Nonlinear Change Models in Populations with Unobserved Heterogeneity

Ken Kelley

Department of Management, University of Notre Dame, Notre Dame, IN, USA

Abstract. When unobserved heterogeneity exists in populations where the phenomenon of interest is governed by a functional form of change linear in its parameters, the growth mixture model (GMM) is useful for modeling change conditional on latent class. However, when the functional form of interest is nonlinear in its parameters, the GMM is not very useful because it is based on a system of equations linear in its parameters. The nonlinear change mixture model (NCMM) is proposed, which explicitly addresses unobserved heterogeneity in situations where change follows a nonlinear functional form. Due to the integration of nonlinear multilevel models and finite mixture models, neither of which generally have closed form solutions, analytic solutions do not generally exist for the NCMM. Five methods of parameter estimation are developed and evaluated with a comprehensive Monte Carlo simulation study. The simulation showed that the parameters of the NCMM can be accurately estimated with several of the proposed methods, and that the method of choice depends on the precise question of interest.

Keywords: longitudinal data analysis, analysis of change, growth modeling, growth mixture modeling, heterogeneous population, heterogeneous change, nonlinear growth models, nonlinear change models, functional form of growth, functional form of change

In the behavioral, education, and social sciences, idiographic conceptualizations of change tend to focus on the individual, whereas nomothetic conceptualizations tend to focus on the group (e.g., see Allport, 1937). Muthén and Muthén (2000) discuss that, at least in general, statistical methods designed to help researchers answer questions about behavior are either variable-centered *or* pattern-centered. Statistical methods that are variable-centered focus more on relationships among variables (e.g., structural equation models and its variants), whereas statistical methods that are pattern-centered focus more on the relationships among the individuals (e.g., finite mixture models and its variants). Variable-centered methods are consistent with the "strong concept of growth," because this view states that "a single developmental function can adequately describe the change of all individuals from some population" (Burchinal & Appelbaum, 1991, p. 25), whereas pattern-centered approaches are consistent with the "weak concept of development," because this view seeks to identify intraindividual patterns of change and interindividual differences in those patterns (Burchinal & Appelbaum, 1991; Nesselroade & Baltes, 1979). Idiographic and nomothetic conceptualizations of behavior are traditionally considered antithetical (Dunn, 1994, p. 377) and as such potential benefits of combining the approaches in an integrated fashion often go unrecognized. Some research questions, however, demand an integration of variable-centered and pattern-centered statistical techniques (e.g., Burchinal & Appelbaum, 1991; Dumenci & Windle, 2001;

Magnusson & Bergmen, 1988; Muthén & Muthén, 2000; Muthén et al., 2002; Nagin, 1999; Schulenberg, O'Malley, Bachman, Wadsworth, & Johnston, 1996).

A fully integrated idiographic and nomothetic theory requires both variable-centered and pattern-centered methods in a unified model. Such an integrated method would allow, for example, unknown groupings of individuals to exist in a single heterogeneous population, where variation exists within each group around the group specific parameters. When latent classes of individuals exist in a heterogeneous population, standard methods of longitudinal data analysis may not yield meaningful results. The reflexive application of nomothetic techniques to mixed nomothetic-idiographic situations may not accurately reflect any of the component distributions (i.e., subpopulations) that are mixed together, which potentially results in misspecified models and misleading conclusions. The questions of interest regarding change would best be applied within each of the classes (i.e., conditional on class membership) and not across all of the classes simultaneously.

As discussed by Muthén, "data are frequently analyzed as if they were obtained from a single population, although it is often unlikely that all individuals in our sample have the same set of parameter values" (1989, p. 558). Attempting to solve this deficiency for models that assume homogeneous populations in the context of longitudinal data analysis, the growth mixture model (GMM) was developed (Muthén, 2002; Muthén & Shedden, 1999; Muthén, 2001a; Muthén, 2001b; Muthén & Muthén, 2000). The idea of the

GMM is that individuals are considered nested within a latent (i.e., unknown) class with the latent classes potentially having variability around the class specific population coefficients of change. The GMM model is analogous to a multiple group analysis, but group membership is unknown and must be estimated. Although Muthén's GMM is capable of combining variable-centered and pattern-centered approaches to studying change (Muthén & Muthén, 2000), a major limitation is that it is built on a linear system of equations, where fixed and unique effects cannot both enter a change model nonlinear in its parameters.

An increased appreciation of models of change nonlinear in their parameters is beginning to gain ground in the behavioral, educational, and social sciences (e.g., see Cudeck & Harring, 2007, for a recent review; see also Browne & du Toit, 1991; Burchinal & Appelbaum, 1991; Cudeck, 1996; van Geert, 1991). Models nonlinear in their parameters (e.g., negative exponential, logistic, Gompertz, etc.) can better model some types of change when compared to models linear in their parameters (e.g., straight line, quadratic, cubic, etc.). Nonlinear models of change are models where parameters enter the mathematical function defining the trajectories in a nonlinear fashion. As has been delineated elsewhere (e.g., Browne & du Toit, 1991; Cudeck, 1996; Davidian & Giltinan, 1995; Pinheiro & Bates, 2000; Thissen & Bock, 1990), some advantages of nonlinear change models are that they need not exhibit unlimited growth or decay, oftentimes require fewer parameters than do linear models for change that is asymptotic and/or sigmoidal ("S-shaped"), and the parameters of nonlinear models oftentimes have "real-world" interpretations. Perhaps most importantly, nonlinear models many times conform to a researchers theory about the phenomenon as it changes over time (e.g., learning curves, asymptotic values, points-of-inflection, etc.).

This article attempts to fully integrate the idiographic approach and the nomothetic approach to the analysis of change in the context of nonlinear models of change. In so doing, variable-centered and pattern-centered techniques are combined in an effective and unified way, where the focus is not on either a completely homogeneous or a completely heterogeneous population, but rather on some combination of the two. More specifically, the goal of the present article is to develop and examine the effectiveness of the *nonlinear change mixture model* (NCMM), which is a general model of heterogeneous change that subsumes many commonly used models as special cases.

The Multilevel Model Approach to the Analysis of Change

General MLMs allow parameters to enter the model in a nonlinear fashion (e.g., Davidian & Giltinan, 1995; Pinheiro & Bates, 2000; Vonesh & Chinchilli, 1997). The non-

linear MLM used throughout is a generalization of the formulation proposed by Lindstrom and Bates (1990). The general (univariate) nonlinear MLM is given as

$$y_{ijt} = f(\psi_{ij}, \mathbf{x}_{ij})_t + \epsilon_{ijt}, \qquad (1)$$

where y_{ijt} is the response from the ith individual ($i = 1, \ldots, N$) in the jth group ($j = 1, \ldots, J$), at the tth timepoint ($t = 1, \ldots, T$), $f(\cdot)$ is some functional form of change relating the r parameters in the vector ψ_{ij} and the predictors (or covariates or independent variables) in the vector $\mathbf{x}_{ij}$, and ϵ_{ijt} is the error term of the ith individual in the jth group at the tth measurement occasion (Lindstrom & Bates, 1990; Pinheiro & Bates, 2000). The parameter vector is potentially specific to the individual and such uniqueness is incorporated into the model by defining ψ_{ij} as

$$\psi_{ij} = \mathbf{A}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{v}_{ij}, \qquad (2)$$

where $\boldsymbol{\beta}$ is a vector of length P ($p = 1, \ldots, P$) of population fixed effect parameters, $\mathbf{v}_{ij}$ is a vector of length P with q nonzero elements representing unique effects associated with the ith individual in the jth group, $\mathbf{A}_{ij}$ is an r by P design matrix associated with the fixed effects, and $\mathbf{Z}_{ij}$ is an r by P design matrix with q nonzero elements associated with the unique effects.

Let $\mathbf{y}_{ij} = [y_{ij1}, \ldots, y_{ijT}]'$, $\mathbf{x}_{ij} = [\mathbf{x}_{ij1}, \ldots, \mathbf{x}_{ijN}]'$, and $\boldsymbol{\epsilon}_{ij} = [\epsilon_{ij1}, \ldots, \epsilon_{ijT}]'$. Equation 1 can then be written as

$$\mathbf{y}_{ij} = f(\psi_{ij}, \mathbf{x}_{ij}) + \boldsymbol{\epsilon}_{ij}. \qquad (3)$$

Combining Equations 2 and 3, the full MLM can be written as

$$\mathbf{y}_{ij} = f(\mathbf{A}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{v}_{ij}, \mathbf{x}_{ij}) + \boldsymbol{\epsilon}_{ij}. \qquad (4)$$

Application of the MLM generally places no restriction on $\mathbf{x}_{ij}$, $\mathbf{A}_{ij}$, or $\mathbf{Z}_{ij}$. Thus, unique measurement occasions (i.e., unbalanced data) pose no special problems, nor does data missing at random or completely at random (e.g., Little & Rubin, 1987; Schafer & Graham, 2002; Twisk & de Vente, 2002). Although not strictly necessary, applications of the MLM often assume that ϵ_{ijt} are independent and normally distributed with a mean of zero and common variance σ_ϵ^2. Furthermore, it is generally assumed that conditional on the fixed effects and the predictors, the unique effects are multivariate normally distributed with mean zero and finite covariance matrix.

The aforementioned nonlinear multilevel change model is useful in part because of its generality. However, because it is so general it can be difficult to translate into a specific change model. To illustrate an application of Equation 4, the logistic change curve is used, which is defined as

$$y_{it} = \frac{\alpha_i}{1 + \exp\left(-\frac{(a_{it}-\beta_i)}{\gamma_i}\right)} + \epsilon_{it}, \qquad (5)$$

where α_i is the asymptote for the ith individual as $a_{it} \to \infty$, β_i defines the point-of-inflection on the abscissa for the ith

individual, γ_i defines the ith individual's curvature parameter, and ε_{it} is the error for the ith individual at the tth time-point (see Appendix C.7 of Pinheiro & Bates, 2000). A key point to recognize is that the parameters of Equation 5 have i subscripts, which implies that they are (potentially) specific to the individual. Furthermore, the parameters of Equation 5 can themselves be modeled as a function of other variables in exactly the same manner as is typical in the standard multilevel modeling context (e.g., Raudenbush & Bryk, 2002; Singer & Willett, 2003).

The Nonlinear Change Mixture Model

Because of the deficiency of statistical methods currently available for modeling longitudinal data when interest lies in latent classes that each have their own potentially unique set of parameters for a change model nonlinear in its parameters, the *nonlinear change mixture model* (NCMM) is proposed and developed. This article helps to alleviate a shortcoming in the analysis of change literature by combining existing models into a unified approach for modeling nonlinear change in mixed populations. The NCMM is predicated on a heterogeneous population (like the GMM) explicitly acknowledging the fact that different classes of individuals potentially have their own unique set of parameter values, with or without cross class constraints, and allows within class variability around the class specific fixed effects for change models nonlinear in their parameters (unlike the GMM). The NCMM integrates the idiographic and nomothetic conceptualizations of behavior into a single unified model, where the model is designed for realistic functional forms thought to govern many behavioral, educational, and social phenomenon.

The NCMM assumes that individuals within a class have trajectories of change that are relatively homogeneous, whereas trajectories of change across class are relatively heterogeneous. Of course, because the NCMM assumes that interindividual differences in change are a function of both class membership and individual uniqueness, the degree of within group homogeneity is relative to the degree of across group heterogeneity. Following the formal definition of the NCMM, five estimation methods are proposed. The accuracy of the estimation methods are then explored via a comprehensive Monte Carlo simulation study.

Defining the Nonlinear Change Mixture Model

Let y_{it} be the outcome variable of the ith individual ($i = 1,\ldots, N$) at the tth measurement occasion ($t = 1,\ldots, T_i$), a_{it} be the value of some nonstochastic time dependent basis at the tth measurement occasion for the ith individual, and x_{im}

be the mth time invariant predictor variable ($m = 1,\ldots, M$) for the ith individual. Further let $\mathbf{y}_i = [y_{i1}, \ldots, y_{iT_i}]'$ be a vector of length T_i of the observed scores for the ith individual, $\mathbf{a}_i = [a_{i1}, \ldots, a_{iT_i}]'$ be the vector of length T_i of the time dependent basis for the ith individual, $\mathbf{x}_i = [x_{i1}, \ldots, x_{iM}]'$ be the vector of length M of time invariant predictor variables for the ith individual, and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]'$ be the N by M matrix of predictor variables for all individuals. The functional form governing the trajectory of change for all individuals is denoted $f(\mathbf{a})$, where $\mathbf{a}$ is a generic representation of the time dependent basis with the function $f(\cdot)$ having P change parameters. Given $f(\mathbf{a})$, let $\mathbf{\theta}_i = [\theta_{i1}, \ldots, \theta_{ip}]'$ be a P length vector of true change coefficients defining the trajectory for the ith individual and $\mathbf{\Theta} = [\mathbf{\theta}_i, \ldots, \mathbf{\theta}_N]'$ be the N by P matrix of individual change coefficients. The vector of length N that identifies which of the G classes from a mutually exclusive set the ith individual is a member is denoted $\mathbf{\gamma} = [\gamma_1, \ldots, \gamma_N]'$.

Given $f(\mathbf{a})$, $\mathbf{\Theta}$, and $\mathbf{\gamma}$, let $f(\mathbf{a}_g)$ be a special case of $f(\mathbf{a})$ for the gth class with $\pi_g \left(\sum_{g=1}^{G} \pi_g = 1 \right)$ being the proportion of the population who are members of the gth class and $n_g \left(N = \sum_{g=1}^{G} n_g \right)$ being the number of individuals in the sample who are members of the gth class. The functional form of $f(\mathbf{a}_g)$ is thus equal to or a special case of $f(\mathbf{a})$, however some if not all of the P parameters of change in $f(\mathbf{a}_g)$ are specific to the gth class. Conditional on $\mathbf{\theta}_i$, γ_i, and $\mathbf{\mu}_g$, where $\mathbf{\mu}_g = [\mu_{1g}, \ldots, \mu_{Pg}]'$ is the gth class specific population mean vector for the P *fixed* effect parameters of change, $\mathrm{E}[\mathbf{\upsilon}_i] = 0$ with variance $\mathbf{\Sigma}_g$ and has a P dimensional multivariate normal distribution, where $\mathbf{\upsilon}_i = [\upsilon_{i1}, \ldots, \upsilon_{iP}]'$ are the individual specific unique effects defined as $\upsilon_{ip} = [\theta_{ip} - (\mu_{gp}|\gamma_i)]$.

The NCMM implies that the coefficients of change for the individuals conform to the following probability density function:

$$\mathrm{d}(\mathbf{\theta}_i, \mathbf{x}_i; \mathbf{\Psi}) = \sum_{g=1}^{G} \pi_g \phi_P(\mathbf{\theta}_i, \mathbf{x}_i; \mathbf{\mu}_g, \mathbf{\Sigma}_g), \quad (6)$$

where $\mathbf{\theta}_i$ and $\mathbf{x}_i$ are considered simultaneously, ϕ_P represents a P dimensional multivariate normal probability density function, and $\mathbf{\Psi} = [\mathbf{\Psi}_1', \ldots \mathbf{\Psi}_G']'$, where $\mathbf{\Psi}_g = [\pi_g, \mathbf{\mu}_g', \text{vech}(\mathbf{\Sigma}_g)']'$ that may or may not be saturated with $\text{vech}(\cdot)$ being an operator that stacks the columns of a symmetric matrix by leaving out elements above the main diagonal. Equation 6, which defines the density of the heterogeneous population of change coefficients, is analogous to the general density of a mixture distribution (e.g., McLachlan & Basford, 1988; McLachlan & Peel, 2001).

What is of interest in the NCMM framework is not literally the probability density function of $\mathbf{\theta}_i$ given $\mathbf{x}_i$, rather interest is in the parameters from a nonlinear MLM with multiple classes of individuals. However, class member-

ship is unknown and must itself be estimated. A multiple group nonlinear MLM leads to a MLM of the form

$$\mathbf{y}_{ig} = f(\mathbf{A}_{ig}\boldsymbol{\beta}_g + \mathbf{Z}_{ig}\boldsymbol{\upsilon}_{ig}, \mathbf{x}_{ig}) + \boldsymbol{\epsilon}_{ig}, \qquad (7)$$

where rather than having a known grouping structure, g is latent and must be estimated. The class identification vector, $\boldsymbol{\gamma}$, can be estimated with the optimal rule of classification with a FMM whose probability density is defined by Equation 6. The optimal rule for classification states that an individual entity should be classified into the class that the entity is most likely to belong (McLachlan & Basford, 1988, pp. 11, 45–46; sometimes this is called the Bayes rule, Anderson, 1984, chapter 6). Note the one-to-one relationship between $\boldsymbol{\theta}_i$, the ith individual's change coefficients, in Equation 6 and $\boldsymbol{\upsilon}_{ig}$, the unique effects associated with the ith individual whom is in the gth class, in Equation 7.

Each of the classes in the general NCMM of Equation 7 potentially has a unique set of fixed effects, $\boldsymbol{\beta}_g$, a unique covariance matrix of errors, $\boldsymbol{\Sigma}_{\epsilon_g}$, and has a unique covariance matrix for the unique effects, $\boldsymbol{\Sigma}_{\upsilon_g}$. There are potentially a large number of parameters that can be implied by the model. Momentarily supposing each of the individuals is measured at the same T measurement occasions, there are potentially GP fixed effect parameters, $G(T(T+1))2^{-1}$ error covariance parameters, and $G(P(P+1))2^{-1}$ unique covariance parameters. Thus, with no parameter constraints, there are a total of $G[p + (T(T+1) + p(p+1))2^{-1}]$ parameters defining the NCMM. Of course, in order for the model not to be underidentified, parameter constraints must be imposed. Thus, not all possible parameters can be free to vary, as doing so yields an inestimable model where there are more equations than unknowns. Often a reasonable constraint for reducing a large number of parameters materializes if $\sigma_{\epsilon g}^2 \mathbf{I}$ is the error structure for each of the classes, and even more constrained when $\sigma_{\epsilon 1}^2 = \sigma_{\epsilon 2}^2 = ... = \sigma_{\epsilon G}^2$, where $\mathbf{I}$ is the identity matrix of the appropriate dimension. The potential flexibility of the model necessarily implies more parameters than either a standard MLM or LGC, but not necessarily more than the GMM. Given the difficulties that often exist for estimation of complicated FMMs and especially for nonlinear MLMs, it is no surprise that estimation of the NCMM can be difficult. The next section discusses how likelihood estimation theoretically applies to the parameters of the NCMM.

Parameters of the NCMM

Because the NCMM combines ideas from the FMM with ideas from the general nonlinear MLM, the likelihood function for the NCMM can be written as

$$\mathcal{L}(\mathbf{Y}, \mathbf{X}; \boldsymbol{\Omega}) = \prod_{i=1}^{N} d(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\Omega}) \qquad (8)$$

where $\mathcal{L}$ represents the likelihood function and $\boldsymbol{\Omega}$ is the R

length vector of parameter values defining the particular NCMM. The parameter vector $\boldsymbol{\Omega}$ can be written in a general form:

$$\boldsymbol{\Omega} = [\pi_1, \ldots, \pi_{G-1}, \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_G', \text{vech}(\boldsymbol{\Sigma}_{\epsilon_1})', \ldots, \\ \text{vech}(\boldsymbol{\Sigma}_{\epsilon_G})', \text{vech}(\boldsymbol{\Sigma}_{\upsilon_1})', \ldots, \text{vech}(\boldsymbol{\Sigma}_{\upsilon_G})']', \qquad (9)$$

where $\boldsymbol{\beta}_g$ is the vector of fixed effects for the gth class, $\boldsymbol{\Sigma}_{\epsilon_g}$ is the covariance matrix of errors for the gth class, and $\boldsymbol{\Sigma}_{\upsilon_g}$ is the covariance matrix for the unique effects for the gth class.

An obvious question that arises is how the likelihood function of Equation 8 relates to the probability density function defined in Equation 6. The vector of group specific population means contained in $\boldsymbol{\mu}_g$ from Equation 6 is analogous to the fixed effect change parameters contained in $\boldsymbol{\beta}_g'$ from Equation 9. Furthermore, the class specific population covariances contained in $\boldsymbol{\Sigma}_g$ from Equation 6 are analogous to the class specific covariances contained in $\boldsymbol{\Sigma}_{\upsilon_g}$ from Equation 9. Knowing this, Equation 6 can be rewritten so that it is consistent with the notation used for the likelihood function of the NCMM defined in Equation 8:

$$d(\boldsymbol{\theta}_i, \mathbf{x}_i; \boldsymbol{\Psi}) = \sum_{g=1}^{G} \pi_g \phi_P(\boldsymbol{\theta}_i, \mathbf{x}_i; \boldsymbol{\beta}_g, \boldsymbol{\Sigma}_{\upsilon_g}). \qquad (10)$$

Methods of Parameter Estimation for the NCMM

In large samples when all assumptions are met and the model is correctly specified, parameter estimation by full information maximum likelihood provides many benefits and is generally the optimal method from a statistical point of view. However, for many complicated models, especially nonlinear MLMs, the likelihood function may be difficult or impossible to write analytically given the current state of knowledge and the limitations of statistical theory (Pinheiro & Bates, 2000, section 7.2.1). If the likelihood function cannot be analytically derived, an approximate likelihood function can sometimes be maximized. As Davidian and Giltinan describe in the context of nonlinear MLMs, "the analytical intractability of likelihood inference has motivated many approaches based on approximations" for nonlinear MLMs (2003, p. 403). Thus, given the current limitations of likelihood estimation and inference, carrying out nonlinear MLMs is almost always based on approximate procedures. As a result, parameter estimates based on approximate likelihood functions are themselves necessarily approximate. That being said, they often still yield very accurate estimates and may be the only way to obtain estimates.

Given the approximate nature of parameter estimates from nonlinear MLMs, the estimation of the parameters of the NCMM is necessarily constrained. Because this article does not attempt to improve or propose estimation procedures for the nonlinear MLM, but rather relies on currently

implemented estimation procedures, a known shortcoming of the NCMM (like nonlinear MLMs in general) is that estimation of the parameters is based on estimation methods that do not necessarily have the optimal properties of full information maximum likelihood. Five estimation methods are presented that are conceptually reasonable and theoretically statistically sound in the following five subsections. Each of the five methods are based on multistage procedures with conceptual or theoretical advantages over the other methods. Each stage uses currently implemented statistical methods but combines the methods in a novel fashion in order to operationalize the NCMM.

Estimation Method 1

Method 1 is a three-stage method, where the first-stage of the model is the estimation of the P parameters of $f(\mathbf{a})$ for each of the N individuals. Estimation of each individual's $\boldsymbol{\theta}$ can be done by any acceptable estimation procedure, but because nonlinear least squares using the Gauss-Newton procedure is the most widely recommended estimation method for individual specific nonlinear models (Bates & Watts, 1988; Ratkowsky, 1983), it is the suggested method for Stage 1. Given the N $\hat{\boldsymbol{\theta}}_i$s and thus $\hat{\boldsymbol{\Theta}}$ from Stage 1, estimation of $\boldsymbol{\gamma}$ occurs at Stage 2. Stage 2 involves fitting a FMM with a fixed number of classes, where individuals are classified into a crisp set of G classes based on the highest posterior probability of class membership (i.e., the optimal rule). A crisp set in this context is where individuals are assigned to one and only one class and thus no partial membership exists. Stage 3 involves fitting G separate MLMs, where each class has a unique set of parameter estimates. These G MLMs have parameter estimates that are based only on the individuals who were estimated to be most likely members of the particular class.

Part of the success of this method is linked to the distance between the G P dimensional multivariate distributions – the larger the distance the better the rate of successful classification, holding everything else constant. In particular, if the G distributions representing each of the classes are, relevantly speaking, far apart from one another, few misclassifications will occur and this method will recover the class specific parameters. If the individuals can be correctly classified into the appropriate class, then each of the MLMs will be based on the appropriate set of individuals and thus no contaminated distributions will distort the parameter estimates. As is true with the application of any mixture model, as the overlap between distributions becomes more pronounced, the accuracy of classification is reduced. Conversely, as the misclassification rate approaches zero, Method 1 approaches the point where the NCMM yields estimates with the same accuracy as the multiple group nonlinear MLM.

Previous models in different contexts have been proposed for modeling heterogeneous populations in the context of longitudinal data analysis, where each of the methods was based on crisp sets (e.g., Dumenci & Windle, 2001; McCall, Appelbaum, & Hogarty, 1973; and Pauler & Laird, 2000). Such models restrict the contribution of each individual to only the fixed effects of the class that the individual is most likely to belong. Thus, the crisp set classification procedure of Method 1 is consistent with (at least) three previously proposed models of heterogeneous change. Further justification for Method 1 being a reasonable procedure is based on the theory of multi-stage estimation. Multistage estimation procedures are at times carried out when the likelihood function of the particular model is not analytically tractable and thus when full information maximum likelihood is not implementable (e.g., Chatterjee, 2004; Stukel & Demidenko, 1997). Multistage estimation procedures are also advantageous when there is a concern that the model may be misspecified (e.g., Bollen, 2001) or when a desire exists to evaluate the model at different levels individually (even when a likelihood function is known) rather than evaluating the full model simultaneously (e.g., Yuan & Bentler, 2007). Further, White (1982) reminded researchers that full information maximum likelihood assumes that the model is correctly specified. When models are misspecified, "specification error in one part of the system can spread bias to other parts of the model" (Bollen, 1996, p. 109). Even though in the ideal situation full information maximum likelihood estimation is known to be optimal, when the likelihood function cannot be derived or when the model is in some way misspecified, multi-stage estimation procedures, although themselves not fully efficient, can outperform maximum likelihood methods.

Estimation Method 2

Method 2 is similar to Method 1, however, rather than separating the individuals into G classes and performing a different MLM for each class in isolation at Stage 3, Method 2 is carried out by way of a multiple group MLM. In particular, given $\hat{\boldsymbol{\gamma}}$, obtained with a crisp set in the same way as was done in Method 1, an N by $G-1$ matrix of class identifiers (i.e., a matrix of dummy variables with a reference group specified) would be constructed and used as an identifier of class membership. Each class would then be treated simultaneously in a multiple group MLM.

Method 2 is a three-stage procedure like Method 1, and thus the advantages of multi-stage estimation procedures also apply to Method 2. Rather than estimating class specific parameters without regard to individuals estimated to be from other classes, Method 2 combines limited information from all classes when estimating parameters from each class. Method 2 also allows parameter invariance (such as cross-class constraints) to be imposed for any parameters of interest. Parameter invariance across class is straightforward with Method 2, but difficult with some of the other methods. Method 2 also allows the possibility of testing an interaction between the estimated class membership and

predictor variables (e.g., Is the asymptotic value for Class 1 larger than the asymptotic value in Class 2?), which adds a potentially interesting dimension to the description of change.

Notice that in Method 2 each class is fit in the context of a multiple group design which requires that the functional form of change and covariance structures for the unique effects and errors to generally be kept constant across all of the classes. Information from Stage 1 and Stage 2 is only used to classify the individuals into the G crisp sets. After the individuals have been classified, Method 2 treats the data as though it came from a multiple group design where the grouping factors are known. Because all classes are fit in a multiple group MLM, the standard errors are based on N individuals (rather than n_g for each of the classes as in Method 1). The estimated standard errors are analogous to the pooled error term in a general linear model context when homogeneity of variance is assumed. Basing the error term on all individuals in the analysis generally yields more statistical precision and power when compared to methods that divide the sample into multiple groups.

Estimation Method 3

Estimation of $\boldsymbol{\Theta}$ and $\boldsymbol{\gamma}$ is required prior to estimation of the parameters of the MLM in Method 1 and Method 2. Because $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\gamma}}$ are not likely to be independent of one another, information contained in $\hat{\boldsymbol{\Theta}}$ can provide information about $\hat{\boldsymbol{\gamma}}$, and vice versa. A full information maximum likelihood approach uses information from all levels of an analysis in the estimation of all parameters (which is why bias in one part of the model tends to affect the parameter estimates at all other parts of the model). Thus, all available information is used in the estimation of all parameter estimates in a full information maximum likelihood estimation procedure. Method 3 attempts to capitalize on the ideas of full information maximum likelihood in a multistage procedure. Method 3 is implemented by using $\hat{\boldsymbol{\Theta}}$ in the estimation of $\hat{\boldsymbol{\gamma}}$, and vice versa, in an iterative fashion. Method 3 begins by estimating the individual change coefficients given in $\tilde{\boldsymbol{\Theta}}$. The N by P matrix of initial estimates, obtained in the same way as Method 1 or Method 2, is denoted $\hat{\boldsymbol{\Theta}}_{(0)}$, where the subscript in parentheses implies the number of iterations necessary to obtain that set of estimates, with a zero subscript representing the initial conditions. Given $\hat{\boldsymbol{\Theta}}_{(0)}$, a FMM is carried out in order to obtain $\hat{\boldsymbol{\gamma}}_{(0)}$. A MLM in accord with Method 1 or Method 2 is then carried out. After the MLM estimates are obtained, new estimates of individual change coefficients that are known as empirical Bayes or James-Stein estimates are available and $\hat{\boldsymbol{\Theta}}_{(0)}$ is updated to $\hat{\boldsymbol{\Theta}}_{(1)}$ based on these new estimates. The empirical Bayes estimates in $\hat{\boldsymbol{\Theta}}_{(1)}$ are based in part on the individuals' scores and in part on the fixed effects of the particular class the individual is most likely to belong as a function of the reliability of the individuals' scores. These updated individual change coefficients are now used as input for the FMM

in order to obtain $\hat{\boldsymbol{\gamma}}_{(1)}$. The MLM is then carried out given $\hat{\boldsymbol{\gamma}}_{(1)}$ in order to obtain $\hat{\boldsymbol{\Theta}}_{(2)}$, and so forth. This iterative process continues until some criterion converges (e.g., the estimates minimizing the likelihood function, mean square error, or Akaike Information Criteria is minimized, etc.). To summarize the difference between Method 3 and Methods 1 and 2, after the MLM has been fit at Stage 3, the updated individual change coefficients (i.e., the empirical Bayes estimates) are used as input in an updated FMM. The new classifications (if any) are used to reestimate the fixed effects and the empirical Bayes estimates so that they can again be used as input into the FMM. This iterative process continues until the convergence criterion is realized.

This method has the potential to improve the estimation of parameters obtained in both Method 1 and in Method 2 by reducing the number of individuals that are misclassified. Although both the FMM and the MLM are iterative in nature themselves, Methods 1 and 2 do not combine the iterative process of the FMM and the MLM. The improvement of Method 3 is that the FMM and MLM are used in tandem. Because Method 3 can be based on the estimation procedures of Method 1 or Method 2, Method 3 will be denoted Method 3.1 or Method 3.2, depending on whether estimation using the rationale of Method 1 or Method 2 is used, respectively.

Estimation Method 4

In order to estimate the class specific change parameters, a standard application of the FMM can be applied to the set of estimated change coefficients, $\hat{\boldsymbol{\Theta}}$, which were estimated at the first stage of the analysis (e.g., using the Gauss-Newton procedure). The first stage of Method 4 is thus the same as the first stage in the three previous methods. The difference between Method 4 and the previous methods is that the FMM itself is used to estimate the NCMM parameters, rather than using a nonlinear MLM after individuals have been classified into the mutually exclusive crisp sets based on their posterior probabilities of class membership. The theoretical rationale for this approach is apparent when it is realized that the FMM provides estimates of the class specific means and covariances for the P change coefficients assumed to be a mixture of G multivariate normal distributions. Because what is of interest is the estimation of the means and covariances of the individuals' change coefficients, a FMM on $\hat{\boldsymbol{\Theta}}$ will provide estimates of the parameters of interest for the NCMM: mean vectors for the latent classes corresponding to fixed effects and covariance matrices for the latent classes corresponding to the covariance structure of the unique effects.

A major difference between Method 4 and the previous methods is that Method 4 does not classify individuals into mutually exclusive classes. Rather, all individuals potentially contribute to the class specific parameter estimates from all classes, in what can be considered a "fuzzy set," that is, class membership is not 0 or 1 but rather it is prob-

abilistic and estimates are weighted accordingly. In Method 4, if an individual has a high probability of belonging to the gth class, then the individual's $\boldsymbol{\theta}_i$ is heavily weighted in the estimation of the parameters from the gth class. However, in Method 4 if an individual's posterior probability of belonging to a particular class is small, the individual's scores receive little weight in the estimation of the parameter estimates from that particular class. This is the case because the mean vector and the covariance matrices are weighted by a function of the probability that the individual belongs to the particular class in accord with the theory of FMMs. The theoretical rationale of Method 4 is thus consistent with the GMM in that all individuals potentially influence all parameter estimates. Method 4 is thought to be especially useful when the overlap among the classes leads to probabilities of class membership that are not close to unity or zero (i.e., when classification is not so clear). Furthermore, the FMM allows for covariance structures to be class specific or constrained across the classes.

Summary of Estimation Methods

Because some of the details of the estimation methods differ only in subtle ways, a summary of each of the estimation methods is provided. The summary provided is not meant to supplant the full description provided for each of the methods of estimation, but rather it is meant to supplement the description provided.

Summary of Method 1

– *Stage 1:* Estimate the P parameters of $f(\mathbf{a})$ for each of the N individuals using the Gauss-Newton estimation procedure (or other appropriate estimation procedure), obtain $\hat{\boldsymbol{\Theta}}$.
– *Stage 2:* Fit a FMM with G classes using $\hat{\boldsymbol{\Theta}}$ as input and then classify the N individuals into G crisp sets based on the highest posterior probability, obtain $\hat{\boldsymbol{\gamma}}$.
– *Stage 3:* Using the original data and $\hat{\boldsymbol{\gamma}}$ to identify class membership, fit G separate MLMs, one for each of the estimated classes of individuals, in isolation.

Summary of Method 2

– *Stage 1*: Same as Stage 1 from Method 1.
– *Stage 2*: Same as Stage 2 from Method 1.
– *Stage 3*: Using the original data and $\hat{\boldsymbol{\gamma}}$ to identify class membership, fit a G-group MLM, where each of the individuals is classified into one and only one class.

Summary of Method 3

– *Stage 1*: Same as Stage 1 from Methods 1 and 2.

– *Stage 2*: Same as Stage 2 from Methods 1 and 2.
– *Stage 3*: Same as Stage 3 from Method 1.
– *Stage 4, Method 3.1*: Using the empirical Bayes estimates obtained after G MLMs have been fit in isolation, refit the FMM (Stage 2) using $\hat{\boldsymbol{\Theta}}_{(1)}$ in an effort to improve the classification success. Iterate between Stage 2, Stage 3, and Stage 4 until the specified convergence criterion is met.
– *Stage 4, Method 3.2*: Using the empirical Bayes estimates obtained after a G-group MLM has been fit, refit the FMM (Stage 2) using $\hat{\boldsymbol{\Theta}}_{(1)}$ in an effort to improve the classification success. Iterate between Stage 2, Stage 3, and Stage 4 until the specified convergence criterion is met.

Summary of Method 4

– *Stage 1*: Same as Stage 1 from Methods 1, 2, and 3.
– *Stage 2*: Fit the FMM as is done in Stage 2 from Methods 1, 2, 3.1, and 3.2; use the parameter estimates obtained in the FMM to estimate the parameters of the NCMM.

Evaluation of the Estimation Procedures

The logistic change curve as discussed in Equation 5 is used as the exemplar nonlinear change model in a Monte Carlo simulation study to evaluate the effectiveness of the estimation methods. The reason for choosing a three parameter logistic change curve to represent general nonlinear change models is because: (a) it is a popular change curve in the medical and biological literature, (b) it is representative of sigmoidal forms which are theoretically useful in behavioral, educational, and social research, (c) given certain parameter combinations it can take on characteristics that in some ways are similar to the asymptotic regression change curve (negative exponential), and (d) because the logistic change curve is representative of models in the Richards family of change curves (Richards, 1959), which are quite general and especially useful for modeling limited capacity data (e.g., Browne & du Toit, 1991; van Geert, 1991). It is thought that there is nothing unique about the logistic change curve to imply that the proposed methods would be more or less effective than if another similarly complicated nonlinear change curve were used. Although the logistic change curve is used here, the methods are general and not in any way specific to the logistic change model. Of course, future work on extensions and improvements to the estimation procedures of the NCCM will be necessary to investigate other functional forms.

Although a thorough delineation of the results of the Monte Carlo simulation study conducted to evaluate the estimation procedures of the NCMM is beyond the scope of this
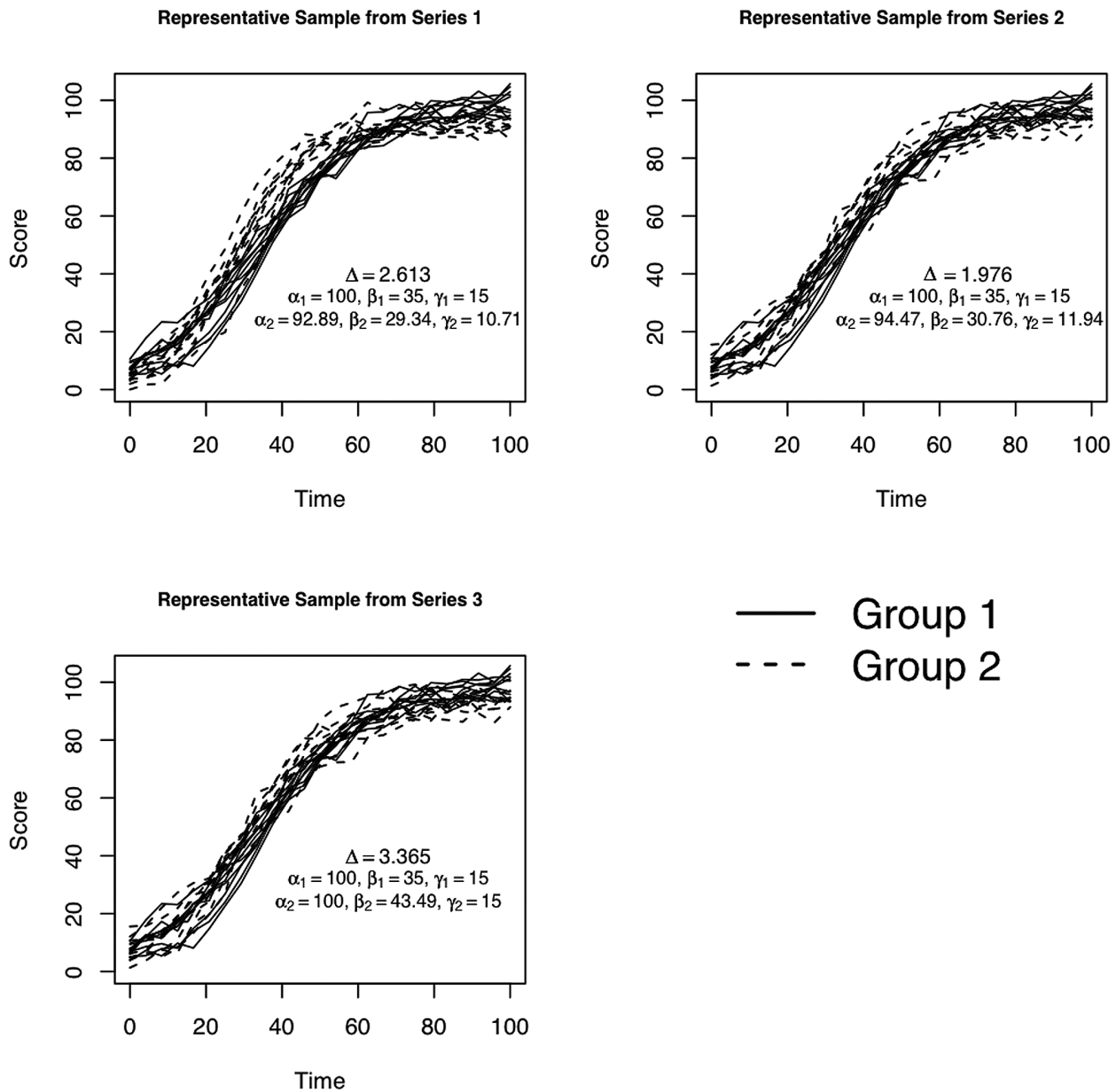
Representative Sample from Series 1

Representative Sample from Series 2

Representative Sample from Series 3

——— Group 1
- - - Group 2

Figure 1. Random samples from a condition within each of the series of fixed effects studied in the Monte Carlo simulation study. *Note:* The value of Δ denoted in the figures is the Mahalanobis distance between the fixed effects given the within group variance with an error variance of 4. The values of the within group parameters are included, with the subscript representing the latent class.

article, a summary of the simulation study and the results obtained is provided. A technical supplemental packet is available from the author that delineates the design and implementation of the Monte Carlo simulation study.

The simulation study consisted of a two (sample size; $N = 200$ and 500) by three (proportion of individuals belonging to each class; $\pi_1 = .5, .75$, and $.875$) by two (the number of measurement occasions; $T = 15$ and 25) by two (the magnitude of the error variance; $\sigma_\varepsilon^2 = 4$ and 8) by three (distance between the *GP* dimensional multivariate normal distributions; square root of the Mahalanobis distance of 2.613, 1.976, and 3.635) fully factorial design,

using a modest correlational structure among the unique effects ($\rho_{\alpha,\beta} = .5, \rho_{\alpha,\gamma} = .4, \rho_{\beta,\gamma} = .3$, where ϱ is the correlation of the subscripted quantities). This design was carried out for each of the five estimation methods yielding a total of 360 conditions. There were two latent classes in actuality as well as fitted. Missing data, conditional models (i.e., those with time invariant concomitant variables), models with time varying covariates, and model misspecifications are not addressed in the simulation study. The goal of the simulation study is to evaluate the effectiveness of the proposed procedures in the estimation of the NCMM parameters under ideal conditions.

A series of 120 situations were tested within each of the three Mahalanobis distances specifications (thus a total of 360 conditions). Figure 1 illustrates 10 random trajectories from both latent classes for each of the three values of the Mahalanobis distance (using the smaller of the two error variances). As can be seen, differences do exist between the latent classes, but the differences are not large. In situations where differences between the latent classes is more pronounced – likely a motivating factor for applying the NCMM – the methods are guaranteed to work better, because if everything else is held constant classification will improve. Each of the situations described in Figure 1 include the Mahalanobis distance (Δ) as well as the class specific logistic change parameters, where the latent class is denoted with a subscript.

Simulations to evaluate the proposed methods were conducted using the R computer program (R Development Core Team, 2005). In addition to the standard functions contained in R, the NLME package (Pinheiro, Bates, DebRoy, & Sarkar, 2004) and the MCLUST package (Fraley & Raftery, 2004) were used. There were 1,000 replications for each of the 360 specific conditions examined. Instances where the particular data set failed to be estimable with any of the methods were not considered in the results. An additional data set was generated for each of the data sets that failed to yield results. Although widespread failures of the methods would cause concern, there were relatively few failures in the simulation study. Across all generated data sets, there were a total of 128 failures (out of a total of 72,128 total, failed and successful, attempts; 0.18% of the total number of attempts). The majority of the failures occurred in situations with smaller sample size, larger error variance, and fewer occasions of measurement.

Operationalizing the Results

The quantification of the effectiveness of the estimation procedures will be in terms of the discrepancy between the known population values and the estimated values obtained from the NCMM. In particular, the bias ($B = \hat{\omega} - \omega$) and relative bias ($P = [\hat{\omega} - \omega]/\omega$) from each condition will be determined empirically from the simulation study, where $\hat{\omega}$ is the estimate of ω, the known population value. Thus, a value of zero would represent a scenario where the estimated value is equal to the population value (i.e., the estimate would be unbiased). Whereas the bias quantifies the raw discrepancy between estimated and population values, the relative bias scales the bias so that the degree of discrepancy is relative to the value of the parameter. Because of the unknown nature of the distribution of the parameter estimates, their skew and (excess) kurtosis will also be calculated.

Accuracy can be conceptualized as the square root of the mean square error (RMS). The RMS measures the effectiveness of an estimator at recovering the population value. The statistical definition of accuracy is given as

$$\sqrt{E[(\hat{\theta} - \theta)^2]} = \sqrt{E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta)^2} = \text{RMS}, \quad (11)$$

where θ is a parameter of interest and $\hat{\theta}$ is an estimate of the parameter (Rozeboom, 1966, p. 500). Notice that the first component under the second radical sign is the precision, whereas the second component under the second radical sign is the squared bias. Thus, Equation 11 implies that accuracy is a function of bias and precision. As a measure becomes more accurate the value of the RMS decreases (and vice-versa) (perfect accuracy implies RMS = 0).

Although evaluating other aspects of the methods' effectiveness is possible, the fixed effects are studied in the simulation study. Without accurate fixed effects, accurate standard errors and/or correlation structures are moot. If the fixed effects are themselves accurately estimated, even generally incorrect standard errors can be corrected with alternative methods that are known to have desirable properties (e.g., the bootstrap or jackknife procedures). Although the covariance structure of the change coefficients is important, it has not yet been extensively studied.

Results

Table 1 shows results of the bias and relative bias for each of the fixed effects for all of the methods in the first condition studied (recall there were 72 situations total), as well as the skew and kurtosis for the parameter estimates. The similarity between the mean and the median discrepancy measures is due to the distributions of estimated change parameters being so small. In fact, the mean of the skewness for the 30 fixed effects (6 parameters by 5 methods) within Table 1 was only $-.036$ (with standard deviation .515), with the largest absolute skewness in Table 1 being -1.184 for γ_1 in Method 4.

The worst case in terms of the relative bias for any of the methods within Table 1, which was for γ_2 from Method 3.2, illustrates the effectiveness of all of the methods at recovering the population values of the fixed effects in the first condition. The worst case within Table 1 had an estimated parameter that was only .103 units smaller than the population value (10.610 when the true value was 10.713). Such a trivial difference between the true and empirically obtained values illustrates that all of the methods were effective at recovering the population values of the fixed effects in this particular situation. Although each of the recovered parameters was shown to be accurate for the first situation, 71 other situations exist. Inspection of the results from the remaining 71 situations ($5 \times 71 = 355$ other conditions) reveal similar findings as those presented in Table 1. Further evidence that the methods are effective at recovering the population fixed effect values across all 360 conditions is that, with

Table 1. Summary of the discrepancies between the estimated population values and the true population values for each of the model parameters for in the the first situation for each of the five estimation methods

Method 1

D	Stat.	α_1	α_2	β_1	β_2	γ_1	γ_2
B	$\hat{\mu}$	0.242	–0.156	0.226	–0.116	0.136	–0.083
	$\hat{M}$	0.256	–0.155	0.255	–0.127	0.15	–0.109
	$\hat{\sigma}^2$	0.409	0.388	0.305	0.266	0.222	0.202
	2.5%	–1.010	–1.409	–0.866	–1.026	–0.801	–0.884
	97.5%	1.443	1.146	1.306	0.936	0.946	0.857
P	$\hat{\mu}$	0.002	–0.002	0.006	–0.004	0.009	–0.008
	$\hat{M}$	0.003	–0.002	0.007	–0.004	0.010	–0.109
	$\hat{\sigma}^2$	<.001	<.001	<.001	<.001	0.001	0.002
	2.5%	–0.010	–0.015	–0.025	–0.035	–0.053	–0.083
	97.5%	0.014	0.012	0.037	0.032	0.063	0.080
Skew		–0.084	0.407	0.031	0.127	–1.173	0.468
Kurtosis		3.134	2.263	2.415	3.263	7.523	6.561

Method 2

D	Stat.	α_1	α_2	β_1	β_2	γ_1	γ_2
B	$\hat{\mu}$	0.247	–0.157	0.229	–0.116	0.137	–0.084
	$\hat{M}$	0.257	–0.155	0.259	–0.127	0.151	–0.110
	$\hat{\sigma}^2$	0.416	0.386	0.308	0.265	0.224	0.201
	2.5%	–1.004	–1.404	–0.868	–1.026	–0.804	–0.880
	97.5%	1.456	1.138	1.321	0.939	0.951	0.856
P	$\hat{\mu}$	0.002	–0.002	0.007	–0.004	0.009	–0.008
	$\hat{M}$	0.003	–0.002	0.007	–0.004	0.010	–0.010
	$\hat{\sigma}^2$	<.001	<.001	<.001	<.001	0.001	0.002
	2.5%	–0.01	–0.015	–0.025	–0.035	–0.054	–0.082
	97.5%	0.015	0.012	0.038	0.032	0.063	0.080
Skew		–0.061	0.414	0.042	0.130	–1.156	0.470
Kurtosis		3.125	2.266	2.434	3.258	7.416	6.566

Method 3.1

D	Stat.	α_1	α_2	β_1	β_2	γ_1	γ_2
B	$\hat{\mu}$	0.225	–0.200	0.195	–0.130	0.121	–0.101
	$\hat{M}$	0.245	–0.185	0.235	–0.142	0.136	–0.125
	$\hat{\sigma}^2$	0.352	0.328	0.277	0.225	0.180	0.168
	2.5%	–0.968	–1.395	–0.844	–0.989	–0.769	–0.893
	97.5%	1.333	0.965	1.234	0.832	0.898	0.753
P	$\hat{\mu}$	0.002	–0.002	0.006	–0.004	0.008	–0.009
	$\hat{M}$	0.002	–0.002	0.007	–0.005	0.009	–0.125
	$\hat{\sigma}^2$	<.001	<.001	<.001	<.001	0.001	0.001
	2.5%	–0.010	–0.015	–0.024	–0.034	–0.051	–0.083
	97.5%	0.013	0.010	0.035	0.028	0.060	0.070
Skew		–0.014	0.218	–0.285	0.424	–0.560	0.479
Kurtosis		1.864	2.132	1.523	1.599	2.453	1.844

Method 3.2

D	Stat.	α_1	α_2	β_1	β_2	γ_1	γ_2
B	$\hat{\mu}$	0.235	–0.197	0.202	–0.132	0.129	–0.103
	$\hat{M}$	0.251	–0.181	0.233	–0.145	0.144	–0.126
	$\hat{\sigma}^2$	0.377	0.337	0.288	0.240	0.189	0.178
	2.5%	–0.988	–1.404	–0.868	–1.026	–0.770	–0.880
	97.5%	1.380	1.007	1.232	0.847	0.931	0.755
P	$\hat{\mu}$	0.002	–0.002	0.006	–0.004	0.009	–0.010
	$\hat{M}$	0.003	–0.002	0.007	–0.005	0.01	–0.012
	$\hat{\sigma}^2$	<.001	<.001	<.001	<.001	0.001	0.002
	2.5%	–0.01	–0.015	–0.025	–0.035	–0.051	–0.082
	97.5%	0.014	0.011	0.035	0.029	0.062	0.070
Skew		–0.153	0.157	–0.278	0.032	–0.524	0.002
Kurtosis		2.641	1.616	1.554	3.477	2.567	5.088

Method 4

D	Stat.	α_1	α_2	β_1	β_2	γ_1	γ_2
B	$\hat{\mu}$	0.072	0.118	0.059	0.088	0.024	0.085
	$\hat{M}$	0.104	0.098	0.088	0.076	0.042	0.058
	$\hat{\sigma}^2$	0.339	0.337	0.254	0.230	0.186	0.176
	2.5%	–1.097	–1.003	–0.985	–0.764	–0.865	–0.633
	97.5%	1.105	1.380	0.993	1.088	0.779	1.001
P	$\hat{\mu}$	0.001	0.001	0.002	0.003	0.002	0.008
	$\hat{M}$	0.001	0.001	0.003	0.003	0.003	0.005
	$\hat{\sigma}^2$	<.001	<.001	<.001	<.001	0.001	0.002
	2.5%	–0.011	–0.011	–0.028	–0.026	–0.058	–0.059
	97.5%	0.011	0.015	0.028	0.0370	0.052	0.093
Skew		–0.475	0.588	–0.370	0.453	–1.184	0.794
Kurtosis		2.587	1.930	1.279	1.537	6.077	4.472

Note: The column heading "D" represents the discrepancy (i.e., a measure of the difference between the estimate and population value) and the column heading "Stat." represents the statistic used to summarize the discrepancies. B represents the bias, whereas P represents the relative bias. Summaries of B and P are given by $\hat{\mu}$, $\hat{M}$, $\hat{\sigma}^2$, 2.5%, and 97.5%, which represent the mean, median, variance, 2.5th and the 97.5th percentile, respectively. Skew and kurtosis for each of the observed parameter estimates is equal to the skew and kurtosis of B and P, respectively, and thus only a single value is given for each of the parameters. The fixed effects correspond to the upper left plot in Figure 1 and is based on the first of the 72 situations studied in the Monte Carlo simulation.

the exception of only two fixed effects, the 2.5% and 97.5% quantiles from the distribution of the discrepancies bracket zero. This implies that each of the methods is close to providing what is functionally an unbiased estimate of the fixed effects, in the sense that zero does not tend to be excluded from the bounds bracketing the discrepancy between the estimated change coefficients and their corresponding population value. The full set of results is available in the aforementioned supplemental packet.

Figure 2 provides an omnibus summary that aggregates accuracy across all situations within the first combination of change coefficients (i.e., all situations within the $\Delta = 2.613$ specification). The height of the bars in the figure represents

accuracy, with the dark gray part of the bars representing the proportion of bias contributing to the overall measure of accuracy and the light gray part of the bars representing the proportion of precision that contributes to the overall measure of accuracy. As can be seen in Figure 2, Method 4 is more accurate than the other methods for each of the change coefficients in Class 1. However, all of the methods yielded approximately the same accuracy for the parameters of Class 2. As can be seen, Method 4 was more biased than the other methods for the fixed effects of Class 2, but more precise, with the end result being essentially the same amount of accuracy across methods for Class 2. Thus, in situations where sample size is small, Method 4 is more biased but more precise than the other methods. The precision of Method 4 usu-
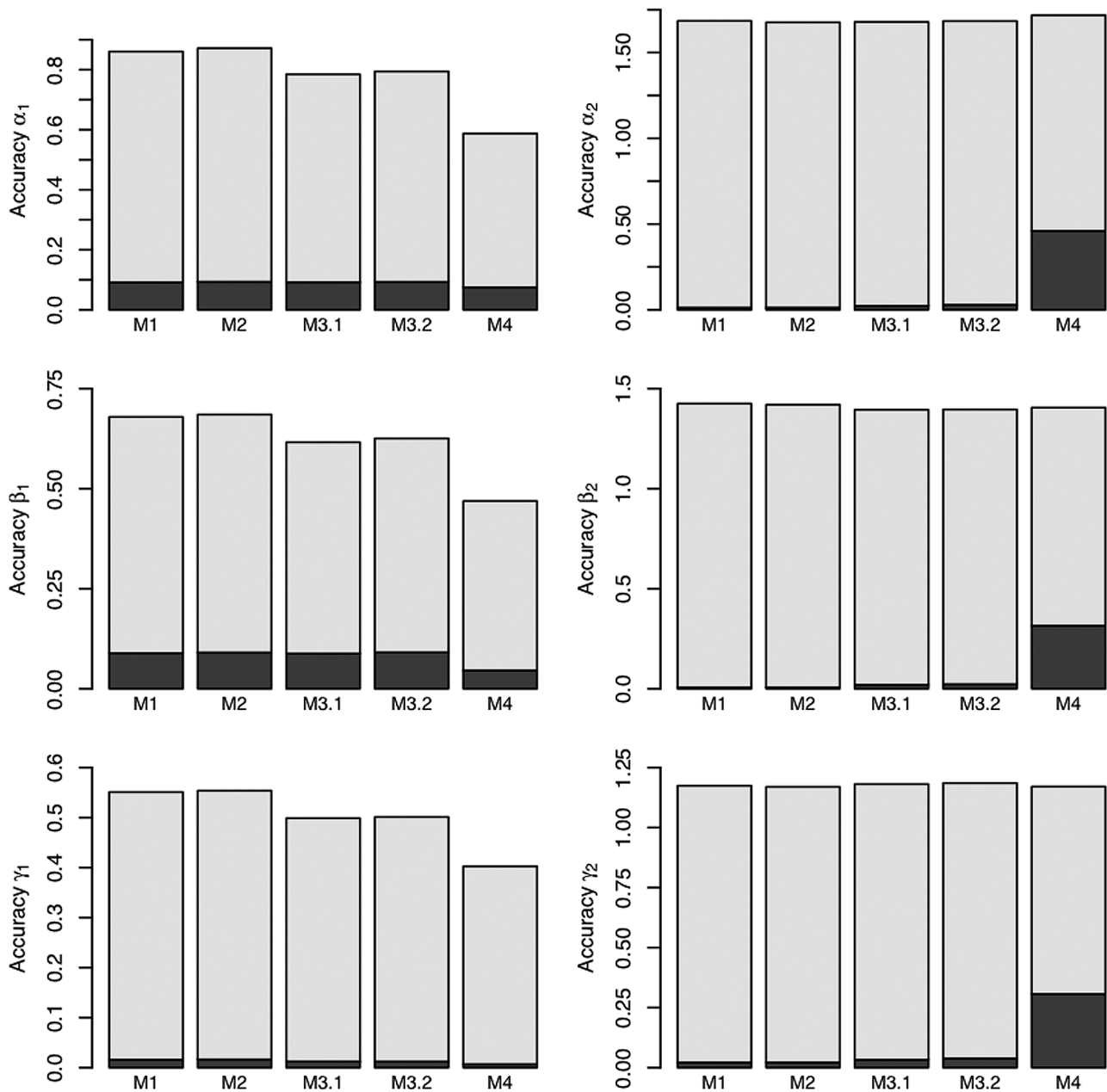
Figure 2. Mean accuracy as a function of bias (dark gray) and precision (light gray) across the first scenario studied for each of the fixed effects.

ally outweighs the bias and thus leads to estimates that are overall still approximately as accurate as the other methods. An analogous pattern emerges for Series 2 when plotted.

Series 3 provides interesting results. This is because there are cross-class constraints leading to fixed effects being based on weighted means for some methods (specifically Methods 1, 3.1, and 4). While at this time it is unclear if such weighted means for constrained parameter estimates is optimal, the results of Series 3 show that Method 4 leads to more accurate estimates for all but one of the fixed effects. In particular, it is shown that β_2 is more accu-

rately estimated with Method 3.2 than it was with Method 4. However, for each of the other three fixed effects Method 4 is more accurate.

Using the Estimation Methods for Implementing the Model

The effectiveness of each method should not be overlooked because of the overall accuracy and success of Method 4.

What should be emphasized is the advantage that each of the methods offers in different situations. The way in which each of the methods attempts to recover the parameters of the NCMM offers interesting possibilities in terms of which method might be the most appropriate for a specific question of interest as it relates to the theoretical framework of interest.

Recall that Method 1 and Method 3.1 treat the G classes in isolation. Although this is not optimal in the sense of obtaining the best estimate of the root mean square error and thus estimation of the standard errors and confidence intervals when the model is identical across class and homogeneity of variance is realized, treating the latent classes in isolation potentially has advantages. In particular, once classification has been carried out, the functional forms of change for the classes need not be constant.

Recall that the NCMM specified $f(\mathbf{a}_g)$ for the change model, meaning that the function is the same across the classes. This is different than allowing $f_g(\mathbf{a})$, where the functional form of change would itself be specific to the group. The NCMM model can be extended so that more fundamental differences between the functional form of change can be realized. Suppose that in a two class situation one latent class could be based on an asymptotic regression change curve whereas the other can be based on a Gompertz change curve. The unresolved issue is that the finite mixture model was based on the change coefficients from a single model. Although nested models do not seem to pose problems when specifying constraints across class differences, problems may arise when fundamental difference exist between the functional forms of change. Although the flexibility of treating the latent classes in isolation could be beneficial, more work is needed to determine if and when fundamental differences in change models across class is tenable for future applications of the NCMM model.

Relatedly, there is no reason that the latent classes are mandated to have the same error structure. For example, one latent class can have a hypothesized auto-regressive error structure whereas the other class can have a hypothesized standard error structure of the form $\sigma_\epsilon^2 \mathbf{I}$. The point is that if some difference in the model occurs across the classes, those differences can be evaluated using Method 1 or Method 3.1.

Method 2 and Method 3.2 are based on a single G-group MLM. Having a single model that is conditional on latent class, yet that also takes into account information from all classes is beneficial when the model is the same across class. When the functional form of change and the covariance structure of the errors is the same for all of the G classes, Method 2 and Method 3.2 are theoretically more powerful and precise than Method 1 and Method 3.1. The power and precision of Method 2 and Method 3.2 is a function of all individuals being used in the estimation of the root mean square error and thus the standard errors of the fixed effects. Perhaps the biggest benefit of using Method 2 or Method 3.2 is that hypothesis tests and confidence intervals can be carried out directly on differences in fixed effect parameters in a multiple group MLM approach. Along these same lines, Method 2 and Method 3.2 can constrain parameters a priori to be set equal to one another. One set of fixed effects examined cross class equality of the asymptotic value (α) and the curvature parameter (γ) in the simulation study. Method 2 and method 3.2 allowed estimated parameters to be based on the data from all individuals. However, Method 1, Method 3.1, and Method 4 required cross-class constraints for those parameters in a post hoc fashion. The ability to test hypotheses about differences in parameters directly and to formerly constrain them makes Method 2 and Method 3.2 very flexible.

The main difference between Method 4 and the other methods is that Method 4 is not based on crisp classification, but rather all individuals contribute to the parameters of all classes. Specifically, the individual contributes to each class specific set of parameter estimates by a function of the likelihood that the individual belongs to that class. Rather than assigning an individual entity to one and only one class, Method 4 allows what is functionally partial inclusion for parameter estimates. This translates into the class specific parameter estimates being based to some degree on data from all of the individuals, based on the probability of that individual belonging to that class. The benefits of using weighted probability for estimation purposes is most prevalent when there is not a large difference in the multivariate distributions of individual change coefficients. When the difference in the distributions are large, the probabilities of class membership approach zero or one, which means that Method 4 functionally becomes more like the other (crisp classification) methods. The real advantage of the estimation procedure of Method 4 is when there is more class ambiguity in the classification of individuals into the most likely class. The clear advantage of Method 4 is in terms of the accuracy with which the fixed effect parameter estimates can be recovered.

Even though the Monte Carlo simulation study showed Method 4 was the overall most accurate method, a blanket statement recommending Method 4 cannot be given because some of the other methods have advantages in certain situations. Due to the positive effect of iterating back and forth between Stage 2 and 3, Method 3.1 and Method 3.2 are universally recommended over Method 1 and Method 2, respectively, as they tended to be more accurate in the simulation study. Although accuracy is important, use of the best method to answer the particular question of interest is also necessary. Because of the success each method showed in the simulation study and the fact that they each have their own advantages, the method of choice depends first and foremost on the research question of interest.

Discussion

This article introduced the *nonlinear change mixture model* (NCMM) and evaluated five estimation procedures so that

some of the limitations of existing change models could be overcome when unknown heterogeneity exists in populations where change follows a functional form nonlinear in its parameters. The flexibility of the NCMM becomes apparent when it is realized that special cases of the NCMM can be made to equal many of the commonly used models of change. Although the estimation methods are different, when the functional form of change is linear in its parameters, the NCMM conceptually reduces to the GMM. When heterogeneity exists and the within class variability is restricted to zero, the NCMM conceptually reduces to the latent class growth model when the functional form specified is linear (e.g., Nagin, 1999). When class membership is known, the NCMM is equivalent to a multiple group nonlinear MLM. When the population is homogeneous, the NCMM is equivalent to a standard nonlinear MLM (or LGC when the functional form of change is linear in its parameters). The NCMM thus fits nicely into the existing analysis of change literature, yet it provides a needed extension.

Five estimation procedures were proposed and evaluated via a comprehensive Monte Carlo simulation study. The results showed that each of the methods were effective at recovering the parameters of the NCMM even when latent class separation was not noticeably large. The fact that each of the methods were generally accurate at recovering the parameters of the model is beneficial, because some of the methods can address fundamentally different types of questions researchers might have. In a standard application of the NCMM where no cross-class constraints are imposed and the functional form of change is constant across the *G* classes, Method 4 was shown to be the most accurate method with regard to fixed effects and is recommended in such situations. When there is an interest in using different functional forms of change for different latent classes, or if there is an interest in specifying different error structures or a different covariance structure for unique effects, Method 3.1 is recommended. When an interest exists in imposing cross-class constraints or testing the difference between parameters of change for different classes and the error structure is constant across class, Method 3.2 is recommended. The suggested method is thus tied to the researcher's question of interest.

Across the 16 fixed effects for the 72 situations, Method 4 was inferior for only one fixed effect. The closest competitor to Method 4 was Method 3.2. Method 3.2 was shown to be more accurate for most of the situations within Series 3, but when aggregating across Series 3 Method 4 was most accurate for three out of the four parameters. Whereas Method 4 was shown to be the most accurate of the methods examined, perhaps more important is the fact that each of the methods recovered the fixed effect parameters well. While there has been much attention devoted to separating the methods by determining which of the methods is most accurate, the effectiveness of Methods 3.1, 3.2, and 4 at recovering the parameters is actually similar in most cases. Because the fit of the models is similar for these

methods, what should be emphasized and considered when using the NCMM is that depending on the question of interest, one estimation method may be most appropriate. Thus, defining a question of interest will tend to dictate which of the three estimation methods should be used.

When heterogeneity exists in an analysis of change context, ignoring the heterogeneity introduces bias into the obtained results. In such a situation the parameter estimates that are supposed to represent a single homogeneous population instead represent a mixture of homogeneous subpopulations. This potentially translates into a model of change that is meant to represent everyone with a single (fixed effect) trajectory but in actuality may fit no one. When theory or data suggest the existence of latent classes in an analysis of change context, combining a MLM or LGC framework with the FMM ensures that parameter estimates are conditional on the latent classes. In cases where the functional form governing change is linear in its parameters, the GMM provides a powerful and elegant model to help understand change. However, in cases where the functional form of change is governed by a nonlinear functional form, which is probably more often than not theoretically true in the behavioral, educational, and social sciences, use of the NCMM is appropriate. The NCMM and its estimation methods can address a diverse set of questions arising from data having a wide range of complicated structures. The hope is that by explicitly modeling unknown heterogeneity in populations where the functional form of change is governed by a nonlinear functional form, the NCMM will help better address the questions asked by researchers who are interested in understanding how a phenomenon changes over time.

Acknowledgments

References

Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Henry Hold and Company.

Anderson, T.W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: Wiley, Inc.

Bates, D.M., & Watts, D.G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.

Bollen, K.A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika, 61*, 109–121.

Bollen, K.A. (2001). Two-stage least squares and latent variable

models: Simultaneous estimation and robustness to misspecifications. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future. A festschrift in honor of Karl Joreskog* (pp. 119–138). Lincolnwood, IL: Scientific Software International, Inc.

Browne, M.W., & du Toit, S.H.C. (1991). Models for learning data. In L.M. Collins & J.L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions.* Washington, DC: American Psychological Association.

Burchinal, M., & Appelbaum, M.I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development, 62*, 23–43.

Chatterjee, N. (2004). A two-stage regression model for epidemiological studies with multivariate disease classification data. *Journal of the American Statistical Association, 99*, 127–138.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research, 31*, 371–403.

Cudeck, R., & Harring, J.R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology, 58*, 615–637.

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

Davidian, M., & Giltinan, D.M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics, 8*, 387–419.

Dumenci, L., & Windle, M. (2001). Cluster analysis as a method of recovering types of intraindividual growth trajectories: A Monte Carlo study. *Multivariate Behavioral Research, 36*, 501–5522.

Dunn, J.G.H. (1994). Toward the combined use of nomothetic and idiographic methodologies in sport psychology: An empirical example. *The Sport Psychologist, 8*, 376–392.

Fraley, C., & Raftery, A.E. (2004). *MCLUST: Software for model based clustering* (Version 2.1–5) [computer software and manual]. Retrieved from http://www.cran.r-project.org/.

Lindstrom, M.J., & Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics, 46*, 673–687.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data.* New York: Wiley, Inc.

Magnusson, D., & Bergmen, L.R. (1988). Individual and variable-based approaches to longitudinal research on early risk factors. In M. Rutter (Ed.), *Studies of psychosocial risk: The power of longitudinal data* (pp. 45–61). New York: Cambridge University Press.

McCall, R.B., Appelbaum, M.I., & Hogarty, P.S. (1973). Developmental changes in mental performance. *Monographs of the Society for Research in Child Development, 38*, 1–83.

McLachlan, G.J., & Basford, K.E. (1988). *Mixture models: Inference and applications to clustering.* New York: Marcel Dekker.

McLachlan, G.J., & Peel, D. (2001). *Finite mixture models.* New York: Wiley, Inc.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.

Muthén, B. (2001a). Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Hillsdale, NJ: Erlbaum.

Muthén, B. (2001b). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L.M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: American Psychological Association.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*, 81–117.

Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C. et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*, 459–475.

Muthén, B., & Muthén, L.K. (2000). Integrating person-centered and variables-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882–891.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463–469.

Nagin, D.S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4*, 137–157.

Nesselroade, J.R., & Baltes, P.B. (1979). (Eds.). *Longitudinal research in the study of behavior and development.* New York: Academic Press.

Pauler, D.K., & Laird, N.M. (2000). A mixture model for longitudinal data with application to assessment of noncompliance. *Biometrics, 56*, 464–472.

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

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2004). *NLME: Linear and nonlinear mixed effects models* (version 3.3–48) [computer software and manual]. Retrieved from http://www.cran.r-project.org/.

R Development Core Team. (2005). *R: A language and environment for statistical computing* [computer software and manual], R foundation for statistical computing. Retrieved from http://www.r-project.org.

Ratkowsky, D.A. (1983). *Nonlinear regression modeling: A unified practical approach.* New York: Marcel Dekker, Inc.

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

Richards, F.J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany, 10*, 290–300.

Rozeboom, W.W. (1966). *Foundations of the theory of prediction.* Homewood, IL: The Dorsey Press.

Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Schulenberg, J., O'Malley, P.M., Bachman, J.G., Wadsworth, K.N., & Johnston, L.D. (1996). Adolescent risk factors for binge drinking during the transition to young adulthood: Variable and pattern-centered approaches to change. *Developmental Psychology, 32*, 659–673.

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

Stukel, T.A., & Demidenko, E. (1997). Two-stage method of estimation for general linear growth curve models. *Biometrics, 53*, 720–728.

Thissen, D., & Bock, R.D. (1990). Linear and nonlinear curve fitting. In A. von Eye (Ed.), *Statistical methods in longitudinal research, Volume II: Time series and categorical longitudinal data* (pp. 289–318). Orlando, FL: Academic Press.

Twisk, J., & de Vente, W. (2002). Attrition in longitudinal studies: How to deal with missing data. *Journal of Clinical Epidemiology, 55*, 329–337.

van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98*, 3–52.

Vonesh, E.F., & Chinchilli, V.M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker.

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

Yuan, K.-H., & Bentler, P.M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology, 37,* 53–82.

Ken Kelley

Department of Management
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556
USA
E-mail kkelley@nd.edu