RASHI 2 (1) : 58-66 (2017)

Challenges in the Covariance Estimation of Longitudinal data: a review

Kiranmoy Das

Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata, India.

ABSTRACT

Longitudinal data, also popularly known as repeated measure data, are challenging from modelling perspective due to the fact that the measurements from the same individual at different time points are inherently correlated. Despite the difficulty, there has been a substantial amount of work done in last twenty years on the covariance estimation of such data. In this article, we will review some standard parametric methods for covariance estimation and then will discuss on some recent developments on it. Powerful semiparametric methods have been proposed in the literature in the last ten years for the joint estimation of the mean trajectories and the covariance matrix. Such methods can guarantee the positive definiteness of the estimated covariance matrix and thus provide reliable parameter estimates. We will also discuss Bayesian semiparametric and nonparametric methods for the covariance estimation of longitudinal data with sparsity. When the data come from the subjects belonging to different related groups, it is desirable that the covariance matrices for the groups will share some parameters and this can be achieved by considering a suitable non-parametric prior distribution on the covariance parameters. We also discuss the joint covariance estimation for bivariate and multivariate longitudinal outcomes. We comment on the adjustment needed for the covariance estimation of the longitudinal data with missing values.

Keywords : Covariance matrix, Longitudinal data, Multivariate data, Positive definiteness, Random effects, Semiparametric approach.

1. Introduction

Repeated measure data or longitudinal data are obtained from the longitudinal studies where the subjects are measured at multiple time points, not necessarily evenly spaced. Such data are also known as cross-sectional time series data. Typically, in statistics, we come across the cross-sectional data where we have *n* subjects and measurements are obtained on one or more variables of interest at one particular time point. On the other hand, time series data are obtained for one or more variables from one subject at multiple time points (*T*). In the longitudinal data, we have *n* subjects and measurements are taken on one or more variables at *T* different time points. Thus, longitudinal data are cross-sectional at each fixed time point and are time series for a fixed subject. In economics and other related social sciences, longitudinal data are also known as panel data.

Longitudinal data occur in a variety of disciplines including but not limited to agriculture, biology, medical science, social science, engineering, and public health. For example, HIV patients are monitored by measuring the CD4 counts at different time points. In health economics, the out of pocket medical expenditures (medical expenditure not reimbursed or paid trough the health insurances) are measured from the aged individuals for different years. A geneticist might be interested in knowing the functional behaviour of certain genes at the different stages of human life and hence measures certain biomarker (body weight, for example) for different subjects at different ages. In a regular longitudinal study all the subjects are measured at the same time points, while in the irregular case, different subjects are measured at different time points and this introduces sparsity in the data.

Statistical analysis of longitudinal data is challenging due to the fact that the measurements obtained from the same subject at different time points are inherently correlated, even when the subjects are independent to each other. Let $Yi_l, \dots, Yi_T$ be the measurements for the *i*-th subject in a longitudinal study, for i = 1, 2, . . . , *n*. Traditionally, one should assume that the vectors $Y_i = (Y_{i1}, \dots, Y_{iT})^T$ are identically and independently distributed with *T*-variate normal distribution with mean vector = μ_i and the covariance matrix = Σ. Modeling of μ_i is essentially a regression problem but the additional complexity in the longitudinal study is to model the unknown covariance matrix Σ. Also, note that the symmetric matrix Σ will have $\frac{T(T+1)}{2}$ number of unknown parameters, and it is not uncommon to have $n < \frac{T(T+1)}{2}$. This introduces the "smaller sample larger parameter" issue in the high-dimensional statistics literature.

For analysing the longitudinal data, Laird and Ware (1982) introduced random effects model and developed a likelihood based estimation approach. This approach is handy but cannot explicitly explain the underlying covariance structure of the variable(s) of interest. Pourahmadi (1999, 2000) provided a model based flexible approach of estimating the covariance matrix

E-mail : kmd@isical.ac.in.

for the univariate longitudinal data. This approach has been used extensively in the literature since it provides a positive definite estimated covariance matrix. Wu and Pourahmadi (2003) proposed a non-parametric approach of estimating the large covariance matrices. In a Bayesian framework, Daniels and Pourahmadi (2002) developed prior distributions which can shrink the underlying unknown covariance matrices to some known structures. Pan and Mackenzie (2003) generalized Pourahmadi's approach to the irregular univariate longitudinal data. In the gene mapping problem, Das *et al*. (2013a) used Pan and Mackenzie's (2003) approach for modelling the longitudinal biomarkers. There is a vast literature on estimating the covariance matrix for the univariate longitudinal data.

For the bivariate or the multivariate settings, although a significant amount of work has been done, the literature is still not vast. The challenge in these settings is to handle the within response and the between response dependence. Sy, Taylor and Cumberland (1997) proposed a treatment for this issue based on the parametric stochastic model for CD4 T-cells and beta-2 microglobulin in AIDS data. This approach can also handle the irregular longitudinal data. Linear mixed models for analysing the bivariate longitudinal data were proposed by Theibaut *et al*. (2002). They also provided SAS package for such modelling. For regular bivariate longitudinal data, a longitudinal model for detecting prescribing change in two drugs with correlated errors was proposed by Sithole and Jones (2007) using a bivariate autoregressive process. This approach was generalized by Das *et al*. (2011) for the irregular sparse longitudinal data. Das *et al*. (2013b) also generalized Pourahmadi's (1999) approach for the bivariate irregular longitudinal data, however, the estimated covariance matrix is not guaranteed to be positive definite. Bandyopadhyay *et al*. (2010), Ghosh and Hanson (2010) used random effects for capturing the longitudinal dependence in the multivariate longitudinal outcomes.

In many applications, we obtain longitudinal data coming from several related groups and it is not uncommon that the covariance features are same and/or similar across the groups under consideration. It is practically impossible to determine such similarity manually. But in a Bayesian framework, suitable prior distributions can be taken for determining such similarity. Gaskins and Daniels (2012) used the matrix stick-breaking process (MSBP), originally proposed by Dunson *et al*. (2008) for non-parametrically modelling the grouped longitudinal matrices for the regular univariate outcomes. Das and Daniels (2014) extended this for the irregular bivariate longitudinal data. In particular, for each response feature, the generalized autoregressive parameters and innovation variances are first expressed as polynomial functions of time, and for the coefficients of these polynomial functions they considered MSBP priors to allow information sharing across the parameters of different groups and to introduce sparsity. Similar approaches have been used in Das *et al.* (2015), Chatterjee *et al*. (2016) for the applications in public health and engineering respectively.

Covariance estimation becomes really challenging in the presence of missingness. In the longitudinal studies, missingness can happen for many reasons, including the death or withdrawal of the subjects from the study. Missingness can be ignorable or non-ignorable (Rubin, 1976). For an efficient covariance estimation, the missing values are to be imputed. Daniels and Hogan (2008) provide different imputation techniques under the ignorable and non-ignorable missingness. Covariance estimation can also be severely affected by the zero-inflation in the outcomes.

The rest of the paper is organized as the following. In Section 2, we discuss the commonly used methods for the covariance estimation in the univariate longitudinal responses. In this section, we also discuss the methods proposed by Pourahmadi (1999, 2000) and the extension of it for the irregular longitudinal case by Pan and Mackenzie (2003). Section 3 provides the generalization and/or extension of the methods discussed in Sections 2 for the bivariate longitudinal case. Bayesian semiparametric covariance estimation for the grouped univariate and bivariate longitudinal data are discussed in Section 4. Finally, Section 5 concludes.

2. Covariance Estimation for Univariate Longitudinal Data

2.1 Traditional parametric approach

For $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})^T$ to be iid $N_T(\mu_i, \Sigma)$, often it is not possible to model and estimate Σ in an unstructured way. This is because the number of unknown parameters to be estimated becomes large even for moderate T. To avoid this issue, traditionally we assume a known structure for Σ and then estimate the underlying parameters. Most commonly used structure for Σ is the auto-regressive structure of order 1, which we will denote by AR(1). The AR(1) structure is given as the following:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \ldots\ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots\ldots & \rho^{T-2} \\ . & . & . & \ldots\ldots & . \\ . & . & . & \ldots\ldots & . \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots\ldots & 1 \end{bmatrix}$$

Note that in AR(1) structure, there are only two parameters, the variance parameters σ^2 and the correlation parameter ρ. Thus the estimation becomes simple and the estimated covariance matrix becomes positive definite. This structure relies on the assumption that the correlations between the observations corresponding to the closer time points are stronger and the strength of the correlation reduces as the time points become far away. Although this assumption sounds reasonable for the longitudinal data, but in the presence of seasonal variation, this assumption is violated.

For large T, sometimes it is meaningful to assume that the longitudinal correlation disappears after some threshold. For example, for some prefixed length w, one can assume correlation $(Y_{it}, Y_{it'}) = \rho^{|t-t'|}$, if $|t-t'| < w$; and 0, otherwise. For w = 3, one will get the following covariance structure:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & 0 & \ldots\ldots & 0 \\ \rho & 1 & \rho & 0 & \ldots\ldots & 0 \\ 0 & \rho & 1 & \rho & \ldots\ldots & 0 \\ . & . & . & \ldots\ldots & . & \\ . & . & . & \ldots\ldots & . & \\ 0 & 0 & 0 & \ldots\ldots & \rho & 1 \end{bmatrix}$$

The optimal band width w is typically not prefixed, but is obtained from the data by cross-validation or some other criterion.

Another popular choice of Σ is the compound symmetry (CS) structure. This structure is based on the assumption that the correlation between measurements corresponding to two different time points is always constant. This will result in the following covariance matrix with two parameters only and the estimated covariance matrix is always positive definite.

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \ldots\ldots & \rho \\ \rho & 1 & \rho & \ldots\ldots & \rho \\ . & . & . & \ldots\ldots & . \\ . & . & . & \ldots\ldots & . \\ \rho & \rho & \rho & \ldots\ldots & 1 \end{bmatrix}$$

In reality, it is not apparent which structure is supported by the data and hence we use the model fitting criteria. For any real data, one should fit both the AR(1) and the compound symmetry structure and compute AIC, BIC values. The best alternative will be chosen based on the smallest AIC, BIC values. Alternatively, one can model Σ as a convex combination of the two structures as the following:

$\Sigma = kAR(1) + (1 - k)\, CS$, for $0 < k < 1$. The optimal value of k is obtained from 10-fold cross-validation. Note that k determines the weights for AR(1) and CS structure. In social sciences, such structures are popularly used for estimating the covariance structure of panel data.

2.2 Pourahmadi's approach

Pourahmadi (1999) developed a regression based approach of estimating the covariance matrix for regular univariate longitudinal data. We consider a particular subject measured at T different time points. For a simple description, we ignore the subscript i for the moment. Without loss of generality, assume the response vector $y = (y_1, \ldots, y_T)$ has mean vector 0 and covariance matrix Σ. Pourahmadi modelled y_t, the response at time t, using its predecessors as the following:

$$y_t = \sum_{t'=1}^{t-1} \phi_{t,t'} y_{t'} + \epsilon_t, \qquad (1)$$

where $\phi_{t,t'}$ is the corresponding regression coefficient. Here $\in_t$ is the prediction error with mean = 0 and σ_t^2 be its variance. Assuming $\in_t$'s to be uncorrelated, we get $cov(\in) = E$, a diagonal matrix with σ_t^2 being the t-th diagonal element, where $\in = (\in_1, \ldots, \in_T)'$, the vector of prediction errors. In matrix representation we get,

$$\in = \mathrm{L}y, \qquad (2)$$

where L is a lower triangular matrix with 1's in diagonal elements and $-\phi_{t,t'}$ in the (t, t′)th off-diagonal position. From the above equation we get,

$$\mathrm{cov}(\in) = \mathrm{L}\,\mathrm{cov}(y)\mathrm{L}^T = \mathrm{L}\Sigma\mathrm{L}^T = \mathrm{E}, \qquad (3)$$

which is similar to the modified Cholesky decomposition of Σ.

Equation (3) essentially guarantees that the estimated covariance matrix will be positive definite. Pourahmadi (1999, 2000) modelled the unconstrained dependence parameters log σ_t^2 and $\phi_{t,t'}$ with a polynomial function of time as the following:

$$\log \sigma_t^2 = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \ldots. + \lambda_g t^g, \qquad (4)$$

$$\phi_{t,t'} = \delta_0 + \delta_1(t - t') + \delta_2(t - t')^2 + \ldots + \delta_h(t - t')^h, \quad (t' = 1, 2, \ldots t - 1). \qquad (5)$$

Note that the estimation of the covariance matrix Σ essentially becomes the estimation of the parameters λs and δs. The optimal order of the above polynomials g and h are obtained from the information criteria AIC, BIC. Pan and Mackenzie (2003) proposed a likelihood based method and estimated the covariance parameters using the Iteratively Reweighted Least Squares (IRLS) algorithm. Note that for the irregular longitudinal measurements, Pan and Mackenzie (2003) simply expressed the subject-specific covariance matrices Σ_i as the following:

$E_i = L_i \Sigma_i L_i^T$, and used the above approach for estimating the covariance matrices Σ_i. Das *et al*. (2013a) used a Bayesian approach and estimated the parameters λs and δs by Markov Chain Monte Carlo (MCMC) method.

3. Covariance Estimation for Bivariate Longitudinal Data

3.1 Parametric approach

In the parametric approach, Sithole and Jones (2007) modelled the covariance matrix of regular bivariate longitudinal data based on the idea similar to the univariate case. The covariance matrix for the bivariate longitudinal response is modelled as a Kronecker product of two known structures. Let Y_{itk} denote the *k*-th response feature measured from the *i*-th subject at time *t*. For $k = 2$, we denote the response vector for the *i*-th subject as the following : $Y_i = [Y_{i11}, Y_{i21}, ..., Y_{iT1}, Y_{i12}, Y_{i22}, ... , Y_{iT2}]$, and we assume that $Y_i s$ are identically and independently distributed as $2T$ variate normal distribution with covariance matrix = Σ. For fixed *k*(=1 and 2), we assume that the longitudinal measurements from the same subject have an AR(1)structure, while the correlation between the two response features remains the same over time. This kind of covariance structure can be expressed as $UN \otimes AR(1)$, where UN is a 2X2 symmetric matrix with σ_1^2 and σ_2^2 as the diagonal elements and σ_{12} as the off-diagonal element. For instance, if we have three repeated measures from a subject, then we assume a covariance matrix with the structure $UN \otimes AR(1)$ where UN and AR(1) are

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

respectively. This parametric structure assumes that intra-response longitudinal correlations are same for both the responses, and also the inter-response correlation is proportional to the intra-response correlation. However, such assumptions are not too restrictive and can be easily modelled using the standard statistical packages (SAS, for example). Das *et al*. (2011) extended this approach for the irregular sparse case while modelling the systolic and diastolic blood pressures in a genome-wide association study. Similarly, one can use $UN \otimes CS$ structure for the same purpose. Similar to the univariate case, in practice, our recommendation is to fit both $UN \otimes AR(1)$ and $UN \otimes CS$ and then choose the optimal one based on the information criteria.

3.2 Regression based approach

This approach was suggested by Das *et al*. (2013b) and it is a generalization of Pourahmadi (1999), and Pan and Mackenzie (2003). Suppose that a particular subject is measured at *T* different time points. For the time being, we suppress the subscript *i*. At time *t*, the residual for the *k*-th response feature ($k = 0,1$) can be modeled in terms of its predecessors, similar to the univariate case in the following way:

$$y_{tk} = \sum_{t'=1}^{t-1} \phi_{t,t'} y_{t'k} + \sum_{t'=1}^{t-1} \psi_{t,t'} y_{t'(1-k)} + \epsilon_{tk}, \quad (6)$$

where $\phi_{t,t'}$ and $\Psi_{t,t'}$ are the corresponding regression coefficients and $\in_{tk}$ is the prediction error with mean 0 and variance σ_{tk}^2. This model considers both the inter-trait and intra-trait correlations over time. Denote $y = \left[y_1^T, y_0^T\right]^T$ and $\in = \left[\in_1^T, \in_0^T\right]^T$,

where $\in_k = (\in_{1k}, ..., \in_{Tk})^{\mathrm{T}}$; $k = 1,0$. Then the matrix representation of the above model becomes

$$\mathbf{y} = \begin{bmatrix} \mathbf{\Phi} & \mathbf{\Psi} \\ \mathbf{\Psi} & \mathbf{\Phi} \end{bmatrix} \mathbf{y} + \boldsymbol{\epsilon}, \quad (7)$$

where Φ and Ψ are both $T \times T$ lower triangular matrices with 0s in the diagonal elements and in the (t, t')-th position, $\phi_{t,t'}$ and $\Psi_{t,t'}$ $(t > t')$ respectively. Thus, we can write,

$$\mathrm{L}y = \in, \quad (8)$$

where $\mathbf{L} = \begin{bmatrix} I - \mathbf{\Phi} & -\mathbf{\Psi} \\ -\mathbf{\Psi} & I - \mathbf{\Phi} \end{bmatrix}$, Following Pourahmadi (1999), we assume that $\in_{tk}$s are uncorrelated and hence we get $cov(\in) = \mathrm{E} = \begin{bmatrix} E_1 & 0 \\ 0 & E_0 \end{bmatrix}$,

where both E_1 and E_0 are $T \times T$ diagonal matrices with the *t*-th diagonal element σ_{t1}^2 and σ_{t0}^2, respectively. Hence, we have,

$$\mathrm{cov}(\in) = \mathrm{L}\mathrm{cov}(y)\mathrm{L}^{\mathrm{T}} = \mathrm{L}\Sigma\mathrm{L}^{\mathrm{T}} = \mathrm{E}. \quad (9)$$

Note that here $\Sigma = L^{-1}E(L^{-1})^T$. Thus we have,

$$\begin{aligned}
\mathbf{\Sigma^{-1}} &= \mathbf{L^T E^{-1} L} \\
&= \begin{bmatrix} (I-\mathbf{\Phi})^{\mathbf{T}} & -\mathbf{\Psi^T} \\ -\mathbf{\Psi^T} & (I-\mathbf{\Phi})^{\mathbf{T}} \end{bmatrix} \begin{bmatrix} \frac{1}{\mathbf{E^{(1)}}} & 0 \\ 0 & \frac{1}{\mathbf{E^{(0)}}} \end{bmatrix} \begin{bmatrix} (I-\mathbf{\Phi}) & -\mathbf{\Psi} \\ -\mathbf{\Psi} & (I-\mathbf{\Phi}) \end{bmatrix} \\
&= \begin{bmatrix} \frac{(I-\mathbf{\Phi})^{\mathbf{T}}(\mathbf{I-\Phi})}{\mathbf{E^{(1)}}} + \frac{\mathbf{\Psi^T\Psi}}{\mathbf{E^{(0)}}} & -\frac{(I-\mathbf{\Phi})^{\mathbf{T}}\mathbf{\Psi}}{\mathbf{E^{(1)}}} - \frac{\mathbf{\Psi^T(I-\Phi)}}{\mathbf{E^{(0)}}} \\ -\frac{\mathbf{\Psi^T(I-\Phi)}}{\mathbf{E^{(1)}}} - \frac{(I-\mathbf{\Phi})^{\mathbf{T}}\mathbf{\Psi}}{\mathbf{E^{(0)}}} & \frac{\mathbf{\Psi^T\Psi}}{\mathbf{E^{(1)}}} + \frac{(I-\mathbf{\Phi})^{\mathbf{T}}(\mathbf{I-\Phi})}{\mathbf{E^{(0)}}} \end{bmatrix} \\
&\hat{=} \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{12}^T} & \mathbf{\Sigma_{22}} \end{bmatrix}
\end{aligned}$$

In a sparse irregular longitudinal setting, the subject-specific covariance matrix will be modeled using $E_i = L_i \Sigma i L i^T$. Hence, for *i*-th subject, we have $\mathbf{\Sigma_i^{-1}} = \begin{bmatrix} \mathbf{\Sigma_{11}^{(i)}} & \mathbf{\Sigma_{12}^{(i)}} \\ \mathbf{\Sigma_{12}^{(i)T}} & \mathbf{\Sigma_{22}^{(i)}} \end{bmatrix}$.

Similar to the univariate case, we express the unconstrained dependence parameters $\log \sigma_{t1}^2$, $\log \sigma_{t0}^2$ $\phi_{t,t'}$ and $\Psi_{t,t'}$ as the following polynomial functions:

$$\log \sigma_{t1}^2 = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + ... + \lambda_g t^g, \qquad (10)$$

$$\log \sigma_{t0}^2 = \delta_0 + \delta_1 t + \delta_2 t^2 + ... + \delta_g t^g, \qquad (11)$$

$$\phi_{t,t'} = \eta_0 + \eta_1(t-t') + \eta_2(t-t')^2 + ... + \eta_h(t-t')^h, \quad (t' = 1, 2, ..., t-1), \qquad (12)$$

$$\Psi_{t,t'} = \theta_0 + \theta_1(t-t') + \theta_2(t-t')^2 + ... + \theta_h(t-t')^h, \quad (t' = 1, 2, ..., t-1), \qquad (13)$$

Here, again, the covariance estimation is essentially the estimation of the parameters λs, δs, ηs and θs. This can be done in a likelihood based framework or in a Bayesian approach using MCMC (Das *et al*., 2013b). The optimal orders of the above polynomials *g* and *h* are determined from the data using the information criteria. Das and Daniels (2014) used a Bayesian approach and obtained the optimal *g* and *h* as the corresponding posterior modes.

4. Covariance Estimation for the grouped longitudinal data

4.1 Univariate regular case

Gaskins and Daniels (2012) developed a non-parametric Bayesian approach for the simultaneous covariance estimation of the univariate longitudinal responses coming from *M* related groups. Such data are obtained from the meta-analyses where a certain experiment is performed at different centres with different sets of subjects over a period of time and the results are noted. Note that since the nature of the experiment is the same, the results obtained from the different centres are likely to be related and this should result in the similarity of the covariance features over the centres. Let Y_{im} be the *T* dimensional response vector for the *i*-th subject belonging to the *m*-th group. The *m*-th group consists of n_m subjects and without loss of generality, we assume that $Y_{im} \sim N_T(0, \Sigma_{\mathrm{m}})$, where Σ_m denotes the covariance matrix for the *m*-th group. Note that traditionally for such data, we either assume the homoscedasticity and assume the same covariance matrix for all the groups. Alternatively, one can assume complete heteroscedasticity and consider completely different Σ_m for each group. However, Gaskins and Daniels (2012) proposed the approach where Σ_m are modelled and estimated simultaneously and the parameters related to Σ_m are similar and/or same over time.

Similar to the Pourahmadi's (1999) approach as described in Section 2.2, The group-specific covariance matrix Σ_m can be decomposed as the following : $\Sigma_m = \Sigma(\Phi_m, \Gamma_m)$ and then from equation (3), we have,

$\Sigma_m^{-1} = L^T(\Phi_m)E^{-1}(\Gamma_m)L(\Phi_m)$ Note that Φ_m and Γ_m respective denote the vector of the auto-regressive parameters and the innovation variances. E is $T \times T$ dimensional diagonal matrix and L is a lower triangular matrix with $\Phi_m = (\phi_{m1}, \phi_{m2}, ... , \phi_{mj})$, where $J = \frac{T(T-1)}{2}$.

The key idea here is to propose a suitable prior distribution of the auto-regressive parameters and the innovation variances such that the groups share information in terms of the covariance parameters. The matrix stick-breaking process (MSBP) prior, developed by Dunson *et al*. (2008) can handle such information exchange across the groups. Gaskins and Daniels (2012) proposed similar priors with proper modification for the covariance parameters as the following.

4.1.1 Prior for Φ

The prior for *Φ*, referred to as the lag-block grouping prior by Gaskins and Daniels (2012), is given as the following:

$$
\begin{aligned}
&\phi_{mj} \sim F_{mj} = \sum_{h=1}^{H_\phi} \pi_{mjh}\delta_{\epsilon_{q(j)h}}(.); \quad m = 1,\ldots,M; \;\; j = 1,\ldots,J;\\
&\epsilon_{qh} \sim \epsilon_q\delta_0(.) + (1-\epsilon_q)N(0,\sigma^2); \;\; q = 1,...T-1; \;\; h = 1,\ldots,H_\phi;\\
&\pi_{mjh} = U_{mh}X_{jh}\prod_{l<h}(1-U_{ml}X_{jl});\\
&U_{mh} \overset{\text{iid}}{\sim} Beta(1,\alpha_\phi), \;\; X_{jh} \overset{\text{iid}}{\sim} Beta(1,\beta_\phi); \;\; h = 1,\ldots,H_\phi - 1.
\end{aligned}
$$

Under this setting, the parameters *ϕmj* are drawn from the random probability measures F_{mj} which are modelled as the truncated MSBP, with a zero-inflated normal base distribution. The mixing probabilities π_{mjh} are split into two components U_{mh} and X_{jh} which allocate the *j*-th parameter from the *m*-th group to the *h*-th cluster as a property of MSBP (Dunson *et al*., 2008). Note that the point masses in F_{mj} are drawn for all parameters *j* of the same lag *q*(*j*). Hence Gaskins and Daniels called this as lag block prior. The statistical properties of the above prior structure are discussed in detail in Gaskins and Daniels (2012).

4.1.2 Prior for Γ

For the innovation variances, the following prior distribution was proposed:

$$
\begin{aligned}
&\gamma_{mj} \sim G_{mj} = \sum_{h=1}^{H_\gamma} \tau_{mjh}\delta_{\eta_{jh}}(.); \quad m = 1,\ldots,M; \;\; j = 1,\ldots,J;\\
&\eta_{jh} = exp(\omega_{jh}); \omega_h = (\omega_{1h},\ldots,\omega_{Th})^T \sim N_T(\psi 1_T, \Omega R(\rho)),\\
&\tau_{mjh} = W_{mh}Z_{jh}\prod_{l<h}(1-W_{ml}Z_{jl});\\
&W_{mh} \overset{\text{iid}}{\sim} Beta(1,\alpha_\gamma), \;\; Z_{jh} \overset{\text{iid}}{\sim} Beta(1,\beta_\gamma); \;\; h = 1,\ldots,H_\gamma - 1.
\end{aligned}
$$

Here the base distribution is an exponential distribution and the mean parameters are jointly modelled in a correlated fashion, *i.e*. drawn from a *T* variate normal density. Note that the parameters Ψ and Ω are scalers and *R*(*ρ*) is essentially AR(1).

Through extensive simulation studies, Gaskins and Daniels (2012) have shown the empirical evidence for the effectiveness of this approach compared to the common covariance (homoscedasticity) and group-specific covariance (complete heteroscedasticity) structure. This approach is quite flexible and robust too.

4.2 Bivariate Irregular Case

The additional difficulty in the bivariate and/or multivariate longitudinal data coming from multiple related groups is the inter-response and the intra-response dependence. Also when there is sparsity in the data, the challenge is to model the subject-specific covariance matrices. Das and Daniels (2014) proposed semiparametric approach for handling such data.

We use the notations similar to Section 3.2. At time *t*, the measurement for *k*-th response feature for a subject in group m is modelled as the following,

$$
y_{tkm} = \sum_{t'=1}^{t-1}\phi^{(k)}_{t,t',m}y_{t'km} + \sum_{t'=1}^{t-1}\psi^{(1-k)}_{t,t',m}y_{t'(1-k)m} + \epsilon_{tkm}, \tag{14}
$$

where $\phi^{(k)}_{t,t',m}$ and $\Psi^{(1-k)}_{t,t',m}$ are autoregressive coefficients and $\in_{\text{tkm}}$ is the prediction error with mean 0 and variance σ^2_{tkm}. Let $y_m = \left[\text{y}^{\text{T}}_{1\text{m}}, \text{y}^{\text{T}}_{0\text{m}}\right]^{\text{T}}$ be the response vector of dimension 2*T* ×1 and $\in_m = \left[\in^{\text{T}}_{1\text{m}}, \in^{\text{T}}_{0\text{m}}\right]^{\text{T}}$. The above equation then can be expressed as:

$$\mathbf{y}_m = \begin{bmatrix} \mathbf{\Phi_{m1}} & \mathbf{\Psi_{m0}} \\ \mathbf{\Psi_{m1}} & \mathbf{\Phi_{m0}} \end{bmatrix} \mathbf{y}_m + \boldsymbol{\epsilon}_m, \qquad (15)$$

where Φ_{mk} and Ψ_{mk} are both $T \times T$ lower triangular matrices with 0's in the diagonal elements and $\phi^{(k)}_{t,t',m}$ and $\Psi^{(k)}_{t,t',m}$ (t > t′) in the (t, t′)-th position respectively. Thus we have,

$$\mathrm{L_m Y_m} = \epsilon_\mathrm{m} \qquad (16)$$

where $\mathbf{L}_m = \begin{bmatrix} \mathbf{I} - \mathbf{\Phi_{m1}} & -\mathbf{\Psi_{m0}} \\ -\mathbf{\Psi_{m1}} & \mathbf{I} - \mathbf{\Phi_{m0}} \end{bmatrix}$. Assuming that ϵ_{tkm}'s are uncorrelated, we obtain cov(ϵm) = $\mathbf{E}_m = \begin{bmatrix} \mathbf{E}_m^{(1)} & 0 \\ 0 & \mathbf{E}_m^{(0)} \end{bmatrix}$, where $\mathrm{E}_m^{(1)}$ and $\mathrm{E}_m^{(0)}$ are $T \times T$ diagonal matrices with t-th diagonal element σ^2_{t1m} and σ^2_{t0m}, respectively. We have from equation (16),

$$\mathrm{cov}(\epsilon_\mathrm{m}) = \mathrm{L_m cov(y_m)}\ \mathrm{L_m^T} = \mathrm{L_m \Sigma_m L_m^T} = \mathrm{E_m} \qquad (17)$$

The subject-specific covariance matrices will be modelled using $\mathrm{E_{im}} = \mathrm{L_{im}}\ \Sigma_\mathrm{im} \mathrm{L_m^T}$. The generalized autoregressive parameters ($\phi^{(k)}$'s and $\Psi^{(k)}$'s) and the logarithm of the innovation ($\log\sigma_1^2$ and $\log\sigma_0^2$) are modelled as polynomial functions as discussed in Section 3.2.

$$\log \sigma^2_{t1m} = \lambda_{0m} + \lambda_{1m} t + \lambda_{2m} t^2 + ... + \lambda_{gm} t^g, \qquad (18)$$

$$\log \sigma^2_{t0m} = \delta_{0m} + \delta_{1m} t + \delta_{2m} t^2 + ... + \delta_{gm} t^g, \qquad (19)$$

$$\phi^{(1)}_{t,t',m} = \eta^{(1)}_{0m} + \eta^{(1)}_{1m}(t-t') + \eta^{(1)}_{2m}(t-t')^2 + ... + \eta^{(1)}_{hm}(t-t')^h,\ (t' = 1, 2, ...t-1) \qquad (20)$$

$$\phi^{(0)}_{t,t',m} = \eta^{(0)}_{0m} + \eta^{(0)}_{1m}(t-t') + \eta^{(0)}_{2m}(t-t')^2 + ... + \eta^{(0)}_{hm}(t-t')^h,\ (t' = 1, 2, ...t-1) \qquad (21)$$

$$\psi^{(1)}_{t,t',m} = \theta^{(1)}_{0m} + \theta^{(1)}_{1m}(t-t') + \theta^{(1)}_{2m}(t-t')^2 + ... + \theta^{(1)}_{hm}(t-t')^h,\ (t' = 1, 2, ...t-1) \qquad (22)$$

$$\psi^{(0)}_{t,t',m} = \theta^{(0)}_{0m} + \theta^{(0)}_{1m}(t-t') + \theta^{(0)}_{2m}(t-t')^2 + ... + \theta^{(0)}_{hm}(t-t')^h,\ (t' = 1, 2, ...t-1) \qquad (23)$$

Note that we have 2(g+2h+3) covariance parameters for each group. Even for moderate values of g, h and m, this results in quite a large number of parameters to estimate. As a result, the matrix stick breaking process (Dunson *et al.*, 2008) was used for the covariance parameters λ_m, δ_m, $\eta_m^{(1)}$, $\eta_m^{(0)}$, $\theta_m^{(1)}$ and $\theta_m^{(0)}$ to effectively reduce the dimension.

4.2.1 Priors for $\eta_m^{(1)}$, $\eta_m^{(0)}$, $\theta_m^{(1)}$ and $\theta_m^{(0)}$

Similar to the univariate case, the MSBP prior was proposed for the covariance parameters for sharing the parameters across M groups. For j' = 0, . . . , h, assume

$$\eta^{(1)}_{j'm} \sim F^{\eta^{(1)}}_{j'm} = \sum_{q=1}^{N_{\eta^{(1)}}} \pi^{\eta^{(1)}}_{j'mq} \delta_{\epsilon^{\eta^{(1)}}_{j'q}}(.);\ \ m = 1, \ldots, M;\ \ j' = 0, \ldots, h;$$

$$\epsilon^{\eta^{(1)}}_{j'q} \overset{\text{iid}}{\sim} F^{\eta^{(1)}}_{0j'},$$

where δ_χ is a point mass at χ. Define $\varepsilon^\eta = \left(\in^{\eta(1)}_{j'q}\right)$ as an $(h+1) \times \mathrm{N}_{\eta(1)}$ matrix of random atoms. In the above truncated MSBP representation, the rows of ε^η correspond to the parameters having base distribution $\mathrm{F}^{\eta(1)}_{0'q}$ and the columns correspond to the clusters. The weights, $\pi^{\eta(1)}_{j''mq}$ are defined as

$$\pi^{\eta^{(1)}}_{j'mq} = V^{\eta^{(1)}}_{j'mq} \prod_{l<q} (1 - V^{\eta^{(1)}}_{j'ml}), \qquad (24)$$

with $V^{\eta(1)}_{j''mq} = U^{\eta(1)}_{j''mq} W^{\eta(1)}_{j'q}$ and $U^{\eta(1)}_{j''mq} \overset{iid}{\sim} Beta\,(1, \alpha_{\eta(1)})$ and $W^{\eta(1)}_{j'q} \overset{iid}{\sim}\ Beta\ (1, \beta_{\eta(1)})$. We take $V^{\eta(1)}_{j'mN_{\eta(1)}} = 1$; for all m

and j' to make $F^{\eta(1)}_{j'm}$ a valid probability measure. The full matrix stick-breaking process corresponds to the limiting case, $N_{\eta(1)} = \infty$. The current formulation is a truncated version of that.

Das and Daniels (2014) proposed a zero-inflated normal base distribution for $\eta_m^{(1)}$ as in Gaskins and Daniels (2012),

$$\begin{aligned}
&\epsilon^{\eta^{(1)}}_{j'q} \sim (1 - B^{\eta^{(1)}}_{j'})\delta_0(.) + B^{\eta^{(1)}}_{j'} N(0, \sigma^2_{\eta^{(1)}}); \quad q = 1, \ldots, N_{\eta^{(1)}}; \quad j' = 0, \ldots, h;\\
&B^{\eta^{(1)}}_{j'} = \prod_{l=0}^{j'} A^{\eta^{(1)}}_{l}; \quad A^{\eta^{(1)}}_{j'} | \pi^{\eta^{(1)}}_{j'} \overset{\text{ind}}{\sim} Bernoulli(\pi^{\eta^{(1)}}_{j'});\\
&\pi^{\eta^{(1)}}_{j'} \overset{\text{iid}}{\sim} Uniform(0,1); j' = 0, \ldots, h;
\end{aligned} \tag{25}$$

where $B^{\eta(1)}_{j'}$ is a binary random variable (for each j′) taking value 1 only when each of the independent Bernoulli random variables $\left\{A_l^{\eta^{(1)}} : l = 0, \ldots, j'\right\}$ take value 1. Here, for each j', $B^{\eta(1)}_{j'}$ is 1 only when all the lower lag coefficients are non-zero. This specification implicitly allows the data to select the order of the polynomials in (20) in an automated way such that a non-zero higher order term cannot appear with zero lower order terms and avoids the need for a two-step approach of selecting the order and then fitting the model (Pan and Mackenzie, 2003).

For $\eta_m^{(0)}$, $\theta_m^{(1)}$ and $\theta_m^{(0)}$, the same prior specification is assumed.

4.2.2 Priors for λ_m and δ_m

We note that λ_m and δ_m are the parameters to model the innovation variances. For λ_m, the proposed prior was:

$$\begin{aligned}
&\lambda_{mj'} \sim F^{\lambda}_{mj'} = \sum_{q=1}^{N_\lambda} \pi^{\lambda}_{mj'q} \delta_{\epsilon^{\lambda}_{j'q}}(.); \quad m = 1, \ldots, M; \quad j' = 0, \ldots, g;\\
&\epsilon^{\lambda}_{j'q} \sim (1 - B^{\lambda}_{j'})\delta_0(.) + B^{\lambda}_{j'} N(0, \sigma^2_\lambda); \quad q = 1, ..., N_\lambda; \quad j' = 0, \ldots, g;\\
&B^{\lambda}_{j'} = \prod_{l=0}^{j'} A^{\lambda}_{l}; \quad A^{\lambda}_{j'} | \pi^{\lambda}_{j'} \overset{\text{ind}}{\sim} Bernoulli(\pi^{\lambda}_{j'}); \quad \pi^{\lambda}_{j'} \overset{\text{iid}}{\sim} Uniform(0,1); \quad j' = 0, \ldots, g;\\
&\pi^{\lambda}_{mj'q} = U^{\lambda}_{mq} W^{\lambda}_{j'q} \prod_{l<q} (1 - U^{\lambda}_{ml} W^{\lambda}_{j'l});\\
&U^{\lambda}_{mq} \overset{\text{iid}}{\sim} Beta(1, \alpha_\lambda), \quad W^{\lambda}_{j'q} \overset{\text{iid}}{\sim} Beta(1, \beta_\lambda); \quad q = 1, \ldots, N_\lambda - 1.
\end{aligned}$$

Similar to the prior for the autoregressive parameters, this prior will choose the order of the polynomial functions from (18-19) and allow equality across groups. Following the Bayesian tradition, an inverse gamma prior was taken for σ^2_λ. The same prior structure was proposed for δ_m.

Through the extensive simulation studies, Das and Daniels (2014) showed the superiority of the above approach compared to the traditional homoscedasticity and the complete heteroscedasticity approaches. Also they analysed data from the Framingham Heart Study (FHS) and compared the results from the above method of the covariance estimation and the traditional parametric models as discussed in Section 3.1.

5 Discussion

The complexity in the covariance estimation of the longitudinal outcomes is not completely solved yet and is a challenge even to the experts. The nature of the complexity depends on the nature of the experiment and the way the data was generated. In this article, we only reviewed a few standard parametric and some advanced non-parametric and semiparametric approaches of handling this issue. However, this is still an ongoing research topic and many flexible advanced methods are being proposed every year.

For longitudinal study, a common issue is the missingness. This occures due to death and/or withdrawal of the subjects from the study for various reasons. The missingness can be monotone in the sense that if a subject is missing at time t, then it will also be missing at time t', for all $t' > t$. Monotone missingness is typically easier to handle. In statistics, there is a rich literature on modelling the missing values and imputing the missing values under various restrictions. Longitudinal data with the missing values should not be treated as the irregular longitudinal data and hence a completely different approach has to be made for handling such data. This is indeed an interesting research area in our times. Some discussions on this topic can be found in Daniels and Hogan (2008).

Another level of challenge occur due to zero-inflation in the longitudinal outcomes. This is a common scenario in health economics, business and management. For such data, typically the analysis is done by considering a random effects model but estimation of the actual covariance structure under this setting in not well addressed yet in the literature. Of course, the complexity will increase if we have multivariate zero-inflated longitudinal data with missingness. And the difficulty level can be further increased if the data come from multiple related groups. Generalisation of the approaches discussed in this article will not be a straight forward extension and need a careful study. Potential researchers can think of these issues based on real applications in various disciplines.

REFERENCES

Bandyopadhyay, D., Lachos, V.H., Abanto-Valle, C.A., and Ghosh, P. 2010. Linear Mixed Models for Skew-Normal/Independent bivariate responses with an application to Peri- odontal Disease. Stat. Medicine **29** : 2643-55.

Chatterjee, A., Venkateswaran, P., and Das, K. 2016. Simultaneous State Estimation for Clustered Based Wireless Sensor Networks. IEEE Transactions on Wireless Communications **15**, 7985–95.

Daniels, M.J., and Pourahmadi, M. 2002. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* **89** : 553-66.

Daniels, M.J. and Hogan, J.W. 2008 Missing data in longitudinal studies: Strategies for Bayesian Modeling and Sensitivity Analysis. Chapman and Hall (CRC Press).

Das, K., Li, J., Fu, G., Wang, Z., and Wu, R. 2011. Genome-wide association studies for bivariate sparse longitudinal data. *Human Heredity* **72 :** 110-20.

Das, K., Li, J., Fu, G., Wang, Z., Li, R., and Wu, R. 2013a. Dynamic semi-parametric Bayesian models for genetic mapping of complex traits with irregular longitudinal data. Stat. Medicine **32** : 509–23.

Das, K., Li, R., Sengupta, S., and Wu, R. 2013b. A Bayesian Semi-parametric Model for Bivariate Sparse Longitudinal Data. Stat. Medicine **32** : 3899–10.

Das, K., and Daniels, M. 2014. A Semi-parametric Approach to Simultaneous Covariance Estimation for Bivariate Sparse Longitudinal Data. *Biometrics* **70** : 33–43.

Das, K., Afriyie, P., and Spirko L. 2015. A semiparametric Bayesian model for analyzing longitudinal data from multiple related groups. The Int. J. Biostatistics **11** : 273–84.

Dunson, D.B., Xue, Y., and Carin, L. 2008. The matrix stick-breaking process: Flexible Bayes meta-analysis. J. of American Stat. Asso. **103 :** 317-27.

Gaskins, J.T., and Daniels, M.J. 2012. A nonparametric prior for simultaneous covariance estimation. *Biometrika*, **100 :** 125-38.

Ghosh, P., and Hanson, T. A. 2010. Semiparametric Bayesian approach to multivariate longitudinal data. Australian and New Zealand Journal of Statistics **52 :** 275–88.

Laird, N.M., and Ware, J.H. 1982. Random effects model for longitudinal data. *Biometrics* **38 :** 963-74.

Pan, J., and Mackenzie, G. 2003. On modelling mean-covariance structures in longitudinal studies. *Biometrika* **90 :** 239-44.

Pourahmadi, M. 1999. Joint mean-covariance model with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86** : 677-90.

Pourahmadi, M. 2000. Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika* **87 :** 425-35.

Rubin, D. 1976. Inference and missing data. Biometrika 63, 581–592. Sithole, J.S., and Jones, P.W. 2007. Bivariate Longitudinal Model for Detecting Prescribing Change in Two Drugs Simultaneously with Correlated Errors. J. Applied Statistics **34 :** 339-52.

Sy, J.P., Taylor, J.M.G., and Cumberland, W.G. 1997. A Stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53 :** 542-55.

Thiebaut, R., Jacqmin-Gadda, H., Chene, G., Leport, C., and Commenges, D. 2002. Bivariate linear mixed models using SAS PROC MIXED. Computer Methods and Programs in Biomedicine **69** : 249–56.

Wu,W.B., and Pourahmadi, M. 2003. Nonparametric estimation of large covariance matrices for longitudinal data. *Biometrika* **90 :** 831-44.