Journal of Data Science **6**(2008), 573-589

Analysis of Covariance Structures in Time Series

Jennifer S. K. Chan[1] and S. T. Boris Choy[2]
[1] *The University of Sydney and* [2] *University of Technology, Sydney*

Abstract: Longitudinal data often arise in clinical trials when measurements are taken from subjects repeatedly over time so that data from each subject are serially correlated. In this paper, we seek some covariance matrices that make the regression parameter estimates robust to misspecification of the true dependency structure between observations. Moreover, we study how this choice of robust covariance matrices is affected by factors such as the length of the time series and the strength of the serial correlation. We perform simulation studies for data consisting of relatively short (N=3), medium (N=6) and long time series (N=14) respectively. Finally, we give suggestions on the choice of robust covariance matrices under different situations.

Key words: Longitudinal data, robustness, serial correlation.

1. Introduction

In ordinary regression, the error terms are assumed to be independently and identically distributed with a constant variance. However, these assumptions are not always true in practice. For example, when subjects are measured repeatedly over time in longitudinal studies, observations from each subject are usually serially correlated. Such time series are common in clinical trials and stock markets. Then, it is customary to consider a generalized regression model in which errors are modeled by a covariance matrix with non-constant variances and non-zero covariances. Lindsey (1993) points out the importance of estimating such a covariance matrix. However, the number of unknown elements in the matrix can increase quadratically with its dimension causing considerable difficulties in estimation. Various approaches have been suggested, see for example, Efron and Morris (1976) who estimated the inverse of a covariance matrix using a loss function, Yang and Berger (1994) who applied a spectral decomposition to a covariance matrix, Chiu *et al.* (1996) who applied a matrix exponential transformation to a covariance matrix and estimated the parameters in the transformed matrix as well as the mean model using an ML approach and Barnard *et al.* (2000) who

modeled the covariance matrix in terms of its standard deviations and correlation matrix and estimated the parameters using a Bayesian approach. Instead of resorting to this complicated techniques, we propose the generalized regression models in which the covariance matrices are based on widely used forms, and we seek those structures that are robust to misspecification of the true structures which are often unknown in practice.

We use the PROC MIXED procedure in SAS, the Statistical Analysis Software, to fit the generalized regression models for continuous outcomes, incorporating different choices of covariance structures with parameters estimated from the restricted maximum likelihood (REML) method (Patterson and Thompson, 1971). Fortran programs are used to simulate data sets of time series from multivariate normal distributions with different covariance matrices. Then the simulation experiments are carried out by fitting models adopting different covariance matrices repeatedly using the PROC MIXED procedures written in SAS macros.

In section 2, we describe the model and estimation procedures. We introduce some common covariance structures and their interpretation in terms of dependency structures between observations. Three criteria for assessing goodness-of-fit are also given. Performance of the models adopting different covariance matrices can be compared through simulation. Since the length of the time series and the strength of the serial correlation may affect the choice of robust covariance structures, simulation experiments are performed in section 3, using two data sets: the blood glucose data for relatively short and medium time series with weak serial correlation (N=3 and 6 and $\rho = 0.36$ in the covariance matrix of AR1 type), and the plasma citrate concentration data available for relatively longer time series with stronger serial correlation (N=14 and $\rho = 0.70$). True values of parameters in the simulation experiments are set equal to the estimates obtained from fitting the two data sets. Moreover we also consider other sets of the parameter values in the covariance matrix in order to study the interaction effect of the two factors: the length of the time series and the strength of the serial correlation on the choice of robust covariance structure. Results from simulations are analyzed and compared in section 4. Finally, in section 5, conclusions are drawn on the best covariance structures that give consistently the best fit, regardless of the true covariance structures.

2. The Model

We assume a generalized regression model of the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y} = (Y_{11}, Y_{12}, \ldots, Y_{MN})^T$ is a vector of MN observed outcomes based on M time series each containing responses from one subject on N equally spaced

time points, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{P_b})^T$ is a vector of $(P_b + 1)$ parameters, $\boldsymbol{X}$ is a $MN \times (P_b+1)$ design matrix, $\boldsymbol{\varepsilon}$ is a vector of MN residuals such that $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is the block diagonal matrix with covariance matrix $\boldsymbol{\Sigma}_0$ for each subject. The covariance matrix $\boldsymbol{\Sigma}_0$ with P_v variance parameters and P_c covariance or correlation parameters describes the relationship between observations within each subject; we assume that the correlations between observations from different subjects are zero. There are $P = P_b + P_v + P_c + 1$ parameters in total; emphasis is on the estimation of $\boldsymbol{\beta}$.

Table 1: The seven types of the covariance matrix

Type		**Properties**
Simple (SIM)	$\sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	Constant variance $Var(\varepsilon_i) = \sigma^2$ Zero covariance $Cov(\varepsilon_i, \varepsilon_j) = 0$ $P_v = 1$ & $P_c = 0$
Equal correlation (EC)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$	Constant variance $Var(\varepsilon_i) = \sigma^2$ Constant covariance $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2\rho$ $P_v = 1$ & $P_c = 1$
Unstructured independent (UN1)	$\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$	Non-constant variance $Var(\varepsilon_i) = \sigma_i^2$ Zero covariance $Cov(\varepsilon_i, \varepsilon_j) = 0$ $P_v = N = 3$ & $P_c = 0$
Unstructured 2 bands (UN2)	$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ 0 & \sigma_{23} & \sigma_3^2 \end{bmatrix}$	Non-constant variance $Var(\varepsilon_i) = \sigma_i^2$ Non-zero covariance for given time lag=1 $Cov(\varepsilon_i, \varepsilon_j) = \sigma_{ij}$, $\text{abs}(i-j) = 1$ $P_v = N = 3$ & $P_c = N - 1 = 2$
First-order autoregressive (AR1)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$	Constant variance $Var(\varepsilon_i) = \sigma^2$ Equal covariance for given time lag k $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2\rho^k$, $\text{abs}(i-j) = k$ $P_v = 1$ & $P_c = 1$
Toeplitz (TOEP)	$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}$	Constant variance $Var(\varepsilon_i) = \sigma^2$ Equal covariance for given time lag k $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2\rho_k$, $\text{abs}(i-j) = k$ $P_v = 1$ & $P_c = N - 1 = 2$
Toeplitz) 2 bands (TOEP2)	$\sigma^2 \begin{bmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_1 \\ 0 & \rho_1 & 1 \end{bmatrix}$	Constant variance $Var(\varepsilon_i) = \sigma^2$ Equal covariance for given time lag=1 $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2\rho_1$, $\text{abs}(i-j) = 1$ $P_v = 1$ & $P_c = 1$

The restricted maximum likelihood (REML) estimates of the $P_v + P_c$ parameters in $\boldsymbol{\Sigma}_0$ can be obtained by maximizing the reduced log-likelihood

$$\begin{aligned} L^* &= -\frac{1}{2}\left\{\log\left[\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right]^T\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right]+M\log(|\boldsymbol{\Sigma}_0|)\right\} \\ &\quad -\frac{1}{2}\log(|\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|) \end{aligned} \tag{2.1}$$

and the REML estimate of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y} \tag{2.2}$$

where $\widehat{\boldsymbol{\Sigma}}$ is obtained by substituting the block diagonal matrix $\boldsymbol{\Sigma}_0$ by $\widehat{\boldsymbol{\Sigma}}_0$. As the REML estimate, $\widehat{\boldsymbol{\Sigma}}_0$ using (2.1), depends on $\widehat{\boldsymbol{\beta}}$ and the REML estimate, $\widehat{\boldsymbol{\beta}}$ using (2.2), depends on $\widehat{\boldsymbol{\Sigma}}_0$ in $\widehat{\boldsymbol{\Sigma}}$, iterations are required. Moreover, as $\widehat{\boldsymbol{\beta}}$ depends on $\boldsymbol{\Sigma}_0$, we seek some covariance matrices for $\boldsymbol{\Sigma}_0$ that consistently give the best estimate of $\boldsymbol{\beta}$ and the best fit to data if the true dependency structure is misspecified.

For short time series of length $N = 3$ say, the covariance matrix $\boldsymbol{\Sigma}_0$ can be modeled by $J = 7$ types, as listed in Table 1.

Note that we discard the unstructured (UN) covariance matrix with no constraints imposed on its entities σ_{ij} because it contains too many parameters and offers no summary of information. Practically, correlation between observations with large time lag is very small and hence it can be well approximated by UN2 especially for time series of medium to long in length. For time series of medium length, $N = 6$ say, SIM and EC can be represented by $\sigma^2\boldsymbol{I}_6$ ($P_v = 1,\ P_c = 0$) and $\sigma^2(1-\rho)\boldsymbol{I}_6 + \sigma^2\rho\boldsymbol{J}_6$ ($P_v = 1,\ P_c = 1$) respectively where $\boldsymbol{I}_6$ and $\boldsymbol{J}_6$ denote respectively a 6×6 identity matrix and a 6×6 matrix with all elements being 1. The remaining matrices are:

UN1

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6^2 \end{bmatrix}$$

$(P_v = 6,\ P_c = 0)$

UN2

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & 0 & 0 & 0 & 0 \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & 0 & 0 & 0 \\ 0 & \sigma_{32} & \sigma_3^2 & \sigma_{43} & 0 & 0 \\ 0 & 0 & \sigma_{43} & \sigma_4^2 & \sigma_{54} & 0 \\ 0 & 0 & 0 & \sigma_{54} & \sigma_5^2 & \sigma_{65} \\ 0 & 0 & 0 & 0 & \sigma_{65} & \sigma_6^2 \end{bmatrix}$$

$(P_v = 6,\ P_c = 5)$

AR1

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

$(P_v = 1,\ P_c = 1)$

TOEP

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_3 & \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_4 & \rho_3 & \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_5 & \rho_4 & \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

$(P_v = 1,\ P_c = 5)$

TOEP2

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & 0 & 0 & 0 & 0 \\ \rho_1 & 1 & \rho_1 & 0 & 0 & 0 \\ 0 & \rho_1 & 1 & \rho_1 & 0 & 0 \\ 0 & 0 & \rho_1 & 1 & \rho_1 & 0 \\ 0 & 0 & 0 & \rho_1 & 1 & \rho_1 \\ 0 & 0 & 0 & 0 & \rho_1 & 1 \end{bmatrix}$$

$(P_v = 1,\ P_c = 1)$

For time series of longer length, $N = 14$ say, the covariance matrices of $J = 7$ types are similarly defined as in the cases of $N = 3$ and $N = 6$. Some of these covariance matrices can be directly linked to particular models. Let Y_{mn} denote the outcome of the n-th measurement from subject m. Then EC corresponds to a *random intercept model* $Y_{mn} = \mu_{mn} + U_m + Z_{mn},\ m = 1, \ldots, M$ and $n = 1, \ldots, N$ where $\mu_{mn} = E(Y_{mn})$, $U_m \sim N(0, \sigma^2\rho)$, $Z_{mn} \sim N(0, \sigma^2(1-\rho))$ and U_m and Z_{mn} are independent. Hence we have $Var(Y_{mn}) = \sigma^2$, $Cov(Y_{mn}, Y_{mn'}) = \sigma^2\rho$ and $\mathbf{\Sigma}_0$ can be rewritten as $\sigma^2(1-\rho)\boldsymbol{I} + \sigma^2\rho\boldsymbol{J}$. (See Diggle et al. 1996) On the other hand, AR1 (see Diggle et al. 1996) corresponds to the model $Y_{mn} = \mu_{mn} + \varepsilon_{mn}$ and $\varepsilon_{mn} = \rho\varepsilon_{m,n-1} + e_{mn}$ where $e_{mn} \sim N(0, \sigma^2(1-\rho^2))$, $Var(\varepsilon_{mn}) = \sigma^2$ and $Cov(\varepsilon_{mn}, \varepsilon_{m,n-1}) = \sigma^2\rho$ and is one of the *exponential correlation models*. Here σ_{mn}, denoting the mn-th element in $\mathbf{\Sigma}_0$, is defined to be $\sigma^2 \exp(-\phi|m-n|) = \sigma^2\rho^{|m-n|}$ where ϕ is a constant showing the rate of decay. It implies that covariance between a pair of measurements on the same subject decays to zero as the time between measurements increases.

3. Goodness-of-fit Tests

In order to facilitate comparison between different covariance structures, data set simulated from model adopting the i-th ($i = 1, \ldots, I$) true covariance structures is fitted to models adopting each of the j-th ($j = 1, \ldots, J$) structures and there are $K = 200$ replicates for each (i, j) combination ($I = J = 7$). Then we assess the goodness-of-fit of the J models adopting different covariance structures based on three criteria $R_{hj},\ h = 1, 2, 3$: the estimated parameters, their standard errors and the Akaike's Information Criterion (AIC) (Akaike, 1973). Let x_{ij} denote the value of any of the three criteria in each (i, j) combination. For each

criterion, ranking is applied twice; firstly across the J fitted structures for each of the i-th true structure as given by $Rank_j(x_{ij})$ such that $Rank_j(x_{ij}) = 1$ when $x_{ij} < x_{ij'}$, for all $j' \neq j$ and $Rank_j(x_{ij}) = J$ when $x_{ij} > x_{ij'}$. When there are ties, we take the average of ranks. Then $Rank_j(x_{ij})$ is summed over i and second ranking R_{hj} is applied across the sums, $\sum_i Rank_j(x_{ij})$, for the J structures in the fitted model. The first ranking $Rank_j(x_{ij})$ is necessary because $\sum_i Rank_j(x_{ij})$ rather than $\sum_i x_{ij}$ will not be affected by the scale of x_{ij} as well as any outlying x_{ij}. If model with covariance structure j gives the smallest R_j, it provides the best fit to data of different true covariance structures and hence is considered the most robust covariance structure.

C1 **Rank of mean squared error**(RMSE):

$$R_{1j} = Rank_j\left(\sum_{i=1}^{J} RMSE_{ij}\right),$$

where

$$RMSE_{ij} = \sum_{p=0}^{P} Rank_j\left(\frac{1}{K}\sum_{k=1}^{K}(\widehat{\beta}_{ijkp} - \beta_p)^2\right),$$

and $\widehat{\beta}_{ijkp}$ is the parameter estimate for β_p in the k-th replicated data set which is simulated from model adopting the i-th true covariance structure and is fitted to model adopting the j-th covariance structure. This is a useful criterion when our objective is parameter estimation.

C2 **Ratio of standard error and standard derivation**:

$$R_{2j} = Rank_j\left(\sum_{i=1}^{I} Ratio_{ij}\right),$$

where

$$\begin{aligned}
Ratio_{ij} &= Rank_j\left(\prod_{p=1}^{P}\frac{\max(SD_{ijp}, SE_{ijp})}{\min(SD_{ijp}, SE_{ijp})}\right), \\
SD_{ijp} &= \left\{\frac{1}{K-1}\sum_{k=1}^{K}\left[\widehat{\beta}_{ijkp} - \frac{1}{K}\left(\sum_{k=1}^{K}\widehat{\beta}_{ijkp}\right)\right]^2\right\}^{\frac{1}{2}} \text{ and} \\
SE_{ijp} &= \frac{1}{K}\sum_{k=1}^{K} SE(\widehat{\beta}_{ijkp}).
\end{aligned}$$

Note that $SE(\widehat{\beta}_{ijkp})$ is the standard error of β_{ijkp}. The ratio $Ratio_{ij}$ between SD_{ijp} and SE_{ijp} measures the relative magnitude between the standard error (SE) and the standard derivation (SD) based on β_{ijkp}. We expect $Ratio_{ij}$ to be close to 1 (SD close to SE) if the models fit the data well.

C3 **Average of AIC**:

$$R_{3j} = Rank_j \left(\sum_{i=1}^{I} AIC_{ij} \right),$$

where

$$AIC_{ij} = \ Rank_j \left(\frac{1}{K} \sum_{k=1}^{K} AIC_{ijk} \right)$$

and AIC_{ijk} is the AIC value in the k-th replicated data set when the fitting involved the i-th true covariance structure and the j-th fitted structure. This is a useful criterion when our objective is data fitting.

An overall measure of the goodness-of-fit of models adopting the J structures based on these three criteria $R_{hj},\ h = 1, 2, 3$ are

$$R_j = Rank_j \left(\sum_{h=1}^{3} R_{hj} \right).$$

4. Simulation studies

4.1 Blood glucose data

Data of inter and intra individual variation of blood glucose levels (Andrews and Herzberg 1985, P.211) obtained from registrants for pre-natal care at Boston City Hospital, USA, are used to estimate the variation of blood glucose for women on the pregnant and non-pregnant states. The data contain two parts. There are 53 women in non-pregnant state. Each of them undertook an annual glucose tolerance test over a period of six years. In each year, a fasting blood glucose test and an one-hour-post-fasting blood glucose test were conducted. There are also 52 women in pregnant state, each having three fasting blood glucose tests and three one-hour-post-fasting blood glucose tests. Outcomes are differences in blood glucose concentration during fasting and one hour post fasting. Measurements are in mg/100 ml. Variables in the analyses include

Dependent variable:
Y: the difference between their fasting and one hour post blood glucose level,

Independent variables:
X_1: the fasting blood glucose level,
X_2: the indicator of pregnancy.

1. **Data of pregnant and non-pregnant states women**

   Data from both pregnant and non-pregnant states women are used in the analysis of short time series. It contains 315 ($3 \times (53 + 52)$) observations coming from $M = 105$ subjects each repeatedly measured $N = 3$ times. For the 53 non-pregnant women, only the first three fasting and one-hour-post-fasting blood glucose levels are used in the analysis. The overall means of Y and X_1 are respectively -21.72 and 76.08 whereas the corresponding means for the pregnant group are -34.94 and 72.88 and for the non-pregnant group are -8.74 and 79.21. The generalized regression model has $P_b = 2$. We use the PROC MIXED in SAS to fit the generalized regression models with 7 different covariance matrices. Table 2 shows that all parameter estimates are significant and the best model is EC according to AIC. By using the parameter estimates in Table 2 as true values for each of the $J = 7$ covariance structures, we simulate $K = 200$ data sets (or totally 1400 data sets) each having 105 trivariate normal vectors or 315 observations. Then we fit each of the 1400 data sets to $J = 7$ covariance structures. The fittings are done automatically using a SAS macro program. Moreover, in order to study the interaction effect of the two factors: the length of the time series and the strength of the serial correlation on the choice of robust covariance structure, we repeat the whole simulation experiment with an adjusted set of true values of which $\boldsymbol{\beta}'$ remain the same but the level of serial correlation is higher: ρ' in AR1 and EC and ρ_1' in TOEP and TOEP2 equal to 0.7, $\sigma_{21}' = 0.7\sigma_2\sigma_1$ and $\sigma_{32}' = 0.7\dfrac{\rho_{32}}{\rho_{21}}\sigma_3\sigma_2$ in UN2 and $\rho_2' = 0.7\dfrac{\rho_2}{\rho_1}$ in TOEP. Obviously, models fitting to data simulated from the same covariance structure are generally the best according to AIC. On the other hand, Table 3 reports the sum of ranks ($\sum_{i=1}^{7} RMSE_{ij}$, $\sum_{i=1}^{7} Ratio_{ij}$, $\sum_{i=1}^{7} AIC_{ij}$) and the ranks R_{hj}, $h = 1, 2, 3$ respectively. Across levels of ρ (0.36 and 0.7 in AR1 say), the ranks R_{2j} according to ratio of SD and SE are identical, the ranks R_{1j} according to MSE are similar but the ranks R_{3j} according to AIC are quite different. Table 3 also reports the overall rank R_j for each structure according to the three criteria. The choices are not the same across levels of ρ but AR1 generally performs well according to both MSE and AIC. Hence AR1 is a good choice of robust covariance

structure for both data fitting and parameter estimation.

Table 2: Parameter estimates for models with different covariance structures for the blood glucose data ($N = 3$ with an original low level and an adjusted high level of ρ)

Type	Estimates (S.E.)	Variance	Covariance/correlation (adjusted)	AIC
SIM	$\beta_0 = -26.1112$ (9.75) $\beta_1 = 0.2192$ (0.12) $\beta_2 = -24.8136$ (2.54)	$\sigma^2 = 460.2936$		2821.0
EC	$\beta_0 = -38.3411$ (9.53) $\beta_1 = 0.3737$ (0.12) $\beta_2 = -23.8374$ (3.24)	$\sigma^2 = 464.6207$	$\rho = 0.3445$ (0.7)	2791.3
UN1	$\beta_0 = -25.3785$ (9.83) $\beta_1 = 0.2100$ (0.12) $\beta_2 = -24.6771$ (2.53)	$\sigma_1^2 = 508.6234$ $\sigma_2^2 = 429.6811$ $\sigma_3^2 = 442.6674$		2824.2
UN2	$\beta_0 = -31.7091$ (9.66) $\beta_1 = 0.2866$ (0.12) $\beta_2 = -24.2122$ (2.88)	$\sigma_1^2 = 506.3417$ $\sigma_2^2 = 416.4010$ $\sigma_3^2 = 446.1486$	$\sigma_{21} = 175.2339$ (321.4221) $\sigma_{32} = 74.3218$ (136.3245)	2794.5
AR1	$\beta_0 = -34.6805$ (9.22) $\beta_1 = 0.3247$ (0.12) $\beta_2 = -24.4292$ (3.11)	$\sigma^2 = 464.4223$	$\rho = 0.3636$ (0.7)	2795.2
TOEP	$\beta_0 = -37.7984$ (9.55) $\beta_1 = 0.3662$ (0.12) $\beta_2 = -23.9514$ (3.24)	$\sigma^2 = 464.8787$	$\rho_1 = 0.3651$ (0.7) $\rho_2 = 0.3060$ (0.5612)	2792.9
TOEP2	$\beta_0 = -31.2468$ (9.66) $\beta_1 = 0.2811$ (0.12) $\beta_2 = -24.6771$ (2.89)	$\sigma^2 = 455.5407$	$\rho_1 = 0.2802$ (0.7)	2801.3

2. **Data of non-pregnant state women**

This data set containing times series of medium length ($N = 6$) is obtained by including all blood glucose tests from the six fasting and one-hour-post-fasting from 53 non-pregnant women. There are 318 (6×53) observations coming from $M = 53$ subjects each repeatedly measured $N = 6$ times. The dependent variable is again Y and the only independent variable is X_1. The means for Y and X_1 over 318 observations are -14.39 and 79.72 respectively. Table 4 reports that parameter estimates for models with $J = 7$ different covariance matrices are all significant and the best model is TOEP according to AIC. Similarly by setting the parameter estimates as true values for $I = 7$ different covariance matrices, we simulate $K = 200$ data sets (or totally 1400 data sets) each having 53 normal vectors of $N = 6$

in length. Next, we fit each of the 1400 data sets to models with $J = 7$ different covariance matrices. From Table 5, the ranks R_{1j} and R_{3j} are similar to those of the previous data with $N = 3$ and low level of ρ, according to MSE and AIC respectively but they are quite different according to ratio of SD and SE. Moreover the first and second choices of robust covariance structure are identical to those of the previous data. Again, AR1 generally performs well according to both MSE and AIC.

Table 3: Ranks of the three criteria for the blood glucose data ($N = 3$)

$RMSE_{ij}$	SIM	EC	UN1	UN2	AR1	TOEP	TOEP2
Original low level of ρ							
$\sum_{i=1}^{7} RMSE_{ij}$	88.5	68.5	105.5	92.5	60	82	63
R_{1j}	5	3	7	6	1	4	2
$\sum_{i=1}^{7} Ratio_{ij}$	42	25	39	22	24	17	27
R_{2j}	7	4	6	2	3	1	5
$\sum_{i=1}^{7} AIC_{ij}$	32	19	45	34	19	27	20
R_{3j}	5	1.5	7	6	1.5	4	3
$\sum_{h=1}^{3} R_{hj}$	17	8.5	20	14	5.5	9	10
R_j	6	2	7	5	1	3	4
Adjusted high level of ρ							
$\sum_{i=1}^{7} RMSE_{ij}$	100	74	134	95	57	54	71.5
R_{1j}	6	4	7	5	2	1	3
$\sum_{i=1}^{7} Ratio_{ij}$	40	27	38	20	24	14	33
R_{2j}	7	4	6	2	3	1	5
$\sum_{i=1}^{7} AIC_{ij}$	34	24	44	19	24	24	27
R_{3j}	6	3	7	1	3	3	5
$\sum_{h=1}^{3} R_{hj}$	19	11	20	8	8	5	13
R_j	6	4	7	2.5	2.5	1	5

Table 4: Parameter estimates for models with different covariance structures for the blood glucose data ($N = 6$ with an original low level of ρ)

Type	Estimates (S.E.)	Variance	Covariance/correlation	AIC
SIM	$\beta_0 = -24.2516$ (9.68)	$\sigma^2 = 436.73$		2835.9
	$\beta_1 = 0.1237$ (0.12)			
EC	$\beta_0 = -36.1006$ (9.68)	$\sigma^2 = 316.63$	$\rho = 0.3920$	2798.9
	$\beta_1 = 0.2723$ (0.12)			
UN1	$\beta_0 = -24.3246$ (9.17)	$\sigma_1^2 = 293.64$		2840.7
	$\beta_1 = 0.1354$ (0.11)	$\sigma_4^2 = 432.82$		
		$\sigma_2^2 = 443.97$		
		$\sigma_5^2 = 529.53$		
		$\sigma_3^2 = 402.06$		
		$\sigma_6^2 = 521.74$		
UN2	$\beta_0 = -35.9661$ (8.83)	$\sigma_1^2 = 289.00$	$\sigma_{21} = 133.93$	2820.7
	$\beta_1 = 0.2860$ (0.11)	$\sigma_4^2 = 443.72$	$\sigma_{54} = 92.10$	
	$\sigma_2^2 = 426.97$	$\sigma_{32} = 29.33$		
	$\sigma_5^2 = 499.70$	$\sigma_{65} = 213.88$		
	$\sigma_3^2 = 402.85$	$\sigma_{43} = 125.23$		
	$\sigma_6^2 = 533.08$			
AR1	$\beta_0 = -40.7659$ (9.45)	$\sigma^2 = 440.42$	$\rho = 0.3766$	2799.5
	$\beta_1 = 0.3313$ (0.12)			
TOEP	$\beta_0 = -40.8435$ (9.50)	$\sigma^2 = 445.22$	$\rho_1 = 0.3911$	2797.1
	$\beta_1 = 0.3662$ (0.12)		$\rho_2 = 0.3179$	
			$\rho_3 = 0.2618$	
			$\rho_4 = 0.1627$	
			$\rho_5 = 0.0777$	
TOEP2	$\beta_0 = -36.6278$ (9.52)	$\sigma^2 = 427.91$	$\rho_1 = 0.2668$	2810.2
	$\beta_1 = 0.2791$ (0.12)			

Table 5: Ranks of the three criteria for the blood glucose data ($N = 6$)

$RMSE_{ij}$	SIM	EC	UN1	UN2	AR1	TOEP	TOEP2
	Original low level of ρ						
$\sum_{i=1}^{7} RMSE_{ij}$	74.5	53	77	62.5	35	53	37
R_{1j}	6	3.5	7	5	1	3.5	2
$\sum_{i=1}^{7} Ratio_{ij}$	29	19	33	36	29	21	29
R_{2j}	4	1	6	7	4	2	4
$\sum_{i=1}^{7} AIC_{ij}$	32	21	40	35	20	27	21
R_{3j}	5	2.5	7	6	1	4	2.5
$\sum_{h=1}^{3} Rank_{hj}$	15	7	20	18	6	9.5	8.5
R_j	5	2	7	6	1	4	3

4.2 Plasma citrate concentration data

An experiment involving 10 subjects (Andrews and Herzberg, 1985, P.237) was carried out to study the variation of plasma citrate concentration during a day. The concentration of citrate in plasma (in μmol per litre) of each subject was measured hourly from 8am to 9pm ($N = 14$) during a day. Meals were given at 8am, at noon and at 5pm. The mean concentration over the complete set of

Table 6: Parameter estimates for models with different covariance structures for the citrate concentration data ($N = 14$ with an original low level and an adjusted high level of ρ)

Type	Estimates (S.E.)	Variance	Covariance/correlation (adjusted)	AIC
SIM	$\beta_0 = 134.23$ (4.23)	$\sigma^2 = 402.43$		1228.2
	$\beta_1 = -10.086$ (4.31)			
	$\beta_2 = -1.6959$ (0.44)			
EC	$\beta_0 = 134.23$ (5.59)	$\sigma^2 = 382.61$	$\rho = 0.6247$ (0.3)	1148.6
	$\beta_1 = -10.086$ (2.90)			
	$\beta_2 = -1.6959$ (0.30)			
UN1	$\beta_0 = 131.94$ (4.09)	$\sigma_1^2 = 408.21$		2840.7
	$\beta_1 = -10.663$ (3.98)	$\sigma_2^2 = 536.82$		
	$\beta_2 = -1.4439$ (0.41)	$\sigma_3^2 = 681.54$		
		$\sigma_4^2 = 606.36$		
		$\sigma_5^2 = 635.54$		
		$\sigma_6^2 = 226.64$		
		$\sigma_7^2 = 258.51$		
		$\sigma_8^2 = 356.57$		
		$\sigma_9^2 = 422.88$		
		$\sigma_{10}^2 = 221.07$		
		$\sigma_{11}^2 = 464.94$		
		$\sigma_{12}^2 = 388.46$		
		$\sigma_{13}^2 = 197.22$		
		$\sigma_{14}^2 = 235.81$		
UN2	$\beta_0 = 130.78$ (4.70)	$\sigma_1^2 = 417.73$	$\sigma_{2,1} = 251.31$ (114.99)	1173.3
	$\beta_1 = -14.295$ (0.60)	$\sigma_2^2 = 351.73$	$\sigma_{3,2} = 358.28$ (163.94)	
	$\beta_2 = -1.1267$ (0.48)	$\sigma_3^2 = 777.91$	$\sigma_{4,3} = 186.51$ (85.34)	
		$\sigma_4^2 = 538.59$	$\sigma_{5,4} = 417.47$ (191.02)	
		$\sigma_5^2 = 696.06$	$\sigma_{6,5} = 57.97$ (26.52)	
		$\sigma_6^2 = 287.85$	$\sigma_{7,6} = 160.18$ (73.29)	
		$\sigma_7^2 = 174.78$	$\sigma_{8,7} = 133.39$ (61.04)	
		$\sigma_8^2 = 332.27$	$\sigma_{9,8} = 95.64(43.76)$	
		$\sigma_9^2 = 402.57$	$\sigma_{10,9} = 198.88$ (91.00)	
		$\sigma_{10}^2 = 247.51$	$\sigma_{11,10} = -90.06$ (-41.21)	
		$\sigma_{11}^2 = 677.45$	$\sigma_{12,11} = 375.17$ (163.43)	
		$\sigma_{12}^2 = 361.59$	$\sigma_{13,12} = -6.68$ (-3.06)	
		$\sigma_{13}^2 = 169.84$	$\sigma_{14,13} = 55.69$ (25.48)	
		$\sigma_{14}^2 = 533.08$		

Table 6 (continued): Parameter estimates for models with different covariance structures for the citrate concentration data ($N = 14$ with an original low level and an adjusted high level of ρ)

Type	Estimates (S.E.)	Variance	Covariance/correlation (adjusted)	AIC
AR1	$\beta_0 =$ 130.88 (6.33) $\beta_1 = -7.661$ (2.36) $\beta_2 = -1.4208$ (0.68)	$\sigma^2 = 407.80$	$\rho = 0.7026$ (0.3)	1142.4
TOEP	$\beta_0 =$ 136.22(6.13) $\beta_1 = -7.971$(2.40) $\beta_2 = -1.8164$(0.48)	$\sigma^2 = 474.44$	$\rho_1 = 0.7449$ (0.3) $\rho_2 = 0.6441$ (0.2594) $\rho_3 = 0.5815$ (0.2342) $\rho_4 = 0.5259$ (0.2118) $\rho_5 = 0.5714$ (0.2301) $\rho_6 = 0.6036$ (0.2431) $\rho_7 = 0.6194$ (0.2495) $\rho_{\ 8} = 0.6302$ (0.2538) $\rho_{\ 9} = 0.6405$ (0.2579) $\rho_{10} = 0.5049$ (0.2034) $\rho_{11} = 0.3984$ (0.1605) $\rho_{12} = 0.2726$ (0.1098) $\rho_{13} = 0.0899$ (0.0362)	1136.5
TOEP2	$\beta_0 =$ 131.87 (4.53) $\beta_1 = -6.915$ (2.83) $\beta_2 = -1.5244$ (0.51)	$\sigma^2 = 349.46$	$\rho_1 = 0.4230$ (0.3)	1176.9

140 observations is 119.35 μmol per litre whereas it is 115.10 μmol per litre over the 30 meal times.

The regression models with 2 covariates, namely indicator of meal time (Meal) and time from 1 to 14 (Time), and using each of the $J = 7$ covariance structures are fitted to the data. Table 6 shows that all parameter estimates are significant and the best model is again TOEP according to AIC. Using the $I = 7$ sets of parameter estimates as true values, we simulate $K = 200$ data sets (or totally 1400 data sets) each having 10 normal vectors of $N = 14$ in length. Lastly we fit each of the 1400 data sets to models with $J = 7$ different covariance structures.

Again, we repeat the whole simulation experiment with an adjusted set of true values of which $\boldsymbol{\beta}'$ remain the same but the level of serial correlation is lower: ρ' in AR1 and EC and ρ'_1 in TOEP and TOEP2 equal to 0.3 and σ'_{ij} in UN2 and ρ'_i in TOEP are similarly calculated as in the blood glucose data when $N = 3$. From Table 7, across levels of ρ (0.7 and 0.7 in AR1 say), the ranks R_{1j} and R_{2j} are very similar according to MSE and ratio of SD and SE but are quite different according to AIC. The table also reports the first and second choices of robust covariance structure according to the total rank R_j using all the three criteria.

Across levels of ρ, the choices are not the same but SIM and EC generally perform well according to MSE. Hence when our objective is parameter estimation, SIM and EC are good choices of robust covariance structure.

Table 7: Ranks of the three criteria for the citrate concentration data ($N = 14$)

	Original low level of ρ						
$\sum_{i=1}^{7} RMSE_{ij}$	57.5	57.5	86	135	74	107	71
R_{1j}	1.5	1.5	5	7	4	6	3
$\sum_{i=1}^{7} Ratio_{ij}$	29	21	41	45	16	23	21
R_{2j}	4	2.5	6	7	1	5	2.5
$\sum_{i=1}^{7} AIC_{ij}$	31	20	44	40	17.5	25	18.5
R_{3j}	5	3	7	6	1	4	2
$\sum_{h=1}^{3} Rank_{hj}$	10.5	7	18	20	6	15	7.5
R_j	4	2	6	7	1	5	3
	Adjusted low level of ρ						
$\sum_{i=1}^{8} RMSE_{ij}$	51	51	89	147	63	120	67
R_{1j}	1.5	1.5	5	7	3	6	4
$\sum_{i=1}^{8} Ratio_{ij}$	23	14	38	49	17	35	20
R_{2j}	4	1	6	7	2	5	3
$\sum_{i=1}^{8} AIC_{ij}$	38	15	27	19	43	45	9
R_{3j}	5	2	4	3	6	7	1
$\sum_{h=1}^{3} R_{hj}$	10.5	4.5	15	17	11	18	8
R_j	3	1	5	6	4	7	2

5. Results

The first two simulation experiments using the blood glucose data consider time series of short and medium length and of weaker serial correlation ($\rho = 0.36$ and 0.38 in AR1 for $N = 3$ and 6 respectively) whereas the third simulation experiment using plasma citrate concentration data considers longer time series of stronger serial correlation ($N = 14$ and $\rho = 0.70$ in AR1). In order to study the interaction effect of the length of the time series and the strength of the serial correlation on the choice of robust covariance structure, two more simulation studies have been done with level of ρ in AR1 set to 0.7 and 0.35 for the two data sets containing time series of $N = 3$ and $N = 14$ respectively. Table 8 summarizes the first and second choices of robust covariance structure according to all the three criteria including MSE, ratio of SD to SE and AIC in a 2×2 contingency table. For data of relatively shorter time series and with weaker serial correlation, AR1 is the most robust structure and EC is also good. Both structures assume a constant variance and a consistent correlation between pairs of lag-k observations,

given by ρ^k and ρ respectively. Models adopting the two covariance structures describe well the gradual decline of serial correlation across 'lag'. For data with stronger serial correlation, TOEP and AR1 (or UN2) rank the first and second respectively. Their lag-k correlations, ρ_k and ρ^k (or non-zero lag-1) respectively, drop to zero across 'lag' at a rate which depends more on the data. For data of relatively longer time series with stronger serial correlation, AR1 and EC are the first and second choices of robust covariance structure. They are exactly identical to those of shorter time series with weaker serial correlation. Their lag-k correlations, ρ^k and ρ respectively, are simple with less parameters and non-zero with constant (ρ) or consistent (ρ_k) correlations between pairs of observations. On the other hand, for data with weaker serial correlation, EC is the most robust structure and TOEP2 is also good. The two correlation structures model correlation between pairs of lag-k observations by a constant ρ and a non-zero lag-1 correlation ($\rho_1 \neq 0$) respectively. They describe the decline of correlation across 'lag'. Note also that SIM, UN1 and UN2 are usually worse especially for data containing longer time series or with stronger serial correlation because they assume non-constant variances as well as zero or non-constant covariances resulting in a large number of parameters. On the other hand, robust structures generally assume constant variances and consistent covariances between observations of a given time lag.

Table 8: Summary of choices for robust covariance structure cross-classified by the length of time series and strength of serial correlation

Serial correlatlon	Low ($\rho \approx 0.35$ in AR1)		High ($\rho \approx 0.70$ in AR1)	
Length of time series	First choice	Second choice	First choice	Second choice
Short ($N = 3$)	AR1	EC	TOEP	AR1 or UN2
Long ($N = 14$)	EC	TOEP2	AR1	EC

6. Conclusion

Results suggest that both the length of time series and the strength of serial correlation affect the choice of robust covariance structure. For data containing relatively shorter time series, AR1 is the choice of the most robust covariance structure irrespective of the strength of serial correlation. AR1 is simple with only two parameters, the constant variance σ^2 and the auto-regressive parameter ρ which describes the strength of serial correlation. Moreover, EC and TOEP are also good choices for data of weak and strong serial correlation respectively. AR1, EC and TOEP all have $P_v = P_c = 1$ but with different assumptions on

lag-k correlation, namely ρ^k, ρ and ρ_k respectively. For data of longer time series, modeling the covariance structure becomes more difficult because of the possibility of more complicate covariance structures. Simulation experiments show that EC is the most robust choice irrespective of the strength of serial correlation. Moreover, AR1 and TOEP2 are also good choices for data of strong and weak serial correlation respectively. The three chosen structures, namely EC, AR1 and TOEP2 are similar to those of short time series except that TOEP is replaced by TOEP2 with less parameters. In general, robust covariance structures for data containing longer time series have constant variances and covariances for observations of equal time lag and zero covariances for observations of higher time lag.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (Edited by B.N. Petrov and F. Csaki), 267-281. Akademiai Kiado.

Andrews, D. F. and Herzberg, A. M. (1985). *Data*. Springer-Verlag.

Barnard, J., McCulloch, R. and Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10** 1281-1311.

Chiu, T. Y. M., Leonard, T. and Tsui, K. (1996). The matrix-logarithm covariance model. *Journal of the American Statistical Association* **81**, 310-320.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1996). *Analysis of Longitudinal Data*. Oxford Science Publications.

Efron, B. and Morris, C. (1976). Multivariate empirical Bayes estimation of covariance matrices. *Annals of Statistics* **4**, 22-32.

Lindsey, J. (1993). *Models for Repeated Measurements*. Clarendon Press.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-554.

Yang, R. and Berger, J. (1994). Estimation of the covariance matrix using the reference prior. *Annals of Statistics* **22**, 1195-1211.

Received April 19, 2007; accepted June 18, 2007.

Jennifer S. K. Chan
School of Mathematics and Statistics
The University of Sydney
NSW 2006, Australia
jchan@maths.usyd.edu.au

Boris S.T. Choy
Department of Mathematical Sciences
University of Technology
Sydney, P.O. Box 123
Broadway, NSW 2007, Australia
boris.choy@uts.edu.au