

Semiparametric hierarchical model with heteroscedasticity

CHUOXIN MA, MAOZAI TIAN, AND JIANXIN PAN*

Recent work on hierarchical data analysis mainly focuses on the multilevel structure of the mean response. Little research for hierarchical heteroscedasticity was done in the literature. In this paper, we propose a class of hierarchical models with heteroscedasticity and then investigate the semi-parametric statistical inferences. Laplace's approximation is employed to obtain an approximated marginal likelihood function and splines method is used to estimate the unknown functions. We also provide the consistency of the estimators. Simulation studies and real data analysis show that the proposed estimation procedures work well.

KEYWORDS AND PHRASES: Heteroscedasticity, Hierarchical models, Semiparametric inference, Laplace's approximation.

1. INTRODUCTION

Hierarchical models are effective in modelling repeated measurements, longitudinal data, clustered data and hierarchical data. Applications in different scientific research areas have led to several specific models, such as mixed-effects models, multilevel models and covariance component models. A great deal of theoretical results of these models can be found in the literature and most of them focus on the estimation of the mean functions. Though several research suggested the existence of heterogeneity in practical application [15, 9, 20, 10] and the necessity to introduce a heteroscedasticity function connecting the intra-individual variation with certain covariates [29], statistical inferences for heteroscedasticity in hierarchical data are less well addressed.

Our research is motivated by an indomethacin analysis. It is based on a pharmacokinetic research in Kwan [13] where the plasma concentration of indomethacin after bolus intravenous injection was investigated to study the metabolic disposition of indomethacin. The data came from an experimental study in which participants received single doses of indomethacin and their serial plasma concentration samples were measured repeatedly at 11 time points subsequently. Kwan et al. [13] used two-compartment open models to analysis the metabolic process. Davidian and Giltinan [5] carried

out subsequent analysis and found that heterogeneous variability is evident for the indomethacin data through studentized residuals plot. To account for the heterogeneity, Davidian and Giltinan [5] extended the classic random coefficient models by adding a heteroscedasticity function as follows,

$$\begin{aligned} Y_{ij} &= f(\mathbf{X}_{ij}, \boldsymbol{\beta}_i) + \sigma g\{f(\mathbf{X}_{ij}, \boldsymbol{\beta}_i), \boldsymbol{\theta}\} \varepsilon_{ij}, \\ \boldsymbol{\beta}_i &\stackrel{iid}{\sim} N(\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, 1), \end{aligned}$$

where f and g are known functions and $\boldsymbol{\theta}$ is a vector of parameters. Their analysis reveals an increase relationship between variance and mean response.

Modelling the heterogeneity and identifying covariates that are related to variance can fully characterize the intra-individual variation [5] and provide a better understanding of the research problems [16]. Studying the heteroscedasticity of the model is not only of practical interest, but is also of important theoretical significance. It is well known that ignoring the heteroscedasticity will lead to a great loss of efficiency of the fixed effects estimators. To improve the performance of the estimators for the mean functions, the heterogeneity of variance must be adequately modelled [7]. Heteroscedasticity particularly exists in data which is best fit by a nonlinear model [1]. Hence when constructing nonlinear models for hierarchical data, nonhomogeneous intra-individual variances should be addressed.

Several approaches were suggested to analyse the heteroscedasticity in hierarchical models. Davidian and Giltinan [5] extended the generalized least squares (GLS) estimator by pooling information across individuals to estimate the intra-individual variability in nonlinear mixed effects models. In another research, they introduced several estimation methods for nonlinear mixed models with heterogeneity [4], including two stage method and linearization method. Vonesh [26] also extended several estimation procedures for nonlinear mixed models in which the variance was related to the mean response. Lin, et al. [16] estimated the regression coefficients, variance components, and heterogeneity parameters in linear mixed models based on quasi-likelihood and method of moments. Other researchers employed Bayesian approaches [7, 15] and quantile regression methods [29] to assess the heterogeneity of residual variances in mixed models and multilevel models. Cao and Lin [2] further studied the hypothesis test for variance heterogeneity and autocorrelation in longitudinal nonlinear models.

*Corresponding author.

However, most models mentioned above were assumed to be parametric, which is restrictive in practical application. Model misspecification can lead to incorrect inferences and misleading conclusions. This problem is especially serious when modelling the heterogeneity as there is usually less mechanistic theories or prior information for the analysis of variance. Thus a more robust and flexible method such as nonparametric or semiparametric regression should be considered.

In this article, we concentrate on the semiparametric statistical inference of the hierarchical models with heteroscedasticity. The rest of this paper is organised as follows. In Section 2 we introduce the hierarchical models with heteroscedasticity and in Section 3 we give the estimation procedures based on Laplace's approximation. We also present the large sample properties of the proposed estimators. Section 4 gives a simulation study to evaluate the performance of our estimation method. A real analysis based on the in-domethacin data set is conducted in Section 5. In Section 6, we make conclusions and provide discussion of the proposed models and estimation method.

2. HIERARCHICAL MODELS WITH HETEROSCEDASTICITY

Assume that we have observations $\{(Y_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij}, T_{ij}, S_{ij}, \mathbf{W}_i, \mathbf{K}_i), i = 1, \dots, m, j = 1, \dots, n_i\}$ where Y_{ij} is the response value of the j th observation in subject i . \mathbf{X}'_{ij} s and \mathbf{W}'_i s are known p by 1 vectors of level-1 predictors and p by r matrices of level-2 predictors in the mean functions, respectively. \mathbf{Z}'_{ij} s and \mathbf{K}'_i s are known q by 1 vectors of level-1 predictors and q by l matrices of level-2 predictors in the heteroscedasticity function, respectively. In practical applications where data are organized hierarchically, it is straightforward to see which explanatory variables are level-1 predictors and which ones are level-2 predictors. For example, in medical research, patients may be grouped into clinics or hospitals. In this case, measurements from individuals like weight or blood pressure are considered to be level-1 predictors while characteristics describing the large unit such as the capacity of hospital are considered to be level-2 predictors. T_{ij} and S_{ij} are covariates in the unknown functions. For simplicity, they are considered as scalars in the rest of the paper. The relationship of Y_{ij} and the first level predictors \mathbf{X}_{ij} , \mathbf{Z}_{ij} , T_{ij} and S_{ij} can be presented according to the following equation:

$$(1) \quad Y_{ij} = f(\mathbf{X}_{ij}; \boldsymbol{\beta}_i, \phi(T_{ij})) + g(\mathbf{Z}_{ij}; \boldsymbol{\theta}_i, \nu(S_{ij}))\varepsilon_{ij},$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T$, and $\boldsymbol{\varepsilon}_i$'s are independent normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Omega}_i$, which depends on i only through the number of observations within a subject. $\boldsymbol{\beta}_i$ and $\boldsymbol{\theta}_i$ are unknown coefficients with random effects. $f(\cdot)$ and $g(\cdot)$ are known functions characterizing the mean and the intra-individual variances respectively. In practice, the function forms of $f(\cdot)$ and $g(\cdot)$

are usually determined by theory or experience in specific fields. For example, in HIV dynamics analysis, it is common to use a biexponential function to describe the relationship between viral load and measurement time. To account for unexplained covariate effects in $f(\cdot)$ and $g(\cdot)$, such as time-varying decay rates in the biexponential models for HIV dynamics, we introduce $\phi(T_{ij})$ and $\nu(S_{ij})$ into $f(\cdot)$ and $g(\cdot)$, which are both unknown smooth functions associated with the fixed effects. Throughout this paper, we only consider nonparametric inference in the fixed effects, similar as that in Ke and Wang [12]. For the feasibility of estimation, we should specify the model space for $\phi(T_{ij})$ and $\nu(S_{ij})$. We assume that $\phi(T_{ij})$ and $\nu(S_{ij})$ are both square integrable functions from the Sobolev space.

To capture the variation between individuals and the hierarchical structure of the data, we further propose the following level-2 models

$$\begin{aligned} \boldsymbol{\beta}_i &= \mathbf{W}_i \boldsymbol{\gamma} + \mathbf{U}_i, & \boldsymbol{\theta}_i &= \mathbf{K}_i \boldsymbol{\eta} + \mathbf{R}_i, \\ \mathbf{U}_i &\stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}), & \mathbf{R}_i &\stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Lambda}), \end{aligned}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are fixed effects parameters. \mathbf{U}_i and \mathbf{R}_i are mutually independent. $(\mathbf{U}_i^T, \mathbf{R}_i^T)$ and ε_{ij} are independent.

Model (1) is flexible for analyzing hierarchical data with heteroscedasticity. It provides insights for the cause of non-homogeneous variability in responses as well as individual-specific characteristics in the mean. Within the semiparametric framework, both the known components and unobserved covariate effects, such as time-varying coefficients, can be addressed adequately.

3. ESTIMATION

Our goal is to estimate $\boldsymbol{\gamma}, \boldsymbol{\eta}, \phi(T_{ij}), \nu(S_{ij}), \boldsymbol{\Omega}_i, \boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ by maximizing the marginal likelihood function. However, there is usually no explicit form of the marginal likelihood function due to the nonlinear structures of the mean level function and heteroscedasticity function. Considerable attention has been paid to this problem and several methods have been proposed to find approximated solutions, such as EM algorithm [28], first order linearization [22, 17] and Laplace's approximation [12]. Both of the EM algorithm and the first order linearization methods are simple to implement, but have their limitations. The former one is computationally intensive as it incurs large amount of calculation for Monte-Carlo integration and slow convergence for the procedure itself. This problem is particular serious in our proposed heteroscedastic model as it involves estimation of parameters in the intra-individual heterogeneity which may lead to complex inverse matrix operations. The latter method only gives reasonable approximation when the inter-individual variation is small [24, 27], which makes it inflexible in practical application. Hence we propose to use the Laplace method, which provides faster speed and reasonable accuracy, to obtain the approximated marginal likelihood function.

For the estimation of the unknown functions $\phi(T_{ij})$ and $\nu(S_{ij})$ in model (1), kernel smoothing [35], smoothing splines [12], smoothing spline analysis of variance [31] and local fitting [25] are commonly-used methods. In consideration of computing speed and simplicity, we employ spline methods in our nonparametric inferences below.

3.1 The approximated likelihood function

We first approximate $\phi(T_{ij})$ and $\nu(S_{ij})$ by two separate systems of basis functions

$$\begin{aligned}\mathbf{Y}_k(T_{ij}) &= [v_0(T_{ij}), v_1(T_{ij}), \dots, v_{k-1}(T_{ij})]^\top, \\ \mathbf{\Psi}_d(S_{ij}) &= [\psi_0(S_{ij}), \psi_1(S_{ij}), \dots, \psi_{d-1}(S_{ij})]^\top,\end{aligned}$$

where k and d are the numbers of basis functions. Let $\boldsymbol{\mu}_k = (\mu_0, \dots, \mu_{k-1})^\top$ and $\boldsymbol{\xi}_d = (\xi_0, \dots, \xi_{d-1})^\top$ be the vectors of coefficients, then we have the approximations

$$\begin{aligned}(2) \quad \phi(T_{ij}) &\approx \phi_k(T_{ij}) = \sum_{j=0}^{k-1} \mu_j v_j(T_{ij}) = \mathbf{Y}_k(T_{ij})^\top \boldsymbol{\mu}_k, \\ \nu(S_{ij}) &\approx \nu_d(S_{ij}) = \sum_{j=0}^{d-1} \xi_j \psi_j(S_{ij}) = \mathbf{\Psi}_d(S_{ij})^\top \boldsymbol{\xi}_d.\end{aligned}$$

Substituting (2) into (1), we obtain the following approximate parametric hierarchical model

$$(3) \quad Y_{ij} = f(\mathbf{X}_{ij}; \boldsymbol{\beta}_i, \phi_k(T_{ij})) + g(\mathbf{Z}_{ij}; \boldsymbol{\theta}_i, \nu_d(S_{ij})) \varepsilon_{ij}.$$

Let $\boldsymbol{\delta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\eta}^\top, \text{vec}(\boldsymbol{\Sigma}), \text{vec}(\boldsymbol{\Lambda}), \text{vec}(\boldsymbol{\Omega}_i), \boldsymbol{\mu}_k^\top, \boldsymbol{\xi}_d^\top)$ be the vector containing all the parameters to be estimated, where $\text{vec}(\mathbf{A})$ represents a vector of parameters in the variance-covariance matrix \mathbf{A} . Denote $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$, $\mathbf{f}_i = [f(\mathbf{X}_{i1}; \boldsymbol{\beta}_i, \phi_k(T_{i1})), \dots, f(\mathbf{X}_{in_i}; \boldsymbol{\beta}_i, \phi_k(T_{in_i}))]^\top$, $\mathbf{G}_i = \text{diag}\{g(\mathbf{Z}_{i1}; \boldsymbol{\theta}_i, \nu_d(S_{i1})), \dots, g(\mathbf{Z}_{in_i}; \boldsymbol{\theta}_i, \nu_d(S_{in_i}))\}$, $\mathbf{V}_i = \mathbf{G}_i \boldsymbol{\Omega}_i \mathbf{G}_i$. The marginal likelihood based on Model (3) is

$$(4) \quad L(\boldsymbol{\delta}) = \prod_{i=1}^m \int p_{\mathbf{Y}}(\mathbf{Y}_i | \boldsymbol{\beta}_i, \boldsymbol{\theta}_i, \boldsymbol{\delta}) p_{\boldsymbol{\beta}}(\boldsymbol{\beta}_i | \boldsymbol{\gamma}, \boldsymbol{\Sigma}) p_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i | \boldsymbol{\eta}, \boldsymbol{\Lambda}) d\boldsymbol{\beta}_i d\boldsymbol{\theta}_i,$$

where

$$\begin{aligned}p_{\mathbf{Y}}(\mathbf{Y}_i | \boldsymbol{\beta}_i, \boldsymbol{\theta}_i, \boldsymbol{\delta}) &= (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{1}{2}(\mathbf{Y}_i - \mathbf{f}_i)^\top \mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{f}_i)\right\}, \\ p_{\boldsymbol{\beta}}(\boldsymbol{\beta}_i | \boldsymbol{\gamma}, \boldsymbol{\Sigma}) &= (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_i - \mathbf{W}_i \boldsymbol{\gamma})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \mathbf{W}_i \boldsymbol{\gamma})\right\}, \\ p_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i | \boldsymbol{\eta}, \boldsymbol{\Lambda}) &= (2\pi)^{-\frac{q}{2}} |\boldsymbol{\Lambda}|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_i - \mathbf{K}_i \boldsymbol{\eta})^\top \boldsymbol{\Lambda}^{-1}(\boldsymbol{\theta}_i - \mathbf{K}_i \boldsymbol{\eta})\right\}.\end{aligned}$$

Let $\boldsymbol{\omega} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_m^\top, \boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_m^\top)^\top$. $N = \sum_{i=1}^m n_i$. Then the marginal likelihood (4) can be written as a function of $\boldsymbol{\omega}$ as follows

$$(5) \quad L(\boldsymbol{\delta}) \propto |\boldsymbol{\Sigma}|^{-\frac{m}{2}} |\boldsymbol{\Lambda}|^{-\frac{m}{2}} \int \exp\{-\rho(\boldsymbol{\omega})\} d\boldsymbol{\omega},$$

where

$$\begin{aligned}\rho(\boldsymbol{\omega}) &= \frac{1}{2} \sum_{i=1}^m \left\{ \log |\mathbf{V}_i| + (\mathbf{Y}_i - \mathbf{f}_i)^\top \mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{f}_i) \right. \\ &\quad \left. + (\boldsymbol{\beta}_i - \mathbf{W}_i \boldsymbol{\gamma})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_i - \mathbf{W}_i \boldsymbol{\gamma}) \right. \\ &\quad \left. + (\boldsymbol{\theta}_i - \mathbf{K}_i \boldsymbol{\eta})^\top \boldsymbol{\Lambda}^{-1}(\boldsymbol{\theta}_i - \mathbf{K}_i \boldsymbol{\eta}) \right\}.\end{aligned}$$

Denote $\boldsymbol{\omega}_0 = (\boldsymbol{\beta}_{01}^\top, \dots, \boldsymbol{\beta}_{0m}^\top, \boldsymbol{\theta}_{01}^\top, \dots, \boldsymbol{\theta}_{0m}^\top)^\top$ as the solution to the equation $\frac{\partial}{\partial \boldsymbol{\omega}} \rho(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}$. Note that $\boldsymbol{\omega}_0$ is actually the vector of the estimated values of $\boldsymbol{\beta}_i$'s and $\boldsymbol{\theta}_i$'s, which maximize the joint likelihood of the response and the mixed effects coefficients. Obviously, the log-Laplace's approximated marginal likelihood is proportional to

$$(6) \quad -\frac{m}{2} \log |\boldsymbol{\Sigma}| - \frac{m}{2} \log |\boldsymbol{\Lambda}| - \frac{1}{2} \left| \rho^{(2)}(\boldsymbol{\omega}_0) \right| - \rho(\boldsymbol{\omega}_0).$$

where $|\rho^{(2)}(\boldsymbol{\omega}_0)|$ is the determinant of $\frac{\partial^2}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top} \rho(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}$. The explicit expressions of the first and second derivatives of $\rho(\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ are shown in the Appendix.

3.2 Iterative procedures

Maximizing the log-likelihood in (6) with respect to the fixed effects parameters leads to the following estimating equations

$$\begin{aligned}\sum_{i=1}^m -\mathbf{W}_i^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_{0i} - \mathbf{W}_i \boldsymbol{\gamma}) &= \mathbf{0}, \\ \sum_{i=1}^m -\mathbf{K}_i^\top \boldsymbol{\Lambda}^{-1}(\boldsymbol{\theta}_{0i} - \mathbf{K}_i \boldsymbol{\eta}) &= \mathbf{0},\end{aligned}$$

where $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\theta}_{0i}$ are the estimates of the mixed effects coefficients $\boldsymbol{\beta}_i$ and $\boldsymbol{\theta}_i$ as indicated in Section 3.1. Given the estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$, the estimators of the fixed effects can be expressed as

$$(7) \quad \begin{aligned}\hat{\boldsymbol{\gamma}} &= \left(\sum_{i=1}^m \mathbf{W}_i^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{W}_i \right)^{-1} \left[\sum_{i=1}^m \mathbf{W}_i^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\beta}_{0i} \right], \\ \hat{\boldsymbol{\eta}} &= \left(\sum_{i=1}^m \mathbf{K}_i^\top \hat{\boldsymbol{\Lambda}}^{-1} \mathbf{K}_i \right)^{-1} \left[\sum_{i=1}^m \mathbf{K}_i^\top \hat{\boldsymbol{\Lambda}}^{-1} \boldsymbol{\theta}_{0i} \right].\end{aligned}$$

Estimation of a full variance-covariance matrix is challenging when its dimension is not small. This is common in clinical trials or education research where the size of a class or the number of successive records of a subject can often be large. If the number of the replicated observations

is not substantial enough, it is suggested to assume certain structures for the residuals to estimate a limited numbers of covariances [19]. Hence we assume structured matrices for Ω_i . Common structures such as variance components and autoregressive(1) (AR(1)) are employed in the subsequent simulation and real analysis. Details for these structured variance-covariance matrices can be found in Jenrich and Schluchter [11]. Compared to the fixed effects estimators, explicit expressions can be hardly found for the variance components. Numerical methods such as Newton-Raphson algorithm or Nelder-Mead algorithm can be used to find approximated estimates.

As for $\phi_k(T_{ij})$ and $\nu_d(S_{ij})$, we use natural cubic splines in our simulation study and real analysis. The number of knots is determined by the BIC criteria and the location of knots is selected by the percentile-based knot placing rule (Wu and Zhang [33], Liu and Wu [18]). Estimates of the spline coefficients can also be obtained via numerical methods.

We estimate $\gamma, \eta, \Sigma, \Lambda, \Omega_i, \mu_k$ and ξ_d iteratively in the following four steps:

1. Given the current estimate of δ , update β_{0i} 's and θ_{0i} 's by solving $\frac{\partial}{\partial \omega} \rho(\omega) = \mathbf{0}$ with $\rho(\omega)$ in (5).
2. Given the current estimates of Σ, Λ , update $\hat{\gamma}$ and $\hat{\eta}$ by (7).
3. Given the current estimates of γ, η, μ_k and ξ_d , update $\hat{\Sigma}, \hat{\Lambda}$ and $\hat{\Omega}_i$ by maximizing (6) with respect to the variance component parameters.
4. Given the current estimates of Σ, Λ and Ω_i , update $\hat{\mu}_k$ and $\hat{\xi}_d$ by maximizing (6) with respect to the spline coefficients.

3.3 Bootstrap standard errors and confidence band

Asymptotic variance-covariance matrix is a popular choice for evaluating the uncertainty of estimators, e.g. the Hessian of likelihood function in nonlinear mixed models and sandwich estimator for generalized mixed models. However, the sample size of our motivated example mentioned in Section 1 is quite small and the asymptotic approaches may fail due to substantial bias. To overcome this problem, we follow the idea of Sherman and Le Cessie [23] to propose ‘‘all block bootstrap’’ to calculate the confidence intervals of the parametric estimators and the confidence bands of the nonparametric functions. The bootstrap procedure is outlined as follows

1. Match all of the covariates from each subject i and then form the blocks $(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{T}_i, \mathbf{Z}_i, \mathbf{K}_i, \mathbf{S}_i)$, $i = 1, \dots, m$, where $\mathbf{X}_i = (\mathbf{X}_{i1}^T, \dots, \mathbf{X}_{in_i}^T)^T$. $\mathbf{Z}_i, \mathbf{T}_i$ and \mathbf{S}_i are similarly defined. Draw a sample of size m from the m blocks with replacement and denote it as $\{(\mathbf{Y}_1^*, \mathbf{W}_1^*, \mathbf{T}_1^*, \mathbf{Z}_1^*, \mathbf{K}_1^*, \mathbf{S}_1^*), \dots, (\mathbf{Y}_m^*, \mathbf{W}_m^*, \mathbf{T}_m^*, \mathbf{Z}_m^*, \mathbf{K}_m^*, \mathbf{S}_m^*)\}$. The sampling probability is $F_m = \frac{1}{m}$ for each block.
2. Fit the bootstrap sample by model (3).

3. Repeat steps 1 and 2 n_B times. Then calculate the standard errors and pointwise confidence bands based on these bootstrap estimates.

3.4 Asymptotic results

In this section, we investigate the consistency of the proposed estimators assuming model (3), including the fixed effects, the variance components, the coefficients of the basis functions and the nonparametric functions. Similar to the results of Vonesh [27], the convergence rates of the parametric estimators are governed by two factors, the accuracy of the Laplace's approximation and the accuracy of the standard maximum likelihood estimators.

Let δ_0 be the true value of δ . The true value of μ_k can be interpreted as follows. Suppose we truncate the expansion of the true (but unknown) function $\phi(T_{ij}) = \sum_{j=0}^{\infty} \mu_{0j} v_j(T_{ij})$ at the k th term and denote μ_{0k} as the vector of the first k coefficients. Then μ_{0k} is the true value of μ_k for the given integer k . ξ_{0d} can be defined similarly.

Let $\hat{\delta} = (\hat{\gamma}^T, \hat{\eta}^T, \text{vec}(\hat{\Sigma}), \text{vec}(\hat{\Lambda}), \text{vec}(\hat{\Omega}_i), \hat{\mu}_k^T, \hat{\xi}_d^T)^T$ be the estimator of δ , where $\hat{\gamma}$ and $\hat{\eta}$ are obtained by (7). $\text{vec}(\hat{\Sigma}), \text{vec}(\hat{\Lambda}), \text{vec}(\hat{\Omega}_i), \hat{\mu}_k$ and $\hat{\xi}_d$ are the estimators which maximize the log-approximated likelihood function in (6). Denote Θ as the parameter space of δ . Let $L_i(\delta)$ be the marginal likelihood function of δ given \mathbf{Y}_i and $L_i^*(\delta)$ be the Laplace's approximated likelihood function given \mathbf{Y}_i . $l_i(\delta) = \log L_i(\delta)$, $\nabla l_i(\delta) = \partial l_i(\delta) / \partial \delta$, $\nabla^2 l_i(\delta) = \partial^2 l_i(\delta) / \partial \delta \partial \delta^T$, $l_m(\delta) = \sum_{i=1}^m l_i(\delta)$, $\nabla l_m(\delta) = \partial l_m(\delta) / \partial \delta$, $\nabla^2 l_m(\delta) = \partial^2 l_m(\delta) / \partial \delta \partial \delta^T$. $l_i^*(\delta) = \log L_i^*(\delta)$, $\nabla l_i^*(\delta) = \partial l_i^*(\delta) / \partial \delta$, $\nabla^2 l_i^*(\delta) = \partial^2 l_i^*(\delta) / \partial \delta \partial \delta^T$, $l_m^*(\delta) = \sum_{i=1}^m l_i^*(\delta)$, $\nabla l_m^*(\delta) = \partial l_m^*(\delta) / \partial \delta$, $\nabla^2 l_m^*(\delta) = \partial^2 l_m^*(\delta) / \partial \delta \partial \delta^T$. Throughout the rest of the paper, \xrightarrow{p} refers to convergence in probability and \xrightarrow{d} means convergence in distribution.

We first give the consistency of the parameter estimators, including the fixed effects, the variance components and the coefficients of the basis functions in Theorem 3.1. To establish the asymptotic properties, we propose the following regularity conditions, which are similar as those discussed in Vonesh [27], Liu and Wu [18], Wang and Wu [30], Casella and Berger [3]:

- (C1) The necessary conditions for demonstrating the order of accuracy associated with the Laplace's approximation as discussed in [32] are satisfied.
- (C2) The first two partial derivatives of the score function $\nabla l_m(\delta)$ with respect to δ exist almost everywhere in Θ .
- (C3) There are positive numbers ζ_{hs} such that the element in the h th row and s th column of $E[\nabla l_i(\delta) \nabla l_i(\delta)^T]$ is positive and bounded by ζ_{hs} for all $h = 1, \dots, a$ and $s = 1, \dots, a$, where a is the dimension of δ .

- (C4) For each $\boldsymbol{\delta}$ in the neighbourhood of $\boldsymbol{\delta}_0$ there exist functions $I_i(\boldsymbol{\delta})$ and $H_i(\boldsymbol{\delta})$ such that $\int I_i(\boldsymbol{\delta}) d\boldsymbol{\delta} < \infty$, $\int H_i(\boldsymbol{\delta}) d\boldsymbol{\delta} < \infty$. And the absolute values of the first two partial derivatives of $l_i(\boldsymbol{\delta})$ are bounded by $I_i(\boldsymbol{\delta})$ and $H_i(\boldsymbol{\delta})$ respectively for $i = 1, \dots, m$.
- (C5) $E[\nabla l_i(\boldsymbol{\delta})\nabla l_i(\boldsymbol{\delta})^T]$ and $-m^{-1} \sum_{i=1}^m E[\nabla^2 l_i(\boldsymbol{\delta})]$ are finite and positive definite matrices for every $\boldsymbol{\delta}$ in Θ .
- (C6) $m^{-1}\nabla^2 l_m(\boldsymbol{\delta}_0)$ is bounded.
- (C7) The fifth order derivatives of $f(\mathbf{X}_{ij}, \boldsymbol{\beta}_i, \phi_k(T_{ij}))$ and $g(\mathbf{Z}_{ij}, \boldsymbol{\theta}_i, \nu_d(S_{ij}))$ in model (3) with respect to $\boldsymbol{\delta}$ exist and are continuous in an open neighbourhood of $\boldsymbol{\delta}$ for $i = 1, \dots, m$, $j = 1, \dots, n_i$.
- (C8) There exists a constant P such that $1/L_i(\boldsymbol{\delta}) < P$ for every $\boldsymbol{\delta} \in \Theta$, $i = 1, \dots, m$.

In the above regularity conditions, condition (C2) ensures that the score function has a Taylor expansion with respect to $\boldsymbol{\delta}$. Conditions (C3)–(C5) are similar to those in Wang and Wu [30], in which conditions (C3) and (C4) ensure the convergence of the second derivative of the log-likelihood function, using the Markov law of large numbers for non-i.i.d. observations, as discussed in [8]. Condition (C5) ensures the stochastic boundness of the score function following the central limit theorem for non-i.i.d. observations in Serfling [21]. Condition (C7) ensures that the Laplace’s approximation-based score function share the same order of accuracy as that of the Laplace-based marginal likelihood function and condition (C8) is proposed to derive the order of accuracy of the log-approximated likelihood functions.

Theorem 3.1. Let $n = \min_i \{n_i\}$. For model (3), under regularity conditions (C1)–(C8), we have:

$$\hat{\boldsymbol{\delta}} = \boldsymbol{\delta}_0 + O_p \left\{ \max \left[m^{-\frac{1}{2}}, n^{-1} \right] \right\}.$$

Proof. The proof can be found in the Appendix. \square

We further show the consistency of the estimators of the unknown functions in Theorem 3.2. The definitions of norm and inner product are similar as those in [18].

Theorem 3.2. With the same assumptions as in Theorem 3.1, if the following conditions are further satisfied,

- (i) $k/m \rightarrow 0, d/m \rightarrow 0$ when $k \rightarrow \infty, d \rightarrow \infty, m \rightarrow \infty$,
- (ii) For fixed k and d , there exist finite constants C_1 and C_2 , such that $\|\mathbf{Y}_k\| < C_1$ and $\|\boldsymbol{\Psi}_d\| < C_2$ as $m \rightarrow \infty, n \rightarrow \infty$,
- (iii) For any fixed k and d , as $m \rightarrow \infty, n \rightarrow \infty$, $E(\hat{\boldsymbol{\mu}}_k) \rightarrow \boldsymbol{\mu}_{0k}$, $E(\hat{\boldsymbol{\xi}}_d) \rightarrow \boldsymbol{\xi}_{0d}$, $\text{Cov}(\sqrt{m}\hat{\boldsymbol{\mu}}_k) \rightarrow \mathbf{J}_k$ and $\text{Cov}(\sqrt{m}\hat{\boldsymbol{\xi}}_d) \rightarrow \mathbf{Q}_d$, where \mathbf{J}_k and \mathbf{Q}_d are some semidefinite positive matrices with $k^{-1}\text{tr}(\mathbf{J}_k)$ and $d^{-1}\text{tr}(\mathbf{Q}_d)$ bounded,

then as $k \rightarrow \infty, d \rightarrow \infty, m \rightarrow \infty$ and $n \rightarrow \infty$, the estimators of the unknown functions in (6) have the following properties:

$$\|\hat{\phi}_k - \phi\| \xrightarrow{P} 0, \quad \|\hat{\nu}_d - \nu\| \xrightarrow{P} 0.$$

Proof. The proof can be found in the Appendix. \square

Asymptotic normality of $\hat{\boldsymbol{\delta}}$ is summarized in the following theorem, in which similar regularity conditions can be found in [14].

Theorem 3.3. For model (3), in addition to the regularity conditions of Theorem 3.1, suppose that there exists an open subset of $\boldsymbol{\delta}$ containing $\boldsymbol{\delta}_0$ such that $l_i^*(\boldsymbol{\delta})$ is twice continuously differentiable and as $m \rightarrow \infty, n \rightarrow \infty$,

- (i) $\sum_{i=1}^m E\|\nabla l_i^*(\boldsymbol{\delta}_0)\|^{2+\alpha} = o(m^\alpha)$ for some $\alpha > 0$,
- (ii) $\left\| \sum_{i=1}^m E\nabla l_i^*(\boldsymbol{\delta}_0) \right\| = o(\sqrt{m})$,
- (iii) $m^{-1} \sum_{i=1}^m \nabla l_i^*(\boldsymbol{\delta}_0) [\nabla l_i^*(\boldsymbol{\delta}_0)]^T \xrightarrow{P} \boldsymbol{\Gamma}_1$,
- (iv) $m^{-1} \sup_{\|\boldsymbol{\delta} - \boldsymbol{\delta}_0\| \leq c} \left\| \sum_{i=1}^m \nabla^2 l_i^*(\boldsymbol{\delta}) - \sum_{i=1}^m \nabla^2 l_i^*(\boldsymbol{\delta}_0) \right\| \xrightarrow{P} 0$ as $c \rightarrow 0$, $-m^{-1} \sum_{i=1}^m \nabla^2 l_i^*(\boldsymbol{\delta}_0) \xrightarrow{P} \boldsymbol{\Gamma}_2$, where $\boldsymbol{\Gamma}_2$ is positive definite.

Then as $m \rightarrow \infty$ and $n \rightarrow \infty$,

$$\sqrt{m}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_2^{-1} \boldsymbol{\Gamma}_1 \boldsymbol{\Gamma}_2^{-1}).$$

Proof. The proof can be found in the Appendix. \square

4. SIMULATION STUDIES

In this section, we conduct a simulation study to evaluate the performance of our proposed estimation procedures. Data are generated from the following model which mimics the indomethacin study in the real analysis.

$$\begin{aligned} (8) \quad y_{ij} &= \{ \exp(\beta_{i1} - \beta_{i2}t_{ij}) + \exp(\beta_{i3} - \phi(t_{ij})t_{ij}) \} \\ &\quad + \exp(\theta_{i1} - \nu(t_{ij})t_{ij})\varepsilon_{ij}, \\ \boldsymbol{\beta}_i &= \boldsymbol{\gamma} + \mathbf{u}_i, \quad \mathbf{u}_i \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}), \\ \theta_i &= \eta + r_i, \quad r_i \stackrel{iid}{\sim} N(0, \lambda^2), \\ \boldsymbol{\varepsilon}_i &= (\varepsilon_{i1}, \dots, \varepsilon_{in_i}) \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Omega}), \\ i &= 1, \dots, 20, \quad j = 1, \dots, 20, \end{aligned}$$

where t_{ij} are taken as equal spaced when $0.25 \leq t \leq 2$ and then $t = 3, 4, 6, 8$ when $t > 2$. The true values are set as $\phi(t) = 0.2t^{-1/2}$, $\nu(t) = 0.1t^{-1/2} + 0.1$. $\boldsymbol{\gamma} = (\log(2), 1.5, \log(0.3))$, $\eta = \log(0.3)$, $\lambda^2 = 0.05$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.03 & 0.01 & -0.01 \\ 0.01 & 0.02 & -0.01 \\ -0.01 & -0.01 & 0.05 \end{pmatrix}.$$

Both variance component and AR(1) structures of $\boldsymbol{\Omega}$ are investigated in the following two scenarios. We fit the data

Table 1. Simulation results of the parametric estimators; scenario 1

		γ_1	γ_2	γ_3	η	Σ_{11}	Σ_{22}	Σ_{33}	λ^2	σ^2
Heter	relBias	0.0184	0.0079	0.0524	0.0131	0.3524	6.4903	2.2935	2.3822	0.7904
	relSD	0.0730	0.0570	0.1139	0.4637	1.3056	11.6010	4.2691	3.6344	1.300
	95%CIW	0.2096	0.3395	0.5685	2.1607	0.1132	0.6796	0.4964	0.5543	0.2423
Homo	relBias	0.0696	0.0151	0.0662	-	0.1475	2.9238	1.6401	-	0.9022
	relSD	0.0978	0.0807	0.1954	-	0.5802	5.3494	3.2038	-	0.1087
	95%CIW	0.2980	0.4927	0.8862	-	0.0775	0.4365	0.5637	-	0.0208

Heter: estimates of the true heteroscedasticity model (8), Homo: estimates of a misspecified homogeneous model.

with our proposed method assuming model (8) and compare the results to those assuming homogeneity models. All computations here are implemented in Matlab. Natural cubic splines are used to estimate $\phi(t)$ and $\mu(t)$. To select the number of knots for the estimated functions, we calculate the BIC values of 16 combinations with the knots of $\hat{\phi}_k(t)$ and $\hat{\nu}_d(t)$ ranging from 1 to 4 respectively. Log-Cholesky parameterizations are employed to enforce the positive semidefinite conditions for variance-covariance matrices. In both scenarios, 100 Monte Carlo datasets are conducted. The performance of the parametric estimators are assessed via the relative bias(relBias), the relative Monte Carlo standard deviation (relSD) and the Monte Carlo 95% confidence interval width (95%CIW),

$$\text{relBias} = \left| \frac{\text{bias of estimate}}{\text{true parameter}} \right|,$$

$$\text{relSD} = \left| \frac{\text{SD of estimate}}{\text{true parameter}} \right|,$$

$$95\% \text{CIW} = 97.5\% \text{ quantile} - 2.5\% \text{ quantile}.$$

As for the nonparametric functions, the performance is assessed with 50 equal spaced grid points in the interval $[0.25, 8]$. We use the average and standard deviation of mean absolute deviation (MAD) and 95% pointwise confidence band to evaluate their curve fitting performance. The MAD of a function $f(t)$ can be calculated as

$$\text{MAD} = n_0^{-1} \sum_{k=1}^{n_0} |\hat{f}(t_k) - f(t_k)|,$$

where $\{t_k, k = 1, \dots, n_0\}$ are the grid points at which $f(t)$ are estimated.

Scenarios I $\Omega = \text{diag}(\sigma^2), \sigma^2 = 0.05$.

Table 1 shows the estimation results of the fixed effects parameters and the variance components. For Σ , we only report the estimates of the diagonal elements. As is shown in Table 1, our proposed estimators of the fixed effects work well in terms of relative bias and relative standard deviation. The estimates under model (8) and those under a misspecified homogeneous model have similar bias. However the latter ones have larger standard deviation in general,

Table 2. Simulation results of the nonparametric functions; scenario 1

	$\hat{\phi}(t)$		$\hat{\nu}(t)$	
	av.MAD	std.MAD	av.MAD	std.MAD
Heter	0.0447	0.0471	0.1149	0.0934
Homo	0.0949	0.0544	-	-

which implies that the estimation method ignoring the heteroscedasticity function leads to less efficient results. It is also found that the variance estimators tend to have larger relative SD compared to the fixed effects estimators. This is not surprising as the estimation of second moments is usually more sophisticated.

The simulation results of the smooth functions is summarized in Table 2. The spline method yields good estimators with significantly small average MAD and standard errors of MAD. We further plot out the 95% pointwise confidence band and the average fitted curve in Figure 1 to see whether our proposed estimators are efficient in capturing the unknown smooth functions. The left panel of Figure 1 is $\hat{\phi}(t)$ in model (8) and the middle panel is $\hat{\nu}(t)$ in model (8). The right panel is $\hat{\phi}(t)$ assuming homogeneity model. The dashed lines are the estimates. The solid lines are the true functions and the dotted lines are the 95% pointwise confidence bands. Most part of the $\hat{\phi}(t)$ and $\hat{\nu}(t)$ from model (8) coincide with the true values. The estimators of unknown functions in the mean have similar performance in both model (8) and the misspecified homogeneous model, except that $\hat{\phi}(t)$ have wider confidence band when the heteroscedasticity function is ignored.

Scenarios II

$$\sigma^2 = 0.05, \rho = 0.8, n = 20,$$

$$\Omega = \sigma^2 \times \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & & & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 & \end{pmatrix}$$

The simulation in scenario 2 aims to investigate the performance of our estimators when observations within a subject is correlated and hence the intra variance-covariance matrix has a more complex structure. Results are shown in

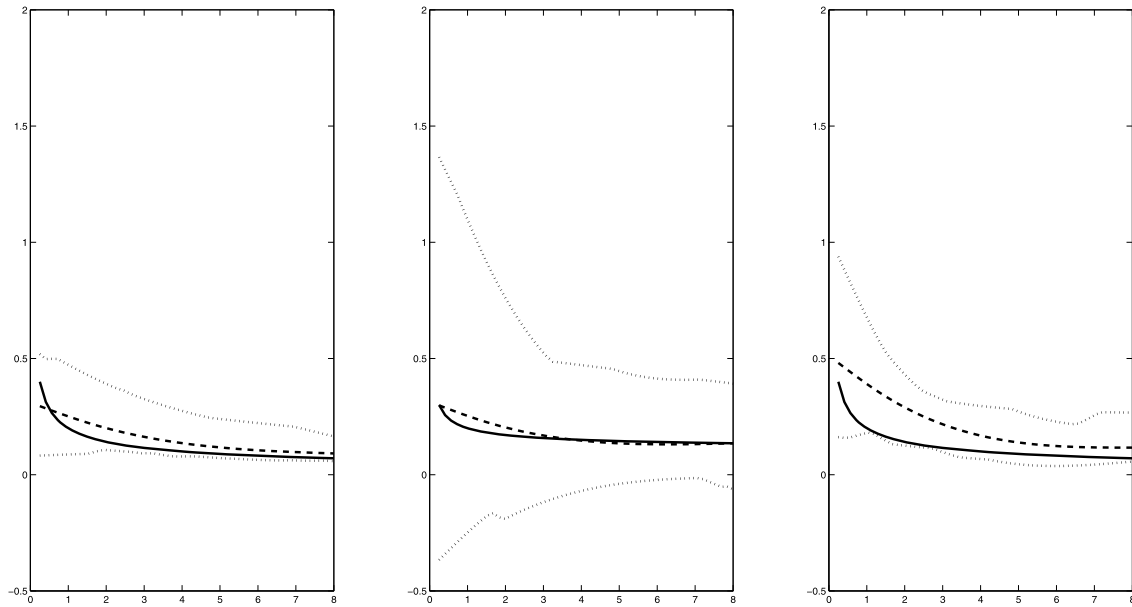


Figure 1. Fitted curves of the unknown functions in scenario 1 of the simulation.

Table 3. Simulation results of the parametric estimators; scenario 2

		γ_1	γ_2	γ_3	η	Σ_{11}	Σ_{22}	Σ_{33}	λ^2	σ^2	
Heter	relBias	0.0239	0.0087	0.0557	0.0964	0.3090	6.1617	2.1130	3.7834	2.1597	0.0442
	relSD	0.1214	0.0997	0.1647	0.7780	1.4766	11.1543	4.4971	17.2934	3.5185	0.1397
	95%CIW	0.2920	0.3586	0.7960	3.2775	0.0745	0.6772	0.5212	0.5452	0.6999	0.4095
Homo	relBias	0.0045	0.0774	0.0586	-	0.4745	4.0728	1.9321	-	0.8476	0.0957
	relSD	0.0785	0.0522	0.2312	-	7.6606	13.5921	2.7406	-	0.1146	0.0935
	95%CIW	0.2343	0.3162	0.9375	-	0.0990	1.0057	0.5716	-	0.0178	0.3180

Heter: estimates of the true heteroscedasticity model (8), Homo: estimates of a misspecified homogeneous model.

Table 3 and 4. The parametric estimates have small bias similar to scenario 1. It is noted that greater standard deviations of parameter estimates from both the heteroscedastic model and homogeneous model are found in scenario 2, especially for the parameters in the variance function. It may be explained by the more complicated structures of the intra variance-covariance matrices. From Figure 2, we also observe wider confidence band for $\hat{\phi}(t)$ in the misspecified homogeneous model as the way the mean interacts with the variance is more complicated (The left panel is $\hat{\phi}(t)$ in model (8) and the middle panel is $\hat{\nu}(t)$ in model (8). The right panel is $\hat{\phi}(t)$ assuming homogeneity model. The dashed lines are the estimates. The solid lines are the true functions and the dotted lines are the 95% pointwise confidence band.).

5. REAL ANALYSIS

In this section, our proposed models and estimation procedures are applied to the indomethacin medical study discussed earlier. The dataset include plasma concentration of 6 individuals measured at 11 time points ranging from 15 minutes to 8 hours after treatment. Davidian and Giltinan

Table 4. Simulation results of the nonparametric functions; scenario 2

	$\hat{\phi}(t)$		$\hat{\nu}(t)$	
	av.MAD	std.MAD	av.MAD	std.MAD
Heter	0.0559	0.2248	0.1522	0.1768
Homo	0.1200	0.0683	-	-

[5] used the following biexponential model with power variance function to fit the data.

$$Y_{ij} = f(t_{ij}; \beta_i) + \sigma^2 f(t_{ij}; \beta_i)^\theta \varepsilon_{ij},$$

$$\beta_i \stackrel{iid}{\sim} N(\beta, \Sigma),$$

where Y_{ij} is the plasma concentration of indomethacin at the j th time point in the i th subject and t_{ij} is the corresponding measurement time. $f(t_{ij}; \beta_i) = e^{\beta_{i1}} \exp(-e^{\beta_{i2}} t_{ij}) + e^{\beta_{i3}} \exp(-e^{\beta_{i4}} t_{ij})$. $\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$. β , θ , Σ and σ^2 are the parameters to be estimated.

It is reasonable to assume that the intra-subject variation is a function of the mean response. However, this leads

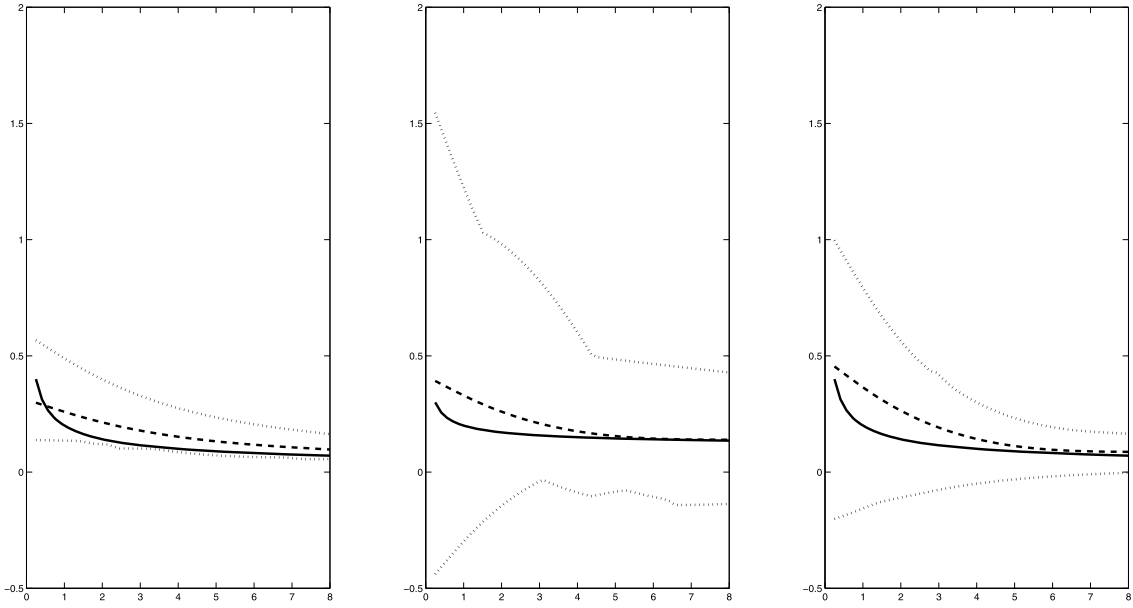


Figure 2. Fitted curves of the unknown functions in scenario 2 of the simulation.

Table 5. Parameter estimates of indomethacin data analysis

		γ_1	γ_2	γ_3	η	Σ_1	Σ_2	Σ_3	λ^2	σ^2
Heter	Est	0.8695	1.9034	-0.5597	0.9577	0.0295	0.3155	0.0767	0.2402	0.0062
	S.E.	0.1113	0.2287	0.1978	0.5411	0.0932	0.2196	0.1571	0.1486	0.2419
	BIC	-336.6686								
Homo	Est	1.0038	2.5391	-0.4329	-	0.0769	0.4008	0.1146	-	0.0056
	S.E.	0.1819	0.5502	0.4194	-	0.0413	0.3690	0.2070	-	0.0045
	BIC	-185.8370								

Heter: estimates of the true heteroscedasticity model (9), Homo: estimates of a misspecified homogeneous model.

to a very complex model with two biexponential components, one in the mean and the other in the variance functions. Such complex models may cast doubt on the reliability of the estimators and the robustness of the model. Hence we remain the biexponential form in the mean function while simply assume a monoexponential form for the variance function. We also incorporate random effects to account for the variations between subjects. As discussed in the HIV dynamics research from Wu and Zhang [34], where biexponential models are also employed, the second decay rate often changes with time. To account for this effect, we assume the decay rates to be time-dependent. Hence we propose the following models for analysing the indomethacin data

$$\begin{aligned}
 (9) \quad Y_{ij} &= f(t_{ij}; \beta_i, \phi(t_{ij})) + g(t_{ij}; \theta_i, \nu(t_{ij})) \varepsilon_{ij}, \\
 \beta_i &\stackrel{iid}{\sim} N(\gamma, \Sigma), \quad \theta_i \stackrel{iid}{\sim} N(\eta, \lambda^2), \\
 i &= 1, \dots, 6, \quad j = 1, \dots, 11,
 \end{aligned}$$

$$\begin{aligned}
 g(t_{ij}; \theta_i, \nu(t_{ij})) &= \exp(\theta_i - \nu(t_{ij})t_{ij}), \\
 f(t_{ij}; \beta_i, \phi(t_{ij})) &= e^{\beta_{i1}} \exp(-\beta_{i2}t_{ij}) + e^{\beta_{i3}} \exp(-\phi(t_{ij})t_{ij}),
 \end{aligned}$$

$\varepsilon_{ij} \stackrel{iid}{\sim} N(\mathbf{0}, \sigma^2)$. $\phi(\cdot)$ and $\nu(\cdot)$ are unknown functions to be estimated. γ , η , Σ , λ^2 and σ^2 are unknown parameters. We assume Σ to be unstructured. Note that Model (9) is a special form of Model (1) in the sense that the second level design matrices \mathbf{W}_i and \mathbf{K}_i are all set to be identity matrices.

We obtain estimates based on both the heteroscedasticity model (9) and homogeneous models where $g(\cdot)$ in (9) equal to 1. Standard errors are calculated via bootstrap method. Results are summarized in Table 5, where Est refers to the estimated values of parameters and S.E. refers to the bootstrap standard errors. Plots of fitted curves are presented in Figure 3. The left panel is $\hat{\phi}(t)$ in model (9) and the middle panel is $\hat{\nu}(t)$ in model (9). The right panel is $\hat{\phi}(t)$ assuming homogeneity model.

From Table 5 we can see that both model (9) and the homogeneous model have similar estimated values for γ in the mean but the former model outperforms the other in terms of the BIC value. In model (9), $\hat{\gamma}_2$ is positive, which indicates that the plasma concentration of indomethacin decrease in the first stage. This decay rate is slightly higher than the results in Davidian and Giltinan's paper [5].

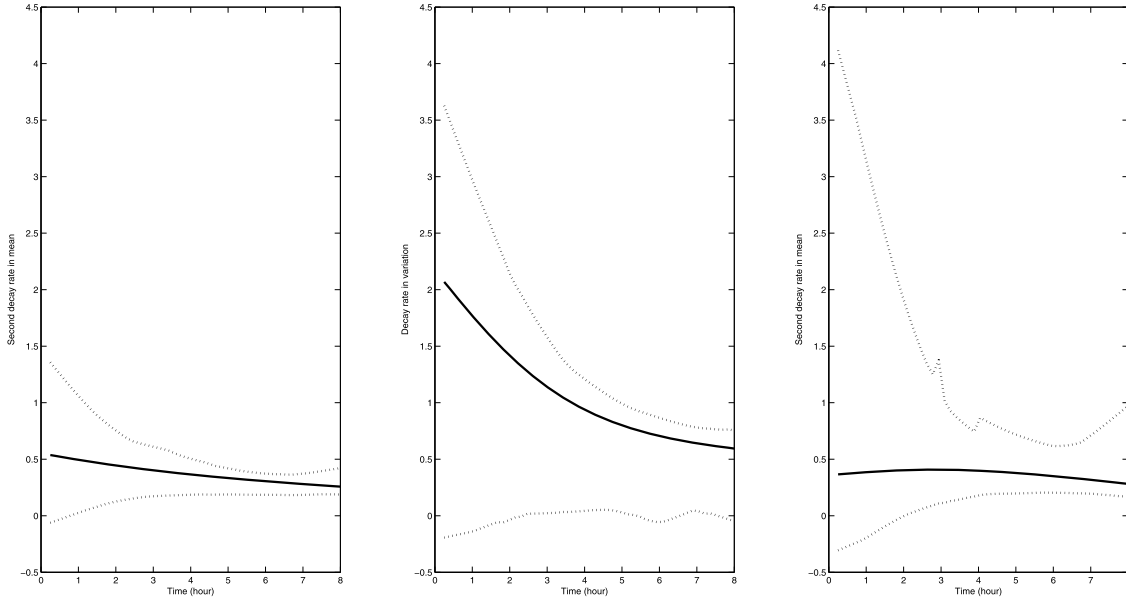


Figure 3. Fitted curves of the unknown functions in the indomethacin data analysis.

Results from model (9) also reveal the time-varying features of the metabolic process of indomethacin. As found in Figure 3, the second decay rate in the mean gradually decrease as time passes by and still remain positive at the end. This means that the level of indomethacin in plasma concentrate after injection decline along time with a reducing rate and this trend may probably go on after 8 hours. The decay rate in variation have similar trend but with a sharper drop than that in the mean. Hence we can infer that the fluctuation of indomethacin reduces along the metabolic process and the declining effect is much more significant at the start.

6. CONCLUSION

To capture the heteroscedasticity in hierarchical data, we have proposed a class of semiparametric hierarchical models with heteroscedasticity. Laplace-based likelihood estimators are obtained and their asymptotic properties are investigated. The simulation study and the real analysis suggest that our proposed estimating procedures work well.

The methodology considered in this paper is only for continuous data with normal distributions for the random effects and model disturbances. The assumption of normality may be relaxed in the subsequent research. In addition, the estimation of the hierarchical models with heteroscedasticity is based on the condition that the structure for the variance-covariance matrices is known in advance, which may limit the use of the proposed methods in some practices. This requires more studies in the future.

Hypothesis testing procedures comparing the proposed semiparametric models with simple parametric models are an interesting topic and will be investigated in our future work.

APPENDIX

A.1 Derivatives of $\rho(\omega)$ in (6)

Denote the first and second derivatives of \mathbf{V}_i with respect to the s th element in θ_i as

$$\begin{aligned}\dot{\mathbf{V}}_{is} &= \mathbf{G}_i \Omega \frac{\partial \mathbf{G}_i}{\partial \theta_{is}} + \frac{\partial \mathbf{G}_i}{\partial \theta_{is}} \Omega \mathbf{G}_i, \\ \ddot{\mathbf{V}}_{is', is} &= \frac{\partial \mathbf{G}_i}{\partial \theta_{is}} \Omega \frac{\partial \mathbf{G}_i}{\partial \theta_{is'}} + \frac{\partial \mathbf{G}_i}{\partial \theta_{is'}} \frac{\partial \mathbf{G}_i}{\partial \theta_{is}} \Omega \mathbf{G}_i \\ &\quad + \mathbf{G}_i \Omega \frac{\partial \mathbf{G}_i}{\partial \theta_{is'} \partial \theta_{is}} + \frac{\partial \mathbf{G}_i}{\partial \theta_{is'}} \Omega \frac{\partial \mathbf{G}_i}{\partial \theta_{is}}, \\ i &= 1, \dots, m, \quad s = 1, \dots, q, \quad s' = 1, \dots, q.\end{aligned}$$

Then the first derivatives of $\rho(\omega)$ with respect to ω can be written as:

$$\begin{aligned}\rho^{(1)}(\omega) &= [\mathbf{A}^T, \mathbf{B}^T]^T \\ \mathbf{A}_{ih} &= \frac{\partial}{\partial \beta_{ih}} \rho(\omega) \\ &= -(\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \frac{\partial \mathbf{f}_i}{\partial \beta_{ih}} + (\beta_i - \mathbf{W}_i \gamma)^T \Sigma^{-1} \mathbf{L}_h, \\ \mathbf{B}_{is} &= \frac{\partial}{\partial \theta_{is}} \rho(\omega) \\ &= -\frac{1}{2} (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{f}_i) \\ &\quad + \frac{1}{2} \text{tr} \left\{ \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is} \right\} + (\theta_i - \mathbf{K}_i \eta)^T \Lambda^{-1} \mathbf{L}_s, \\ h &= 1, \dots, p, \\ s &= 1, \dots, q,\end{aligned}$$

where \mathbf{A}_{ih} is the $(i-1) \times p + h$ component of a $m \times p$ by 1 vector \mathbf{A} and \mathbf{B}_{is} is the $(i-1) \times q + s$ component of a $m \times q$ by 1 vector \mathbf{B} . \mathbf{L}_h is the h th column of a p dimensional identity matrix and \mathbf{L}_s is the s th column of a q dimensional identity matrix. $\text{tr}(\mathbf{A})$ denotes the trace of matrix \mathbf{A} . The second derivatives can be similarly expressed as

$$\begin{aligned} \rho^{(2)}(\boldsymbol{\omega}) &= \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix} \\ \mathbf{C}_{ih', ih} &= \frac{\partial^2}{\partial \boldsymbol{\beta}_{ih'} \partial \boldsymbol{\beta}_{ih}} \rho(\boldsymbol{\omega}) \\ &= \frac{\partial \mathbf{f}_i^T}{\partial \boldsymbol{\beta}_{ih'}} \mathbf{V}_i^{-1} \frac{\partial \mathbf{f}_i}{\partial \boldsymbol{\beta}_{ih}} - (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \frac{\partial^2 \mathbf{f}_i}{\partial \boldsymbol{\beta}_{ih'} \partial \boldsymbol{\beta}_{ih}} \\ &\quad + \mathbf{L}_{h'}^T \boldsymbol{\Sigma}^{-1} \mathbf{L}_h, \\ \mathbf{D}_{ih, is} &= \frac{\partial^2}{\partial \boldsymbol{\beta}_{ih} \partial \boldsymbol{\theta}_{is}} \rho(\boldsymbol{\omega}) \\ &= (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is} \mathbf{V}_i^{-1} \frac{\partial \mathbf{f}_i}{\partial \boldsymbol{\beta}_{ih}}, \\ \mathbf{E}_{is, ih} &= \frac{\partial^2}{\partial \boldsymbol{\theta}_{is} \partial \boldsymbol{\beta}_{ih}} \rho(\boldsymbol{\omega}) \\ &= (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is} \mathbf{V}_i^{-1} \frac{\partial \mathbf{f}_i}{\partial \boldsymbol{\beta}_{ih}}, \\ \mathbf{F}_{is', is} &= \frac{\partial^2}{\partial \boldsymbol{\theta}_{is'} \partial \boldsymbol{\theta}_{is}} \rho(\boldsymbol{\omega}) \\ &= (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is'} \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{f}_i) \\ &\quad - \frac{1}{2} \text{tr} \left(\mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is'} \mathbf{V}_i^{-1} \dot{\mathbf{V}}_{is} \right) \\ &\quad - \frac{1}{2} (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{V}_i^{-1} \ddot{\mathbf{V}}_{is', is} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{f}_i) \\ &\quad + \frac{1}{2} \text{tr} \left(\mathbf{V}_i^{-1} \ddot{\mathbf{V}}_{is', is} \right) + \mathbf{L}_{s'}^T \boldsymbol{\Lambda}^{-1} \mathbf{L}_s, \end{aligned}$$

where $\mathbf{C}_{ih', ih}$ is the element in the $(i-1) \times p + h'$ row and $(i-1) \times p + h$ column of a $m \times p$ by $m \times p$ matrix \mathbf{C} . $\mathbf{D}_{ih, is}$, $\mathbf{E}_{is, ih}$, $\mathbf{F}_{is', is}$ can be similarly defined.

A.2 Proof of Theorem 3.1

Proof. Similar to Liu and Wu [18], we investigate the large sample properties of our proposed estimators based on the approximate parametric models in (3). The marginal likelihood function of \mathbf{Y}_i can be written as an integral with respect to the mixed coefficients $\boldsymbol{\omega}_i = (\boldsymbol{\beta}_i^T, \boldsymbol{\theta}_i^T)^T$

$$\begin{aligned} L(\boldsymbol{\delta} | \mathbf{Y}_i) &= \int \exp\{n_i J_i(\boldsymbol{\omega}_i)\} d \boldsymbol{\omega}_i \\ &= \int \exp\{-n_i [-J_i(\boldsymbol{\omega}_i)]\} d \boldsymbol{\omega}_i, \end{aligned}$$

where $J_i(\boldsymbol{\omega}_i) = n_i^{-1} [\log p_{\mathbf{Y}}(\mathbf{Y}_i | \boldsymbol{\beta}_i, \boldsymbol{\theta}_i, \boldsymbol{\delta}) + \log p_{\boldsymbol{\beta}}(\boldsymbol{\beta}_i | \boldsymbol{\gamma}, \boldsymbol{\Sigma}) + \log p_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i | \boldsymbol{\eta}, \boldsymbol{\Lambda})]$ and the detailed expressions of $p_{\mathbf{Y}}(\cdot)$, $p_{\boldsymbol{\beta}}(\cdot)$ and $p_{\boldsymbol{\theta}}(\cdot)$ can be found in Section 3.1. Applying Laplace's approximation we can obtain the following marginal likeli-

hood function

$$\begin{aligned} L(\boldsymbol{\delta} | \mathbf{Y}_i) &= \exp\{n_i J_i(\boldsymbol{\omega}_{0i})\} \left[\frac{(2\pi)^{\frac{p+q}{2}}}{\left| -n_i J_{i, \boldsymbol{\omega}_{0i} \boldsymbol{\omega}_{0i}}^{(2)} \right|^{\frac{1}{2}}} \right] [1 + O(n_i^{-1})], \end{aligned}$$

where $\boldsymbol{\omega}_{0i} = (\boldsymbol{\beta}_{0i}^T, \boldsymbol{\theta}_{0i}^T)^T$ is a vector of estimates of the mixed effects coefficients, satisfying $\frac{\partial}{\partial \boldsymbol{\omega}_i} J_i(\boldsymbol{\omega}_i) |_{\boldsymbol{\omega}_i = \boldsymbol{\omega}_{0i}} = \mathbf{0}$. $J_{i, \boldsymbol{\omega}_{0i} \boldsymbol{\omega}_{0i}}^{(2)} = \frac{\partial^2}{\partial \boldsymbol{\omega}_i \partial \boldsymbol{\omega}_i^T} J_i(\boldsymbol{\omega}_i) |_{\boldsymbol{\omega}_i = \boldsymbol{\omega}_{0i}}$. Denote $x = O(n_i^{-1})$ and consider $l_i^*(\boldsymbol{\delta})$ as a function of x , then we have

$$l_i^*(\boldsymbol{\delta}) = \log [L_i(\boldsymbol{\delta}) + x].$$

By Taylor's theorem and condition (C8), we have

$$\begin{aligned} l_i^*(\boldsymbol{\delta}) &= \log [L_i(\boldsymbol{\delta})] + \frac{1}{L_i(\boldsymbol{\delta})} O(n_i^{-1}) \\ &\quad + \frac{1}{-2(L_i(\boldsymbol{\delta}) + x^*)^2} [O(n_i^{-1})]^2 \\ &= \log [L_i(\boldsymbol{\delta})] + O(n_i^{-1}). \end{aligned}$$

where x^* is on the line segment joining 0 to x . Hence

$$l_m^*(\boldsymbol{\delta}) = l_m(\boldsymbol{\delta}) + O_p(mn^{-1}).$$

$\hat{\boldsymbol{\delta}}$ satisfies $\nabla l_m^*(\hat{\boldsymbol{\delta}}) = \mathbf{0}$. Assume $\hat{\boldsymbol{\delta}}$ is an interior point in a neighbourhood containing $\boldsymbol{\delta}_0$ and take a first order Taylor series expansion of $\nabla l_m(\hat{\boldsymbol{\delta}})$ about $\boldsymbol{\delta}_0$ then we can obtain

$$\begin{aligned} m^{-1} \nabla l_m(\hat{\boldsymbol{\delta}}) &= m^{-1} \nabla l_m(\boldsymbol{\delta}_0) + m^{-1} \nabla^2 l_m(\boldsymbol{\delta}_0) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \\ &\quad + O_p(1) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)^T (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \mathbf{1}_a, \end{aligned}$$

where $\mathbf{1}_a$ is a vector with all the elements equalling to 1 and a is the dimension of $\boldsymbol{\delta}$. Given condition (C6) and the assumption that $\hat{\boldsymbol{\delta}}$ is an interior point in a neighbourhood containing $\boldsymbol{\delta}_0$, we can show that $m^{-1} \nabla^2 l_m(\boldsymbol{\delta}_0) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) + O_p(1) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)^T (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \mathbf{1}_a = O_p(1) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)$, where $\mathbf{1}_a$ is defined as the same as above. Combining these results, it is shown that

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = m^{-1} [\nabla l_m(\hat{\boldsymbol{\delta}}) - \nabla l_m(\boldsymbol{\delta}_0)] / O_p(1).$$

Under conditions (C3) and (C4), by the Markov's law of large numbers for non-i.i.d. observations, the second derivative of the log-likelihood function converges to a matrix as below

$$-m^{-1} \nabla^2 l_m(\boldsymbol{\delta}) \xrightarrow{p} \boldsymbol{\Pi}.$$

Then together with condition (C5) and the Lindeberg condition, it follows that

$$m^{-\frac{1}{2}} \nabla l_m(\boldsymbol{\delta}) \xrightarrow{d} N(0, \boldsymbol{\Pi}).$$

Hence

$$m^{-1}\nabla l_m(\boldsymbol{\delta}) = O_p(m^{-\frac{1}{2}}).$$

Moreover, by conditions (C7), we have

$$m^{-1}\nabla l_m(\hat{\boldsymbol{\delta}}) = m^{-1}\nabla l_m^*(\hat{\boldsymbol{\delta}}) + O_p(n^{-1}).$$

Hence

$$\begin{aligned} \hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 &= m^{-1}\nabla l_m^*(\hat{\boldsymbol{\delta}}) + O_p(n^{-1}) + O_p(m^{-\frac{1}{2}}) \\ &= O_p\left\{\max\left[m^{-\frac{1}{2}}, n^{-1}\right]\right\}. \quad \square \end{aligned}$$

A.3 Proof of Theorem 3.2

Proof. The proof of the consistency of $\hat{\nu}_d$ is similar to that of $\hat{\phi}_k$, hence we only demonstrate the consistency of $\hat{\phi}_k$ in this section. Let $\phi_{0k} = \sum_{j=0}^{k-1} \mu_{0j}v_j$ be the truncation of the expansion of the true unknown function ϕ as described in Section 3.4. Following Liu and Wu [18], it can be shown that

$$\begin{aligned} E\|\hat{\phi}_k - \phi\|^2 &\leq 2\{E\|\hat{\phi}_k - \phi_{0k}\|^2 + \|\phi_{0k} - \phi\|^2\} \\ &= 2\{C_1E\|\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_{0k}\|^2 + \|\phi_{0k} - \phi\|^2\} \\ &= 2\{C_1\text{tr}[\text{Cov}(\hat{\boldsymbol{\mu}}_k)] + C_1\|E(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{\mu}_{0k}\|^2 \\ &\quad + \|\phi_{0k} - \phi\|^2\} \\ &= 2\left\{\frac{C_1}{m}\text{tr}[\text{Cov}(\sqrt{m}\hat{\boldsymbol{\mu}}_k)] + C_1\|E(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{\mu}_{0k}\|^2\right. \\ &\quad \left. + \|\phi_{0k} - \phi\|^2\right\} \\ &= 2\left\{\frac{C_1}{m}\{\text{tr}[\mathbf{J}_k + o(1)]\} + C_1\|E(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{\mu}_{0k}\|^2\right. \\ &\quad \left. + \sum_{j=k}^{\infty} \mu_{0j}^2\right\} \\ &= \left\{\frac{C_1k}{m} \frac{\text{tr}[\mathbf{J}_k]}{k} + C_1\|E(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{\mu}_{0k}\|^2 + \sum_{j=k}^{\infty} \mu_{0j}^2\right\} \\ &\quad + o\left(\frac{k}{m}\right). \end{aligned}$$

Under conditions (i)–(iii) as stated in Theorem 3.2, as $m \rightarrow \infty$, $n \rightarrow \infty$, $k \rightarrow \infty$,

$$\frac{k}{m} \frac{\text{tr}[\mathbf{J}_k]}{k} \rightarrow 0, \quad \|E(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{\mu}_{0k}\|^2 \rightarrow 0, \quad \sum_{j=k}^{\infty} \mu_{0j}^2 \rightarrow 0.$$

Hence

$$E\|\hat{\phi}_k - \phi\|^2 \rightarrow 0,$$

which implies

$$\|\hat{\phi}_k - \phi\| \xrightarrow{P} 0.$$

As we have proved in Theorem 3.1, $\hat{\boldsymbol{\mu}}_k$ is the consistent estimator of $\boldsymbol{\mu}_{0k}$, hence $\hat{\phi}_k$ converges to ϕ in probability in L^2 norm. \square

A.4 Proof of Theorem 3.3

Proof. Since $\sum_{i=1}^m \nabla l_i^*(\hat{\boldsymbol{\delta}}) = 0$, by Taylor series expansion of

$\sum_{i=1}^m \nabla l_i^*(\hat{\boldsymbol{\delta}})$ about $\boldsymbol{\delta}_0$, we have

$$\sum_{i=1}^m \nabla l_i^*(\boldsymbol{\delta}_0) = -\sum_{i=1}^m \nabla^2 l_i^*(\boldsymbol{\delta}^*)(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0),$$

where $\boldsymbol{\delta}^*$ is on the line segment joining $\boldsymbol{\delta}_0$ to $\hat{\boldsymbol{\delta}}$. It follows from assumptions (i)–(iii) in Theorem 3.3 and the Liapounov central limit theorem in [8] that as $m \rightarrow \infty$,

$$m^{-\frac{1}{2}} \sum_{i=1}^m \nabla l_i^*(\boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_1).$$

By condition (iv) in Theorem 3.3 and the consistency of $\hat{\boldsymbol{\delta}}$, we further have

$$-m^{-1} \sum_{i=1}^m \nabla^2 l_i^*(\boldsymbol{\delta}^*) \xrightarrow{P} \boldsymbol{\Gamma}_2.$$

Combining these results and using Slutsky's theorem, we have

$$\sqrt{m}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1})$$

as $m \rightarrow \infty$ and $n \rightarrow \infty$. \square

Received 21 February 2016

REFERENCES

- [1] BEAL, S. L. and SHEINER, L. B. (1988). Heteroscedastic nonlinear regression. *Technometrics* **30** 327–338.
- [2] CAO, C. Z. and LIN, J. G. (2012). Heteroscedasticity and/or autocorrelation checks in longitudinal nonlinear models with elliptical and AR(1) errors. *Acta. Math. Appl. Sin.-E.* **28** 49–62. [MR2864351](#)
- [3] CASELLA, G. and BERGER, R. L. (2001). *Statistical inference*, 2nd ed. Duxbury Press.
- [4] DAVIDIAN, M. and GILTINAN, D. M. (1993). Some general estimation methods for nonlinear mixed-effects models. *J. Biopharm. Stat.* **3** 23–55.
- [5] DAVIDIAN, M. and GILTINAN, D. M. (1993). Some simple methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics* **49** 59–73.
- [6] DIGGLE, P. J., HEAGERTY, P., LIANG, K. Y. and ZEGER, S. L. (2002). *Analysis of longitudinal data*, 2nd ed. Oxford University Press, USA. [MR2049007](#)
- [7] FOULLEY, J. L., CRISTOBAL, M. S., GIANOLA, D. and IM, S. (1992). Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Comput. Stat. Data Anal.* **13** 291–305.
- [8] GREENE, W. H. (2011). *Econometric analysis*, 7th ed. Pearson Education.

- [9] HARLOW, S. D. and MATANOSKI, G. M. (1991). The association between weight, physical activity, and stress and variation in the length of the menstrual cycle. *Am. J. Epidemiol.* **133** 38–49.
- [10] HEDEKER, D., MERMELSTEIN, R. J., BERBAUM, M. L. and CAMPBELL, R. T. (2009). Modeling mood variation associated with smoking: an application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction*. **104** 297–307.
- [11] JENNRICH, R. I. and SCHLUCHTER, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42** 805–820. [MR0872961](#)
- [12] KE, C. L. and WANG, Y. D. (2001). Semiparametric nonlinear mixed-effects models and their applications. *J. Amer. Statist. Assoc.* **96** 1272–1281. [MR1946577](#)
- [13] KWAN, K. C., BREAUULT, G. O., UMBENHAUER, E. R., MCMAHON, F. G. and DUGGAN, D. E. (1976). Kinetics of indomethacin absorption, elimination and enterohepatic circulation in man. *J. Pharmacokinetic. Biopharm.* **4** 255–280.
- [14] LAI, T. L. and SHIH, M. C. (2003). A hybrid estimator in nonlinear and generalized linear mixed effects models. *Biometrika* **90** 859–879. [MR2024762](#)
- [15] LI, B. Y., BRUYNEEL, L. and LESAFFRE, E. (2014). A multivariate multilevel Gaussian model with a mixed effects structure in the mean and covariance part. *Stat. Med.* **33** 1877–1899. [MR3256909](#)
- [16] LIN, X. H., RAZ, J. and HARLOW, S. D. (1997). Linear mixed models with heterogeneous within-cluster variances. *Biometrics* **53** 910–923. [MR1475054](#)
- [17] LINDSTROM, M. J. and BATES, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46** 673–687. [MR1085815](#)
- [18] LIU, W. and WU, L. (2010). Some asymptotic results for semiparametric nonlinear mixed-effects models with incomplete data. *J. Stat. Plan. Infer.* **140** 52–64. [MR2568121](#)
- [19] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional data analysis*, 2nd ed. Springer, New York. [MR2168993](#)
- [20] RODBARD, D. and FRAZIER, G. R. (1975). Statistical analysis of radioligand assay data. *Methods Enzymol.* **37** 3–22.
- [21] SERFLING, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. [MR0595165](#)
- [22] SHEINER, L. B. and BEAL, S. L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J. Pharmacokinetic. Biopharm.* **8** 553–71.
- [23] SHERMAN, M. and LE CESSIE, S. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Commun. Statist.-Simula.* **26** 901–925. [MR1467108](#)
- [24] SOLOMON, P. J. and COX, D. R. (1992). Nonlinear component of variance models. *Biometrika* **79** 1–11. [MR1158513](#)
- [25] TIAN, M. Z., TANG, M. L. and CHAN, P. S. (2009). Semiparametric quantile modelling of hierarchical data. *Acta. Math. Sin-E.* **25** 597–616. [MR2495512](#)
- [26] VONESH, E. F. (1992). Non-linear models for the analysis of longitudinal data. *Stat. Med.* **11** 1929–1954.
- [27] VONESH, E. F. (1996). A note on the use of Laplace’s approximation for nonlinear mixed-effects models. *Biometrika* **83** 447–452. [MR1439795](#)
- [28] WALKER, S. (1996). An EM algorithm for nonlinear random effects models. *Biometrics* **52** 934–944. [MR1411741](#)
- [29] WANG, H. J. (2009). Inference on quantile regression for heteroscedastic mixed models. *Stat. Sinica.* **19** 1247–1261. [MR2536154](#)
- [30] WANG, T. and WU, L. (2013). Multivariate one-sided tests for nonlinear mixed-effects models. *Can. J. Stat.* **41** 453–465. [MR3101594](#)
- [31] WANG, Y. D. (1998). Mixed-effects smoothing spline analysis of variance. *J. Roy. Stat. Soc. B.* **60** 159–174. [MR1625640](#)
- [32] WONG, R. (2001). *Asymptotic approximations of integrals*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia. [MR1851050](#)
- [33] WU, H. L. and ZHANG, J. T. (2002). The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Stat. Med.* **21** 3655–3675.
- [34] WU, H. L. and ZHANG, J. T. (2006). *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. John Wiley & Sons, Inc. [MR2216899](#)
- [35] ZEGER, S. L. and DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

Chuoxin Ma
 School of Mathematics
 University of Manchester
 Manchester, M13 9PL
 UK
 E-mail address: chuoxin.ma@manchester.ac.uk

Maozai Tian
 Center for Applied Statistics
 School of Statistics
 Renmin University of China
 Beijing
 100872
 P.R. China
 School of Statistics
 Lanzhou University of Finance and Economics
 Lanzhou
 730101
 Gansu
 P.R. China
 E-mail address: mztian@ruc.edu.cn

Jianxin Pan
 School of Mathematics
 University of Manchester
 Manchester, M13 9PL
 UK
 E-mail address: Jianxin.Pan@manchester.ac.uk