UC Davis

UC Davis Previously Published Works

Title

Fitting correlated residual error structures in nonlinear mixed-effects models using SAS PROC NLMIXED

Permalink

<https://escholarship.org/uc/item/67c3z2v7>

Journal

Behavior Research Methods, 46(2)

ISSN

1554-3528

Authors

Harring, Jeffrey R
Blozis, Shelley A

Publication Date

2013-10-09

DOI

10.3758/s13428-013-0397-z

Peer reviewed

Fitting correlated residual error structures in nonlinear mixed-effects models using SAS PROC NL MIXED

Jeffrey R. Harring · Shelley A. Blozis

Published online: 9 October 2013
© Psychonomic Society, Inc. 2013

Abstract Nonlinear mixed-effects (NLME) models remain popular among practitioners for analyzing continuous repeated measures data taken on each of a number of individuals when interest centers on characterizing individual-specific change. Within this framework, variation and correlation among the repeated measurements may be partitioned into interindividual variation and intraindividual variation components. The covariance structure of the residuals are, in many applications, consigned to be independent with homogeneous variances, $\sigma^2 \mathbf{I}_{n_i}$, not because it is believed that intraindividual variation adheres to this structure, but because many software programs that estimate parameters of such models are not well-equipped to handle other, possibly more realistic, patterns. In this article, we describe how the programmatic environment within SAS may be utilized to model residual structures for serial correlation and variance heterogeneity. An empirical example is used to illustrate the capabilities of the module.

Keywords Nonlinear mixed-effects model · Intraindividual variation · Serial correlation · Random effects · Subject-specific

Introduction

The intraindividual-level residuals in nonlinear mixed-effects (NLME) models represent measures of misfit between an individual's data and their own fitted function (see, e.g., Davidian & Giltinan, 2003). For numerous alternative residual

covariance structures that have been considered for mixed-effects models (see, e.g., Jennrich & Schluchter, 1986; Wolfinger, 1993a, 1996), when considered in conjunction with the covariance structure of the random effects, a very simple structure like the conditional-independence model (Laird & Ware, 1982) with homogeneous variances, $\sigma^2 \mathbf{I}_{n_i}$ is often adopted. This simple error structure is parsimonious and, in some instances, works quite well in summarizing the pattern of intraindividual variability for which individual fluctuations are of the same magnitude. One situation where this occurs with some frequency is when the lag between measurement occasions is substantial (e.g., yearly measurements). In other modeling contexts, within-individual measurements over time are likely to be positively correlated within an individual, with the strength of the correlation diminishing as observations are more separated in time. When this type of serial correlation, or variance heterogeneity, is suspected, a simple structure will be deficient and, even more worrisome, could lead to problems in testing fixed effects.

Two reasons support the justification of the statistical investment in selecting an adequate residual covariance structure. First, the choice of covariance structure impacts the fit of the model. In other words, a model with a particular response function may be viewed as adequate or inadequate, depending on the form of the associated covariance structure (Davidian & Giltinan, 2003). Second, the inability to adequately account for intraindividual variability can impact interindividual variance and covariance estimates, which, in turn, could lead to inefficiencies in inferential testing of population fixed effects. This has been demonstrated in a series of Monte Carlo simulation studies for linear mixed-effects models (Chi & Reinsel, 1989) and linear latent growth curve models (see, e.g., Ferron, Dailey, & Yi, 2002; Kwok, West, & Green, 2007; Sivo, Fan, & Witta, 2005), in which an autocorrelated residual error structure was used to generate data but a model specifying a conditional-independence structure was actually fitted. Despite evidence from studies that suggest that alternative structures are more appropriate in some cases, use of the

J. R. Harring (✉)
Measurement, Statistics and Evaluation, University of Maryland,
1230 Benjamin Building, College Park, MD 20742-1115, USA
e-mail: harring@umd.edu

S. A. Blozis
University of California, Davis, One Shields Avenue, Davis,
CA 95616, USA
e-mail: sablozis@ucdavis.edu

conditional-independence structure seems routine. This choice may be, in part, a function of limitations of current software regularly employed in the estimation of NLME models.¹ SAS PROC NLMIXED, a popular software module for fitting NLME models, is not, at first glance, equipped to fit error structures with greater complexity than the conditional-independence model. It is within this context that we demonstrate how the programmatic environment in SAS PROC NLMIXED can be utilized to fit intraindividual residual error structures that account for serial correlation and variance heterogeneity.

The rest of the article progresses in the following way. In the next section, the nonlinear mixed-effects model is briefly introduced, with an emphasis on the residual error structure. Subsequent sections will introduce the mathematical foundations needed to implement these methods in SAS PROC NLMIXED. An empirical example is provided to demonstrate how complex error structures that account for within-individual correlation and variance heterogeneity can be implemented.

A nonlinear mixed-effects model

We consider a slightly modified version of the NLME model described by Davidian and Giltinan (1995), which can be viewed as a hierarchical model. From this perspective, it has been demonstrated (see, e.g., Crowder & Hand, 1990; Cudeck, 1996) in some circumstances to subsume both the linear mixed-effects model (Laird & Ware, 1982) and the typical nonlinear regression model for independent data (Seber & Wild, 1989). In the individual-level (or level 1) model, the j th observation on the i th individual is modeled as

$$y_{ij} = f(t_{ij}, \beta_i, \mathbf{z}_i) + e_{ij} \quad i = 1, \dots, m \quad j = 1, \dots, n_i, \quad (1)$$

where f is a nonlinear function governing intraindividual behavior that depends on individual-specific parameter vector β_i , time t_{ij} , and possible person characteristics $\mathbf{z}_i$, characteristics (e.g., age, gender, treatment condition) specific to the i th individual. The total number of individuals is m , and n_i is the number of observations for individual i . The n_i subscript permits each individual to be measured at possibly different times. Without a loss of generality, we assume a balanced design where the timing of the repeated measures is common to the m individuals, yet allow for the possibility of

missing data (under the assumption that the missingness is MAR). Intraindividual deviations, $e_{ij} = y_{ij} - f(t_{ij}, \beta_i, \mathbf{z}_i)$, reflect uncertainty in the response of the i th individual at time t_{ij} and are assumed to satisfy $E[e_{ij}|\beta_i] = 0$ for all i and j . Conditioned on β_i , the variance of $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$ is captured in $\Theta_i(\beta_i, \xi)$, an $n_i \times n_i$ matrix function of individual-specific regression coefficients, β_i , and fixed-effects, ξ . This is typically referred to as the *intraindividual* or *level 1* covariance structure.

In an empirical example to follow, data were collected that described the improvement in task performance for a sample of individuals. Learning data like these tend to follow nonlinear monotonic trajectories that level off at later measurement occasions. Subjects' learning profiles can often be effectively modeled and may differ in initial status, asymptotic response, or rate of change. Due to the increasing nature of performance on the task coupled with a leveling-off at later trials, a logistic or exponential model (Browne, 1993) that includes an asymptote and random effects may be effective in characterizing the within-subjects response. A logistic function (cf. Browne, 1993), for example, may be formulated as

$$f(t_{ij}, \beta_i, \mathbf{z}_i) = \frac{\beta_{1i}\beta_{2i}}{[\beta_{1i} + (\beta_{2i} - \beta_{1i})\exp\{-\beta_{3i}(t_{ij} - 1)\}]}, \quad (2)$$

where β_{1i} represents initial performance of individual i at $t_{ij} = 1$. Potential performance is asymptotic and captured by β_{2i} and occurs after several occasions for large n_i . The rate governing change between initial and potential performance on the response, what Browne (1993) termed learning speed, is given by β_{3i} . The logistic function is inherently nonlinear, since at least one of the derivatives of the expectation function, f , with respect to the parameters depends on at least one of the parameters (Bates & Watts, 1988).

Because the individual regression coefficients characterizing f in Eq. 1 often correspond to scientifically relevant facets of the underlying change process, a primary goal of many NLME analyses is to attempt to understand individual differences in these variables. Toward that end, a submodel may be specified for each. At the population level (level 2), a general specification of an individual coefficient is a potentially nonlinear function of fixed parameters, (β) , covariates $(\mathbf{z}_i)$, and random effects $(\mathbf{b}_i)$ (see, e.g., Cudeck & Harring, 2007). For individual coefficient k , this is

$$\beta_{ki} = g_k(\mathbf{z}_i, \beta_k, \mathbf{b}_{ki}),$$

where g_k is a flexible function of the arguments and could be distinct for each coefficient. The number of individual coefficients, fixed effects, and random effects is usually unequal to allow for optimal flexibility in the form the coefficients take. Define these, respectively, as r , p , and q . Initially, NLME

¹ In contrast to linear mixed-effects models where a variety of programs are available that have options making it relatively easy for users to estimate level 1 error covariance structures (e.g., SAS PROC MIXED, SAS PROC GLIMMIX), few alternatives exist for fitting these same error structures in the context of NLME models.

models are typically fit with a simple additive model with no covariates

$$g_k(\mathbf{z}_i, \beta_k, b_{ki}) = \beta_k + b_{ki},$$

although individual covariates are often added at a later stage of the analysis to account for between-individual heterogeneity in the coefficients. In the population of individuals, it is almost always assumed that the random effects are normally distributed with null mean vector and covariance matrix Φ . This is called the *interindividual* or *level 2* covariance structure,

$$(b_{1i}, \dots, b_{ki})' \sim N(0, \Phi). \quad (3)$$

In the specific case of $q = 3$, like that of the logistic model in Eq. 2, the covariance matrix among random effects is

$$\Phi = \begin{pmatrix} \text{var}(b_{1i}) & \text{cov}(b_{2i}, b_{1i}) & \text{cov}(b_{3i}, b_{1i}) \\ \text{cov}(b_{2i}, b_{1i}) & \text{var}(b_{2i}) & \text{cov}(b_{3i}, b_{2i}) \\ \text{cov}(b_{3i}, b_{1i}) & \text{cov}(b_{3i}, b_{2i}) & \text{var}(b_{3i}) \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{21} & \varphi_{31} \\ \varphi_{21} & \varphi_{22} & \varphi_{32} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} \end{pmatrix}. \quad (4)$$

The diagonal elements of Φ are variances that summarize the extent to which the random effects are dispersed around zero. Off-diagonal elements of Φ describe the linear association between pairs of random effects. Lastly, it is assumed that the e_{ij} are independent of $\mathbf{b}_i$.

Within-individual covariance structures

Specifying a model for the intraindividual variation may often be an afterthought for many researchers, not only because the focus may be on parameters defining the change process itself, but also because the literature has not been very informative of how to examine this type of variation. Diggle, Heagerty, Liang, and Zeger (2001) provided a conceptualization of within-individual variation that was later discussed by and elaborated on by other authors (see, e.g., Davidian & Giltinan, 2003; Fitzmaurice, Laird, & Ware, 2011). Of the three sources of random variation identified by these authors, two define intraindividual variation: (1) serial correlation and (2) measurement error. Figure 1 depicts these two within-individual sources for the logistic model in Eq. 2 on task performance (y) over time (t) for a single individual i .

At least part of an individual's observed repeated measurements on an outcome may be a response to time-varying stochastic processes operating within the individual. For example, task performance may be influenced by, among other things, level of concentration, motivation, memory, or other behavioral traits that fluctuate randomly across time. This type of stochastic variation may result in a correlation between pairs of measurements on the same individual, the magnitude of which depends on the time lag between measurement pairs. This correlation becomes weaker as the lag between measurements increases.

Fig. 1 Graphical representation of two sources of within-individual variability in repeated measures for a hypothetical individual. The solid line is the “inherent” trajectory, the black dots denote repeated measures free of measurement error, and the open dot represents observed responses with measurement error

Measurement error is the other source of random variation usually thought to occur at the individual level, where the measurement process may itself add a component of variation to the data. Two measurements of some behavioral process taken simultaneously from the same individual would likely produce two different results. This would likely occur using an instrument that demonstrates imperfect reliability. Reliability is the consistency, or reproducibility, of an individual's score (often conceptualized as measuring a latent quantity) obtained over repeated simultaneous assessments. Scores gathered repeatedly from instruments with low reliability have attenuated correlations among the data and, thus, can influence the fit of the model.

Both of these sources are illustrated in Fig. 1. Here, the repeated measures of an individual are shown to vary around their true underlying response. If an instrument yields perfectly reliable scores, the hypothetical responses might follow the corresponding jagged curve and would produce error-free repeated measures represented in the figure as solid dots. Due to the imprecise nature of the instrument or assessment procedure, however, the repeated measures that were realized vary about the jagged line.

Modeling serial correlation and measurement error

Greater generality than a conditional-independence structure may be necessary to characterize complex patterns of within-individual variation that may arise with nonlinear repeated

measures (Davidian & Giltinan, 2003).² Thus, the modeling environment must be sufficiently flexible to accommodate both serial correlation and measurement error that may comprise the overall pattern of intraindividual variability.

To clarify how this may be accomplished, the intraindividual covariance structure is rewritten as

$$\Theta_i(\xi) = \Omega_i^{1/2}(\omega) \mathbf{P}_i(\rho) \Omega_i^{1/2}(\omega), \quad (5)$$

where $\Omega_i(\omega)$ is either a scalar or an $(n_i \times n_i)$ diagonal matrix. In both specifications, the elements are defined as $\text{var}(e_{ij}|\beta_i)$ depending on parameters ω . Correlation matrix $\mathbf{P}_i(\rho)$ has dimensions $(n_i \times n_i)$ with (j, j') elements defined as $\text{cor}(e_{ij}, e_{ij'}|\beta_i)$ depending on parameters, ρ .

For the upcoming example, we examine a number of autocorrelation error covariance structures, and thus, we define and show how each may fit into the general structure of Eq. 5. Assume four equally spaced measurement occasions that are common to all individuals. The covariance structure for the residuals, $\mathbf{e}_i$, thought of as arising from a first-order autoregressive process has general element

$$[\Theta_i]_{jk} = \sigma^2 \rho^{|j-k|} \quad \sigma^2 > 0; \quad 0 < \rho < 1,$$

and furthermore, could be decomposed as in Eq. 5:

$$\begin{aligned} \Theta_i(\xi) &= \Omega_i^{1/2}(\omega) \mathbf{P}_i(\rho) \Omega_i^{1/2}(\omega) \\ &= \sigma \cdot \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \cdot \sigma, \end{aligned} \quad (6)$$

where $\xi = (\omega, \rho)' = (\sigma^2, \rho)'$.

A compound symmetry (equicorrelation) structure is much like the first-order autoregressive structure, except that it specifies that the correlation between any two time points is the same no matter how far apart the measurements were taken. This structure is defined as

$$\begin{aligned} \Theta_i(\xi) &= \Omega_i^{1/2}(\omega) \mathbf{P}_i(\rho) \Omega_i^{1/2}(\omega) \\ &= \sigma \cdot \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho & \rho & 1 & \\ \rho & \rho & \rho & 1 \end{pmatrix} \cdot \sigma, \end{aligned} \quad (7)$$

² Davidian and Giltinan (2003) refer to residual structures that permit dependence on i through an individual's mean response, given β_i . Indeed, this provides a method of modeling the variances across time that may depend on an individual's trajectory. However, because the focus of this article is on modeling the serial correlation, from this point forward we will suppress the β_i notation and assume that the within-individual covariance structure $\Theta_i(\xi)$ may vary across individuals in terms of its dimension, but not otherwise.

where again, $\xi = (\omega, \rho)' = (\sigma^2, \rho)'$. Like the autoregressive structure in Eq. 6, a Toeplitz covariance structure specifies a different level of covariation for different time lags but restricts the level of covariation to be the same within any one time lag. This structure can be specified as

$$\begin{aligned} \Theta_i(\xi) &= \Omega_i^{1/2}(\omega) \mathbf{P}_i(\rho) \Omega_i^{1/2}(\omega) \\ &= \sigma \cdot \begin{pmatrix} 1 & & & \\ \rho_1 & 1 & & \\ \rho_2 & \rho_1 & 1 & \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix} \cdot \sigma, \end{aligned} \quad (8)$$

where $\xi = (\omega, \rho)' = (\sigma^2, \rho_1, \rho_2, \rho_3)'$. For simplicity, we suppress the parameter notation, ξ , from the intraindividual covariance structure, with the implicit understanding that the matrix is a function of fundamental parameters accounting for the variances and covariances.

Other autocorrelation structures, such as a banded Toeplitz tridiagonal structure (Schott, 1983; Sutradhar & Kumar, 2003), which specifies a particular correlation between adjacent measurements, whereas correlations between non-adjacent measurements are constrained to zero, are also possible. In the absence of theory, we recommend fitting and testing several covariance structures in the early stages of an analysis. However, specifying something other than a simple residual covariance structure for NLME models, like an autocorrelation structure, is not straightforward in many software packages, including SAS PROC NLMIXED. We demonstrate now how this can be accomplished.

The loglikelihood

An instructive way to begin discussing how to implement an autocorrelation structure in SAS PROC NLMIXED requires examining how the SAS module will execute the estimation of the model. Maximum likelihood, a relatively standard method of estimation for linear mixed-effects models, is much less common in the estimation of nonlinear models like that in Eq. 1, (Cudeck, 1996), because the marginal distribution of $\mathbf{Y}_i$ cannot be analytically derived. The likelihood is based on the joint density of response $\mathbf{Y}_i$ and random effects $\mathbf{b}_i$ given parameter ϑ and covariate $\mathbf{Z}_i$ and is expressed as

$$\begin{aligned} L(\vartheta|\mathbf{y}) &= \prod_{i=1}^m p(\mathbf{y}_i|\vartheta, \mathbf{z}_i) \\ &= \prod_{i=1}^m \int p(\mathbf{y}_i, \mathbf{b}_i|\vartheta, \mathbf{z}_i) d\mathbf{b}_i \\ &= \prod_{i=1}^m \int p(\mathbf{y}_i|\vartheta_f, \mathbf{b}_i, \mathbf{z}_i) p(\mathbf{b}_i|\vartheta_b) d\mathbf{b}_i, \end{aligned} \quad (9)$$

where overall parameter vector $\vartheta = (\vartheta_f, \vartheta_b)$ is decomposed into parameters associated with the conditional data distribution and those associated with the random effects density and where $p(\mathbf{y}_i | \vartheta_f, \mathbf{b}_i, \mathbf{z}_i)$ is the conditional density of $\mathbf{Y}_i$ given the parameters pertaining to the conditional distribution, ϑ_f , $\mathbf{b}_i$ and covariates $\mathbf{Z}_i$. The density $p(\mathbf{b}_i | \vartheta_b)$ is the marginal distribution of $\mathbf{b}_i$ given parameters pertaining to the random effects distribution, ϑ_b . Note that the likelihood is not a function of the random effects $\mathbf{b}_i$, since they are *marginalized* out of the expression through the integral. Furthermore, note that the likelihood is a function of the fixed parameter ϑ , and the goal of the maximum likelihood procedure is to find the single value of ϑ that maximizes the function and is, therefore, considered to be the value that most likely generated the data.

The integral in Eq. 9 generally does not have a closed-form expression if the model f is nonlinear in $\mathbf{b}_i$. Several approaches have been proposed that handle the intractability of the integral, and many of these methods are available options in SAS PROC NLMIXED. These include a method that circumvents dealing with the integration directly like a first-order linearization method (`method=firo` in SAS PROC NLMIXED) via a Taylor series expansion of the nonlinear function, f , around the expected value of the random effects (Beal & Sheiner, 1982; Wolfinger & Lin, 1997). Alternatively, advances in computational power have permitted techniques that maximize the likelihood in Eq. 9 directly using either a deterministic or a sto-

chastic approximation to handle the integral. If $p(\mathbf{b}_i | \vartheta_b)$ is a normal density, numerical approximation of the integral may be achieved by Gauss-Hermite quadrature. This is a standard deterministic method of approximating an integral by a weighted average of the integrand evaluated at suitably chosen points over a grid, where accuracy increases with the number of grid points (Davidian & Giltinan, 1993). As the integrals over $\mathbf{b}_i$ in Eq. 9 are q -dimensional, Pinheiro and Bates (2000) proposed an approach they referred to as adaptive Gaussian quadrature, where the grid for choosing the points of integration are centered around the mode of the random effects, $\hat{\mathbf{b}}_i$, and scaled in a way that allows suitable accuracy with fewer grid points to evaluate, thus decreasing the computational burden of maximizing the likelihood. The adaptive quadrature method is the default for approximating the integrals in SAS PROC NLMIXED, whereas the former Gauss-Hermite quadrature is obtained via the `method=gauss noad` option. For a more thorough discussion of the estimation approaches used by PROC NLMIXED, the interested reader is directed to Wolfinger and Lin (1997), Davidian and Giltinan (2003), and Pinheiro and Bates (1995, 2000).

Letting u_j and w_j for $j = 1, \dots, Q$ represent, respectively, the abscissas and weights for the one-dimensional Gaussian quadrature rule with Q points based on the kernel density $N(0,1)$, the loglikelihood under the nonadaptive quadrature approach is specified according to Pinheiro and Bates (1995) as

$$l(\vartheta | \mathbf{y}) = -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln |\Theta_i| + \sum_{i=1}^m \ln \left\{ \sum_{j=1}^Q \exp \left[-\frac{1}{2} \left(\mathbf{y}_i - f(\beta, \Phi^{T/2} u_j) \right)' \Theta_i^{-1} \left(\mathbf{y}_i - f(\beta, \Phi^{T/2} u_j) \right) \right] \prod_{j=1}^Q w_j \right\}. \quad (10)$$

Implementation in SAS

The default intraindividual covariance structure fit in SAS PROC NLMIXED is conditional independence, $\Theta_i = \sigma^2 \mathbf{I}_{n_i}$. To fit an autocorrelation structure requires utilizing the programming capabilities within the module to construct the loglikelihood function. The MODEL statement using the `general()` option³ allows the user the greatest flexibility in building a loglikelihood from elemental pieces comprising the mean function and the covariance structure. The random effects distribution is assumed to be normal with zero mean

vector and unstructured covariance matrix Φ , as in Eq. 4. The MODEL statement specifies the dependent variable and its conditional distribution given the random effects. For the present problem, the conditional distribution, $p(\mathbf{y}_i | \vartheta_f, \mathbf{b}_i, \mathbf{z}_i)$, is normal and has density

$$p(\mathbf{y}_i | \vartheta_f, \mathbf{b}_i, \mathbf{z}_i) = (2\pi)^{-\frac{n_i}{2}} |\Theta_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_i - f(t_{ij}, \beta_i, \mathbf{z}_i)]' \Theta_i^{-1} [\mathbf{y}_i - f(t_{ij}, \beta_i, \mathbf{z}_i)] \right\},$$

where Θ_i is defined as in Eq. 5, in which the variances of the level 1 residuals and a matrix modeling serial correlation are separable. The importance of separating these facets

³ Only quadrature methods are allowed within PROC NLMIXED when building the loglikelihood function using the `general()` option.

making up the overall structure is that both the inverse and determinant of Θ_i can be constructed by combining the inverse and determinant of the variance matrix and correlation matrix, which for autocorrelation structures have a relatively simple form.

Inverse and determinant of Θ_i

Building a loglikelihood function requires both the inverse and determinant of the intraindividual covariance structure, Θ_i , that depends on both Ω_i and $\mathbf{P}_i$. Fortunately, for patterned matrices, like an autoregressive, banded Toeplitz or compound symmetric structure, the inverse and determinant of the matrix $\mathbf{P}_i$ have a form involving few parameters, and thus, determining the inverse and determinant of Θ_i can be done in a straightforward manner. We examine five covariance structures in the subsequent analysis: (1) conditional independence, (2) heterogeneous variances, (3) first-order autoregressive, (4) compound symmetry, and (5) symmetric tridiagonal Toeplitz. The inverse and determinant for each of these structures will be given next.

Conditional independence

For the $n_i \times n_i$ matrix Θ_i having a conditional independence structure, define $\Omega_i = \sigma^2 \mathbf{I}_{n_i}$ and $\mathbf{P}_i = \mathbf{I}_{n_i}$, where $\mathbf{I}$ is an identity matrix of dimension n_i . Then, the inverse and determinant are

$$\begin{aligned}\Theta_i^{-1} &= \Omega_i^{-1} = \sigma^{-2} \mathbf{I}_{n_i} \\ |\Theta_i| &= |\Omega_i| = (\sigma^2)^{n_i}.\end{aligned}$$

Heterogeneous variances

For the $n_i \times n_i$ matrix Θ_i that permits distinct variances at each time point, define $\Omega_i = \sigma_j^2 \mathbf{I}_{n_i}$ and $\mathbf{P}_i = \mathbf{I}_{n_i}$. Consequently, the inverse and determinant are given as

$$\begin{aligned}\Theta_i^{-1} &= \Omega_i^{-1} = \sigma_j^{-2} \mathbf{I}_{n_i} \\ |\Theta_i| &= |\Omega_i| = \sigma_1^2 \times \cdots \times \sigma_{n_i}^2.\end{aligned}$$

Autoregressive structure

Another type of autocorrelation structure that appears often in the literature is a first-order autoregressive structure like that previously defined in Eq. 6 with general element, $[\Theta_i]_{jk} = \sigma^2 \rho^{|j-k|}$. Graybill (1983) showed that the inverse of this patterned matrix is $\Theta_i^{-1} = \Omega_i^{-1} \times \mathbf{P}_i^{-1}$. Grenander and Szegö (1958) noted that $\mathbf{P}_i$ has a simple tridiagonal form and is a function of one parameter, ρ ; thus, its inverse will also have a simple form—namely,

$$\mathbf{P}_i^{-1} = (1-\rho^2)^{-1} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}. \quad (11)$$

Note that diagonal elements 1 and n_i are $1/(1-\rho^2)$, diagonal elements $2, \dots, (n_i - 1)$ are $(1+\rho^2)/(1-\rho^2)$, and the subdiagonal element is $-\rho/(1-\rho^2)$. If $\rho = 0.5$ and $n_i = 4$, for example, then $\mathbf{P}_i$ and $\mathbf{P}_i^{-1}$ would be

$$\mathbf{P}_i = \begin{pmatrix} 1 & .5 & .25 & .125 \\ .5 & 1 & .5 & .25 \\ .25 & .5 & 1 & .5 \\ .125 & .25 & .5 & 1 \end{pmatrix} \quad \mathbf{P}_i^{-1} = \begin{pmatrix} 1.333 & -.667 & 0 & 0 \\ -.667 & 1.667 & -.667 & 0 \\ 0 & -.667 & 1.667 & -.667 \\ 0 & 0 & -.667 & 1.333 \end{pmatrix}.$$

Graybill (1983) also showed that the determinant of the covariance matrix, Θ_i , could be computed as $|\Theta_i| = (\sigma^2)^{n_i} (1-\rho^2)^{n_i-1}$.

Compound symmetric structure

Let the $n_i \times n_i$ matrix Θ_i have a compound symmetric structure with $\Omega_i = \sigma^2 \mathbf{I}_{n_i}$ and $\mathbf{P}_i = (1-\rho) \mathbf{I}_{n_i} + \rho \mathbf{J}_{n_i}$, where $\mathbf{J}$ is a matrix of ones of dimension n_i . The inverse of Θ_i is the product of the inverses Ω_i^{-1} and $\mathbf{P}_i^{-1}$. The inverse $\mathbf{P}_i^{-1}$, a function of a single parameter ρ , from Rao (1973) is given by $\mathbf{P}_i^{-1} = (a-b) \mathbf{I}_{n_i} + b \mathbf{J}_{n_i}$, where $a = \{1 + (n_i - 2)\rho\} / [(1-\rho)\{1 + (n_i - 1)\rho\}]$ and $b = -\rho / [(1-\rho)\{1 + (n_i - 1)\rho\}]$. Then the jk -th element is given by

$$[\Theta_i^{-1}]_{jk} = \sigma^{-2} [\mathbf{P}_i^{-1}]_{jk}.$$

Sutradhar and Kumar (2003) showed that the determinant of Θ_i is the product of the determinants $|\Omega_i| |\mathbf{P}_i|$ and can be expressed as

$$\begin{aligned}|\Theta_i| &= |\Omega_i| |\mathbf{P}_i| \\ &= (\sigma^2)^{n_i} (1-\rho)^{n_i} (1 + \rho(n_i - 1)).\end{aligned} \quad (12)$$

Symmetric banded Toeplitz structure

The banded Toeplitz structure in Eq. 8 allows the covariances to differ between residuals spaced with different time lags. Unfortunately, the inverse of this structure does not have a form that translates into the programmable language needed in SAS NLMIXED in a straightforward manner. However, the symmetric tridiagonal Toeplitz structure, which specifies a covariance for measurements separated by a single time lag only, does. The symmetric tridiagonal Toeplitz structure has a general diagonal element σ^2 and subdiagonal and superdiagonal elements $[\Theta_i]_{jk} = \sigma^2 \rho$ for $|j-k| = 1$ and $[\Theta_i]_{jk} = 0$ for $|j-k| > 1$. The inverse of this matrix can be

Table 1 Sample mean vector and covariance matrix for learning task scores across nine measurement occasions, $m = 140$

Variables	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	Mean
$y_{t=1}$	94.32									20.12
$y_{t=2}$	80.21	94.24								25.52
$y_{t=3}$	73.41	88.35	96.42							29.32
$y_{t=4}$	63.07	77.49	82.47	86.67						32.40
$y_{t=5}$	53.73	64.26	70.11	71.12	74.20					34.19
$y_{t=6}$	44.14	51.01	54.83	56.85	58.77	59.58				35.60
$y_{t=7}$	35.44	41.39	46.35	45.82	50.26	46.49	54.56			37.73
$y_{t=8}$	40.13	44.42	46.44	44.39	48.69	46.18	49.68	60.96		39.03
$y_{t=9}$	35.75	39.45	42.57	39.71	43.26	43.26	47.46	54.96	66.66	38.79

formulated as the product of inverses $\Theta_i^{-1} = \Omega_i^{-1} \times \mathbf{P}_i^{-1}$, where $\mathbf{P}_i^{-1}$ was derived considering a moving average

process of order 1, MA(1), by Sutradhar and Kumar (2003), and whose jk -th element is given as

$$[\mathbf{P}_i^{-1}]_{jk} = \left\{ \frac{1 + \theta^2}{1 - \theta^2} \right\} \left[\left\{ \theta^{|j-k|} - \theta^{2n-j-k+2} \right\} - \frac{\theta^{j+k}}{1 - \theta^{2n-2}} \left\{ (1 - \theta^{2n-2j-2})(1 - \theta^{2n-2k-2}) \right\} \right],$$

where $-1 < \theta < 1$ is the parameter of the process and $\rho = -\theta/(1 + \theta^2)$. The determinant of Θ_i depends on n_i and is the product of the determinants $|\Omega_i| |\mathbf{P}_i|$. We now fit these structures to the NLME model for data from a learning study. Annotated SAS NLMIXED code for fitting the NLME model with the AR(1) error structure is located in the Appendix. An additional SAS script to estimate a NLME model with the five previously outlined level 1 error structures may be found in the online Appendix at <http://education.umd.edu/EDMS/fac/Harring/Misc/BRM-Supplement-Final.sas>.

Task performance

To illustrate how to estimate an NLME model that includes one of several possible autocorrelation residual error structures using SAS PROC NLMIXED, data are presented from a computerized learning task that simulated the activities of an air traffic controller. These data have been used in past methodological articles (see, e.g., Browne, 1993; Choi, Harring, & Hancock, 2009) to highlight statistical methods and models that incorporate nonlinear functions. Kanfer and Ackerman (1989) developed a computerized learning task simulating the duties of an air traffic controller. Their primary objective was to examine the performance of subjects to safely bring in planes. The task was continuous, and the response variable represented the number of planes brought in to land safely every 10 min. Subjects were allowed 10-min breaks following completion of each set of three subsequent trials after the initial trial (i.e., after trial 4 and trial 7) to minimize massed

practice effects. Subjects were not permitted to confer with one another during the breaks. The researchers administered this task to multiple samples, and the sample employed here consists of $m = 140$ subjects.

Scores were recorded from the task continuously for a period of 100 min, yielding 10 scores; however, scores from the first trial were discarded, allowing for an adjustment period to the task. The sample covariance matrix and mean vector of the individual-level responses across the nine time points, labeled for the purposes of this example as $t = 1$ through $t = 9$, are shown in Table 1. Figure 2 shows a 20 % sample of randomly selected individual profiles. It is clear that

Fig. 2 Spaghetti plot of a 20 % random sample of individuals

Table 2 Measures of model fit and number of parameters for alternative functions fitted to the empirical means using nonlinear least squares regression and nonlinear mixed-effects model fitted with a first-order linearization algorithm

Nonlinear Least Squares Regression				
Within-Individual Function		Parameters	MSR	R^2
Linear:	$f_j = \beta_1 + \beta_2 t_j$	2	4.192	.912
Quadratic:	$f_j = \beta_1 + \beta_2 t_j + \beta_3 t_j^2$	3	0.348	.994
Cubic:	$f_j = \beta_1 + \beta_2 t_j + \beta_3 t_j^2 + \beta_4 t_j^3$	4	0.252	.995
Exponential-1:	$f_j = \beta_2 - (\beta_2 - \beta_1) \exp\{-\beta_3 t_j\}$	3	0.174	.997
Exponential 2:	$f_j = \beta_2 - (\beta_2 - \beta_1) \cdot 2^{\left(\frac{-t_j}{\beta_3}\right)}$	3	0.174	.997
Logistic:	$f_j = \beta_1 \beta_2 / [\beta_1 + (\beta_2 - \beta_1) \exp\{-\beta_3 t_j\}]$	3	0.315	.994
Gompertz:	$f_j = \beta_2 \exp[\ln(\beta_1/\beta_2) \exp\{-\beta_3 t_j\}]$	3	0.227	.996
Linear-Linear:	$f_j = \beta_1 + \beta_2 t_j + (\beta_2 - \beta_3)(\beta_4 - t_j)_+$	4	0.626	.991
Nonlinear Mixed-Effects Model				
Within-Individual Function		Parameters	$-2\ln L$	AIC
Linear:	$f_{ij} = \beta_1 + \beta_2 t_{ij}$	6	7,873.4	7,885.4
Quadratic:	$f_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2$	10	7,498.0	7,518.0
Cubic:	$f_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 t_{ij}^3$	15	7,743.7	7,773.7
Exponential 1:	$f_{ij} = \beta_2 - (\beta_2 - \beta_1) \exp\{-\beta_3 t_{ij}\}$	10	7,439.5	7,459.5
Exponential 2:	$f_{ij} = \beta_2 - (\beta_2 - \beta_1) \cdot 2^{\left(\frac{-t_{ij}}{\beta_3}\right)}$	10	7,439.5	7,459.5
Logistic:	$f_{ij} = \beta_1 \beta_2 / [\beta_1 + (\beta_2 - \beta_1) \exp\{-\beta_3 t_{ij}\}]$	10	7,432.0	7,452.0
Gompertz:	$f_{ij} = \beta_2 \exp[\ln(\beta_1/\beta_2) \exp\{-\beta_3 t_{ij}\}]$	10	7,431.9	7,451.9
Linear-Linear:	$f_{ij} = \beta_1 + \beta_2 t_{ij} + (\beta_2 - \beta_3)(\beta_4 - t_{ij})_+$	15	7,488.1	7,518.1

the rate of increase in learning for a majority of individuals is not constant, with the rate of increase more rapid at earlier trials and leveling off at later trials. Thus, a nonlinear function that attends to the curvilinearity in the data needs to be ascertained.

Analysis and results

Characterizing within-individual behavior, such as learning, with a fairly simple mathematical function can be as much art as science. Still, there are a number of approaches to narrowing down the number of alternative models. One method involves fitting several candidate functions to each individual's data, separately. This may be accomplished by conducting a series of nonlinear least squares (NLS) analyses. As convergence to a proper solution for NLS problems is known to be susceptible to good starting values of the parameters, initial coefficients can be chosen on a case-by-case basis using the Nelder–Mead simplex algorithm (Nelder & Mead, 1965). In addition, the empirical means should also be fitted by the candidate functions deemed adequate at the individual level. Of course, both of these methods do not take into account the dependencies in the data; therefore, it may be prudent to also fit several candidate functions to the repeated measure data as well.

Table 2 shows the fit of several candidate functions to the empirical means of the learning data via NLS measured by conventional mean square residual (MSR) and R^2 . A first-order linearization method was used to examine the fit of these functions as well, with deviance ($-2\ln L$) and AIC (Akaike, 1974) recorded for each. On the basis of MSR and R^2 , several functions fit the empirical means equally well. The same conclusion could be reached regarding the fit of nonlinear mixed-effects models, although on the basis of AIC values, some functions clearly fit the repeated measures data better than do others. In addition to model fit, deciding on a final function may hinge upon other theoretical considerations or on the appropriateness of interpretation

Table 3 Alternative within-individual covariance structures for the logistic NLME model

Covariance Structure	Parameters	$-2\ln L$	AIC
Conditional independence	10	7,410.7	7,430.7
Heterogeneous variances	18	7,379.1	7,415.1
Compound symmetry	11	7,372.4	7,394.4
Banded Toeplitz	11	7,321.5	7,343.5
First-order autoregressive	11	7,300.0	7,322.0

Table 4 Maximum likelihood estimates and standard errors for the logistic NLME model with first-order autoregressive level 1 error structure

Parameter	Estimate	SE	<i>p</i> -value
$\hat{\beta}_1$	16.94	0.40	<.001
$\hat{\beta}_2$	38.82	0.69	<.001
$\hat{\beta}_3$	0.75	0.05	<.001
$\hat{\varphi}_{11}$	125.20	6.28	<.001
$\hat{\varphi}_{21}$	56.48	8.47	<.001
$\hat{\varphi}_{22}$	61.32	9.10	<.001
$\hat{\varphi}_{31}$	-2.51	0.47	<.001
$\hat{\varphi}_{32}$	-1.57	0.35	<.001
$\hat{\varphi}_{33}$	0.13	0.02	<.001
$\hat{\sigma}^2$	13.04	1.02	<.001
$\hat{\rho}$	0.41	0.05	<.001

of the coefficients. We decided on the logistic function because its pattern closely aligns with the theoretic pattern of learning that occurs during skill acquisition operationalized in the task performance.

The population model for all models was defined with all regression coefficients to be the sum of fixed and random effects:

$$g_k(\mathbf{z}_i, \beta_k, b_{ki}) = \beta_k + b_{ki} \text{ for } k = 1, 2, 3.$$

Several intraindividual covariance structures were fitted to the data using the methods outlined previously with the results summarized in Table 3. Relative fit of the models was assessed using AIC. Relatively small AIC values denote better fit. Among the pool of alternative covariance structures, the first-order autoregressive structure fit the best (AIC = 7,322.0) and was adopted for the remainder of this analysis. Maximum likelihood estimates and their accompanying standard errors for the final model⁴ are summarized in Table 4.

The results indicate that the initial performance level for the typical individual at $t_{i1} = 1$ is a little under 17 planes brought in safely $[\hat{\beta}_1 = 16.94(0.40)]$. Potential performance for the typical individual, as $t_{ij} \rightarrow \infty$, is a little under 40 planes brought in safely $[\hat{\beta}_2 = 38.82(0.69)]$. The rate of change (or learning speed) from initial to potential performance is $[\hat{\beta}_3 = 0.75(0.05)]$. Individual differences in the three facets of performance can be ascertained by examining the variances and covariances of the random coefficients. From $\hat{\Phi}$, the variance estimates of the random coefficients are all statistically

significant at the nominal $\alpha = 0.05$. The correlation between initial performance and growth rate is

$$\text{corr}(b_{1i}, b_{3i}) = \frac{-2.51}{\sqrt{125.20 \cdot 0.13}} = -0.62,$$

and indicates a moderately strong trend that those individuals who began with lower task performance scores tended to increase (learn) at a faster rate than do those individuals whose initial status was higher. Correlations between the other random effects could be computed and interpreted in a similar manner.

Modeling longitudinal data with subject-specific models like the NLME model is valuable when individual trajectories vary considerably and focus is on the individuals. To give some indication of how well the model fit, it is instructive to examine the fitted functions of individuals. Figure 3 displays the fitted functions for six individuals whose within-individual behavior differs markedly. The fitted trajectory for the typical individual is superimposed to highlight individual differences in profiles.

Discussion

In this article, we demonstrated how several autocorrelation within-individual covariance structures could be fitted to data for an NLME model using the SAS PROC NLMIXED module. Because this necessarily requires the user to construct the loglikelihood function, the `general()` specification must be utilized, which itself requires the use of Gaussian quadrature to handle the multidimensional integration. Other methods to facilitate maximum likelihood estimation, such as linearization approaches, may also be used to fit NLME models with autocorrelated errors and heterogeneous variances. These methods are appealing because they rely on procedures that are commonly implemented for the estimation of purely linear models (Davidian & Giltinan, 1995). In the R program, for example, the `nlme` function in the package with the same name fits NLME models in the formulation described in Lindstrom and Bates (1990). Additionally, a SAS macro, NLINMIX, that was developed for the estimation of NLME models using generalized estimating equations implements the nonlinear regression procedure PROC NLIN, along with the linear mixed-effects models procedure PROC MIXED (Littell, Milliken, Stroup, & Wolfinger, 1996; Wolfinger, 1993b). An advantage of using the SAS macro is its reliance on PROC MIXED, which accommodates a very wide range of covariance structures, including those discussed in this article. An obvious drawback, however, to these linearization methods

⁴ The final model comprised the logistic function defined in Equation 2, an additive level 2 model of fixed and random effects, and a first-order autoregressive error structure for the level 1 residuals.

Fig. 3 Fitted logistic function for six selected individuals (—). The trajectory for the typical individual (···) was superimposed. The NLME model was fitted with $Q = 20$ quadrature points

is that the Taylor series approximation may be poor or if n_i is not large. Additionally, problems may exist if intraindividual variation is large, as compared with interindividual variation. These situations do occur with some frequency in the social and behavioral sciences in studies attempting to quantify human behavior or attributes. The accuracy of methods of estimation that directly handle the integral, such as Gauss-Hermite quadrature used by PROC NLMIXED, are not dependent upon n_i and can be made arbitrarily accurate at the expense of greater computational intensity.

Correlated residuals in NLME models can be scientifically meaningful, since they can represent carryover effects that were not accounted for by the random effects covariance structure. This may occur more readily in situations where the lag between measurement occasions is short. In general, both intra- and interindividual terms contribute to the overall pattern

of correlation among responses on the same individual. It is important to recognize that intraindividual variance and correlation are relevant even if scope of inference is limited to a given individual only. As has been noted by Diggle et al. (2001) and Verbeke and Molenberghs (2000), in some applications, the preponderance of variation in the repeated measures [i.e., $\text{var}(\mathbf{y}_i|\mathbf{z}_i)$] is attributable to interindividual variation. This may explain why many published applications of NLME models adopt simple, diagonal models for $\Theta_i(\beta_i, \xi)$ that emphasize measurement error (i.e., $\sigma^2 \mathbf{I}_{n_i}$). In this scenario, how one models intraindividual correlation or, imprudently, disregards it may have a negligible impact on inference (Davidian & Giltinan, 1995). The responsibility falls on the shoulders of the data analyst to evaluate critically the rationale for and consequences of adopting a simplified model in a particular application. In situations where a simple residual covariance structure is not

realistic, we have shown how the programmatic environment in SAS PROC NLMIXED can be used to fit NLME models that account for intraindividual serial correlation and variance heterogeneity above and beyond what can be accounted for by the random effects covariance structure.

Acknowledgements We would like to thank Dr. Robert Cudeck for his thoughtful discussion of early versions of the manuscript, especially the coding of missing data in SAS.

Appendix

SAS code for running the NLME model with a logistic function and first-order autoregressive within-individual covariance structure. Note that the data are read in by SAS NLMIXED in the wide (cases by variables) format, which is atypical of how data for NLME models are usually used by the program. The variables to be read in for each subject are the responses, Y_i , time t_i , and individual identification variable *subj*.

```
proc nlmixed data=acker noad qpoints=20 tech=newrap maxit=1000;

#The proc nlmixed statement includes information regarding
the estimation method, optimization algorithm, iteration
and convergence criteria

array y{*} y1-y9;
array x{*} t1-t9;
array e{*} e1-e9;

#Setting up placeholders for responses, time, and residuals

parms b1 39.7 b2 20 b3 .36
      s1 50
      c21 26 s2 80
      c31 -.9 c32 -.6 s3 .09
      rho .48 se 10;

#Starting values for the parameters

bounds s1 >= 0, s2 >= 0, s3 >= 0, se >= 0;

#Using the bounds statement to constrain variance
parameters to be greater or equal to zero

random u1 u2 u3 ~ normal([0,0,0],[s1,c21,s2,c31,c32,s3])
subject=subj;

#Random statement to specify the random effects distribution

ni=9;
ln2pi = 1.8378770664;
rho2 = rho * rho;
```

```

ka = 1 / (1 - rho2);
kb = -ka * rho;
kc = ka * (1 + rho2);

#Quantities making up the elements of P#

yh = ((b1+u1)*(b2+u2))/((b1+u1)+
      ((b2+u2)-(b1+u1))*exp(-(b3+u3)*x[1]));
e[1] = y[1] - yh;

w2=0; w3=0;
do j = 2 to ni;
  yh = ((b1+u1)*(b2+u2))/((b1+u1)+
      ((b2+u2)-(b1+u1))*exp(-(b3+u3)*x[j]));
  e[j] = y[j] - yh;
  w2 = w2 + e[j]*e[j-1];
  w3 = w3 + e[j]*e[j];
end;

#Loop to compute the lag 1 residual sum of squares e[j]*e[j-1]
and usual residual sum of squares e[j]*e[j]

w1 = e[1] * e[1] + e[ni] * e[ni];
w3 = w3 - e[ni]*e[ni];
quad = (ka*w1 + 2*kb*w2 + kc*w3) / se;

#Building up the quadratic term which includes the residual
sum of squares and inverse the within-subject covariance
structure

lndet = ni*log(se) + (ni-1) * log(1-rho2);

#Computing the natural log of the determinant of the
within-subject covariance structure

Li = -0.5 * (ni * ln2pi + lndet + quad);

#Calculating the contribution of the loglikelihood
for individual i

dm=1; model dm ~ general(Li);
run;
#Using the general ( ) notation to specify a user-specific
loglikelihood

```


References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *AC-19*, 716–723.
- Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.
- Beal, S. L., & Sheiner, L. B. (1982). Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering*, *8*, 195–222.
- Browne, M. W. (1993). Structured latent curve models. In C. M. Cuadras & R. Rao (Eds.), *Multivariate analysis: Future directions* (2nd ed., pp. 171–197). Amsterdam: Elsevier Science.
- Chi, E. M., & Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*, *84*, 452–459.
- Choi, J., Harring, J. R., & Hancock, G. R. (2009). Latent growth modeling for logistic response functions. *Multivariate Behavioral Research*, *44*, 620–645.
- Crowder, M. J., & Hand, D. J. (1990). *Analysis of repeated measures*. London: Chapman & Hall.
- Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures. *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. (1993). Some simple methods for estimating intra-individual variability in nonlinear mixed effects models. *Biometrics*, *49*, 59–73.
- Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. Boca Raton, FL: Chapman & Hall.
- Davidian, M., & Giltinan, D. M. (2003). Nonlinear models for repeated measures data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, *8*, 387–419.
- Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2001). *Analysis of longitudinal data* (2nd ed.). Oxford: Oxford University Press.
- 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.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: Wiley.
- Graybill, F. A. (1983). *Matrices with applications in statistics* (2nd ed.). Belmont, CA: Wadsworth.
- Grenander, U., & Szegő, G. (1958). *Toeplitz forms and their applications*. Los Angeles, CA: University of California Press.
- Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, *42*, 805–820.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657–690.
- 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, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.
- Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, *46*, 673–687.
- Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute Inc.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.
- Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the loglikelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics*, *4*, 12–35.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-plus*. New York: Springer-Verlag.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
- Schott, J. R. (1983). *Matrix analysis for statistics* (2nd ed.). Hoboken, NJ: Wiley.
- Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.
- Sivo, S., Fan, X., & Witta, L. (2005). The biasing effects of unmodeled ARMA time series processes on latent growth curve model estimates. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 215–231.
- Sutradhar, B. C., & Kumar, P. (2003). The inversion of a correlation matrix for MA(1) process. *Applied Mathematics Letters*, *16*, 317–321.
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.
- Wolfinger, R. D. (1993a). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation*, *22*, 1079–1106.
- Wolfinger, R. D. (1993b). Laplace's approximation for nonlinear mixed models. *Biometrika*, *80*, 791–795.
- Wolfinger, R. D. (1996). Heterogeneous variance: Covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*, 205–230.
- Wolfinger, R. D., & Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis*, *25*, 465–490.