MULTILEVEL MODELS FOR INTENSIVE LONGITUDINAL DATA WITH HETEROGENEOUS ERROR STRUCTURE: COVARIANCE TRANSFORMATION AND VARIANCE FUNCTION MODELS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

SEUNGMIN JAHNG

Dr. Phillip K. Wood, Dissertation Supervisor

DECEMBER 2008

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

MULTILEVEL MODELS FOR INTENSIVE LONGITUDINAL DATA WITH HETEROGENEOUS ERROR STRUCTURE: COVARIANCE TRANSFORMATION AND VARIANCE FUNCTION MODELS

presented by Seungmin Jahng,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Phillip K. Wood

Professor Timothy J. Trull

Professor Jianguo Sun

Professor Douglas L. Steinley

Professor Hans-Friedrich Koehn

To my three girls, Miyoung, Jennifer, and Sarah

ACKNOWLEDGEMENTS

I would like to first thank Professor Phillip Wood for his advising me over the whole period of my study in University of Missouri. Since the first year of my graduate training in quantitative psychology program, he has supported me both academically and psychologically in every aspect of my life. He has continually guided, encouraged and stimulated me to study broad topics in analysis of longitudinal data. I also would like to thank Professor Tim Trull for providing me an opportunity to work for an ecological momentary assessment study with him. I have learned a lot about EMA studies from his lab and the project, which encouraged me to extensively study statistical models for intensive longitudinal data in this dissertation. Professor Douglas Steinley also provided me an opportunity to work with him in the area of multivariate statistical models, especially in cluster analysis of longitudinal data, which improved my knowledge of statistical models in many ways. Professor Hans-Friedrich Koehn and Professor Jianguo Sun gave me many precious comments and grateful supports to enhance my study and dissertation writing. Maria Solhan and Whitney Brown, my lab mates, also helped me to understand many important issues in the analysis of EMA data and ILD. I could have not completed my dissertation successfully without any of them and really appreciate their help.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

MULTILEVEL MODELS FOR INTENSIVE LONGITUDINAL DATA WITH HETEROGENEOUS ERROR STRUCTURE: COVARIANCE TRANSFORMATION AND VARIANCE FUNCTION MODELS

Seungmin Jahng

Dr. Phillip K. Wood, Dissertation Supervisor

ABSTRACT

Recent developments in data collection methods in the behavioral and social sciences, such as Ecological Momentary Assessment (EMA) enables researchers to gather intensive longitudinal data (ILD) and to examine more detailed features of intraindividual variation of a variable(s) over time. Due to its high intensity of assessments within individuals, ILD often has different characteristics from traditional longitudinal data with a few measurement occasions and requires different assumptions of statistical models in use. In the present thesis, issues in the analysis of ILD and problems of current use of statistical models for the analysis of ILD are discussed and investigated. Specifically, the issue of heterogeneity of autocorrelation and variance across individuals in ILD is extensively studied for multilevel models (MLMs). In chapter 2, a brief introduction to multilevel models and issues in modeling residual covariance structure in MLMs are provided and discussed. In chapter 3, it is shown that bias in estimation of parameters in MLMs under homogeneity assumption is not ignorable when autocorrelation differs across individuals and its average is high. It is also shown that a transformation method, which multiplies variables in the model by the inverse of Cholesky factor of individual-specific error covariance, attenuates the bias for ILD. Chapter 4 reviews variance function models for heterogeneous variance and introduces a two-step MLM approach for modeling heterogeneous variance using squared residuals. A simulation study showed that the two-step MLM does not suffer from non-convergence and is applicable to ILD.

1. Introduction

Psychological research has focused on between-individual variation (or interindividual variation) on one or more dimensions to understand characteristics or relational patterns of psychological phenomena of interest. The findings from between-individual variation have been often generalized to those of within-individual change. Because this generalization requires very strict conditions, called *ergodicity* (e.g., stationarity and non-cyclicity, see Molenaar, 2004), however, longitudinal studies have had their own right in psychology to describe characteristics of within-individual change.

Traditional longitudinal studies often involve a small to moderate number of repeated observations (usually less than 10 occasions) across many individuals. Accordingly, the number of individuals is typically much greater than the number of observations within each individual. In such cases, the prediction of a response variable as a function of the within-individual covariates (including time for growth model), the between-individual covariates, and the interactions among the covariates are often of interest. However, with a small to moderate number of observations for each individual, more detailed investigation of the dynamic process of a response within individuals is restricted. If interest is more focused on intraindividual changes in response (e.g., mood fluctuations across time), many repeated observations are required for each individual, that is an intensive longitudinal study is called for.

Recent developments in data collection methods in the behavioral and social sciences, such as Ecological Momentary Assessment (EMA) (Hufford, Shiffman, Paty, & Stone, 2001; Stone & Shiffman, 1994), enables researchers in this area to gather intensive longitudinal data (ILD) and to examine more detailed features of intraindividual variation over time. For statistical analysis, multilevel models (MLMs) are useful and widely used tools for analysis of traditional longitudinal data and ILD in particular (Schwartz & Stone, 2007; Walls & Schafer, 2006). Due to its high intensity of assessments within individuals, however, ILD often has different characteristics or requires different assumptions from traditional longitudinal data (e.g., heterogeneity of error structure) and applications of traditional MLMs used in longitudinal data with a few measurement occasions to ILD may produce inaccurate statistical inferences. Additionally, when heterogeneity of intraindividual processes other than the effects of a covariate on the mean function are of interest (e.g., autocorrelation, variability, or instability), traditionally used MLMs as often implemented do not regularly examine the interindividual heterogeneity in those intraindividual variation patterns and such models must be extended or modified to accommodate ILD studies. In the present thesis, issues in the analysis of intensive longitudinal data and problems of current use of MLMs for the analysis of ILD are discussed and investigated. In addition, new developments in MLMs to model heterogeneous residual process are suggested and evaluated. A brief introduction of general characteristics of longitudinal data introduces readers to central conceptual issues in ILD analysis.

Cattell's Data Box and the Analysis of Longitudinal Data

Design and analysis of a scientific research require consideration of the structure of probable patterns of relation among phenomena of interest, which in turn calls for understanding the structure of the data under consideration. Cattell (1946, 1988) provided one such framework for the organization and analysis of multivariate data in psychological research (applicable to other fields of science as well). In its original form, Cattell (1946) proposed the Covariation Chart (or the data box) with three coordinates (for persons, for tests, and for occasions), but later (Cattell, 1988) modified the original data box to consist of four dimensions: dimensions for organisms, states, stimuli/situations, and responses. The person dimension of the original box corresponds to the organism dimension, the occasion corresponds to the state (and situation in part), and the test dimension is split into two dimensions, stimulus/situation and response.

This approach differentiates external or situational conditions from responses and states (or time). Clearly, more than four dimensions can be conceived and empirically studied (e.g., addition of an "observer" dimension or separation of *focal* stimuli from *background* situation). In its final form, Cattell (1988) proposed a ten dimensional Basic Data Relation Matrix (BDRM) which is, he thought, sufficient for defining a behavioral event. The first five dimensions are person, stimulus, background, response, and observer. Cattell classified these five dimensions as time invariant *proto-types*, such as *trait*. The other five dimensions are time varying *variants* or *states* corresponding to the first five dimensions, i.e., state of the person, variant of the stimulus, phase of the background situation, style of the response, and condition of the observer.

Although Cattell's ten dimensional BDRM is useful as a comprehensive description and conceptual framework for possible data relationships, it is, as he noted, neither feasible nor necessary to obtain data or test hypotheses across all ten dimensions in a given study. First, not all dimensions are of interest. In such cases, in any given study, some of the ten dimensions are measured (and analyzed) but others are not and, as such, these missing dimensions serve to highlight limitations to the generality of study conclusions (Cattell, 1988, pp. 95-100). Measured dimensions are assumed to be controlled for or fixed at a point on the unmeasured dimensions, or allowed to vary across all the points (or grids) on the unmeasured dimensions (i.e., marginalized or integrated out across the unmeasured dimensions in mathematical term). For example, if we relate scores of a set of individuals over a set of stimuli without temporal information, it may be assumed that each score is made on one occasion or averaged out across occasions. Moreover, some of the dimensions can be described in terms of other dimensions. For instance, different observers can be thought of as a situational factor. As a result, a four dimensional data box with person, state/time, response, and stimulus/situation can be considered a successful reduction of the ten dimensions in that the four dimensions form a core relational structure of the phenomena of interest. Accordingly, Ozer (1986) argued for the four dimensions (persons, situations, responses, and time) as a simplification of Cattell's ten dimensional BDRM. Analytically, responses are variables of interest, persons are the units of analysis, stimuli/situations are covariates, and times are occasions of observations within units (Biesanz, West, & Kwok, 2003). Figure 1.1 illustrates a four dimensional data box with two situations, three responses, four persons, and four time points.

The degree of dimensionality and presence of (or variation in) each dimension in the data box determines the range of possible relational patterns to be investigated as well as the nature of appropriate statistical models. The response dimension implies multivariate observations and is required when we are interested in a profile or

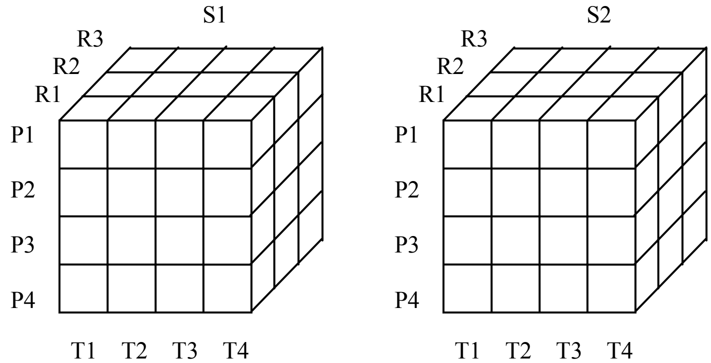


Figure 1.1. An illustration of a four dimensional data box with two situations (S), three response measures (R), four persons (P), and four time points (T).

multivariate characteristics of variables. Variation in the person dimension allows an investigation of (between) individual differences. With multiple assessments across time dimension, within individual change can be described and inferred. Variation in situations or stimuli is required to investigate the effect of covariate(s) on a response or responses. Researchers may combine two or more dimensions, resulting in a data matrix or a (hyper-) data box, to examine more complicated relational patterns.

In terms of statistical models, linear models (e.g., regression or ANOVA), for example, are often used for the analyses of data matrix with the person and the situation dimensions while time series models (e.g., ARIMA model or spectral analysis) are appropriate to investigate the time dimension with or without other dimensions. Data matrix with the response dimension and the person dimension is typically analyzed by multivariate statistical models (e.g., factor analysis or structural equation models). If three or four dimensions are involved in a study, MLMs (e.g., linear mixed model or multilevel structural equation models) are useful tools for the analysis of the data box.

In general, psychological research is designed to investigate systematic differences in response across individuals (e.g., group difference or individual difference) or changes of a response(s) within individuals (e.g., intervention effect or growth trajectory). These two different components of variability have been termed interindividual variation (IEV) and intraindividual variation (IAV), respectively (Molenaar, 2004; Molenaar, Huizenga, & Nesselroade, 2003). Traditional psychological studies have examined either IEV or IAV separately, due to the characteristics of data available for a particular study (i.e., lack of one or more of dimensions in the data box) and the limitations in the application of statistical models intended for IEV to modeling IAV, or vice versa. Interindividual variation is typically analyzed using linear models (e.g., regression, ANOVA, or structural equation models) that usually assume independence of observations among individuals while intraindividual variation is commonly analyzed using time-series regression models that allow serial correlations of observations within individuals. However, application of longitudinal studies (e.g., EMA study) and appropriate statistical models (e.g., MLMs) enables researchers to investigate both IEV and IAV simultaneously using one statistical model.

Quantitatively, longitudinal studies are those which, in the data box, assess at least the time and person dimensions and often include the situation dimension due to interest in the effect of a covariate(s). Further, if the response dimension is added, the result is a multivariate longitudinal study. In longitudinal studies, many useful research questions can be posed concerning relational patterns among (and within) the dimensions. For example, questions concerning intraindividual variation (across the time dimension) of a response and interindividual differences (across the person dimension) in such intraindividual variation may be of interest: We may be interested in how a response changes across time for individuals in different groups (growth model), or how time varying covariates are differentially related to change in mean response over time across individuals (multilevel regression model).

Intensive Longitudinal Data and Heterogeneous Covariance Structure

The questions described concern the effect of covariates as well as heterogeneity of their effects across individuals and are appropriately analyzed by multilevel models or latent variable models in which heterogeneous effects across individuals are treated as random effects or latent variables. In traditional longitudinal analysis using MLMs, inference concerning the effect of the covariate (for both fixed and random) is made under the assumption that the (within-individual) residual distribution is identical across individuals (nested within or controlled for other covariates), i.e., the residual covariance structure (variance and/or autocovariance components) is assumed to be homogeneous across individuals.

This assumption is made not because researchers believe within-individual covariance structure to actually be homogenous in nature, but rather because individual-specific parameters of the covariance are, if present, nuisance parameters (i.e., not the parameters of interest). Alternatively, it is often not efficient to estimate individual-level covariance parameters when only a small number of observations are present for each individual. In intensive longitudinal studies where many repeated assessments are available for each individual, however, the heterogeneous within-individual covariance structure may be both estimable and of theoretical interest as well. When this characteristic of ILD is ignored, it may cause bias in statistical inference on the parameters of interest (i.e., the effect of covariates) under MLM's as traditionally specified for longitudinal data and therefore modification which adjust for the biasing effects of such processes are necessary.

It is well-known that autocorrelation in residuals adversely affects the efficiency of the OLS estimation for covariates effects in the context of regression on a single time-series (Cochrane & Orcutt, 1949; Watson, 1955). It is also known that misspecified covariance structure in MLMs produce inaccurate statistical inference in both fixed effects and variance components (Ferron, Dailey, & Yi, 2002; Jacqmin-Gadda, Sibillot, Proust, Molina, & Thiébaut, 2007; Kwok, West, & Green, 2007; Lange & Laird, 1989). However, little is known about how violation of the homogeneity assumption of within-individual covariance structure affects statistical inference on parameters in multilevel models for intensive longitudinal data (or even, for that matter, for longitudinal data in general). Heterogeneity of within-individual variance and autocorrelations may need to be taken into account in order to estimate the parameters of interest more efficiently. We may consider a model that allows heterogeneity of individual-level covariance structure or a new model that corrects for the adverse effects of heterogeneous autocorrelation.

Modeling Heterogeneous Variances in ILD

Individual differences in variance and/or serial correlation are of substantive interest in their own right in some ILD studies. Affective variability (as evidenced by within-individual variance) and instability (as a function of the variance and the autocorrelation), for example, are defining characteristics of psychological disorders such as Borderline Personality Disorder and the individual difference or heterogeneity of those parameters can be investigated using EMA study (Cowdry, Gardner, O'Leary, Leibenluft, & Rubinow, 1991; Ebner-Priemer et al., 2007; Stein, 1996; Trull et al., 2008; Woyshville, Lackamp, Eisengart, & Gilliland, 1999).

At least two sources are responsible for heterogeneity of variance across individuals: Difference in within-individual variance may be caused by (1) difference in the effect of a time varying covariate on a response (i.e., random effect) and (2) difference in individual level characteristics (e.g., individual differences in impulsivity that cause different variability in mood fluctuations). Of these two sources of heterogeneity of variance, (1) can be modeled in a MLM by adding individual-level random effects, as a part of modeling mean responses, while (2) requires modeling variance as a function of predictors, i.e., variance function models. Variance function models are rather unfamiliar and strange models to many researchers in psychology. Although the history of variance function modeling is rather long, a recent development by Hedeker, Mermelstein, and Demirtas (2008) enables researchers to model random variance. In this model, variance of a response, not the response itself, is modeled in multilevel equations and random mean responses are also modeled as well. Simultaneous estimation of mean function and variance function increases flexibility of the model while the increased complexity may cause difficulties in numerical optimization of the model.

Alternatively, a two-step approach can be applied to modeling variance function. After fitting a multilevel model with specified factors to data, residuals from the model can be estimated. Given that the mean of squared residuals is approximately the variance, we may fit a new multilevel model on the squared residuals obtained and take their mean as an estimate of variance. The regression on squared residuals as a variance function modeling has been suggested by many researchers (Goldfeld & Quandt, 1972; Hildreth & Houck, 1968; Jobson & Fuller, 1980). Originally, this approach was suggested for modeling within-individual variance as a function of within-individual covariate (i.e., modeling heterogeneous variance across observations within individuals). Nevertheless, there is no barrier to applying the regression-on-squared-residual approach to multilevel models as well, especially if the goal is to identify the effect of individual-level covariates on intraindividual variances. Because the two-step approach estimates the mean function and the variance function separately, the complexity of the model is reduced and we may less suffer from problems in optimization.

Summary of Introduction and Topics of Following Chapters

Intensive longitudinal study is an emerging area of psychological research. Due to recent developments in data collection, statistical modeling, and computing technology, it is now possible to collect intensive longitudinal data and conduct a proper analysis on ILD. Multilevel models provide a great deal of flexibility in modeling such complex data and are considered as the prevailing approach to ILD by many researchers (Schafer, 2006; Schwartz & Stone 2007; Walls, Höppner, & Goodwin, 2007). However, some properties of statistical models typically used in ILD analysis are not fully investigated yet. Moreover, new developments and deliberate applications in statistical modeling are needed to examine interesting research questions unique to ILD analysis.

Heterogeneous variance and autocorrelation across individuals are likely to exist in most of ILD and may raise serious problems in parameter estimation and interpretation of the mean function. In chapter 2, a brief introduction to multilevel models will be provided as a basis of the current issues. In addition, the issues in modeling residual covariance structure in MLMs will also be discussed. In chapter 3, a multilevel modeling approach that transforms an autocorrelated error structure to an independent structure will be introduced. The transformation is designed to provide a legitimate application of MLM to a serially, and differently, correlated intensive longitudinal data. Using a simulation study, the suggested procedure will be compared with other commonly used MLM approaches that misspecify covariance structure.

In many applications of ILD, heterogeneous variance and autoregressive processes are of significant interest in itself. Chapter 4 will review different approaches that model heterogeneous variance available in current MLM approaches and introduce a two step MLM approach to model heterogeneous variance. A simulation study is conducted to verify the validity of the procedure.

2. Multilevel Models for Longitudinal Data and Modeling Error Covariance Structure

Multilevel models, also known as random effects models (Laird & Ware, 1982), general linear mixed models (Goldstein, 1986), mixed effects models (Pinheiro & Bates, 2000), random coefficient models (Longford, 1993), or hierarchical linear models (Raudenbush & Bryk, 2002), are linear models to analyze data with a multilevel or hierarchical structure where units of analysis are nested within higher level(s), such as students in classes, classes in schools, and so on. Although their use generally applicable to any type of hierarchical or multileveled data, they are widely used for longitudinal data analysis as originated in Laird and Ware (1982). In this chapter, multilevel models for the analysis of longitudinal data are introduced as a basis of following chapters. In addition, issues in modeling covariance structure are also discussed, in the context of autocorrelation, misspecification, and heterogeneity.

Multilevel Models for Longitudinal Data

In longitudinal design, observations can be thought of as, for example, representations of two level data structure where repeated observations (level-1) are nested within individuals (level-2). Suppose there are n_i repeated observations of a variable y_{ti} for individual i, where $t = 1, 2, \ldots, n_i$ and $i = 1, 2, \ldots, N$, with J-1 time varying (within-individual) covariate z_{jti}. A linear regression model for y_{ti} is

$$y_{ti} = \beta_{0i} + \sum_{j=1}^{J-1} \beta_{ji} z_{jti} + e_{ti} \ , \tag{2.1}$$

Suppose also the intercept β_{0i} and the regression coefficients β_{ji} in (2.1) are random and have linear regression models with K-1 time invariant (between-individual) covariate w_{ki} as

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \sum_{k=1}^{K-1} \gamma_{0k} w_{ki} + u_{0i} \\ \beta_{ji} &= \gamma_{j0} + \sum_{k=1}^{K-1} \gamma_{jk} w_{ki} + u_{ji}. \end{aligned} \tag{2.2}$$

In (2.1), β_{0i} is the level-1 intercept and β_{ji} is the regression coefficient of the j^{th} variable z_{jti} within individual i, respectively, and e_{ti} is the residual for occasion t within individual i. Equations in (2.2) are the level-2 equations where γ_{j0} is the individual level intercept, γ_{jk} is the regression coefficient of the k^{th} level-2 variable w_{ki}, and u_{ji} is the residual for individual i for β_{ji}, for j = 0, 1, 2, …, J-1, respectively. By substituting (2.2) into (2.1), the level-1 and level-2 equations can be combined into a single equation form as,

$$\begin{aligned} y_{ti} &= \gamma_{00} + \sum_{k=1}^{K-1} \gamma_{0k} w_{ki} + u_{0i} + \sum_{j=1}^{J-1} (\gamma_{j0} + \sum_{k=1}^{K-1} \gamma_{jk} w_{ki} + u_{ji}) z_{jti} + e_{ti} \\ &= \sum_{k=0}^{K-1} \gamma_{0k} w_{ki} + u_{0i} + \sum_{j=1}^{J-1} (\sum_{k=0}^{K-1} \gamma_{jk} w_{ki} + u_{ji}) z_{jti} + e_{ti} \\ &= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \gamma_{jk} w_{ki} z_{jti} + \sum_{j=0}^{J-1} u_{ji} z_{jti} + e_{ti}, \end{aligned} \tag{2.3}$$

where $w_{0i} = z_{0ti} = 1$.

In matrix terms, (2.1), (2.2), and (2.3) are expressed as

$$\mathbf{y_i} = \mathbf{Z_i} \boldsymbol{\beta}_\mathbf{i} + \mathbf{e_i}, \tag{2.4}$$

$$\boldsymbol{\beta}_\mathbf{i} = \mathbf{W_i} \boldsymbol{\gamma} + \mathbf{u_i} \tag{2.5}$$

$$\mathbf{y_i} = \mathbf{Z_i W_i} \boldsymbol{\gamma} + \mathbf{Z_i u_i} + \mathbf{e_i}, \tag{2.6}$$

respectively, where $\mathbf{Z_i} = \begin{bmatrix} 1 & z_{11i} & \cdots & z_{(J-1)1i} \\ 1 & z_{12i} & \cdots & z_{(J-1)2i} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1n_1 i} & \cdots & z_{(J-1)n_1 i} \end{bmatrix}$, $\mathbf{W_i} = \mathbf{I}_J \otimes \mathbf{w}'_\mathbf{i}$, $\mathbf{w}'_\mathbf{i} = (1, w_{1i}, \ldots, w_{(K\text{-}1)i})$,

and $\boldsymbol{\gamma}' = (\gamma_{00}, \ldots, \gamma_{0(K\text{-}1)}, \gamma_{10}, \ldots, \gamma_{1(K\text{-}1)}, \ldots, \gamma_{(J\text{-}1)0}, \ldots, \gamma_{(J\text{-}1)(K\text{-}1)})$, $\mathbf{u}' = (u_{0i}, u_{1i}, \ldots, u_{(j\text{-}1)i})$, and $\otimes$ denotes the Kronecker product. Equation (2.6) can also be written as

$$\mathbf{y_i} = \mathbf{X_i} \boldsymbol{\gamma} + \mathbf{Z_i u_i} + \mathbf{e_i}, \tag{2.7}$$

where $\mathbf{X_i} = \mathbf{Z_i W_i}$, and is called multilevel models or linear mixed models.

Unlike general linear models (e.g., linear regression or ANOVA), MLMs have two error terms: $\mathbf{u_i}$ and $\mathbf{e_i}$. These two types of error terms are assumed normally distributed with $E\begin{bmatrix} \mathbf{u_i} \\ \mathbf{e_i} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ and $Var\begin{bmatrix} \mathbf{u_i} \\ \mathbf{e_i} \end{bmatrix} = \begin{bmatrix} \mathbf{G_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{R_i} \end{bmatrix}$. The mixture of the two normal distributions results in a multivariate normal distribution of $\mathbf{y_i}$, $\mathbf{y_i} \sim N(\mathbf{X_i}\boldsymbol{\gamma}, \mathbf{V_i})$, where $\mathbf{V_i} = \mathbf{Z_i G_i Z'_i} + \mathbf{R_i}$. [1] In practice, $\mathbf{G_i}$ and $\mathbf{R_i}$ are assumed homogeneous across all level-2 individuals (i.e., $\mathbf{G_1} = \mathbf{G_2} = \ldots = \mathbf{G_I}$ and $\mathbf{R_1} = \mathbf{R_2} = \ldots = \mathbf{R_I}$) in most of applications.

For the entire observations $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_I)'$, the MLM is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Zu} + \mathbf{e}, \tag{2.8}$$

where $\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \vdots \\ \mathbf{X_I} \end{bmatrix}$, $\mathbf{Z}$ is the block diagonal matrix of $\mathbf{Z_i}$, i.e., $\mathbf{Z} = \begin{bmatrix} \mathbf{Z_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z_I} \end{bmatrix}$,

$\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_I)'$, and $\mathbf{e} = (\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_I)'$. $\mathbf{u}$ and $\mathbf{e}$ are normally distributed with

[1] Alternatively, $\mathbf{y_i}$ can also be expressed to have a normal distribution as $\mathbf{y_i} \sim N(\mathbf{X_i}\boldsymbol{\gamma} + \mathbf{Z_i u_i}, \mathbf{R_i})$.

$E\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$ and $Var\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$, where **G** and **R** are the block diagonal matrices of $\mathbf{G_i}$ and $\mathbf{R_i}$, resulting in a multivariate normal distribution of **y**, $\mathbf{y} \sim N(\mathbf{X\gamma}, \mathbf{V})$, where $\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$ is the block diagonal matrices of $\mathbf{V_i}$.

Given normality assumptions of **u** and **e**, maximum likelihood (ML) or restricted maximum likelihood (REML) estimators of **G** and **R** can be obtained by maximizing the corresponding log-likelihood functions as follows:

$$l_{ML}(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{r'V}^{-1}\mathbf{r} - \frac{T}{2}\log(2\pi) \tag{2.9}$$

$$\begin{aligned} l_{REML}(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X'V}^{-1}\mathbf{X}| \\ -\frac{1}{2}\mathbf{r'V}^{-1}\mathbf{r} - \frac{T-P}{2}\log(2\pi), \end{aligned} \tag{2.10}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})\mathbf{X'V}^{-1}\mathbf{y}$, $T = \sum_{i=1}^{N} n_i$ is the total number of observations, and $P = JK$ is the rank of **X**.

Estimates of $\boldsymbol{\gamma}$ and **u** can be obtained by solving following mixed model equations (Henderson, 1984; Rao, 2003):

$$\begin{bmatrix}\mathbf{X'R}^{-1}\mathbf{X} & \mathbf{X'R}^{-1}\mathbf{Z}\\ \mathbf{Z'R}^{-1}\mathbf{X} & \mathbf{Z'R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\gamma}}\\ \hat{\mathbf{u}}\end{bmatrix}=\begin{bmatrix}\mathbf{X'R}^{-1}\mathbf{y}\\ \mathbf{Z'R}^{-1}\mathbf{y}\end{bmatrix} \tag{2.11}$$

which is also written as

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= (\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} &= \mathbf{GZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\gamma}}) \end{aligned}. \tag{2.12}$$

If **G** and **R** are known, $\hat{\boldsymbol{\gamma}}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\gamma}$ and $\hat{\mathbf{u}}$ is the best linear unbiased predictor (BLUP) of **u**, in that $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{u}}$ have the minimum variances among all possible linear unbiased estimators of $\boldsymbol{\gamma}$ and linear unbiased predictors of **u**, respectively. If **G** and **R** are unknown and should be estimated from data through ML or REML estimation by maximizing (2.9) or (2.10), for example, the empirical BLUE (EBLUE) of $\boldsymbol{\gamma}$ and the empirical BLUP (EBLUP) of **u** can be obtained by replacing **G** and **R** with their corresponding ML or REML estimates $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$ in (2.11) or (2.12). The covariance matrix of the EBULE of $\boldsymbol{\gamma}$ and the EBLUP of **u** is given by

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-1}, \tag{2.13}$$

which is the approximation of the true covariance matrix of $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{u}}$:

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1}. \tag{2.14}$$

Autocorrelation, Misspecification, and Heterogeneity in Covariance Structure of MLMs

Common use of MLMs for longitudinal data, including linear growth models where the intercept and the linear slope (and/or higher order of polynomials) over time are estimated, assumes an unstructured $\mathbf{G_i}$ matrix, that allows to estimate variance and covariance for all random effects (e.g., $\mathbf{G_i} = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u0u1} & \sigma^2_{u1} \end{bmatrix}$), and an independent and identical $\mathbf{R_i}$ matrix (ID: $\mathbf{R_i} = \sigma^2_e \mathbf{I}$). However, it is likely that residuals of an MLM for repeated observations have serial correlations across time (over and beyond the fixed and random effects in the model). If this is the case, independence assumption in $\mathbf{R_i}$ is not appropriate and a suitable covariance structure that models autocorrelations between consecutive residuals should be specified in $\mathbf{R_i}$. If observations are measured at equally spaced time, a covariance matrix generated by a first order autoregressive (AR(1)) process can be used to model such a structure. The AR(1) covariance structure with four occasions, for example, is modeled as

$$\mathbf{R_i} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}, \tag{2.15}$$

where ρ is the first order autoregressive parameter (or autocorrelation between observations measured at time t and t+1). Another covariance structure modeling autocorrelations is a covariance matrix that generated by a first order autoregressive and moving average (ARMA(1,1)) process (see chapter 3 for more details about AR and ARMA process). The ARMA(1,1) covariance structure with four occasions, for example, is modeled as

$$\mathbf{R_i} = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}. \tag{2.16}$$

Because **V** is determined by two covariance matrices **G** and **R** as well as a design matrix of random effect **Z**, misspecification of **R** may affect the estimation of **G**, or vice versa, and thus, $\boldsymbol{\gamma}$ and **u** as well. The effect of misspecification of **R** or **G** on the estimations of $\boldsymbol{\gamma}$ (and **G)** has been studied by several researchers, in the context of linear growth models (Ferron et al., 2002; Jacqmin-Gadda et al., 2007; Kwok et al., 2007; Lange & Laird, 1989). A study by Lange and Laird (1989) has shown that misspecification of **G** in a linear growth model, where the mean intercept and the linear slope are estimated (i.e., $\boldsymbol{\gamma} = [\gamma_{00}, \gamma_{10}]'$), affects estimation of the standard error of the estimated means of the intercept and the slope. Specifically, they found that if the intercept is in fact random (i.e., $\mathbf{G_i} = \sigma^2_{u0}$) but misspecified as fixed (i.e., $\mathbf{G_i} = 0$), the standard error of the estimated mean intercept is always underestimated and that of the slope estimate is always overestimated. Moreover, if the true model has both random intercepts and random slopes (i.e., $\mathbf{G_i} = \begin{bmatrix} \sigma^2_{u0} & 0 \\ 0 & \sigma^2_{u1} \end{bmatrix}$) but is misspecified as a model only with a random intercept (i.e., $\mathbf{G_i} = \sigma^2_{u0}$), the standard error of the slope estimate is always underestimated while that of the intercept estimate is not biased (Lange & Laird, 1989).

Other researchers have been more interested in the effect of misspecification of **R** on the estimation of $\boldsymbol{\gamma}$ and **G**. Ferron et al. (2002) found that misspecification of AR(1) residual covariance structure as ID structure in linear growth models results in overestimation of both σ^2_{u0} and σ^2_{u1} in **G** when ρ = .3 or .6, although bias in estimation of σ^2_{u1} is much smaller than that of σ^2_{u0}. They also found 95% confidence intervals for the slope did not cover the true value when the number of individuals is small (N = 30 in the study). Following the previous findings, Kwok et al. (2007) investigated the effect of misspecification in $\mathbf{R_i}$ for various covariance structures, such as ID, AR(1), ARMA(1,1), and TOEP(2) (second banded Toeplitz or first order moving average) structure, and found that underspecification (i.e., misspecification of a covariance structure as a nested structure with smaller number of parameters, such as AR(1) as ID, ARMA(1,1) as AR(1) or ID) produced minor overestimation of the standard errors of estimation for the intercept and the slope as well as noticeable overestimation of variance estimates in **G**.

Jacqmin-Gadda et al. (2007) showed that the estimation of $\boldsymbol{\gamma}$ under the normal ID assumption of $\mathbf{R_i}$ is robust to heteroscedastic residuals (unless the residual variance is a function of individual level covariates) and non-normal residuals. When residuals are serially correlated, however, estimation of $\boldsymbol{\gamma}$ was biased: The coverage rates of 95% confidence intervals for intercept, slope, individual level covariate, and the interaction of the last two were significantly smaller than the nominal value of .95.

In summary, previous research has argued that, for linear growth models, misspecification of **G** matrix produces biased standard errors of the estimated mean intercept and the linear slope. In addition, a falsely assumed ID structure of $\mathbf{R_i}$ causes overestimation of variance components in **G**, especially for the variance of random intercept, and inflated standard errors of $\boldsymbol{\gamma}$ for certain conditions, when there are positive serial correlations of residuals in the true model. Note that all the results mentioned were obtained from linear growth models for traditional longitudinal data where the number of observations for each individual is small to moderate (3 to 12). In such models, time varying covariates are time itself in a polynomial form and the research questions of interest focused on individual differences in a systematic increase or decrease of a variable over time. In intensive longitudinal studies, however, of interest are often the fixed and the random regression effect of time varying covariates other than polynomial linear change.

All the models investigated in the literature above assumed homogeneous $\mathbf{G_i}$ and $\mathbf{R_i}$ across individuals, except for the heteroscedastic conditions in the study by Jacqmin-Gadda et al. (2007). Because $\mathbf{G_i}$ models between-individual variations, the assumption of homogeneous $\mathbf{G_i}$ across individuals looks appropriate in most situations. The homogenous $\mathbf{R_i}$, however, is assumed not because it is strongly believed that residual covariance (i.e., variance and autocorrelations) is same for all individuals. Instead, it is assumed mainly because accurate estimation of individual covariance structure is not plausible with a small to moderate number of observations within individuals and heterogeneity of the covariance structure is not the major interest in traditional longitudinal studies. In addition, it is thought that, if not severe, violation of the homogenous variance assumption does not produce significant bias in estimation of regression parameters as seen in Jacqmin-Gadda et al. (2007). For these reasons, most applications of MLMs in longitudinal study assume homogenous $\mathbf{R_i}$ in practice. In case of intensive longitudinal studies, however, this common practice is questionable on two grounds. First, for such data, it is possible to reliably estimate individual-level covariance structure given the massive number of observations within individuals typically in ILD and, as a result, heterogeneity of $\mathbf{R_i}$ is likely to be found. Second, heterogeneous autocovariance functions within individuals, which require a sizable number of observations to be estimated, may produce significant bias in estimation of $\mathbf{G}$, $\boldsymbol{\gamma}$, and $\mathbf{u}$. More importantly, in many intensive longitudinal studies, heterogeneity of variance and autocovariance between individuals, is of substantive interest and not merely a nuisance factor, making modeling $\mathbf{R_i}$ (instead of or in addition to the mean of $\mathbf{y_i}$) necessary.

In the following chapters, multilevel modeling of intensive longitudinal data in the presence of heterogeneous variance-covariance is investigated. In Chapter 3, the effect of misspecification of heterogeneous covariance structure on the estimation of mean function parameters is investigated along with an introduction and evaluation of a correction procedure for heterogeneous autocorrelation to perform a valid estimation of

MLMs. Chapter 4 introduces two newly developed MLMs to model heterogeneous random variance across individuals and will evaluate their performances as estimators as well through a simulation study.

3. Misspecification and correction of heterogeneous covariance structure in Multilevel Models for ILD

As seen in the previous chapter, in the analysis of longitudinal data, especially of ILD, presence of autocorrelation between successive observations at the individual level is a major concern and needs to be taken into account in the statistical model of interest. However, current use of MLMs for analysis of longitudinal data, including ILD, almost always assumes homogenous residual covariance structure. Violation of homogenous covariance structure may result in bias in estimation of parameters of interest. Given the bias, one possible approach when observations for each individual are collected intensively across time, is to relax the homogeneity assumption in the model and estimate $\mathbf{R_i}$ separately for each i. For studies examining several individuals, however, this relaxation produces too many estimated parameters. For example, if there are 50 individuals in the data and we assume heterogeneous AR(1) structure across individuals, this approach estimates 100 parameters in the covariance structure $\mathbf{R}$. Because these parameters are simultaneously estimated, the large number of parameters may cause optimization problems such as failure to convergence, convergence to local optima, and/or convergence to improper solutions. Moreover, this approach is usually only available in most of statistical programs for a restricted number of simple autocovariance models (e.g., AR(1) or ARMA(1,1)).

Alternatively, a transformation method for regression with autocorrelated error can be applied to the analysis of ILD in this context. In this chapter, a transformation method to model ILD with heterogeneous covariance structure is introduced and its performance is evaluated by a simulation study. In addition, the effects of misspecification of heterogeneous $\mathbf{R_i}$ as ID structure or homogenous AR(1) structure are also investigated. Although ILD consist of a set of time series within multilevel structure (i.e., each time series are nested within individuals), the problem of autocorrelation exists at the first level of time series. As such, we consider a mathematical description of autocorrelation in time series and the transformation method in a single time series first.

Regression with Autocorrelated Errors in a Single Time Series

For single time series data, the effect of autocorrelated errors on estimation of regression parameters is well known and estimation of regressions when residuals are autocorrelated has long been of interest to statisticians (Chipman, 1979; Cochrane & Orcutt, 1949; Harvey 1981; Koreisha & Fang, 2001; Maeshiro, 1980; Park & Mitchell, 1980; Watson, 1955). Although traditional OLS estimation for linear regression model assumes independence of observations, time series data usually violate this assumption. Consider a regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{3.1}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_t, \ldots, y_n)'$, $\mathbf{X}$ is a $n \times q$ design matrix (of input variables), $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{q-1})'$ is a $q \times 1$ regression parameter vector, and $\mathbf{e} = (e_1, e_2, \ldots, e_t, \ldots, e_n)'$ is a $n \times 1$ random residual vector with an $n \times n$ covariance matrix $\boldsymbol{\Sigma} = \sigma_e^2 \mathbf{V}_e$. If e_t is independent and has constant variance across t, i.e., $\boldsymbol{\Sigma} = \sigma_e^2 \mathbf{I}$ or $\mathbf{V}_e = \mathbf{I}$, where $\mathbf{I}$ is an $n \times n$ identity matrix, we can apply ordinary least squares (OLS) to estimate $\boldsymbol{\beta}$ such that

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{3.2}$$

and its covariance matrix is $\sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}$, where the square root of each diagonal element is the standard error of estimation for the corresponding parameter in $\boldsymbol{\beta}$. The OLS estimator $\hat{\boldsymbol{\beta}}$ in this case is known as an unbiased and efficient estimator, or the best linear unbiased estimator (BLUE), in the sense that it has the smallest variance among all possible linear unbiased estimators.

If e_t is serially correlated, i.e., $\boldsymbol{\Sigma} \neq \sigma_e^2\mathbf{I}$, however, (3.2) is no longer efficient. In such cases, generalized least squares (GLS) is used to estimate $\boldsymbol{\beta}$ such that

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}. \tag{3.3}$$

Alternatively, a suitable transformation of $\mathbf{y}$ can also be used. In that case, multiplying (3.1) by a transformation matrix $\mathbf{A}$, such that $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \sigma_w^2\mathbf{I}$, gives

$$\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}\mathbf{e} = \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \mathbf{w} \tag{3.4}$$

where $\mathbf{w}$ is a white noise vector with covariance matrix $\sigma_w^2\mathbf{I}$. (3.4) then can be expressed as

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{w}, \tag{3.5}$$

where $\mathbf{y}^* = \mathbf{A}\mathbf{y}$ and $\mathbf{X}^* = \mathbf{A}\mathbf{X}$. (3.4) or (3.5) provides a valid OLS estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}_w = (\mathbf{X}'\mathbf{A}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}'\mathbf{A}\mathbf{y} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \tag{3.6}$$

because $\sigma_w^2\boldsymbol{\Sigma}^{-1} = \mathbf{A}'\mathbf{A}$. The transformation matrix $\mathbf{A}$ is obtained as $\mathbf{A} = \mathbf{L}^{-1}$, where $\mathbf{L}$ denotes the Cholesky root of $\mathbf{V}_w$, where $\mathbf{V}_w = (\sigma_w^2)^{-1}\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Sigma} = \sigma_w^2\mathbf{V}_w$), that is $\mathbf{V}_w = \mathbf{L}\mathbf{L}'$ with $\mathbf{L}$ lower triangular (Note that $\mathbf{V}_w \neq \mathbf{V}_e$). If we know the covariance matrix $\boldsymbol{\Sigma}$ or $\mathbf{V}$, (3.3) and (3.6) can directly be applied and will produce identical estimates of $\boldsymbol{\beta}$

(i.e., $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_w$). When we don't know the covariance matrix **Σ**, the problem is how to estimate **Σ**.

One possible approach for estimation of **Σ** (and thus a transformation matrix **A**) is to construct **Σ** from a known autocorrelation structure. Pioneering work in this approach was done by Cochrane and Orcutt (1949) for the simple Markov process. For a Markov process, $e_t = \phi_1 e_{t-1} + w_t$, $w_t \sim N(0, \sigma^2_w)$ (i.e., the first order autoregressive process), autocovariance $\gamma(h)$ is well known to be expressed as

$$\gamma(h) = \frac{\sigma_w^2 \rho^h}{1-\rho^2}, \tag{3.7}$$

where $\rho = \phi_1$ is the first order autocorrelation (see Shumway & Stoffer, 2006, pp. 86-87). Thus, autocovariance matrix **Σ** is then

$$\boldsymbol{\Sigma} = \frac{\sigma_w^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^n \\ \rho & 1 & \rho & \cdots & \rho^{n-1} \\ \rho^2 & \phi & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^n & \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}, \tag{3.8}$$

and its inverse matrix is

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_w^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \tag{3.9}$$

Assuming a simple Markov process, the GLS estimator $\tilde{\boldsymbol{\beta}}$ can be obtained by (3.3), where $\boldsymbol{\Sigma}^{-1}$ is specified as in (3.9).

To form a transformation matrix **A**, the inverse of Cholesky factor of $\mathbf{V}_w$ (i.e., the transpose of Cholesky factor of $\mathbf{V}_w^{-1}$, where $\mathbf{V}_w^{-1} = \sigma_w^2 \mathbf{\Sigma}^{-1}$) then can be obtained from $\mathbf{\Sigma}^{-1}$, given by

$$\mathbf{L}^{-1} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \tag{3.10}$$

Using (3.10), a valid OLS estimator $\hat{\boldsymbol{\beta}}_w$ is obtained through (3.6) (Judge, Griffiths, Hill, Lütkepohl, & Lee, 1985). Cochrane and Orcutt (1949) did not provide the exact form of (3.10) but a similar idea of transformation was offered. The following is the generalization of Cochrane and Orcutt's approach to general ARMA(p,q) process. Interested readers may consult chapter 3 and 5 of Shumway and Stoffer (2006) and chapter 8 of Judge et al. (1985) for more details.

There are two well-known processes that produce autocorrelations in a (stationary) time series: autoregressive process and moving average process. An autoregressive process, by definition, means that the current value of the series, e_t, can be explained as a linear function of a unique component of independent normal process and p past values, e_{t-1}, e_{t-2}, …, e_{t-p}. As such, the autoregressive model of order p, or AR(p), can be expressed as a regression model:

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \ldots + \phi_p e_{t-p} + w_t, \tag{3.11}$$

where w_t is a normal white noise process, i.e., $w_t \sim \mathrm{N}(0, \sigma^2_w)$.

Equation (3.11) can be simplified as

$$\phi(B)e_t = w_t, \tag{3.12}$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the autoregressive operator and B^p is a backshift operator, i.e., $B^p e_t = e_{t-p}$. Although autoregressive model in (3.12) assumes that white noise w_t is a linear combination of e_{t-i} of order p, an alternative model represents the observed data e_t as a linear combination of w_{t-i} of order q, i.e., the moving average model of order q or MA(q). The MA(q) model is expressed as

$$e_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}, \tag{3.13}$$

and simplified as

$$e_t = \theta(B) w_t, \tag{3.14}$$

where the moving average operator $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$. The two process can be combined in one model, denoted by ARMA(p,q) and expressed as

$$e_t = \phi_1 e_{t-1} + \ldots + \phi_p e_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}, \tag{3.15}$$

The equation (3.15) is simplified as

$$\phi(B)e_t = \theta(B) w_t, \tag{3.16}$$

Parameters in equation (3.15) or (3.16) can be identified by maximum likelihood or least square estimation. For entire observations, (3.16) is written as in matrix term

$$\boldsymbol{\varphi}(\mathbf{B})^* \mathbf{e}^* = \boldsymbol{\theta}(\mathbf{B})^* \mathbf{w}^* \tag{3.17}$$

where $\mathbf{e}^* = (e_{max(p,q)+1}, \ldots, e_{n-1}, e_n)$ and $\mathbf{w}^* = (w_{max(p,q)+1}, \ldots, w_{n-1}, w_n)$, and $\boldsymbol{\varphi}(\mathbf{B})^*$ and $\boldsymbol{\theta}(\mathbf{B})^*$ are submatrices of $\boldsymbol{\varphi}(\mathbf{B}) = \begin{bmatrix} \boldsymbol{\varphi}(\mathbf{B})^{\sim} \\ \boldsymbol{\varphi}(\mathbf{B})^* \end{bmatrix}$ and $\boldsymbol{\theta}(\mathbf{B}) = \begin{bmatrix} \boldsymbol{\theta}(\mathbf{B})^{\sim} \\ \boldsymbol{\theta}(\mathbf{B})^* \end{bmatrix}$, consisting of rows of $\boldsymbol{\varphi}(\mathbf{B})$ and $\boldsymbol{\theta}(\mathbf{B})$ lower than $max(p,q)$th row, respectively, where

$$
\boldsymbol{\varphi}(\mathbf{B}) = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ -\phi_1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\phi_p & -\phi_{p-1} & \cdots & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi_p & \cdots & -\phi_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\phi_p & -\phi_{p-1} & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & -\phi_p & \cdots & -\phi_1 & 1 \end{bmatrix} \tag{3.18}
$$

and

$$
\boldsymbol{\theta}(\mathbf{B}) = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \theta_1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \theta_q & \theta_{q-1} & \cdots & 1 & 0 & \cdots & 0 & 0 \\ 0 & \theta_q & \cdots & \theta_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \theta_q & \theta_{q-1} & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & \theta_q & \cdots & \theta_1 & 1 \end{bmatrix}. \tag{3.19}
$$

Once the estimates $\hat{\phi}(B)$ and $\hat{\theta}(B)$ are obtained, we can apply (3.4) with **A** = $\left[\hat{\boldsymbol{\theta}}(\mathbf{B})^*\right]^- \hat{\boldsymbol{\varphi}}(\mathbf{B})^*$, where $^-$ denotes a generalized inverse, to get a valid OLS estimator of parameter $\boldsymbol{\beta}$ in (3.4) or, identically a GLS estimator of $\boldsymbol{\beta}$ in (3.1). Notice that, for AR(1) model, the transform matrix $\mathbf{L}^{-1}$ and $\boldsymbol{\varphi}(\mathbf{B})^*$ are identical except that the first row of $\mathbf{L}^{-1}$ is excluded in $\boldsymbol{\varphi}(\mathbf{B})^*$.

The algorithm described above is a direct generalization of Cochrane and Orcutt's approach. Although the algorithm can be used to model both AR and MA process, AR(p) models are widely used in practice. For AR(p) model, the original response values in **y** are transformed as $y_t^* = y_t + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}$. However, direct application of the transformation matrix $\hat{\boldsymbol{\varphi}}(\mathbf{B})^*$ (in case of AR(p) model) excludes first p observations in the estimation, which produces biased estimations, especially in small samples. Use of the transformation matrix $\mathbf{L}^{-1}$, such as (3.10), can avoid exclusion of the several initial observations. In practice, the transformation is carried out using a Kalman filter (Harvey & Phillips, 1979; Jones, 1980), and the lower triangular matrix $\mathbf{L}^{-1}$ is never directly computed. For these reasons, the transformation approach is exclusively used to correct higher order autoregressive error process, because the GLS approach requires a complicated (nonlinear) parameterization of $\mathbf{V}$ or $\mathbf{V}^{-1}$.

Correction for Heterogeneous Autocorrelations for ILD

So far the regression-with-autocorrelation problem has been addressed for a single time series. In ILD, this problem must be extended to multiple (or multilevel) time series and expressed in multilevel models. In the analysis of intensive longitudinal data, the concern for the heterogeneous autocorrelation is high. In applications of MLM for ILD, uncorrected residual dependency may produce bias in estimation and inference. As described above, simultaneous estimation for heterogeneous individual covariance structure may not successfully address this problem. Alternatively, we may apply the transformation approach used in a single time series to multilevel intensive longitudinal data.

Application of the transformation method to ILD is straightforward. With a number of repeated observations for each individual, transformation of each single series may correct autocorrelated errors estimated separately by individual. A time series model (3.1) for individual i is written as

$$\mathbf{y_i} = \mathbf{Z_i}\boldsymbol{\beta}_\mathbf{i} + \mathbf{e_i}, \tag{3.20}$$

where $\mathbf{Z_i}$ is time varying covariates (instead of $\mathbf{X_i}$) as in (2.4). Alternatively, (3.20) can be written in a multilevel format as

$$\mathbf{y_i} = \mathbf{X_i}\boldsymbol{\gamma} + \mathbf{Z_i}\mathbf{u_i} + \mathbf{e_i}, \tag{3.21}$$

where $\mathbf{X_i} = \mathbf{Z_i}\mathbf{W_i}$,, $\mathbf{W_i}$ is the matrix of individual level covariates, $\boldsymbol{\gamma}$ is the fixed effect, and $\mathbf{u_i}$ is the random effect, respectively, as specified in (2.5) through (2.7). The random effect $\mathbf{u_i}$ and the residual $\mathbf{e_i}$ are assumed to be normal distributed with $E\begin{bmatrix}\mathbf{u_i}\\ \mathbf{e_i}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$ and

$Var\begin{bmatrix}\mathbf{u_i}\\ \mathbf{e_i}\end{bmatrix} = \begin{bmatrix}\mathbf{G_i} & \mathbf{0}\\ \mathbf{0} & \mathbf{R_i}\end{bmatrix}$.

If each individual has one's own (heterogeneous) covariance for $\mathbf{e_i}$, (i.e., $\mathbf{R_1} \neq \mathbf{R_2} \neq \ldots \neq \mathbf{R_I}$, for $i = 1, 2, \ldots, I$), the transformed equation for each individual will be given by

$$\begin{aligned}\mathbf{A_i}\mathbf{y_i} &= \mathbf{A_i}\mathbf{X_i}\boldsymbol{\gamma} + \mathbf{A_i}\mathbf{Z_i}\mathbf{u_i} + \mathbf{A_i}\mathbf{e_i}\\ &= \mathbf{A_i}\mathbf{X_i}\boldsymbol{\gamma} + \mathbf{A_i}\mathbf{Z_i}\mathbf{u_i} + \mathbf{w_i}\end{aligned}, \tag{3.22}$$

where random effect $\mathbf{u_i}$ and residual $\mathbf{w_i}$ are then normal distributed with $E\begin{bmatrix}\mathbf{u_i}\\ \mathbf{w_i}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$

and $Var\begin{bmatrix}\mathbf{u_i}\\ \mathbf{w_i}\end{bmatrix} = \begin{bmatrix}\mathbf{G_i} & \mathbf{0}\\ \mathbf{0} & \sigma^2_{w_i}\mathbf{I}_{n_i\times n_i}\end{bmatrix}$. $\mathbf{A_i}$ can be obtained by solving (3.12), if AR(p) model is

assumed, where $\mathbf{A_i} = \hat{\boldsymbol{\varphi}}\mathbf{(B)}^*_\mathbf{i}$, or solving (3.16) or (3.17) in more general cases, where $\mathbf{A_i}$ =

$\left[\hat{\boldsymbol{\theta}}\mathbf{(B)}^*_\mathbf{i}\right]^- \hat{\boldsymbol{\varphi}}\mathbf{(B)}^*_\mathbf{i}$. For the entire system, the transformed equation is written as

$$\begin{aligned}\mathbf{Ay} &= \mathbf{AX}\boldsymbol{\gamma} + \mathbf{AZu} + \mathbf{Ae}\\ &= \mathbf{AX}\boldsymbol{\gamma} + \mathbf{AZu} + \mathbf{w}\end{aligned}, \tag{3.23}$$

where $\mathbf{A}$ is the block diagonal matrix of $\mathbf{A_i}$ and $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{u}$ are specified as in (2.8). Assuming homogeneous variance of the transformed residuals ($\sigma_{w_1}^2 = \sigma_{w_2}^2 = \ldots = \sigma_{w_k}^2$), the transformed variables ($\mathbf{Ay}$, $\mathbf{AX}$, and $\mathbf{AZ}$) can be modeled in a multilevel model with the ID residual structure. Correction (3.22) or (3.23) is expected to reduce bias, if any, in estimation of the parameters when there are heterogeneous autocorrelations in the data in use.

The correction procedure of multilevel models with heterogeneous autocorrelations is summarized as

(1) Fit (3.20) by OLS estimation for each individual. Obtain $\hat{\mathbf{y}}_\mathbf{i} = \mathbf{Z_i}\hat{\boldsymbol{\beta}}_\mathbf{i}$

(2) Calculate residuals as $\hat{\mathbf{e}}_\mathbf{i} = \mathbf{y}_\mathbf{i} - \hat{\mathbf{y}}_\mathbf{i}$ and investigate autocorrelations for $\hat{\mathbf{e}}_\mathbf{i}$.

(3) Define order *p* of AR(*p*) model for each individual by investigating autocorrelation patterns in residuals.

(4) Apply transformation procedure by solving (3.12) for each individual to correct defined autocovariance structures.

(5) Fit the intended MLM with $\mathbf{R_i} = \sigma_w^2\mathbf{I}$. Obtain estimates for parameters of interest

Simply speaking, the above procedure consists of two steps. In the first step, a transformation matrix is obtained from a regression-with-autoregressive-error model for each individual. In the second step, the intended MLM is applied to the transformed variables obtained in the first step. We call this correction method as the two-step multilevel model with transformation (TS MLM-T). The TS MLM-T has several strengths in correction of autocorrelation in error structure. First, correction is applied for each individual, without unrealistic homogeneity assumption in ILD. Second, individual specific correction is made avoiding risk in optimization process in MLM with heterogeneous covariance where a number of parameters are estimated simultaneously.

Third, time intervals between successive observations are not restricted to be equal across different individuals, although time intervals within individuals are restricted to be similarly spaced as in the ARMA models. Last, and most importantly, the correction is applicable even higher order autoregressive error structure and allows different orders of autoregressive processes across individuals.

Performance in Estimation of MLMs with Heterogeneous Autoregressive Errors: A Simulation Study

As seen above, heterogeneity of residual covariance structure is likely in ILD and in many applications, likely affects the estimation of the parameters of interest in MLMs. Currently, common practice in using MLMs for such data is to ignore this plausible violation of the homogeneity assumption. As such, we focus on the effect of misspecification of residual covariance structure $\mathbf{R}$ on the estimation of fixed effects and variance components of the random effects, in the sense that if MLMs assume homogeneous $\mathbf{R_i}$, when it is in fact heterogeneous, it may cause bias in estimation of the parameters in MLMs. To this end, a simulation study was conducted where data generated from a longitudinal multilevel structure with heterogeneous autoregressive error process were analyzed by MLMs with homogeneous assumption. The two most commonly used covariance structures for the analysis of longitudinal data (i.e., ID and AR(1)) were used to build misspecified models. In addition, the performance of the TS MLM-T was evaluated in terms of reduction of bias in estimation.

Method

For simplicity, the following (two-level) linear mixed model was used to generate data:

$$y_{ti} = \gamma_{00} + \gamma_{10}x_{ti} + u_{0i} + x_{ti}u_{1i} + e_{ti} \tag{3.24}$$

where γ_{00} is the fixed intercept, γ_{10} is the fixed effect of a time varying covariate x_{ti}, u_{0i} is the random effect of intercept, and u_{1i} is the random effect of x_{ti} that is generated from a normal distribution as $x_{ti} \sim N(0,1)$. The parameters γ_{00} and γ_{10} were set to 1. The random effect u_{0i} and u_{1i} were distributed multivariate normal as $\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{bmatrix}\right)$,

where $\sigma_{u0}^2 = .5$, $\sigma_{u1}^2 = .5$, and $\sigma_{u0u1} = .15$ (i.e., $r_{u0u1} = .3$). The errors were generated with a first order autoregressive model, $e_{ti} = \rho_i e_{ti} + w_{ti}$, $w_{ti} \sim N(0,1)$.

The autoregressive parameter ρ_i was allowed to vary across individuals, generated from a uniform distribution as $\rho_i \sim U(\rho - .3, \rho + .3)$, where $\rho = 0, .3$, or $.6$, providing that $E(\rho) = 0, .3$, or $.6$, respectively. Each data set was completely balanced with L (series length, or the number of observations within individuals) = 10, 20, 50, or 100 and N (the number of individuals) = 20, 50, or 100. Accordingly, 3 (ρ) × 4 (L) × 3 (N) = 36 conditions were obtained. In each condition, 500 data sets were simulated, resulting in total of 18000 data sets. Each data set was analyzed three times separately by three different MLMs: MLM with ID covariance structure (ID), MLM with homogeneous AR(1) covariance structure (AR[1]), and the TS MLM-T (TST) introduced above. For the transformation procedure in the first step of TS MLM-T, regression with autoregressive error models was fitted for each individual and variables were transformed by using the AUTOREG procedure in SAS. After transformation, the transformed variables were fitted by a MLM in the second step. All three MLM were properly modeled as in (3.24) and fit using the MIXED procedure in SAS with RMLE estimation.

Bias of parameter estimation was investigated in terms of relative bias for the fixed effects γ_{00} and γ_{10}, and variance components, σ_{u0}^2, σ_{u1}^2, and σ_{u0u1}. Relative bias was calculated as $\frac{1}{R}\sum_{r=1}^{R}\frac{\hat{\theta}_r-\theta}{\theta}$, where θ is the true parameter value, $\hat{\theta}_r$ is the corresponding sample estimate of rth sample, and R is the number of replications converged in each condition. Bias in the estimated standard error was also investigated. Relative bias of standard error for the fixed effects was calculated as $\frac{1}{R}\sum_{r=1}^{R}\frac{\hat{\theta}_r-\theta}{\theta}$, where θ is the true standard error obtained from (2.14) and $\hat{\theta}_r$ is the estimated standard error obtained from (2.13) for rth sample.

No significant bias was expected in the estimates of the fixed effects obtained by the two misspecified MLMs (i.e., ID model and AR[1] model) because misspecified error covariance structure is unlikely to influence bias in point estimation of the fixed effects. Estimates of the variance components of random effects, however, are likely to be biased for the two MLMs with misspecified covariance structures, especially for the variance of random intercept and with high serial correlations. This bias was expected to be greater for the ID model than the AR(1) model because the first is more restricted by the independence assumption. In addition, TS MLM-T is expected to reduce bias of estimates of the variance components in some conditions, but not in other conditions. Specifically, because a successful correction of the TS MLM-T depends on the valid estimation of the transformation matrix (i.e., the autoregressive parameters) in the first step, which requires enough number of observations for each individual, a less biased estimation in the second step was expected not for the data with a small to moderate number of observations (e.g., L = 10 or 20) but for the data with a large number of observations for each individual (e.g., L = 50 or 100). Bias in estimation of the standard error of estimation for the fixed effects is also more likely in the ID model and the AR(1) model than from the TST model, because the estimated standard error is a function of the estimated covariance as seen in (2.13). Because the estimation of random effects is also affected by the covariance structure in the model, a better performance in the estimation of the random effects and the corresponding standard error was expected for the TST model than the others models when series length is long enough.

Results[2]

Bias in fixed effects. Relative bias (*RB*) for the fixed effects γ_{00} and γ_{10} are presented in Table 3.1. Null bias was tested using t-test for each method. There was no significant bias in the estimation of the fixed intercept for all the three methods: *RB* = -0.00, $t_{(17999)} = -0.91$, $p = .36$ for ID; *RB* = -0.00, $t_{(17999)} = -0.80$, $p = .43$ for AR(1); *RB* = -0.00, $t_{(17998)} = -0.63$, $p = .53$ for TST. To test effects of the sample size (N), series length (L), degree of autocorrelation (ρ), and their interactions on relative bias, a $3(N)\times4(L)\times3(\rho)$ ANOVA was conducted for each method. Across all the methods, no significant effect was found except for the interaction effect of sample size by series length: $F_{(6,17964)} = 2.30$, $p < .05$ for ID; $F_{(6,17964)} = 2.40$, $p < .05$ for AR(1); $F_{(6,17963)} = 2.47$, $p < .05$ for TST. However, the effect sizes (η^2) of the $N \times L$ interaction effect for all the three methods

[2] Of the total of 3×18000 analyses, only one analyzed by TS MLM-T on a sample from the condition of ρ = .3, N = 20, L = 10 did not converge.

were less than .001. The results for the fixed regression effect were almost identical. All the three methods did not show any significant overall bias in the estimation of the fixed regression effect: *RB* = -0.00, $t_{(17999)}$ = -0.62, *p* = .54 for ID; *RB* = -0.00, $t_{(17999)}$ = -0.30, *p* = .76 for AR(1); *RB* = -0.00, $t_{(17998)}$ = -0.38, *p* = .70 for TST. The ANOVA for the

Table 3.1
Relative Bias of $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ under Heterogeneous Autocorrelation in Covariance for the Three Different MLMs

		ρ = .0			ρ = .3			ρ = .6		
N	L	ID	AR(1)	TST	ID	AR(1)	TST	ID	AR(1)	TST
					$\hat{\gamma}_{00}$					
20	10	-0.02	-0.02	-0.01	-0.02	-0.02	-0.02	0.00	0.00	0.00
	20	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00
	50	0.00	0.00	0.00	0.01	0.01	0.01	-0.01	-0.01	-0.01
	100	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.01
50	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	20	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01
	50	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01
	100	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.00	0.00	0.00
100	10	0.00	0.00	0.00	-0.01	-0.01	-0.01	0.00	0.00	0.00
	20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	50	0.00	0.00	0.00	0.01	0.01	0.00	-0.01	-0.01	-0.01
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
					$\hat{\gamma}_{10}$					
20	10	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02	-0.01	-0.01
	20	0.01	0.01	0.01	0.00	0.00	0.00	-0.01	0.00	0.00
	50	0.00	0.00	0.00	0.01	0.00	0.01	-0.01	-0.01	-0.01
	100	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01
50	10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	20	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	-0.01	-0.01
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	100	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.00	0.00	0.00
100	10	0.00	0.00	0.00	0.01	0.01	0.00	-0.01	-0.01	-0.01
	20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note. ID: MLM with homogenous variance assumption; AR(1) MLM with homogeneous first order autoregressive error structure; TST: Two-step MLM with transformation.
ρ = average autocorrelation, N = sample size and L = series length.

relative bias also did not present any significant effects except for the $N \times L$ interaction effect: $F_{(6,17964)} = 2.42$, $p < .05$ for ID; $F_{(6,17964)} = 2.32$, $p < .05$ for AR(1); $F_{(6,17963)} = 2.16$, $p < .05$ for TST. The effect sizes (η^2) of the $N \times L$ interaction effect for the three methods did not exceed .001. The result suggests that the estimates of the fixed effects obtained by MLMs with homogenous covariance assumption are not biased when the error covariance structure is in fact heterogeneous. This is true whether sample size is small or large, series length is short or long, and the average error autocorrelation is null or high, at least for $N = (20,100)$, $L = (10,100)$, and $\rho = (.0,.6)$. It is also suggested that the transformation procedure does not produce biased estimates for the fixed effects in MLMs.

Bias in variance and covariance of random effects. Relative bias for the variance of the random intercept (σ_{u0}^2) is presented in Table 3.2. There was a significant bias in the estimation of σ_{u0}^2 for all the three methods: $RB = 0.29$, $t_{(17999)} = 68.68$, $p < .0001$ for ID; $RB = 0.09$, $t_{(17999)} = 30.85$, $p < .0001$ for AR(1); $RB = 0.16$, $t_{(17998)} = 47.27$, $p <$ .0001for TST. ANOVA showed significant effects of series length ($F_{(3,17964)} = 1417.01$, $p < .0001$, $\eta^2 = .10$), autocorrelation ($F_{(2,17964)} = 7074.03$, $p < .0001$, $\eta^2 = .34$), and their interaction ($F_{(6,17964)} = 843.01$, $p < .0001$, $\eta^2 = .12$) for the ID model. For the AR(1) model, significant effects of series length ($F_{(3,17964)} = 124.56$, $p < .0001$, $\eta^2 = .02$), autocorrelation ($F_{(2,17964)} = 820.45$, $p < .0001$, $\eta^2 = .08$), their interaction ($F_{(6,17964)} = 86.14$, $p < .0001$, $\eta^2 = .03$), and the interaction of sample size by series length ($F_{(6,17964)} = 2.87$, $p < .001$, $\eta^2 = .00$) were found. The TST model showed significant effects of series length ($F_{(3,17963)} = 1241.06$, $p < .0001$, $\eta^2 = .12$), autocorrelation ($F_{(2,17963)} = 2218.61$, $p < .0001$, $\eta^2 = .15$), and their interaction ($F_{(6,17963)} = 625.51$, $p < .0001$, $\eta^2 = .13$). ANOVA

Table 3.2

Relative Bias of $\hat{\sigma}^2_{u0}$ under Heterogeneous Autocorrelation in Covariance for the Three Different MLMs

		ρ = .0			ρ = .3			ρ = .6		
N	L	ID	AR(1)	TST	ID	AR(1)	TST	ID	AR(1)	TST
20	10	-0.01	-0.01	0.02	0.25	0.04	0.19	1.44	0.40	0.98
	20	0.00	0.00	0.00	0.14	0.03	0.05	0.97	0.33	0.45
	50	0.01	0.01	0.00	0.06	0.01	0.01	0.48	0.18	0.14
	100	0.00	0.00	0.00	0.03	0.00	0.00	0.26	0.10	0.06
50	10	0.02	0.02	0.06	0.26	0.06	0.20	1.44	0.45	1.03
	20	0.02	0.02	0.03	0.11	0.00	0.03	0.95	0.33	0.47
	50	0.02	0.02	0.01	0.06	0.02	0.01	0.44	0.15	0.11
	100	0.00	0.00	0.00	0.01	-0.01	-0.02	0.23	0.08	0.03
100	10	0.01	0.01	0.06	0.25	0.05	0.19	1.45	0.52	1.05
	20	0.00	0.00	0.00	0.13	0.02	0.05	0.93	0.33	0.45
	50	-0.01	-0.01	-0.01	0.06	0.02	0.01	0.43	0.14	0.09
	100	-0.01	-0.01	-0.01	0.02	0.00	0.00	0.23	0.08	0.03

Note. ID: MLM with homogenous variance assumption; AR(1) MLM with homogeneous first order autoregressive error structure; TST: Two-step MLM with transformation. ρ = average autocorrelation, N = sample size and L = series length.

tables of relative bias for σ^2_{u0} across the three models are presented in Table 3.3. The result showed that the misspecified homogenous covariance models produced biased estimates for the variance of random intercept. It also showed that the bias is large when autocorrelation is high, series length is short, or both. This bias was higher in the ID model than the AR(1) model, $F_{(1,17964)} = 28685.8$, $p < .0001$, $\eta^2 = .04$, analyzed by 2(method) ×3(N)×4(L)×3(ρ) repeated ANOVA.[3] On the other hand, the transformation method did not completely eliminate the bias of the estimates of σ^2_{u0}. Instead, it produced larger bias than the AR(1) model when series length is 10 or 20. However, the TST model showed smaller bias than the AR(1) model when series length is 50 or 100 (Figure 3.1). When the two models were compared, there were considerable interaction effects of

[3] Because each method was fitted to the same data sets, the method factor should be considered as a within-subject factor instead of a between-subject factor.

method by series length, $F_{(3,17963)} = 2132.06$, $p < .0001$, $\eta^2 = .02$, method by autocorrelation, $F_{(2,17963)} = 997.02$, $p < .0001$, $\eta^2 = .00$, and method by series length by autocorrelation, $F_{(6,17963)} = 896.31$, $p < .0001$, $\eta^2 = .01$, as well as a main effect of method, $F_{(1,17963)} = 2413.18$, $p < .0001$, $\eta^2 = .01$.

By contrast, no significant bias was found in the estimation of the variance of the

Table 3.3

Analysis of Variance for Relative Bias of σ^2_{u0} *by Three Models*

Source	*df*	SS	MS	*F*	η^2
		ID			
Sample size (N)	2	0.26	0.13	0.88	0.00
Series length (L)	3	617.88	205.96	1417.01**	0.10
N × L	6	0.55	0.09	0.63	0.00
Autocorrelation (ρ)	2	2056.39	1028.20	7074.03**	0.34
N × ρ	4	0.85	0.21	1.47	0.00
L × ρ	6	735.18	122.53	843.01**	0.12
N × L × ρ	12	0.63	0.05	0.36	0.00
Error	17964	2611.04			
		AR(1)			
Sample size (N)	2	0.11	0.05	0.37	0.00
Series length (L)	3	54.60	18.20	124.56**	0.02
N × L	6	2.51	0.42	2.87*	0.00
Autocorrelation (ρ)	2	239.78	119.89	820.45**	0.08
N × ρ	4	0.48	0.12	0.83	0.00
L × ρ	6	75.53	12.59	86.14**	0.03
N × L × ρ	12	2.17	0.18	1.24	0.00
Error	17964	2625.06			
		TST			
Sample size (N)	2	0.05	0.03	0.20	0.00
Series length (L)	3	466.12	155.37	1241.06**	0.12
N × L	6	1.57	0.26	2.09	0.00
Autocorrelation (ρ)	2	555.52	277.76	2218.61**	0.15
N × ρ	4	0.38	0.09	0.75	0.00
L × ρ	6	469.86	78.31	625.51**	0.13
N × L × ρ	12	1.40	0.12	0.93	0.00
Error	17963	2248.884			

Note. ID: MLM with homogenous variance assumption; AR(1) MLM with homogeneous first order autoregressive error structure; TST: Two-step MLM with transformation

* $p < .05$, ** $p < .0001$

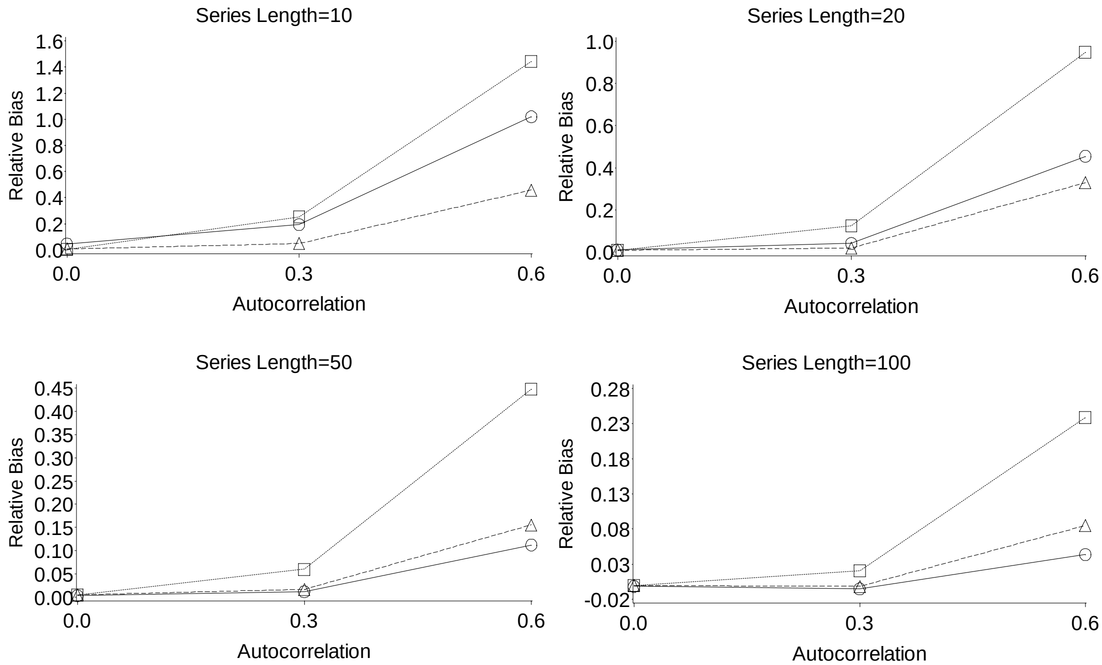


Figure 3.1. Line plots of relative bias of $\hat{\sigma}^2_{u0}$ produced by three models across autocorrelations and series lengths. The square, triangle, and circle represent a relative bias for the ID model, the AR(1) model and the TST model, respectively.

random regression effect and the covariance of the random intercept and the random regression effect for the two misspecified models: The ID model did not show any significant bias for the variance σ^2_{u1}, *RB* = 0.00, $t_{(17999)} = 0.86$, $p = .39$, or for the covariance σ_{u0u1}, *RB* = 0.00, $t_{(17999)} = 1.22$, $p = .22$. The AR(1) model also did not produce biased estimates for σ^2_{u1}, *RB* = 0.00, $t_{(17999)} = 1.37$, $p = .17$, or for σ_{u0u1}, *RB* = 0.00, $t_{(17999)} = 1.10$, $p = .27$ (see Table 3.4). No main effects or interaction effects of *N*, *L*, and ρ significantly explained the variation of the relative bias of σ^2_{u1} and σ_{u0u1} for the two models. On the contrary, the TST model produced a significant bias in the estimation of σ^2_{u1}, *RB* = 0.03, $t_{(17998)} = 13.66$, $p < .0001$. There also was a significant effect of series length on the relative bias of the estimates, $F_{(3,17963)} = 112.88$, $p < .0001$, $\eta^2 = .02$. The

Table 3.4

Relative Bias of $\hat{\sigma}^2_{u1}$ *and* $\hat{\sigma}_{u0u1}$ *under Heterogeneous Autocorrelation in Covariance for the Three Different MLMs*

		ρ = .0			ρ = .3			ρ = .6		
N	L	ID	AR(1)	TST	ID	AR(1)	TST	ID	AR(1)	TST
						$\hat{\sigma}^2_{u1}$				
20	10	-0.01	-0.01	0.08	0.01	0.01	0.10	-0.01	0.00	0.07
	20	0.00	0.00	0.02	0.02	0.02	0.04	-0.01	0.00	0.01
	50	-0.01	-0.01	-0.01	0.00	0.00	0.00	-0.02	-0.02	-0.02
	100	0.02	0.02	0.02	0.00	0.00	0.00	0.01	0.01	0.01
50	10	0.02	0.02	0.12	-0.01	-0.01	0.09	0.01	0.01	0.08
	20	0.01	0.01	0.03	-0.01	-0.01	0.01	-0.01	-0.01	0.00
	50	-0.01	-0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.02
100	10	0.01	0.01	0.10	0.00	0.00	0.09	0.01	0.01	0.08
	20	0.00	0.00	0.02	-0.01	-0.01	0.01	0.01	0.01	0.02
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	100	0.01	0.01	0.01	-0.01	-0.01	-0.01	-0.01	0.00	0.00
						$\hat{\sigma}_{u0u1}$				
20	10	-0.02	-0.02	-0.01	0.01	0.01	0.01	0.01	0.01	0.00
	20	-0.02	-0.01	-0.01	0.01	0.01	0.01	0.00	-0.01	-0.01
	50	0.01	0.01	0.01	-0.01	-0.01	-0.01	0.02	0.02	0.01
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.02
50	10	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.01
	20	0.01	0.01	0.01	-0.01	-0.01	-0.01	0.01	0.01	0.01
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01
100	10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	20	0.00	0.00	0.00	-0.01	-0.01	-0.01	0.00	0.00	0.00
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01

Note. ID: MLM with homogenous variance assumption; AR(1) MLM with homogeneous first order autoregressive error structure; TST: Two-step MLM with transformation. ρ = average autocorrelation, N = sample size and L = series length.

effect of series length implies that the bias is significant only when the series length is short, as seen in Table 3.4. There was no significant bias of the estimates of covariance σ_{u0u1}, *RB* = 0.00, $t_{(17998)}$ =0.90, p = .37, and no effects of *N*, *L*, and ρ for the TST model.

Bias in the standard error of estimation for fixed effects. Because of the bias in the estimated variance of the random intercept, the standard error of estimation of fixed effects was also expected to be biased. Table 3.5 presents relative bias of the estimated standard error for $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$. Significant relative bias of the estimated standard error of

Table 3.5

Relative Bias of the Standard Error of $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ under Heterogeneous Autocorrelation in Covariance for the Three Different MLMs

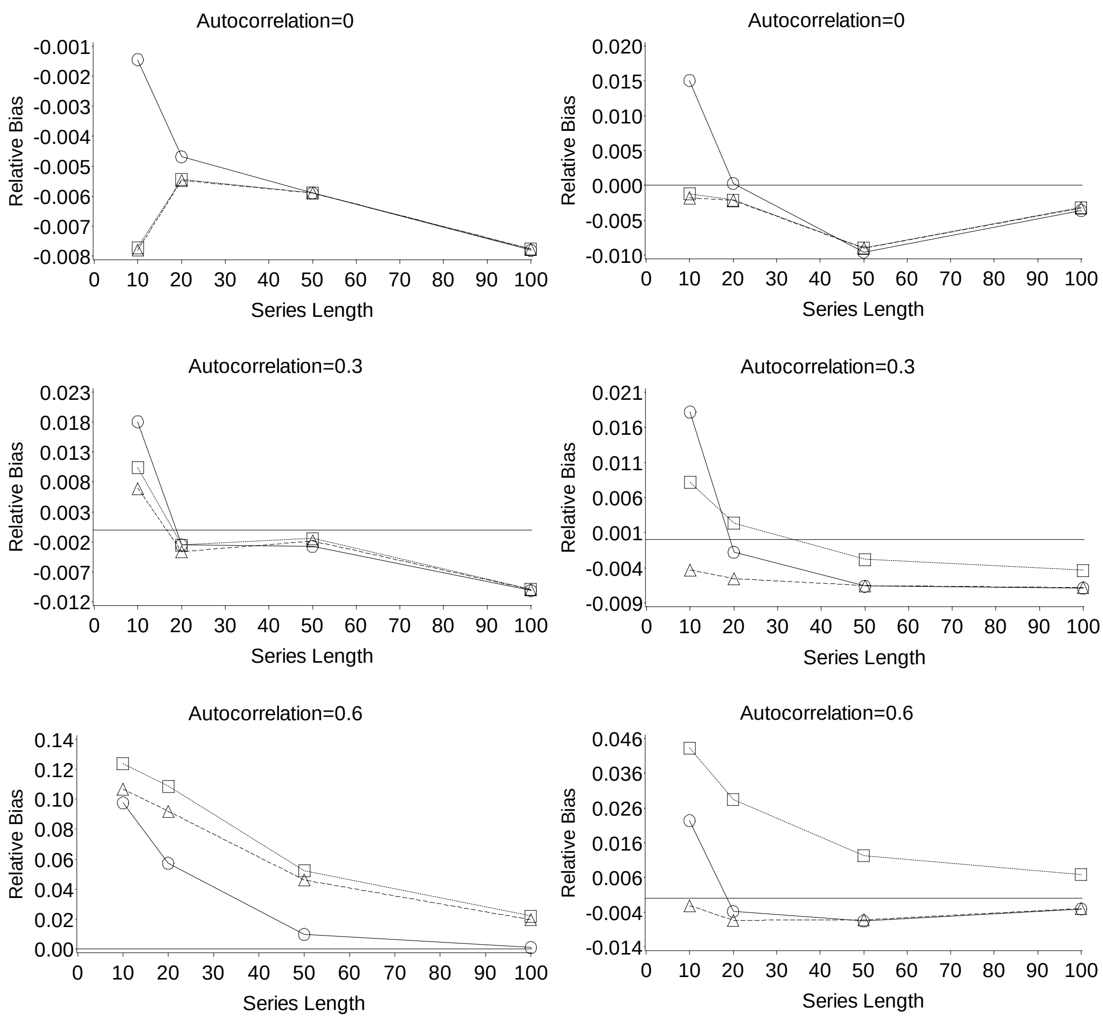
		ρ = .0			ρ = .3			ρ = .6		
N	L	ID	AR(1)	TST	ID	AR(1)	TST	ID	AR(1)	TST
					$SE(\hat{\gamma}_{00})$					
20	10	-0.02	-0.02	-0.02	0.00	0.00	0.01	0.11	0.10	0.08
	20	-0.02	-0.02	-0.02	0.00	-0.01	0.00	0.10	0.09	0.04
	50	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.05	0.05	0.01
	100	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.02	0.02	0.00
50	10	0.00	0.00	0.01	0.01	0.01	0.02	0.13	0.11	0.10
	20	0.00	0.00	0.01	-0.01	-0.01	-0.01	0.11	0.10	0.07
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.01
	100	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.02	0.02	0.00
100	10	0.00	0.00	0.01	0.01	0.01	0.02	0.13	0.11	0.11
	20	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.09	0.06
	50	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.05	0.04	0.01
	100	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.02	0.02	0.00
					$SE(\hat{\gamma}_{10})$					
20	10	-0.01	-0.01	0.00	0.00	-0.01	0.01	0.03	-0.01	0.01
	20	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.02	-0.01	-0.01
	50	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	0.00	-0.02	-0.02
	100	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	-0.01
50	10	0.01	0.01	0.02	0.01	0.00	0.02	0.05	0.00	0.02
	20	0.00	0.00	0.00	0.00	-0.01	0.00	0.03	-0.01	-0.01
	50	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	0.02	0.00	0.00
	100	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.01	0.00	0.00
100	10	0.00	0.00	0.02	0.01	0.00	0.02	0.05	0.00	0.03
	20	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
	100	0.00	0.00	0.00	0.00	-0.01	-0.01	0.00	0.00	0.00

Note. ID: MLM with homogenous variance assumption; AR(1) MLM with homogeneous first order autoregressive error structure; TST: Two-step MLM with transformation; SE: standard error.

ρ = average autocorrelation, N = sample size and L = series length.

the fixed intercept ($\hat{\gamma}_{00}$) was found in all the methods: *RB* = 0.02, $t_{(17999)} = 23.31$, $p <$ .0001 for ID; *RB* = 0.02, $t_{(17999)} = 19.74$, $p < .0001$ for AR(1); *RB* = 0.01, $t_{(17998)} = 13.05$, < .0001 for TST. For the ID model, there were significant effects of sample size ($F_{(2,17964)} = 9.31$, $p < .0001$, $\eta^2 = .00$), series length ($F_{(3,17964)} = 97.30$, $p < .0001$, $\eta^2 = .01$), autocorrelation ($F_{(2,17964)} = 837.60$, $p < .0001$, $\eta^2 = .08$), and the interaction of the last two ($F_{(6,17964)} = 64.87$, $p < .0001$, $\eta^2 = .02$). For the AR(1) model, significant effects of sample size ($F_{(2,17964)} = 9.40$, $p < .0001$, $\eta^2 = .00$), series length ($F_{(3,17964)} = 70.65$, $p < .0001$, $\eta^2 = .01$), autocorrelation ($F_{(2,17964)} = 655.74$, $p < .0001$, $\eta^2 = .07$), and the interaction of the last two ($F_{(6,17964)} = 47.66$, $p < .0001$, $\eta^2 = .01$) were found. The TST model also showed significant effects of sample size ($F_{(2,17963)} = 17.01$, $p < .0001$, $\eta^2 = .00$), series length ($F_{(3,17963)} = 113.80$, $p < .0001$, $\eta^2 = .02$), their interaction ($F_{(6,17963)} = 2.42$, $p < .05$, $\eta^2 = .00$), autocorrelation ($F_{(2,17963)} = 253.12$, $p < .0001$, $\eta^2 = .03$), and the interaction of series length and autocorrelation ($F_{(6,17963)} = 50.62$, $p < .0001$, $\eta^2 = .02$).

As in the bias of $\hat{\sigma}^2_{u0}$, there was a significant mean difference between the ID model and the AR(1) model ($F_{(1,17964)} = 1341.52$, $p < .0001$, $\eta^2 = .00$) as well as significant interaction effects of method by series length ($F_{(3,17964)} = 177.30$, $p < .0001$, $\eta^2 = .00$), method by autocorrelation ($F_{(2,17964)} = 945.17$, $p < .0001$, $\eta^2 = .00$), and method by series length by autocorrelation ($F_{(6,17964)} = 111.90$, $p < .0001$, $\eta^2 = .00$). There also was a significant mean difference between the AR(1) model and the TST model ($F_{(1,17963)} = 449.96$, $p < .0001$, $\eta^2 = .00$) as well as significant interaction effects of method by series length ($F_{(3,17963)} = 117.61$, $p < .0001$, $\eta^2 = .00$), method by autocorrelation ($F_{(2,17963)} = 812.47$, $p < .0001$, $\eta^2 = .00$), and method by series length by autocorrelation ($F_{(6,17964)} = 30.19$, $p < .0001$, $\eta^2 = .00$). The results support the transformation method is less biased in the estimation of the standard error of $\hat{\gamma}_{00}$ than the misspecified models, especially when the average autocorrelation is high and the series length is long (see Figure 3.2). The reduced bias achieved by the transformation method is expected to increase the accuracy of interval estimation and test of the fixed intercept.



SE($\hat{\gamma}_{00}$) SE($\hat{\gamma}_{10}$)

Figure 3.2. Line plots of relative bias of the standard error of $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ produced by three models across autocorrelations and series lengths. The square, triangle, and circle represent a relative bias for the ID model, the AR(1) model and the TST model, respectively.

On the other hand, a significant bias of the standard error of $\hat{\gamma}_{10}$ was found for the misspecified models (RB = 0.01, $t_{(17999)}$ = 7.38, p < .0001 for ID; RB = -0.00, $t_{(17999)}$ = -5.47, p < .0001 for AR[1]), but not for the TST model (RB = 0.00, $t_{(17999)}$ = 1.32, p = .19). ANOVA showed significant effects of sample size ($F_{(2,17964)}$ = 17.39, p < .0001, η^2 = .00), series length ($F_{(3,17964)}$ = 21.46, p < .0001, η^2 = .00), autocorrelation ($F_{(2,17964)}$ = 86.26, p < .0001, η^2 = .01), and the interaction of the last two ($F_{(6,17964)}$ = 6.25, p < .0001, η^2 = .00) for the ID model. For the AR(1) model, only the effect of sample size ($F_{(2,17964)}$ = 15.32, p < .0001, η^2 = .00) was significant. The TST model showed significant effects of sample size ($F_{(2,17963)}$ = 14.82, p < .0001, η^2 = .00) and series length ($F_{(3,17963)}$ = 45.54, p < .0001, η^2 = .01). The ID model and the AR(1) model were significantly different in their mean ($F_{(2,17964)}$ = 3352.51, p < .0001, η^2 = .00) and the effects of series length ($F_{(3,17964)}$ = 310.82, p < .0001, η^2 = .00), autocorrelation ($F_{(2,17964)}$ = 1724.30, p < .0001, η^2 = .00), and the interaction of the last two ($F_{(6,17964)}$ = 145.66, p < .0001, η^2 = .00). The AR(1) model and the TST model were significantly different in their mean ($F_{(2,17963)}$ = 1211.69, p < .0001, η^2 = .00) and the effects of series length ($F_{(3,17963)}$ = 924.47, p < .0001, η^2 = .00), autocorrelation ($F_{(2,17964)}$ = 15.40, p < .0001, η^2 = .00), and the interaction of the last two ($F_{(6,17964)}$ = 10.58, p < .0001, η^2 = .00). Overall all models are downwardly biased for the estimates of the standard error of $\hat{\gamma}_{10}$. However, overestimation occurred when autocorrelation was .3 and series length was short or autocorrelation was .6 for the ID model, and when series length was 10 for the TST model. Figure 3.2 (right column) presents relative bias in the estimation of the standard error of $\hat{\gamma}_{10}$ for the three models across series length and autocorrelation.

Conclusions

In general, all three models failed to produce highly biased estimates of the fixed intercept and regression effects. This was also true for estimation of the variance of the random regression effect. Estimation of the variance of the random intercept, however, was severely biased when the average autocorrelation was .3 or .6. The AR(1) model was less biased than the ID model in these conditions but the amount of bias was still unsatisfactory. The TST model was also biased in the estimation of the variance of the random intercept. In fact, the bias of the TST model was higher than AR(1) model when the number of observations within individuals was 10 or 20. When the number of observation was large (50 or 100), however, the TST model was less biased than the AR(1) model. This was expected before analysis of the data. Because the transformation procedure, actually an estimation of autoregressive parameters for each individual, requires a large number of observations (within individuals), performance of the TS MLM-T depends critically on the number of observations. If this is not the case, the first step of TS MLM-T may fail to identify a valid transformation matrix, resulting in poor estimates in the second step. Once enough number of observations are available and analyzed for each individual, however, TS MLM-T produces better estimates than the other misspecified models under heterogeneous covariance structure. Bias in the estimation of the variance of the random intercept resulted in bias in estimation of the standard error of the fixed intercept. As seen in the results, TS MLM-T may reduce this bias if a large number of observations for each individual (say 50 or more) are available.

4. A Two-Step Multilevel Random Variance Model for Heterogeneous Individual Variance

In the previous chapter, we have seen how heterogeneous covariance structure may affect estimation and statistical inference in multilevel models which contain effects of covariates on a response variable of interest. In such models, the parameters of interest are regression coefficients that predict or explain the change of a response. Characteristics of error structure (e.g., variance or autocorrelation) are generally not of major substantive concern. As such, heterogeneity of error covariance is often ignored, or at least modeled to enhance the estimation of parameters for the mean function of a response, as done by TS MLM-T (introduced in the previous chapter). In some studies with intensive longitudinal data, however, heterogeneity of residual variance (as an index of variability) or autocorrelation (as an index of temporal dependency) across individuals is of interest in its own right and individual level factors (e.g., group membership or impulsivity level) that explain the heterogeneity of variance and autocorrelation (e.g., fluctuations of negative mood state or alcohol use) are also of interest. In this chapter, modeling heterogeneous variance in the context of variance function model is discussed. A new model, suggested by Hedeker et al. (2008), that allows random variation of the variance across individuals is now introduced. In addition, an alternative multilevel model for modeling variance function and random variance is proposed and evaluated by comparison with the model suggested by Hedeker et al. in a simulation study.

Variance Function Models and the Mixed-Effect Location Scale Model

Modeling variance as a function of within-individual or between-individual covariates has been suggested by many researchers (Goldstein, 1995; Hedeker et al., 2008; Hedeker & Mermelstein, 2007; Pinheiro & Bates, 2000; Raudenbush & Bryk, 1987; Raudenbush & Bryk, 2002). A basic form of variance function is suggested in the context of the single level linear regression model (Carroll & Ruppert, 1988; Davidian & Giltinan, 1995; Harvey, 1976). Suppose there is a response variable y_t whose expected value is a function of a vector of covariates $\mathbf{x_t}$ and its corresponding parameter $\boldsymbol{\beta}$. A general specification for variance function g of y_t is

$$Var(y_t) = \sigma^2 g^2(\mu_t, \mathbf{v_t}, \boldsymbol{\theta}), \tag{4.1}$$

where $\mu_t = E(y_t) = f(\mathbf{x_t}, \boldsymbol{\beta})$, $\mathbf{v_t}$ is a vector of covariates predicting the variance of y_t, and $\boldsymbol{\theta}$ is a parameter vector relating μ_t and $\mathbf{v_t}$ to the variance of y_t. Initially, the variance functions were suggested to relate mean and variance for heteroscedastic regression models. Examples of variance function g in this context include the power-of-the-mean model

$$g(\mu_t, \boldsymbol{\theta}) = \mu_t^{\theta}, \tag{4.2}$$

the exponential model

$$g(\mu_t, \boldsymbol{\theta}) = \exp(\mu_t \theta), \tag{4.3}$$

and a two-component model

$$g^2(\mu_t, \boldsymbol{\theta}) = \theta_1 + \mu_t^{2\theta_2}. \tag{4.4}$$

Notice that if $\theta = 1$ in (4.2), σ^2 is the coefficient of variation. Variance of y_t is sometimes thought to depend on a covariate. In this case, an exponential model can be used as

$$g(\mathbf{v_t}, \boldsymbol{\theta}) = \exp(\theta v_t). \tag{4.5}$$

This single level variance function has been extended to the linear mixed models or the MLMs by several researchers (Davidian & Giltinan, 1995; Goldstein, 1995; Hedeker & Mermelstein, 2007; Pinheiro & Bates, 2000; Raudenbush & Bryk, 2002). Assume a two level regression model (2.3) for a variable y_{ti}, rewritten as

$$y_{ti} = f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i}) + e_{ti}, \tag{4.6}$$

where $\mathbf{z_{ti}}$ is a vector of within-individual covariates at time t for individual i, $\mathbf{w_i}$ is a vector of between individual covariates for i, $\boldsymbol{\gamma}$ is a vector of fixed effects and $\mathbf{u_i}$ is a vector of random effects for i. The variance function g of this model is expressed as

$$Var(y_{ti}) = \sigma^2 g^2(\mu_{ti}, \mathbf{v_{ti}}, \boldsymbol{\theta}) \tag{4.7}$$

where $\mu_{ti} = E(y_{ti}) = f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i})$, $\mathbf{v_{ti}}$ is a vector of within- and/or between-individual covariates predicting the variance of y_{ti}, and $\boldsymbol{\theta}$ is a parameter vector relating μ_{ti} and $\mathbf{v_{ti}}$ to the variance of y_{ti}. $\mathbf{v_{ti}}$ may or may not include $\mathbf{z_{ti}}$ and $\mathbf{w_i}$. As in the single level case, variance function g may have the form of either the power-of-the-mean model or the exponential model or a combination of both (Pinheiro & Bates, 2000). A major difference between the variance functions of the single level regression and the multilevel model is that $\mathbf{v_{ti}}$ may include one or more between-individual covariates to explain the heterogeneity of variance in a multilevel variance function but not in a single level variance function. Notice that the parameter $\boldsymbol{\theta}$ in variance function (4.7) is also fixed as in the single level variance function.

The variance function (4.7) can be, however, further extended to a function with random parameters $\boldsymbol{\delta_i}$. Hedeker et al. (2008) suggested one such model, called the mixed-effects location scale model (MLSM). Consider a multilevel model (4.6) where $f(\mathbf{z_{ti}}, \mathbf{w_i}, \gamma, \mathbf{u_i}) = \mathbf{z}'_{\mathbf{ti}}\mathbf{W_i}\boldsymbol{\gamma} + \mathbf{z}'_{\mathbf{ti}}\mathbf{u_i}$, written by

$$y_{ti} = \mathbf{z}'_{\mathbf{ti}}\mathbf{W}_{\mathbf{i}}\boldsymbol{\gamma} + \mathbf{z}'_{\mathbf{ti}}\mathbf{u}_{\mathbf{i}} + e_{ti}\,, \tag{4.8}$$

where $\mathbf{Z}_{\mathbf{i}} = (\mathbf{z}_{\mathbf{1i}}, ..., \mathbf{z}_{\mathbf{n_i i}})'$, and $\mathbf{W}_{\mathbf{i}}$, $\boldsymbol{\gamma}$, and $\mathbf{u}_{\mathbf{i}}$ are defined as in (2.6). Hedeker et al. (2008) suggested a multilevel variance function for the variance of y_{ti}

$$\sigma^2_{e_{ti}} = \exp(\mathbf{v}'_{\mathbf{ti}}\boldsymbol{\theta} + \delta_i)\,, \tag{4.9}$$

where $\mathbf{v}_{\mathbf{ti}}$ is a vector of within- and/or between-individual covariates predicting the variance of y_{ti}, $\boldsymbol{\theta}$ is a vector of fixed effects of $\mathbf{v}_{\mathbf{ti}}$, and δ_i is a random intercept distributed as normal (across individuals) with mean 0 and variance σ^2_{δ}. Taking logs both side of (4.9) yields $\log(\sigma^2_{e_{ti}}) = \mathbf{v}'_{\mathbf{ti}}\boldsymbol{\theta} + \delta_i$, meaning that the logarithm of variance is a linear function of $\mathbf{v}_{\mathbf{ti}}$. Note that the estimated variances of (4.9) are guaranteed to be positive. Because δ_i is normally distributed, the variance follows lognormal distribution across individuals. The location (mean) model (4.8) and the scale (variance) model (4.9) can be jointly written as

$$y_{ti} = \mathbf{z}'_{\mathbf{ti}}\mathbf{W}_{\mathbf{i}}\boldsymbol{\gamma} + \mathbf{z}'_{\mathbf{ti}}\mathbf{u}_{\mathbf{i}} + \exp\left\{\frac{1}{2}(\mathbf{v}'_{\mathbf{ti}}\boldsymbol{\theta} + \delta_i)\right\}\varepsilon_{ti}\,, \tag{4.10}$$

where ε_{ti} is a standard normal. The model (4.10) can be estimated by (marginalized) maximum likelihood (ML) method and the ML estimates can be easily obtained by the NLMIXED procedure in SAS, for example.

Although Hedeker et al. (2008) restricted the MLSM in (4.10) to one random effect, this model can easily be extended to a model with two or more random effects by replacing (4.9) with

$$\sigma^2_{e_{ti}} = \exp(\mathbf{v}'_{\mathbf{ti}}\boldsymbol{\theta} + \mathbf{s}'_{\mathbf{ti}}\boldsymbol{\delta}_{\mathbf{i}})\,, \tag{4.11}$$

where $\boldsymbol{\delta}_{\mathbf{i}}$ is a vector of random effects of $\mathbf{s}_{\mathbf{ti}}$. For example, suppose that the variance is a function of a within-individual covariate a_{ti} and a between-individual covariate b_i.

Suppose also that the intercept and the within-individual covariate have both fixed and random effects. In this case, (4.11) can be written in multilevel equations as

$$\begin{aligned} \sigma^2_{e_{ti}} &= \exp(\lambda_{0i} + \lambda_{1i} a_{ti}) \\ \lambda_{0i} &= \theta_{00} + \theta_{01} b_i + \delta_{0i} \\ \lambda_{1i} &= \theta_{10} + \delta_{1i} \end{aligned} \tag{4.12}$$

and a single equation form is given by

$$\sigma^2_{e_{ti}} = \exp(\theta_{00} + \theta_{01} b_i + \theta_{10} a_{ti} + \delta_{0i} + \delta_{1i} a_{ti})\ . \tag{4.13}$$

Notice that in (4.13) there are two random effects: δ_{0i} and δ_{1i}. Notice also that there is no level-1 random error in this model.

A Two-Step Multilevel Random Variance Model

Carroll and Ruppert (1988) discussed alternative methods of variance function estimation other than introduced above, in the context of a single level regression model. One such method is a regression model where responses are squared residuals and the regression function is the variance function introduced in (4.1). For example, a variance function for a single level regression model can be modeled as

$$[y_t - f(\mathbf{z_t}, \boldsymbol{\gamma})]^2 = \sigma^2 g^2(\mu_t, \mathbf{v_t}, \boldsymbol{\theta}) + \omega_t, \tag{4.14}$$

where $y_t = f(\mathbf{z_t}, \boldsymbol{\gamma}) + e_t$ and $\mu_t = E(y_t) = f(\mathbf{z_t}, \boldsymbol{\gamma})$. This method is based on the fact that the expectation of squared residuals is approximately the variance. As such, fitting (4.14) will provide a good approximation of the variance function g. For normal e_t, (4.14) is a nonlinear model with heteroscedastic variance proportional to $\sigma^4 g^4(\mu_t, \mathbf{v_t}, \boldsymbol{\theta})$. Given $e_t \sim N(0, \sigma^2_e)$, e_t^2 is distributed as a scaled χ^2 with $df = 1$, which is a special case of the gamma distribution. If there is a function h such that $h(\pi_t) = \theta_1\mu_t + \mathbf{v}_t'\boldsymbol{\theta}^*$, $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}^*)$ and $\pi_t = \mathrm{E}(e_t^2) = \sigma^2 g^2(\mu_t, v_t, \boldsymbol{\theta})$, (4.14) can be fitted by a generalized linear model with gamma error distribution and a link function of h.

In multilevel models, the same method is easily applicable. For a multilevel model (4.8), a multilevel random variance model is written by

$$\left[y_{ti} - f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i})\right]^2 = \sigma^2 g^2(\mu_{ti}, \mathbf{v_{ti}}, \mathbf{s_{ti}}, \boldsymbol{\theta}, \boldsymbol{\delta}_\mathbf{i}) + \omega_{ti}, \tag{4.15}$$

where notations are defined as in (4.6), (4.7) or (4.11). The residual e_{ti} is obtained by fitting (4.8) first, and then subtracting the expected (or predicted) value $f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i})$ from the observed value y_{ti}. Because the fitted value $f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i})$ is a function of EBLUP estimates, that is also empirical Bayes estimates, the obtained residuals are empirical Bayes estimates of the true residuals.

As in the single level model, e_{ti}^2 is distributed as a scaled χ^2 with $df = 1$, that is a gamma distribution, and can be fitted by a generalized linear mixed model with gamma distribution. For squared residuals $e_{ti}^2 = \left[y_{ti} - f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i})\right]^2$, we write

$$e_{ti}^2 \mid \alpha_{ti}, \beta_{ti} \sim Gamma(\alpha_{ti}, \beta_{ti}) \tag{4.16}$$

to denote that e_{ti}^2 has a gamma distribution with a shape parameter α_{ti} and a scale parameter β_{ti}, $\alpha_{ti} > 0$ and $\beta_{ti} > 0$. The mean and the variance of (4.16) are $E(e_{ti}^2 | \alpha_{ti}, \beta_{ti}) = \pi_{ti} = \alpha_{ti}\beta_{ti}$ and $Var(e_{ti}^2 | \alpha_{ti}, \beta_{ti}) = \alpha_{ti}\beta_{ti}^2$, respectively. As such, the variance function $\sigma^2 g^2(\mu_{ti}, \mathbf{v_{ti}}, \mathbf{s_{ti}}, \boldsymbol{\theta}, \boldsymbol{\delta}_\mathbf{i})$ in (4.15) is equated to π_{ti}, that is the expected value of e_{ti}^2 or the approximation of the variance. Notice that when e_{ti} is distributed as $N(0,1)$, e_{ti}^2 follows

$Gamma(1/2, 2)$, i.e., a χ^2 distribution with *df* = 1. If e_{ti} is distributed as $N(0, \sigma_i^2)$, e_{ti}^2 follows $Gamma(1/2, 2\sigma_i^2)$.

If there is a function *h* such that $h(\pi_{ti}) = \eta_{ti}$, where η_{ti} is a linear function of μ_{ti}, $\mathbf{v_{ti}}$, and $\mathbf{s_{ti}}$, it can be used as a link function for (4.15). For example, if we define the variance function as (4.11), that is $\sigma_{e_{ti}}^2 = \sigma^2 g^2(\mu_{ti}, \mathbf{v_{ti}}, \mathbf{s_{ti}}, \boldsymbol{\theta}, \boldsymbol{\delta_i}) = \exp(\mathbf{v'_{ti}}\boldsymbol{\theta} + \mathbf{s'_{ti}}\boldsymbol{\delta_i})$, and set $e_{ti} = y_{ti} - f(\mathbf{z_{ti}}, \mathbf{w_i}, \boldsymbol{\gamma}, \mathbf{u_i})$, (4.15) is rewritten as

$$e_{ti}^2 = \exp(\mathbf{v'_{ti}}\boldsymbol{\theta} + \mathbf{s'_{ti}}\boldsymbol{\delta_i}) + \omega_{ti} \tag{4.17}$$

In this case, the log link ensures linear relationship between logarithm of variance and predictors:

$$\eta_{ti} = \log(\pi_{ti}), \tag{4.18}$$

$$\eta_{ti} = \mathbf{v'_{ti}}\boldsymbol{\theta} + \mathbf{s'_{ti}}\boldsymbol{\delta_i} \tag{4.19}$$

where η_{ti} is the log of the expected value of e_{ti}^2.

Random effects $\boldsymbol{\delta_i}$ is assumed to be multivariate normal. For (4.17), a normal distribution of $\boldsymbol{\delta_i}$ results in the lognormal distribution of π_{ti}. (4.15) or (4.17) can also be expressed via multilevel equations. For example, (4.12) can be expressed as (4.17) where level-1 and level-2 equations are written by

$$\begin{aligned} e_{ti}^2 &= \exp(\lambda_{0i} + \lambda_{1i} a_{ti}) + \omega_{ti} \\ \lambda_{0i} &= \theta_{00} + \theta_{01} b_i + \delta_{0i} \\ \lambda_{1i} &= \theta_{10} + \delta_{1i} \end{aligned} \tag{4.20}$$

Where $E(e_{ti}^2) = \pi_{ti} = \alpha_{ti}\beta_{ti}$ and $\begin{pmatrix} \delta_{0i} \\ \delta_{1i} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \varsigma_{00} & \\ \varsigma_{01} & \varsigma_{11} \end{bmatrix} \right)$. The model (4.17), or (4.15) in general, can be estimated by (marginalized) ML and the ML estimates can be easily obtained using the NLMIXED procedure in SAS, for example.

The MLSM of (4.11) and the model (4.17) are almost equivalent to each other: (1) The log of the variance is modeled as a linear function of predictors. (2) Variance is modeled in two levels. (3) Variance is allowed to vary systematically and/or randomly. (4) Variance is modeled to follow a lognormal distribution. Two major differences of the two approaches are: (1) MLSM models an exact function of variance without assuming measurement or sampling errors of variance while (4.17) models variance function in terms of a regression model allowing errors. (2) MLSM fits mean function and variance function simultaneously while (4.17) fits variance function on the squared residuals obtained from a previous analysis. As such, we call the latter approach the Two-Step Multilevel Random Variance Models (TS-MRVM). In the following section the performance of the MLSM and the TS-MRVM will be investigated by a simulation study.

Performance in Estimation of MLMs for Random Variance Function: A Simulation Study

MLSM introduced by Hedeker et al. (2008) is a rather new model of the variance function. Hedeker et al. stated that the MLSM often did not converge with relatively small sample size (e.g., 20 subjects with 5 observations each) in their small simulation study. A plausible problem in the MLSM is complexity of the model in the sense of simultaneous estimation of both mean function and variance function. By contrast, the TS-MRVM may not suffer from this problem because it estimates parameters in mean function and variance function separately. Compensation may occur by producing bias in estimation.

Method

The following models were used to generate data sets. For the mean function, y_{ti} was set to be a function of a time varying covariate z_{ti} and a between-individual covariate w_i. For simplicity, only the intercept was allowed to randomly vary in the model. As such, the mean function MLM was given by

$$y_{ti} = \gamma_{00} + \gamma_{10} z_{ti} + \gamma_{01} w_i + u_{0i} + e_{ti} \; . \tag{4.21}$$

The time varying covariate z_{ti} and a between-individual covariate w_i were generated from a normal distribution with mean 0 and variance 1, independently. The fixed effects were set to a value: γ_{00} = .5, γ_{10} = .5, and γ_{01} = .5. The random intercept u_i was generated from a normal distribution with the mean 0 and the variance σ^2_{u0} = .25. The error e_{ti} was generated from a distribution with the mean 0 and the variance $\sigma^2_{e_{ti}}$, that is modeled by the following function.

The residual variance was modeled as a function of the between-individual covariate w_i. Only the intercept was modeled to randomly vary. Accordingly, the variance function model was given as

$$\sigma^2_{e_{ti}} = \exp(\theta_{00} + \theta_{10} w_i + \delta_{0i}) \tag{4.22}$$

The fixed intercept θ_{00} and θ_{01} were set to .5 and the random effect of δ_{0i} was generated from a normal distribution with mean 0 and the variance $\sigma^2_{\delta 0}$ = .25.

Each data set was completely balanced with L (the number of observations within individuals) = 10, 20, 50, or 100 and N (the number of individuals) = 20, 50, or 100. Accordingly, 3 (L) × 4 (N) = 12 conditions were obtained. In each condition, 500 data sets were simulated, resulting total of 6000 data sets. Each data set was analyzed by the MLSM and the TS-MRVM separately. For the TS-MRVM, a MLM was fitted for the mean function in the first step using the MIXED procedure in SAS. After fitting the mean function, empirical Bayes (EB) estimates of the residuals were obtained. In the second step, a multilevel variance model was fitted on the squared (EB) residuals for the variance function using the NLMIXED procedure in SAS. For the MLSM, both the mean function and the variance function were fitted simultaneously using the NLMIXED procedure in SAS. When fitting the NLMIXED procedure for both methods, true parameter values were provided as starting values. Relative bias was calculated as in chapter 3 to investigate bias in the estimation of parameters in mean model and variance model.

Results

Convergence rate. The TS-MRVM converged in all the data sets in all conditions. As expected, however, the MLSM showed poor performance in convergence as seen in Figure 4.1. Clearly, the coverage rate decreased as the series length increased (Wald χ^2 = 35.00, $p < .001$) or the number of individuals increased (Wald χ^2 = 8.53, $p < .05$). When the number of individuals was 20, convergence rate decreased from 100 % through 96% as the series length increased from 10 through 100. For the sample size of 50, convergence rate declined from 99.2% to 90.2% as the series length raised from 10 to 100. When N = 100, convergence rate decreased from 99% through 82%.

Bias in mean model. The performance of the estimation of parameters in mean function was investigated. Relative bias (*RB*) for the fixed effects γ_{00}, γ_{10}, and γ_{01} and variance of random intercept σ_{u0}^2 are presented in Table 4.1. Because, for TS-MRVM, the mean function was fit by an MLM in which the variance function was not modeled,

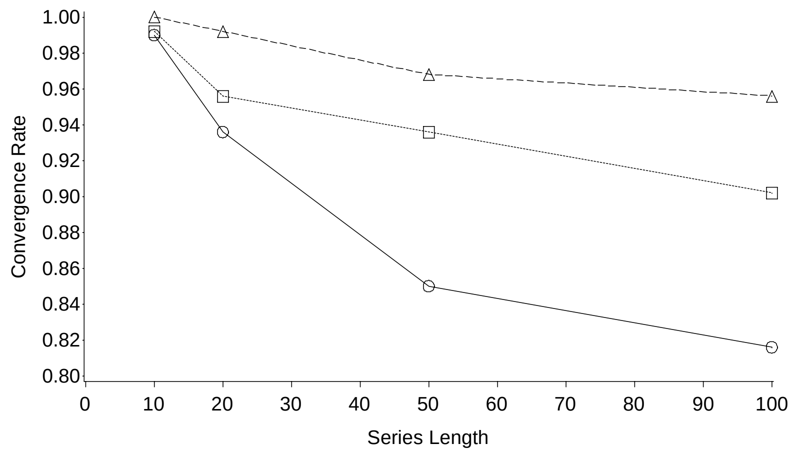


Figure 4.1. Line plots of convergence rate of MLSM across sample size and series length. The square, triangle, and circle represent a convergence rate for sample size of 20, 50, and 100, respectively.

the first step of TS-MRVM is a model with misspecified variance structure. Nevertheless, there was no significant bias of estimates of the fixed effects in the mean function for the first step of TS-MRVM: *RB* = 0.00, $t_{(5999)} = -0.33$, $p = .74$ for γ_{00}; *RB* = 0.00, $t_{(5999)} = 0.33$, $p = .74$ for γ_{10}; *RB* = 0.00, $t_{(5999)} = 0.89$, $p = .37$ for γ_{01}. Likewise, no significant bias was found in estimates of the fixed effects in the mean function for MLSM: *RB* = 0.00, $t_{(5646)} = -0.37$, $p = .71$ for γ_{00}; *RB* = 0.00, $t_{(5646)} = -0.23$, $p = .82$ for γ_{10}; *RB* = 0.00, $t_{(5646)} = 1.10$, $p = .27$ for γ_{01}. Neither sample size nor series length has significant effects on the relative bias of the fixed estimates in both models. On the other hand, there was a significant bias in the estimation of the variance of random intercept in the MLSM, *RB* = -0.02, $t_{(5646)} = -16.28$, $p < .0001$. The bias decreased as sample size increased, $F_{(2,5635)} = 42.07$, $p < .0001$, $\eta^2 = .01$. In contrast, the misspecified MLM (i.e., the first step of TS-MRVM) did not produce biased estimates of σ^2_{u0}, *RB* = 0.00, $t_{(5999)} = 1.74$, $p = .08$. The unbiased estimates of the parameters in the mean function in the misspecified MLM support use of the empirical Bayes residuals obtained the first step of TS-MRVM in the second step.

Table 4.1

Relative Bias of $\hat{\gamma}_{00}$, $\hat{\gamma}_{10}$, $\hat{\gamma}_{10}$, *and* $\hat{\sigma}^2_{u0}$ *for the Two Random Variance MLMs*

		$\hat{\gamma}_{00}$		$\hat{\gamma}_{10}$		$\hat{\gamma}_{10}$		$\hat{\sigma}^2_{u0}$	
N	L	MLSM	MRVM	MLSM	MRVM	MLSM	MRVM	MLSM	MRVM
20	10	0.00	0.00	0.00	0.00	0.00	0.00	-0.04	0.01
	20	0.00	-0.01	0.00	0.00	0.01	0.01	-0.03	0.00
	50	-0.01	-0.01	0.00	0.00	-0.01	-0.01	-0.03	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	-0.03	0.00
50	10	0.01	0.01	0.00	0.00	0.01	0.01	-0.01	0.00
	20	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	-0.01
	50	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00
100	10	0.00	0.00	0.00	0.00	0.01	0.01	-0.01	0.00
	20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	50	-0.01	0.00	0.00	0.00	0.00	0.00	-0.01	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00

Note. MLSM: Mixed-effects Location Scale Model; MRVM: Two-Step Multilevel Random Variance Model; N = sample size; L = series length.

Bias in variance model. Unlike estimation of the parameters in the mean function, the TS-MRVM produced biased estimates for the parameters in the variance function (Table 4.2). The TS-MRVM produced bias in the estimates of the fixed intercept $\hat{\theta}_{00}$, *RB* = -0.02, $t_{(5999)}$ = -18.97, p < .0001, and $\hat{\theta}_{10}$, *RB* = -0.01, $t_{(5999)}$ = -6.77, p < .0001. The bias was the function of series length in the both parameters, $F_{(2,5988)}$ = 27.73, p < .0001, η^2 = .01 for $\hat{\theta}_{00}$, and $F_{(2,5988)}$ = 30.22, p < .0001, η^2 = .01 for $\hat{\theta}_{10}$. By contrast, the MLSM did not show any bias in the estimation of $\hat{\theta}_{00}$, *RB* = -0.00, $t_{(5646)}$ = -0.54, p = .58, and $\hat{\theta}_{10}$, *RB* = 0.00, $t_{(5646)}$ = 0.45, p = .65. For the estimation of the variance of random intercept (i.e., $\sigma^2_{\delta 0}$), both the MLSM model and the TS-MRVM showed significant bias: *RB* = -0.02, $t_{(5646)}$ = -19.03, p < .0001 for MLSM, and *RB* = -0.03, $t_{(5999)}$ = -26.76, p < .0001 for TS-MRVM. For the MLSM, the bias was a function of sample size, $F_{(2,5635)}$ = 55.69, p <

Table 4.2

Relative Bias of $\hat{\theta}_{00}$, $\hat{\theta}_{10}$, and $\hat{\sigma}^2_{\delta 0}$ for the Two Random Variance MLMs

		$\hat{\theta}_{00}$		$\hat{\theta}_{10}$		$\hat{\sigma}^2_{\delta 0}$	
N	L	MLSM	MVM	MLSM	MVM	MLSM	MVM
20	10	0.00	-0.04	-0.01	-0.04	-0.06	-0.07
	20	0.01	-0.02	0.00	-0.01	-0.03	-0.04
	50	-0.01	-0.02	0.01	0.00	-0.02	-0.03
	100	0.00	-0.01	0.00	0.00	-0.03	-0.03
50	10	0.00	-0.04	0.00	-0.03	-0.02	-0.04
	20	0.00	-0.03	0.00	-0.01	-0.02	-0.02
	50	0.00	-0.02	-0.01	-0.01	-0.01	-0.01
	100	0.00	-0.01	0.01	0.01	-0.02	-0.02
100	10	0.00	-0.04	0.01	-0.01	-0.01	-0.03
	20	0.00	-0.03	0.00	-0.01	-0.01	-0.01
	50	0.00	-0.02	0.00	0.00	-0.01	-0.01
	100	0.00	-0.01	0.00	0.00	-0.01	-0.01

Note. MLSM: Mixed-effects Location Scale Model; TS-MRVM: Two-Step Multilevel Random Variance Model; N = sample size; L = series length.

.0001, η^2 = .02, series length, $F_{(2,5635)}$ = 10.99, p < .0001, η^2 = .01, and the interaction of the two, $F_{(2,5635)}$ = 3.13, p < .01, η^2 = .00. Similarly, bias in the TS-MRVM was also the function of sample size, $F_{(2,5988)}$ = 83.19, p < .0001, η^2 = .03, series length, $F_{(2,5988)}$ = 56.56, p < .0001, η^2 = .03, and the interaction of the two, $F_{(2,5988)}$ = 3.80, p < .001, η^2 = .00. Figure 4.2 presents line plots of relative bias for the parameters in the variance function across series length, the major effect in each case. Clearly, the biases produced in the TS-MRVM converged to 0 as series length increased.

Conclusions

The two multilevel modeling approaches for random variance function were introduced and their estimations were evaluated by a simulation study. The Mixed-effects Location

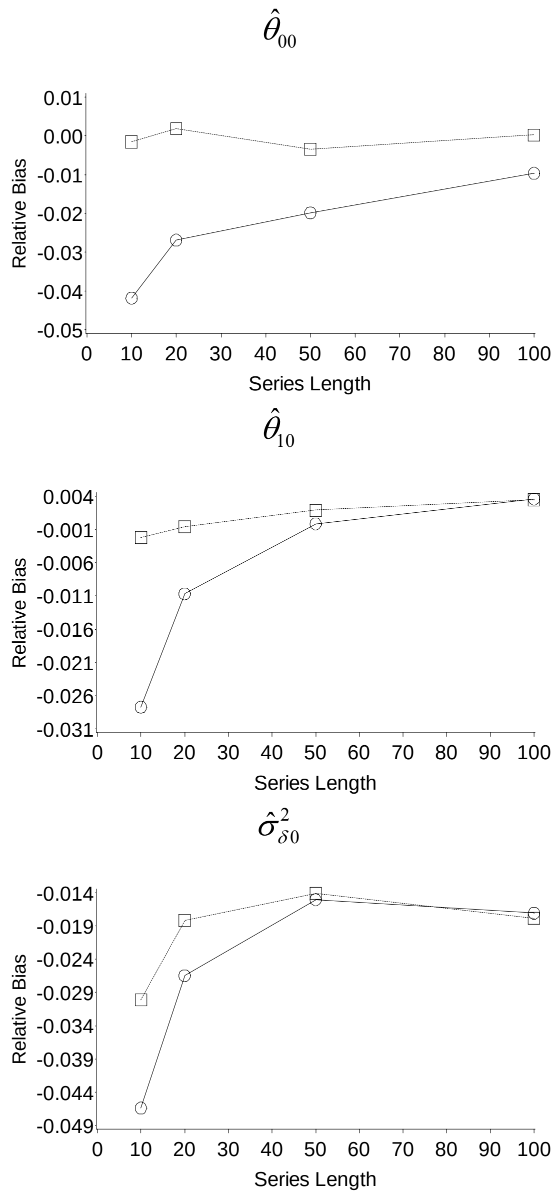


Figure 4.2. Line plots of relative bias of $\hat{\theta}_{00}$, $\hat{\theta}_{10}$, *and* $\hat{\sigma}^2_{\delta 0}$ produced by two models across series lengths. The square and circle represent a relative bias for the MLSM and the TS-MRVM, respectively.

Scale Model (MLSM) proposed by Hedeker et al. (2008) is a useful model in that it models variance as a function of within-individual and between-individual covariates. It also allows variance to vary randomly across individuals as well as systematically (as a function of predictors). The MLSM models both multilevel mean function and multilevel variance function simultaneously. This feature may produce more accurate estimation on the parameters in the model because the estimations of parameters in both functions are connected to each other. At the same time, the complexity of the simultaneous estimation of parameters in both functions may raise problems in optimization, such as convergence. The result of the simulation study showed this problem found in the MLSM. The convergence rate was very unsatisfactory when we consider that the true parameter values were given as starting values and increased series length and sample size made convergence rate even worse.

On the other hand, the Two-Step Multilevel Random Variance Model (TS-MVM) is an alternative MLSM. Although TS-MRVM estimates mean function and variance function of the model separately, which may suffer from inaccurate estimation, the result of the present simulation study showed no bias in estimation of the parameters in the mean function. In addition, the bias found in the estimation of the parameters of the variance function, obtained by the TS-MRVM, decreased as series length increased, meaning that the TS-MRVM produces asymptotically unbiased estimates.

One limitation of TS-MRVM is that it can not model covariance (or correlation) between a random effect in the mean function and a random effect in the variance function, because the estimation of each function was separated. It is also noted that the current study did not investigate the bias in the estimate of standard error. However, given that the estimates of variance of random effects were not much biased, especially when the series length is 50 or more, it is not highly suspected.

5. Discussion

In this paper, we discussed issues related to the heterogeneity of residual covariance in analysis of intensive longitudinal data. In chapter 3, we found that if homogeneous covariance is incorrectly assumed, MLMs produce highly biased estimates of variance of the random intercept when the average autocorrelation is high. It is also found that biased estimates of random intercept variance also create biased estimates of the standard error of the fixed effects. For intensive longitudinal data (e.g., 50 or more observations for each individual), we saw that application of MLMs to variables, transformed by the inverse of Cholesky factor of individual specific residual covariance, can be used to reduce the bias. We also found that squared residuals can be successfully modeled within multilevel modeling framework to estimate variance function. Although the residual-based approach produced biased estimates when the number of observations within individuals was small to moderate, it is applicable to intensive longitudinal data without much concern. More importantly, this approach does not suffer non-convergence problem on which a flexible alternative (i.e., the mixed-effects location scale model) has much concern. Because the concern of the alternative increases as the number of observations increases, the suggested model is more appropriate for ILD than longitudinal data with a small to moderate number of observations.

When researchers are interested in the estimation of fixed effects but not in the variance of random effects, there are alternative ways to optimally estimate the fixed effects in ILD. For example, one can fit a multilevel model with unstructured error covariance matrix without modeling covariance of random effects. Unstructured covariance structures in which every element is freely estimated from the data may represent a complicated correlational pattern among occasions. This approach, however, has several limitations when applied to ILD. First, this approach also assumes homogenous covariance across individuals and estimates each element in covariance matrix by pooling observations at each occasion across individuals. As such, although the number of parameters in unstructured covariance matrix (i.e., $n(n+1)/2$, where n is the number of occasions) is large in ILD, the pooled covariance matrix may not well represent a simple covariance structure (e.g., AR(1)) if it differs across individuals. In addition, unstructured covariance for ILD has too many parameters to be estimated, because the total number of parameters in the covariance matrix depends on the number of occasions. For example, if the number of observations within each individual is 100, the number of parameters in the unstructured covariance is 5050. If this is the case, it is likely to suffer from under-identification, non-convergence and/or improper solutions. It also requires a large number of individuals relative to the number of occasions, which is not feasible in some ILD studies.

Another method for optimal estimation of fixed effects for ILD is generalized estimating equations (GEE) (Liang & Zeger, 1986). GEE is a generalization of quasi-likelihood and robust estimation for longitudinal data. It is well known that robust or sandwich estimation provides an asymptotically unbiased estimate of the covariance of the estimates of fixed parameters even when constant variance across individuals is suspected and the pattern of heteroscedasticity is unknown. If correlations among observations at different occasions, in addition to the heteroscedasticity, are expected in longitudinal data, then GEE can be applied to obtain an accurate estimate of the covariance of the parameter estimates of fixed effects. GEE does not assume homogenous error covariance structure and can be applied to non-normally distributed variables, such as a binary or count variable. GEE approach also has limitations, however, including need for a large number of individuals to achieve its asymptotical properties and a restricted number of autocorrelation patterns available as well as lack of information of interindividual variability of random effects (Schafer, 2006).

On the other hand, Bayesian approach can be used to model multilevel random variance. In general, Bayesian statistical models find a posterior probability distribution of parameters given data by incorporating assumed prior distributions of parameters and likelihood function of data using Bayes' theorem. The assumed hierarchical structures commonly used in Bayesian modeling well suit to hierarchical data and multilevel models, and thus the application of Bayesian multilevel regression models is natural and common (Gelman, Carlin, Stern, & Rubin, 2004; Gelman & Hill, 2007). Although a basic application of Bayesian multilevel regression models assumes distribution of random variance across replications not across individuals, extension to models with interindividual-specific random variance is also possible. This approach is especially useful when the number of individuals is small in ILD because ML or REML estimation of the variance of random effects, based on asymptotic theory, assumes a large number of individuals and often obtain unreliable solution from small samples while Bayesian approach may avoid unreliability by giving weight more on prior distribution than on the likelihood of data when sample size is small. A major difficulty of applications of Bayesian approach is unfamiliarity with Bayesian statistical models and software programs to most of social scientists. If sample size and the number of observations within individuals are large enough, differences between likelihood-based approach and Bayesian approach is minimal. Therefore, if ILD with enough sample size are modeled, the benefits of Bayesian approaches, including random effects models and random variance models, are not impressive.

Several limitations of the present studies are acknowledged. First of all, because both suggested procedures (MLMs for transformed variables or squared residuals) consist of two steps and as such suffer from problems common to any two-step approach. The transformation method requires estimation of autocorrelation or autoregressive parameters of individual time series in the first step and application of the intended MLM on the transformed variables in the second step. Multilevel modeling of squared residuals, as a multilevel variance function model, in the second step requires estimation of valid residuals in the first step. As such, a major limitation of the suggested models is high dependency of the performance in the first step analysis. If a poor transformation matrix or invalid residuals are obtained in the first step, the final intended MLMs will produce biased estimation of the parameters of interest. For intensive longitudinal data, however, this concern is not critical because performance of the first step analysis gets better as the number of observations within individuals increases. However, for longitudinal data with small to moderate number of observations within individuals, the suggested two-step approaches produce biased estimation and should be limited to use. An iterative method alternating the two steps may converge to a better solution.

Results of the simulation studies conducted in chapter 3 and chapter 4 are restricted to generalization in several ways. First, the models used to generate data did not vary widely. For example, the parameters in true model were set to one value for each in both studies (e.g., γ_{00} =1, γ_{10} =1, σ_{u0}^2 = .5, σ_{u1}^2 = .5, and σ_{u0u1} = .15 in chapter 3). In addition, conditions used were not comprehensive. We only used sample size of 20 through 100 and series length of 10 through 100 in the both simulation studies. We did not investigate the effect of negative autocorrelations in chapter 3. Observations in all conditions are balanced and equally spaced temporally. As such, all the results found in the studies need cautions for generalization to other situations.

Another limitation is related to the assessed time intervals between successive measurements. The error covariance structures used in the transformation method in chapter 3 assume constant and equally spaced time of measurements. Because ILD are often measured at randomly prompted times (e.g., within-day random assessments of electronic diary), transformation methods introduced in chapter 3 can not be applied directly to ILD with random time intervals.

The suggested transformation method should be extended to ILD with random time intervals. In such cases, heterogeneous covariance with autocorrelation that exponentially decreases over random time intervals may be modeled and estimated by individual and transformation of original variables by multiplying the inverse of the Cholesky factor of the estimated covariance matrix can be applied to get a valid estimation of fixed effects and the variance of random effects. A simulation study can be conducted to evaluate the performance of the transformation method on ILD with random time intervals and heterogeneous autocorrelation.

Multivariate analysis, especially using Latent Variable Models (LVMs), is another immediate extension of heterogeneous within-individual covariance structure. For example, multivariate intraindividual covariance has been modeled using Dynamic

Factor Analysis (DFA) that applies traditional Confirmatory Factor Analysis to multivariate data with time series structure (Browne & Nesselroade, 2005; Molenaar, 1985; Wood & Brown, 1994). On the other hand, LVMs have been generalized to multivariate data with multilevel structure, called multilevel LVMs or multilevel Structural Equation Models (Goldstein & Browne, 2002; Goldstein & McDonald, 1987; Mehta & Neale, 2005; Rabe-Hesketh, Skrondal & Pickles, 2004). This line of research is based on the recognition that, with some restrictions, a univariate multilevel model or linear mixed model can be expressed in a multivariate latent variable model (Bauer, 2003; Curran, 2003; MacCallum, Kim, Malarkey, & Keicolt-Glaser, 1997; Meredith & Tisak, 1990; Rovine & Molenaar, 2000). Because LVMs are about modeling covariance and multilevel approach aims to modeling heterogeneity, heterogeneity of covariance structure is of direct interest in multilevel LVMs. Extending DFA to multilevel data structure, however, has not been fully developed in this context, because estimation of the multilevel time series structure is difficult to be solved by maximum likelihood based method commonly used in estimation of SEMs.

Interestingly, it is known that state space models have great flexibility in analysis of intensive longitudinal data. It can be thought of not only as a generalization of time series models but also as a generalization of other approaches. For example, it is well known that a linear mixed model has its state-space form (for two-level model). In addition, it is also known that DFA has its state-space form (Ho, Shumway, & Ombao, 2006). This means that state space models can be used as a general model for the analysis of multi-dimensional hierarchical data structure, especially in intensive longitudinal study. Although state-space forms for more complicated models of other approaches (e.g., three-level MLM) and the combinations of more than two approaches (e.g., multivariate longitudinal MLM) are not fully investigated yet, the application of state space models on intensive longitudinal study with high dimensional data is a promising topic of study in ILD analysis.

Intensive longitudinal data are useful to investigate various patterns of intraindividual processes. Although we restricted our discussion of analysis of ILD to the multilevel models for mean function and variance function, there are still other possibilities of modeling intraindividual processes using ILD. For example, heterogeneity of autocorrelation can be modeled in MLMs. In this regard, an interesting extension of MLM has been suggested by Rovine and Walls (2006). Rovine and Walls showed that the autoregressive parameters can be modeled as predictors in MLM, which allow estimation of individual specific autoregressive parameters as a random effect, as well as higher order AR(p) process. Jahng, Wood, and Trull (2008) suggested a multilevel random instability model for EMA data where instability is equated as frequent and extreme fluctuations over time and expressed as a function of variance and autocorrelation. Other possibilities for modeling heterogeneous intraindividual process include time varying regression effects (Fan, & Gijbels, 1996; Li, Root, & Shiffman, 2006), nonlinear multilevel models (Davidian & Giltinan, 1995; Fok & Ramsay, 2006), state space models (Durbin & Koopman, 2001; Ho, Shumway, & Ombao, 2006), and differential equation models (Boker, 2001; Boker & Laurenceau, 2006; Ramsay, 2006).

In summary, intensive longitudinal data enable researchers to fully investigate the time dimension and the person dimension as well as other dimensions. Using ILD, more detailed investigation on the intraindividual dynamic process is available. Understanding time series process plays a key role in the investigation of intraindividual dynamics. In addition, unlike a result obtained from single time series data, heterogeneity of idiographic dynamic process likely existing in ILD can be modeled to provide more general descriptions of the processes using multilevel models. Moreover, the heterogeneity of intraindividual dynamic process is not restricted to the mean function but extended to other characteristics, such as variance and autocorrelation. This heterogeneity of variance and autocorrelation can be of direct interest or be a factor influencing estimation of mean change. In both cases, applications of MLM for the analysis of ILD require extensions and modifications of existing models. The transformation procedure of correction for autocorrelated error and multilevel random variance models on squared residuals are such an extension or modification which seems to be appropriately applied to intensive longitudinal data.

Due to recent developments in data collection methods (e.g., electronic diaries) and statistical models (e.g., multilevel models), intensive longitudinal studies and the analysis of ILD are gaining popularity across many areas of psychology. As such, development of new methods and evaluation and proper applications of existing methods for the analysis of ILD become more important than ever. Studies on statistical models for intensive longitudinal data such as this will provide better tools to understand intraindividual change and interindividual difference of psychological phenomena and even stimulate interesting studies that should be assisted by the development and proper use of the quantitative methods.

REFERENCES

Bauer, D. J. (2003). Estimating multilevel linear models as structural models. *Journal of Educational and Behavioral Statistics, 28,* 135–167.

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.

Boker, S. M. (2001). Differential structural modeling of intraindividual variability. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 3-28). Washington, DC: American Psychological Association.

Boker, S. M., & Laurenceau, J. (2006). Dynamical systems modeling: An application to the regulation of intimacy and disclosure in marriage. In T. A. Walls & J. L. Schafer (Eds.). *Models for intensive longitudinal data* (pp. 195-216). New York: Oxford University Press.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics*. Mahwah, NJ: Erlbaum.

Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression.* London: Chapman and Hall.

Cattell, R. B. (1946). *Description and measurement of personality*. New York: World Book Company.

Cattell, R. B. (1988). The data box: Its ordering of total resources in terms of possible relational systems. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of Multivariate Experimental Psychology* (pp. 69-130). New York: Plenum Press.

Chipman, J. S. (1979). Efficiency of least squares estimation of linear trend when residuals are autocorrelated. *Econometrica, 47,* 115-128.

Cochrane, D. & Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association, 44,* 32-61.

Cowdry, R. W., Gardner, D. L., O'Leary, K. M., Leibenluft, E., & Rubinow, D. R. (1991). Mood variability: A study of four groups. *American Journal of Psychiatry*, *148*, 1505-1511.

Cowdry, R. W., Gardner, D. L., O'Leary, K. M., Leibenluft, E., & Rubinow, D. R. (1991). Mood variability: A study of four groups. *American Journal of Psychiatry*, *148*, 1505-1511.

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

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data.* New York: Chapman & Hall.

Durbin, J., & Koopman, S. J. (2001). Time series analysis by state-space models. New York: Oxford University Press.

Ebner-Priemer, U. W., Kuo, J., Kleindienst, N., Welch, S. S., Reisch, T., Reinhard, I., Lieb, K., Linehan, M. M., & Bohus, M. (2007). State affective instability in borderline personality disorder assessed by ambulatory monitoring. *Psychological Medicine, 37,* 1-10.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76,* 662-676.

Fan, J. & Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman & Hall.

Farmer, R. F., Nash, H. M. & Dance, D. (2004). Mood patterns and variations associated with personality disorder pathology. *Comprehensive Psychiatry, 45,* 289–303.

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.

Fok, C. C. T., & Ramsay, J. O. (2006). Fitting curves with periodic and nonperiodic trends and their interactions with intensive longitudinal data. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 109-123). New York: Oxford University Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, Florida: Chapman & Hall/CRC.

Goldfeld, S. M., & Quandt, R. E. (1972). *Nonlinear methods in econometrics*. Amsterdam: North-Holland.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73,* 43-56.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). New York: John Wiley.

Goldstein, H., & Browne, W. J. (2002). Multilevel factor analysis modeling using Markov Chain Monte Carlo (MCMC) estimation. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 225-243). Hillsdale, NJ: Erlbaum.

Goldstein, H., & McDonald, R. P. (1987). A general model for the analysis of multilevel data. *Psychometrika, 53,* 455-467.

Gorman, B. S., & Wessman, A. E. (1974). The relationship of cognitive styles and moods. *Journal of Clinical Psychology, 30,* 18-25.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrics, 44,* 461-465.

Harvey, A. C. (1981). *The econometric analysis of time series*. New York: John Wiley & Sons.

Harvey, A. C., & Phillips, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika, 66,* 49-58.

Hedeker, D., & Mermelstein, R. J. (2007). Mixed-effects regression models with heterogeneous variance: Analyzing ecological momentary assessment (EMA) data. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 183-206). Mahwah, NJ: Erlbaum.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics, 64,* 627-634.

Henderson, C. R. (1984), *Applications of linear models in animal breeding*, Ontario, Canada: University of Guelph.

Hildreth, C., & Houck, J. P. (1968). Some estimates for a linear model with random coefficients. *Journal of the American Statistical Association, 63,* 584-595.

Ho, M. R., & Shumway, R., & Ombao, H. (2006). The state-space approach to modeling dynamic processes. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 148-175). New York: Oxford University Press.

Hoffman, B. U., & Meyer, T. M. (2006). Mood fluctuations in people putatively at risk for bipolar disorders. *British Journal of Clinical Psychology, 45,* 105–110.

Hufford, M. R., Shiffman, S., Paty, J., & Stone, A. A. (2001). Ecological momentary assessment: real-world, real-time measurement of patient experience. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment: Computer-assisted psychological and psych physiological methods in monitoring and field studies* (pp. 69-92). Kirkland, WA: Hogrefe & Huber.

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J., & Thiébaut, R. (2007). Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics and Data Analysis, 51*, 5142-5154.

Jahng, S., Wood, P. K., & Trull, T. J. (in press). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*.

Jobson, J. D., & Fuller, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *Journal of the American Statistical Association, 75,* 176-181.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics, 22,* 389-395.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: John Wiley & Sons.

Koreisha, S. G., & Fang, Y. (2001). Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society. Series B, 63,* 515-531.

Kwok, O., 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.

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

Lange, N., & Laird, N. (1989). The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Journal of the American Statistical Association, 84,* 241-247

Larson, R. W., Csikszentmihalyi. M., & Graef. R. (1980). Mood variability and the psychosocial adjustment of adolescents. *Journal of Youth and Adolescence, 9,* 469-489.

Li, R., Root, T. L., & Shiffman, S. (2006). A local linear estimation procedure for functional multilevel modeling. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 63-83). New York: Oxford University Press.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13-22.

Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University Press.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Keicolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32,* 215–253.

Maeshiro, A. (1980). Small sample properties of estimators of distributed lag models. *International Economic Review, 21,* 721-733.

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

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

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50,* 181-202.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2,* 201-218.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of inter-individual and intraindividual variability: A theoretical and empirical vindication of Developmental Systems Theory. In U. M.

Staudinger & U. Lindenberger (Eds.), *Understanding human development* (pp. 339-360). Dordrecht, The Netherlands: Kluwer.

Ozer, D. J. (1986). *Consistency in personality: A methodological framework*. Berlin: Springer-Verlag.

Park, R. E., & Mitchell, B. M. (1980). Estimating the autocorrelated error model with trended data. *Journal of Econometrics, 13,* 185-201.

Penner, L. A., Shiffman, S., Paty, J. A., & Fritzsche, B. A. (1994). Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology, 66,* 712-721.

Pinheiro, J. C., & Bates, M. D. (2000). *Mixed-effects models in S and S-plus*. New York: Springer.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69,* 167-190.

Ramsay, J. O. (2006). The control of behavioral input/output systems. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 176-194). New York: Oxford University Press.

Rao, J. N. K. (2003). *Small area estimation.* Hoboken, NJ: John Wiley & Sons.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Bryk, A. S. (1987). Examining correlates of diversity. *Journal of Educational Statistics*, *12*, 241-269.

Rovine, M. J. & Walls, T. A. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. In T. A. Walls & J. L.

Schafer (Eds), *Models for intensive longitudinal data* (pp. 124-147). New York: Oxford University Press.

Rovine, M. J., & Molenaar, P. C. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Schafer, J. L. (2006). Marginal modeling of intensive longitudinal data by generalized estimating equations. In T. A. Walls & J. L. Schafer (Eds), *Models for intensive longitudinal data* (pp. 38-62). New York: Oxford University Press.

Schwartz, J. E., & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In A. Stone, S. Shiffman, A. Arienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press.

Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications* (2nd ed.). New York: Springer.

Stein, K. F. (1996). Affect instability in adults with a borderline personality disorder. *Archives of Psychiatric Nursing, 10,* 32–40.

Stone, A.A. & Shiffman, S. (1994). Ecological momentary assessment in behavioral medicine. *Annals of Behavioral Medicine, 16*, 199-202.

Walls, T. A., & Schafer, J. L. (Eds.), (2006). *Models for intensive longitudinal data*. New York: Oxford University Press.

Walls, T. A., Höppner, B. B., & Goodwin, M. S. (2007). Statistical issues in intensive longitudinal data analysis. In A. Stone, S. Shiffman, A. Arienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 338-360). New York: Oxford University Press.

Watson, G. S. (1955). Serial correlation in regression analysis. I. *Biometrika, 42,* 327-341.

Wood, P. K., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin, 116,* 166-186.

Woyshville, M. J., Lackamp, J. M., Eisengart, J. A. & Gilliland, J. A. (1999). On the meaning and measurement of affective instability: Clues from chaos theory. *Biological Psychiatry, 45,* 261–269.

Zeigler-Hill, B., & Abraham, J. (2006). Borderline personality features: Instability of self-esteem and affect. *Journal of Social and Clinical Psychology, 25,* 668-687.

VITA

Seungmin Jahng was born in Seoul, South Korea in 1974. He has earned his B.A. in psychology in 1998 and M.A. in personality psychology in 2000 at Seoul National University. During his master's training at Seoul National University, he studied on human emotion and wrote a thesis on the effect of relational motivation on anger experience and expression. After Jahng decided to develop his career in quantitative methods, he started his doctoral training in quantitative psychology at the Department of Psychological Sciences in University of Missouri in 2003. During his doctoral training, he has earned master degrees in quantitative psychology in 2007 and statistics in 2008, and finally completed doctoral degree in quantitative psychology in 2008. He developed his expertise in multilevel models and structural equation models, especially for the analysis of longitudinal data. His current research interest focuses on statistical models on the analysis of intensive longitudinal data as extensively studied in this dissertation.