



The Effects of Covariance Structures on Modelling of Longitudinal Data

Yin Chen¹, Yu Fei², Jianxin Pan^{3*}

¹School of Insurance, University of International Business and Economics, Beijing, China

²School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, China

³School of Mathematics, University of Manchester, Manchester, UK

Email: chenyin@uibe.edu.cn, feiyukm@aliyun.com, *jianxin.pan@manchester.ac.uk

Received 20 September 2015; accepted 24 October 2015; published 28 October 2015

Copyright © 2015 by authors and OALib.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Extending the general linear model to the linear mixed model takes into account the within-subject correlation between observations with introduction of random effects. Fixed covariance structures of random error and random effect are assumed in linear mixed models. However, a potential risk of model selection still exists. That is, if the specified structure is not appropriate to real data, we cannot make correct statistical inferences. Joint modelling method removes all specifications about covariance structures and comes over the above risk. It simply models covariance structures just like modelling the mean structures in the general linear model. Our conclusions include: a) The estimators of fixed effects parameters are similar, that is, the expected mean values of response variables are similar. b) The standard deviations from different models are obviously different, which indicates that the width of confidence interval is evidently different. c) Through comparing the AIC or BIC value, we conclude that the data-driven mean-covariance regression model can fit data much better and result in more precise and reliable statistical inferences.

Keywords

Linear Mixed Models, Correlated Random Effects, Mean-Covariance Modeling, Longitudinal Data

Subject Areas: Mathematical Statistics, Statistics

1. Introduction

After the within subject correlation is taken into account in general linear models, the random errors are not assumed independent. Linear mixed models are setting up to solve this problems. Previous literature work on linear models and linear mixed models has mainly focused on setting up a list of possible covariance matrix

*Corresponding author.

structures of random errors and random effects. However, a potential vital risk still exists. If no one structure in list can fit data properly, even the best one in that list may distort statistical inference. Misspecification of covariance structures can lead to a great loss of efficiency of the parameter estimates (Wang and Carey, 2003, [1]).

Joint modelling (Pourahmadi, (1999), [2]) is a data-driven methodology for modelling the variance-covariance matrix similar to the general linear model framework for mean modelling. The main idea of joint modelling is considering mean and variance components in equally important level, modelling the mean, variance and correlation of response variables in three submodels, building three design matrices for mean, variance and correlation respectively, and estimating three parameters vectors for these three submodels simultaneously. The advantage of these method is to reduce the high dimensionality of computation for the covariance matrix. Modified Cholesky decomposition of the inverse of covariance matrix, Σ^{-1} , guarantees that Σ^{-1} and Σ must be positive definite. The resulting parameters estimates are obtained and compared to those obtained using the above approaches. Assumed structures and joint modelling are compared and discussed further in terms of their ability to accurately estimate both regression coefficients and variance components.

In this paper, we aim to model covariance structures of random errors in the linear models in Section 2, model covariance structures of random errors and random effects in linear mixed models in Section 3 and model mean and covariance simultaneously based on modified Cholesky decomposition (Pourahmadi, (2000), [3]) in Section 4. Cattle data published by Kenward in 1987 [4] were analyzed using the above three models in Section 5. The core conclusion is that real data analysis confirms the superiority of joint modelling of mean-covariance. Further discussions are followed in Section 6.

2. General Linear Models with Correlated Errors

The linear regression model assumes that the response variables and the explanatory variables are related through

$$Y_i = X_i\beta + \epsilon_i, \tag{1}$$

$$\epsilon_i \sim N_{n_i}(\mathbf{0}, \sigma^2 I_{n_i}),$$

where, for the i^{th} subject of of m subjects, Y_i is an $n_i \times 1$ responses, X_i is an $n_i \times p$ design matrix, β is a $p \times 1$ fixed effect, random error ϵ_i is an $n_i \times 1$ random vector and assumed as normal distributed with mean zero and covariance matrix identity structure, I_{n_i} is an $n_i \times n_i$ identity matrix. Under Gauss assumption, $Y_i \sim N_{n_i}(X_i\beta, \sigma^2 I_{n_i})$,

However, response in real data may be not independent. The inferences of the regression parameters of primary interest take account of the likely correlation in the data. So linear models with correlated errors is considered,

$$\epsilon_i \sim N_{n_i}(\mathbf{0}, \Sigma_i(\alpha)). \tag{2}$$

Under this assumption, $Y_i \sim N_{n_i}(X_i\beta, \Sigma_i(\alpha))$, where $\Sigma_i(\alpha) = \sigma^2 V_i$.

2.1. Random Errors in AR(1) Covariance Structure

In this model, for the i^{th} subject, the $(j, k)^{th}$ element $v_{jk} = Cov(Y_{ij}, Y_{ik})$ in V_i has the form

$$v_{jk} = \sigma^2 \exp(-\phi |t_j - t_k|). \tag{3}$$

The correlation between a pair of measurements on the same subject decays towards zero as the time separation between the measurements increases. The rate of decay is faster for larger values of ϕ . Note that if the observation times, t_j , are equally spaced, say $t_{j+1} - t_j = d$ for all j , then Equation (3) is expressible as

$$v_{jk} = \sigma^2 \rho^{j-k}. \tag{4}$$

where $\rho = \exp(-\phi d)$ is the correlation between successive observations on the same subject.

A justification of (4) is to represent the random variables Y_{ij} as

$$Y_{ij} = x'_{ij}\beta + W_{ij}, i = 1, \dots, m; j = 1, \dots, n_i, \tag{5}$$

with

$$W_{ij} = \rho W_{ij-1} + Z_{ij}, \tag{6}$$

and Z_{ij} are mutually independent $N\{0, \sigma^2(1-\rho^2)\}$ random variables, to give $Var(Y_{ij}) = Var(W_{ij}) = \sigma^2$ as required. In view of (5) and (6), the exponential correlation model is sometimes called the *first order autoregressive model*, because (6) is the standard definition of a discrete-time first-order autoregressive process. By substituting $W_{ij} = Y_{ij} - x'_{ij}\beta$ into (6), we obtain the conditional distribution of Y_{ij} , given the preceding response, Y_{ij-1} , as

$$Y_{ij} | Y_{ij-1} \sim N(x'_{ij}\beta + \rho(Y_{ij-1} - x'_{ij-1}\beta), \sigma^2(1-\rho^2)). \tag{7}$$

This form treats both the explanatory variables and prior response as explicit predictors of the current outcome. So we can get the analytical form of V_i :

$$V_i = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i-1} \\ \rho & 1 & \rho & \dots & \rho^{n_i-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n_i-1} & \rho^{n_i-2} & \rho^{n_i-3} & \dots & 1 \end{bmatrix}_{n_i \times n_i} \tag{8}$$

2.2. Log Likelihood

The log of the likelihood is

$$\begin{aligned} \ell(\beta, \sigma^2, V_i) &= \log L(\beta, \sigma^2, V_i) = \log \prod_{i=1}^m f(y_i) \\ &= \sum_{i=1}^m \log \left[(2\pi)^{-\frac{1}{2}n_i} |\sigma^2 V_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y_i - X_i \beta)' (\sigma^2 V_i)^{-1} (y_i - X_i \beta) \right\} \right], \end{aligned} \tag{9}$$

for given V_i , the maximum likelihood estimator for β and σ is, namely,

$$\hat{\beta}(V_i) = \left(\sum_{i=1}^m X_i' V_i^{-1} X_i \right)^{-1} \sum_{i=1}^m X_i' V_i^{-1} y_i = (X' V^{-1} X)^{-1} X' V^{-1} y \tag{10}$$

$$\hat{\sigma}^2 = \frac{\{y - X \hat{\beta}\}' V^{-1} \{y - X \hat{\beta}\}}{\sum_{i=1}^m n_i}. \tag{11}$$

where $V = \text{diag}(V_1, V_2, \dots, V_m)$.

2.3. AIC Criterion

AIC and BIC Criterion are defined as before:

$$\begin{aligned} \text{AIC} &= 2 \log(\hat{L}) + 2K, \\ \text{BIC} &= 2 \log(\hat{L}) + \log(N)K, \end{aligned} \tag{12}$$

where \hat{L} is the maximum likelihood, K the number of parameters to be estimated in the model, and N the sample size. High-order terms are ignored here.

3. Linear Mixed Models and Restricted Maximum Likelihood Estimation (REML)

The distributional assumption about ϵ in Model (2) may be too restrictive. For example, when data are collected over time on an individual such as in repeated measurements data, it is more likely that observed measurements have a high correlation within individual. The linear mixed model extends the general linear model by allowing a more flexible specification of the covariance matrix of ϵ . In other word, it allows for both correlation and heterogenous variances, although we still assume normality.

Laird and Ware (1982) [5] provided the following description of a linear mixed model

$$Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i, \tag{13}$$

where, for the i^{th} subject of of m subjects, Y_i is an $n_i \times 1$ responses, X_i is an $n_i \times p$ design matrix, β is a $p \times 1$ fixed effect. Z_i is an $n_i \times q$ design matrix, γ_i is a $q \times 1$ random effect, $\gamma_i \sim N_q(\mathbf{0}, G_i)$, random effects are independent between subjects. Random error ϵ_i is an $n_i \times 1$ random vector and assumed as normal distributed with mean zero and covariance matrix identity structure R_i , R_i is an $n_i \times n_i$ identity matrix. $\epsilon_i \sim N_{n_i}(\mathbf{0}, R_i)$, random errors are independent between subjects.

The model also can be written in

$$Y = X\beta + Z\gamma + \epsilon, \tag{14}$$

$$\begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right], \tag{15}$$

$$\Sigma = Cov(Y) = ZGZ' + R, \tag{16}$$

$$Y \sim N(X\beta, \Sigma), \tag{17}$$

where Y is an $n \times 1$ responses, X is an $n \times p$ design matrix, β is a $p \times 1$ fixed effect, Z is an $n \times mq$ design matrix, γ is a $mq \times 1$ random effect, ϵ is an $n \times 1$ random error.

The method of restricted maximum likelihood, or REML estimation, was introduced by Patterson and Thompson (1971) [6] as a way of estimating variance components in a general linear model. The objection to the standard maximum likelihood procedure is that it produces biased estimators. In REML, the degree of freedom of fixed effects are taken into account implicitly, unlike in the ML where they are not. That is why the ML estimator are biased.

REML is an alternative maximum likelihood procedure which involves applying the ML idea to linear combinations of y . Denote $A = I - X(X'X)^{-1}X'$ and B the $n \times (n-p)$ matrix defined by the requirements that $BB' = A$ and $B'B = I$, where I denote the $(n-p) \times (n-p)$ identity matrix. Finally, let $Z = B'Y$.

For fixed α , where α is variance components parameters, defined as in Section 2, the maximum likelihood estimator for β is

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}Y = DY. \tag{18}$$

Also, the respective pdf of Y and $\hat{\beta}$ are

$$f(y) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y - X\beta)' \Sigma^{-1} (y - X\beta) \right\}, \tag{19}$$

and

$$g(\hat{\beta}) = (2\pi)^{\frac{1}{2}p} |X'\Sigma^{-1}X|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\hat{\beta} - \beta)' (X'\Sigma^{-1}X) (\hat{\beta} - \beta) \right\}. \tag{20}$$

Furthermore, $E(Z) = 0$ and Z and $\hat{\beta}$ are independent, whatever the value of β , as we now show. Firstly, we prove $E(Z) = 0$,

$$E(Z) = B'E(Y) = B'X\beta = B'BB'X\beta.$$

since $B'B = I$. But since $BB' = A$, this gives

$$E(Z) = B'AX\beta$$

$$AX = \left\{ I - X(X'X)^{-1}X' \right\} X = X - X = 0$$

Hence $E(Z) = 0$ as required.

Secondly, we prove $Cov(Z, \hat{\beta}) = 0$

$$\begin{aligned} Cov(Z, \hat{\beta}) &= E\left\{ Z(\hat{\beta} - \beta)' \right\} = E\left\{ B'Y(Y'D' - \beta') \right\} = B'E(Y Y')D' - B'E(Y)\beta' \\ &= B'\left\{ Var(Y) + E(Y)E(Y)' \right\}D' - B'E(Y)\beta' \end{aligned}$$

Now, substituting $E(Y) = X\beta$ into this last expression gives

$$Cov(Z, \hat{\beta}) = B'(\Sigma + X\beta\beta'X')D' - B'X\beta\beta'$$

Also,

$$X'D' = X'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} = I$$

and

$$B'\Sigma D' = B'\Sigma\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} = B'X(X'\Sigma^{-1}X)^{-1} = 0$$

Where $B'X = B'AX = 0$ as in the proof that $E(Z) = 0$.

It follows that $Cov(Z, \hat{\beta}) = B'X\beta\beta' - B'X\beta\beta' = 0$. Finally, in the multivariate Gaussian setting, zero covariance is equivalent to independence. It follows that the algebraic form of the (singular) multivariate Gaussian pdf of Z , expressed in term of Y , is proportional to the ratio $\frac{f(y)}{g(\hat{\beta})}$. To obtain the explicit form of this ratio, we use

the following standard result for the general linear mixed model that

$$(y - X\beta)' \Sigma^{-1} (y - X\beta) = (y - X\hat{\beta})' \Sigma^{-1} (y - X\hat{\beta}) + (\hat{\beta} - \beta)' (X'\Sigma^{-1}X) (\hat{\beta} - \beta).$$

Then, the pdf of $Z = B'Y$ is proportional to

$$\frac{f(y)}{g(\hat{\beta})} = (2\pi)^{-\frac{1}{2}(n-p)} |\Sigma|^{-\frac{1}{2}} |X'\Sigma^{-1}X|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (y - X\hat{\beta})' \Sigma^{-1} (y - X\hat{\beta}) \right\}, \tag{21}$$

where the omitted constant of proportionality is the Jacobian of the transformation from Y to $(Z, \hat{\beta})$. Harville (1974) [7] shows that the Jacobian reduces to $|XX'|^{-\frac{1}{2}}$, which does not depend on any of the parameters in the model and can therefore be ignored for making inferences about α and β . Note that the right hand side of (21) is independent of A , and the same result would therefore hold for any Z such that $E(Z) = 0$ and $Cov(Z, \hat{\beta}) = 0$.

The practical implication of (21) is that the REML estimator, $\tilde{\alpha}$, maximizes the log likelihood

$$\begin{aligned} l^*(\alpha) &= -\frac{n-p}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma| - \frac{1}{2} \ln|X'\Sigma^{-1}X| - \frac{1}{2} (y - X\hat{\beta})' \Sigma^{-1} (y - X\hat{\beta}) \\ &= -\frac{n-p}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma_{(\alpha)}| - \frac{1}{2} \ln|X'\Sigma_{(\alpha)}^{-1}X| \\ &\quad - \frac{1}{2} \left[y - X(X'\Sigma_{(\alpha)}^{-1}X)^{-1}X'\Sigma_{(\alpha)}^{-1}y \right] \Sigma_{(\alpha)}^{-1} \left[y - X(X'\Sigma_{(\alpha)}^{-1}X)^{-1}X'\Sigma_{(\alpha)}^{-1}y \right]. \end{aligned}$$

whereas the maximum likelihood estimator, MLE, $\hat{\alpha}$ maximizes

$$\begin{aligned} l^*(\alpha) &= -\frac{n-p}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \ln |X' \Sigma^{-1} X| - \frac{1}{2} (y - X \hat{\beta})' \Sigma^{-1} (y - X \hat{\beta}) \\ &= -\frac{n-p}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_{(\alpha)}| - \frac{1}{2} \ln |X' \Sigma_{(\alpha)}^{-1} X| \\ &\quad - \frac{1}{2} \left[y - X (X' \Sigma_{(\alpha)}^{-1} X)^{-1} X' \Sigma_{(\alpha)}^{-1} y \right]' \Sigma_{(\alpha)}^{-1} \left[y - X (X' \Sigma_{(\alpha)}^{-1} X)^{-1} X' \Sigma_{(\alpha)}^{-1} y \right]. \end{aligned}$$

4. Joint Mean-Covariance Model for Longitudinal Data

Response variables Y_i was assumed Gaussian distribution

$$Y_i \sim N_{n_i}(\mu_i, \Sigma_i).$$

4.1. Modified Cholesky Decomposition for Longitudinal Data

The subject-specific covariance matrices $\Sigma_i = Cov(Y_i)$ can be composed into

$$T_i \Sigma_i T_i' = D_i,$$

where T_i is a lower triangular matrix with 1's as diagonal entries and D_i is a diagonal matrix with positive diagonal entries.

Denote

$$T_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{i21} & 1 & 0 & \cdots & 0 \\ -\phi_{i31} & -\phi_{i32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{in_1} & -\phi_{in_2} & -\phi_{in_3} & \cdots & 1 \end{pmatrix}_{n_i \times n_i},$$

$$D_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{i2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{in_i}^2 \end{pmatrix}_{n_i \times n_i},$$

ϕ_{ijk} and σ_{ij}^2 have clear statistical interpretations:

a) ϕ_{ijk} are *autoregressive coefficients* in the autoregression models

$$\hat{y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \phi_{ijk} (y_{ik} - \mu_{ik}),$$

b) σ_{ij}^2 are the *innovation variances*

$$\sigma_{ij}^2 = Var(y_{ij} - \hat{y}_{ij}),$$

where $j = 1, \dots, n_i$ and $i = 1, \dots, m$.

4.2. Mean-Covariance Regression Model for Longitudinal Data

The unconstrained parameters μ_{ij} , ϕ_{ijk} and $\varsigma_{ij} \equiv \log \sigma_{ij}^2$ can be modelled in terms of regression models

$$\begin{cases} \mu_{ij} = x'_{ij} \beta \\ \phi_{ijk} = z'_{ijk} \gamma \\ \varsigma_{ij} = h'_{ij} \lambda \end{cases} \quad (22)$$

where β , γ and λ are p -, q - and d -dimensional parameter vectors, associated with the covariates x_{ij} , z_{ijk} and h_{ij} , respectively (Pourahmadi, 1999, [2]).

For growth curve data, such as the cattle data which we want to analyse, the covariates may take the forms:

$$\begin{cases} x_{ij} = \left(1, t_{ij}^{(1)}, t_{ij}^{(2)}, \dots, t_{ij}^{(p-1)}\right)' \\ z_{ijk} = \left(1, (t_{ij} - t_{ik})^{(1)}, (t_{ij} - t_{ik})^{(2)}, \dots, (t_{ij} - t_{ik})^{(q-1)}\right)' \\ h_{ij} = \left(1, t_{ij}^{(1)}, t_{ij}^{(2)}, \dots, t_{ij}^{(d-1)}\right)' \end{cases}$$

where the superscript “()” stands for the orthogonalization vector.

Linear mean-covariance model for growth curve data is

$$\begin{cases} \mu_{ij} = \beta_0 + \beta_1 t_{ij}^{(1)} + \beta_2 t_{ij}^{(2)} + \dots + \beta_{p-1} t_{ij}^{(p-1)} \\ \phi_{ijk} = \gamma_0 + \gamma_1 (t_{ij} - t_{ik})^{(1)} + \gamma_2 (t_{ij} - t_{ik})^{(2)} + \dots + \gamma_{q-1} (t_{ij} - t_{ik})^{(q-1)} \\ \varsigma_{ij} = \lambda_0 + \lambda_1 t_{ij}^{(1)} + \lambda_2 t_{ij}^{(2)} + \dots + \lambda_{d-1} t_{ij}^{(d-1)} \end{cases} \quad (23)$$

where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n_i$, and $1 \leq k \leq j - 1$.

4.3. Advantages of Joint Mean-Covariance Model

First, mean-covariance regression models have compact form and clear interpretations, so that they can be understood easily.

Second, the numerical linear algebra method can reduce the high dimensional $\frac{n_i(n_i + 1)}{2}$ parameters in Σ_i into $q + d$ dimensional parameters in terms of γ and λ . This approach can decrease the computational time to a great extent.

Third, modified Cholesky decomposition of Σ^{-1} can guarantee that Σ is positive definite.

Fourth, this model can be used commonly. For special cases, this approach can also be valid with certain constraints placed on T_i and D_i . For example, AR(1) structure

$$\Sigma_i = \sigma^2 \left(\rho^{|k-l|} \right)_{k,l=1}^{n_i},$$

$$T_i = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}_{n_i \times n_i},$$

$$D_i = \begin{bmatrix} \frac{\sigma^2}{1-\rho^2} & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix}_{n_i \times n_i}.$$

4.4. Maximum Likelihood Estimation for Joint Modelling

For clarity and simplicity of presentation, we only consider standard multivariate data or balanced longitudinal data (Diggle and Kenward, 1994) [8]. For $i = 1, 2, \dots, m$, we assure that $Y_i \sim N_{n_i}(X_i \beta, \Sigma_i)$ are independent

n_i -dimension vectors, where Y_i stand for the n_i repeated measurements on the i^{th} subject, $i=1, \dots, m$. X_i is $n_i \times p$ design matrix and β is $p \times 1$ vector of mean parameters. We assume that the components T_i and D_i of the covariance matrix Σ_i , are modelled as (22), so that the likelihood function has three representations corresponding to the three submodels in (22).

Log-likelihood function is

$$\begin{aligned} \ell(y_1, y_2, \dots, y_m, \beta, \gamma, \lambda) &= \log L(y_1, y_2, \dots, y_m, \beta, \gamma, \lambda) \\ &= \sum_{i=1}^m -\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^m \log |\Sigma_i| + \sum_{i=1}^m \left\{ -\frac{1}{2} (y_i - X_i \beta)' \Sigma_i^{-1} (y_i - X_i \beta) \right\}. \end{aligned} \tag{24}$$

4.5. Algorithm for Calculating Parameters

Step 1. Select an initial value $\tilde{\beta}$ for β .

Step 2. Compute $S^2 = (1/m) \sum_{i=1}^m (Y_i - X_i \tilde{\beta})(Y_i - X_i \tilde{\beta})'$ and its factors \tilde{T} and \tilde{D} to be used as initial values for T and D in the next step.

Step 3. For the inner loop, compute $\tilde{\theta} = (\tilde{\gamma}', \tilde{\lambda}')'$ by solving the last two equations above using Newton-Raphson method. At convergence, compute $T(\tilde{\gamma})$ and $D(\tilde{\lambda})$ and form $\tilde{\Sigma}^{-1} = T(\tilde{\gamma})' D^{-1}(\tilde{\lambda}) T(\tilde{\gamma})$.

Step 4. Update $\hat{\beta}$ using

$$\hat{\beta} = \left(\sum_{i=1}^m X_i \tilde{\Sigma}^{-1} X_i \right)^{-1} \sum_{i=1}^m X_i \tilde{\Sigma}^{-1} Y_i.$$

Step 5. Stop the process if $\hat{\beta} \approx \tilde{\beta}$. Take $\hat{\beta}$ as an estimate of β and form the estimate $\hat{\Sigma} \approx \tilde{\Sigma}$. Otherwise, repeat Steps 2 - 4 replacing $\tilde{\beta}$ by $\hat{\beta}$.

5. Cattle Data Analysis

Kenward (1987) [4] reported a set of data on cattle’s growth. 60 cattle were randomly allocated into two groups, each of which would hold 30 animals. The animals were put out to pasture at the start of the grazing season, and the members of each group received one of two different treatments, A and B say, respectively. The weight (response) of each animal was then recorded 11 times to the nearest kilogram over a 133-day period. The first 10 measurements were made at two-week intervals and the final measurement was made after one-week interval.

We extract the best structure in different models from different models and put them in the **Table 1**. The second column gives the result for classical linear model with identical matrix of random errors in Section 2. The third column shows the result for linear model with AR(1) structure covariance matrix of random errors in linear model in Section 2. The fourth column represent the Model (13) (AR(1) structure for random effect and AR(1) structure for random error) in Section 3. The last column lists the joint modelling of mean-covariance regression in Section 4.

6. Conclusions and Discussions

The main conclusions of this paper include:

- a) The estimated mean value for the response variable and the estimators of fixed effects parameters are similar whatever the covariance structure is chosen.
- b) The standard deviation of fixed effects parameters estimates is relatively different, which indicates the confidence intervals depend on the specification of covariance structure.
- c) As a model selection criterion, AIC or BIC value is very important. The smaller the AIC or BIC value is, the better the model is. We conclude that the joint mean-covariance regression model is the best model for analysing cattle data. Note that in all of these models, the sample size in AIC or BIC value is 660.

Apart from the above three conclusions, we can see some other hiding connotations. In the first column of **Table 1**, we assume all measurements are independent, and then the maximum marginal log-likelihood value is -2782.644 . When extending this model and assuming some correlations in successive measurements within individual, for example, the first order autoregressive model AR(1), the result shown in the second column becomes much better. The maximum marginal log-likelihood value increase from -2782.644 to -2289.697 .

Table 1. Comparison among the best structures in different models.

	Classical Linear Model	General Linear Model With AR(1)	LMM AR(1) (random effect) AR(1) (random error) REML	Linear Joint Model
$\hat{\mu} = X\hat{\beta}$	222.4591	225.2165	225.2492	219.2690
	233.6470	233.9385	234.0066	230.8766
	247.0486	245.8849	245.9751	244.5267
	261.7197	259.9175	260.0181	259.3281
	276.7159	274.8982	274.9988	274.3896
	291.0929	289.6890	289.7807	288.8202
	303.9065	303.1519	303.2270	301.7286
	314.2123	314.1488	314.2012	312.2236
	321.0661	321.5416	321.5666	319.4143
	323.5234	324.1923	324.1864	322.4093
322.8083	323.3837	323.3616	320.3175	
$\hat{\beta}$	283.47273	283.26935	283.3247	281.20939
	116.57233	116.54085	116.4581	117.19641
	-18.32619	-14.58151	-14.68762	-22.29320
	-10.52633	-12.68589	-12.67034	-11.67378
Std of $\hat{\beta}$	0.6403	0.1093	1.6366	1.8487
	2.1235	0.1630	2.6177	2.5327
	2.1235	0.0933	1.6960	1.4091
	2.1235	0.0665	1.3390	0.9784
ℓ	-2782.644	-2289.697	-2273.92	-2251.517
AIC	5575.288	4591.395	4563.841	4527.033
BIC	5597.749	4618.348	4599.779	4580.94

Simultaneously, AIC value reduces from 5575.288 to 4591.395 and BIC from 5597.749 to 4618.348. Obviously, these two values reduce remarkably. Large reduction implies that the correlation between successive observations should not be ignored. It also implies that the joint mean-covariance model provides correct information from the increasing of log likelihood or from changes of AIC and BIC value.

What's more, we expect that if the structure of variance is AR(2), the result will be better than AR(1) structure. This part of work can be done as further work.

When we separate the variance-covariance matrix for each animal, Σ_i , into two parts, G_i for random effect between subjects and R_i for random error within subject. The maximum log-likelihood value increases very little (from -2289.697 to -2273.92) compared to the last increase from column 2 to column 3. If R_i is not AR(1) structure, for example, we try identity and identity for R_i , the log-likelihood value in the linear mixed model is not better than that in general linear model, which indicates that the AR(1) structure for R_i is more appropriate than other specifications.

Another further work on the cattle data analysis that may be carried out is to assume that different groups of animals may have different covariance structures, unlike in this dissertation where we assume the same cova-

riance structure applied in both groups of animals. Other modelling methods, for example, nonlinear regression model might be also used to analyse cattle data.

Acknowledgements

We thank the editors and the reviewers for their comments. This research is funded by the National Social Science Foundation No. 12CGL077, National Science Foundation granted No. 71201029, No. 71303045 and No. 11561071. This support is greatly appreciated.

References

- [1] Wang, Y.G. and Carey, V. (2003) Working Correlation Structure Misspecification, Estimation and Covariate Design: Implications for Generalised Estimating Equations Performance. *Biometrika*, **90**, 29-41. <http://dx.doi.org/10.1093/biomet/90.1.29>
- [2] Pourahmadi, M. (1999) Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterization. *Biometrika*, **86**, 677-690. <http://dx.doi.org/10.1093/biomet/86.3.677>
- [3] Pourahmadi, M. (2000) Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, **87**, 425-435. <http://dx.doi.org/10.1093/biomet/87.2.425>
- [4] Kenward, M.G. (1987) A Method for Comparing Profiles of Repeated Measurements. *Applied Statistics*, **36**, 296-308. <http://dx.doi.org/10.2307/2347788>
- [5] Laird, N.M. and Ware, J.J. (1982) Random-Effects Models for Longitudinal Data. *Biometrics*, **38**, 963-974. <http://dx.doi.org/10.2307/2529876>
- [6] Patterson, H.D. and Thompson, R. (1971) Recovery of Interblock Information When Block Sizes Are Unequal. *Biometrika*, **58**, 545-554. <http://dx.doi.org/10.1093/biomet/58.3.545>
- [7] Harville, D. (1974) Bayesian Inference for Variance Components Using Only Error Constrasts. *Biometrika*, **61**, 383-385. <http://dx.doi.org/10.1093/biomet/61.2.383>
- [8] Diggle, P.J. and Kenward, M.G. (1994) Informative Drop-Out in Longitudinal Data Analysis. *Applied Statistics*, **43**, 49-93. <http://dx.doi.org/10.2307/2986113>