in a multi-indicator latent growth curve (i.e., a growth curve which controls for measurement error), von Oertzen, Hetzog, Lindenberger and Ghisletta (2010) showed both analytically and through simulations that increasing the number of time points is most effective at increasing power when model misfit is not due to measurement error (see also Willett, 1989). These authors provide formulas to help determine the optimal number of time points and number of indicators per time point when model misfit is a combination of both measurement and non-measurement error.

All of the studies mentioned above, which looked at the power to detect nonzero slope variances or covariances as a function of number of time points, also manipulated the sample size. In each case, smaller sample sizes yielded lower power to detect nonzero

effects. Likewise, small sample sizes have the drawback of making some of the parameter estimates biased. Maas and Hox (2005) showed in a simulation study that small level-2 samples (less than 50) in multilevel modeling lead to negatively biased level-2 standard errors. In the case of growth curves, this corresponds to a situation where few individuals are sampled. Similarly, variance components are underestimated in small samples. This is because the maximum likelihood estimate of the variance of normal distributions is negatively biased, and no adjustment is made in growth curve modeling (Enders, 2010). All in all, small sample sizes can not only yield low power, but also affect the accuracy of certain variance estimates.

With regards to the relative location of time points, some relevant information stems from simulation studies using planned missing data designs. In planned missing data designs, the researcher does not administer all the measures to all the participants. In the case of longitudinal designs, this means that some participants are missing one or more assessments by design. Simulation studies have shown that "not all time points are created equal." Graham, Taylor and Cumcille (2001) found that designs in which the missing data were concentrated in the middle time points, allowing for the use of the full sample at the extreme time points, yielded greater power to detect a binary predictor of the linear slope in a linear growth model than when the extreme time points had missing data. Mistler and Enders (2012) replicated these results looking at power to detect either linear or quadratic change over time for different planned missing data designs. Specifically, increasing the variance of the time points by imposing missing data on the middle time points and leaving the extreme time points intact decreased standard errors, which resulted in higher power. Willett (1989) derived analytically the same conclusions

for complete-cases designs. Specifically, he showed that a researcher can increase growth curve reliability by either increasing the number of time points, increasing the variance of the time points (sampling time points as far away as possible from the mean time point), or both.

**Treating Time-Unstructured Data as Time-Structured**

Some studies have evaluated the consequences of incorrectly analyzing longitudinal data (e.g., by misspecifying the level-1 error structure; Kwok, West, & Green, 2007; Vallejo, Ato, & Valdés, 2008; Wu & West, 2010), but few studies have investigated the impact of treating time-unstructured data as time-structured, which is the focus of the current study. Singer and Willett (2003, Ch. 5) present the results of a study analyzed in one of two ways: 1) using the participants' actual age at the time of assessment; 2) using the participants' *expected* age at each assessment (the authors had planned to assess the participants when they were around 6.5, 8.5, and 10.5 years old, so they used these age values for every participant). In the second analysis, they found that, relative to when the data were analyzed correctly, the linear slope was overestimated, as were the variances of the intercept and linear slope. Mehta and West (2000) presented a similar scenario, where they ignored individual variability in age at the start of the study (they used wave instead of age as a way to track time; see also Hertzog & Nesselroade, 2003). They found that the estimate of the intercept variance was overestimated relative to the known population value, and that the negative covariance between intercept and linear slope was closer to zero.

Both of these studies analyzed only one dataset each; none of the characteristics of the dataset (e.g., sample size) were systematically varied through simulations. Also,

one of the studies (Singer & Willett, 2003) did not make use of population values, making it impossible to determine just how badly the incorrect analysis performed. Moreover, the analysis reported by that study used *incorrect* wave values, given that the *expected* age at each time point did not correspond to the *mean* age. These studies have not looked at a situation where there is no error introduced in the mean time values at each wave, but where individual differences in time values are merely ignored instead. The current study aims to overcome these shortcomings.

**Current Study**

The goal of the current study is to use Monte Carlo simulation to investigate the impact of ignoring individual differences in times of assessment in growth curve models. In other words, this study seeks to establish the consequences of analyzing time-unstructured data as if it were time-structured. I refer to this simplification as a *model misspecification* hereafter.

Specifically, I look at the ability of linear growth curve models to recover population parameter values when either medium or large amounts of individual differences in time points are ignored. To do so, I add a random deviation to each time value for each person, but analyze the data as if these deviations were not present. I define a "medium" deviation as any deviation between -0.2 and 0.2 (up to 20% of one unit on the time scale), and a "large" deviation as any deviation between -0.5 and 0.5 (up to 50% of one unit on the time scale). In the simulation, I also vary sample size and number of time points which, as discussed above, have been shown to affect the accuracy of results and power to detect significant effects. I generate datasets in which there is either no or some linear change over time, and investigate how these factors affect Type I

error rate and power for the linear slope, respectively. I also look at how the manipulated factors affect parameter bias for all parameters, and the likelihood of achieving a converged, admissible solution.

My hypotheses are as follows. Obtaining a converged, admissible solution should be less likely when ignoring individual differences in time points. When solutions do converge, there should be no bias in the mean intercept and linear slope, since the loadings used when ignoring individual differences in time points correspond to the expected mean time points. However, given the model misspecification caused by ignoring individual differences in time points, the standard error of the linear slope should be larger, which in turn should lead to less power to detect linear change over time when it is nonzero, and a lower Type I error rate when it is zero. Similarly, the misspecification should lead the intercept and slope variances to be overestimated; given the compensatory relationship between residual and growth factor variances (see Kwok et al., 2007), this could also lead the residual variance to be underestimated. These problems should be more salient when large amounts of individual differences in time points are ignored compared to medium amounts. Similarly, obtaining a converged solution, achieving adequate power to detect a nonzero linear slope, and estimating the intercept and slope variances without bias should be more often achieved as sample size and number of time points increase.

# Chapter 2

# Method

## Simulation Design

Table 1 shows the factors that were varied during the simulation. There are 2 magnitudes of individual differences in time points that are ignored (up to either 20% or 50% of one time unit per individual for each time point), 4 sample sizes (from 30 to 500), 4 numbers of time points (from 3 to 9), and 2 linear slopes (0 or 2), yielding 64 different conditions (2x4x4x2). Figure 1 shows the population model for the condition with a large (50%) misspecification, 3 time points, and nonzero linear slope.

| Factor | Values |
| --- | --- |
| Extent of misspecification | 20%, 50% |
| Sample size | 30, 50, 200, 500 |
| Number of time points | 3, 5, 7, 9 |
| Linear slope | 0, 2 |

*Table 1*. Values taken by the manipulated factors during the simulation.
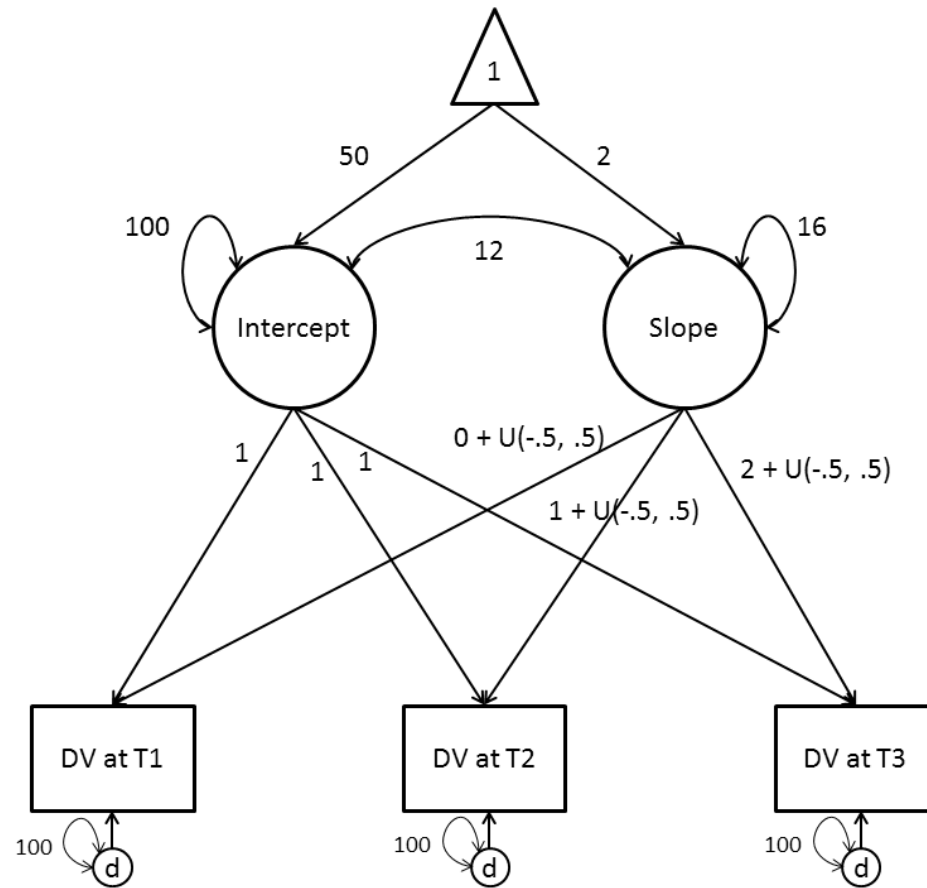
*Figure 1*. Path diagram for population model in simulation condition with large individual differences in time points, nonzero linear slope, and three time points.

Population values for the mean intercept and nonzero linear slope, intercept and slope variances, intercept and slope covariance, and residual variance were chosen to be representative of published research. Values at time 0 are $T$ scores (see also Hertzog et al., 2006, 2008; von Oertzen et al., 2010). $T$ scores are a linear transformation of $z$ scores: $T = 10z + 50$. Therefore, the population mean intercept is 50, and intercept variance is 100. The mean nonzero slope is 2, corresponding to a change of one fifth of a standard deviation from one assessment to the next. The slope variance was chosen so as to be smaller than the intercept variance and was set to 16, meaning that the intercept variance is 6.25 times larger than the slope variance. Given the mean slope and slope variance, we

can compute a standardized effect size ($\delta$) for the linear slope (Raudenbush & Liu, 2001) that is interpreted like a Cohen's *d*:

$$\delta = \frac{\alpha_2}{\sqrt{\Psi_{22}}} = \frac{2}{\sqrt{16}} = 0.5$$

which corresponds to a medium effect size. The covariance between the intercept and slope was set to 12, which corresponds to an intercept-slope correlation of +0.30. Finally, the residual variance was set to 100, implying a low growth curve reliability of 50% at time 0.

Ten thousand datasets were generated in each condition, and analyzed either while ignoring individual differences in times of assessments (slope loadings in Figure 1), or while using the correct times of assessments for each individual.

**Outcomes**

In each condition of the simulation, I look at the following outcomes.

*Convergence rate and proportion of inadmissible solutions*. The convergence rate is the proportion of datasets that yield a converging solution. I also report the proportion of datasets that yield inadmissible solutions in the form of negative variances and out-of-bounds correlations (see Analysis of Datasets).

*Parameter bias*. Parameter bias is computed as

$$B = \frac{\bar{\bar{\theta}} - \theta}{\theta}$$

where $\bar{\bar{\theta}}$ is the average parameter estimate across replications, and $\theta$ is the population value for that parameter. Parameter bias was computed for all estimated parameters, namely, mean intercept and linear slope, intercept and slope variances, intercept and slope covariance, and residual variance.

*Power*. Power to detect linear change over time (for conditions where the linear slope is nonzero) was computed as the proportion of replications in each condition where the Wald $z$ test of the slope factor is significant at an alpha level of .05.

*Type I error rate*. Type I error rate for the slope factor (for conditions where the linear slope is zero) was computed as the proportion of replications in each condition where the Wald $z$ test of the slope factor is significant at an alpha level of .05.

**Data Generation**

Ten thousand datasets were generated for each condition of the simulation design (Table 1). Datasets were sampled from these populations using R v3.0.2 (R Core Team, 2013), according to Equation 5. For example, the population model shown in Figure 1 uses the following matrices for each person:

$$\tau_y = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Lambda_{yi} = \begin{bmatrix} 1 & 0 + U(-0.5, 0.5) \\ 1 & 1 + U(-0.5, 0.5) \\ 1 & 2 + U(-0.5, 0.5) \end{bmatrix}$$

$$\alpha = \begin{bmatrix} 50 \\ 2 \end{bmatrix}$$

$$\varepsilon_i = \begin{bmatrix} N(0, 100) \\ N(0, 100) \\ N(0, 100) \end{bmatrix}$$

$$\zeta_i \sim MVN(\mathbf{0}, \mathbf{\Psi}), \text{ where } \mathbf{\Psi} = \begin{bmatrix} 100 & 12 \\ 12 & 16 \end{bmatrix}$$

The terms $U(-0.5, 0.5)$ represent a value sampled from a theoretical uniform distribution with a minimum of -0.5 and a maximum of 0.5. The terms $N(0, 100)$ represent a value sampled from a theoretical normal distribution of mean 0 and variance

100. The dimensions of the $\boldsymbol{y}_i$, $\boldsymbol{\tau}_y$, $\boldsymbol{\Lambda}_{yi}$, and $\boldsymbol{\varepsilon}_i$ matrices (see Equation 5) vary according to the number of time points per individual. To allow for each person to have their own time values (i.e., for $\boldsymbol{\Lambda}_y$ in Equation 5 to vary across persons), data were generated for one person at a time (hence the added $i$ subscript in $\boldsymbol{\Lambda}_{yi}$).

**Analysis of Datasets**

Each dataset was analyzed in two ways: 1) by constraining the factor loadings to be equal across individuals (e.g., set the slope loadings to 0, 1, and 2 in the 3-wave condition, which are the expected mean loadings at each wave); and 2) by allowing each person to have his or her own factor loadings. Analyses were done using Mplus v.7.0 (Muthén & Muthén, 2012). To achieve the second analysis, the individually-varying times of observations method described in Mehta and West (2000) as implemented in Mplus was used. In all cases, the residual variances at each time point were constrained to equality.

Mplus does not distinguish between admissible and inadmissible solutions, so long as the estimation converges to a solution. The R package MplusAutomation (Hallquist & Wiley, 2013) was used to loop through each analysis, and flag datasets that yielded inadmissible solutions in the form of negative variances or out-of-bounds correlation between the intercept and slope factors for each simulation condition. Moreover, Mplus does not automatically provide the correlation between intercept and slope when using individually-varying times of observations. Therefore, for each analysis, the correlation was computed as (see Singer & Willett, 2003, p. 100):

$$\hat{\rho} = \frac{\widehat{\Psi}_{12}}{\sqrt{\widehat{\Psi}_{11}\widehat{\Psi}_{22}}}$$

Only the datasets that yielded admissible solutions were used in pooling the results within each condition of the simulation.

Listings of the computer code used to generate and analyze the data are presented in the Appendix.

**Chapter 3**

**Results**

**Convergence Rate and Proportion of Admissible Solutions**

Figure 2 shows the proportion of converged solutions (row 1), proportion of admissible solutions (row 2), and proportion of usable solutions (solutions that are both converged and admissible; row 3), as a function of number of time points (x axis), extent of misspecification (points), type of analysis (lines), and sample sizes (panels), when the mean slope is nonzero. The first row of Figure 2 shows that the time-structured analysis virtually always converged to a solution, for all time points, misspecifications, and sample sizes. The time-unstructured analysis almost always converged to a solution, with a few exceptions when there were only 3 time points and sample size was 30 or 50, where the convergence rate was 96% in both cases.
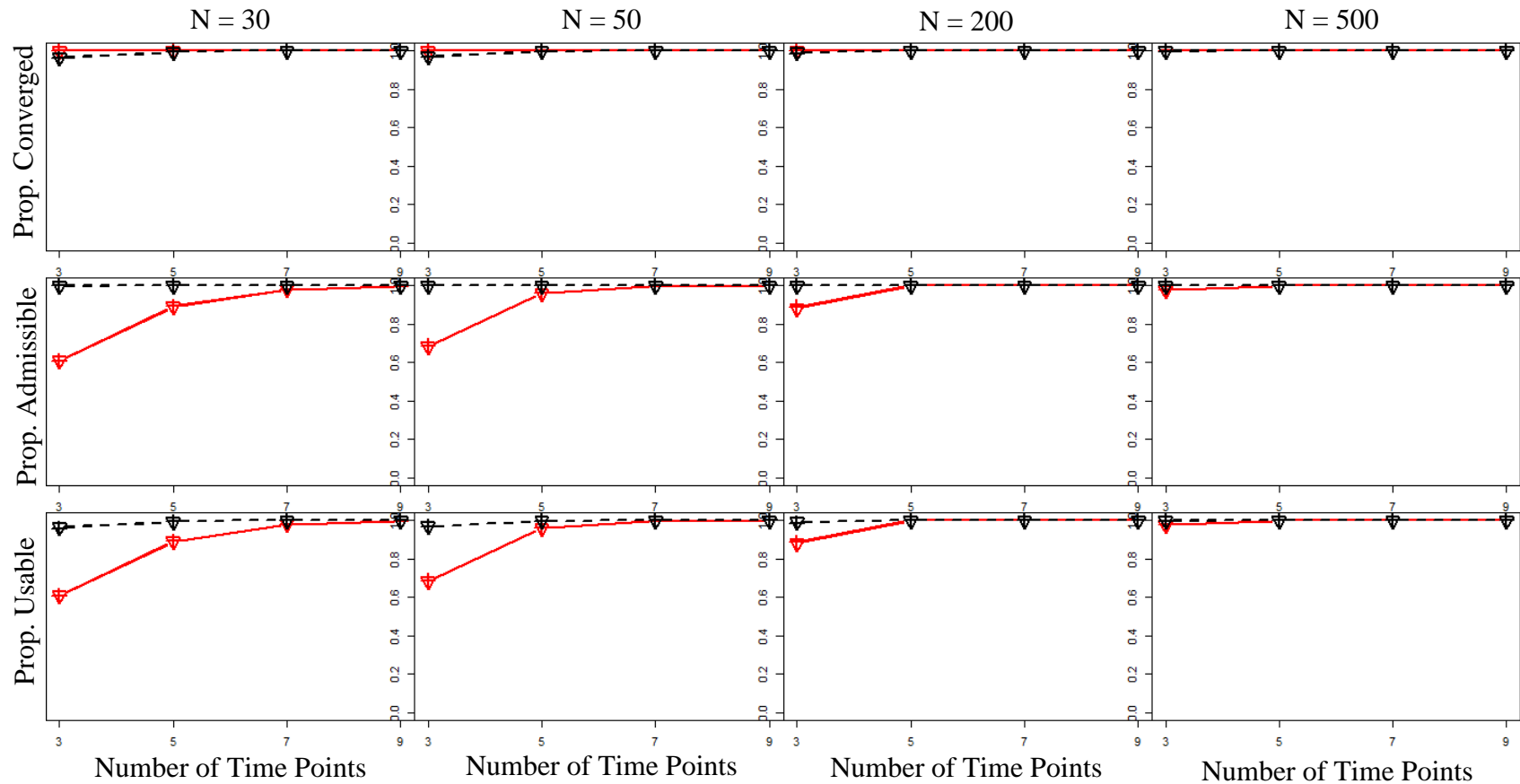
*Figure 2.* Proportion of analyses that yielded a converged, admissible, and usable solution as a function of number of time points (x axis), type of analysis (solid line = time structured, dotted line = time unstructured), magnitude of individual differences in time points (+ = 20%, ∇ = 50%), and sample size (columns), when slope is nonzero.

The second row of Figure 2 shows that the time-unstructured analysis virtually always converged to an admissible solution. Conversely, the time-structured analysis sometimes yielded inadmissible solutions, and this was a function of number of time points and sample size: The time-structured analysis was more likely to yield an inadmissible solution for 3 or 5 time points, and for samples of 30 or 50 observations. The likelihood of obtaining an inadmissible solution in a time-structured analysis was unrelated to whether misspecification was medium or large (the two solid lines overlap in Figure 2).

The third row of Figure 2 shows the combination of these results, i.e. a solution is usable if it both converged and was admissible. Overall, if a researcher were to collect one sample from this population, he or she could proceed with the time-unstructured analysis almost 100% of the time, regardless of number of time points, sample size, and magnitude of individual differences in times of assessment. On the other hand, the researcher could proceed with the time-structured analysis only some of the time, depending on number of time points and sample size. Comparing rows 1 and 2 from Figure 2 shows that this difficulty is due to the time-structured analysis yielding inadmissible solutions in the form of negative variances and out-of-bounds correlations.

**Parameter Bias**

*Growth factor means*. Figure 3 is analogous to Figure 2, and shows parameter bias (in percent) for the mean intercept (row 1) and mean slope (row 2) as a function of number of time points, extent of misspecification, type of analysis, and sample size. Figure 3, row 1 shows that the mean intercept is generally estimated without bias, regardless of any of the manipulated factors. Row 2, on the other hand, shows a more ambiguous pattern. Mean slope tends to be slightly underestimated for 3 time points and

samples of 30 or 50 observations. Bias in mean slope does not seem to vary systematically as a function of either type of analysis or magnitude of individual differences in times of assessment.

*Figure 3*. Mean intercept and mean slope bias (in %) as a function of number of time points (x axis), type of analysis (solid line = time structured, dotted line = time unstructured), magnitude of individual differences in time points (+ = 20%, ∇ = 50%), and sample size (columns), when slope is nonzero.

*Growth factor variances and covariance*. Figure 4 shows parameter bias (in percent) for intercept variance (row 1), slope variance (row 2), and intercept-slope covariance (row 3). Intercept variance is overestimated when there are only 3 time points and 30 or 50 observations, but much more so when using a time-structured analysis than when using a time-unstructured analysis, regardless of the magnitude of individual differences in times of assessment that are ignored. With 5 time points and 30 observations, the time-structured analysis yields an unbiased estimate of intercept variance, whereas the time-unstructured analysis *under*estimates the intercept variance. With 30 observations and 7 or 9 time points, or with 50 observations and 5, 7, or 9 time points, intercept variance is slightly underestimated regardless of type of analysis or magnitude of individual differences in time points. For larger sample sizes, intercept variance is estimated without bias regardless of number of time points, type of analysis, or magnitude of individual differences in time points, except with 3 time points and 200 observations, where the time-structured analysis slightly overestimates intercept variance.

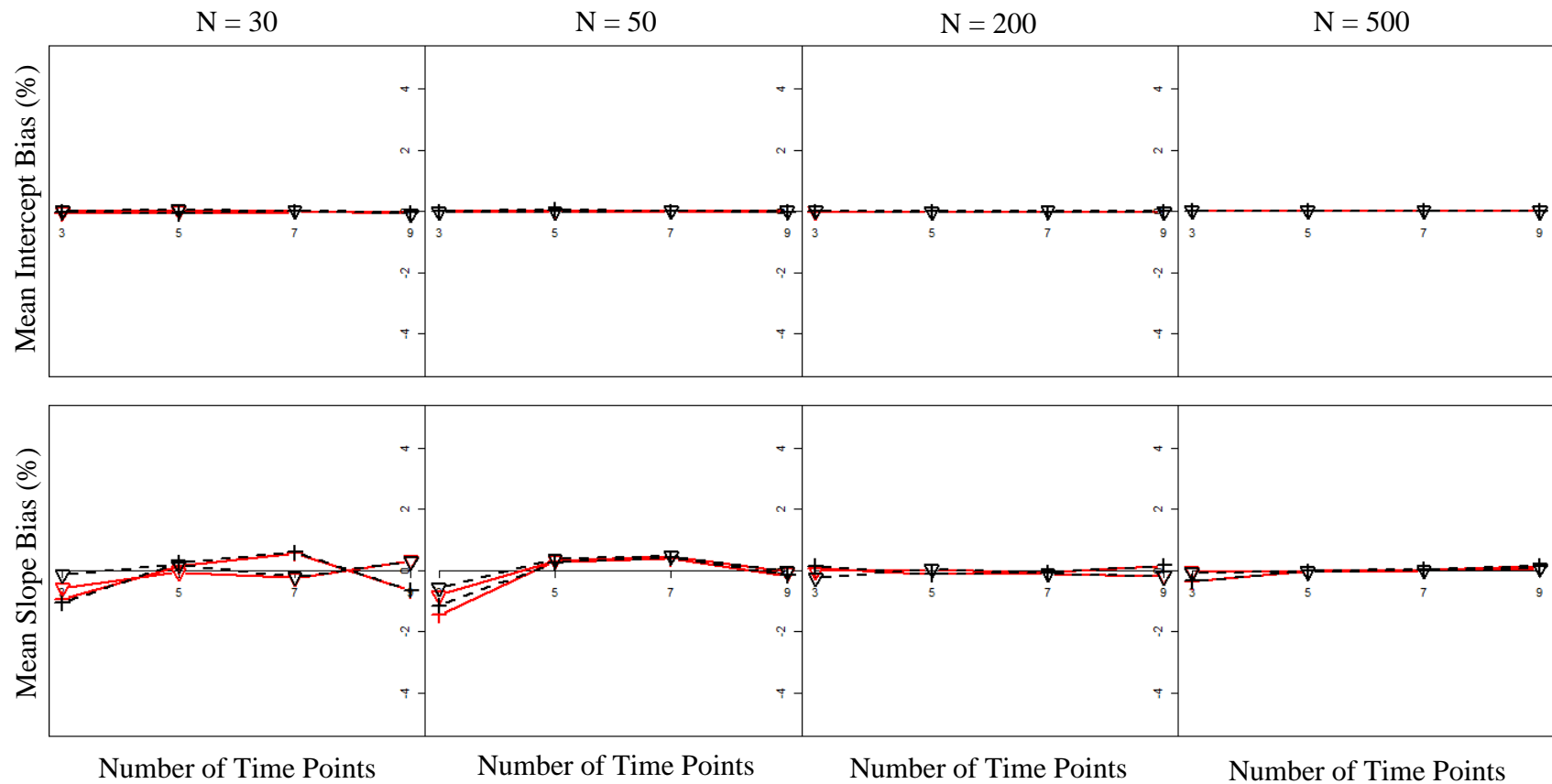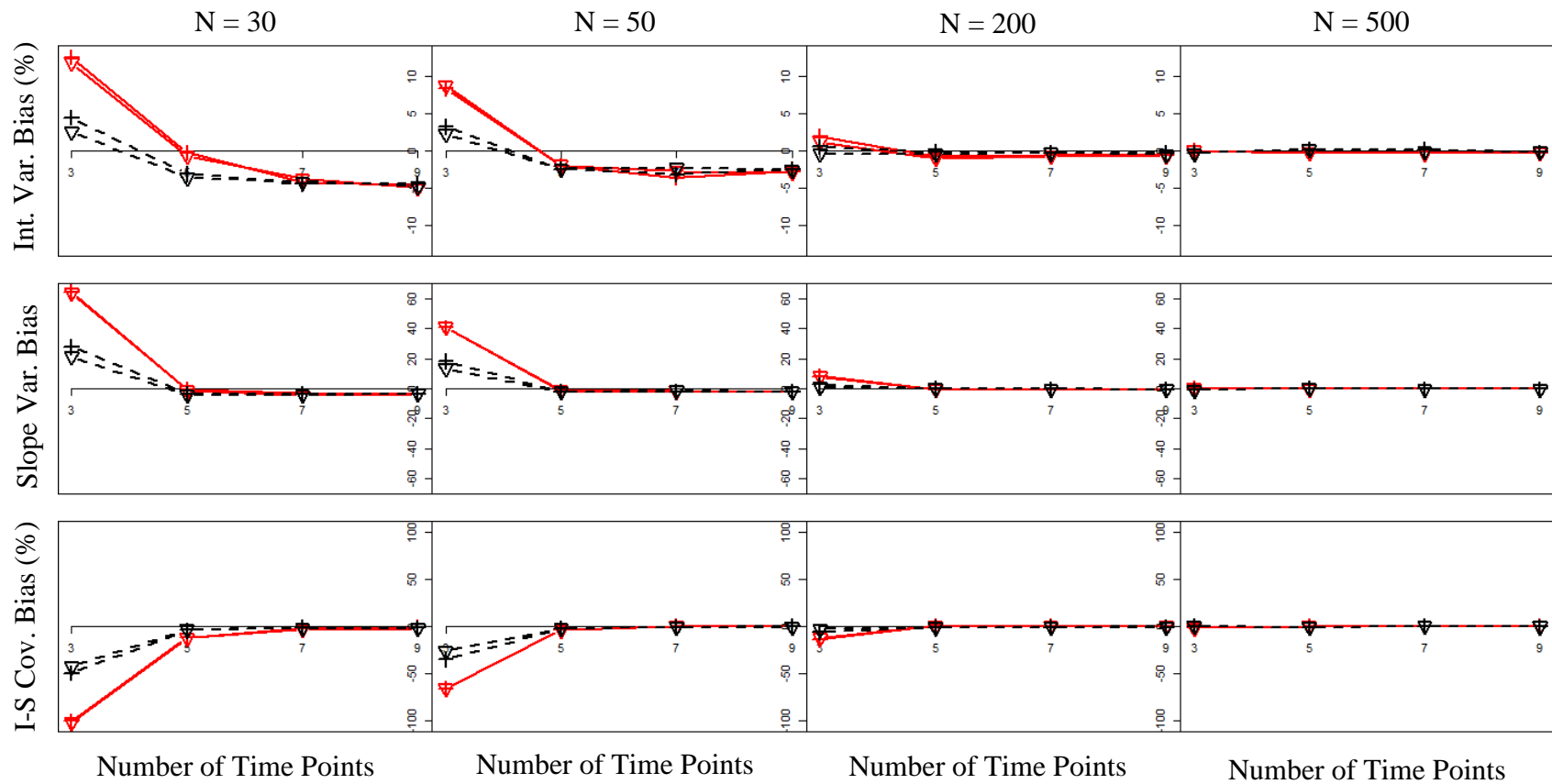*Figure 4*. Intercept and slope variances and covariance bias (in %) as a function of number of time points (x axis), type of analysis (solid line = time structured, dotted line = time unstructured), magnitude of individual differences in time points (+ = 20%, ∇ = 50%), and sample size (columns), when slope is nonzero.

Figure 4, row 2 shows a very similar pattern for bias of slope variance, with one important difference: Slope variance is more extremely overestimated, sometimes by over 60% (see y axis). As was the case for intercept variance, slope variance is overestimated with 3 time points and 30 or 50 observations, and much more so when using a time-structured analysis rather than a time-unstructured analysis. As before, it did not matter whether the magnitude of individual differences in time points was medium or large. Finally, as was the case for intercept variance, slope variance is overestimated with 3 time points and 200 observations, but only with a time-structured analysis, and not with a time-unstructured analysis.

Figure 4, row 3 shows parameter bias for the intercept-slope covariance. In terms of bias, the intercept-slope covariance is the parameter that suffers the most from a small number of time points and small sample sizes (see y axis). Whereas both intercept and slope variance were overestimated with few time points and small sample sizes, intercept-slope covariance is *under*estimated with 3 points and 30 or 50 observations, particularly so when using a time-structured analysis instead of a time-unstructured analysis. Intercept-slope covariance is also underestimated with 5 time points and 30 observations, and with 3 time points and 200 observations, but only with a time-structured analysis, and not with a time-unstructured analysis. As was the case for both intercept and slope variance, there is no bias with a sample size of 500, regardless of number of time points, type of analysis, or magnitude of individual differences in times of assessment.

*Residual variance*. Figure 5 shows parameter bias for residual variance as a function of the manipulated factors. Residual variance is underestimated with 3 time

points and 30 or 50 observations, particularly so when using a time-structured analysis instead of a time-unstructured analysis. With 5, 7, or 9 time points, bias for residual variance is close to 0%, except in one situation: When large individual differences in times of assessments are ignored, a time-structured analysis tends to slightly overestimate residual variance, regardless of sample size (even with 500 observations).
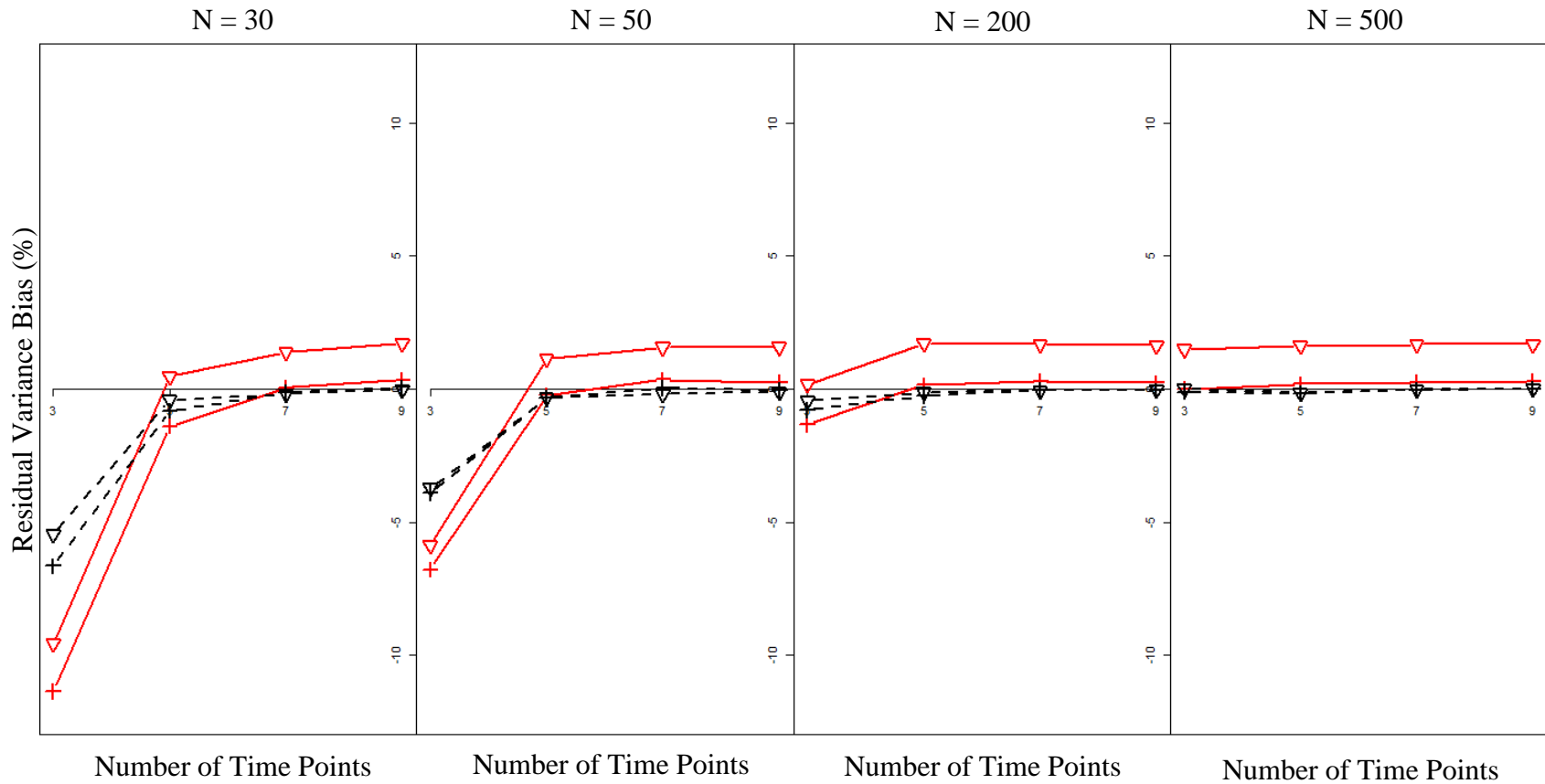
*Figure 5*. Residual variance bias (in %) as a function of number of time points (x axis), type of analysis (solid line = time structured, dotted line = time unstructured), magnitude of individual differences in time points (+ = 20%, ∇ = 50%), and sample size (columns), when slope is nonzero.

**Power and Type I Error Rate**

Figure 6 shows power (row 1) and Type I error rate (row 2) for the mean linear slope as a function of the manipulated factors. Power to detect the mean linear slope increases as a function of number of time points and sample size, but is unaffected by type of analysis or magnitude of individual differences in times of assessment. Power never reaches .80 with a sample size of 30, whereas power reaches .80 with a sample size of 50 when there are 5 or more time points. Power is well above .80 when there are 200 or 500 observations, regardless of number of time points.
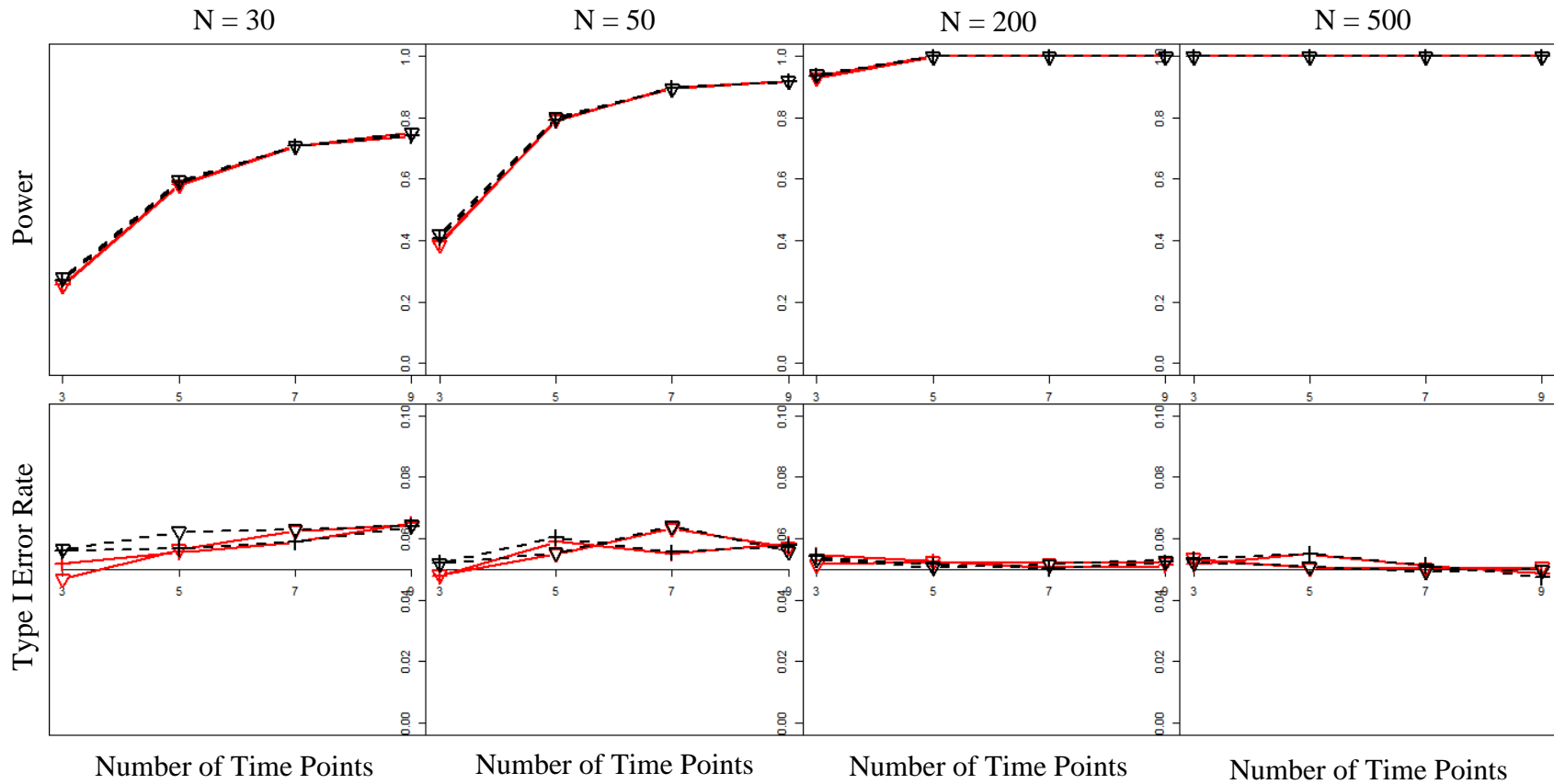
*Figure 6*. Power and Type I error rate as a function of number of time points (x axis), type of analysis (solid line = time structured, dotted line = time unstructured), magnitude of individual differences in time points (+ = 20%, ∇ = 50%), and sample size (columns), when slope is nonzero (first row) or zero (second row).

Figure 6, row 2 shows that Type I error rate (when mean slope is 0) is slightly above the nominal value of .05 when sample size is small (30 or 50), especially when there are more than 3 time points per subject. On the other hand, Type I error rate is around the nominal value of .05 when sample size is larger (200 or 500), regardless of number of time points, sample size, type of analysis, or magnitude of individual differences in times of assessment.

**Chapter 4**

**Discussion**

The goal of this study was to investigate the effects of ignoring individual differences in times of assessment when modeling change over time. Specifically, I looked at how treating a time-unstructured dataset as though it were time-structured in growth curve modeling affects parameter estimates, convergence and admissibility of solution, and power and Type I error rate for the linear slope.

I hypothesized that obtaining a converged and admissible solution, with no negative variances and out-of-bound correlations, would be more difficult to achieve when ignoring individual differences in times of assessment (time-structured analysis) than when accounting for these differences (time-unstructured analysis). The hypothesis was partially supported. Solutions almost always converged when using a time-unstructured analysis, and solutions always converged when using a time-structured analysis. However, the time-structured analysis frequently yielded inadmissible solutions; this was not the case for the time-unstructured analysis, which virtually always converged on an admissible solution. As expected, the time-structured analysis was more likely to yield an inadmissible solution when there were few time points and when sample size was small. Contrary to expectations, the proportion of inadmissible solutions in the time-structured analysis was not affected by the magnitude of the individual differences in time points that were ignored. Overall, removing the information relative to the times of assessment of each subject led to difficulties in estimation, even when the loss of information was not extremely large.

Regarding parameter bias, I hypothesized that mean intercept and slope would be estimated without bias in the time-structured analysis. Mean intercept was estimated without bias regardless of type of analysis, magnitude of individual differences in time points, sample size, or number of time points. Mean slope was generally estimated without bias, though it appeared to be slightly underestimated with few time points and small sample sizes, although any bias was unrelated to type of analysis and magnitude of individual differences in time points. Singer and Willett (2003) found that the mean slope in their study, with a sample size of 89, was larger when using a time-structured analysis instead of a time-unstructured analysis. This might be due to the fact that the time values used in their time-structured analysis did not correspond to the mean time value at each time point, whereas they did in the current study. Conversely, Mehta and West (2000) used the correct mean time values at each time point, as was done here, and similarly found that their estimate of the mean linear slope was unbiased.

I further hypothesized that both intercept and slope variance would be overestimated in a time-structured analysis. As expected, both variances were overestimated when there were few time points and sample size was small, more so when using the time-structured analysis. Therefore, it appears that random error due to the misspecification in the time-structured analysis is confounded with intercept and slope variance. However, it again did not matter whether the magnitude of the individual differences that were being ignored was medium or large; compared to the size of the individual differences in time points in the data, the type of analysis used was more predictive of bias in the estimate of the intercept and slope variances. One unexpected result was the slight negative bias in the intercept and slope variances in small samples

when there were more than 3 time points per subject, which occurred in both the time-structured and time-unstructured analyses. Neither Singer and Willett (2003) nor Mehta and West (2000) report such a negative bias in intercept and slope variance. However, the sample sizes in those studies were 89 and 250 (respectively), well above the samples sizes of 30 and 50 at which intercept and slope variances were underestimated in this study. Some studies mentioned previously (Hertzog et al., 2008; von Oertzen et al., 2010) focused specifically on slope variance, but looked at power only, not bias. Future research could focus on bias in slope—and intercept—variance, and how such bias varies with combinations of sample size and number of time points in the context of growth curve modeling.

In this study, I found that the intercept-slope covariance was also estimated with bias in small samples with few time points, though the bias was negative. This contrasts with the positive bias observed for the intercept and slope variances. Both the negative bias in the intercept-slope covariance and the positive bias in the intercept and slope variances would lead the analyst to believe that the intercept-slope correlation is closer to zero than it should be, on average. As was the case for the intercept and slope variances, the time-structured analysis worsened the bias of the intercept-slope covariance relative to the time-unstructured analysis, while the specific magnitude of individual differences in time points that were ignored did not affect bias. Mehta and West (2000) also found that the intercept-slope covariance in their study was closer to zero when using a time-structured analysis compared to a time-unstructured analysis. However, they had a sample size of 250 (and 4 time points). With this sample size, the intercept-slope covariance showed no bias in this study (see Figure 4, row 3). Future research should

determine whether this is due to differences in one or several population values that were not manipulated in this study (such as growth curve reliability or intercept or slope variance).

In terms of residual variance, I hypothesized that there would be a compensatory relationship between the intercept and slope variances, on the one hand, and the residual variance, on the other hand, replicating previous research (see Kwok et al., 2007). As expected, where intercept and slope variances were overestimated for small samples with few time points, residual variance was underestimated (the compensatory relationship seemed to be limited to these specific conditions in this study, however). In those cases, residual variance was more extremely underestimated when using a time-structured analysis. The underestimated residual variance coupled with the overestimated intercept and slope variances can give the impression of *better* fit of the growth model to the data than would be warranted. This is especially true when using a time-structured analysis rather than a time-unstructured analysis, because the intercept and slope variances are more overestimated, and residual variance is more underestimated, at least when sample size is small and time points are few. Indeed, when sample size is small and time points are few, a researcher who would compute the ratio of reliable (intercept and slope) variance to residual variance would obtain, on average, a larger ratio with a time-structured analysis than with a time-unstructured analysis. Whether this illusion of a better fit is reflected in fit indexes remains to be investigated in future research. Preliminary results indicate that fit indexes might be able to correctly reflect the model misspecification: Singer and Willett (2003) report the deviance, AIC, and BIC fit statistics for both their time-unstructured and time-structured analysis, and all three fit

statistics indicate better model fit when using the time-unstructured analysis. (Mehta and West [2000] found no bias in their estimate of residual variance, but they do not provide fit statistics that can be compared across analyses.)

With more than 3 time points per subject, residual variance was more often than not estimated without bias, except when a time-structured analysis was used and there were large individual differences in times of assessment, in which case residual variance was overestimated. Surprisingly, this is the only time in this study where magnitude of individual differences in time points actually mattered.

Contrary to what was hypothesized, power and Type I error rate were both unaffected by type of analysis and, yet again, by magnitude of individual differences in time points. As expected, power to detect a nonzero linear slope increased as number of time points and sample size increased, as was previously found for power to detect linear slope variance (Hertzog et al., 2008) and correlation between two linear slopes (Hertzog et al., 2006). The Type I error rate for the linear slope was around .05 for larger samples (200 and 500), but was found to be unexpectedly higher than .05 for smaller samples (30 and 50). This is likely due to the use of the Wald $z$ test to establish significance of the linear slope, which is the test provided by Mplus. This test assumes that the sampling distribution of the mean linear slope is normal, but the sampling distribution is less likely to be normal for small samples. An alternative to the Wald $z$ test is the likelihood ratio test, which does not make any assumption regarding the form of the sampling distribution (Enders, 2010; see also Raudenbush & Bryk, 2002). Hertzog et al. (2008) compared the Type I error rate for the slope variance when using the Wald $z$ test and two different likelihood ratio tests, and found that the two likelihood ratio tests maintained the Type I

error rate at around .05, but that the Wald $z$ test was more liberal, and yielded an average Type I error rate of .08 for a sample size of 100. Another option would be for the analyst to compare the Wald $z$ statistic to a $t$ distribution rather than a normal distribution. Most software packages such as SPSS, SAS, and the package `nlme` in R, provide a $p$ value for the linear slope based on a $t$ distribution instead of a normal distribution, but Mplus does not. Future studies should compare the Type I error rates for the mean linear slope when basing conclusions on a Wald $z$ test, a $t$ test, or a likelihood ratio test.

Regardless of the type of analysis, the use of small samples and few time points was associated with bias in most parameter estimates and low power to detect the nonzero linear slope. Many researchers are cognizant of the pitfalls of small samples, but perhaps fewer are aware of the difficulties associated with few time points. One recommendation for researchers designing a longitudinal study would be to move beyond the typical pre-post design with small samples, and to aim to increase not only the sample size, but also the number of assessments per participant.

**Limitations and Future Research**

This study was limited by a number of factors. As must be the case in simulation studies, some population values were fixed and not varied during the simulation. These population values were chosen to be representative of values found in published research. For example, the slope variance was set to be smaller than the intercept variance, and growth curve reliability was low (50%) at the first time point. However, it is possible that the results presented here are not generalizable to other population values (or combination of population values). More simulation studies using different population values are needed.

Another factor might limit the generalization of the results, perhaps more so than the choice of particular population values. In this study, deviations around the real time points were assumed to be independent from one time point to the next. However, this is unlikely to always be the case in longitudinal research. As the study progresses, deviations around the true time values tend to get larger, a phenomenon dubbed "occasion creep" (Singer & Willett, 2003). Future simulation studies could compare the consequences of ignoring individual differences in time points when errors are nonadditive with a more realistic situation where errors are additive.

In this study, I used full-information maximum likelihood to obtain parameter estimates. Both previous studies that have investigated the effects of ignoring individual differences in times of assessment in growth curve modeling (Mehta & West, 2000; Singer & Willett, 2003) also used full-information maximum likelihood. However, another estimation method, restricted maximum likelihood estimation, is often used in practice, and is indeed the default estimation method in many software packages, like SAS and `nlme` and `lme4` in R. The two methods are asymptotically equivalent, but can sometimes yield different results, particularly in small samples (Hox, 2010). It remains to be seen how the results presented here change when a different estimation method is used.

It is noteworthy that the magnitude of individual differences in time points that were being ignored had very little impact on the outcomes of interest in this study, except for residual variance bias. This lack of effect is potentially due to the low growth curve reliability (i.e. the relatively high proportion of residual variance) used in this study, a factor that has been manipulated in other simulation studies on growth curve modeling

(Hertzog et al., 2006, 2008; von Oertzen et al., 2010). Since there was already much residual variance relative to reliable (intercept and slope) variance, the addition of more error created by the miscoding of time did not add much misfit to the data, regardless of whether the miscoding was medium or large. I suspect that when the model fits the data better, the deleterious effects of the miscoding of time would be increased, because the added misfit to the data would be larger relative to the already existing misfit.

In the same vein, the effects of the miscoding of time might have been more salient in the presence of missing data. In this study, all samples were free of missing data, though previous simulation studies on growth curve modeling have successfully manipulated missing data (e.g., Mistler & Enders, 2012; Rhemtulla et al., 2013). Modern missing-data handling techniques yield unbiased estimates only when variables related to the missingness are included in the model (see Enders, 2010). One variable that is related to the missingness in longitudinal studies is often time itself: More participants are lost to the study as the study goes on. In growth curve modeling, the time variable is indeed included in the model; however, if the analyst uses a time-structured analysis instead of a time-unstructured analysis, then the time variable does not take the correct values, particularly if individual differences in assessment times are large. Therefore, I suspect that missing data would worsen the effects of ignoring individual differences in time points.

In conclusion, the current study is, to the best of my knowledge, the first study to systematically investigate the consequences of ignoring individual differences in times of assessment on solutions obtained when conducting growth curve modeling. Further, I provided several ways to improve on the study in future research. All in all, the current

study shows that it is advisable to use the correct time values where possible, even if the analyst believes that the deviations around time points are not too large. Singer and Willett (2003, p. 146) phrase the advice thus: "The bottom line: never 'force' an unstructured dataset to be structured."

# APPENDIX

## Code Used to Generate and Analyze the Data

```
#########################################################
################## DATA GENERATION ##################
library(MASS)

#number of timepoints
Ts <- c(3,5,7,9)

#factor means
alpha1 <- 50
alpha2 <- c(2, 0)

#factor covariance matrix (psi)
psi<-matrix(c(100, 12, 12, 16), 2, 2)

#level-1 residual variance
eps <- 100


d <- list() #store generated datasets
N <- c(30, 50, 200, 500) #sample sizes
misspecification <- c(0.2, 0.5)
nrep <- 1000 #number of datasets per condition

for (t in Ts)
{
        .colnames <- c(paste("y", 1:t, sep=""), paste("t", 1:t, sep=""))

        #factor loadings (lambda)
        lambda<-matrix(NA, t, 2)
        lambda[,1] <- 1

        for (s in alpha2)
        {
                for (n in N)
                {

                        for (mis in misspecification)
                        {
                                print(c(mis,n,s,t)) #start of this condition

                                for (r in 1:nrep)
                                {
                                        #level-2 residuals: zeta (sample from bivariate
normal)
                                        zeta <- mvrnorm(n, mu=c(0,0), Sigma=psi) #equation 7

                                        #level-1 residuals vector (epsilon)
                                        epsilon <- matrix(rnorm(t*n,0,sqrt(eps)), t, n,
byrow=F) #equation 7

                                        #factor loadings slope (time points)
                                        timepts <- matrix(runif(t*n,-1*mis,mis), t, n,
byrow=F)

                                        .d<-matrix(NA,0,t*2) #(empty) vector of scores for
one person

                                        #each person in sample for repetition r
                                        for (i in 1:n)
                                        {
                                                y <- vector()
                                                #factor loadings
                                                lambda[,2] <- 0:(t-1) + timepts[,i]

                                                #eta
```

```
                                                        eta <- c(alpha1, s) + zeta[i,] #equation 6

                                                        #compute vector of scores for observation
                                                        y <- lambda%*%eta + epsilon[,i] #equation 5
                                                        .d <- rbind(.d, c(as.vector(y), lambda[,2])
)
                                            }

                                            .d <- as.data.frame(.d) #n rows, Ts columns
                                            colnames(.d) <- .colnames

                                            ######export data for Mplus
                                            .file <-
paste("c:/.../datasets/t",t,"s",s,"n",n,"mis",mis,"_",r,".dat", sep="")
                                            prepareMplusData(.d, .file, overwrite=T)
                                            ######

                                            d[[length(d)+1]] <- .d #stores rth dataset in list
of datasets
                                }
                        }
                }
        }
}

#########################################################


#########################################################
############# MPLUS: CREATE INPUT FILES, ##############
#############    RUN, AND LOAD OUTPUTS    #############

library(MplusAutomation)

#create Mplus input files from template
createModels("c:/.../MplusTemplate3.txt")

#run models through Mplus
runModels(directory="c:/.../inputs", recursive=T)

#get Mplus outputs into R
outputs <- readModels("c:/.../inputs", recursive=T)

#save outputs
save(outputs, file="c:/.../outputs.RData")

#load outputs
outputs <- get(load(file="c:/.../outputs.RData"))

#######################################################


############################################################################
############# CREATE TABLE WITH ONE LINE PER OUTPUT    #############
############# WITH PARAM. ESTIMATES AND OTHER OUTCOMES #############


#####nonconverged
nonconverged <- vector(length=length(outputs)); nonconverged[] <- FALSE
for (i in 1:length(outputs))
{
        if (is.null(outputs[[i]]$summaries$LL)) nonconverged[i] <- TRUE
}


#####inadmissible (AFTER nonconverged)
inadmissible <- vector(length=length(outputs)); inadmissible[] <- FALSE
for (i in 1:(length(outputs)))
{
```

```
        .vars <- subset(outputs[[i]]$parameters$unstandardized,
subset=paramHeader%in%c("Variances", "Residual.Variances"), select="est")
        .cor <- subset(outputs[[i]]$parameters$unstandardized,
subset=paramHeader=="S.WITH", select="est")/(.vars[1,]*.vars[2,])^0.5
        if (nonconverged[i] == FALSE & (.vars[1,]<0 | .vars[2,]<0 | .vars[3,]<0 | .cor[1,]
> 1 | .cor[1,] < -1)   ) inadmissible[i] <- TRUE
}

inadmissible[which(nonconverged)] <- NA

#####




###########create table with: output #; t; n; slope; mis; converged; admissible; usable

#get condition
cond.ana = cond.t = cond.n = cond.mis = cond.s <- vector()
for (i in 1:length(outputs))
{
        .title <- outputs[[i]]$input$title #e.g.: "TUNt9s0n500mis0.5 "

        #ana
        if (grepl("TSt", .title)) cond.ana[i] <- "TS" else if (grepl("TUNt", .title))
cond.ana[i] <- "TUN"

        #t
        if (grepl("t3", .title)) cond.t[i] <- 3 else if (grepl("t5", .title)) cond.t[i] <-
5 else if (grepl("t7", .title)) cond.t[i] <- 7 else if (grepl("t9", .title)) cond.t[i] <-
9

        #n
        if (grepl("n30mis", .title)) cond.n[i] <- 30 else if (grepl("n50mis", .title))
cond.n[i] <- 50 else if (grepl("n200mis", .title)) cond.n[i] <- 200 else if
(grepl("n500mis", .title)) cond.n[i] <- 500

        #mis
        if (grepl("mis0.2", .title)) cond.mis[i] <- 0.2 else if (grepl("mis0.5", .title))
cond.mis[i] <- 0.5

        #s
        if (grepl("s0n", .title)) cond.s[i] <- 0 else if (grepl("s2n", .title)) cond.s[i]
<- 2
}

cond <- paste(cond.ana, "t", cond.t, "n", cond.n, "mis", cond.mis, "s", cond.s, sep="")

#create table
outputs.info <- data.frame(number=1:length(outputs), ana=cond.ana, t=cond.t, n=cond.n,
mis=cond.mis, s=cond.s, cond=cond, converged=!nonconverged, admissible=!inadmissible)
outputs.info$usable <- outputs.info$converged & outputs.info$admissible

#get estimated parameter values for each converged & admissible solution
SwithI=meanI=meanS=varI=varS=resid=pSlope <- rep(NA, length(outputs))
for (i in which(outputs.info$usable==TRUE))
{
        .ests <- subset(outputs[[i]]$parameters$unstandardized,
subset=paramHeader%in%c("Means","S.WITH","Variances","Residual.Variances"))
        .ests <- .ests[order(.ests$param, .ests$paramHeader),]
        meanI[i] <- .ests[1,"est"]
        SwithI[i] <- .ests[2,"est"]
        varI[i] <- .ests[3,"est"]
        meanS[i] <- .ests[4,"est"]
        varS[i] <- .ests[5,"est"]
        resid[i] <- .ests[6,"est"]
        pSlope[i] <- .ests[4,"pval"]

}

#update outputs table with parameter estimates from each output
```

```
outputs.info <- data.frame(outputs.info, meanI, meanS, varI, varS, SwithI, resid, pSlope)


#save
save(outputs.info, file="c:/.../outputs.info.RData")


#load
o <- get(load(file="c:/.../outputs.info.RData")) #500
nrow(o) #1,280,000
###########

########################################################################



########################################################################
############# CREATE TABLE WITH ONE LINE PER CONDITION #############
#############  WITH PARAMETER BIAS AND OTHER OUTCOMES   #############


###compute parameter bias in each condition
o.bias <- with(o, aggregate(cbind(meanI, meanS, varI, varS, SwithI, resid), by=list(ana,
mis, t, n, s), FUN=function(x){mean(x,na.rm=T)})      )
names(o.bias)[1:5] <- c("ana", "mis", "t", "n", "s")

o.bias$meanI <- (o.bias$meanI-50)/50*100
o.bias$meanS[o.bias$s==2] <- (o.bias$meanS[o.bias$s==2] - 2)/2*100
o.bias$meanS[o.bias$s==0] <- NA
o.bias$varI <- (o.bias$varI -100)/100*100
o.bias$varS <- (o.bias$varS -16)/16*100
o.bias$SwithI <- (o.bias$SwithI -12)/12*100
o.bias$resid <- (o.bias$resid -100)/100*100
###end bias


####compute other outcomes: prop. converged, admissibile, usable; power, Type I error
o.other <- with(o, aggregate(cbind(converged, admissible, usable), by=list(ana, mis, t,
n, s), FUN=function(x){      return(sum(x, na.rm=T) / sum(complete.cases(x)))    })
        )
names(o.other)[1:5] <- c("ana", "mis", "t", "n", "s")

#function to compute power
get.power <- function(p, alpha=.05)
{
        p <- p[complete.cases(p)]
        return(sum(p<alpha)/length(p))
}

#compute power or Type I error in each condition
o.power <- with(o, aggregate(pSlope, by=list(ana, mis, t, n, s), FUN=get.power)   )
names(o.power)[1:6] <- c("ana", "mis", "t", "n", "s", "power")

#get a dataset with bias and power/Type I error for each condition combined
o.all <- merge(x=o.bias, y=o.other, all.x=T, all.y=T, by=c("ana", "mis", "t", "n", "s"))
o.all <- merge(x=o.all, y=o.power, all.x=T, all.y=T, by=c("ana", "mis", "t", "n", "s"))
####end other outcomes

########################################################################



######################################################
####################### PLOTS #######################

##outcome: row
##sample size: column (4)
##mis(2) and ana(2): lines (4)
##t: x axis (4)
##restrict to slope=2
##one graph: x=t, y=bias, lines=ana*mis; slope=2, sample=one at a time
```

```
.params1 <- c("meanI", "meanS") #growth factor means
.params2 <- c("varI", "varS", "SwithI") #growth factor variances & covariance
.params3 <- c("resid") #residual variance
.plots4 <- c("converged", "admissible", "usable")
.plots5 <- c("power", "typeI")
.listparams <- list(.params1, .params2, .params3, .plots4, .plots5) #5 different figures
.ylim <- c(-100,100) #y axis (temporary)
.margins <- c(1,0,1,0)+0.0 #margins


for (.params in .listparams)
{

        windows() #open new plot window
        par(mfrow=c(length(.params),4), mar=.margins) #one row per outcome, 4 panels
(columns) per row

        for (param in .params) #one row at a time
        {
                .slope <- ifelse(param=="typeI", 0, 2) #if looking at Type I, use
conditions where slope=0; otherwise, use conditions where slope=2
                if (param=="typeI") param <- "power"
                .d <- subset(o.all, s==.slope, select=c("t", param))
                .ylim <- c(    floor( min(.d[,2:(length(param)+1)], na.rm=TRUE)    ),
ceiling(     max(.d[,2:(length(param)+1)], na.rm=TRUE)    )        ) #find lowest and
highest value for this particular outcome (param)
                .ylim <- c(max(abs(.ylim),5)*-1, max(abs(.ylim),5)) #set y-axis limits to
± the largest value in abs. value
                if (param %in% c("converged", "admissible", "usable", "power")) .ylim <-
c(0,1) #set y-axis for proportions
                if (param=="power" & .slope==0) .ylim <- c(0,0.1) #set y-axis for Type I
error plot

                for (ssize in c(30,50,200,500))
                {
                        #margins
                        if (ssize == 500)     par(mar=.margins+c(0,0,0,0.05)) #increase
right margin in rightmost panel
                        else par(mar=.margins)

                        plot(1, type="n", xlab=NA, ylab=NA, ylim=.ylim, xlim=c(3,9),
xaxt="n")

                        #position of x axis
                        if (param %in% c("meanI", "meanS", "varI", "varS", "SwithI",
"resid")) .xaxis.pos <- 0 else if (param=="power" & .slope==0) .xaxis.pos <- 0.05 else if
(param=="power" & .slope==2) .xaxis.pos <- NA else .xaxis.pos <- NA
                        axis(1, at=c(3,5,7,9), labels=T, tick=T, pos=.xaxis.pos) #add x-
axis

                        for (a in c("TS", "TUN"))
                        {
                                for (m in c(0.2, 0.5))
                                {
                                        .d <- subset(o.all, s==.slope & n==ssize & ana==a &
mis==m, select=c("t", param)) #get the 4 values for particular line to plot
                                        print(paste(ssize, a, m, sep=","))
                                        print(.d)
                                        lines(x=.d[,1], y=.d[,2], type="b", lty=nchar(a)-1,
pch=m*10+1, col=switch(a, TS="red", TUN="black"), lwd=2, cex=2) #plot points and line,
with different symbol and color depending on misspecification and type of analysis

                                }
                        }
                }
        }
}

#####################################################
```

## Template used to generate Mplus input files (MplusTemplate3.txt)

```
[[init]]
iterators = ana t slope n mis r;
ana = 1:2;
anaName#ana = TS TUN;
t = 3 5 7 9;
slope = 2 0;
n = 30 50 200 500;
mis = 0.2 0.5;
r = 1:1000;
filename = "[[r]].inp";
outputDirectory = C:/.../inputs/[[anaName#ana]]/t[[t]]/s[[slope]]/n[[n]]/mis[[mis]];
[[/init]]

TITLE: [[anaName#ana]]t[[t]]s[[slope]]n[[n]]mis[[mis]]

DATA: FILE = "c:/.../datasets/t[[t]]s[[slope]]n[[n]]mis[[mis]]_[[r]].dat";
VARIABLE:
NAMES =
[[t==3]]
y1-y3 t1-t3;
[[/t==3]]
[[t==5]]
y1-y5 t1-t5;
[[/t==5]]
[[t==7]]
y1-y7 t1-t7;
[[/t==7]]
[[t==9]]
y1-y9 t1-t9;
[[/t==9]]
[[ana == 1]]
USEVARIABLES =
[[t==3]]
y1-y3;
[[/t==3]]
[[t==5]]
y1-y5;
[[/t==5]]
[[t==7]]
y1-y7;
[[/t==7]]
[[t==9]]
y1-y9;
[[/t==9]]
[[/ana == 1]]
[[ana == 2]]
TSCORES =
[[t==3]]
t1-t3;
[[/t==3]]
[[t==5]]
t1-t5;
[[/t==5]]
[[t==7]]
t1-t7;
[[/t==7]]
[[t==9]]
t1-t9;
[[/t==9]]
[[/ana == 2]]

ANALYSIS:
[[ana==2]]
TYPE = RANDOM;
[[/ana==2]]
ESTIMATOR = ML;

MODEL:
```

```
[[ana == 1]]
[[t==3]]
i s | y1@0 y2@1 y3@2;
[[/t==3]]
[[t==5]]
i s | y1@0 y2@1 y3@2 y4@3 y5@4;
[[/t==5]]
[[t==7]]
i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6;
[[/t==7]]
[[t==9]]
i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8;
[[/t==9]]
[[/ana == 1]]
[[ana == 2]]
[[t==3]]
i s | y1-y3 AT t1-t3;
[[/t==3]]
[[t==5]]
i s | y1-y5 AT t1-t5;
[[/t==5]]
[[t==7]]
i s | y1-y7 AT t1-t7;
[[/t==7]]
[[t==9]]
i s | y1-y9 AT t1-t9;
[[/t==9]]
[[/ana == 2]]

[[t==3]]
y1 y2 y3 (resid);
[[/t==3]]
[[t==5]]
y1 y2 y3 y4 y5 (resid);
[[/t==5]]
[[t==7]]
y1 y2 y3 y4 y5 y6 y7 (resid);
[[/t==7]]
[[t==9]]
y1 y2 y3 y4 y5 y6 y7 y8 y9 (resid);
[[/t==9]]

OUTPUT:
```

**EXAMPLE OF MPLUS INPUT FILE FOR THE TIME-STRUCTURED ANALYSIS AND A DATASET WITH 5 TIME POINTS, SLOPE OF 2, SAMPLE SIZE OF 50, AND MISSPECIFICATION OF 0.5**

```
TITLE: TSt5s2n50mis0.5

DATA: FILE = "c:/.../t5s2n50mis0.5_1.dat";
VARIABLE:
NAMES =
y1-y5 t1-t5;
USEVARIABLES =
y1-y5;

ANALYSIS:
ESTIMATOR = ML;

MODEL:
i s | y1@0 y2@1 y3@2 y4@3 y5@4;
```

```
y1 y2 y3 y4 y5 (resid);

OUTPUT:
```

**EXAMPLE OF MPLUS INPUT FILE FOR THE TIME-UNSTRUCTURED ANALYSIS AND A DATASET WITH 5 TIME POINTS, SLOPE OF 2, SAMPLE SIZE OF 50, AND MISSPECIFICATION OF 0.5**

```
TITLE: TUNt5s2n50mis0.5

DATA: FILE = "c:/.../t5s2n50mis0.5_1.dat";
VARIABLE:
NAMES =
y1-y5 t1-t5;
TSCORES =
t1-t5;

ANALYSIS:
TYPE = RANDOM;
ESTIMATOR = ML;

MODEL:
i s | y1-y5 AT t1-t5;

y1 y2 y3 y4 y5 (resid);

OUTPUT:
```

# References

Adolph, K. E., & Robinson, S. R. (2011). Sampling development. *Journal of Cognition and Development*, *12*, 411-423.

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*, 527-543.

Cole, D. A., & Maxwell, S. E. (2009). Statistical methods for risk-outcome research: Being sensitive to longitudinal structure. *Annual Review of Clinical Psychology*, *5*, 71-96.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505-528.

Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence*, *68*, 85-96.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529-569.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121-136.

Ellis-Davies, K., Sakkalou, E., Fowler, N. C., Hilbrink, E. E., & Gattis, M. (2012). CUE: The continuous unified electronic diary method. *Behavior Research Methods*, *44*, 1063-1078.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335−353). Washington, DC: APA.

Hallquist, M., & Wiley, J. (2013). MplusAutomation: Automating Mplus model estimation and interpretation (Version 0.6-1). Retrieved from http://CRAN.R-project.org/package=MplusAutomation

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*, 244-252.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, *18*, 639-657.

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling*, *15*, 541-563.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). New York: Routledge.

Kaplan, D., Kim, J. S., & Kim, S. Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 592-613). Thousand Oaks, CA: SAGE.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford.

Kwok, O. M., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research*, *42*, 557-592.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 86-92.

McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods*, *2*, 403-435.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*, 259-284.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*, 23-43.

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

Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. P. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of Developmental Research Methods* (pp. 742-754). New York: Guilford.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual changes. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 92–105). Washington, DC: APA.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: SAGE.

R Core Team (2013). R: A language and environment for statistical computing (Version 3.0.2). Retrieved from http://www.r-project.org

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

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.

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, *24*, 778.

Rhemtulla, M., Jia, F., Wu, W., & Little, T. D. (2013). *Planned missing designs to optimize the efficiency of latent growth parameter estimates*. Manuscript submitted for publication.

Selig, J. P., Hoy, R., & Little, T. D. (in press). Temporal design in organizational research. In E. Paavilainen-Mäntymäki & M. Hassett (Eds.), *Handbook of Longitudinal Research Methods in Studies of Organizations*. Northampton, MA: Elgar Publishing.

Selig, J. P., Preacher, K. J., & Little, T. D. (2012). Modeling Time-Dependent Association in Longitudinal Data: A Lag as Moderator Approach. *Multivariate Behavioral Research*, *47*, 697-716.

Shiyko, M. P., & Ram, N. (2011). Conceptualizing and estimating process speed in studies employing ecological momentary assessment designs: A multilevel variance decomposition approach. *Multivariate Behavioral Research*, *46*, 875-899.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Tan, X., Shiyko, M. P., Li, R., Li, Y., & Dierker, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychological Methods*, *17*, 61-77.

Vallejo, G., Ato, M., & Valdés, T. (2008). Consequences of misspecifying the error covariance structure in linear mixed models for longitudinal data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*, 10-21.

von Oertzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology*, *63*, 627-646.

Vouloumanos, A. (2011). Are U-Shaped Developmental Trajectories Illusory? *Journal of Cognition and Development*, *12*, 154-158.

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, *49*, 587-602.

Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research*, *45*, 420-452.