REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Longitudinal Data Regression Analysis Using Semiparametric Modelling

Authors: Abdulla Mamun
– Department of Mathematics, Gonzaga University,
Spokane, WA, USA
mamun@gonzaga.edu

Sudhir Paul ✉
– Department of Mathematics and Statistics, University of Windsor,
Windsor, ON, Canada
smjp@uwindsor.ca

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- Zhang, Leng and Tang [1] propose joint parametric modelling of the means, variances, and the correlations by decomposing the correlation matrix via hyperspherical co-ordinates and show that this results unconstrained parameterization, fast computation, easy interpretation of the parameters, and model parsimony. With unconstrained structures, they also suggest future research on modelling the mean, the variance, and the correlations non-parametrically and semiparametrically. In this paper we explore semiparametric modelling via simulations and data analysis. Extensive simulations show that the semiparametric modelling produces similar bias and efficiency properties of the parameter estimates as those by the parametric modelling. However, model selection, using the AIC and the BIC, through the analysis of two real biomedical data sets show significant improvement in model parsimony.

Keywords:

- *B-spline, Hyperspherical co-ordinates, Joint mean-covariance models, Longitudinal data, Model parsimony, Penalized spline, Semiparametric models.*

AMS Subject Classification:

- 62F99, 62G99.

✉ Corresponding author

1. INTRODUCTION

In the statistical literature, the methods to understand the relationship of explanatory variables on each individual outcome variable are well developed and widely applied. However, in most health-related studies given the technological advancement and sophisticated methods of obtaining and storing data, a need to perform joint analysis of mean and covariance parameters simultaneously and accounting for the correlations is in high demand since a good covariance modelling approach improves statistical inference of the mean of interest ([2]- [4]). Furthermore, the covariance structure itself may be of scientific interest [5].

Model parsimony in regression analysis, specially in the context of longitudinal data regression analysis, is important in biomedical fields. Zhang *et al.* [1] propose joint parametric modelling of the means, the variances, and the correlations by decomposing the correlation matrix via hyperspherical co-ordinates and show that this results unconstrained parameterization, fast computation, easy interpretation of the parameters, and model parsimony. Zhang *et al.* ([1], pp 237) comment that the decomposition of the correlation matrix opens many new avenues for future research and that with unconstrained structures, we can model the mean, the variance, and the correlation non-parametrically and semi-parametrically. In this paper we deal with the semiparametric modelling. Our main aim, though, is to find whether semi-parametric modelling improves model parsimony over the parametric modelling approach.

Suppose longitudinal measurements $\mathbf{y}_i = (y_{i1}, \ldots, y_{im_i})'$ and covariate vectors $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i})'$ $(i = 1, \ldots, n)$, with $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ for $j = 1, \ldots, m_i$, collected from n subjects, are observed at times $\mathbf{t}_i = (t_{i1}, \ldots, t_{im_i})'$. In longitudinal data analysis it is important that statistical analysis takes into account that the repeated observations y_{ij}, $j = 1, \ldots, m_i$ are correlated ([2]- [4]). Accordingly we assume that $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \Sigma_i)$, where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})'$, $\Sigma_i = D_i R_i D_i$, $D_i = \text{diag}(\sigma_{i1}, \ldots, \sigma_{im_i})$, and $R_i = (\rho_{ijk})_{j,k=1}^{m_i}$ is the correlation matrix of $\mathbf{y}_i$ with $\rho_{ijk} = \text{corr}(y_{ij}, y_{ik})$ being the correlation between the j^{th} and k^{th} measurements of the i^{th} subject. The main purpose in such longitudinal studies is to estimate the parameters involved in the means, the variances, and the correlation matrices. This can be done by maximizing the log-likelihood or by solving the maximum likelihood estimating equations. However, the constraints involved in the correlation parameters create a challenge. This can be overcome by decomposing the correlation matrix by Cholesky decomposition.

Zhang *et al.* [1] proposed to parametrize the correlation matrix R_i for subject i (we suppress i) via hyperspherical co-ordinates by the Cholesky decom-

position $R = TT'$, where $T = (T_{jk})$ is a lower triangular matrix given by

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ c_{21} & s_{21} & 0 & \cdots & 0 \\ c_{31} & c_{32}s_{31} & s_{32}s_{31} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ c_{m1} & c_{m2}s_{m1} & c_{m3}s_{m2}s_{m1} & \cdots & \prod_{l=1}^{m-1} s_{ml} \end{pmatrix},$$

with $c_{jk} = \cos(\phi_{jk})$ and $s_{jk} = \sin(\phi_{jk})$ are trigonometric functions of angles ϕ_{jk}.

For subject i, the total number of angles ϕ_{ijk} $(1 \leq k < j \leq m_i)$ is $m_i(m_i - 1)/2$, which is the same as that of the free parameters in the correlation matrix. The decomposition of R has several advantages (i) diagonal elements of TT' are 1, and all other elements fall between -1 and 1, (ii) TT' is always non-negative definite, satisfying the requirements of a correlation matrix, and (iii) the angles ϕ_{jk} of T as parameters are unconstrained in the range $[0, \pi)$. It also establishes a hierarchical connection between the correlations and the angles (for further discussion on this see Zhang *et al.*, [1]). They then propose a joint regression model for the means, the variances, and the correlations as

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta}, \quad \log \sigma^2_{ij} = \mathbf{z}'_{ij}\boldsymbol{\lambda}, \quad \text{and} \quad \phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma}, \tag{1.1}$$

where $\mathbf{x}_{ij}$ are the usual known covariates as mentioned earlier, $\mathbf{z}_{ij}$ and $\mathbf{w}_{ijk}$ may contain baseline covariates, as well as polynomials in time (time related to longitudinal data) and their interactions. The unknown regression parameters $\boldsymbol{\beta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ are of dimensions $p \times 1$, $d \times 1$, and $q \times 1$ respectively. In practice we may choose $\mathbf{w}_{ijk}$ as a polynomial of time lag $(t_{ik} - t_{ij})$. Zhang *et al.* [1] estimate the parameters $\boldsymbol{\beta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ of the model in equation (1) via quasi-Fisher scoring algorithm, a review of which is given in Section 2.

For investigating the properties of the estimates of the regression parameters through semiparametric modelling of the means and variances and to study the impact of this to model parsimony we consider three models.
Model 1: the parametric model (Zhang *et al.* [1]) given above (equation (1.1)),
Model 2: a model in which only the means are modelled semiparametrically, which is,

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + f_1(t_{ij}), \quad \log(\sigma^2_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\lambda}, \quad \text{and} \quad \phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma},$$

Model 3: a model in which the means and the variances are modelled semiparametrically, which is,

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + f_1(t_{ij}), \quad \log(\sigma^2_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\lambda} + f_2(t_{ij}), \quad \text{and} \quad \phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma}.$$

In Model 2 and Model 3, $f_1(\cdot)$ and $f_2(\cdot)$ are smooth functions parametrized by regression splines.

As in Zhang *et al.* [1] we decompose the correlation matrix via hyperspherical co-ordinates and as in Lang, Zhang and Pan [6] we use B-spline to estimate

the unknown functions $f_1(\cdot)$ in Model 2 and $f_1(\cdot)$ and $f_2(\cdot)$ in Model 3. Five further investigations were conducted. First of these is to see the performance of the estimators of the regression parameters in terms of bias and efficiency and to see the effect of fixing the knots in spline smoothing. The second is to study the performance of the estimators of the regression parameters when some covariates are correlated, the third is a robustness study where the normality assumption of the error distribution is replaced by a mixture of normal distributions, the fourth is to see whether use of penalized spline results in improved estimation of the non-parametric functions in comparison to using B-spline, and the fifth is to see the effects of estimators by fitting the data generated from the semiparametric model to the parametric model and vice versa.

Section 2 deals with the method of estimation of the parameters of Model 3 in which the correlation matrix is decomposed via hyperspherical co-ordinates and the unknown functions $f_1(\cdot)$ and $f_2(\cdot)$ are estimated using B-spline basis functions. Estimation procedures in Model 2 and Model 1 are discussed as special cases of Model 3. An extensive simulation study is conducted and results are summarized in Section 3. Detailed analysis of two real data sets is given in Section 4 and a discussion follows in Section 5.

2. ESTIMATION IN JOINT SEMIPARAMETRIC MODELS

2.1. Estimation in Model 3 based on B-spline

We estimate the regression parameters of Model 3 (given above) where the means and the variances are modelled semiparametrically. As in Zhang *et al.* [1], we parametrize R_i via hyperspherical co-ordinates. For simplicity, we assume that f_1 and f_2 have the same smoothness property. Without loss of generality, we assume that the domain of t_{ij} is in the interval $[0, 1]$ with partitions $0 = a_0 < a_1 < \cdots < a_{k_n} < a_{k_n+1} = 1$. Using the a_i's as knots, we have $K = k_n + l$ normalized B-spline basis functions of order l that form a basis for the linear spline space. The B-spline basis of order l, $B_i^{(l)}(t)$, is defined as

$$B_i^{(l)}(t) = \frac{t - a_i}{a_{i+l-1} - a_i} B_i^{(l-1)}(t) + \frac{a_{i+l} - t}{a_{i+l} - a_{i+1}} B_{i+1}^{(l-1)}(t), \quad \text{and}$$

$$B_i^{(1)}(t) = \begin{cases} 1, & a_i \le t < a_{i+1} \\ 0, & \text{otherwise.} \end{cases}$$

Note that $B_i^{(l)}(t)$ is a polynomial function of degree $l - 1$. More details on the construction of B-spline basis can be found in Schumaker [7]. Thus $f_1(t)$ and $f_2(t)$ are approximated by $\boldsymbol{\pi}'(t)\boldsymbol{\alpha}$ and $\boldsymbol{\pi}'(t)\tilde{\boldsymbol{\alpha}}$, respectively, where $\boldsymbol{\pi}(t) = (B_1^{(l)}(t), \ldots, B_K^{(l)}(t))'$ is the vector of basis functions and $\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}} \in \mathbb{R}^K$ are the spline coefficient vector. Let $\boldsymbol{\pi}_{ij} = \boldsymbol{\pi}(t_{ij})$. With this notation, the nonlinear regression

models can be linearized as in what follows

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{\pi}'(t_{ij})\boldsymbol{\alpha} = \Pi'_{ij}\boldsymbol{\theta} \quad \text{and} \quad \log(\sigma^2_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\lambda} + \boldsymbol{\pi}'(t_{ij})\tilde{\boldsymbol{\alpha}} = \Upsilon'_{ij}\boldsymbol{\rho},$$

where $\Pi'_{ij} = (\mathbf{x}'_{ij}, \boldsymbol{\pi}'_{ij})$, $\Upsilon'_{ij} = (\mathbf{z}'_{ij}, \boldsymbol{\pi}'_{ij})$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$, and $\boldsymbol{\rho} = (\boldsymbol{\lambda}', \tilde{\boldsymbol{\alpha}}')'$. Suppose $\boldsymbol{\Pi}_i = (\Pi'_{i1}, \Pi'_{i2}, \ldots, \Pi'_{im_i})'$ and $\boldsymbol{\Upsilon}_i = \left(\Upsilon'_{i1}, \Upsilon'_{i2}, \ldots, \Upsilon'_{im_i}\right)'$.

Thus, now the parameters of interest are $\boldsymbol{\theta}$, $\boldsymbol{\rho}$, and $\boldsymbol{\gamma}$. Let $\mathbf{r}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$. Then, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})' = T_i^{-1}D_i^{-1}\mathbf{r}_i \sim N(\mathbf{0}, I_{m_i})$. Denoting l to be the log-likelihood apart from a constant it can be shown that

$$-2l = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(\log\sigma^2_{ij} + \log T^2_{ijj} + \epsilon^2_{ij}).$$

From this, omitting details, by usual derivations, we can show that the maximum likelihood estimating equations for $\boldsymbol{\theta}, \boldsymbol{\rho}$, and $\boldsymbol{\gamma}$ are

$$\begin{aligned}
\boldsymbol{U}_1 &= \sum_{i=1}^{n}\boldsymbol{\Pi}'_i\Delta_i\Sigma_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i\boldsymbol{\theta})) = 0,\\
\boldsymbol{U}_2 &= \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\Upsilon}'_i(\mathbf{q}_i - \mathbf{1}_{m_i}) = 0, \quad \text{and}\\
\boldsymbol{U}_3 &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}}(\epsilon^2_{ij} - 1) + \epsilon_{ij}\sum_{k=1}^{j-1}b_{ijk}\epsilon_{ik}\right] = 0
\end{aligned}$$

respectively, where $\Delta_i = \Delta_i(\boldsymbol{\Pi}_i\boldsymbol{\theta}) = \text{diag}\{\dot{g}^{-1}(\Pi'_{i1}\boldsymbol{\theta}), \ldots, \dot{g}^{-1}(\Pi'_{im_i}\boldsymbol{\theta})\}$, $\dot{g}^{-1}(\cdot)$ is the derivative of the inverse link function $g^{-1}(\cdot)$, $\mu(\cdot) = g^{-1}(\cdot)$, $\mathbf{q}_i = \text{diag}(R_i^{-1}D_i^{-1}\mathbf{r}_i\mathbf{r}'_iD_i^{-1})$, and $b_{ijk} = \sum_{l=k}^{j}\frac{\partial T_{ilk}}{\partial\gamma}a_{ijl}$ with a_{ijl} being the (j,l) element of T_i^{-1}. As in Zhang *et al.* [1] these equations are solved by the quasi-Fisher scoring algorithm which is described in Appendix A.

2.2. Estimation in Model 2 based on B-spline

If we consider a semiparametric model with only the mean having semiparametric term as $g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + f_1(t_{ij})$, $\log(\sigma^2_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\lambda}$, and $\phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma}$, then we need to estimate $\boldsymbol{\theta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ which are obtained by solving

$$\begin{aligned}
\boldsymbol{V}_1 &= \sum_{i=1}^{n}\boldsymbol{\Pi}'_i\Delta_i\Sigma_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i\boldsymbol{\theta})) = 0,\\
\boldsymbol{V}_2 &= \tfrac{1}{2}\sum_{i=1}^{n}\mathbf{Z}'_i(\mathbf{q}_i - \mathbf{1}_{m_i}) = 0, \quad \text{and}\\
\boldsymbol{V}_3 &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\tfrac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}}(\epsilon^2_{ij} - 1) + \epsilon_{ij}\sum_{k=1}^{j-1}b_{ijk}\epsilon_{ik}\right] = 0.
\end{aligned}$$

These equations can also be solved using the algorithm in Appendix A. At convergence the variance-covariance of $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\gamma}}$ are obtained by inverting the Fisher information matrix given in Appendix B. Parameters of Model 1, are estimated, of course, by setting $f_1(t_{i,j}) = 0$ in the above 3 equations.

2.3. Estimation in Model 2 based on penalized Spline

The B-spline methodology, in some applications, produces over fitting of the data (Carroll and Ruppert, [8]). In such cases penalized spline has been used to overcome this (Eilers and Marx, [9]). So, here, we further use the penalized spline in Model 2 instead of the B-spline to see whether it produces improvement in estimation of the non-parametric functions. As in Section 2.2, $f_1(\cdot)$ can be approximated by $\boldsymbol{\pi}'(t)\boldsymbol{\alpha}$. Now, we impose a penalization upon the parameters $\alpha_1, \alpha_2, \ldots, \alpha_K$, so that they are constrained such that $\sum_{i=1}^{K} \alpha_i^2 \leq C$.

With this constraint the log-likelihood apart from a constant can be written as

$$-2l = \sum_{i=1}^{n} \left[\log|\Sigma_i| + (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i\boldsymbol{\theta}))'\Sigma_i^1(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i\boldsymbol{\theta}))\right] - \tau\boldsymbol{\theta}' D\boldsymbol{\theta},$$

where $\tau > 0$ is a constant and $D = \begin{bmatrix} 0_{p\times p} & 0_{p\times K} \\ 0_{K\times p} & I_{K\times K} \end{bmatrix}$.

Using Lagrange multiplier the score equations for $\boldsymbol{\theta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ can be written as

$$\begin{aligned}
\boldsymbol{W}_1 &= -\sum_{i=1}^{n} \boldsymbol{\Pi}_i'\boldsymbol{\Delta}_i\Sigma_i^{-1}\boldsymbol{\Delta}_i(\mathbf{y}_i - \boldsymbol{\mu}_i) + \tau D\boldsymbol{\theta} = 0, \\
\boldsymbol{W}_2 &= \tfrac{1}{2}\sum_{i=1}^{n} \mathbf{Z}_i'(\mathbf{q}_i - \mathbf{1}_{m_i}) = 0, \quad \text{and} \\
\boldsymbol{W}_3 &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \left[\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}}(\epsilon_{ij}^2 - 1) + \epsilon_{ij}\sum_{k=1}^{j-1} b_{ijk}\epsilon_{ik}\right] = 0.
\end{aligned}$$

All these equations can then be solved using the same algorithm as what we use in Section 2.1. All of the block components of the Fisher information matrix remain the same as Model 2 except I_{11} which in this case is

$$I_{11} = -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \sum_{i=1}^{n} \boldsymbol{\Pi}_i'\boldsymbol{\Delta}_i\Sigma_i^{-1}\Delta_i\boldsymbol{\Pi}_i + \tau D.$$

2.4. Knot Selection

The importance of knot selection in spline smoothing work has been well described in two pioneering papers by He, Fung and Zhu [4] and Leng *et al.* [6]. These authors found that knot selection is less critical for the estimation of $\boldsymbol{\beta}$ than for the estimation of the nonparametric functions involved in Model 1 and Model 2 discussed in Section 1. However, in most situations, they found it appropriate to use the sample quantiles of $\{t_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i\}$ as knots. We follow their suggestion and note that the number of knots are not prespecified, rather, it depends on the total sample size $N = \sum_{i=1}^{n} m_i$. Through detailed asymptotic theoretical study Leng *et al.* [6] show that the number of internal knots to be used is the integer part of $N^{1/5}$.

3. SIMULATION STUDY

As indicated in Section 1, an extensive simulation investigation is conducted in this section. Our purpose in this simulation, in addition to the study of the performance of the estimators of the regression parameters in terms of bias and efficiency, to study the effect of fixing the number of knots, the effect of correlated covariates, the effect of misspecifying the error distribution (robustness study), and compare performance of the estimation methods using B-spline and penalized spline. These simulations are performed in sections, 3.1, 3.2, 3.3, and 3.4. A further study, in section 3.5, is conducted to compare the estimators by fitting the data generated from the semiparametric model to the parametric model and vice versa.

3.1. Study 1: Properties of the regression parameters

For this purpose we generate response data from each of the 3 models (Model 1, Model 2, and Model 3)

$$\begin{aligned} y_{ij} &= x_{ij1}\beta_1 + x_{ij2}\beta_2 + e_{ij}, \quad \log(\sigma_{ij}^2) = z_{ij1}\lambda_1 + z_{ij2}\lambda_2; \\ y_{ij} &= x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_1(t_{ij}) + e_{ij}, \quad \log(\sigma_{ij}^2) = z_{ij1}\lambda_1 + z_{ij2}\lambda_2; \end{aligned}$$

and

$$y_{ij} = x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_1(t_{ij}) + e_{ij}, \quad \log(\sigma_{ij}^2) = z_{ij1}\lambda_1 + z_{ij2}\lambda_2 + f_2(t_{ij}).$$

For each model, values of parameters considered were $(\beta_1, \beta_2) = (1, 0.5)$, $(\gamma_1, \gamma_2) = (0.35, 0.5)$ and $(\lambda_1, \lambda_2) = (-0.5, 0.2)$. Following Leng *et al.* [6] we generate the observation times as in what follows.
For each individual we consider a set of scheduled time points $\{0, 1, 2, \ldots, 12\}$.

At each scheduled time, except time 0, each individual has a 20% probability of missing a fixed time point. To make it irregular and unequal time distances for different individuals a uniform $[0, 1]$ random variable is added to a non skipped scheduled time. This results in different observed time points t_{ij} per subject. However, while analysis t_{ij} is transformed onto $[0, 1]$.

For covariates, we take $x_{ij1} = t_{ij} + \delta_{ij}$, where δ_{ij} follows the standard normal distribution and x_{ij2} is generated from a Bernoulli(0.5) distribution. The nonparametric functions are taken as $f_1(t) = \cos(\pi t)$ and $f_2(t) = \sin(\pi t)$. The error $(e_{i1}, \ldots, e_{im_i})$ is generated from a multivariate normal distribution with mean 0 and covariance $\Sigma_i = D_i R_i D_i$, where $R_i = T_i T_i'$ with $\mathbf{w}_{ijk} = (1, t_{ij} - t_{ik})'$, and $\mathbf{z}_{ij} = \mathbf{x}_{ij}$. The expected sample size (for the calculation of the number of knots) is about 1040 (=$100 \times 13 \times 0.8$). The number of the knots is taken to be $4 \approx 1040^{1/5}$ (He *et al.*, [4]). Here, as can be seen, the number of knots is not prespecified. Parameters of the above models were estimated using the R package jmcm for the simulation studies.

Bias of the estimates of the parameters of all three models along with their standard errors, and MSE (S) of the non-parametric functions f_1 and f_2, based on 1000 replications, are given in Table 1.

Par.	Tr. val.	Zhang *et al.* (2015) (Model 1)		Semi. with mean (Model 2)		Semi. with mean and var (Model 3)		Semi. with mean and var when knot=10		Semi. with mean and var when knot=20	
		Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
β_1	1.0	0.00	0.02	-0.02	0.22	0.09	0.27	0.00	0.24	-0.01	0.25
β_2	0.5	0.00	0.05	0.01	0.11	-0.03	0.14	0.01	0.12	0.01	0.13
γ_1	0.35	-0.27	1.30	6.61	1.42	3.14	2.25	-21.51	2.42	-41.93	3.41
γ_2	0.5	-0.35	2.29	6.62	2.53	5.56	4.12	-3.62	4.08	-5.74	5.04
λ_1	-0.5	0.00	0.07	0.00	0.00	0.00	0.08	0.00	0.07	0.00	0.07
λ_2	0.2	0.00	0.13	-0.01	0.01	0.01	0.15	0.01	0.14	0.00	0.14
S($\hat{f}_1$)					11.89		10.90		496.59		516.37
S($\hat{f}_2$)							112.87		373.95		642.26

Table 1: Bias and standard error of the estimated parameters based on 1000 replications; all the results are multiplied by a factor of 10^3

Table 1 shows that our semiparametric methods yield similar bias property of the estimates as compared to that for the parametric model. We redo the simulations by fixing the number of knots as $k_n = 10$ and $k_n = 20$. The results show that as the number of knots increase, the MSE of the estimated functions f_1 and f_2, bias, and standard error of the estimates of the parameters also increase.

3.2. Study 2: Properties of the regression parameters when some covariates are correlated

We have done a simulation study where all the covariates are correlated. The data sets are generated from the model

$$y_{ij} = x_{ij0}\beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_1(t_{ij}) + e_{ij} \quad (i = 1, \ldots, n; j = 1, \ldots, m_i),$$
$$\phi_{ijk} = \gamma_0 + w_{ijk1}\gamma_1 + w_{ijk2}\gamma_2, \quad \text{and}$$
$$\log(\sigma_{ij}^2) = z_{ij0}\lambda_0 + z_{ij1}\lambda_1 + z_{ij2}\lambda_2 + f_2(t_{ij}).$$

where $(x_{ij0}, x_{ij1}, x_{ij2})'$ is generated from a multivariate normal distribution with mean 0, marginal variance 1, and correlation 0.5. We take $\boldsymbol{z}_{ij} = \boldsymbol{x}_{ij}$ and $\mathbf{w}_{ijk} = (1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2)'$. t_{ij}, e_{ij}, $f_1(t)$, and $f_2(t)$ are same as study 3.1.

Parameter	True value	Parametric		Semiparametric	
		Bias	SE	Bias	SE
β_0	1.0	0.00	0.02	0.05	0.12
β_1	-0.5	0.00	0.02	-0.05	0.11
β_2	0.5	0.00	0.01	0.03	0.02
γ_0	0.35	-0.04	1.43	3.42	2.34
γ_1	0.5	-0.08	3.41	5.71	4.81
γ_2	-0.3	0.05	1.93	-3.74	2.82
λ_0	-0.5	0.00	0.01	0.00	0.05
λ_1	0.5	0.00	0.02	0.00	0.04
λ_2	-0.3	0.00	0.01	0.00	0.05
MSE($\hat{f}_1$)					9.62
MSE($\hat{f}_2$)					98.84

Table 2: Bias and standard error of the estimated parameters based on 1000 replications when the covariates are correlated; all the results are multiplied by a factor of 10^3

Table 2 shows that both parametric and semiparametric methods yield similar bias and standard error property of the estimates.

3.3. Study 3: A robustness study. Properties of the regression parameters and the functions f_1 and f_2 when error follows mixture of normal distributions

Again, another study, similar to what was done in study 3.1, has been conducted, in which, for generating the response data of Model 2 and Model 3, we consider a mixture of two multivariate normal distributions with error distributions $N_{m_i}(0, \Sigma_i)$ and $N_{m_i}(0, 0.25^2\Sigma_i)$ with equal probability. The results for the simulation study are displayed in Table 3 which show that the bias and the standard errors of the estimates, and the MSE of f_1 remain almost the same as those in study 3.1. However, MSE of the fitted function f_2 in Model 3 increases

significantly. Note that the function f_2 is associated with the variance and the function f_1 is associated with the mean.

Note that the mixed modeling affects only the MSE of the estimates of f_2 in the semi-parametric model 3, but not f_1. Although we are not aware of a proof as to why mixed modelling affects only the MSE of the estimates of f_2, but not f_1, experience in this context and other similar contexts show that parameters or functions associated with the variance parameter are more difficult to estimate and show more bias and MSE than parameters or functions associated with the mean parameter. See, for example, Zhang *et al.* ([1], pp 234) and Saha and Paul ([10], pp 183). In estimating the mean parameter m and the dispersion parameter c, Saha and Paul [10] find that estimates of the mean parameter m or regression parameters show much smaller bias and standard error than that of the dispersion parameter c.

Par.	True value	Semi. (Model 2) (without mixed)		Semi. (Model 2) (with mixed)		Semi. (Model 3) (without mixed)		Semi. (Model 3) (with mixed)	
		Bias	SE	Bias	SE	Bias	SE	Bias	SE
β_1	1.0	-0.02	0.22	-0.02	0.21	0.09	0.27	0.11	0.22
β_2	0.5	0.01	0.11	0.01	0.13	-0.03	0.14	-0.04	0.11
γ_1	0.35	6.61	1.42	6.61	6.32	3.14	2.25	8.84	2.36
γ_2	0.5	6.62	2.53	-8.57	8.17	5.56	4.12	13.48	4.39
λ_1	-0.5	0.00	0.00	0.00	0.1	0.00	0.08	0.00	0.09
λ_2	0.2	-0.01	0.01	0.00	0.1	0.01	0.15	-0.02	0.18
MSE($\hat{f}_1$)			11.89		11.89		10.90		11.00
MSE($\hat{f}_2$)							112.87		693.49

Table 3: Simulation results for Study 3.3 in Model 2 and Model 3 over 1000 replications when error terms follow mixture of normal distribution; $n = 100$; all the results are multiplied by a factor of 10^3

3.4. Study 4: Comparison of using B-spline and penalized spline in Model 2

A further study is conducted to compare the B-spline and the penalized spline to estimate the nonparametric function f_1 in Model 2. The estimation procedure to estimate all parameters are discussed in Section 2.3. We have used generalized cross-validation approach to find the penalty parameter for penalized splines. To generate response variable we consider the same mean and variance models as those in study 3.1 and the results are presented in Table 4. Results in Table 4 show no advantage of using the penalized spline over the B-spline.

Parameter	True value	Semiparametric Model 2			
		B spline		Penalized spline	
		Bias	SE	Bias	SE
β_1	1.0	-0.02	0.22	-0.01	0.22
β_2	0.5	0.01	0.11	0.01	0.11
γ_1	0.35	6.61	1.42	6.60	1.38
γ_2	0.5	6.62	2.53	6.50	2.51
λ_1	-0.5	0.00	0.00	0.00	0.08
λ_2	0.2	-0.01	0.01	-0.02	0.15
MSE($\hat{f}_1$)			11.89		11.89

Table 4: Bias and standard error of the estimated parameters in Model 2 based on 1000 replications using B-spline and penalized spline; all the results are multiplied by a factor of 10^3

3.5. Study 5: Comparison of the estimators by fitting the data generated from the semiparametric model to the parametric model and vice versa

We have done a simulation study by fitting the data generated from the semiparametric model to the parametric model and vice versa. We generate the response variable from Model 1 similar to study 3.2 and estimate the regression parameters by fitting Model 3 and vice versa. The results for the simulation study are displayed in Table 5 which show that bias and standard errors of the estimates are reasonable by fitting the data generated from the parametric model and estimating the regression parameters by the semiparametric model. In contrast, bias and standard errors increase significantly if we generate response variable using the semiparametric model and estimate the regression parameters by the parametric model.

Parameter	True value	Parametric to Semiparametric		Semiparametric to Parametric	
		Bias	SE	Bias	SE
β_0	1.0	0.00	0.06	-46.48	7.63
β_1	-0.5	0.00	0.06	46.87	7.64
β_2	0.5	0.00	0.04	-27.46	7.84
γ_0	0.35	-1.74	2.53	473.47	15.79
γ_1	0.5	-2.83	5.47	-40.41	13.92
γ_2	-0.3	1.64	3.13	702.96	133.61
λ_0	-0.5	0.00	0.05	95.04	16.74
λ_1	0.5	0.00	0.05	-92.93	16.71
λ_2	-0.3	0.00	0.05	57.25	16.72

Table 5: Bias and standard error of the estimated parameters based on 1000 replications by fitting the data generated from the semiparametric model to the parametric model and vice versa; all the results are multiplied by a factor of 10^3

A note on the simulation results is that the choice of the values of the parameters, such as, β_1 and β_2 in Table 1, does not affect the results on the properties of the estimates. For example, the simulations for the results in Table 1 for Model 3 was re-run with $\beta_1 = 0.5$ and $\beta_2 = 2.5$ instead of $\beta_1 = 1$ and

$\beta_2 = 0.5$ as in Table 1. The bias and SE of the estimates of all the parameters were virtually unchanged.

4. ANALYSIS OF TWO REAL DATA SETS ARISING IN BIOMEDICAL/ENVIRONMENTAL STUDIES

The purpose of this section is to study the impact of semiparametric modelling on model parsimony through the analysis of two real life longitudinal data sets arising in biomedical/environmental studies. The first data set is regarding progesterone metabolite (pregnanediol-3-glucuronide, PdG) measures that was obtained by the Institute for Toxicology and Environmental Health at the University of California, Davis in collaboration with the Reproductive Epidemiology Section of the California Department of Health Services, Berkeley (Brumback and Rice, [11]). The second data set is regarding CD4 cell counts of 369 HIV-infected men obtained by Kaslow *et al.* [12]. The first data set involves polynomials in time and the second data set involves polynomials in time as well as real covariates. Full description and analysis of the these two data sets are given in Sections 4.1 and 4.2.

In order to select the most parsimonious model we need to identify the best integer triple representing, namely, the degrees of the three polynomial functions for the mean structure, the correlation structure and the variance structure. Similar to Pourahmadi [13] and Pan and Mackangee [14], we use the Bayesian information criterion, BIC, to identify the best triple as follows

$$\text{BIC}(p, q, d) = -2\hat{l}_{\max}/n + (p + q + d + 3)\log(n)/n,$$

where $\hat{l}_{\max}$ is the maximum of the log-likelihood, p, q and d lie in the range of 0 to $(m-1)$ with $m = \max_{1\le i\le n}\{m_i\}$.

Thus, (p', q', d') is the optimal triplet which minimizes $\text{BIC}(p, q, d)$. Shao [15] and Shi and Tsai [16] demonstrated that BIC-criterion can identify the true model consistently. Although, in the literature BIC is preferred over the AIC (Akakie Information Criterion)

$$\text{AIC}(p, q, d) = -2\hat{l}_{\max}/n + 2(p + q + d + 3)/n,$$

we include the later here for comparison in our model selection.

4.1. Progesterone metabolite data

The Brumback and Rice [11] data consist of repeated progesterone metabolite (pregnanediol-3-glucuronide, PdG) measures from day -8 to day 15 in the

menstrual cycle (day 0 denotes ovulation day) on a sample of 22 conceptive cycles from 22 women and 29 non-conceptive cycles from another 29 women to study of early pregnancy loss. Altogether 1130 observations were obtained from 51 women, with each woman contributing 9 to 24 observations over time. The data are reproduced in Table 1 of the Supplementary Material.

As in Brumback and Rice [11], we take a log transformation of these data to make the normality assumption reasonable. Using our semiparametric modelling, the full model is

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{ij}\beta_1 + \cdots + x_{ij}^p\beta_p + f_1(t_{ij}), \\ \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1 + \cdots + w_{ijk}^q\gamma_q, \end{aligned}$$

and

$$\log(\sigma_{ij}^2) = \lambda_0 + z_{ij}\lambda_1 + \cdots + z_{ij}^d\lambda_d + f_2(t_{ij}),$$

where $x_{ij} = z_{ij} = t_{ij}$, $w_{ijk} = t_{ij} - t_{ik}$ and p, q, and d lie in the range of 1 to 23. The parametric model, of course, will not have the non-parametric functions.

We now analyze the data using these parametric and semiparametric models. The results are given in Table 6. Note that here we give 9 most parsimonious models to capture the model having smallest BIC and/or smallest AIC. Results in Table 6 show that the most parsimonious model having the smallest BIC ($BIC = 0.47$) and the smallest AIC ($AIC = -0.29$) obtained by parametric modelling has $(p, q, d) = (7, 3, 7)$. Where as the corresponding most parsimonious model by semiparametric modelling has $(p, q, d) = (2, 2, 1)$ with BIC = -0.61 and AC=-0.92. Thus the final models for these data are

$$\begin{aligned} g(\mu_{ij}) &= f_1(t_{ij}) + \beta_0 + x_{ij}\beta_1 + x_{ij}^2\beta_2, \\ \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1 + w_{ijk}^2\gamma_2, \\ &\text{and} \\ \log(\sigma_{ij}^2) &= f_2(t_{ij}) + \lambda_0 + z_{ij}\lambda_1, \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $w_{ijk} = t_{ij} - t_{ik}$. Note that both the BIC and the AIC criteria choose the same model for these data.

4.2. Analysis of CD4 cell data

The data comprise CD4 cell counts of 369 HIV-infected men with six covariates including time since seroconversion (t_{ij}), age (relative to arbitrary origin, x_{ij1}), packs of cigarettes smoked per day (x_{ij2}), recreation drug use (x_{ij3}), number of sexual partners (x_{ij4}), and mental illness score (x_{ij5}). In total there are 2376 observations with multiple repeated measurements taken for each individual at different times, covering a period of approximately eight and a half years.

(p,q,d)	No. of par.	Parametric			Semiparametric		
		$\hat{l}_{\max}$	BIC	AIC	$\hat{l}_{\max}$	BIC	AIC
(2,2,1)	8	-71.66	3.43	3.12	31.34	-0.61	-0.92
(2,2,2)	9	-66.59	3.31	2.96	31.34	-0.54	-0.88
(2,1,1)	7	-129.58	5.62	5.36	16.01	-0.09	-0.35
(2,1,4)	10	-66.98	3.40	4.53	16.38	0.12	-0.25
(3,1,4)	11	-28.97	1.98	3.09	16.38	0.21	-0.21
(3,1,5)	12	-28.20	2.03	3.00	19.58	0.16	-0.30
(2,1,6)	12	-65.84	3.51	4.42	23.73	-0.005	-0.46
(3,2,3)	11	-32.85	2.14	1.81	31.34	-0.38	-0.80
(7,3,7)	20	27.31	0.47	-0.29	42.06	-0.11	-0.87

Table 6: Progesterone hormone data: A comparison of various models using parametric (Zhang *et al.*, [1]) and semiparametric approaches

The number of measurements for each individual varies from 1 to 12 taken at unequally spaced time points.

Now, for these data m_i varies from 1 to 12, so p, q, and d all lie in the range of 0 to 11. The values of p, q, and d for the most saturated model are 11. So, we have a total $11 \times 11 \times 11 = 1331$ models of which we present 6 most parsimonious models.

4.2.1. Analysis using polynomials in time

This data set has been analyzed by others in the past (for example, Zeger and Diggle, [17] and Ye and Pan, [18]) using polynomials in time. Most recently Zhang *et al.* [1], in order to jointly model the mean, correlation and variance structures, fit polynomial regressions of time

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{ij}\beta_1 + \cdots + x_{ij}^p\beta_p, \\ \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1 + \cdots + w_{ijk}^q\gamma_q, \text{and} \\ \log(\sigma_{ij}^2) &= \lambda_0 + z_{ij}\lambda_1 + \cdots + z_{ij}^d\lambda_d, \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $\mathbf{w}_{ijk} = t_{ij} - t_{ik}$.

We now analyze these data using the above parametric models and our method of semiparametric modelling Model 3 developed in Section 2.1. The results, in terms of $\hat{l}_{max}$, BIC and AIC of the parametric and semiparametric models are given in Table 6. Note that here we provide all relevant information of 6 most parsimonious models. Of these, the most parsimonious model obtained from parametric modelling has $(p,q,d) = (8,1,1)$ with $\hat{l}_{max} = -4892.72$, BIC =26.72, and AIC =26.58 (smallest BIC as well as smallest AIC among all these 6 models). This finding is in agreement with what was found by Zhang *et al.* [1]. The semiparametric modelling, however, shows $(p,q,d) = (4,1,1)$ with $\hat{l}_{max} = -4877.41$, BIC =26.58, and AIC =26.48 (smallest BIC as well as smallest AIC among all these 6 models). Including the intercept parameters the semiparametric

model has 9 parameters as opposed to 13 parameters of the model obtained by the parametric model. Thus, the semiparametric modelling shows a significant improvement in model parsimony as compared to the parametric modelling and the most parsimonious model for these data is

$$\begin{aligned} g(\mu_{ij}) &= f(t_{ij}) + \beta_0 + x_{ij}\beta_1 + x_{ij}^2\beta_2 + x_{ij}^3\beta_3 + x_{ij}^4\beta_4, \\ \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1, \\ &\text{and} \\ \log(\sigma_{ij}^2) &= g(t_{ij}) + \lambda_0 + z_{ij}\lambda_1, \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $\mathbf{w}_{ijk} = t_{ij} - t_{ik}$.

Note that the same parsimonious model is obtained by both the BIC and the AIC procedures. To investigate the stability of AIC and BIC, a simulation study is conducted by choosing p, q, and d from $1, 2, \ldots, 11$ and obtain that the stability of AIC and BIC are very similar which is 62%. From the same simulation study, we further investigate the parametric and semi-parametric models performances in terms of AIC and BIC and found that parametric models perform better than semiparametric models.

(p,q,d)	No. of par.	Parametric			Semiparametric		
		$\hat{l}_{\max}$	BIC	AIC	$\hat{l}_{\max}$	BIC	AIC
(4,1,1)	9	-4910.87	26.76	26.67	-4877.41	26.58	26.48
(3,1,1)	8	-4926.88	26.83	26.74	-4882.11	26.59	26.50
(8,1,1)	13	-4892.72	26.72	26.58	-4874.40	26.63	26.49
(8,3,1)	15	-4890.44	26.75	26.59	-4881.66	26.69	26.64
(3,3,3)	12	-4919.52	26.85	26.73	-4879.37	26.64	26.51
(8,3,3)	17	-4886.36	26.76	26.58	-4872.74	26.68	26.50

Table 7: CD4 cell data: A comparison of various models using parametric (Zhang *et al.*, [1]) and semiparametric approaches with polynomials in time

4.2.2. Analysis using polynomials in time and the covariates

We now analyze the CD4 data using the Model 3 regression models after including 5 covariates in the mean model. The results, in terms of $\hat{l}_{max}$, BIC, and AIC of the parametric and the semiparametric models, are given in Table 8. Here also we provide all relevant information of 6 best (parsimonious) models. Of these, the most parsimonious model by the parametric modelling has $(c, p, q, d) = (3, 8, 1, 1)$ with 16 parameters by both the BIC and the AIC procedures, where c represents the number of covariates in the model. However, the most parsimonious model by the semiparametric modelling has $(c, p, q, d) = (3, 2, 1, 1)$ with 10 parameters using the BIC procedure (BIC=26.47)and has $(c, p, q, d) = (3, 4, 1, 1)$ with 12 parameters using the AIC procedure (AIC=26.38).

(c,p,q,d)	No. of par.	Parametric			Semiparametric		
		$\hat{l}_{\max}$	BIC	AIC	$\hat{l}_{\max}$	BIC	AIC
(3,2,1,1)	10	-4944.54	26.91	26.84	-4863.43	26.47	26.40
(5,2,1,1)	12	-4938.40	26.88	26.80	-4862.52	26.47	26.39
(3,4,1,1)	12	-4891.12	26.65	26.56	-4858.68	26.48	26.38
(5,4,1,1)	14	-4889.44	26.65	26.55	-4857.85	26.47	26.38
(5,3,1,1)	13	-4905.57	26.72	26.63	-4862.52	26.48	26.40
(3,8,1,1)	16	-4873.51	26.62	26.49	-4857.16	26.53	26.40

Table 8: CD4 cell data: A comparison of various models using parametric (Zhang *et al.*, [1]) and semiparametric approaches with polynomials in time and covariates

Thus, the most parsimonious models for the CD4 data involving time and the covariates is

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{ij}\beta_1 + x_{ij}^2\beta_2 + x_{ij2}\beta_3 + x_{ij3}\beta_4 + x_{ij5}\beta_5 + f(t_{ij}), \\ \phi_{ijk} &= \gamma_0 + w_{ij}\gamma_1, \\ &\text{and} \\ \log(\sigma_{ij}^2) &= \lambda_0 + z_{ij}\lambda_1 + g(t_{ij}), \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $\mathbf{w}_{ijk} = t_{ij} - t_{ik}$.

Our finding is that both the BIC and the AIC select same or similar models, although, BIC has some advantage over the AIC in the sense that, in some cases, the model selected by the BIC is slightly more parsimonious than the model selected by the AIC. Further, the semiparametric modelling obtains a model that is much more parsimonious than the model obtained by the parametric modelling approach.

5. DISCUSSION

We develop joint estimation procedure for the mean (regression) and the variance parameters in longitudinal data using semiparametric modelling of the mean and the variance, regression spline, and by decomposing the correlation matrix via hyperspherical co-ordinates. Through an extensive simulation study we compare our method with the parametric method by Zhang *et al.* [1]. Further, the effect of the misspecification of the error distribution and of the number of knots used in the estimation of the nonparametric functions, and whether the penalized spline procedure improves the estimation of the nonparametric functions over the B-spline are investigated. Furthermore two real data sets arising from biomedical/environmental study are analyzed.

The main findings of the simulation study are: (a) the parametric modelling and the semiparametric modelling produce similar bias and efficiency property of the regression parameters, (b) good choice of number of knots can have significant impact of estimation and inference on non-linear trends (e.g., the nonparametric

functions) using regression B-splines, and (c) use of the penalized spline does not improve the efficiency of the estimates of the nonparametric functions.

Through data analysis: (i) our findings regarding the model selection procedures are that both the BIC and the AIC select same or similar models, although, BIC has some advantage over the AIC in the sense that, in some cases, the model selected by the BIC is slightly more parsimonious than the model selected by the AIC, (ii) the main advantage of the semiparametric modelling over the parametric modelling is that the former produces a much more parsimonious model than the latter.

We also analyzed some other real bio-medical data sets and obtained similar conclusions, the details of which are not given here.

ACKNOWLEDGMENTS

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada. We also acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Zhang, W.; Leng, C. and Tang, Y.C. (2015). A joint modelling approach for longitudinal studies, *Journal of the Royal Statistical Society* B, **77**, 219–238.

[2] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

[3] Diggle P.J.; Heagerty, P.; Liang, K.Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data*, 2nd edition, Oxford University Press.

[4] He X.M.; Fung W.K. and Zhu, Z.Y. (2005). Robust estimation in generalized partial linear models for clustered data, *Journal of the American Statistical Association*, **100**, 1176–1184.

[5] Lin, X. and Carroll, R.J. (2006). Semiparametric estimation in general repeated measures problems, *Journal of the Royal Statistical Society* B, **68**, 69–88.

[6] Leng, C.; Zhang, W. and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data, *Journal of the American Statistical Association*, **105**, 181–193.

[7] Schumaker, L.L. (1981). *Spline Functions*, Wiley, New York, **182, 190**.

[8] Carroll, R.H; Macam, J.D. and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error, *Biometrika*, **86**, 541–554.

[9] Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with b-splines and penalties, *Statist. Sci.*, **11**, 89–121.

[10] Saha, K.K. and Paul, S.R. (2005). Bias-corrected maximum likelihood estimator of the intraclass correlation parameter for binary data, *Statistics In Medicine*, **24**, 3497–3512.

[11] Brumback, B.A. and Rice, J.A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion), *Journal of the American Statistical Association*, **93**, 961–994.

[12] Kaslow, R.A; Ostrow, D.G.; Detels, R. *et al.* (1987). The multicenter AIDS cohort study: Rationale, organization and selected characteristics of the participants, *American Journal of Epidemiology*, **126**, 310–318.

[13] Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix, *Biometrika*, **87**, 425–435.

[14] Pan, J. and Mackenzie, G. (2003). Model selection for joint mean-covariance structures in longitudinal studies, *Biometrika*, **90**, 239–244.

[15] Shao, J. (1997). An asymptotic theory for linear model selection, *Statistica Sincia*, 221–264.

[16] Shai, P. and Tsai, C.L. (2002). Regression model selection-a residual likelihood approach, *Journal of the Royal Statistical Society* B, **64**, 237–252.

[17] 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.

[18] Ye, H. and Pan, J. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data, *Biometrika*, **93**, 927–941.

Appendix A: Solution of the Estimating Equations of Model 3

We apply the quasi-Fisher scoring algorithm to solve estimating equations of Model 3 where the parameters $\boldsymbol{\theta}, \boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ solve sequentially one by one with other parameter keep fixed in optimization:

Step 1 : Choose initial values of the parameters as $\boldsymbol{\theta}^{(0)}, \boldsymbol{\rho}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$. Set $k = 0$

Step 2 : Calculate $\boldsymbol{\Sigma}_i$ by using $\boldsymbol{\rho}^{(k)}$ and $\boldsymbol{\gamma}^{(k)}$. Update $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + I_{11}^{-1}\boldsymbol{U}_1|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}$$

Step 3 : Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k+1)}$, update $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$ by using

$$\begin{pmatrix} \boldsymbol{\gamma}^{(k+1)} \\ \boldsymbol{\rho}^{(k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}^{(k)} \\ \boldsymbol{\rho}^{(k} \end{pmatrix} + \left[\begin{pmatrix} I_{22} & I_{23} \\ I_{32} & I_{33} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{U}_3 \\ \boldsymbol{U}_2 \end{pmatrix} \right] |_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(k)}, \boldsymbol{\rho}=\boldsymbol{\rho}^{(k)}}$$

Step 4 : Set $k \leftarrow k+1$ and repeat steps 2 and 3 until a desired convergence criteria is satisfied.

Note that block components of Fisher information matrix I are:

$$
\begin{aligned}
I_{11} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \sum_{i=1}^{n}\boldsymbol{\Pi}_i'\boldsymbol{\Delta}_i\Sigma_i^{-1}\Delta_i\boldsymbol{\Pi}_i,\\
I_{12} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}'}\right] = -\sum_{i=1}^{n}\left[\boldsymbol{\Pi}_i'\Delta_i\frac{\partial\Sigma_i^{-1}}{\partial\boldsymbol{\gamma}'}(E(\mathbf{y}_i)-\boldsymbol{\mu}_i)\right] = 0,\\
I_{13} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\rho}'}\right] = 0,\\
I_{22} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[2\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}}\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}'}+\sum_{k=1}^{j-1}b_{ijk}b_{ijk}'\right],\\
I_{23} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\rho}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}}\Upsilon_{ij}'+\frac{1}{2}\sum_{k=1}^{j-1}b_{ijk}\sum_{l=k}^{j}a_{ijl}T_{ilk}\Upsilon_{il}'\right],\\
I_{33} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'}\right] = \frac{1}{4}\sum_{i=1}^{n}\boldsymbol{\Upsilon}_i'\left[I_{m_i}+R_i^{-1}\circ R_i\right]\boldsymbol{\Upsilon}_i,
\end{aligned}
$$

where '$\circ$' represents the Hadamard product.

Appendix B: Block Components of Fisher Information Matrix of the Estimating Equations of Model 2

$$
\begin{aligned}
I_{11} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \sum_{i=1}^{n}\boldsymbol{\Pi}_i'\boldsymbol{\Delta}_i\Sigma_i^{-1}\Delta_i\boldsymbol{\Pi}_i,\\
I_{12} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}'}\right] = -\sum_{i=1}^{n}\left[\boldsymbol{\Pi}_i'\Delta_i\frac{\partial\Sigma_i^{-1}}{\partial\boldsymbol{\gamma}'}(E(\mathbf{y}_i)-\boldsymbol{\mu}_i)\right] = 0,\\
I_{13} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}'}\right] = 0,\\
I_{22} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[2\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}}\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}'}+\sum_{k=1}^{j-1}b_{ijk}b_{ijk}'\right],\\
I_{23} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\lambda}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{\partial\log T_{ijj}}{\partial\boldsymbol{\gamma}}Z_{ij}'+\frac{1}{2}\sum_{k=1}^{j-1}b_{ijk}\sum_{l=k}^{j}a_{ijl}T_{ilk}Z_{il}'\right],\\
I_{33} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right] = \frac{1}{4}\sum_{i=1}^{n}\mathbf{Z}_i'\left[I_{m_i}+R_i^{-1}\circ R_i\right]\mathbf{Z}_i,
\end{aligned}
$$

where '$\circ$' represents the Hadamard product.