Kiranmoy Das*, Prince Afriyie and Lauren Spirko

A Semiparametric Bayesian Approach for Analyzing Longitudinal Data from Multiple Related Groups

DOI 10.1515/ijb-2015-0002

Abstract: Often the biological and/or clinical experiments result in longitudinal data from multiple related groups. The analysis of such data is quite challenging due to the fact that groups might have shared information on the mean and/or covariance functions. In this article, we consider a Bayesian semiparametric approach of modeling the mean trajectories for longitudinal response coming from multiple related groups. We consider matrix stick-breaking process priors on the group mean parameters which allows information sharing on the mean trajectories across the groups. Simulation studies are performed to demonstrate the effectiveness of the proposed approach compared to the more traditional approaches. We analyze data from a one-year follow-up of nutrition education for hypercholesterolemic children with three different treatments where the children are from different age-groups. Our analysis provides more clinically useful information than the previous analysis of the same dataset. The proposed approach will be a very powerful tool for analyzing data from clinical trials and other medical experiments.

Keywords: cholesterol level, linear mixed model, longitudinal response, matrix stick-breaking process, MCMC

1 Introduction

Almost all biological, agricultural and clinical experiments give longitudinal responses where we get multiple measurements from each subject at different time points. In clinical trials, treatments are given to subjects under study and the outcomes (typically continuous but not necessarily) are measured at many different time points to infer the effectiveness or toxicity of the drugs administered. Longitudinal data provide more powerful inference than a single time-point data for obvious reasons.

Despite the difficulty, many parametric, nonparametric and semiparametric approaches have been proposed for modeling the mean trajectory and the underlying covariance structure of the longitudinal response. Pourahmadi [1] proposed a likelihood based approach of estimating the mean function and the covariance matrix based on Cholesky decomposition. Pan and Mackenzie [2] extended Pourahmadi's approach to irregular sparse longitudinal data. Mao et al. [3] proposed a joint mean-covariance estimation with basis function approximation. Lin and Pan [4] proposed a modified local linear smoothing estimation technique for nonparametric mean-covariance estimation of longitudinal data. Das et al. [5, 6] proposed semiparametric approaches of modeling the covariance structures for univariate and bivariate sparse longitudinal data.

In many clinical trials, response comes from multiple related groups and hence the features of the mean structure (parameters related to the mean function) might be similar or same across the groups. This similarity might occur for various (unknown) reasons, e.g. the similarity in the covariate values which makes it almost impossible to measure such similarity manually. However, in the presence of such similarity in the mean structure, the traditional approach of considering either exactly the same or completely different means for the groups might result in poor inference. The matrix stick-breaking process (MSBP) proposed in Dunson et al. [7] addressed this issue with a Bayesian approach. More recently matrix stick-breaking priors are adapted to simultaneous covariance estimation for longitudinal data [8, 9]. These authors have shown the superiority of matrix stick-breaking priors for information sharing across the groups. Here we consider a similar semiparametric approach for efficiently modeling the mean trajectories of the related groups.

***Corresponding author: Kiranmoy Das,** Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata 700108, India, E-mail: kiranmoy.das@gmail.com

Prince Afriyie, Lauren Spirko, Department of Statistics, Temple University, Philadelphia PA, 19122

We analyze the data from nutrition education for the hypercholesterolemic children of the United States. Tershakovec et al. [10] analyzed this data as a collaborative research project of Children's Hospital of Philadelphia, The Pennsylvania State University and Abington Memorial Hospital. The goal of the project was to study the effectiveness of the newly developed parent-child autotutorial program over the usual counseling program. A cholesterol screening program was conducted to identify "at-risk" children having total cholesterol level above 176 μg/dL and LDL cholesterol level 107–164 μg/dL for boys and 112–164 μg/dL for girls. These children were randomized into one of the two nutrition education intervention groups (parent-child autotutorial and counseling) or "at-risk control" group. Food habits, lipid intake and the LDL cholesterol level were assessed for all groups at the baseline and after 3, 6 and 12 months. The participants were majorly white americans belonging to middle and upper socioeconomic levels and most of the children were living with both biological parents. We refer to Tershakovec et al. [10] for a more detailed explanation of the dataset and the protocol.

Previous study of this data [10] used a simple linear mixed model and inferred that the parent-child autotutorial is not much effective in the long run for lowering LDL cholesterol of the children. However, the children were not from the same age group and thus the maturity level of the participating children might differ across their race and gender. Here we categorize the children into 3 different age groups (<5 years, 5–8 years and >8 years) and study the effect of the treatments on each group.

We use a linear mixed model for our analysis following Laird and Ware [11], Li et al. [12] and Das and Daniels [8], where the LDL cholesterol is taken as the response variable. Race and gender are taken as covariates and the fixed and the random effects of time are also considered in the model. For the fixed effects of time, we consider a smooth function (twice differentiable) which is modeled by Penalized splines (P-splines). Random effects are assumed to be normally distributed. We consider an autoregressive of order 1 (AR [1]) covariance structure for simplicity. We perform a Bayesian analysis where the model parameters are estimated and evaluated by MCMC.

The current article is organized as the following. In section 2, we explain the proposed model and the appropriate prior structures for the matrix-stick breaking process. Results from the simulation studies are shown and discussed in section 3. In section 4, we perform the data analysis and compare the results to Tershakovec et al. [10]. Some concluding remarks are given in section 5.

2 Model and methods

2.1 Linear mixed model

Suppose we have longitudinal response measurements from M related groups and the m-th group consists of n_m subjects. The total number of subjects under study is n and thus $n = \sum_{m=1}^{M} n_m$. In our data analysis, we have three different age groups. Let the i-th subject belonging to the m-th group receive the k-th treatment and we get T measurements from each subject. The longitudinal response from the i-th subject belonging to the m-th group receiving the k-th treatment can be modeled as the following:

$$y_{im}^{(k)}(t) = f_m^{(k)}(t) + x_i^T \boldsymbol{\beta} + z_i^T \boldsymbol{b}_{\mathbf{i}} + e_{im}(t); \tag{1}$$

where the smooth (twice differentiable) function $f_m^{(k)}$ is the general effect of time on the m-th group, $\boldsymbol{\beta}$ is the vector of covariate effects and x_i is the corresponding design vector. The vector of subject-specific random effects is $\boldsymbol{b_i}$ with the design vector z_i. In matrix notation, we can write the above model as:

$$\mathbf{Y}_{im}^{(k)} = \mathbf{f}_m^{(k)} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{b}_{\mathbf{i}} + \mathbf{e}_{im}, \tag{2}$$

where $\mathbf{X}$ and $\mathbf{Z}_i$ are the design matrix corresponding to the covariates and subject-specific random effects and the vector of residuals e_{im} are assumed to independently follow multivariate normal with mean vector 0

and covariance matrix Σ. We model Σ as a simple autoregressive of order 1 structure with parameters σ^2 and ρ. We also assume that $\boldsymbol{b_i}$s are independently distributed as $N(0, D)$ for some unknown covariance matrix D for which we consider a Wishart prior.

2.2 Modeling the general effect of time and prior structures

We model the group-specific general effect of time $f_m^{(k)}(.)$ in a flexible approach using penalized splines. We consider P-splines of degree r with knots $(\tau_1, \ldots, \tau_S)$. The following regression model is used for our modeling:

$$f_m^{(k)}(t) = a_{0m}^{(k)} + a_{1m}^{(k)} t + a_{2m}^{(k)} t^2 + \ldots + a_{rm}^{(k)} t^r + \sum_{s=1}^{S} c_{sm}^{(k)} (t - \tau_s)_+^r, \tag{3}$$

where $(x)_+^r = x^r I(x > 0)$ and $(\tau_1 < \tau_2 < \ldots < \tau_S)$ is a fixed set of knots. Following Ruppert, Wand and Caroll [13], knots are selected based on evenly spaced sample quantiles of time. Note that S can also be left as an unknown parameter to be estimated from the data [14]. To avoid the computational complexity, we use the standard approach of selecting the optimal S. The optimal r is chosen from the posterior distribution as discussed below.

The proposed prior for $\boldsymbol{a}_m^{(k)} = (a_{0m}^{(k)}, a_{1m}^{(k)}, \ldots, a_{rm}^{(k)})^T$ is based on the matrix stick-breaking process (MSBP) originally proposed by Dunson et al. [7]. This process allows information sharing across M related groups (different age-groups in our case) by incorporating dependence in the prior distributions of $\mathbf{a}_m^{(k)}$, which is an $(r+1)$-dimensional parameter vector for the m-th group. For $j = 0, \ldots, r$, we consider the following prior:

$$a_{jm}^{(k)} \sim F_{jm}^{(k)} = \sum_{g=1}^{N^*} \pi_{jmg}^{(k)} \delta_{\varepsilon_{jg}^{(k)}}(.); \;\; m = 1, \ldots, M; \;\; j = 0, \ldots, r; \;\; k = 1, \ldots, K;$$

$$\varepsilon_{jg}^{(k)} \overset{\text{iid}}{\sim} F_{0j}^{(k)},$$

where δ_c is just a point mass at c. Let $\varepsilon^{(k)} = (\varepsilon_{jg}^{(k)})$ be an $(r+1) \times N^*$ matrix of random atoms. We note that the rows of $\varepsilon^{(k)}$ correspond to the parameters having base distribution $F_{0j}^{(k)}$ and the columns correspond to N^* clusters. Following Dunson et al. [7], $\pi_{jmg}^{(k)}$ are split as the following:

$$\pi_{jmg}^{(k)} = V_{jmg}^{(k)} \prod_{l<g} (1 - V_{jml}^{(k)}), \tag{4}$$

where $V_{jmg}^{(k)} = U_{mg}^{(k)} W_{jg}^{(k)}$ and $U_{mg}^{(k)} \overset{\text{iid}}{\sim} Beta(1, \alpha_1)$ and $W_{jg}^{(k)} \overset{\text{iid}}{\sim} Beta(1, \alpha_2)$. Here we note that stick breaking weights are partitioned into two components, $U^{(k)}$ and $W^{(k)}$, assigning the coefficients from m-th group and j-th parameter component respectively to the g-th cluster. Note that $F_{jm}^{(k)}$ will be a valid probability measure when $V_{jmN^*}^{(k)} = 1$, for all m and j. We consider a finite N^*, which is typically called a truncated MSBP. When $N^* = \infty$, this is called a full matrix stick-breaking process.

The most important task in the truncation approximation of the matrix stick breaking process is the selection of the optimum N^*. This is typically done by making the expected approximation error arbitrarily small [15]. In our case, following Dunson et al. [7], we can show that $E\left(\sum_{g=N^*}^{\infty} \pi_{jmg}^{(k)}\right) = \left[1 - \frac{1}{(1+\alpha_1)(1+\alpha_2)}\right]^{N^*-1}$. Thus, we choose N^* for which the r.h.s. of the above expression is below 0.01.

Following Gaskins and Daniels [9], Das and Daniels [8], we consider a zero-inflated normal base distribution for $\boldsymbol{a}_m$,

$$\varepsilon_{jg}^{(k)} \sim (1 - B_j^{(k)})\delta_0(.) + B_j^{(k)} N(0, \sigma_a^2);\ g = 1, \ldots, N^*;\ j = 0, \ldots, r;$$
$$B_j^{(k)} = \prod_{l=0}^{j} A_l^{(k)};\ A_j^{(k)} | \pi_j^{(k)} \overset{\text{ind}}{\sim} Bernoulli(\pi_j^{(k)});\ \pi_j^{(k)} \overset{\text{iid}}{\sim} Uniform(0,1); j = 0, \ldots, r. \tag{5}$$

We note that in the above prior formulation, $B_j^{(k)}$ is a product of $j+1$ independent Bernoulli random variables; thus, $B_j^{(k)}$ is binary taking values 0 and 1 only. Note that $B_j^{(k)}$ is 1 only when all the lower lag coefficients are 1 and 0 otherwise. This prior, thus, will select a model where there will be no nonzero higher order terms with zero lower order terms. The optimal r can be obtained from the posterior mode, and hence we can avoid the traditional two-step method where the optimal r is chosen using AIC/BIC and then the model fitting is performed. The prior probability of getting an r-th order spline function is:

$$Pr(Order = r) = Pr(B_r^{(k)} = 1,\ B_{r+1}^{(k)} = 0) = \prod_{i=0}^{r} \pi_i^{(k)} (1 - \pi_{r+1}^{(k)}).$$

We assume a multivariate normal prior with mean vector $\mathbf{0}$ and covariance matrix Σ_c for the vector $c_m^{(k)} = [c_{1m}^{(k)}, \ldots, c_{Sm}^{(k)}]^T$, where Σ_c is a diagonal matrix with each diagonal element $\frac{1}{\lambda}$. We take Gamma(α^*, β^*) prior for λ. This prior specification will shrink the roughness of splines towards 0, which is indeed the Bayesian equivalence of frequestist approach to penalized splines. For σ_a^2 and σ^2, we consider inverse gamma priors. Uniform [–1, 1] prior is taken for ρ, and we consider MVN$(0, \Sigma_\beta)$ prior for $\boldsymbol{\beta}$. Following the traditional approach, we consider Wishart prior for the random effects covariance matrix D.

Posterior distributions of the model parameters are obtained following Dunson et al. [7] and Das and Daniels [8]. Model parameters are estimated by MCMC. Sensitivity analyses are performed routinely to assess the effect of the priors on the parameter estimates. Convergence of the chains are assessed following Brooks and Gelmen [16]. Details of these computations related to our data analysis are provided in section 4.

3 Simulation Studies

3.1 Simulation 1

The operating characteristics of the proposed modeling approach are investigated through simulation studies. We consider a sample of size 100 from three related groups with sample sizes 25, 35 and 40, respectively. We measure the longitudinal trait at 5 evenly spaced time points and simulate data from the following linear mixed model:

$$y_{im}(t) = f_m(t) + \beta_1 x_1 + \beta_2 x_2 + b_{i0} + b_{i1} t + e_{im}(t), \tag{6}$$

where $f_m(t) = a_{0m} + a_{1m} t + a_{2m} t^2$, $t = 1, \ldots, 5$, $m = 1, 2, 3$. Two covariates x_1 and x_2 are considered in the model. The continuous covariate x_1 is simulated from Gamma(1.5,3) for all 100 subjects and the other covariate x_2 is simulated from a Bernoulli distribution with $p = 0.4$. For the simulation purpose, we take $\beta_1 = 1.35$ and $\beta_2 = 2.6$. We assume the vector of random effects $b_i = [b_{i0}, b_{i1}]^T$ has MVN$(0, D)$ distribution where D comes from Wishart$(V, 3)$ prior. V is a 2×2 matrix with diagonal elements 2.5 and off-diagonal elements 1. Also we assume MVN$(0, \Sigma)$ distribution for residual vectors where Σ is AR [1] with $\sigma^2 = 2.25$ and $\rho = 0.6$. We simulate data using the following values for the fixed effect parameters:

$$a_{01} = 1.13,\ a_{11} = 1.18,\ a_{21} = 0.87;\ a_{02} = 1.15,\ a_{12} = 1.23,\ a_{22} = 0.84;\ a_{03} = 1.10,\ a_{13} = 1.16,\ a_{23} = 0.90.$$

Note that some of these parameters are similar/same across the three groups.

We fit three different models to the simulated data for assessing the effectiveness of the proposed approach. First, we assume that the parameters related to the general effect of time are exactly the same across all the groups (Model I). Second, we fit the model where the above mentioned parameters are completely different for all groups (Model II). Finally we fit the proposed model which allows information sharing across the groups (Model III). For each approach, we estimate the underlying model parameters and calculate the residual sum of squares (RSS).

We consider 200 replicated datasets and for each data, RSS is computed. We finally calculate the average RSS based on the 200 replicates. Table 1 shows the average RSS for the three modeling approaches. We note that RSS is maximum for Model II which considers the parameters to be completely different for all the groups. The RSS for Model I is lower than that of Model II. Model III provides the smallest RSS among all three models. In Table 2, we provide the average width of the credible intervals, estimated biases and coverage probabilities of the estimated mean parameters for group-1. It shows that Model III gives the parameter estimates with shorter credible intervals, much lower biases and comparable coverage probabilities than the other two models. Similar results are obtained for the other two groups. Thus the effectiveness of the proposed approach (in presence of the covariates) is assessed when the parameters are actually similar/same across the groups under consideration.

Table 1: Average RSS for different models in Simulation 1.

Model	RSS
Model I	28.73
Model II	35.91
Model III	19.64

Table 2: Effects of different models on the average mean parameters for group-1 in Simulation 1.

Model	a_0		a_1		a_2	
	Bias	Width of C.I. (Cov. Prob.)	Bias	Width of C.I. (Cov. Prob.)	Bias	Width of C.I. (Cov. Prob.)
Model I	0.38	0.94 (0.96)	0.44	0.83(0.95)	0.49	1.12 (0.95)
Model II	0.54	1.04 (0.96)	0.63	0.99(0.95)	0.68	1.24 (0.96)
Model III	0.25	0.69 (0.95)	0.31	0.62(0.94)	0.32	0.79 (0.94)

In Figure 1, we show the posterior distribution of r, the optimal degree of the P-splines. As we see, the posterior mode corresponds to the true order ($r = 2$). Hence our proposed prior structure selects the optimal degree avoiding the traditional two-step approach using the information criteria (AIC/BIC).

3.2 Simulation 2

We perform a second simulation study to investigate the usefulness of the proposed approach for irregular longitudinal measurements. Here we assume that different subjects are measured at different time points and the number of measurements vary from 1 to 5. Thus for the i-th subject, we first randomly select an integer number from 1 to 5 (including 1 and 5) which gives the number of measurements (t_i) for that subject. Next, we again randomly select t_i integer points ($1 \le t_i \le 5$) and get the time points at which this subject is measured. The rest of the simulation method is very similar to Simulation 1, i.e. we consider 100 subjects from 3 different groups. The longitudinal trait values are then simulated from the following mixed model:

$$y_{im}(t_{ij}) = f_m(t_{ij}) + \beta_1 x_1 + \beta_2 x_2 + b_{i0} + b_{i1} t_{ij} + e_{im}(t_{ij}), \tag{7}$$

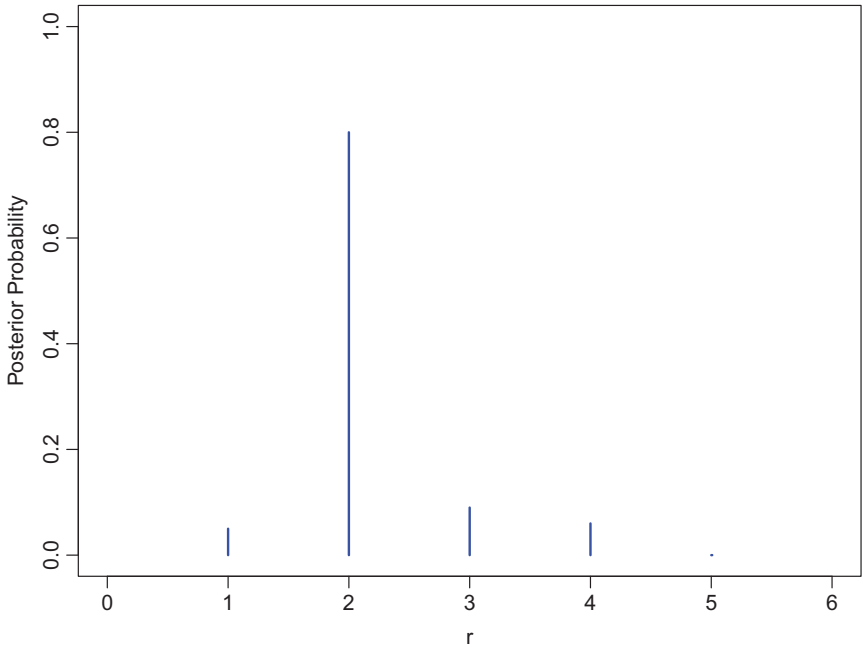

Figure 1: Posterior probabilities for r in Simulation 1.

where $f_m(t_{ij}) = a_{0m} + a_{1m}t_{ij} + a_{2m}t_{ij}^2$, $i = 1, \ldots, 100$, $j = 1, \ldots, t_i$, $m = 1, 2, 3$. The covariates x_1 and x_2 are simulated as in Simulation 1. The residual vectors are assumed to follow MVN$(0, \Sigma_i)$, where Σ_i is measured by an auto-regressive structure of order 1. For simulating the trait values, we consider the same parameters values as the previous simulation study. The same procedures are applied for estimating the model parameters for three different models described earlier.

For assessing the effectiveness of our approach in case of the irregular longitudinal data, we compute average RSS based on 200 replicated datasets. The average RSS values for Model I and Model II are 39.61 and 46.58 respectively. For Model III, the average RSS is computed as 24.19. Note that the smallest RSS values are obtained for Model III. Although the RSS values for this case are higher than the regular longitudinal case, the performance of Model III is still better than the other two models. In Table 3, we report the average biases, average widths of the credible intervals and the estimated coverage probabilities for the three parameters related to the general effect of time. Clearly, it is observed that the proposed model (Model III) gives the best estimates among the three models under consideration.

Table 3: Effects of different models on the average mean parameters for group-1 in Simulation 2.

Model	a_0		a_1		a_2	
	Bias	Width of C.I. (Cov. Prob.)	Bias	Width of C.I. (Cov. Prob.)	Bias	Width of C.I. (Cov. Prob.)
Model I	0.43	1.06(0.95)	0.51	0.94(0.95)	0.48	1.10(0.95)
Model II	0.56	1.13(0.96)	0.72	1.16(0.96)	0.74	1.35(0.95)
Model III	0.29	0.78(0.94)	0.35	0.77(0.95)	0.36	0.78(0.94)

Comparing Tables 2 and 3, we notice that for the regular longitudinal data, the estimated biases are lower, the credible intervals are shorter than the irregular data case. This is possibly due to the measurement error due to the irregular time points. Nevertheless the performance of the proposed approach is consistent in both cases.

4 Data analysis

We analyze the data from the nutrition education program discussed in Tershakovec et al. [10]. A total of 3,652 children (of age < 11 years) were screened and 997 were found to have an elevated total cholesterol level. Out of those, 924 children met the eligibility criteria. Next step was to perform a confirmatory test on the children meeting the eligibility criteria and only 458 agreed to participate in the confirmatory test. After the test, 271 children were confirmed to have a high LDL cholesterol level and were randomized into one of the three treatments (Control, Counseling and Auto-tutorial). After being informed on the random assignment, 38 children never returned and thus we had a total of 233 children who completed the program.

Our dataset consists of 233 children from 3 age-groups (<5 years, 5–8 years and >8 years). This classification is based on the information of the "type of school" the children belong to; <5 is typically pre-kindergarten age, 5–8 is the kindergarten age (kindergarten age is typically 5–7 though, but we include 8 just to make the groups more or less of equal size) and >8 is the age for the upper-grade elementary school. The goal is to see the effect of the treatments on the different age-groups allowing the possibility of information sharing on the mean trajectories across the groups.

4.1 Model and priors

We fit the following linear mixed model:

$$y_{im}^{(k)}(t) = f_m^{(k)}(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{i0} + b_{i1} t + e_{im}(t), \tag{8}$$

where $y_{im}^{(k)}(t)$ is the LDL cholesterol level at time t from the i-th child belonging to the m-th age-group (m = 1,2,3) receiving the k-th treatment (k = 1,2,3). Also $f^{(k)}(.)$ is the average effect of time which is modelled using P-splines as discussed earlier. The covariates x_1 (1 for Whites, 2 for Blacks and 3 for Others) and x_2 (1 for Males and 2 for Females) denote the race and gender of the subject respectively, b_{i0} and b_{i1} denote the subject-specific random intercept and random slope respectively. The residual vectors e_{im} are assumed to follow MVN$(0, \Sigma)$ where we model Σ with AR [1] structure with parameters ρ and σ^2. We also assume that the vectors $b_i = [b_{i0}, b_{i1}]^T$ are iid $N(0, D)$.

Knots of the P-splines are selected from evenly spaced sample quantiles. In our analysis, we consider 2 knots (K = 2). This specification is chosen since it has the smallest DIC value (defined later in section 4.2). We consider Uniform [–1, 1] prior for ρ, inverse gamma (1.5,2.5) prior for σ^2 and independent gamma (1,1) prior for α_1, α_2. Gamma (2,3.5) prior is taken for the penalty parameter λ of the P-splines. We consider the commonly used prior distributions for the P-spline parameters [5, 6]. For D, we consider Wishart$(V, 3)$ prior [8] where V is a 2×2 matrix with diagonal elements 3 and off-diagonal elements 1.8.

4.2 Model selection and computational details

Following the traditional approach, we draw the starting values of the parameters for the MCMC algorithm from their respective prior distributions. We run the chains with 110,000 iterations and discard the first 10,000 `burn-in' iterations. Also we thin the chains by keeping every 10-th iteration. We follow the recommendation in Brooks and Gelman [16] to assess the convergence of the chains by considering 5 independent chains for each parameter. All the computed potential scale reduction factors were below 1.3, indicating that the chains have converged.

Note that for the matrix stick-breaking prior for $\mathbf{a}_m^{(k)}$, we need to know the value of the number of clusters N^*. Here we consider $N^* = 20$ since this keeps the expected approximation error smaller than 0.01. We perform sensitivity analyses by investigating the effect of the priors on the estimated parameter values.

Table 4: Sensitivity analysis results for $a_0^{(1)}$.

Parameter	Prior	Estimate (group-1)	Estimate (group-2)	Estimate (group-3)
ρ	Unif [−1,1]	0.87	0.74	0.68
–	Beta [1,2]	0.88	0.72	0.69
–	Unif [2,5]	0.87	0.73	0.70
σ^2	IG (1.5,2.5)	0.86	0.73	0.71
–	IG (3,5.5)	0.88	0.72	0.70
–	IG [2,4]	0.85	0.75	0.69
α_1	Gamma [1,1]	0.87	0.73	0.70
–	Gamma [3,5]	0.88	0.72	0.73
–	Gamma (1.5,4.5)	0.86	0.73	0.69
α_2	Gamma [1,1]	0.86	0.75	0.69
–	Gamma [2,5]	0.84	0.74	0.69
–	Gamma (1.5,5)	0.87	0.75	0.70
λ	Gamma (2,3.5)	0.87	0.74	0.71
–	Gamma (4,5.5)	0.89	0.73	0.70
–	Gamma (2,4.5)	0.88	0.72	0.71

As shown in Table 4, for different gamma and/or inverse gamma priors for the hyperparameters, the estimated parameter values don't differ much for $a_0^{(1)}$. Similar results are obtained for other mean parameters for each treatment and age-group (results not shown).

We use Deviance Information Criteria (DIC) for model selection. Here we consider three different model specifications (common mean, group-specific mean and proposed approach) and need to choose the optimal one for the data under consideration. Since we are using a mixed model for our analysis, we consider conditional DIC as proposed in Celeux et al. [17]. This DIC is based on the conditional likelihood $l(\mathbf{Y}|\mathbf{b})$ and is given by DIC $= -4E[\log l(\mathbf{Y}|\mathbf{b})] + 2\log l(\mathbf{Y}|\hat{\mathbf{b}})$, where $\mathbf{b}$ denotes the vector of random effects and $\hat{\mathbf{b}}$ is the corresponding estimate (posterior mean). In Table 5, we provide the DIC values and the corresponding effective number of parameters (p_D) for different models. As we see, the proposed approach provides the smallest DIC value compared to the other models.

Table 5: DIC from different model specifications in data analysis.

Model	p_D	DIC
Common mean	8.2	25,732
Group-specific mean	11.6	27,153
Proposed structure	9.4	21,386

p_D = Effective number of parameters in DIC.

The proposed approach also estimates the posterior distribution of r, the optimal polynomial order. In Figure 2, we show the posterior distribution and by considering the posterior mode, we get $r = 2$ as the optimal order. As mentioned earlier, this avoids the traditional two-step method of selecting the optimal r.

4.3 Results

In Figure 3, we show the posterior mean curves of the LDL cholesterol levels for the three different age-groups under consideration. We note that for the children participants below 5 years, PCAT worked very well from baseline till the first evaluation (3 months from baseline) but after the second evaluation (6 months from baseline) counseling worked better for lowering the LDL cholesterol. For the second group

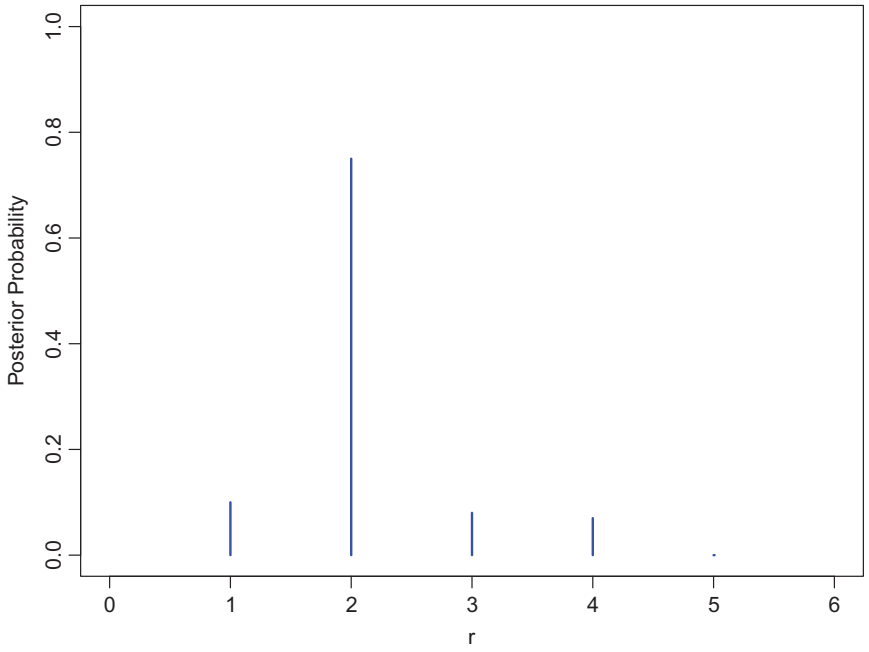

Figure 2: Posterior probabilities for r in data analysis.

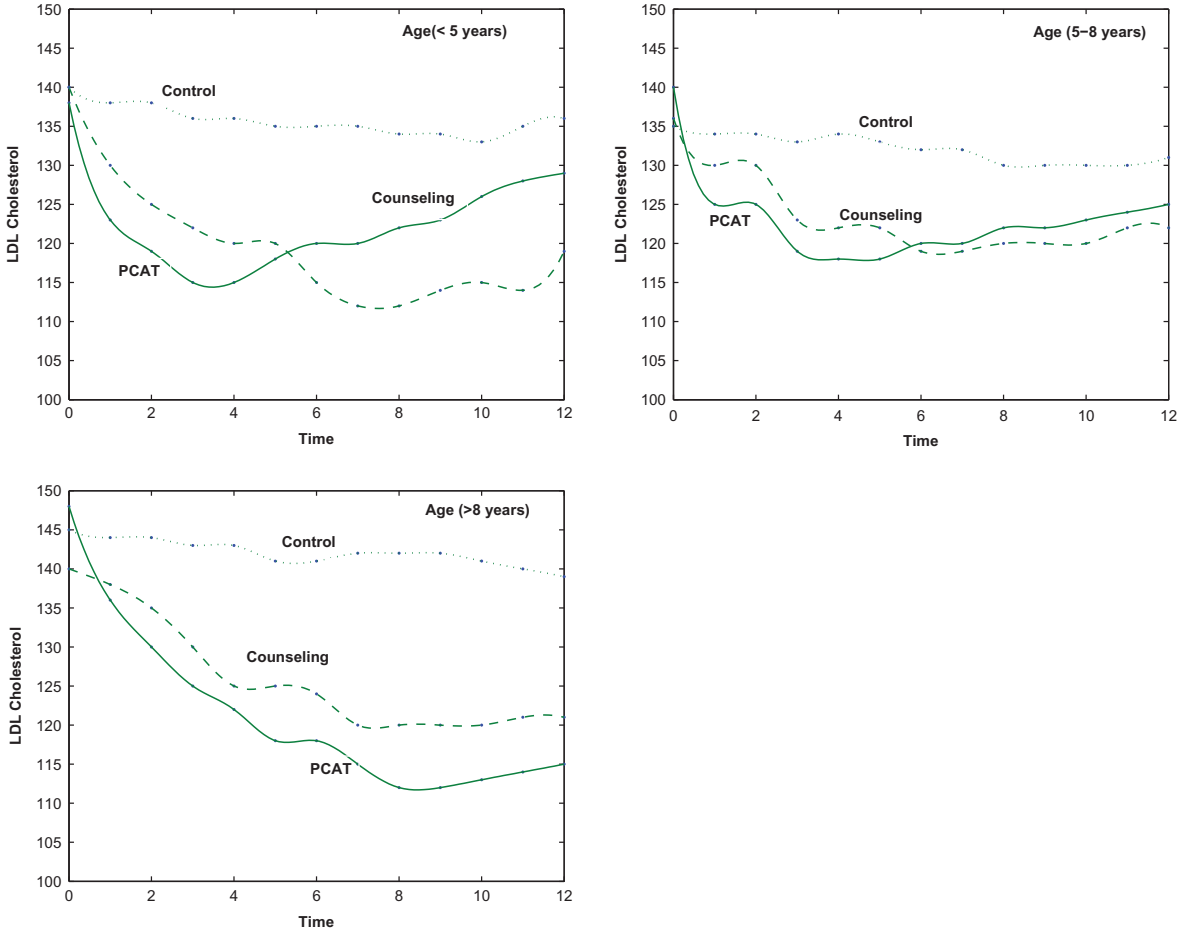

Figure 3: The posterior mean trajectories for three age-groups in data analysis.

(children between 5 and 8 years), initially PCAT worked slightly better but after the second evaluation counseling and PCAT worked almost in the same way. However, for the children participants belonging to the upper age group (above 8 years), PCAT worked better than counseling consistently throughout the study time. Figure 3 clearly explains that it is important to study the effects of the three treatments over different age-groups separately for meaningful inference.

Our results demonstrate that for the younger participants (<5 years) PCAT might be an acceptable initial alternative to nutrition counseling but the effect of autotutorial is not retained for a longer time. For the second group, PCAT and counseling are equally fruitful and for the elder participants (>8 years), PCAT is definitely a better alternative. We note that Tershakovec et al. [10] did a combined analysis of all the age-groups and concluded that PCAT had a significant short-term effect but not-significant long term effect. Therefore they recommended that the national cholesterol education program should consider developing booster educational materials for parents and children. However, our approach suggests that PCAT is more effective for children with age more than 5 years and have a long term effect for these children.

In Figure 4, we show the grouping nature of the proposed priors for the mean parameters. We show the posterior probabilities of $Pr(a_{jm} = a_{jm'})$, for each m, m', j, combination for the children treated by PCAT. The size of each box is proportional to the corresponding posterior matching probabilities. The boxes on the line $y = x$ diagonal correspond to the probability 1. We refer age group <5 years, 5–8 years and >8 years by

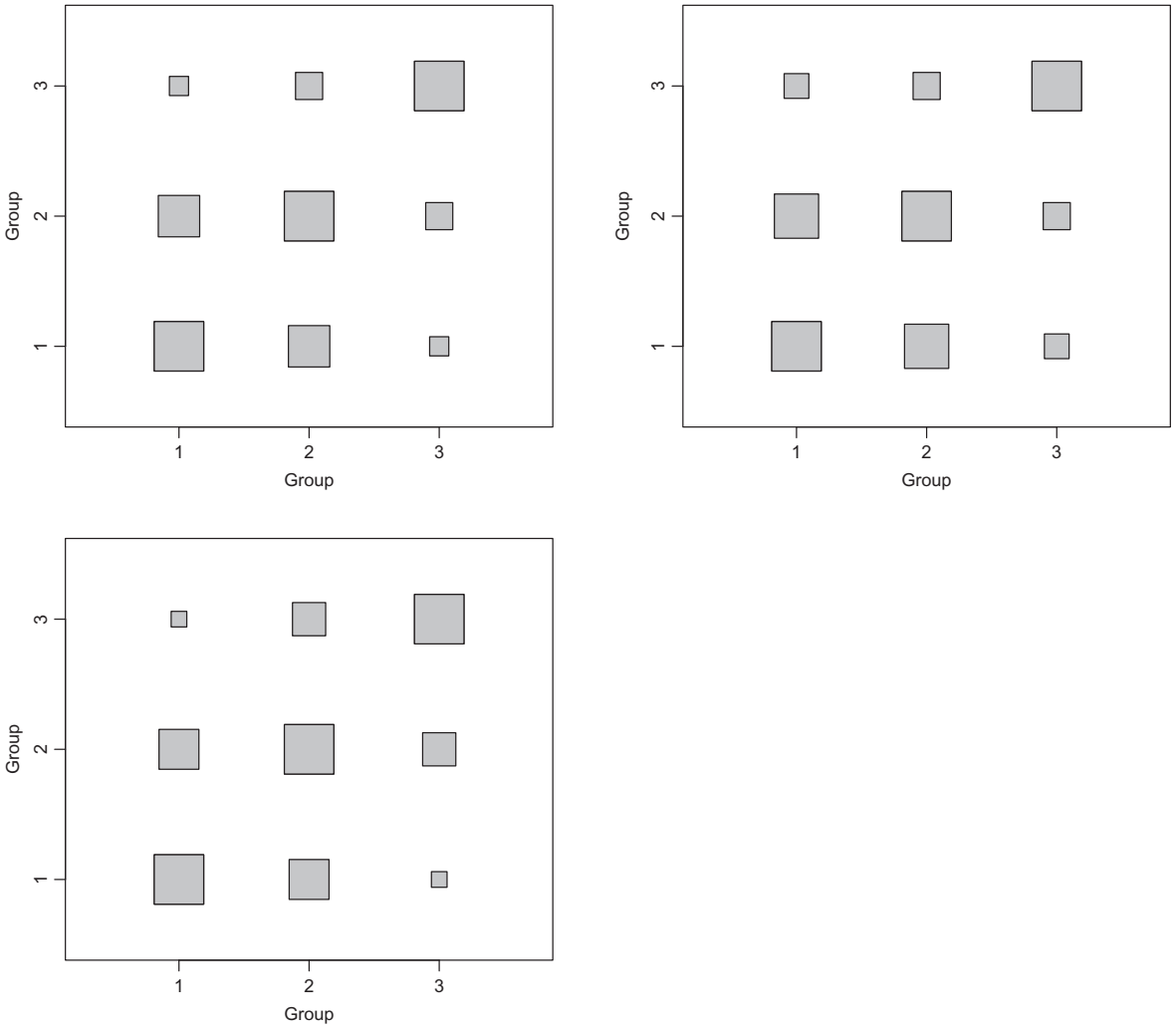

Figure 4: The posterior probabilities of matching for a_0, a_1 and a_2 respectively for PCAT treatment group.

group 1,2 and 3 respectively. We notice that there are higher matching probabilities for the first and second age group. We obtain similar results for children treated by counseling and the control group (results not shown). This may be due to the higher maturity-level of the third group (> 8 years) compared to the other groups. Thus the proposed approach helps to find the structural similarity across the groups in a data-dependent way.

5 Discussions

The proportion of people with a high LDL cholesterol level has increased significantly in the United States over the last decade [18]. Higher LDL cholesterol levels might cause heart and blood vessel diseases, stroke etc. Often, due to family history, high LDL cholesterol begins in childhood. Early detection and precautions (e.g. weight loss, lower fat intake, more physical activities etc.) might control the level of bad cholesterol. Nutrition education programs for the hypercholesterolemic children is an essential tool for preventing future death due to high LDL cholesterol levels. Our current analysis shows that the PCAT had a significant short term effect but a less significant long term effect for children below 5 years. However, for the children above 8 years, PCAT had a long term effect for lowering LDL cholesterol.

In this article, we proposed a semiparametric model for analyzing the data from a one-year follow up of a nutrition education program. Our approach allows different age-groups to share information on mean trajectories and thus provides a more reliable and powerful inference. When different groups actually share information on mean parameters, the proposed approach is more efficient since a smaller number of parameters are to be estimated. When groups differ from each other, the proposed approach becomes close to the model which assumes different parameters for each group.

We note that in this analysis, we did not consider any missing data. In reality one might have a lot of missing responses. However, if the missingness is ignorable (e.g. missing at random), a simple data-augmentation step needs to be added to the proposed approach. For non-ignorable missingness, one needs more complex as approaches proposed in Daniels and Hogan [19].

Acknowledgment: The authors thank Prof. Vern Chinchilli, The Pennsylvania State University, for sharing the data from one-year nutrition education.

References

1. Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. Biometrika 1999;86:677–690.
2. Pan J, Mackenzie G. On modelling mean-covariance structures in longitudinal studies. Biometrika 2003;90:239–44.
3. Mao J, Zhu Z, Fung WK. Joint estimation of mean-covariance model for longitudinal data with basis function approximations. Comput Stat Data Anal 2011;55:983–92.
4. Lin H, Pan J. Nonparametric estimation of mean and covariance structures for longitudinal data. Can J Stat 2013;41:557–74.
5. Das K, Li J, Fu G, Wang Z, Li R, Wu R. Dynamic semi-parametric Bayesian models for genetic mapping of complex traits with irregular longitudinal data. Stat Med 2013a;32:509–23.
6. Das K, Li R, Sengupta S, Wu R. A Bayesian semi-parametric model for bivariate sparse longitudinal data. Stat Med 2013b;32:3899–3910.
7. Dunson DB, Xue Y, Carin L. The matrix stick-breaking process: flexible Bayes meta-analysis. J Am Stat Assoc 2008;103:317–27.
8. Das K, Daniels MJ. A semiparametric approach to simultaneous covariance estimation for bivariate sparse longitudinal data. Biometrics 2014;70:33–43.
9. Gaskins JT, Daniels MJ. A nonparametric prior for simultaneous covariance estimation. Biometrika 2012;100:125–138.
10. Tershakovec AM, Shannon BM, Achterberg CL, et al. One-year follow-up of nutrition education for hypercholesterolemic children. Am J Public Health 1998;88:258–261.

11. Laird NM, Ware JH. Random effects model for longitudinal data. Biometrics 1982;38:963–974.
12. Li Y, Lin X, Müller P. Bayesian inference in semiparametric mixed models for longitudinal data. Biometrics 2010;66:70–78.
13. Ruppert D, Wand MP, Carroll RJ. Semiparametric regression. New York: Cambridge University Press, 2003.
14. Wang H-B. Bayesian estimation and variable selection for single index models. Comput Stat Data Anal 2009;53:2617–2627.
15. Ishwaran H, James LF. Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. J Comput Graph Stat 2002;11:1–26.
16. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 1998;7:434–455.
17. Celeux G, Forbes F, Robert CP, Titterington DM. Deviance information criteria for missing data models. Bayesian Anal 2006;1:651–74.
18. Tattersall MC, Gangnon RE, Karmali KN, et al. Trends in low-density lipoprotein cholesterol goal achievement in high risk united states adults: longitudinal findings from the 1999–2008 national health and nutrition examination surveys. PLOS One 2013;8:1–7.
19. Daniels MJ, Hogan JW. Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis. Chapman and Hall (CRC Press), 2008.

Article note: The authors Prince Afriyie and Lauren Spirko contributed equally to this paper.