Spring 2013

Analysis of Continuous Longitudinal Data with ARMA(1, 1) and Antedependence Correlation Structures

Sirisha Mushti
Old Dominion University

ANALYSIS OF CONTINUOUS LONGITUDINAL DATA WITH ARMA(1, 1) AND ANTEDEPENDENCE CORRELATION STRUCTURES

by

Sirisha Mushti
B.Sc. June 2004, Osmania University, India
M.Sc. June 2006, University of Hyderabad, India

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
MAY 2013

Approved by:

N. Rao Chaganty (Director)

Larry D. Lee (Member)

Nak-Kyeong Kim (Member)

Tina D. Cunningham (Member)

ABSTRACT

ANALYSIS OF CONTINUOUS LONGITUDINAL DATA WITH ARMA(1, 1) AND ANTEDEPENDENCE CORRELATION STRUCTURES

Sirisha Mushti
Old Dominion University, 2013
Director: Dr. N. Rao Chaganty

Longitudinal or repeated measure data are common in biomedical and clinical trials. These data are often collected on individuals at scheduled times resulting in dependent responses. Inference methods for studying the behavior of responses over time as well as methods to study the association with certain risk factors or covariates taking into account the dependencies are of great importance. In this research we focus our study on the analysis of continuous longitudinal data. To model the dependencies of the responses over time, we consider appropriate correlation structures generated by the stationary and non-stationary time-series models. We develop new estimation procedures depending on the correlation structures considered and compare those procedures with the existing methods.

The first part of this dissertation focuses on the robust correlation structure generated by the first-order autoregressive-moving average (ARMA(1, 1)) stationary time-series model. ARMA(1, 1) correlation structure is characterized by two correlation parameters and this correlation structure reduces to the AR(1), MA(1) and CS structures in special cases. Although standard efficient procedures are preferable to estimate the correlation parameters, there are computational challenges in implementing them. To overcome these challenges we employ an alternative estimation procedure based on pairwise likelihoods. A typical advantage of this approach is that the inference procedure does not involve complex computations and it results in a closed form expressions for the estimators of the correlation parameters. We show that the estimates obtained using the pairwise likelihood method for ARMA(1, 1) correlation structure are highly efficient asymptotically when compared to that of maximum likelihood.

The second part of the dissertation studies correlation structures generated by non-stationary time-series model known as antedependence models of first order. These correlation structures are capable of modeling the non-constant correlations between the same-lagged observations. Note that while this correlation structure has been extensively studied in the case of heterogeneous variance, we model homogenous variance and use a recent and new method known as quasi-least squares to estimate the correlation parameters. A major advantage of the quasi-least squares method is that it yields closed form expressions for the estimators of correlation parameters unlike the maximum likelihood method. We provide the asymptotic and small-sample properties of these estimators and compare their performance using relative efficiencies.

Dedicated to my parents for their unending love and support
and to my advisor Dr. N. Rao Chaganty.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my mentor and dissertation advisor Professor N. Rao Chaganty for his continuous support and guidance. I am grateful for the opportunity to work as his student. His guidance, continuous support and patience were tremendous in pursuing this research and completing this dissertation.

I would like to thank professors Dr. Larry Lee, Dr. Nak-kyeong Kim and Dr. Tina Cunningham for serving on my dissertation committee. I am very thankful for their thoughtful reviews in refining this dissertation. Special thanks are due to Dr. Larry Lee for an in-depth review of my dissertation and numerous suggestions. I would like to express my gratitude to the Late Dr. Dayanand Naik for his enormous support during my graduate studies. Thanks are also due to Dr. N. Rao Chaganty for arranging research support at the Old Dominion University Department of Mathematics and Statistics and at the Eastern Virginia Medical School. Teaching the statistics undergraduate courses at Old Dominion University and the statistical consulting projects at the Eastern Virginia Medical School were a rewarding and learning experience.

I would also like to thank professors, Dr. Norou Diawara, Dr. Micheal Doviak and Dr. Richard Noren, for the knowledge gained through their insightful teachings during my graduate studies. Also, thanks are due to the staff members in the Department of Mathematics and Statistics at Old Dominion University, and the Department Chair, Dr. Mark Dorrepaal, for providing a pleasant and stimulating environment. I also wish to acknowledge the staff members Barbara Jeffrey, Miriam Venable and Sheila Hegwood for their help during my graduate studies.

This entire dissertation would not have been possible without the help of many individuals, my fellow students, friends and experiences in my life. Their collective support, guidance and inspiration made this possible. I would like to thank each one of them for making this journey joyful and memorable. Special thanks are due to Dr. Raghavendra Kurada for his continuous support, inspiration and numerous statistical discussions we had during the last four years.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

Chapter

1. INTRODUCTION 1
   1.1 LITERATURE REVIEW 1
   1.2 MODEL SPECIFICATIONS 2
   1.3 CORRELATION STRUCTURES 4
   1.4 ESTIMATION PROCEDURES 8
   1.5 ASYMPTOTIC THEORY 12
   1.6 DATA EXAMPLES 14
   1.7 OUTLINE OF THE THESIS 17

2. PAIRWISE LIKELIHOOD METHOD 19
   2.1 INTRODUCTION 19
   2.2 PROPERTIES OF CORRELATION STRUCTURES 20
   2.3 PARAMETER ESTIMATION 26
   2.4 ASYMPTOTIC VARIANCES 32
   2.5 ASYMPTOTIC COMPARISONS 39
   2.6 ANALYSIS OF REAL DATA 46

3. ANALYSIS OF ANTEDEPENDENCE MODELS 49
   3.1 ANTEDEPENDENCE CORRELATION STRUCTURES 50
   3.2 MOTIVATING EXAMPLE 53
   3.3 ESTIMATION PROCEDURES 55

3.4 ASYMPTOTIC THEORY 59
3.5 ASYMPTOTIC EFFICIENCY COMPARISONS 64
3.6 SMALL SAMPLE EFFICIENCIES 65
3.7 ANALYSIS OF CATTLE GROWTH DATA 68

4. SUMMARY 77

REFERENCES 78

APPENDICES

A. SOME FORMULAS AND DERIVATIVES 80
A.1 CLOSED FORM ROOT OF A CUBIC POLYNOMIAL 80
A.2 MATRIX DERIVATES 80
A.3 DERIVATIVES FOR ANTEDEPENDENCE CORRELATION MATRIX 81

B. R PROGRAMS 83
B.1 PAIRWISE LIKELIHOOD ESTIMATORS 83
B.2 QUASI-LEAST SQUARES AND MAXIMUM LIKELIHOOD ESTIMATORS 91

VITA 99

LIST OF TABLES

Table		Page
1	Ozone Changes Data	15
2	Oxygen Saturation Study	16
3	Weights of Cattle from Growth Study	17
4	Asymptotic Relative Efficiency of γ for ARMA(1, 1)	41
5	Asymptotic Relative Efficiency of ρ for ARMA(1, 1)	43
6	Asymptotic Relative Efficiency of γ for MA(1) and CS Structures with $t = 4$ and $\phi = 3$	44
7	Parameter Estimates for the Ozone Data	47
8	Parameter Estimates for the Oxygen Saturation Data	48
9	Sample Correlation Matrix for Cattle Growth Data	54
10	Asymptotic Relative Efficiency of ρ_1 for Antedependence	70
11	Asymptotic Relative Efficiency of ρ_2 for Antedependence	71
12	Asymptotic Relative Efficiency of ϕ for Antedependence	72
13	Asymptotic Relative Efficiency of λ and ϕ for Antedependence when $t = 4$	73
14	Small-Sample Efficiency of ρ_1 and ρ_2 for Antedependence	74
15	Small-Sample Efficiency of ρ_1, ρ_2 and ρ_3 for Antedependence	75
16	Parameter Estimates for the Cattle Growth Data	76

LIST OF FIGURES

Figure Page

1 Positive definite range for 3×3 case 22

2 Positive definite range for different values of t. 23

(a) $t = 3,\ t = 4$ 23

(b) $t = 5,\ t = 7$ 23

(c) $t = 9,\ t = 12$ 23

3 ARE of γ for $t = 4$ and $\phi = 3$ 42

4 ARE of ρ for $t = 4$ and $\phi = 3$ 45

5 ARE of γ in MA(1) for $t = 4$ and $\phi = 3$ 45

6 ARE of γ in CS for $t = 4$ and $\phi = 3$ 46

7 ARE of ρ_1 for $t = 3$ and $\phi = 3$. 65

8 ARE of ρ_2 for $t = 3$ and $\phi = 3$. 66

9 ARE of ϕ for $t = 3$. 67

10 ρ_1 RE for QLS and MLE for $t = 3$ 68

11 ρ_2 RE for QLS and MLE for $t = 3$ 69

CHAPTER 1

INTRODUCTION

1.1 LITERATURE REVIEW

Longitudinal data occur frequently in many scientific studies including clinical trials, psychology and public health studies. In these studies, data are typically collected by following individuals over a period of time. Valid inference methods for longitudinal data are of great importance in scientific research. In longitudinal studies, data is collected on the variables of interest on individuals at scheduled times. The analysis in longitudinal studies usually focuses on how the variables change over time and how they are associated with certain risk factors or covariates. Various statistical models and methods for longitudinal data analysis have been developed over the past few decades. Davis (2002) provides a comprehensive introduction to a wide variety of statistical methods for the analysis of repeated measurements. Others authors Liang and Zeger (1986) and Diggle et al. (2002) discussed further the statistical analysis methods for continuous and discrete data.

As stated in Davis (2002), key strength of longitudinal studies is that this is the only type of design in which it is possible to obtain information concerning individual patterns of change. This type of design also economizes on the number of subjects. However, there are difficulties associated with analyzing such kind of data. For instance, the analysis becomes complicated by the natural dependence among the responses observed on the same subject, and in some instances the response from a subject may be missing, resulting in an unbalanced or partially incomplete data, which further complicates the analysis. Many approaches have been studied to address these complications.

Weiss (2005) discusses the importance of the choice of correlation structures to model the dependency among the responses observed on the same subject. In

This dissertation follows the style of *Journal of the American Statistical Association* particular Weiss (2005) elaborates on the techniques to choose an appropriate model or generate additional correlation models. Choosing the correct correlation model is crucial in longitudinal studies since it has an impact on the precision of regression parameter estimator. The regression coefficients measure effect of the covariates on the responses, which is the main objective in any statistical analysis. Hence, several methods were suggested to estimate the correlation parameters accurately in the literature.

Recent works in the field of longitudinal studies concentrate on developing robust estimation procedures for estimating the association parameters. Some of these works use the composite likelihood method as an alternative to standard likelihood-based inference to overcome the difficulties caused by high-dimensional interdependencies. For example, Kuk and Nott (2000), Varin and Vidonib (2006) and Zhao and Joe (2005) discuss the idea of composite likelihood and some modifications in order to model different types of data.

1.2 MODEL SPECIFICATIONS

In this section we introduce some basic notation to facilitate the discussion in this dissertation. Recall that longitudinal data consists of responses or measurements taken at different time points on several independent subjects in a study. Let y_{ij} denote the response observed on subject i observed at time point j for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, t_i$. We represent the response vector on subject i as a t_i dimensional vector $Y_i = (y_{i1}, y_{i2}, \ldots, y_{it_i})'$. Suppose $x_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp})'$ be the $p \times 1$ vector of covariates observed along with y_{ij}. For convenience we restrict our attention to the balanced data case, that is, we assume $t_i = t$ for all i. Extension of our results to the unbalanced case is straightforward. In addition we also assume that the responses y_{ij} are continuous but the covariates x_{ijk} could be either discrete or continuous. Let $X_i = (x_{i1}, x_{i2}, \ldots, x_{it})'$ be the $t \times p$ matrix of covariates observed on subject i. The main interest in longitudinal studies is to study the relationship between response Y_i and covariates X_i by taking into consideration the within subject dependency among the responses.

Specification of a multivariate family of probability density functions $f(Y_i|X_i;\theta)$, indexed by the parameter θ often involves modeling the marginal distribution

$f(y_{ij}|x_{ij};\theta)$. A natural and convenient set of candidate models for the marginal distribution is the exponential family given by

$$f(y_{ij}|x_{ij};\theta) = \exp\left[\{y_{ij}\nu_{ij} - b(\nu_{ij})\}/a(\phi)\right]\, h(y_{ij};\phi),$$

where $a(.)$, $b(.)$ and $h(.)$ are some known functions, ϕ is a constant that quantifies over-dispersion. Here ν_{ij} is a function of the parameter θ and it is known as the canonical parameter. It is easy to check that $E(y_{ij}) = \mu_{ij} = b'(\nu_{ij})$ and $Var(y_{ij}) = a(\phi)b''(\nu_{ij})$. In generalized linear models we assume that the mean μ_{ij} is a function of the covariates given by

$$g(\mu_{ij}) = x'_{ij}\,\beta,$$

where $g(.)$ is a known monotone and differentiable function, commonly known as the link function since it links the mean with the covariates. Here $\beta = (\beta_1, \ldots, \beta_p)'$ is an unknown $p \times 1$ vector of regression coefficients. When $g(\cdot)$ equals the inverse of $b'(\cdot)$, it is known as the canonical link function and it satisfies

$$\nu_{ij} \;=\; (b')^{-1}(\mu_{ij}) = g(\mu_{ij}) = x'_{ij}\,\beta. \tag{1}$$

If we assume that y_{ij}'s $(1 \leq j \leq t)$ are independent then the multivariate density $f(Y_i|X_i;\theta)$ is simply

$$f(Y_i|X_i;\theta) = \prod_{j=1}^{t} f(y_{ij}|x_{ij};\theta).$$

However, the independent assumption is invalid for longitudinal setting since the observations y_{ij} are dependent with respect to j for each value of i. Therefore, various types of multivariate distributions are proposed to feature different association structures present in longitudinal data. For instance, in this dissertation, we assume that the responses are continuous and hence multivariate normal distribution can be employed to construct the joint distribution. Furthermore, under the assumption of continuous responses the canonical link function in equation (1) can be taken as the identity function. Thus, we model $E(Y_i) = \mu_i = X_i'\beta$ and $Cov(Y_i) = \phi R(\lambda)$ where λ is a vector of parameters that determine the correlation matrix and ϕ is a scale parameter that does not change with j and i. Hence, the parameter vector in our model is $\theta = (\beta', \phi, \lambda')'$.

In any statistical analysis, estimation of the regression coefficient β plays a vital role since it quantifies the effect of the covariates on the responses. However, to obtain accurate estimates for regression coefficients it also becomes equally important that we estimate the correlation parameters λ accurately and efficiently. There are several estimation methods proposed in literature such as method of moments, maximum likelihood method and many other robust estimation procedures to estimate the correlation parameters and to calculate their standard errors. However, before studying the estimation procedures, emphasis is laid on practical aspects of choosing an appropriate correlation structure. It is desirable to choose the correlation model that fits the data best since the use of the correct or optimal correlation model increases the efficiency of the regression estimator. However, increase in the number of parameters of the correlation structure corresponds to increase in the number of estimating functions. Hence, choosing a suitable and parsimonious correlation structure is equally important. In the following section we describe few such potential correlation structures.

1.3 CORRELATION STRUCTURES

The selection of the correlation structure to model the complex dependency among the correlated responses observed on the same subject plays a vital role. In the literature, a large number of correlation models have been proposed as well as methods to generate additional covariance models have been studied. However, in this research we focus on the popular parameterized correlation models which are generated by the time series models. A parameterized correlation matrix $R(\lambda)$ is one where all correlations in the matrix are functions of a parameter λ whose dimension q is usually small.

The primary class of correlation structures we consider in this dissertation are autoregressive-moving average of first order, autoregressive of order one, moving average of first order and compound symmetry structure. We also consider a special case of non-stationary correlation models known as antedependence correlation structures. We discuss the motivation and properties of each correlation structure. We also study the positive definite ranges of the parameters involved particularly for the first-order autoregressive-moving average correlation structure.

1.3.1 AUTOREGRESSIVE MOVING AVERAGE STRUCTURE

A general structure for the correlation matrix considered for analyzing real world longitudinal data is the first-order autoregressive-moving average or ARMA(1, 1), which is a generalization of the first-order autoregressive (AR(1)), moving average (MA(1)) and compound symmetry (CS) structures. The ARMA(1, 1) correlation matrix of dimension t is given by

$$R(\lambda) = R(\gamma, \rho) = \begin{pmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 & \cdots & \gamma\rho^{t-2} \\ \gamma & 1 & \gamma & \gamma\rho & \cdots & \gamma\rho^{t-3} \\ \gamma\rho & \gamma & 1 & \gamma & \cdots & \gamma\rho^{t-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma\rho^{t-2} & \gamma\rho^{t-3} & \gamma\rho^{t-4} & \gamma\rho^{t-5} & \cdots & 1 \end{pmatrix}, \qquad (2)$$

where $\lambda = (\gamma, \rho)$. We can see that this correlation structure is characterized by two parameters $\lambda_1 = \gamma$ and $\lambda_2 = \rho$. The first parameter γ is known as the lag one correlation because

$$\text{Corr}(y_{ij}, y_{i(j+1)}) = \gamma,$$

whereas the second parameter ρ is the additional decrease in correlation for each additional lag. Thus, lag k correlation is given by

$$\text{Corr}(y_{ij}, y_{i\,(j+k)}) = \gamma\rho^{k-1}.$$

The rate of decrease of lag k correlation is directly proportional to k, and it depends on the value of ρ. The ARMA(1, 1) structure is appropriate for longitudinal data that exhibit this correlation pattern.

1.3.2 AUTOREGRESSIVE, MOVING AVERAGE AND COMPOUND SYMMETRY STRUCTURES

As discussed in Section 1.3.1 the AR(1), MA(1) and CS structures are special cases of the autoregressive-moving average correlation structure. Unlike the ARMA(1, 1) correlation structure, these three correlation structures are determined by only one correlation parameter and hence they are more parsimonious correlation models. The idea of reducing the correlation models to a simpler form depends on how we want to characterize the correlations between the repeated measurements or in other words which pattern better fits the data. These correlation structures were studied extensively in the literature and hence in this dissertation we discuss briefly their theoretical properties.

First-order autoregressive or AR(1) correlation structure often is an adequate model for longitudinal data. This correlation model arises from first-order Markov process and is extensively studied in time series analysis, see Fuller (1996). In case of t repeated measurements, the AR(1) correlation matrix is given by

$$R(\lambda) = R(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{t-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{t-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{t-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \rho^{t-4} & \cdots & 1 \end{pmatrix}. \tag{3}$$

In the AR(1) model the correlation between two observations y_{ij} and $y_{i(j+k)}$ depends on the absolute value of the time difference between them, that is,

$$\text{Corr}(y_{ij}, y_{i(j+k)}) = \rho^k \text{ for } k > 1.$$

Notice that the correlation decreases exponentially with the time lag. The farther apart two observations are, the lower is the correlation between them. In longitudinal studies, it is common for correlations to diminish as the lag between the time points increases and AR(1) structure serves as a potential candidate model to account for the dependency.

Another simpler correlation structure is the first-order moving average in which each response is assumed to be correlated only with its succeeding and preceding responses. The MA(1) correlation matrix is given by

$$R(\lambda) = R(\rho) = \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{4}$$

Here ρ characterizes the correlations between two neighboring responses, that is,

$$\text{Corr}(y_{ij}, y_{i(j\pm1)}) = \rho$$

and any two responses separated by lag more than one are uncorrelated. Compound symmetry (CS) correlation structure is another correlation model used commonly in clustered data studies, and occasionally in longitudinal studies. The compound symmetry correlation matrix has the following form

$$R(\lambda) = R(\rho) = \begin{pmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{pmatrix}. \tag{5}$$

Here $\mathrm{Corr}(y_{ij}, y_{ik}) = \rho$ for all $j \neq k$. Even though this structure is employed in some longitudinal studies, it is most appropriate for clustered data where there is no reason to believe that some pairs of observations have strong or weak correlations than other pairs.

1.3.3 ANTEDEPENDENCE STRUCTURE

In Sections 1.3.1 and 1.3.2 we discussed the commonly used correlation models for analyzing longitudinal data. These are stationary models because the correlations only depend on the lag rather than the time point at which the measurements were taken. However, there are instances where the correlations vary according to the time point and vary even for two pairs of observations with the same lag. Such dependencies are captured by the antedependence correlation models. A first-order antedependence correlation structure accommodating different correlations between observations lagged same distance apart, is given by

$$R(\lambda) = R(\rho_1 \rho_2 \cdots \rho_{t\text{-}1}) = \begin{pmatrix} 1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2\rho_3 & \cdots & \prod_{j=1}^{t-1}\rho_j \\ \rho_1 & 1 & \rho_2 & \rho_2\rho_3 & \cdots & \prod_{j=2}^{t-1}\rho_j \\ \rho_1\rho_2 & \rho_2 & 1 & \rho_3 & \cdots & \prod_{j=3}^{t-1}\rho_j \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \prod_{j=1}^{t-1}\rho_j & \prod_{j=2}^{t-1}\rho_j & \prod_{j=3}^{t-1}\rho_j & \prod_{j=4}^{t-1}\rho_j & \cdots & 1 \end{pmatrix}. \tag{6}$$

This t dimensional antedependence structure is characterized by $t-1$ correlation parameters. Although parsimony is being compromised compared to the stationary correlation models, it is essential to consider these models to analyze the varying correlations between equally lagged observations. A detailed description of the antedependence structure and its corresponding time series concepts will be discussed in Chapter 3.

1.4 ESTIMATION PROCEDURES

Methods for estimating the regression and correlation parameters have been well researched in the literature for longitudinal data with AR(1), MA(1) and CS correlation models. However, methods of estimating the parameters of the ARMA(1, 1) correlation structure in the context of longitudinal data are scanty. Thus, focusing on the ARMA(1, 1) correlation structure we will first study parameter estimation in a repeated measures setting in Section 1.2. Later, we extend the estimation methods to other correlation structures defined in Section 1.3.2 and study robust properties of the methods.

Under the normality assumption the maximum likelihood estimate of the regression parameter and the scale parameter are in a closed form. Besides the maximum likelihood estimates are well known to be optimal. The maximum likelihood method is the standard estimation procedure, nevertheless, applying this method to estimate the correlation parameter λ involves intense computations in most cases. Therefore we employ a new algorithm for the estimation of λ depending on the correlation structure considered. The alternative methods discussed in this dissertation are pairwise likelihood method for ARMA(1, 1) and quasi-least squares method for first-order antedependence correlation structure.

Pairwise likelihood method is a two-stage approach which first maximizes the complete likelihood of Y_i to obtain the estimate of the regression coefficient β and the scale parameter ϕ. The second stage consists of maximizing likelihood composed of two-dimensional densities, resulting in the estimator of λ. As a result of the first stage, the functional form of the estimates of β and ϕ turn out to be the same as that of maximum likelihood since both methods maximize the same likelihood function.

The pairwise likelihood method is discussed elaborately in Chapter 2 along with the mathematical details involved. In this dissertation, we also study the asymptotic efficiency of the pairwise likelihood method with respect to the optimal maximum likelihood method.

The quasi-least squares method is also a well known two-step estimation procedure developed to overcome the computational shortcomings of the maximum likelihood. This method, based on the principle of generalized least squares, is developed as a distribution-free estimation procedure. Detailed discussion of the quasi-least squares method and its performance compared to maximum likelihood method is given in Chapter 3.

1.4.1 MAXIMUM LIKELIHOOD ESTIMATION

Assume that Y_i is distributed as a t dimensional multivariate normal with mean $X_i\beta$ and covariance $\mathrm{Cov}(Y_i) = \phi R(\lambda)$ for $i = 1, 2, \ldots, n$, where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_q)$, is a q-dimensional vector of unknown correlation parameters. Further assume that Y_i, $1 \leq i \leq n$ are independent. Let $\theta = (\beta', \phi, \lambda')'$. The likelihood function is given by

$$\begin{aligned} L_n(\theta|Y_1, \ldots, Y_n) &= \prod_{i=1}^{n} (2\pi)^{-\frac{t}{2}} \left|\phi R(\lambda)\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\phi} (Y_i - X_i\beta)' R^{-1}(\lambda) (Y_i - X_i\beta)\right\} \\ &= (2\pi)^{-\frac{nt}{2}} \phi^{-\frac{nt}{2}} \left|R(\lambda)\right|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\phi} \mathrm{tr}\left(R^{-1}(\lambda) Z_n\right)\right\}, \end{aligned} \tag{7}$$

where $Z_n = \sum_{i=1}^{n} (Y_i - X_i\beta)\ (Y_i - X_i\beta)'$. The loglikelihood function is

$$\begin{aligned} \ell_n(\theta) &= \log L_n(\theta|Y_1, \ldots, Y_n) \\ &= \mathrm{const} - \frac{nt}{2} \log(\phi) - \frac{n}{2} \log\left|R(\lambda)\right| - \frac{1}{2\phi} \mathrm{tr}\left(R^{-1}(\lambda)\, Z_n\right). \end{aligned} \tag{8}$$

To find the maximum likelihood estimate of θ we equate to zero the partial derivatives of (8) with respect to β and ϕ. Thus, estimators $\widehat{\beta}$ and $\widehat{\phi}$ are given by

$$\begin{aligned} \widehat{\beta} &= \left(\sum_{i=1}^{n} X_i' R^{-1}(\widehat{\lambda}) X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' R^{-1}(\widehat{\lambda}) Y_i\right), \\ \widehat{\phi} &= \frac{1}{nt} \sum_{i=1}^{n} \left(Y_i - X_i\widehat{\beta}\right)' R^{-1}(\widehat{\lambda}) \left(Y_i - X_i\widehat{\beta}\right), \end{aligned} \tag{9}$$

where $\widehat{\lambda}$ is an estimate of λ. We obtain the estimates of the correlation parameters by equating to zero the partial derivative of (8) with respect to each λ_j for $j = 1, 2, \ldots, q$. This results in the following score equations

$$-\frac{n}{2}\text{tr}\left(R^{-1}(\lambda)\frac{\partial R(\lambda)}{\partial \lambda_j}\right) + \frac{1}{2\phi}\,\text{tr}\left(R^{-1}(\lambda)\frac{\partial R(\lambda)}{\partial \lambda_j}R^{-1}(\lambda)\, Z_n\right) = 0, \tag{10}$$

for $1 \leq j \leq q$. For different parameterized correlation matrices $R(\lambda)$ we can obtain simplified expressions for tr $\left(R^{-1}(\lambda)\frac{\partial R(\lambda)}{\partial \lambda_j}\right)$ and $\left(R^{-1}(\lambda)\frac{\partial R(\lambda)}{\partial \lambda_j}R^{-1}(\lambda)\right)$. Further details about the maximum likelihood estimation for different correlation structures are given in Section 2.3.1 of Chapter 2, along with an overview of the computational complexities involved.

In the literature there are other estimation procedures such as restricted maximum likelihood method, where θ is estimated by maximizing the following modified loglikelihood function

$$\begin{aligned} \log RL_n(\theta) = \text{const} - \frac{nt}{2}\log(\phi) - \frac{n}{2}\log|R(\lambda)| - \frac{n}{2}\log|X'R^{-1}(\lambda)X| \\ - \frac{1}{2\phi}\text{tr}\left(R^{-1}(\lambda)\, Z_n\right). \end{aligned} \tag{11}$$

The function (11) is similar to (8) except that it involves an extra term $-\frac{n}{2}\log|X'R^{-1}(\lambda)X|$. We do not pursue the restricted maximum likelihood in this dissertation.

1.4.2 PAIRWISE LIKELIHOOD METHOD

Pairwise likelihood method is a special case of composite likelihood method which has been studied extensively in the context of discrete longitudinal data. Composite likelihood method was initiated by Besag (1977) and further developed by Lindsay (1988) and many other authors including Zhao and Joe (2005). Composite likelihood provides a useful inference alternative for full likelihood based inference. The inference function is derived by multiplying a collection of component likelihoods, whether or not they are independent; the particular collection used is often determined by the context. The inference function obtained by this method retains some properties of the likelihood from a complete model.

The simplest composite marginal likelihood is the inference function constructed under independence assumption of the marginals. This inference function permits inference only on marginal parameters. If parameters related to association are also of interest it is necessary to model pairs of observations as in the pairwise likelihood, see Varin and Vidonib (2006). The inference function in this case is

$$\log PL_n(\theta) = \sum_{i=1}^{n} \sum_{j<j'=1}^{t} \log f(y_{ij}, y_{ij'}|x_{ij}, x_{ij'}; \theta).$$

Thus, instead of specifying the full distribution, we only need to specify a partial structure for the distribution $f(Y_i|X_i;\theta)$. This eases the issues related to complex modeling and the advantages become more obvious when the dimension of Y_i increases.

Construction of the above likelihood and solving for the estimates is the second stage estimation in the pairwise likelihood method. To obtain the estimates of the parameters we derive the estimating equations with respect to each of the correlation parameters. In addition to that, for continuous responses, based on the correlation structures the above likelihood is being modified accordingly with inference focused on the dependence parameters. The pairwise likelihood estimation method is explained in detail in the later chapters depending on the correlation model that is being considered.

1.4.3 QUASI-LEAST SQUARES METHOD

The quasi-least squares method developed by Chaganty (1997), Shults and Chaganty (1998) and Chaganty and Shults (1999), is an extension of the method of generalized least squares. This method uses the quasi-score or quasi-loglikelihood function (12) to obtain the estimates of correlation parameters. For the regression parameter estimates, this method uses the same functional estimates as the maximum likelihood method. However the method differs from the maximum likelihood in the way the correlation parameters are estimated. The advantage of this method is that it results in closed form expressions for the correlation parameter estimates for many commonly used correlation structures. Chaganty (2003) establishes the consistency and efficiency of the correlation parameters.

The quasi-loglikelihood function used in the quasi-least squares method, under the assumption of homogeneous variance, is given by,

$$Q(\theta) = \sum_{i=1}^{n} (Y_i - X_i\beta)' R^{-1}(\lambda) (Y_i - X_i\beta) = \text{tr}\left(R^{-1}(\lambda) Z_n\right) \quad (12)$$

A minimization of (12) with respect to λ is done to obtain the first step estimate $\widetilde{\lambda}$, which is the solution of the equation

$$\frac{\partial Q(\theta)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left(\text{tr}\left(R^{-1}(\lambda) \widetilde{Z}_n\right)\right) = \text{tr}\left(\frac{\partial R^{-1}(\lambda)}{\partial \lambda} \widetilde{Z}_n\right) = 0, \quad (13)$$

where $\widetilde{Z}_n$ is Z_n evaluated at the estimate $\widetilde{\beta}$ of β.

It is shown in Chaganty and Shults (1999) that $\widetilde{\lambda}$ is a biased estimate of λ due to the fact that the expectation of (13) evaluated at $\widetilde{\lambda}$ in not zero, that is,

$$E\left[\text{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \lambda} \widetilde{Z}_n\right)\right] = \text{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \lambda} E(\widetilde{Z}_n)\right) = \phi\text{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \lambda} R(\lambda)\right) \neq 0.$$

Thus, we try to adjust the bias by equating the above expression, to zero and solve for λ for fixed value of $\widetilde{\lambda}$. The drawbacks associated with considering the original expression with ϕ is discussed in Chaganty and Shults (1999). Quasi-least squares estimation method overcomes these drawbacks by considering only the trace term. Thus at the second stage of the quasi-least squares method, we solve the equation

$$\text{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \lambda} R(\lambda)\right) = 0, \quad (14)$$

to get the estimate $\widehat{\lambda}$, which is asymptotically unbiased. Thus, to obtain the final quasi-least squares estimates we iterate between (9), (13) and (14) until convergence. We then use $\widehat{\lambda}$ to obtain the final estimates of $\widehat{\beta}$ and $\widehat{\phi}$. Chaganty (2003) discusses the estimation of correlation parameters for patterned correlation matrices and studied their asymptotic distributions.

1.5 ASYMPTOTIC THEORY

In this section we briefly discuss the statistical properties of the estimates derived using the maximum likelihood, pairwise likelihood and the quasi-least sqaures method. In the case of n independent observations $Y_1, \ldots, Y_n$ from the model

$f(Y_i|X_i;\theta)$ and $n \to \infty$ with t fixed, some standard asymptotic results are summarized below.

1.5.1 MAXIMUM LIKELIHOOD METHOD

The asymptotic theory for the maximum likelihood estimators is well known. Under the normality assumption, the asymptotic covariance matrix for the ML estimators can be derived using the loglikelihood function given in equation (8). For a sample of n observations, the Fisher information matrix $\mathcal{I}_\ell(\theta)$ is found by taking the negative expectation of the second derivative of the loglikelihood function $\ell_n(\theta)$ with respect to θ. Then, under the regularity conditions, $\mathcal{I}_\ell(\theta)$ is given by

$$\mathcal{I}_\ell(\theta) = -E\left(\frac{\partial^2 \ell_n(\theta)}{\partial\theta\partial\theta'}\right).$$

According to Cramer's theorem, the maximum likelihood estimate $\widehat{\theta}_{\mathrm{ML}}$ of θ is approximately distributed as multivariate normal with mean θ and covariance matrix $\mathcal{I}_\ell^{-1}(\theta)$, for large n.

1.5.2 PAIRWISE LIKELIHOOD AND QUASI-LEAST SQUARES METHODS

The theory of unbiased estimating equations can be used to derive the asymptotic variances and covariances for the pairwise likelihood and quasi-least squares estimators. For these alternative methods discussed in this dissertation, we make use of the following theorem given in Joe (1997), to obtain the asymptotic covariance matrix of the parameter estimates.

Suppose $h_i(\theta)$ is an unbiased equation for θ, that is, $E(h_i(\theta)) = 0$. Let $\widehat{\theta}_{alt}$ be the root of the equation $\sum_{i=1}^n h_i(\theta) = 0$. Under the regularity conditions, for t fixed and large n, we have $\widehat{\theta}_{alt}$ is approximately distributed as multivariate normal with mean θ and covariance matrix $\mathcal{G}^{-1}(\theta)/n$, where $\mathcal{G}(\theta)$ is known as the Godambe information matrix given by $\mathcal{G}(\theta) = D(\theta)\ M^{-1}(\theta)\ (D(\theta))'$, where

$$D(\theta) = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_i(\theta)}{\partial\theta}\right) \quad \text{and} \quad M(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(h_i(\theta)).$$

This result due to Godambe, is actually a generalization of the Cramer's theorem.

The estimating equations in the pairwise likelihood method are obtained by taking the derivative of the pairwise likelihoods constructed as a linear combination of the scores associated with each loglikelihood term. Thus the large sample properties of the pairwise likelihood estimator, $\widehat{\theta}_{PL}$, follow from the above stated theory of unbiased estimating equations. Similarly, by considering the unbiased estimating equations used to obtain the quasi-least squares estimates, we can establish the corresponding asymptotic theory. Particular cases of the above two methods are discussed in detail in Sections 2.4.2 and 3.4.2 respectively.

1.6 DATA EXAMPLES

We will use three motivational real data examples to illustrate the methods in this dissertation. These data examples exhibit the dependency patterns similar to the correlation patterns described in Section 1.3.

1.6.1 OZONE DATA

The first example is the Ozone data. In recent years, the amount of ozone in the atmosphere has decreased which results in a direct exposure to the harmful sun rays and its one of the reasons why ozone is being looked at a lot by scientists who study climate and changes in earth systems. Hence, a study of the ozone in the atmosphere helps in understanding the change in the pattern of the levels.

This data set records ozone over a three-day period during late July 1987 at 20 locations (ANAH, AZUS,...,WSLA) in and around Los Angeles, California, USA. Five hourly recordings during the peak hours of the day are recorded starting from 1pm to 5pm giving us a set of $20 \times 6 \times 3$ ozone readings. Measurement units are in parts per hundred million. The data has $60 = 20 \times 3$ records with 5 longitudinal measures each. Table 1 shows a subset of the Ozone data.

The objective in this study is to assess the effect of day on ozone readings. The repeated response variable is the ozone readings observed on each day between 1pm to 5pm at an hourly interval. We consider day as the covariate having three levels(Day-1, Day-2 and Day-3) and the data is balanced.

Table 1. Ozone Changes Data

Site	Day	Hour 13	14	15	16	17
ANAH	1	10	9	7	6	6
	2	8	8	7	8	6
	3	10	12	11	9	7
AZUS	1	13	13	13	10	8
	2	18	17	12	13	10
	3	19	24	21	16	13
BURK	1	8	9	8	7	5
	2	15	14	11	7	4
	3	16	17	15	10	6
CELA	1	8	8	7	6	5
	2	8	9	8	6	4
	3	8	10	11	8	5
CLAR	1	14.7	16.9	17.3	16.8	13.3
	2	18.7	23.9	22.9	18.8	15.8
	3	19.4	25.2	28.6	26.3	17.6
FONT	1	12	16	18	16	13
	2	15	19	22	19	16
	3	11	17	14	23	21
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
WSLA	1	8	6	7	6	5
	2	7	7	5	5	5
	3	9	9	7	6	5

1.6.2 OXYGEN SATURATION DATA

The second example is the oxygen saturation data. This data is the result of a study to check the effectiveness of three different methods of suctioning an endotracheal tube. The first is standard suctioning, and the second is a new method using a special vacuum, and the third is manual bagging of the patient while suctioning. The outcome is the oxygen saturation levels measured at five time points: baseline, first suctioning pass, second suctioning pass, third suctioning pass, and 5 mins post suctioning. The covariate is method of suctioning which has three levels. Twenty-five

ICU patients were randomized to each of the three methods. Table 2 displays the oxygen saturation data.

Table 2. Oxygen Saturation Study

		Time				
ID	Method	s1	s2	s3	s4	s5
1	1	95	96	94	97	95
2	1	94	94	92	93	95
3	1	94	93	92	91	93
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	1	98	100	100	100	96
25	1	91	90	92	92	92
1	2	94	95	95	95	94
2	2	96	96	95	95	94
3	2	92	94	93	94	92
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	2	95	96	97	96	94
25	2	94	93	94	94	94
1	3	92	97	98	97	91
2	3	96	99	97	99	99
3	3	94	96	96	98	96
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	3	95	94	97	98	97
25	3	95	93	92	93	95

1.6.3 CATTLE GROWTH DATA

The third data example considered in this dissertation is the Cattle growth data Kenward (1987). In this experiment, cattle were given two treatments, labeled A and B, to treat intestinal parasites. The response, weights (in Kgs) of the cattle, were recorded 11 times over a 133-day period with the first 10 measurements on each cow recorded at two-week intervals and the last measurements recorded one-week after the tenth measurement. Thirty cattle were randomized to each of the two treatments. Thus, here we have a longitudinal data measured at 11 time points on each subject(cow), with treatment as the only covariate with two-levels. The objective is to study the effects of the treatment on the cattle weight loss. Table 3 displays the partial data for Treatment-A and Treatment-B.

Table 3. Weights of Cattle from Growth Study

		Week										
Treatment	Cow	0	2	4	6	8	10	12	14	16	18	19
A	1	233	224	245	258	271	287	287	287	290	293	297
	2	231	238	260	273	290	300	311	313	317	321	326
	3	232	237	245	265	285	298	304	319	317	334	329
	4	239	246	268	288	308	309	327	324	327	336	341
	5	215	216	239	264	282	299	307	321	328	332	337
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	29	233	241	252	273	301	316	332	336	339	348	345
	30	221	219	231	251	270	272	287	294	292	292	299
B	1	210	215	230	244	259	266	277	292	292	290	264
	2	230	240	258	277	277	293	300	323	327	340	343
	3	226	233	248	277	297	313	322	340	354	365	362
	4	233	239	253	277	292	310	318	333	336	353	338
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	29	221	232	251	274	284	295	300	323	319	333	322
	30	233	238	254	266	282	294	295	310	320	327	326

1.7 OUTLINE OF THE THESIS

This dissertation is organized as follows. In Chapter 2, we study the ARMA(1, 1), AR(1), MA(1) and compound symmetry correlation structures. We provide a brief summary of the time series model which generates the ARMA(1, 1), AR(1),MA(1) and compound symmetry correlation structures and discuss the positive definite range of the parameters characterizing each of the four correlation structures. We introduce the alternative approach known as pairwise likelihood method to estimate the correlation parameters incase of each correlation structure and define the estimation procedure in particular for each correlation structure. Next we discuss the asymptotic properties of the estimators obtained using pairwise likelihood method and compare these estimators with maximum likelihood estimators using the asymptotic relative efficiency. We analyze the ozone data in Table 1 and oxygen saturation data in Table 2 using the pairwise likelihood and maximum likelihood methods and compare the results. The R code that we developed is used to compute the pairwise likelihood estimates of the parameters in the data analysis.

In Chapter 3, we introduce first-order antedependence correlation structure. We provide, using a real life data given in Table 3, the appropriateness of using the first-order antedependence correlation structure to model the dependencies among the responses that exhibits a peculiar correlation patterns. Next, we study some of the properties of these first-order antedependence correlation structures. We discuss two methods of estimation procedures, maximum likelihood and quasi-least squares method to estimate the correlation parameters under the assumption of first-order antedependence correlation model. In this chapter, we also study the large-sample properties of the estimators obtained using each of these estimation methods and discuss the large-sample and small-sample efficiencies between the maximum likelihood and quasi-least squares estimators. Using these two estimation methods, we analyze the data introduced in Table 3.

In Chapter 4, we summarize the alternative approaches used in case of each correlation structure and the advantages associated with these methods compared to the standard estimating procedures in each case.

Finally, we end this dissertation with an Appendix that contains the important formulae used in deriving the parameter estimators and their corresponding asymptotic properties. We have also provided the matrix derivatives used in case of first-order antedependence correlation structure and R programs that compute the estimates and the asymptotic covariance matrices in case of ARMA(1, 1) and first-order antedependence correlation structures.

CHAPTER 2

PAIRWISE LIKELIHOOD METHOD

2.1 INTRODUCTION

In this chapter, we focus on the inference for parametric models where the likelihood function is difficult to evaluate due to complex dependencies involved in longitudinal studies. In these situations, alternative approaches based on modifications of the likelihood have been advocated by several authors. Well known examples include the pseudo-likelihood, partial-likelihoods and composite likelihoods and many more. A related idea is to use approximate likelihoods by compounding low-dimensional marginal densities, for example, on the univariate or bivariate marginal distributions. In this dissertation, the term pairwise likelihood method is proposed for this class of low-dimensional likelihoods. The pairwise likelihood is a special case of a more general class of composite likelihoods given in Lindsay (1988).

In light of the growing interest in solving different complex applications, we emphasize on the development of the theory of pairwise likelihood method which uses only pairwise joint distributions to construct the likelihoods. Later we outline the efficiency and robustness of pairwise likelihood method by discussing the asymptotic and small sample properties of the estimators obtained using pairwise likelihood method.

In the context of longitudinal data analysis, the robustness of pairwise likelihood method depends on the choice of correlation structure used to model the complex dependencies among the serially correlated responses of the same subject. Hence, the next section attempts to study the properties of the correlation structure and their impact on revising the pairwise likelihood method.

2.2 PROPERTIES OF CORRELATION STRUCTURES

In Section 1.3 we introduced several correlation structures that could be used for analyzing longitudinal data. In this section we study the properties of each of these correlation structures to an extent relevant for their applicability in the analysis.

2.2.1 AUTOREGRESSIVE MOVING AVERAGE OF ORDER (1,1) CORRELATION STRUCTURE

Recall that ARMA(1, 1) correlation structure given by equation (2), has two correlation parameters γ and ρ. The lag one correlation is quantified by γ, $\text{Corr}(y_{ij}, y_{i(j-1)}) = \gamma$, and ρ is the additional decrease in correlation for each additional lag. Thus the lag k correlation is given by $\text{Corr}(y_{ij}, y_{i(j-k)}) = \gamma\rho^{k-1}$.

Derivation of the ARMA(1, 1) Correlation structure

The correlation matrix (2) is generated by the autoregressive moving average time series model of order (1, 1). Time series models are studied to account for the fact that data taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for. Similarly, in longitudinal studies we are interested in studying the dependencies among the data collected over time, however, in these studies the data is observed only for a small fixed number of time points but observations are collected on large number of independent subjects.

ARMA(1, 1) time series model is defined as the sequence $\{\zeta_s : s \in (0, 1, 2, ...)\}$, satisfying the difference equation

$$\zeta_s - \rho\,\zeta_{s-1} = \epsilon_s + \varphi\epsilon_{s-1},$$

where ϵ_s is a sequence of uncorrelated random variables with mean zero and variance σ^2. For any $k \geq 1$, the auto correlation function is

$$\text{Corr}(\zeta_s, \zeta_{s+k}) = \frac{(\varphi+\rho)(1+\varphi\rho)}{(1+\varphi^2+2\varphi\rho)}\,\rho^{k-1} = \gamma\rho^{k-1}.$$

Note $\gamma = (\varphi+\rho)(1+\varphi\rho)/(1+\varphi^2+2\varphi\rho)$ is a reparametrization of the coefficients involved in the ARMA(1, 1) time series model.

Properties of ARMA(1, 1) Correlation Structure

Three interesting special cases arises from ARMA(1, 1) correlation structure $R(\gamma,\,\rho)$ depending on the value that ρ takes.

1. When $\rho = 1$, the matrix $R(\gamma,\,\rho)$ reduces to CS structure.
2. When $\rho = \gamma$, the matrix $R(\gamma,\,\rho)$ reduces to AR(1) structure.
3. When $\rho = 0$, the matrix $R(\gamma,\,\rho)$ has MA(1) structure.

Thus, the correlation matrix $R(\gamma,\,\rho)$ given in (2) encompasses the three most commonly used correlation matrices and therefore could be used as a robust model for the correlation between repeated measurements in longitudinal data analysis. It is of interest to know the ranges of the parameters γ and ρ such that the correlation matrix (2) is positive definite. We pursue this in the next section.

Positive Definite Range for ARMA(1, 1) Correlation Parameters

The positive definiteness of a correlation matrix imposes some restrictions on the ranges of the correlation parameters involved. Hence, the correlation parameters, γ and ρ, in the matrix $R(\gamma,\,\rho)$ also have restriction based on the dimension of the matrix. We show these restrictions for $t = 3$ case only, since it is cumbersome to obtain the ranges for general t. For $t = 3$, we have

$$R(\lambda) = R(\gamma, \rho) = \begin{pmatrix} 1 & \gamma & \gamma\rho \\ \gamma & 1 & \gamma \\ \gamma\rho & \gamma & 1 \end{pmatrix}, \tag{15}$$

where $\lambda = (\gamma,\,\rho)$. We want to investigate the ranges of the parameters γ and ρ such that the correlation matrix (15) is positive definite. In order for $R(\gamma,\,\rho)$ to be positive definite all of its leading principal minors should have positive determinants. The determinant of the first leading principal minor is $1-\gamma^2$, which is positive if and only if $-1 < \gamma < 1$. The determinant of the second leading principal minor equals $1 - 2\gamma^2 + 2\gamma^3\rho - \gamma^2\rho^2$. For fixed γ this is quadratic equation in ρ, whose roots are

$$\rho_1 = \frac{1}{\gamma} \quad \text{and} \quad \rho_2 = \frac{(2\gamma^2 - 1)}{\gamma} = 2\gamma - \frac{1}{\gamma}.$$

Note that $\rho_1 > \rho_2$ for $\gamma > 0$ and $\rho_1 < \rho_2$ for $\gamma < 0$. Thus $R(\lambda)$ is positive definite if and only if

$$\text{(i)} \quad 0 < \gamma < 1, \rho_2 < \rho < \rho_1 \quad \text{and} \quad \text{(ii)} \quad -1 < \gamma \leq 0, \quad \rho_1 < \rho < \rho_2. \qquad (16)$$

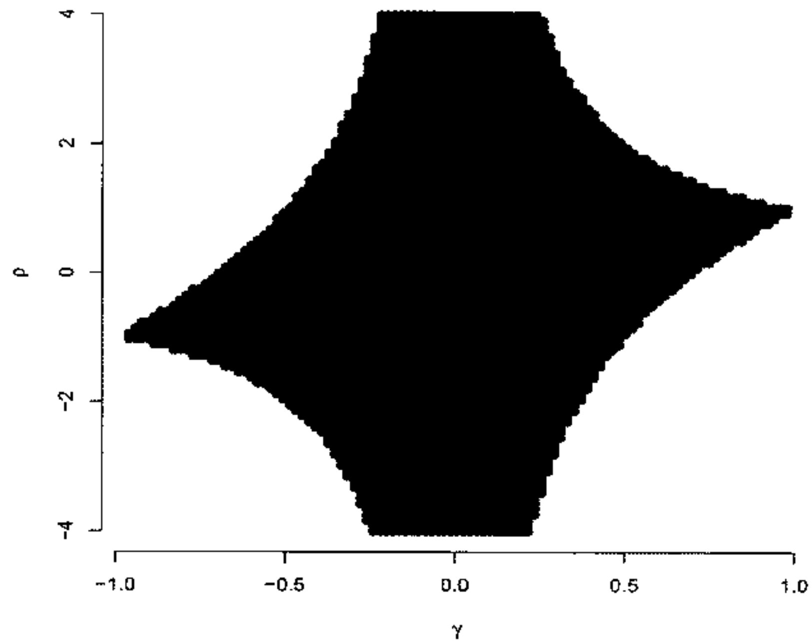
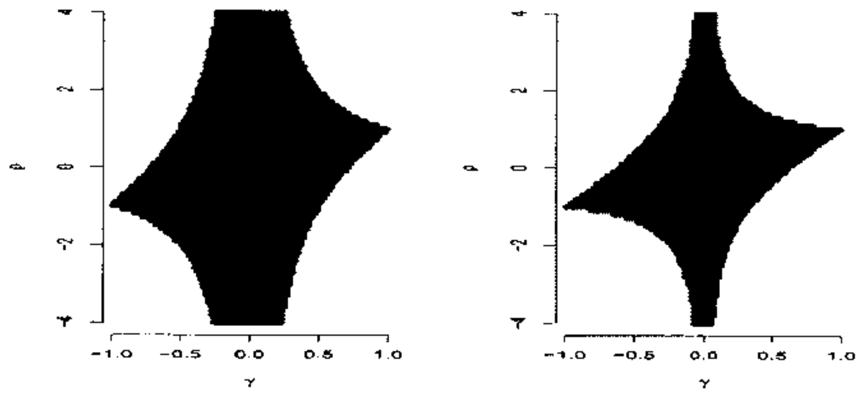


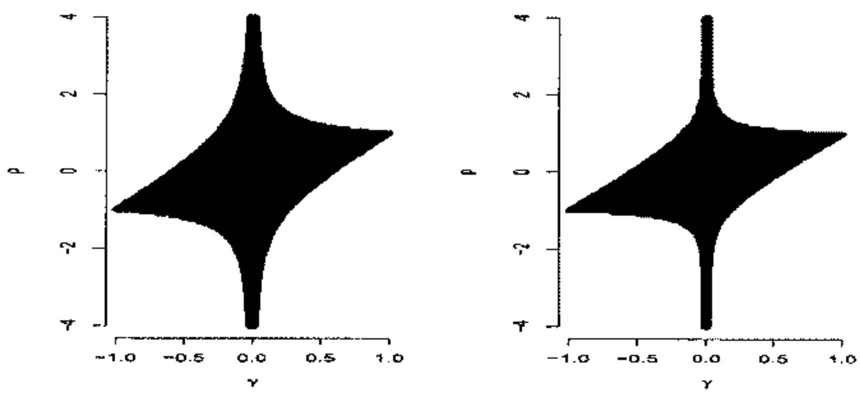
Figure 1. Positive definite range for 3×3 case

Figure 1 displays the graph of the positive definite ranges for the two parameters γ and ρ. Now, when $t = 4$, the correlation matrix is of the form

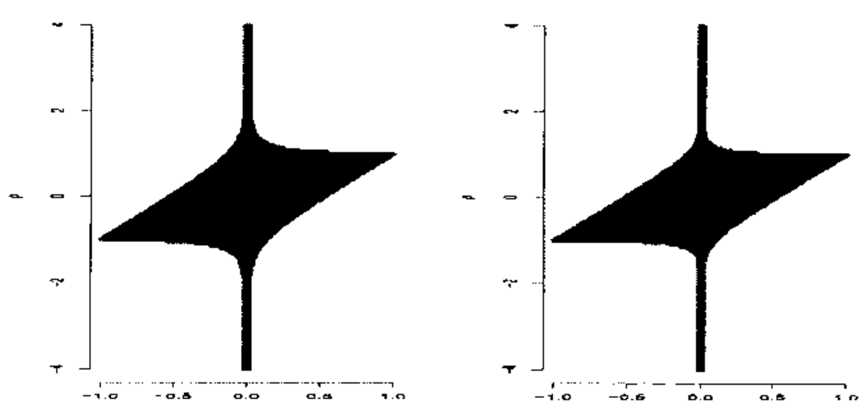
$$R(\lambda) = R(\gamma, \rho) = \begin{pmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{pmatrix}. \qquad (17)$$



(a) $t = 3,\ t = 4$



(b) $t = 5,\ t = 7$



(c) $t = 9,\ t = 12$

Figure 2. Positive definite range for different values of t.

The determinant of (17) is

$$|R(\lambda)| = -\gamma^3\rho^4 + 4\gamma^3\rho^3 - (2\gamma^2 + 4\gamma^4)\rho^2 + 4\gamma^3\rho + (1 - 3\gamma^2 + \gamma^4). \quad (18)$$

As we discussed earlier, $R(\lambda)$ is positive definite if and only if all of its leading principal minors have positive determinants. We have already given the restrictions imposed by the leading principal minors of the sub matrix of dimension three. So, $R(\lambda)$ in (17) is positive definite if the determinant (18) is positive together with the restrictions (16). Note that the (18) is a fourth order degree polynomial in ρ for fixed γ. Thus, it is difficult to analytically obtain the restrictions. Hence, using a trial and error numerical scheme, we investigated the positive definite region of γ and ρ graphically for arbitrary dimension t. Figure 2 displays the positive definite region for γ and ρ for several values of t.

The positive definite range for any general t requires all the principal sub-matrices of lower order to have positive determinants which results in narrower positive definite range for (γ, ρ) as t increases as shown in the Figure 2. However, from the graphs in Figure 2, it can be seen that the positive definite range does not become rapidly narrower after certain value of t.

In the context of time series, stationarity condition restricts ρ to be between -1 and 1, see (Fuller (1996)). This may seem to be contradicting with the presented positive definite range in Figure 1 since the positive definite range includes values greater than 1 for ρ. The reason can be attributed to the difference in the designs considered for a longitudinal and time series data. It is well-known that in time series a response variable for a single subject is observed on a large number of equally spaced time points whereas in longitudinal study, multiple responses from different subjects are observed at a fixed but varying and small number of time points. Hence, for large t in longitudinal study it is expected that the positive definite range for γ and ρ should satisfy the stationarity conditions. This conjecture can be clearly visualized from the Figure 2. For large values of t, we can see in Figure 2 that the positive definite range for ρ becomes smaller and tends to lie within the stationarity bounds of -1 and 1, except at the value of $\gamma = 0$. Since at $\gamma = 0$ the correlation matrix becomes identity matrix which is positive definite irrespective of the value of ρ. As mentioned earlier in this section, special cases of ARMA(1, 1) are AR(1), MA(1) and CS are well studied in the literature. For the sake of completeness, in the next section we provide some properties of these structures.

2.2.2 PROPERTIES OF AUTOREGRESSIVE, MOVING AVERAGE OF FIRST-ORDER AND COMPOUND SYMMETRY CORRELATION STRUCTURES

First-order autoregressive or AR(1) correlation structure depends on only one correlation parameter. In the AR(1) model, ρ defines correlation between two observations y_{ij} and $y_{i(j+k)}$ as $\text{Corr}(y_{ij}, y_{i(j+k)}) = \rho^k$ for $k \geq 1$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, t$. The AR(1) model exhibits an exponentially decreasing correlation pattern. The first-order moving average or MA(1) is also a one parameter correlation model. Here ρ characterizes correlation between successive responses $\text{Corr}(y_{ij}, y_{i(j\pm1)}) = \rho$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, t$. Compound symmetry correlation structure (CS) is another potential one parameter correlation model. Here the correlation between any two observations from the same subject is given by $\text{Corr}(y_{ij}, y_{ij'}) = \rho$ for all $i = 1, 2, \ldots, n$ and $j \neq j' = 1, 2, \ldots, t$.

Derivation of the AR(1), MA(1) Correlation Structures

In this section we define the time series models which generates the AR(1) and MA(1) correlation structures. The correlation matrix (3) is generated by the AR(1) time series model defined by $\zeta_s = \rho\zeta_{s-1} + \epsilon_s$, and the auto correlation function for $k \geq 1$ is

$$\text{Corr}(\zeta_s, \zeta_{s+k}) = \rho^k.$$

Similarly, the correlation matrix (4) is generated by the MA(1) time series model defined by the process $\zeta_s = \epsilon_s + \rho\epsilon_{s-1}$. The auto correlation function is

$$\text{Corr}(\zeta_s, \zeta_k) = \begin{cases} \rho & \text{if } k = s \pm 1, \\ 0 & \text{elsewhere} \end{cases}$$

where ϵ_s is a sequence of uncorrelated random variables with mean 0 and variance σ^2 in both the models.

Positive definite conditions for the AR(1), MA(1) and CS structures

The parameter ρ defines the correlation between the successive observations in case of AR(1) and MA(1) and it defines the correlation between any two observations in case of CS correlation structure. Hence, ρ satisfies the properties of the correlation coefficient and as a result it satisfies the condition $-1 < \rho < 1$ with further restrictions imposed depending on the correlation structure. The positive definite range for ρ in AR(1) correlation structure is $-1 < \rho < 1$. In theory ρ can be negative, but in longitudinal data studies negative correlations are rarely observed and hence we consider the positive definite range to be between $0 < \rho < 1$ only. Further, the positive definite ranges for ρ in case of CS structure is $-\dfrac{1}{t-1} < \rho < 1$ and in case of MA(1) structure, the positive definite range is $-1/(2\cos(\frac{\pi}{t+1})) < \rho < -1/(2\cos(\frac{t\pi}{t+1}))$. Once again, in all cases we assume $\rho > 0$.

As discussed in Section 1.2, we have longitudinal data on n subjects. The response vector is represented as $Y_i = (y_{i1}, y_{i2}, \ldots, y_{it})'$ with the corresponding covariate $X_i = (x_{i1}, x_{i2}, \ldots, x_{it})'$ for $i = 1, 2, \ldots, n$. The primary interest in longitudinal studies is to study the relationship between response Y_i and covariates X_i using the model $E(Y_i) = \mu_i = X_i\beta$ and $Cov(Y_i) = \phi R(\lambda)$. In the previous sections we have discussed few potential correlation models for $R(\lambda)$ generated by stationary time series models. The estimators for β and ϕ are given in (9). We next discuss methods of estimation for the correlation parameter λ.

2.3 PARAMETER ESTIMATION

In this section we discuss the different methods of estimating the parameters involved in the regression model. Although the regression coefficients are of primary interest in the statistical analysis, at this point we concentrate on the methods of estimating the correlation parameters. As discussed in Section 1.4 we study two different methods of estimating the correlation parameters, the maximum likelihood method and the pairwise likelihood method. We have proposed the pairwise likelihood method in order to overcome the complexities that arise due to the maximum likelihood method of estimation. A detailed discussion on each estimation method along with the difficulties associated with the maximum likelihood method are given in the following sections.

Recall, for each estimation procedure the following estimators for β and ϕ are used.

$$\widehat{\beta} = \left(\sum_{i=1}^{n} X_i' R^{-1}(\widehat{\lambda}) X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' R^{-1}(\widehat{\lambda}) Y_i\right),$$

$$\widehat{\phi} = \frac{1}{tn}\sum_{i=1}^{n}\left(Y_i - X_i\widehat{\beta}\right)' R^{-1}(\widehat{\lambda})\left(Y_i - X_i\widehat{\beta}\right)$$

where $\widehat{\lambda}$ is the vector of the correlation parameter estimators obtained using the two estimation procedures outlined next.

2.3.1 MAXIMUM LIKELIHOOD ESTIMATION

As discussed in Section 1.4.1, for any general t, the score equations for the correlation parameters λ_j reduces to

$$-\frac{n}{2}\text{tr}\left(R^{-1}\frac{\partial R}{\partial \lambda_j}\right) + \frac{1}{2\widehat{\phi}}\sum_{i=1}^{n}(Y_i - X_i\widehat{\beta})'\left(R^{-1}\frac{\partial R}{\partial \lambda_j}R^{-1}\right)(Y_i - X_i\widehat{\beta}) = 0 \qquad (19)$$

where expressions for $\widehat{\beta}$, $\widehat{\phi}$ are as defined in equation (9) and we wrote $R(\lambda) = R$ for convenience.

To study the complexities involved in solving equation (19), we consider the simple case of $t = 3$ with R as the ARMA(1, 1) correlation structure. In case of ARMA(1, 1) the dimension of λ is two with $\lambda = (\lambda_1, \lambda_2) = (\gamma, \rho)$. Hence the terms $\text{tr}\left(R^{-1}\frac{\partial R}{\partial \lambda_j}\right)$ and $R^{-1}\frac{\partial R}{\partial \lambda_j}R^{-1}$ in (19) for $(\lambda_1, \lambda_2) = (\gamma, \rho)$ can be expressed as below.

For $\lambda_1 = \gamma$,

$$\text{tr}\left(R^{-1}\frac{\partial R}{\partial \gamma}\right) = \frac{4\gamma}{(2\gamma^2 - \gamma\rho - 1)} + \frac{2\gamma\rho(\gamma - \rho)}{|R|}$$

and

$$R^{-1}\frac{\partial R}{\partial \gamma}R^{-1} = \begin{pmatrix} \frac{2\gamma(1-\gamma\rho)}{\omega|R|} + \frac{2\gamma\rho(1-\gamma^2)(\gamma-\rho)}{|R|^2} & \frac{2\gamma^2}{\omega^2} - \frac{(1-\gamma\rho)}{\omega|R|} & \frac{\gamma^2\rho(\gamma-\rho)^2+\rho(1-\gamma^2)^2}{|R|^2} \\ \frac{2\gamma^2}{\omega^2} - \frac{(1-\gamma\rho)}{\omega|R|} & \frac{-2\gamma^2\rho}{\omega^2} & \frac{2\gamma^2}{\omega^2} - \frac{(1-\gamma\rho)}{\omega|R|} \\ \frac{\gamma^2\rho(\gamma-\rho)^2+\rho(1-\gamma^2)^2}{|R|^2} & \frac{2\gamma^2}{\omega^2} - \frac{(1-\gamma\rho)}{\omega|R|} & \frac{2\gamma(1-\gamma\rho)}{\omega|R|} + \frac{2\gamma\rho(1-\gamma^2)(\gamma-\rho)}{|R|^2} \end{pmatrix},$$

where $\omega = (2\gamma^2 - \gamma\rho - 1)$.

For $\lambda_2 = \rho$ and ω as defined above, we have

$$\text{tr}\left(R^{-1}\frac{\partial R}{\partial \rho}\right) = \frac{2\gamma^2(\gamma - \rho)}{|R|}$$

and

$$R^{-1}\frac{\partial R}{\partial \rho}R^{-1} = \begin{pmatrix} \frac{2\gamma^2(1-\gamma^2)(\gamma-\rho)}{|R|^2} & \frac{\gamma^2(1-\gamma\rho)}{\omega|R|} & \frac{\gamma^3(\gamma-\rho)^2+\gamma(1-\gamma^2)^2}{|R|^2} \\ \frac{\gamma^2(1-\gamma\rho)}{\omega|R|} & \frac{2\gamma^3}{\omega^2} & \frac{\gamma^2(1-\gamma\rho)}{\omega^2} \\ \frac{\gamma^3(\gamma-\rho)^2+\gamma(1-\gamma^2)^2}{|R|^2} & \frac{\gamma^2(1-\gamma\rho)}{\omega|R|} & \frac{2\gamma^2(1-\gamma^2)(\gamma-\rho)}{|R|^2} \end{pmatrix}.$$

Note that when $t = 3$, inverse of $R(\lambda)$ is given by

$$R^{-1}(\lambda) = \begin{pmatrix} \frac{1-\gamma^2}{|R|} & \frac{\gamma}{2\gamma^2-\gamma\rho-1} & \frac{\gamma(\gamma-\rho)}{|R|} \\ \frac{\gamma}{2\gamma^2-\gamma\rho-1} & -\frac{(\gamma\rho+1)}{2\gamma^2-\gamma\rho-1} & \frac{\gamma}{2\gamma^2-\gamma\rho-1} \\ \frac{\gamma(\gamma-\rho)}{|R|} & \frac{\gamma}{2\gamma^2-\gamma\rho-1} & \frac{1-\gamma^2}{|R|} \end{pmatrix}.$$

Hence the score equations for γ and ρ in (19) reduces to

$$-\frac{n}{2}\left(\frac{4\gamma}{(2\gamma^2-\gamma\rho-1)} + \frac{2\gamma\rho(\gamma-\rho)}{|R|}\right) + \frac{1}{2\widehat{\phi}}\sum_{i=1}^{n}(y_i - X_i\widehat{\beta})'\,(R^{-1}\frac{\partial R}{\partial \gamma}R^{-1})\,(y_i - X_i\widehat{\beta}) = 0, \tag{20}$$

and

$$-\frac{n}{2}\left(\frac{2\gamma^2(\gamma-\rho)}{|R|}\right) + \frac{1}{2\widehat{\phi}}\sum_{i=1}^{n}(y_i - X_i\widehat{\beta})'\,(R^{-1}\frac{\partial R}{\partial \rho}R^{-1})\,(y_i - X_i\widehat{\beta}) = 0. \tag{21}$$

It is difficult to solve equations (20) and (21) simultaneously and obtain closed form expression for the estimates of γ and ρ. Alternatively, we can solve the two equations simultaneously using numerical techniques. Apart from computational intensities, these techniques can be highly sensitive to the choice of initial values. Additionally, solving the score equations (20) and (21) becomes computationally intensive as the number of repeated measures increases on each subject. Hence, we propose a more simpler and reliable procedure, the pairwise likelihood method, to come up with estimators for the correlation parameters. This method also relies on the normality assumption for the response vector Y_i similar to the maximum likelihood method.

2.3.2 PAIRWISE LIKELIHOOD METHOD

The motivation for the pairwise likelihood method was discussed in the earlier sections. In Section 1.4.2 we elucidated the method as a general case of composite likelihood method. As explained before, pairwise likelihood and maximum likelihood methods differ conceptually in terms of likelihood construction. In maximum likelihood method, we use the likelihood given in (7) which is constructed using the complete data on each subject. On the other hand, a new likelihood is constructed in pairwise likelihood method based on successive and/or alternative pairs of data on each subject. We then estimate the correlation parameter λ by maximizing this newly constructed likelihood. Although, we lose some information but for some correlation structures the contribution of the observations other than successive and alternative pairs is minimal in estimating the correlation parameters. In the following sections we further elaborate the pairwise likelihood method specific to each correlation structure that we presented in Section 1.3.

Pairwise Likelihood Method for ARMA(1, 1) Correlation Structure

As discussed in Section 1.3.1, the two parameters γ, ρ govern the association in ARMA(1, 1) correlation structure. Thus, we employ the pairwise likelihood method for estimation of $\lambda = (\gamma, \rho)$. The pairwise likelihood estimators are obtained by maximizing the likelihood functions constructed using the pairs formed by taking the successive and the alternate repeated measures, because the successive pairs, y_{ij} and $y_{i(j+1)}$, provides the information about the correlation parameter γ and the alternate pairs, y_{ij} and $y_{i(j+2)}$, involve ρ in their correlations. Although some efficiency loss may occur in the pairwise likelihood method due to omission of the information that can be obtained using the other pairs, an obvious advantage is the substantial gain in the ease of computation due to the fact that the dimension of the correlation structure is much simpler.

To estimate γ, as a first step, consider the successive pairs of the response vector Y_i. Define $Y_{1ij} = (y_{ij}, y_{i(j+1)})'$ for $j = 1, 2, \cdots, t-1$; $i = 1, 2, \cdots, n$. Based on the normality assumption of Y_i we have

$$Y_{1ij} = \begin{pmatrix} y_{ij} \\ y_{i(j+1)} \end{pmatrix} \sim N\left(\begin{pmatrix} x'_{ij}\beta \\ x'_{i(j+1)}\beta \end{pmatrix}, \phi R_1\right) \quad \text{where } R_1 = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}. \tag{22}$$

Define Z_{1ij} as

$$Z_{1ij} = \begin{pmatrix} y_{ij} - x'_{ij}\beta \\ y_{i(j+1)} - x'_{i(j+1)}\beta \end{pmatrix} = \begin{pmatrix} z_{ij} \\ z_{i(j+1)} \end{pmatrix}.$$

Thus, considering all such pairs, the loglikelihood function for γ in case of ARMA(1, 1) model is given as:

$$\ell_1 = \sum_{i=1}^{n}\sum_{j=1}^{t-1} -\log(2\pi) - \log\phi - \frac{1}{2}\log|R_1| - \frac{1}{2\phi}Z'_{1ij}\, R_1^{-1}\, Z_{1ij}. \tag{23}$$

Equating to zero the derivative of the above loglikelihood function with respect to γ results in the following score equation.

$$-\frac{n(t-1)}{2}\text{tr}\left(R_1^{-1}\frac{\partial R_1}{\partial\gamma}\right) + \frac{1}{2\widehat{\phi}}\sum_{i=1}^{n}\sum_{j=1}^{t-1} Z'_{1ij}\left(R_1^{-1}\frac{\partial R_1}{\partial\gamma}R_1^{-1}\right) Z_{1ij} = 0 \tag{24}$$

If we let

$$h_{i1} = -\frac{(t-1)}{2}\text{tr}\left(R_1^{-1}\frac{\partial R_1}{\partial\gamma}\right) + \frac{1}{2\widehat{\phi}}\sum_{j=1}^{t-1} Z'_{1ij}\left(R_1^{-1}\frac{\partial R_1}{\partial\gamma}R_1^{-1}\right) Z_{1ij}, \tag{25}$$

then (24) can be written as $h_1 = \sum_{i=1}^{n} h_{i1}$. Since R_1 given in (22) has simpler form, the terms $\text{tr}\left(R_1^{-1}\frac{\partial R_1}{\partial\gamma}\right)$ and $R_1^{-1}\frac{\partial R_1}{\partial\gamma}R_1^{-1}$ simplifies to a closed form expressions. Thus, equation (24) on further simplification reduces to the following cubic equation in γ.

$$\begin{aligned} &-n\widehat{\phi}(t-1)\gamma^3 + \left[\sum_{i=1}^{n}\sum_{j=1}^{t-1}\widehat{z}_{i,j}\widehat{z}_{i,j+1}\right]\gamma^2 \\ &+ \left[n(t-1)\widehat{\phi} - \sum_{i=1}^{n}\sum_{j=1}^{t-1}\left(\widehat{z}^2_{i,j} + \widehat{z}^2_{i,j+1}\right)\right]\gamma + \left[\sum_{i=1}^{n}\sum_{j=1}^{t-1}\widehat{z}_{i,j}\widehat{z}_{i,j+1}\right] = 0. \end{aligned} \tag{26}$$

It is easy to obtain closed form solutions for the roots of cubic polynomial and find the feasible estimate of γ.

In the next step, we estimate ρ by considering the alternate pairs of the response vector Y_i and follow a similar procedure given in the first step. Define $Y_{2ij} = (y_{ij}, y_{i(j+2)})'$ for $j = 1, 2, \cdots, t-2; i = 1, 2, \cdots, n$. Based on the normality assumption of Y_i,

$$Y_{2ij} = \begin{pmatrix} y_{ij} \\ y_{i(j+2)} \end{pmatrix} \sim N\left(\begin{pmatrix} x'_{ij}\beta \\ x'_{i(j+2)}\beta \end{pmatrix}, \phi R_2\right) \quad \text{where} \quad R_2 = \begin{pmatrix} 1 & \gamma\rho \\ \gamma\rho & 1 \end{pmatrix}.$$

Let

$$Z_{2ij} = \begin{pmatrix} y_{ij} - x'_{ij}\beta \\ y_{i(j+2)} - x'_{i(j+2)}\beta \end{pmatrix} = \begin{pmatrix} z_{ij} \\ z_{i(j+2)} \end{pmatrix}.$$

Thus, considering all such pairs, the loglikelihood function of ρ for a fixed value of γ is given as:

$$\ell_2 = \sum_{i=1}^{n}\sum_{j=1}^{t-1} -\log(2\pi) - \log\phi - \frac{1}{2}\log|R_2| - \frac{1}{2\phi} Z'_{2ij}\, R_2^{-1}\, Z_{2ij}.$$

The score equation of ρ becomes

$$-\frac{n(t-2)}{2}\text{tr}\left(R_2^{-1}\frac{\partial R_2}{\partial\rho}\right) + \frac{1}{2\widehat{\phi}}\sum_{i=1}^{n}\sum_{j=1}^{t-2} Z'_{2ij}\left(R_2^{-1}\frac{\partial R_2}{\partial\rho}R_2^{-1}\right)Z_{2ij} = 0. \tag{27}$$

Let

$$h_{i2} = -\frac{(t-2)}{2}\text{tr}\left(R_2^{-1}\frac{\partial R_2}{\partial\rho}\right) + \frac{1}{2\widehat{\phi}}\sum_{j=1}^{t-2} \widehat{Z}'_{2ij}\left(R_2^{-1}\frac{\partial R_2}{\partial\rho}R_2^{-1}\right)\widehat{Z}_{2ij}, \tag{28}$$

then (27) can be written as $h_2 = \sum_{i=1}^{n} h_{i2}$, which further reduces to the following cubic equation in ρ.

$$-n\phi(t-2)\gamma^4\rho^3 + \left[\sum_{i=1}^{n}\sum_{j=1}^{t-2}\gamma^2 z_{ij}z_{i(j+2)}\right]\rho^2$$
$$+\left[n\phi(t-2)\gamma^2 - \sum_{i=1}^{n}\sum_{j=1}^{t-2}\gamma\left(z_{ij}^2 + z_{i(j+2)}^2\right)\right]\rho + \left[\sum_{i=1}^{n}\sum_{j=1}^{t-2} z_{ij}z_{i(j+2)}\right] = 0. \tag{29}$$

Based on above likelihoods the estimation procedure in case of ARMA(1, 1) correlation structure can be described as follows:

Step 1: Choose initial values β_0 and ϕ_0 for β and ϕ.

Step 2: Solve the cubic equation (26) with β_0, ϕ_0 to obtain a solution for γ, say γ_0.

Step 3: Solve the cubic equation (29) with β_0, ϕ_0 and γ_0 to obtain a solution for ρ, say ρ_0.

Step 4: Using γ_0 and ρ_0, update the value of β and ϕ using the equation (9) and denote them as β_1 and ϕ_1.

Step 5: Replace β_1=β_0 and ϕ_1= ϕ_0 and repeat Steps 2 - 4 until convergence is achieved.

We denote the final pairwise likelihood estimates by $\widehat{\theta}_{PL} = (\widehat{\beta}, \widehat{\phi}, \widehat{\lambda})$.

2.3.3 PAIRWISE LIKELIHOOD METHOD FOR OTHER CORRELATION STRUCTURES

We repeat the same procedure outlined in the previous section to obtain the estimates in case of AR(1), MA(1) and CS correlation structures. However, we make a slight modification to the method keeping in mind the fact that there is only one correlation parameter involved in these correlation structures. Hence, we consider only the successive pairs to construct the likelihood as mentioned in the first step of pairwise likelihood method in case of ARMA(1, 1). The likelihood equation turns out to be same in all the three correlation structures, because they have ρ as the correlation between the successive measures. Thus, the likelihood function for ρ is same as equation (23) with R_1 equals

$$R_1 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The score equation for ρ is given as

$$\begin{aligned} h_1: \quad & -n\phi(t-1)\rho^3 + \left[\sum_{i=1}^{n}\sum_{j=1}^{t-1} z_{ij}z_{i(j+1)}\right]\rho^2 \\ & + \left[n(t-1)\phi - \sum_{i=1}^{n}\sum_{j=1}^{t-1}\left(z_{ij}^2 + z_{i(j+1)}^2\right)\right]\rho + \left[\sum_{i=1}^{n}\sum_{j=1}^{t-1} z_{ij}z_{i(j+1)}\right] = 0. \end{aligned} \tag{30}$$

See A.1 for a feasible root of this cubic polynomial in a closed form. The same iterative procedure in the previous section can be followed to obtain the estimates of β, ϕ and ρ.

2.4 ASYMPTOTIC VARIANCES

To find the efficiency of pairwise likelihood estimators when compared to maximum likelihood estimators we use the asymptotic theory described in Section 1.5. In order to calculate these efficiencies, we need the asymptotic variance of the estimates from each of the estimating procedures. In the following section we derive these asymptotic variances for each of the estimating procedures.

2.4.1 MAXIMUM LIKELIHOOD ESTIMATORS

The asymptotic variances and covariances for the maximum likelihood estimators defined in Section 1.5.1 are given by the inverse of the Fisher Information matrix. The diagonal elements are the variances and the off-diagonals represent the covariances. In general, the Fisher information matrix for parameter $\theta = (\beta,\ \phi,\ \lambda)$ is

$$\mathcal{I}_\ell(\theta) = \begin{pmatrix} I_1(\beta) & 0 \\ 0 & I_2(\phi,\lambda) \end{pmatrix}. \tag{31}$$

Note that the dimension of $I_2(\phi,\lambda)$ depends on the dimension of λ. For example, in the model with ARMA(1, 1) correlation structure the dimension of $I_2(\phi,\lambda)$ is 3. Hence, in this model $\mathcal{I}_\ell(\theta)$ can be represented as

$$\mathcal{I}_\ell(\theta) = \begin{pmatrix} I(\beta) & 0 & 0 & 0 \\ 0 & I(\phi) & I(\phi,\gamma) & I(\phi,\rho) \\ 0 & I(\phi,\gamma) & I(\gamma) & I(\gamma,\rho) \\ 0 & I(\phi,\rho) & I(\gamma,\rho) & I(\rho) \end{pmatrix}. \tag{32}$$

Similarly for the other correlation structures, AR(1), MA(1) and CS, the dimension of $I_2(\phi,\lambda)$ is 2, and $\mathcal{I}_\ell(\theta)$ can be represented as

$$\mathcal{I}_\ell(\theta) = \begin{pmatrix} I(\beta) & 0 & 0 \\ 0 & I(\phi) & I(\phi,\rho) \\ 0 & I(\phi,\rho) & I(\rho) \end{pmatrix}. \tag{33}$$

The individual terms in $\mathcal{I}_\ell(\theta)$ are given below.

$$\begin{aligned}
I(\beta) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \beta^2}\right) &&= \frac{1}{\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i \\
I(\phi) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \phi^2}\right) &&= \frac{nt}{2\phi^2} \\
I(\phi,\gamma) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \gamma \partial \phi}\right) &&= \frac{n}{2\phi}\mathrm{tr}\left(R^{-1}\frac{\partial R}{\partial \gamma}\right) \\
I(\phi,\rho) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \rho \partial \phi}\right) &&= \frac{n}{2\phi}\mathrm{tr}\left(R^{-1}\frac{\partial R}{\partial \rho}\right) \\
I(\gamma) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \gamma^2}\right) &&= \frac{n}{2}\mathrm{tr}\left(R^{-1}\frac{\partial R}{\partial \gamma}R^{-1}\frac{\partial R}{\partial \gamma}\right) \\
I(\gamma,\rho) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \rho \partial \gamma}\right) &&= \frac{n}{2}\mathrm{tr}\left(R^{-1}\frac{\partial R}{\partial \gamma}R^{-1}\frac{\partial R}{\partial \rho}\right) \\
I(\rho) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \rho^2}\right) &&= \frac{n}{2}\mathrm{tr}\left(R^{-1}\frac{\partial R}{\partial \rho}R^{-1}\frac{\partial R}{\partial \rho}\right),
\end{aligned} \tag{34}$$

where R can be any correlation matrix as defined in Section 1.3 and $\ell(\theta)$ as in (8). We left the expressions in (34) in their general form without further simplifications for each correlation structure that R can take.

2.4.2 PAIRWISE LIKELIHOOD ESTIMATORS

The asymptotic variances and covariances for the pairwise likelihood estimators are obtained using Godambe information matrix, $\mathcal{G}(\theta)$, as explained in Section 1.5.2. Recall that the Godambe information matrix for θ is given as

$$\mathcal{G}(\theta) = D(\theta)\ M^{-1}(\theta)\ (D(\theta))'$$

where $D(\theta) = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_i(\theta)}{\partial \theta}\right)$ and $M(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(h_i(\theta))$.

Similar to the Fisher information matrix discussed in the previous section, the dimension of Godambe information matrix also depends on the type of correlation structure. Therefore, we give expressions for the Godambe information matrix for each of the correlation types.

Pairwise Likelihood estimators for ARMA(1, 1)

In the pairwise likelihood estimation procedure for ARMA(1, 1) correlation structure, the vector of unbiased estimating equations $h_i(\theta)$, for $i = 1, 2, \cdots, n$, are given as

$$h_i(\theta) = (g_{i1}(\theta),\, g_{i2}(\theta),\, h_{i1}(\theta),\, h_{i2}(\theta))',$$

where $h_{i1}(\theta)$ and $h_{i2}(\theta)$ are the estimating equations corresponding to the correlation parameters γ and ρ respectively that are defined in equations (25) and (28). The estimating equations $g_{i1}(\theta)$ and $g_{i2}(\theta)$ correspond to the regression coefficients β and the residual variance ϕ respectively. Note that the estimating equations for $g_{i1}(\theta)$ and g_{i2} are same for any of the correlation structures R and are given as

$$g_{i1}(\theta) = \frac{1}{\phi} X_i' R^{-1} Z_i$$

and

$$g_{i2}(\theta) = \frac{-t}{2\phi} + \frac{1}{2\phi^2} Z_i' R^{-1} Z_i.$$

The terms in the Godambe information matrix $D(\theta)$ and $M(\theta)$ for the ARMA(1, 1) correlation structure are as follows.

$$D(\theta) = -\frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial h_i(\theta)}{\partial \theta'}\right] = \begin{pmatrix} D_{11} & 0 & 0 & 0 \\ 0 & D_{22} & D_{23} & D_{24} \\ 0 & D_{32} & D_{33} & 0 \\ 0 & D_{42} & D_{43} & D_{44} \end{pmatrix},$$

where

$$\begin{aligned}
D_{11} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial g_{i1}(\theta)}{\partial \beta}\right) &= \frac{1}{n\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i, \\
D_{22} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial g_{i2}(\theta)}{\partial \phi}\right) &= \frac{t}{2\phi^2}, \\
D_{23} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial g_{i2}(\theta)}{\partial \gamma}\right) &= \frac{1}{2\phi} tr\left(R^{-1}\frac{\partial R}{\partial \gamma}\right), \\
D_{24} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial g_{i2}(\theta)}{\partial \rho}\right) &= \frac{1}{2\phi} tr\left(R^{-1}\frac{\partial R}{\partial \rho}\right),
\end{aligned}$$

$$
\begin{aligned}
D_{32} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_{i1}(\theta)}{\partial \phi}\right) &=& \frac{-\gamma(t-1)}{\phi(1-\gamma^2)^2}, \\
D_{33} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_{i1}(\theta)}{\partial \gamma}\right) &=& \frac{(t-1)(1+\gamma^2)}{(1-\gamma^2)^2}, \\
D_{42} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_{i2}(\theta)}{\partial \phi}\right) &=& \frac{-\gamma^2\rho(t-2)}{\phi(1-\gamma^2\rho^2)^2}, \\
D_{43} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_{i2}(\theta)}{\partial \gamma}\right) &=& \frac{(t-2)\gamma\rho(1+\gamma^2\rho^2)}{(1-\gamma^2\rho^2)^2}, \\
D_{44} &= -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_{i2}(\theta)}{\partial \rho}\right) &=& \frac{(t-2)\gamma^2(1+\gamma^2\rho^2)}{(1-\gamma^2\rho^2)^2}.
\end{aligned}
$$

Next

$$
M(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(h_i(\theta)) = \begin{pmatrix} M_{11} & 0 & 0 & 0 \\ 0 & M_{22} & M_{23} & M_{24} \\ 0 & M_{23} & M_{33} & M_{34} \\ 0 & M_{24} & M_{34} & M_{44} \end{pmatrix},
$$

where

$$
\begin{aligned}
M_{11} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(g_{i1}) &=& \frac{1}{n\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i \\
M_{22} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(g_{i2}) &=& \frac{t}{2\phi^2} \\
M_{23} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(g_{i2}, h_{i1}) &=& \frac{-\gamma(t-1)}{\phi(1-\gamma^2)} \\
M_{24} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(g_{i2}, h_{i2}) &=& \frac{-\gamma^2\rho(t-2)}{\phi(1-\gamma^2\rho^2)} \\
M_{33} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(h_{i1}) &=& \frac{(t-1)(1+\gamma^2)}{(1-\gamma^2)^2} + (t-2)\mathrm{tr}(R_1^* A_1 R_1^* A_1') \\
& & & + \sum_{j_2-j_1>1} \mathrm{tr}(R_1^* A R_1^* A').
\end{aligned}
$$

Here $j_2 > j_1 = 1, 2, \cdots, t-1$ and

$$R_1^* = \frac{1}{(1-\gamma^2)^2} \begin{pmatrix} -2\gamma & 1+\gamma^2 \\ 1+\gamma^2 & -2\gamma \end{pmatrix}, \tag{35}$$

$$A_1 = \begin{pmatrix} \gamma & \gamma\rho \\ 1 & \gamma \end{pmatrix}, \quad A = \begin{pmatrix} \gamma\rho^{j_2-j_1-1} & \gamma\rho^{j_2-j_1} \\ \gamma\rho^{j_2-j_1-2} & \gamma\rho^{j_2-j_1-1} \end{pmatrix}.$$

Now,

$$M_{34} = \frac{1}{n}\sum_1^n \mathrm{Cov}(h_{i1}, h_{i2}) = \frac{1}{2}\sum_{j_1=1}^{t-1}\sum_{j_2=1}^{t-2} \mathrm{tr}\big(R_1^* A_{j_1j_2} R_2^* A'_{j_1j_2}\big),$$

where R_1^* and R_2^* are as defined in (35) and (36). Let γ_{ij} be the (i, j) th element of $R(\lambda)$ defined in (2) then

$$A_{j_1j_2} = \begin{pmatrix} \gamma_{j_1j_2} & \gamma_{j_1(j_2+2)} \\ \gamma_{(j_1+1)j_2} & \gamma_{(j_1+1)j_2+2} \end{pmatrix}.$$

The last term M_{44} is given by

$$\begin{aligned} M_{44} &= \frac{1}{n}\sum_1^n \mathrm{Var}(h_{i2}) \\ &= \frac{(t-2)\gamma^2(1+\gamma^2\rho^2)}{(1-\gamma^2\rho^2)^2} + (t-3)\,\mathrm{tr}(R_2^* A_1 R_2^* A_1') + (t-4)\,\mathrm{tr}(R_2^* A_2 R_2^* A_2') \\ &\quad + \sum_{j_2-j_1>2} \mathrm{tr}(R_2^* A R_2^* A'), \end{aligned}$$

where $j_2 > j_1 = 1, 2, \cdots, t-2$ and

$$R_2^* = \frac{\gamma}{(1-\gamma^2\rho^2)^2} \begin{pmatrix} -2\gamma\rho & 1+\gamma^2\rho^2 \\ 1+\gamma^2\rho^2 & -2\gamma\rho \end{pmatrix}, \tag{36}$$

$$A_1 = \begin{pmatrix} \gamma & \gamma\rho^2 \\ \gamma & \gamma \end{pmatrix},$$

$$A_2 = \begin{pmatrix} \gamma\rho & \gamma\rho^3 \\ 1 & \gamma\rho \end{pmatrix},$$

$$A = \begin{pmatrix} \gamma\rho^{j_2-j_1-1} & \gamma\rho^{j_2-j_1+1} \\ \gamma\rho^{j_2-j_1-3} & \gamma\rho^{j_2-j_1-1} \end{pmatrix}.$$

The asymptotic covariance matrix for the pairwise likelihood estimators of θ is $\frac{1}{n}\mathcal{G}^{-1}(\theta) = \frac{1}{n}D^{-1}(\theta)\, M(\theta)\, (D^{-1}(\theta))'$.

Pairwise Likelihood estimators for other correlation structures

In the pairwise likelihood estimation procedure for AR(1), MA(1) and CS correlation structures, the vector of unbiased estimating equations $h_i(\theta)$, for $i = 1, 2, \cdots, n$, is given as

$$h_i(\theta) = (g_{i1}(\theta),\, g_{i2}(\theta),\, h_{i1}(\theta))',$$

where $h_{i1}(\theta)$ are the estimating equations corresponding to the correlation parameter ρ and they are defined in equation (30) as $h_1 = \sum_{i=1}^{n} h_{i1}(\theta)$. Similar to ARMA(1, 1) case, $g_{i1}(\theta)$ and $g_{i2}(\theta)$ are the estimating equations corresponding to the regression coefficient β and the residual variance ϕ respectively and their expressions are given in Section 2.4.2. Also, $D(\theta)$ and $M(\theta)$ have similar expressions as in Section 2.4.2 but with slight modifications as given below

$$D(\theta) = \begin{pmatrix} \frac{1}{n\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i & 0 & 0 \\ 0 & \frac{t}{2\phi^2} & \frac{1}{2\phi} tr\left(R^{-1}\frac{\partial R}{\partial \rho}\right) \\ 0 & \frac{-\rho(t-1)}{\phi(1-\rho^2)^2} & \frac{(t-1)(1+\rho^2)}{(1-\rho^2)^2} \end{pmatrix},$$

with R as the corresponding AR(1), MA(1) or CS correlation structure. Also, $M(\theta)$ is given by

$$M(\theta) = \begin{pmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & M_{23} \\ 0 & M_{23} & M_{33} \end{pmatrix},$$

where

$$M_{11} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(g_{i1}) = \frac{1}{n\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i$$

$$M_{22} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(g_{i2}) = \frac{t}{2\phi^2}$$

$$M_{23} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(g_{i2}, h_{i1}) = \frac{-\rho(t-1)}{\phi(1-\rho^2)}$$

$$M_{33} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(h_{i1}) = \frac{(t-1)(1+\rho^2)}{(1-\rho^2)^2} + (t-2)\,\mathrm{tr}(R_1^* A_1 R_1^* A_1') + \sum_{j_2 - j_1 > 1} \mathrm{tr}(R_1^* A R_1^* A'),$$

where $j_2 > j_1 = 1, 2, \cdots, t-1$ and

$$R_1^* = \frac{1}{(1-\rho^2)^2}\begin{pmatrix} -2\rho & 1+\rho^2 \\ 1+\rho^2 & -2\rho \end{pmatrix}, \tag{37}$$

$$A_1 = \begin{pmatrix} \rho & \rho^2 \\ 1 & \rho \end{pmatrix}, \quad A = \begin{pmatrix} \rho^{j_2-j_1} & \rho^{j_2-j_1+1} \\ \rho^{j_2-j_1-1} & \rho^{j_2-j_1} \end{pmatrix}.$$

The asymptotic covariance matrix for the pairwise likelihood estimators is

$$\frac{1}{n}\mathcal{G}^{-1}(\theta) = \frac{1}{n}D^{-1}(\theta)\, M(\theta)\, (D^{-1}(\theta))'.$$

2.5 ASYMPTOTIC COMPARISONS

Based on the asymptotic theory given in Section 1.5, we have shown that the pairwise likelihood method yields consistent estimators. We know that the maximum likelihood estimators are consistent asymptotically. Hence, we compare the efficiency of these two estimators using asymptotic relative efficiency. The asymptotic relative efficiency of pairwise likelihood estimator($\widehat{\theta}_{\mathrm{PL}}$) to maximum likelihood estimator($\widehat{\theta}_{\mathrm{ML}}$) is defined as the ratio of asymptotic variance of $\widehat{\theta}_{\mathrm{PL}}$ and $\widehat{\theta}_{\mathrm{ML}}$. These variances are obtained as the diagonal elements of asymptotic covariance matrices $\mathcal{I}_\ell^{-1}(\theta)$ and $\mathcal{G}^{-1}(\theta)$ derived in Section 2.4.

As we have seen, pairwise likelihood method differs from the maximum likelihood method in the way the correlation parameters are estimated. The analytical estimators of β and ϕ remain unchanged irrespective of the method of estimation. Thus, it is sufficient to study the efficiency of only the correlation parameter estimators obtained from the two methods. It can also be observed from $\mathcal{I}_\ell^{-1}(\theta)$ and $\mathcal{G}^{-1}(\theta)$ that the asymptotic covariance matrices for β are identical. As a consequence, the relative efficiency is equal to 1 entailing the equivalent performance of pairwise likelihood and maximum likelihood method for the regression parameter estimators. For each correlation parameter, we compute the efficiency of pairwise likelihood estimators to maximum likelihood estimator by taking the ratio of the corresponding asymptotic variances from $\mathcal{I}_{\ell}^{-1}(\theta)$ and $\mathcal{G}^{-1}(\theta)$.

We can observe that the computations for asymptotic relative efficiency does not depend on the value of n. Hence, to start with, we compute the asymptotic variances for fixed values of $t = 4$ and $\phi = 3$ and we let the values of correlation parameters vary between -1 and 1. In case of ARMA(1, 1), for each of several combinations of γ and ρ in the positive definite range we calculate asymptotic relative efficiency. Later we plot efficiencies for different values of γ and ρ. Similar asymptotic relative efficiency computations are performed for the correlation parameter ρ using AR(1), MA(1) and CS correlation structures.

2.5.1 ASYMPTOTIC RELATIVE EFFICIENCIES FOR PAIRWISE LIKELIHOOD ESTIMATORS IN CASE OF ARMA(1, 1)

We consider the ARMA(1, 1) correlation matrix and compute the asymptotic relative efficiency (ARE) for pairwise likelihood estimators of γ and ρ . We choose several combinations for the pair $(\gamma,\ \rho)$ and calculate the asymptotic variances using the formulas derived in Sections 2.4.1 and 2.4.2.

Table 4 provides the asymptotic relative efficiency of pairwise likelihood estimator of γ when $t = 4$ and $\phi = 3$. In this table, for each value of γ, we present the positive definite range of ρ described in Section 2.2.1. For the values of $(\gamma,\ \rho)$ in the positive definite range, we present the AREs of γ. Here we see the ARE of γ is very high over a wide range of γ and ρ, indicating that the pairwise likelihood estimator variance is almost as small as that of the maximum likelihood estimator, and only for extreme values of γ close to the boundary the efficiency of the pairwise likelihood method decrease when compared to the maximum likelihood method. However, the efficiencies remain more than 0.8.

Figure 3 provides a graphical representation of the asymptotic relative efficiency of γ for all values of γ and ρ in the positive definite range. The interpretations for

Table 4. Asymptotic Relative Efficiency of γ for ARMA(1, 1)

γ	Range of ρ	ρ	ARE γ	γ	Range of ρ	ρ	ARE γ
-0.985	(-1, -0.964)	-0.99	0.9887	0.005	(-1, 1)	0.10	1
		-0.98	0.9815			0.25	0.9989
		-0.97	0.9815			0.35	0.9962
-0.745	(-1, -0.355)	-0.85	0.9932	0.5	(-0.366, 1)	0.28	0.9879
		-0.75	1			0.40	0.9984
		-0.65	0.9827			0.55	0.9997
		-0.55	0.7242			0.75	0.9952
-0.505	(-1, 0.3495)	-0.75	0.9952	0.74	(0.3415, 1)	0.55	0.7815
		-0.50	1			0.65	0.9862
		-0.40	0.9982			0.75	0.9999
		-0.35	0.9952			0.85	0.9929
0.005	(-1, 1)	-0.34	0.9962	0.98	(0.952, 1)	0.96	0.8777
		-0.20	0.9995			0.97	0.9461
		-0.05	1			0.98	1

NOTE: Range of ρ is the positive definite range. The parameter values are $t = 4$ and $\phi = 3$.

AREs of γ given based on Table 4 can be clearly seen in the Figure 3. We next study the AREs of ρ.

Table 5 presents the asymptotic relative efficiency of ρ with values of γ and ρ in the positive definite range for $t = 4$ and $\phi = 3$. We can observe that the ARE of ρ is high when the values of ρ are close to zero and the AREs decreases as we move farther away from zero. The same pattern follows for any fixed value of γ. For values of γ close to zero, the ARE of ρ is almost close to one, indicating that pairwise likelihood method is equivalent to maximum likelihood method.

Figure 4 displays the asymptotic relative efficiency plot for ρ. In this figure, we can observe that the AREs of ρ follows a bell-shaped pattern for any fixed value of γ. The AREs tend to remain small for extreme values of ρ and they increase as the value of ρ approaches zero.

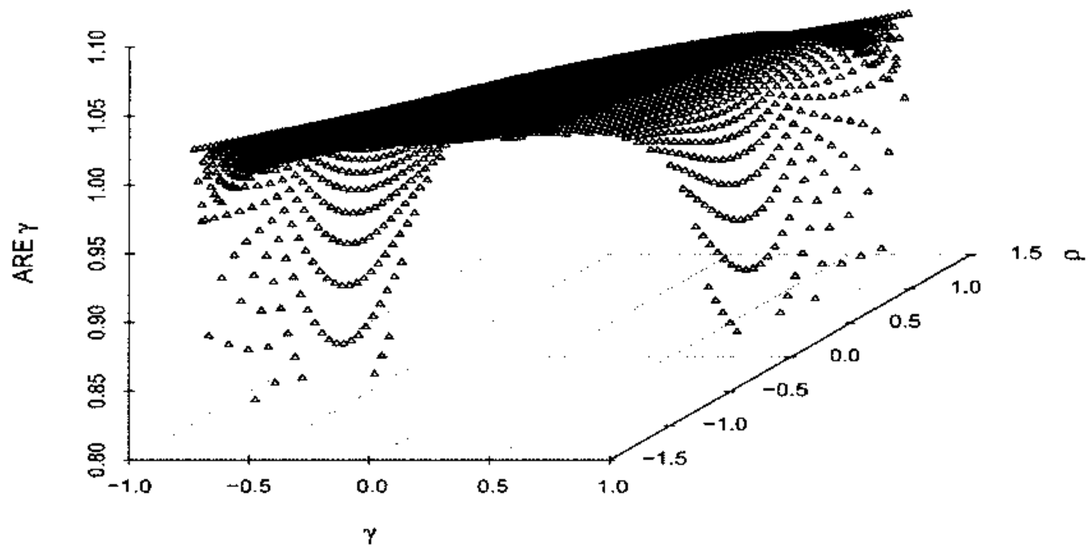


Figure 3. ARE of γ for $t = 4$ and $\phi = 3$

Therefore, after investigating the AREs, we can conclude that asymptotically the pairwise likelihood method is comparable to maximum likelihood method for most of the plausible correlation values.

2.5.2 ASYMPTOTIC RELATIVE EFFICIENCY FOR PAIRWISE LIKELIHOOD ESTIMATORS IN CASE OF OTHER CORRELATION STRUCTURES

We now consider the AR(1), MA(1) and CS correlation structures and compute the asymptotic relative efficiency (ARE) for pairwise likelihood estimators of ρ. For each value of γ in the positive definite range, we calculate the asymptotic variances using the formulas derived in Sections 2.4.1 and 2.4.2.

Table 5. Asymptotic Relative Efficiency of ρ for ARMA(1, 1)

γ	Range of ρ	ρ	ARE ρ	γ	Range of ρ	ρ	ARE ρ
-0.985	(-1, -0.964)	-0.99	0.8801	0.005	(-1, 1)	0.10	0.9823
		-0.98	0.9294			0.25	0.8928
		-0.97	0.9539			0.35	0.8080
-0.745	(-1, -0.355)	-0.85	0.8450	0.5	(-0.366, 1)	0.28	0.8015
		-0.75	0.8896			0.40	0.8836
		-0.65	0.8416			0.55	0.9201
		-0.55	0.8513			0.75	0.7860
-0.505	(-1, 0.3495)	-0.74	0.7895	0.74	(0.3415, 1)	0.55	0.8358
		-0.50	0.9172			0.65	0.8435
		-0.40	0.8791			0.75	0.8905
		-0.35	0.8477			0.85	0.8407
0.005	(-1, 1)	-0.34	0.8083	0.98	(0.952, 1)	0.96	0.6670
		-0.20	0.9225			0.97	0.9707
		-0.05	0.9940			0.98	0.9281

NOTE: Range of ρ is the positive definite range. The parameter values are $t = 4$ and $\phi = 3$.

In case of AR(1), the AREs of γ are all equal to 1, indicating that the pairwise likelihood method is equally efficient as maximum likelihood method for any t.

Table 6 (a) and 6 (b) provides the asymptotic relative efficiency of γ when $t = 4$ and $\phi = 3$ for MA(1) and CS correlation models. In this table, for each value of γ in the positive definite range described in Section 2.2.2, we present the AREs of γ. Here we see that in case of MA(1), the ARE is high for values of γ close to zero, indicating the performance of pairwise likelihood method compared to the maximum likelihood method. Further, in case of CS correlation model, the AREs of γ follow an interesting cubic pattern of increasing efficiency as the value of γ approaches upper boundary value of one.

Table 6. Asymptotic Relative Efficiency of γ for MA(1) and CS Structures with $t = 4$ and $\phi = 3$

(a) MA(1)

γ	ARE of γ
-0.6	0.3751
-0.5	0.1902
-0.45	0.0246
-0.4	0.0432
-0.35	0.2262
-0.3	0.4368
-0.25	0.6216
-0.15	0.8736
-0.1	0.9454
-0.05	0.9866
0	1
0.05	0.9866
0.1	0.9454
0.15	0.8736
0.2	0.7671
0.25	0.6216
0.3	0.4368
0.35	0.2262
0.4	0.0432
0.45	0.0246
0.5	0.1902
0.6	0.3751

(b) CS

γ	ARE of γ
-0.3	0.2716
-0.25	0.0838
-0.2	0.0005
-0.15	0.0861
-0.1	0.2385
-0.05	0.3841
0	0.5000
0.05	0.5855
0.1	0.6459
0.2	0.7147
0.3	0.7426
0.35	0.7485
0.4	0.7517
0.5	0.7552
0.55	0.7579
0.6	0.7623
0.7	0.7802
0.75	0.7957
0.8	0.8171
0.9	0.8841
0.95	0.9343

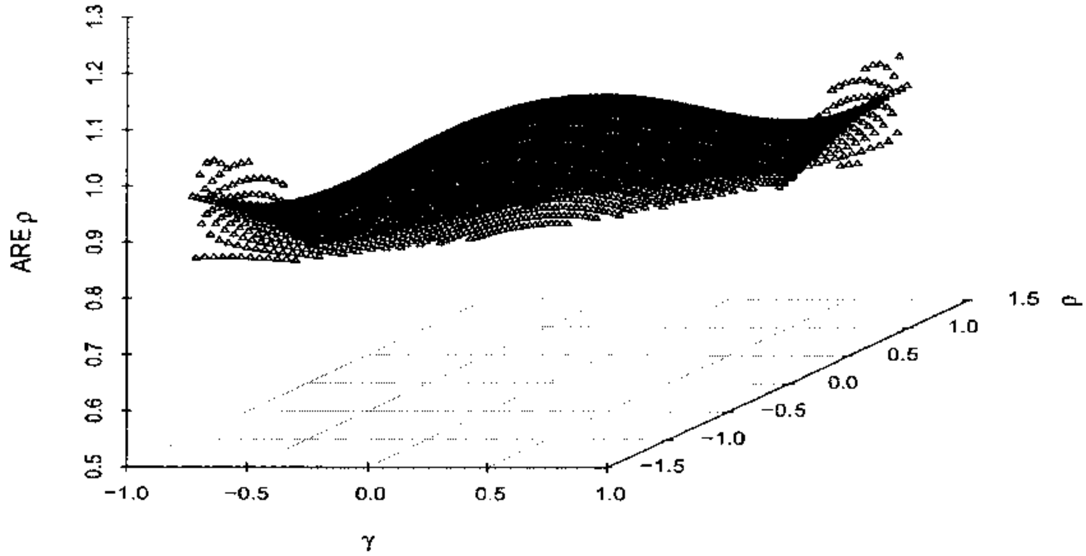


Figure 4. ARE of ρ for $t = 4$ and $\phi = 3$

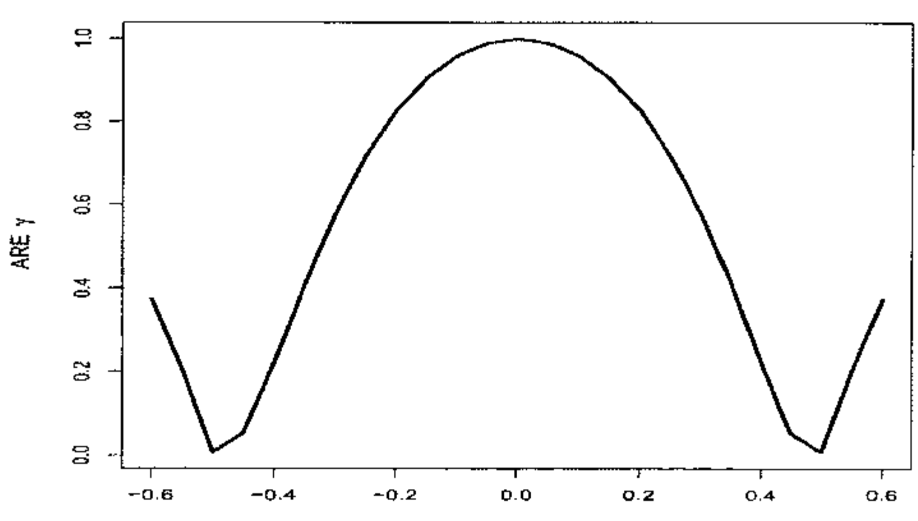


Figure 5. ARE of γ in MA(1) for $t = 4$ and $\phi = 3$

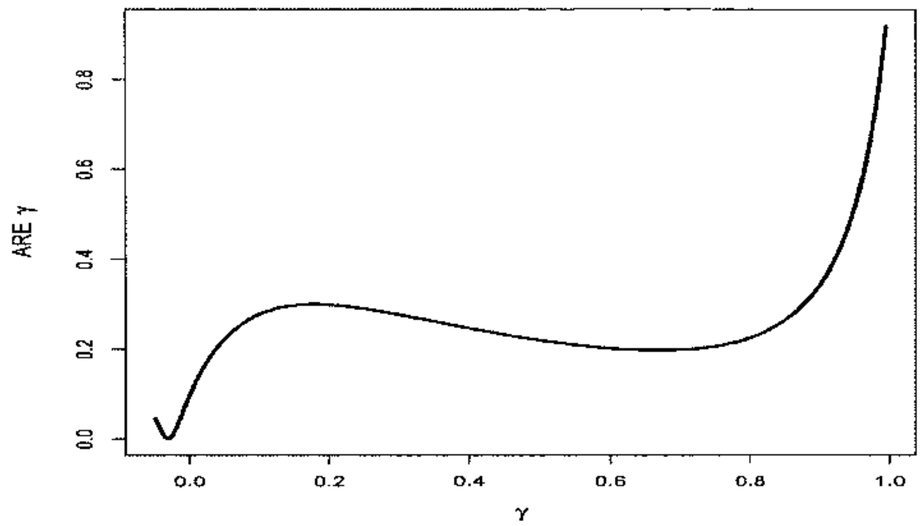


Figure 6. ARE of γ in CS for $t = 4$ and $\phi = 3$

Figures 5 and 6 provides a graphical representation of the asymptotic relative efficiency of γ for all plausible values of γ . The interpretations for AREs of γ given based on Table 6 (a) and 6 (b) can be clearly seen in the Figures 5 and 6.

2.6 ANALYSIS OF REAL DATA

To illustrate the application of the two likelihood estimation methods, in this section we present the analysis of the two continuous longitudinal data that we introduced in Sections 1.6.1 and 1.6.2.

2.6.1 OZONE DATA

Table 1 in Chapter 1 displays a sample subset of the ozone data. The response, ozone levels in the environment are recorded at five different time points beginning from 7 am to 12 noon during one hour interval. These observations are taken on three consecutive days. A research interest focuses on modeling the change in ozone levels during the day. The data we analyze here contain $n = 60$ records. We assume a ARMA(1, 1) correlation structure to model the dependencies among the ozone recordings from the same day.

Table 7. Parameter Estimates for the Ozone Data

	PLE			MLE		
Parameter	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	11.7864	0.9086	< 0.0001	11.6561	0.9582	< 0.0001
Day-1	-2.5743	1.2849	0.0451	-2.5657	1.3550	0.0434
Day-2	-1.4618	1.2849	0.2552	-1.4361	1.3550	0.2937
ϕ	23.5020	3.2509	< 0.0001	26.3972	3.5082	< 0.0001
γ	0.8422	0.0236	< 0.0001	0.8595	0.0195	< 0.0001
ρ	0.7895	0.0429	< 0.0001	0.7747	0.0367	< 0.0001
-2ℓ	1497.5			1487.2		

For the response model in Table 7, the pairwise and the maximum likelihood estimates are approximately close to each other. The p-values from both the approaches reveal that there is significant difference between the Day-1 and Day-3 ozone levels with a negative estimate indicating that the levels are low on Day-1 compared to Day-3. The p-value of 0.2552 implies that the difference between Day-2 and Day-3 is not significant.

2.6.2 OXYGEN SATURATION DATA

Table 2 displays a subset of the oxygen saturation data. The main research interest in this study is to examine the effectiveness of three different methods of suctioning an endotracheal tube: Standard suctioning, a new method using a special vacuum, and manual bagging of the patient while suctioning is taking place. Oxygen saturation was measured at five time points: baseline, first suctioning pass, second suctioning pass, third suctioning pass, and 5 mins post suctioning. Twenty-five ICU patients were randomized to each of the three methods. We assume ARMA(1, 1) correlation and apply both maximum likelihood and pairwise methods to estimate the parameters.

Table 8 provides the point estimates, standard errors and the p-values for both

Table 8. Parameter Estimates for the Oxygen Saturation Data

	PLE			MLE		
Parameter	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	95.892	0.5164	< 0.0001	95.896	0.4739	$< .0001$
New Method	-0.4153	0.7303	0.5695	-0.416	0.6702	0.5368
Manual Bagging	0.2806	0.7303	0.7008	0.288	0.6702	0.6687
ϕ	8.6627	1.6994	< 0.0001	7.7119	0.9514	$< .0001$
γ	0.7001	0.0530	< 0.0001	0.6601	0.0459	$< .0001$
ρ	1.0175	0.037	< 0.0001	1	0	.
-2ℓ	1600			1599.4		

the estimation methods. The estimates and standard errors are similar. The p-values of 0.5695 and 0.7008 for the new method and the manual bagging procedure show that there is no significant difference between the methods. Hence, all the three methods perform equivalently.

CHAPTER 3

ANALYSIS OF ANTEDEPENDENCE MODELS

In Chapter 2, we studied a class of correlation models which exhibits the stationary characteristics defined in terms of the time-series models. Stationary in combination with correlation models implies the equality of variances across time and the correlations depend on the absolute differences between the measurement times. We also discussed consequences of the stationary conditions on the positive definiteness of the correlation models in case of first-order autoregressive moving average. In this chapter, we study the correlation structures generated by antedependence models whose parameters are not constrained due to the restrictions imposed by the stationary conditions.

Antedependence models are one of the several large classes of available correlation models for longitudinal data. The most well-known members of the class are autoregressive models, but antedependence models are much more flexible and general. An interesting property of antedependence models is that they do not impose any stationary assumptions on the parameters unlike the stationary autoregressive counterparts. Thus, antedependence models differ from other models in that they allow the parameters to change over the course of the longitudinal study and as a result they are suitable for modeling longitudinal data that exhibits non-stationary characteristics. For instance, antedependence models can be used to accommodate correlations that not only depend on the lag but also on the time of observation. More general class of antedependence models are defined by considering heterogeneous variances for repeated measures. However, in our work we focus on the case where the variances are homogeneous. A very general class of antedependence models of higher orders exists and have been discussed in the literature. In this dissertation we focus only on the first-order antedependence models.

Much of the existing literature focuses on the parametric modeling of the longitudinal data that exhibits serial correlations using the correlation structures generated by the stationary time series models. As a result, relatively little attention was given to the more generalized antedependence structures in compared to their usefulness. Although, existing software packages permits one to use this structure to model the correlations, they however are prone to convergence problems. In this chapter, we will implement one such generalized structure and discuss an alternative method for estimating the parameters. This alternative method known as quasi-least squares, overcomes several drawbacks of the traditional methods of estimation. This chapter is organized as follows. We discuss the antedependence correlation model in Section 3.1. A motivating example is given in Section 3.2. The maximum likelihood and the quasi-least squares methods of estimation for the antedependence model are discussed in Section 3.3. Asymptotic theory and efficiency calculations are reported in Sections 3.4, 3.5 and 3.6. In Section 3.7, we analyze the data discussed in Section 3.2.

3.1 ANTEDEPENDENCE CORRELATION STRUCTURES

Modeling covariance or correlation structure parsimoniously is crucial to efficiently estimate the data's mean structure and its corresponding standard error. One such parsimonious correlation structure that can be considered for analyzing the longitudinal data is the antedependence correlation structure. Great flexibility is provided by the antedependence structure in modeling correlations. These structures allow the correlation to vary between the observations, which is one of the common scenarios that we encounter in longitudinal studies. The t dimensional first-order antedependence structure which accommodates such a distinctive feature is given by

$$R(\lambda) = R(\rho_1\,\rho_2 \cdots \rho_{t-1}) =$$

$$\begin{pmatrix} 1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2\rho_3 & \cdots & \rho_1\rho_2\cdots\rho_{t-1} \\ \rho_1 & 1 & \rho_2 & \rho_2\rho_3 & \cdots & \rho_2\rho_3\cdots\rho_{t-1} \\ \rho_1\rho_2 & \rho_2 & 1 & \rho_3 & \cdots & \rho_3\rho_4\cdots\rho_{t-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1\rho_2\cdots\rho_{t-1} & \rho_2\rho_3\cdots\rho_{t-1} & \rho_3\rho_4\cdots\rho_{t-1} & \rho_4\rho_5\cdots\rho_{t-1} & \cdots & 1 \end{pmatrix} \qquad (38)$$

From (38), we can see that the parameter vector that characterizes the correlation structure is $\lambda = (\rho_1, \rho_2, \rho_3, \ldots, \rho_{t-1})$. Thus, in the case of balanced data with t time points, there are $t-1$ correlation parameters. The correlation parameter ρ_j is the lag one correlation between observations at time-points j and $j+1$, that is,

$$\mathrm{Corr}(y_{ij}, y_{i(j+1)}) = \rho_j,$$

for $j = 1, 2, \ldots, t-1$ and $i = 1, 2, \ldots, n$. This manifests the non-stationary behavior of the antedependence models. The correlation between two observations depend on the time as well as on the time lag between them. For example, correlation between the observations from the subject i taken at first and second time points is ρ_1, that is, $\mathrm{Corr}(y_{i1}, y_{i2}) = \rho_1$, whereas the correlation between observations taken at second and third time points is given by ρ_2, that is, $\mathrm{Corr}(y_{i2}, y_{i3}) = \rho_2$, and similarly $\mathrm{Corr}(y_{i(t-1)}, y_{it}) = \rho_{t-1}$. Furthermore, higher lag correlations are completely determined by the lag one correlations and they are calculated by taking the product of the intervening lag one correlations. Thus, we have

$$\mathrm{Corr}(y_{il}, y_{ik}) = \rho_{lk} = \prod_{m=k}^{l-1} \rho_m \quad \text{for} \quad l > k = 1, 2, \ldots, t-1;\ i = 1, 2, \ldots, n.$$

As a result clearly the higher lag correlations also vary over time.

3.1.1 TIME SERIES MODEL FOR ANTEDEPENDENCE

The antedependence correlation structures arise from antedependence time series models. For example, a first-order antedependence correlation model arises from the first-order time series antedependence model given by

$$\zeta_s = \rho_{s-1}\zeta_{s-1} + \epsilon_s, \tag{39}$$

where ϵ_s is a sequence of uncorrelated random variables with mean zero and variance σ^2. Note that $k \geq 1$,

$$\mathrm{Corr}(\zeta_s,\ \zeta_{s+k}) = \rho_s \rho_{s+1} \rho_{s+2} \cdots \rho_{s+k-1}$$

which depends on the time point s. From the above model we can see that the number of autoregressive coefficients is one less than the total number of repeated observations. In addition, the autoregressive coefficients are unconstrained. Thus, with these properties, model (39) becomes highly inefficient in the time series context since the series length is usually very long. However, this model turns out to be much more general compared to its stationary counterparts in the context of longitudinal data. There is another definition of antedependence variables due to Gabriel (1962) and Macchiavelli and Arnold (1994).

Definition 3.1. *The sequence of random variables $Y_1, Y_2, \ldots, Y_t$ are said to be antedependent of order p, or AD(p) for $0 \leq p \leq t-1$, if Y_k, given at least p immediately preceding variables, is independent of all further preceding variables for $k = 1, 2, \ldots, t$.*

Several other equivalent but simpler definitions of antedependence variables exists in literature. However, definition 3.1 refers back to the Gabriel's definition. It is noteworthy to mention that the antedependence generally depends on the particular ordering of the variables. Therefore, antedependence conditions on a permuted set is not equivalent to partial independence conditions of the original indexed variables. Nevertheless, this doesn't lead to any major concern since the ordering in the longitudinal data is universal, namely chronological order.

The extreme case $p = 0$ refers to the mutual independence structure and the extreme case $p = t - 1$ is equivalent to full dependence. In addition, it follows from the definition that the AD(p) variables are nested,

$$AD(0) \subset AD(1) \subset AD(2) \subset \cdots \subset AD(t-1)$$

which means that if the random variables $Y_1, Y_2, \ldots, Y_t$ are AD(i) then by definition $Y_1, Y_2, \ldots, Y_t$ also satisfies the conditions for AD(j) for $0 \leq i < j \leq t - 1$.

Thus the antedependence models partition the set of all dependence structures depending on the order.

3.1.2 PROPERTIES OF THE ANTEDEPENDENCE MODELS

Antedependence models generalize the AR(1) model in the sense when $\rho_j = \rho$ for all j, the correlation matrix (38) reduces to the AR(1) correlation structure (3). The determinant of $R(\lambda)$ of the antedependence structure (38) is

$$|R(\lambda)| = \prod_{j=1}^{t-1} \left(1 - \rho_j^2\right). \tag{40}$$

The lower diagonal elements of $R^{-1}(\lambda) = (r^{ij})$ are given by

$$r^{ij} = \begin{cases} \dfrac{1}{(1-\rho_1^2)} & i = j = 1 \\ \dfrac{1}{(1-\rho_{t-1}^2)} & i = j = t \\ \dfrac{1-\rho_{j-1}^2\rho_j^2}{(1-\rho_{j-1}^2)(1-\rho_j^2)} & i = j \neq 1, t \\ \dfrac{-\rho_j}{(1-\rho_j^2)} & i = j+1 \\ 0 & i - j > 1. \end{cases} \tag{41}$$

In the special case $t = 4$, the inverse can be written explicitly as

$$R^{-1}(\lambda) = \begin{pmatrix} \dfrac{1}{1-\rho_1^2} & \dfrac{-\rho_1}{1-\rho_1^2} & 0 & 0 \\ \dfrac{-\rho_1}{1-\rho_1^2} & \dfrac{1-\rho_1^2\rho_2^2}{(1-\rho_1^2)(1-\rho_2^2)} & \dfrac{-\rho_2}{1-\rho_2^2} & 0 \\ 0 & \dfrac{-\rho_2}{1-\rho_2^2} & \dfrac{1-\rho_2^2\rho_3^2}{(1-\rho_2^2)(1-\rho_3^2)} & \dfrac{-\rho_3}{1-\rho_3^2} \\ 0 & 0 & \dfrac{-\rho_3}{1-\rho_3^2} & \dfrac{1}{1-\rho_3^2} \end{pmatrix}.$$

Thus the inverse is a tridiagonal matrix. The nonzero elements are in a closed form and they have a nice pattern. This nice form of the inverse makes it easy to implement alternative methods of estimation.

3.1.3 POSITIVE DEFINITE RANGE

It is easy to see that the antedependence correlation matrix (38) is positive definite if and only if $-1 < \rho_j < 1$ for $j = 1, 2 \ldots, (t-1)$. In practical longitudinal data analysis the occurrence of negative correlations is not common. Hence in this dissertation we restrict the range to $0 \leq \rho_j < 1$ for all j, even though this restriction is not necessary for the estimation procedures that we discuss in later sections.

3.2 MOTIVATING EXAMPLE

We discussed in Chapter 2, different methods of estimating the correlation parameters for several correlation structures. We assume these correlation structures are generated by stationary time series models such as ARMA(1,1), AR(1) and MA(1). However, these correlation models may not be proficient in case of longitudinal data exhibiting non-stationary correlations as illustrated in the example below.

Recall the cattle growth data presented in Table 3, Section 1.6.3. The data consists of weights (in kg) recorded at 11 sequential time points of cattle receiving two treatments, A and B, for intestinal parasites. Consider the weights of cattle receiving treatment A. Table 9 provides sample correlation matrix computed using the 30 observations.

Table 9. Sample Correlation Matrix for Cattle Growth Data

1										
0.824	1									
0.764	0.907	1								
0.658	0.844	0.925	1							
0.635	0.804	0.879	0.942	1						
0.585	0.741	0.846	0.914	0.943	1					
0.524	0.628	0.748	0.824	0.872	0.930	1				
0.529	0.667	0.771	0.837	0.893	0.942	0.932	1			
0.519	0.600	0.712	0.769	0.838	0.904	0.934	0.969	1		
0.475	0.584	0.699	0.734	0.798	0.865	0.884	0.943	0.964	1	
0.478	0.551	0.679	0.713	0.773	0.830	0.864	0.924	0.958	0.984	1

NOTE: Sample correlations for Treatment-A.

From Table 9, we note that the correlations decrease within the columns. Hence, using stationary autoregressive models seems to be an appropriate choice to model such data. However, on further exploration we can observe that same-lag correlations (within each sub-diagonal) are not constant but instead tend to increase over time violating the stationary condition. Thus, for this data the first-order antedependence correlation structure defined in (38) seems to be an appropriate model. We revisit this data analysis in Section 3.7.

In the following sections we discuss the maximum likelihood and quasi-least squares methods for estimating the regression and the correlation parameters in case of antedependence correlation models. In addition, we briefly discuss the complexities involved in using the pairwise likelihood method introduced in Chapter 2.

3.3 ESTIMATION PROCEDURES

We consider the setup described in Section 1.4.1. Data consists of responses $Y_i = (y_{i1}, y_{i2}, \ldots, y_{it})'$ and corresponding covariates $X_i = (x_{i1}, x_{i2}, \ldots, x_{it})'$ on subject i, for $i = 1, 2, \ldots, n$. The subjects are assumed to be independent. We assume $E(Y_i) = \mu_i = X_i\beta$ and $Cov(Y_i) = \phi R(\lambda)$, where the correlation matrix $R(\lambda)$ is of the form given by (39). The estimates of β and ϕ can be obtained using the formulas in (9). In this section we focus on two methods of estimating the correlation parameters in the antedependence correlation matrix, namely, the maximum likelihood and the quasi-least squares method. A detailed discussion on each of the estimating methods are given in the following sections along with the drawbacks of pairwise likelihood method in case of antedependence correlation structure. We also evaluate the performance of the quasi-least squares with respect to the optimal maximum likelihood method of estimation.

3.3.1 MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood estimation procedure for antedependence correlation structure is similar to the maximum likelihood method for ARMA (1, 1) correlation structure given in Section 2.3.1. Using the determinant (40) and the inverse (41) of the antedependence structure $R(\lambda)$, we can reduce the Gaussian loglikelihood in (8) to

$$\begin{aligned} \ell(\theta) &= \text{const} - \frac{nt}{2}\log(\phi) - \frac{n}{2}\sum_{j=1}^{t-1}\log\left(1-\rho_j^2\right) - \frac{1}{2\phi}\sum_{i=1}^{n}\left\{\frac{1}{1-\rho_1^2}\, z_{i1}^2 \right. \\ &\quad \left. -2\sum_{j=1}^{t-1}\frac{\rho_j}{1-\rho_j^2}\, z_{ij}z_{i(j+1)} + \sum_{j=1}^{t-2}\frac{1-\rho_j^2\rho_{j+1}^2}{(1-\rho_j^2)(1-\rho_{j+1}^2)}\, z_{i(j+1)}^2 + \frac{1}{1-\rho_{t-1}^2}\, z_{it}^2\right\} \end{aligned} \tag{42}$$

where $Z_i = (Y_i - X_i\beta) = (z_{i1}, z_{i2}, \ldots, z_{it})'$ for $i = 1, 2, \ldots, n$.

To maximize $\ell(\theta)$ with respect to ρ_j, we take the derivative of (42) with respect to each ρ_j and equate it to zero. The resulting score equation reduces to a cubic equation in ρ_j given by

$$-n\phi\rho_j^3 + \left[\sum_{i=1}^{n} z_{ij} z_{i(j+1)}\right] \rho_j^2 - \left[\sum_{i=1}^{n} \left(z_{ij}^2 + z_{i(j+1)}^2\right) - n\phi\right] \rho_j + \left[\sum_{i=1}^{n} z_{ij} z_{i(j+1)}\right] = 0. \tag{43}$$

The estimate of each correlation parameter ρ_j can be obtained solving the above cubic polynomial and it does not depend on the estimate of $\rho_{j'}$ for $j' \neq j$. Thus the maximum likelihood final estimates are obtained using an iterative procedure. We start with an initial value for β and ϕ and use them to solve (43) for ρ_j resulting in $\widehat{\rho}_j$. Update the value of $\widehat{\beta}$ and $\widehat{\phi}$ using $\widehat{\rho}_j$ in (9). Iterate this process until convergence and obtain $\widehat{\theta}_{ML} = (\widehat{\beta}, \widehat{\phi}, \widehat{\lambda})$ as the final maximum likelihood estimate of θ.

3.3.2 DRAWBACKS OF THE PAIRWISE LIKELIHOOD METHOD

In this section we briefly illustrate the shortcomings of using the pairwise likelihood method introduced in Chapter 2 for estimating the antedependence correlation parameters. As described in Section 2.3.2 the pairwise likelihood estimators for correlation parameters are obtained by maximizing the likelihood functions constructed using the bivariate random vectors formed by considering pairs that involve the corresponding correlation parameter. For the antedependence correlation structure in (38), it can be noted that the correlations are characterized by product of lag-one correlations. Hence, to estimate the correlation parameter ρ_j, we consider the pairs constructed using the components of $Z_i = Y_i - X_i\beta$ that involve ρ_j. For illustration purposes, we describe the pairwise likelihood method for $t = 4$ case. Note that, when $t = 4$, the antedependence correlation matrix is

$$R(\lambda) = R(\rho_1, \rho_2, \rho_3) = \begin{pmatrix} 1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2\rho_3 \\ \rho_1 & 1 & \rho_2 & \rho_2\rho_3 \\ \rho_1\rho_2 & \rho_2 & 1 & \rho_3 \\ \rho_1\rho_2\rho_3 & \rho_2\rho_3 & \rho_3 & 1 \end{pmatrix}.$$

To estimate ρ_1, we need to consider the following three pairs (z_{i1}, z_{i2}), (z_{i1}, z_{i3}) and (z_{i1}, z_{i4}) of observations. Since $\text{Corr}(z_{i1}, z_{i2}) = \rho_1$, $\text{Corr}(z_{i1}, z_{i3}) = \rho_1\rho_2$, and $\text{Corr}(z_{i1}, z_{i4}) = \rho_1\rho_2\rho_3$, the pairwise likelihood constructed from these pairs can be used to derive a score equation for ρ_1. Similarly we can derive a score equation for ρ_2 constructed by the pairwise likelihood using the pairs (z_{i2}, z_{i3}), (z_{i2}, z_{i4}) and (z_{i1}, z_{i3}). Finally we can get a score equation for ρ_3 using the pairs (z_{i3}, z_{i4}), (z_{i2}, z_{i4}) and (z_{i1}, z_{i4}). The simultaneous solution of these three equations will give estimates of the three parameters ρ_1, ρ_2 and ρ_3.

We can quickly notice two problems with this approach. First, the number of pairs required to construct the likelihood function for each ρ_j depends and grows with the value of t. As a consequence, it becomes cumbersome to construct the likelihoods for large values of t. Second, the score equation for each ρ_j involves all the other correlation parameters and there is a need to implement an iterative procedure to solve all the t equations simultaneously. The number of iterations required to solve for the correlation parameter estimates also increases with t thus compromising the advantage of computational ease associated with the pairwise likelihood method. Justification of this failure can be attributed to the non-stationarity characterization of antedependence time series models as described in Section 3.1.1. However, one can explore the pairwise likelihood method in case of first-order antedependence models by considering only the bivariate distributions of successive pairs to construct the likelihood function.

Instead of pairwise likelihood method, we use quasi-least squares method to estimate the correlation parameters in antedependence model, as an alternative to maximum likelihood. Quasi-least squares method was developed in Chaganty (1997) and Chaganty and Shults (1999) and we describe this method in the following section for antedependence correlation structure.

3.3.3 QUASI-LEAST SQUARES

The quasi-least squares method uses the quasi-loglikelihood function which is defined as

$$Q(\theta) = \sum_{i=1}^{n} (Y_i - X_i\beta)' R^{-1}(\lambda) (Y_i - X_i\beta) = \operatorname{tr}\left(R^{-1}(\lambda) Z_n\right),$$

where $Z_n = \sum_{i=1}^{n} (Y_i - X_i\beta)(Y_i - X_i\beta)' = \sum_{i=1}^{n} Z_i Z_i'$. When $R(\lambda)$ is the antedependence structure, using the inverse formula given in (41) we can reduce the above expression to

$$Q(\theta) = \sum_{i=1}^{n} \left\{ \frac{1}{1-\rho_1^2}\, z_{i1}^2 - 2 \sum_{j=1}^{t-1} \frac{\rho_j}{1-\rho_j^2}\, z_{ij} z_{i(j+1)} + \sum_{j=1}^{t-2} \frac{1-\rho_j^2 \rho_{j+1}^2}{(1-\rho_j^2)(1-\rho_{j+1}^2)}\, z_{i(j+1)}^2 \right.$$
$$\left. + \frac{1}{1-\rho_{t-1}^2}\, z_{it}^2 \right\}, \tag{44}$$

where $Z_i = (Y_i - X_i\beta) = (z_{i1}, z_{i2}, \ldots, z_{it})'$ for $i = 1, 2, \ldots, n$.

Recall from Section 1.4.3, that the quasi-least squares estimation method estimates the correlation parameters in 2 steps. In step 1 we minimize the quasi-loglikelihood function (44) with respect to each ρ_j. Equating the derivate of (44) to zero we get

$$\begin{aligned} \frac{\partial Q(\theta)}{\partial \rho_j} &= \operatorname{tr}\left(\frac{\partial R^{-1}(\lambda)}{\partial \rho_j}\, Z_n \right) \\ &= \sum_{i=1}^{n} \frac{2}{(1-\rho_j^2)^2} \left\{ \rho_j^2\, z_{ij} z_{i(j+1)} - \rho_j \left(z_{ij}^2 + z_{i(j+1)}^2 \right) + z_{ij} z_{i(j+1)} \right\} = 0, \end{aligned} \tag{45}$$

for $j = 1, 2, \ldots, t-1$. The quadratic equation is positive at $\rho_j = -1$ and negative at $\rho_j = 1$. Hence there is a unique solution to the equation (45) in the interval $(-1,\ 1)$, and this root is the step 1 estimate of ρ_j given by

$$\widetilde{\rho}_j = \frac{\sum_{i=1}^{n} \left(z_{ij}^2 + z_{i(j+1)}^2 \right) - \sqrt{\left[\sum_{i=1}^{n} \left(z_{ij}^2 + z_{i(j+1)}^2 \right) \right]^2 - 4 \left[\sum_{i=1}^{n} z_{ij} z_{i(j+1)} \right]^2}}{2 \left(\sum_{i=1}^{n} z_{ij} z_{i(j+1)} \right)}. \tag{46}$$

As shown in Chaganty and Shults (1999) the equations (45) are biased at the initial estimates $\widetilde{\rho}_j$ since

$$\begin{aligned} E\left[\operatorname{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_j}\, Z_n \right) \right] &= \operatorname{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_j}\, E(Z_n) \right) \\ &= \phi \operatorname{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_j}\, R(\lambda) \right) \neq 0, \end{aligned} \tag{47}$$

where $\widetilde{\lambda} = (\widetilde{\rho}_1, \ldots, \widetilde{\rho}_j, \ldots, \widetilde{\rho}_{t-1})$. In step 2, the QLS method obtains an asymptotically unbiased estimate of λ by solving the equations

$$\operatorname{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_j} R(\lambda)\right) = 0 \qquad \text{for} \quad j = 1, 2, \ldots, t-1. \tag{48}$$

The above equation can be written in a closed form

$$\frac{2\widetilde{\rho}_j - \left(1 + \widetilde{\rho}_j^{\,2}\right)\rho_j}{\left(1 - \widetilde{\rho}_j^{\,2}\right)^2} = 0, \tag{49}$$

and the solution is given by

$$\widehat{\rho}_j = \frac{2\widetilde{\rho}_j}{\left(1 + \widetilde{\rho}_j^{\,2}\right)}, \tag{50}$$

for $j = 1, 2, \ldots, t-1$. In terms of the z_{ij}'s we have

$$\widehat{\rho}_j = \frac{2\sum_{i=1}^{n} z_{ij} z_{i(j+1)}}{\sum_{i=1}^{n} \left(z_{ij}^2 + z_{i(j+1)}^2\right)}. \tag{51}$$

One of the advantages of the QLS method is that the estimates of β and ρ_j do not depend on ϕ. To summarize, the iterative procedure for obtaining the QLS estimates of β, ϕ and ρ_j for $j = 1, 2, \ldots, t-1$ can be described as follows:

Step 1: Compute Z_n using an initial value of $\widetilde{\beta}$ for β.

Step 2: Compute $\widehat{\lambda} = (\widehat{\rho}_1, \widehat{\rho}_2, \ldots, \widehat{\rho}_{t-1})$ using (51).

Step 3: Update the value of β, say $\widehat{\beta}$, using $\widehat{\lambda}$ and (9).

Step 4: Replace $\widetilde{\beta}$ with $\widehat{\beta}$ and repeat Steps 1 - 3 until convergence.

Step 5: At the end, compute $\widehat{\phi}$ using the formula (9). The QLS estimate of θ is given by $\widehat{\theta}_{\text{QL}} = (\widehat{\beta}, \widehat{\phi}, \widehat{\lambda})$.

Clearly the big advantage of the QLS method is that there are closed form expressions for the estimates of all the parameters in the model. In addition, even though we have stated that it is a two step procedure, it is essentially a one step for the antedependence model since $\widehat{\rho}_j$ can be computed directly from the Z_i's.

3.4 ASYMPTOTIC THEORY

We compare the performance of quasi-least squares estimates with maximum likelihood estimates for antedependence correlation models using asymptotic relative efficiency criterion. In order to calculate these efficiencies, we need the asymptotic variance of the estimates from each of the estimating procedures. The asymptotic variance of quasi-least squares and maximum likelihood estimators can be derived using the asymptotic theory described in Section 1.5.

3.4.1 MAXIMUM LIKELIHOOD

The asymptotic variances and covariances for the maximum likelihood estimators are obtained as the inverse of Fisher information matrix $\mathcal{I}_\ell(\theta)$, which is defined in Section 1.5.1. We derive $\mathcal{I}_\ell(\theta)$ for the antedependence correlation structure. For ease of notation we use R for $R(\lambda)$ in the following expressions. The general expression for Fisher information matrix in case of for any given t is given below. For convenience, only the upper diagonal elements are provided, since the matrix is symmetric.

$$\mathcal{I}_\ell(\theta) = \begin{pmatrix} I(\beta) & 0 & 0 & 0 & \cdots & 0 \\ & I(\phi) & I(\phi,\rho_1) & I(\phi,\rho_2) & \cdots & I(\phi,\rho_{t-1}) \\ & & I(\rho_1) & 0 & \cdots & 0 \\ & & & I(\rho_2) & \cdots & 0 \\ & & & & \ddots & \vdots \\ & & & & & I(\rho_{t-1}) \end{pmatrix} \qquad (52)$$

where the individual elements

$$\begin{aligned} I(\beta) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \beta^2}\right) &= \frac{1}{\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i \\ I(\phi) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \phi^2}\right) &= \frac{nt}{2\phi^2} \\ I(\phi,\rho_j) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \rho_j \partial \phi}\right) &= \frac{-n\rho_j}{(1-\rho_j^2)} \\ I(\rho_j) &= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \rho_j^2}\right) &= \frac{-n(1+\rho_j^2)}{(1-\rho_j^2)^2} \end{aligned}$$

for $j = 1, 2, \ldots, t-1$. When $t = 4$ we have

$$\mathcal{I}_\ell(\theta) = \begin{pmatrix} \frac{1}{\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i & 0 & 0 & 0 & 0 \\ & \frac{nt}{2\phi^2} & \frac{-n\rho_1}{(1-\rho_1^2)} & \frac{-n\rho_2}{(1-\rho_2^2)} & \frac{-n\rho_3}{(1-\rho_3^2)} \\ & & \frac{-n(1+\rho_1^2)}{(1-\rho_1^2)^2} & 0 & 0 \\ & & & \frac{-n(1+\rho_2^2)}{(1-\rho_2^2)^2} & 0 \\ & & & & \frac{-n(1+\rho_3^2)}{(1-\rho_3^2)^2} \end{pmatrix}.$$

Clearly $\mathcal{I}_\ell^{-1}(\theta)$ is the asymptotic covariance matrix for the maximum likelihood estimates.

3.4.2 QUASI-LEAST SQUARES

The asymptotic variances for the quasi-least squares estimators can be obtained using the theory of unbiased estimation equations and the associated Godambe information matrix, $\mathcal{G}(\theta)$ stated in Section 1.5.2. Recall that,

$$\mathcal{G}(\theta) = D(\theta)\, M^{-1}(\theta)\, (D(\theta))',$$

where $D(\theta) = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_i(\theta)}{\partial \theta}\right)$ and $M(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(h_i(\theta))$. For quasi-least squares procedure the vector $h_i(\theta)$ involved in the unbiased estimating equation for θ is given by

$$h_i(\theta) = \left(g_{i1}(\theta),\, g_{i2}(\theta),\, h_{i1}(\theta), \ldots,\, h_{i(t-1)}(\theta)\right)'$$

where $g_{i1}(\theta)$ and $g_{i2}(\theta)$ are the estimating equations corresponding to the regression coefficients β and the residual variance ϕ respectively. Expressions for $g_{i1}(\theta)$ and $g_{i2}(\theta)$ are given in Section 2.4.2. Similarly, $h_{ij}(\theta)$ is the unbiased estimating equation for the correlation parameter ρ_j and it is given by

$$h_{ij}(\theta) = \mathrm{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \widetilde{\rho}_j}\,(Z_i Z_i' - \phi R(\lambda))\right),$$

where $\widetilde{\lambda}$ is the solution to the equation

$$\operatorname{tr}\left(\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_j} R(\lambda)\right) = 0,$$

for $j = 1, 2, \ldots, t-1$. The Godambe information matrix for the quasi-least squares estimate $\widehat{\theta}_{\text{QL}}$ is

$$\mathcal{G}_{\text{QL}}(\theta) = D_{\text{QL}}(\theta) M_{\text{QL}}^{-1}(\theta)(D_{\text{QL}}(\theta))',$$

where the matrix $D_{\text{QL}}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial h_i(\theta)}{\partial \theta}\right)$ has the structure

$$D_{\text{QL}}(\theta) = \begin{pmatrix} D(\beta) & 0 & 0 & 0 & \cdots & 0 \\ 0 & D(\phi) & D(\phi,\rho_1) & D(\phi,\rho_2) & \cdots & D(\phi,\rho_{t-1}) \\ 0 & 0 & D(\rho_1) & 0 & \cdots & 0 \\ 0 & 0 & 0 & D(\rho_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & D(\rho_{t-1}) \end{pmatrix}. \tag{53}$$

The individual elements of $D_{\text{QL}}(\theta)$ are

$$D(\beta) = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial g_{i1}(\theta)}{\partial \beta}\right) = \frac{1}{n\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i$$

$$D(\phi) = -\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{\partial g_{i2}(\theta)}{\partial \phi}\right) = \frac{t}{2\phi^2}$$

$$D(\phi,\rho_j) = -E\left(\frac{\partial^2 g_{i2}(\theta)}{\partial \rho_j \partial \phi}\right) = \frac{-\rho_j}{\phi(1-\rho_j^2)}$$

$$D(\rho_j) = -E\left(\frac{\partial^2 h_{i1}(\theta)}{\partial \rho_j^2}\right) = \frac{-2\phi(1+\widetilde{\rho}_j^2)}{(1-\widetilde{\rho}_j^2)^2}$$

where $\widetilde{\rho}_j$ is the Step 1, QLS estimate of ρ_j. Similarly, $M_{\text{QL}}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \operatorname{Cov}(h_i(\theta))$ has the following form

$$M_{\text{QL}}(\theta) = \begin{pmatrix} M(\beta) & 0 & 0 \\ 0 & M(\phi) & 0 \\ 0 & 0 & M(\lambda) \end{pmatrix}, \tag{54}$$

where

$$M(\beta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(g_{i1}) = \frac{1}{n\phi}\sum_{i=1}^{n} X_i' R^{-1} X_i$$

$$M(\phi) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(g_{i2}) = \frac{t}{2\phi^2}$$

and $M(\lambda) = (m_{lk})$ of order $(t-1)$ is the covariance matrix of h_{ij}'s. The elements are

$$m_{lk} = \begin{cases} \dfrac{2\phi^2\ \rho_k(\rho_{k+1}\cdots\rho_{l-1})^2\rho_l}{(1+\tilde{\rho}_k^2)(1+\tilde{\rho}_l^2)} & \text{for} \quad l > k \\ \dfrac{4\phi^2}{(1+\tilde{\rho}_l^2)} & \text{for} \quad l = k \\ m_{kl} & \text{for} \quad l < k. \end{cases}$$

When $t = 4$, the matrices above can be written as

$$D_{\mathrm{QL}}(\theta) = \begin{pmatrix} \frac{1}{n\phi}\sum_{i=1}^{n} X_i'R^{-1}X_i & 0 & 0 & 0 & 0 \\ 0 & \frac{t}{2\phi^2} & \frac{-\rho_1}{\phi(1-\rho_1^2)} & \frac{-\rho_2}{\phi(1-\rho_2^2)} & \frac{-\rho_3}{\phi(1-\rho_3^2)} \\ 0 & 0 & \frac{-2\phi(1+\tilde{\rho}_1^2)}{(1-\tilde{\rho}_1^2)^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{-2\phi(1+\tilde{\rho}_2^2)}{(1-\tilde{\rho}_3^2)^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{-2\phi(1+\tilde{\rho}_3^2)}{(1-\tilde{\rho}_3^2)^2} \end{pmatrix}$$

and $M_{\mathrm{QL}}(\theta) =$

$$\begin{pmatrix} \frac{1}{n\phi}\sum_{i=1}^{n} X_i'R^{-1}X_i & 0 & 0 & 0 & 0 \\ 0 & \frac{t}{2\phi^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{4\phi^2}{(1+\tilde{\rho}_1^2)} & \frac{2\phi^2\ \rho_1\rho_2}{(1+\tilde{\rho}_1^2)(1+\tilde{\rho}_2^2)} & \frac{2\phi^2\ \rho_1\rho_2^2\rho_3}{(1+\tilde{\rho}_1^2)(1+\tilde{\rho}_3^2)} \\ 0 & 0 & \frac{2\phi^2\ \rho_1\rho_2}{(1+\tilde{\rho}_1^2)(1+\tilde{\rho}_2^2)} & \frac{4\phi^2}{(1+\tilde{\rho}_2^2)} & \frac{2\phi^2\ \rho_2\rho_3}{(1+\tilde{\rho}_2^2)(1+\tilde{\rho}_3^2)} \\ 0 & 0 & \frac{2\phi^2\ \rho_1\rho_2^2\rho_3}{(1+\tilde{\rho}_1^2)(1+\tilde{\rho}_3^2)} & \frac{2\phi^2\ \rho_2\rho_3}{(1+\tilde{\rho}_2^2)(1+\tilde{\rho}_3^2)} & \frac{4\phi^2}{(1+\tilde{\rho}_3^2)} \end{pmatrix}.$$

3.5 ASYMPTOTIC EFFICIENCY COMPARISONS

In this section we present the asymptotic relative efficiencies of the quasi-least squares estimator with respect to the maximum likelihood estimators for the antedependence model. These asymptotic relative efficiencies are calculated by taking the ratio of diagonal elements of the covariance matrices $\mathcal{I}_\ell^{-1}(\theta)$ and $\mathcal{G}_{\mathrm{QL}}^{-1}(\theta)$. Figure 7 has visualization of the asymptotic relative efficiency of ρ_1 for all values of ρ_1 and ρ_2 in the unit interval when $t = 3$ and $\phi = 3$. Clearly, the efficiency surface shows that the quasi-least squares estimate is a good competitor for the optimal maximum likelihood estimator over the entire positive definite range. Selected values of the asymptotic relative efficiency of ρ_1 for several values of ρ_2 are in Table 10. We observe that the ARE of ρ_1 is very high over a wide range of ρ_1 and ρ_2, indicating that the quasi-least squares estimator variance is almost as small as that of the maximum likelihood estimator. Only for extreme values of ρ_1, that is, values close to the boundary $(-1, -0.95)$, the efficiency of the quasi-least squares method is low when compared to the maximum likelihood method. But the efficiency does not fall below 0.85.

Similarly, the AREs of ρ_2 when $t = 3$ for several values of ρ_1 are in Table 11 when $t = 3$ and $\phi = 3$. From the table values we can see that the ARE of ρ_2 is very high over a wide range of ρ_1 and ρ_2, indicating that the quasi-least squares estimator once again estimates even the correlation parameter ρ_2 with as good precision as the maximum likelihood. Only for extreme values of ρ_2 close to the boundary $(0.95, 1)$, we see some drop in the efficiency but it does not drop below 0.85. The surface plot of the efficiency is in Figure 8. The conclusions drawn regarding the ARE for ρ_2 from Table 11 can be seen clearly in this Figure 8.

Next, the asymptotic relative efficiencies of the estimates for the parameter ϕ are given in Table 12 and also in Figure 9. The efficiencies presented here are not similar to the efficiencies calculated in other cases. Here we fix the value of ϕ and then study the efficiencies with respect to values of correlation parameters since variance of ϕ depends on λ as explained in Section 3.4.2. Both the Table 12 and the Figure 9 show the efficiency of ϕ is high as expected since the functional form of $\widehat{\phi}$ is the same for both the methods.

Finally in Table 13 we give the asymptotic relative efficiencies of quasi-least

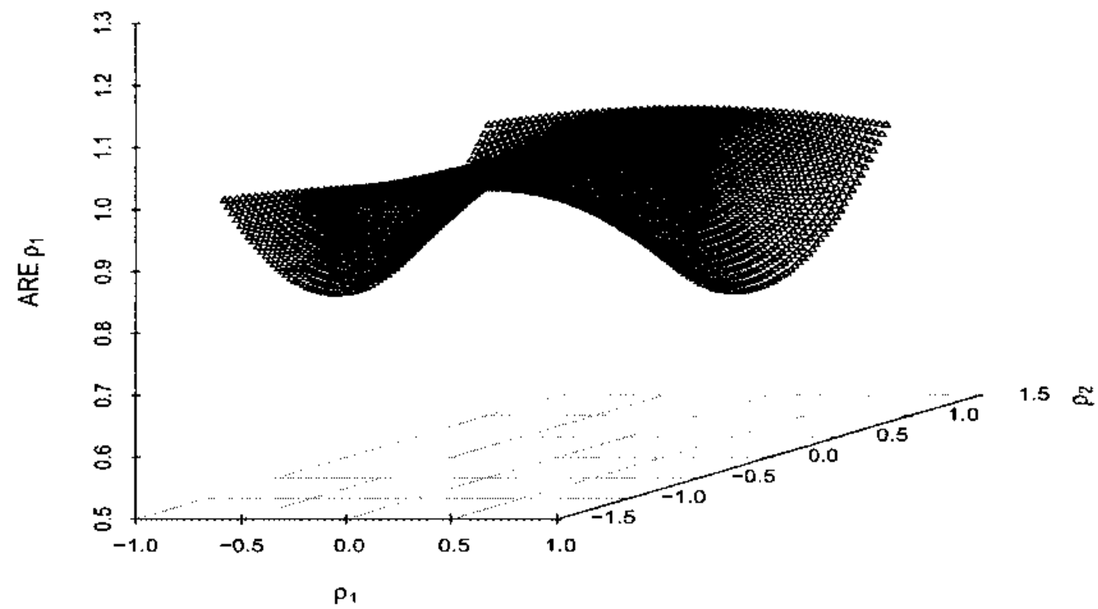


Figure 7. ARE of ρ_1 for $t = 3$ and $\phi = 3$

squares estimates of the three correlation parameters and the scale parameter for $t = 4$. All the efficiency calculations are done using $\phi = 3$. The efficiency pattern that we observed for $t = 3$ seems to hold even for $t = 4$ and we believe this will be the same for large values of t as well. Thus in conclusion we can say that the quasi-least squares method is a good competitor to the maximum likelihood method in case of antedependence correlation models.

3.6 SMALL SAMPLE EFFICIENCIES

In this section we study the small-sample performance of the two estimation procedures, the maximum likelihood and the quasi-least squares, assuming that the

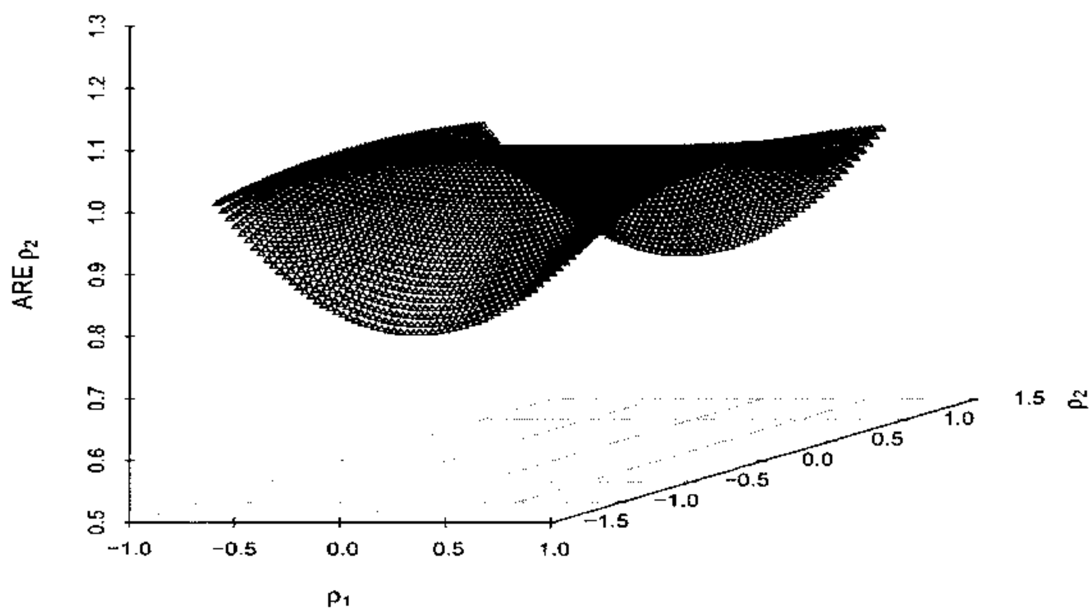


Figure 8. ARE of ρ_2 for $t = 3$ and $\phi = 3$

longitudinal data comes from a multivariate distribution with antedependence correlation structure. In the simulations we took the number of repeated measurements t to be 3. This will enable us to create three dimensional surface plots of the efficiencies because there are only two correlation parameters, ρ_1 and ρ_2.

To evaluate the small-sample performance, we fix the sample size $n = 30$ and set the scale parameter $\phi = 3$. In the simulation model we considered two covariates. The values for the first covariate x_{ij1} are simulated from uniform distribution between 0 and 1. The second covariate x_{ij2} is dichotomous taking values 0 and 1 with equal probability. We fixed the true regression coefficients as the intercept $\beta_0 = 22.5$, the coefficient of x_{ij1} as $\beta_1 = 2.5$ and $\beta_2 = 0.5$ for the coefficient of x_{ij2}. Using these as true parameter values and the covariates we simulated 1000 sets of samples consisting of 30 observations from the three dimensional multivariate normal distribution with

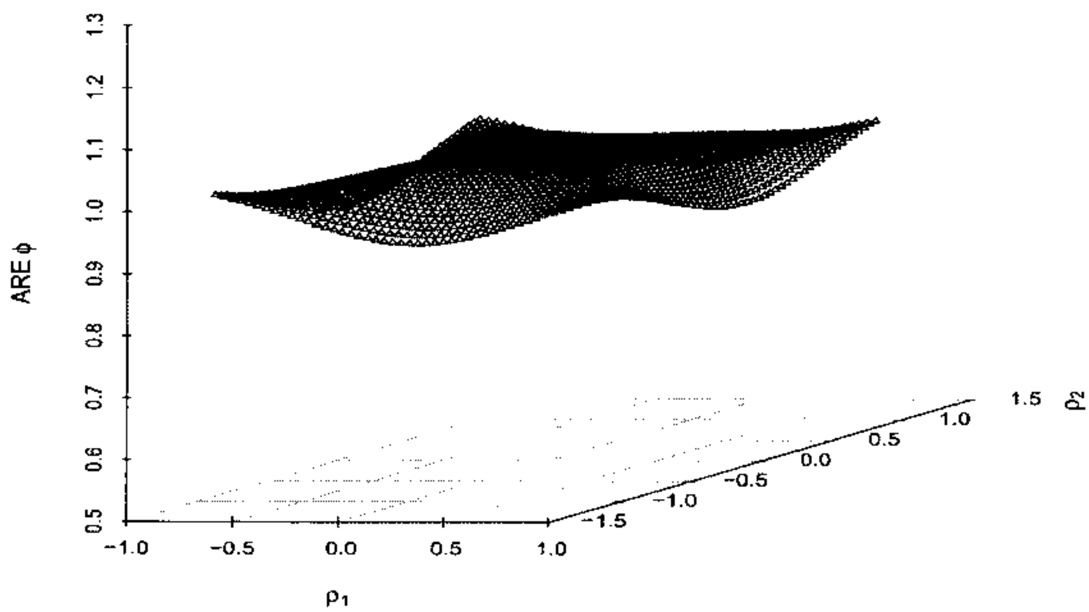


Figure 9. ARE of ϕ for $t = 3$

mean that is specified by the regression parameter and the generated covariates and the antedependence correlation structure for several values of the two correlation parameters ρ_1 and ρ_2 in the positive definite range. For each set of simulated data, we obtained estimates of the regression and the correlation parameters using both maximum likelihood and the quasi-least squares. Since the two methods differ in the estimation of the correlation parameters, we calculated the mean square errors (MSE) for the two correlation parameters ρ_1 and ρ_2 and the relative efficiencies (RE) as the ratio

$$\mathrm{RE} = \frac{\text{MSE of ML estimator}}{\text{MSE of QLS estimator}}.$$

If the ratio is more than 1, we can conclude that the QLS method performs better than the maximum likelihood method and vice versa.

The relative efficiencies (RE) of the QLS estimates of the correlation parameters are presented in Table 14. Some efficiency values in Table 14 are greater than one indicating that the QLS estimates are more efficient than the maximum likelihood estimator for some values in the correlation parameter space. In particular the efficiency of the QLS estimate of ρ_1 is greater than 1 for small values of ρ_1 irrespective of the true value of ρ_2. However the efficiencies are low for large positive and large negative values of ρ_1, that is, at the extreme values of ρ_1. Similarly the relative efficiency of the QLS estimate of the correlation parameter ρ_2 is more than one when $\rho_1 = 0.3$ irrespective of the value of ρ_2. For some values in the parameter space the QLS is less efficient when compared to the maximum likelihood estimate but in case the efficiency falls below 0.71 and for positive values of ρ_1 and ρ_2 it does not fall below 0.75. The relative efficiencies of the QLS estimates of the two correlation parameters ρ_1 and ρ_2 are plotted in Figures 10 and 11 respectively over the entire positive definite range of the correlation matrix.

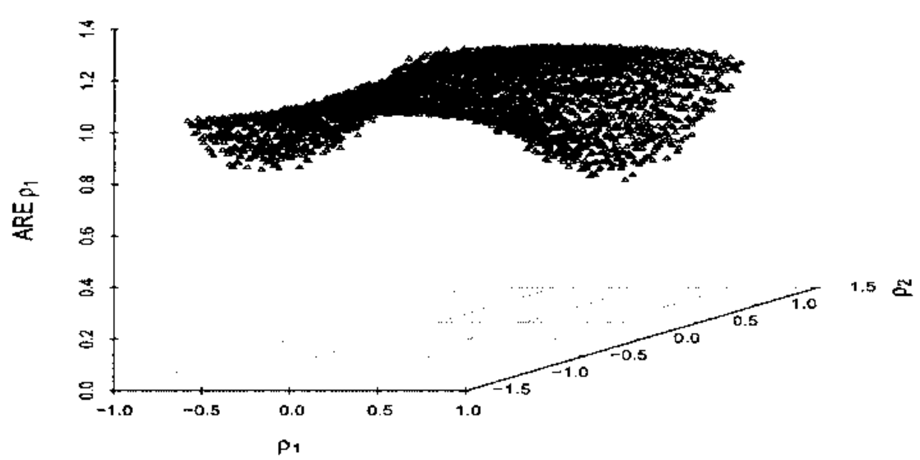


Figure 10. ρ_1 RE for QLS and MLE for $t = 3$

Table 15 displays the small-sample efficiencies of the QLS estimates of the three correlation parameters in the antedependence model for $t = 4$. Once again the values in the table show that QLS outperforms the maximum likelihood in some cases and it is a good competitor in other cases. In the next section we revisit the analysis of the cattle growth data discussed in Section 3.2.

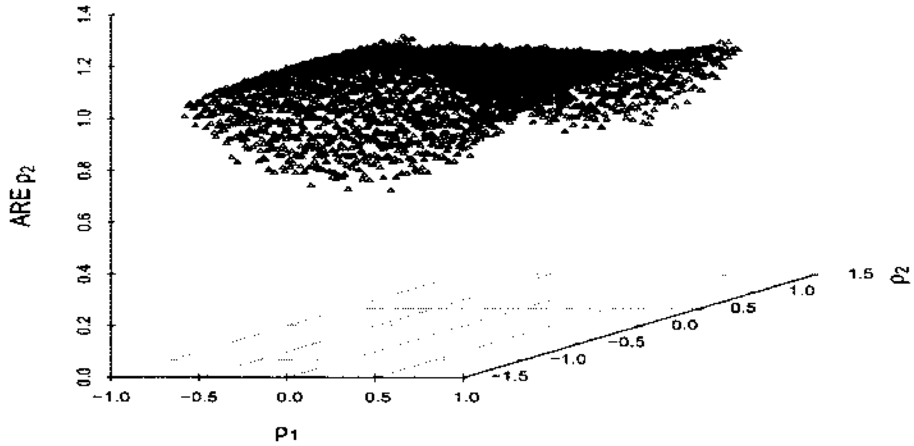


Figure 11. ρ_2 RE for QLS and MLE for $t = 3$

3.7 ANALYSIS OF CATTLE GROWTH DATA

To illustrate the application of two estimation methods, in this section we present the analysis of the continuous longitudinal data that we discussed in Section 1.6.3. In this study sixty cattle are evenly assigned randomly to two treatments A and B. Their weights were taken at eleven different time points. Table 3 displays a subset of the cattle growth data. The main interest in this research study is to examine whether there is a significant difference between the treatment groups A and B. As discussed in Section 3.2 we assume first-order antedependence correlation structure to model the dependencies among the weights. Since the responses are collected at eleven time-points, under the assumed correlation model there are ten different correlation parameters. We used the quasi-least squares and the maximum likelihood methods to estimate the parameters. Table 16 contains the point estimates, standard errors and the p-values for both the estimation methods. From Table 16, we can observe that the two methods give results that are in agreement. The p-value of 0.9202 for the treatment group A implies that there is no significant difference between the two groups regarding the weights of the cattle.

Table 10. Asymptotic Relative Efficiency of ρ_1 for Antedependence

ρ_1	ρ_2	ARE ρ_1	ρ_1	ρ_2	ARE ρ_1
-0.95	-0.95	0.9783	0.005	0.1	1
	-0.7	0.8836		0.5	1
	-0.3	0.7900		0.8	1
	0.005	0.7690	0.3	-0.95	0.9978
	0.1	0.7711		-0.7	0.9870
	0.5	0.8279		-0.3	0.9742
	0.8	0.9183		0.005	0.9709
-0.75	-0.95	0.9861		0.1	0.9712
	-0.7	0.9241		0.5	0.9797
	-0.3	0.8579		0.8	0.9912
	0.005	0.8421	0.6	-0.95	0.9911
	0.1	0.8439		-0.7	0.9500
	0.5	0.8851		-0.3	0.9042
	0.8	0.9473		0.005	0.8929
-0.5	-0.95	0.9938		0.1	0.8941
	-0.7	0.9647		0.5	0.9233
	-0.3	0.9314		0.8	0.9656
	0.005	0.9231	0.85	-0.95	0.9823
	0.1	0.9240		-0.7	0.9045
	0.5	0.9455		-0.3	0.8246
	0.8	0.9759		0.005	0.8059
0.005	-0.95	1		0.1	0.8080
	-0.7	1		0.5	0.8571
	-0.3	1		0.8	0.9333
	0.005	1		0.95	0.9823

NOTE: The parameter values are $t = 3$ and $\phi = 3$.

Table 11. Asymptotic Relative Efficiency of ρ_2 for Antedependence

ρ_1	ρ_2	ARE ρ_2	ρ_1	ρ_2	ARE ρ_2
-0.95	-0.95	0.9783	0.005	0.1	0.9967
	-0.7	0.9879		0.5	0.9231
	-0.3	0.9978		0.8	0.8242
	0.005	1	0.3	-0.95	0.7900
	0.1	0.9998		-0.7	0.8739
	0.5	0.9938		-0.3	0.9742
	0.8	0.9843		0.005	1
-0.75	-0.95	0.9005		0.1	0.9971
	-0.7	0.9433		0.5	0.9314
	-0.3	0.9891		0.8	0.8414
	0.005	1	0.6	-0.95	0.8531
	0.1	0.9988		-0.7	0.9146
	0.5	0.9702		-0.3	0.9831
	0.8	0.9272		0.005	1
-0.5	-0.95	0.8279		0.1	0.9981
	-0.7	0.8984		0.5	0.9545
	-0.3	0.9797		0.8	0.8914
	0.005	1	0.85	-0.95	0.9373
	0.1	0.9977		-0.7	0.9647
	0.5	0.9455		-0.3	0.9933
	0.8	0.8713		0.005	1
0.005	-0.95	0.7690		0.1	0.9993
	-0.7	0.8596		0.5	0.9817
	-0.3	0.9709		0.8	0.9545
	0.005	1		0.95	0.9373

NOTE: The parameter values are $t = 3$ and $\phi = 3$.

Table 12. Asymptotic Relative Efficiency of ϕ for Antedependence

ρ_1	ρ_2	ARE ϕ	ρ_1	ρ_2	ARE ϕ
-0.95	-0.95	0.9908	0.005	0.1	1
	-0.7	0.9684		0.5	0.9890
	-0.3	0.927		0.8	0.9474
	0.005	0.9131	0.3	-0.95	0.9270
	0.1	0.9148		-0.7	0.9731
	0.5	0.9469		-0.3	0.9985
	0.8	0.9782		0.005	0.9984
-0.75	-0.95	0.9734		0.1	0.9985
	-0.7	0.9808		0.5	0.9923
	-0.3	0.9658		0.8	0.9575
	0.005	0.9569	0.6	-0.95	0.9577
	0.1	0.9581		-0.7	0.9825
	0.5	0.9757		-0.3	0.9846
	0.8	0.9799		0.005	0.9793
-0.5	-0.95	0.9469		0.1	0.9800
	-0.7	0.9807		0.5	0.9879
	-0.3	0.9923		0.8	0.9752
	0.005	0.9890	0.85	-0.95	0.9827
	0.1	0.9895		-0.7	0.9761
	0.5	0.9917		-0.3	0.9483
	0.8	0.9699		0.005	0.9369
0.005	-0.95	0.9131		0.1	0.9383
	-0.7	0.9654		0.5	0.9631
	-0.3	0.9984		0.8	0.9803
	0.005	1		0.95	0.9827

NOTE: The parameter value is $t = 3$.

Table 13. Asymptotic Relative Efficiency of λ and ϕ for Antedependence when $t = 4$

ρ_1	ρ_2	ρ_3	ARE ρ_1	ARE ρ_2	ARE ρ_3	ARE ϕ
0.005	0.005	0.005	1	1	1	1
		0.5	1	1	0.8889	0.9877
		0.95	1	1	0.6890	0.9033
	0.5	0.005	1	0.8889	1	0.9877
		0.5	1	0.9000	0.9000	0.9756
		0.95	1	0.9207	0.7132	0.8932
	0.95	0.005	1	0.6890	1	0.9033
		0.5	1	0.7132	0.9207	0.8932
		0.95	1	0.7627	0.7627	0.8237
0.5	0.005	0.005	0.8889	1	1	0.9877
		0.5	0.9000	1	0.9000	1
		0.95	0.9207	1	0.7132	0.9572
	0.5	0.005	0.9000	0.9000	1	0.9756
		0.5	0.9143	0.9143	0.9143	0.9884
		0.95	0.9422	0.9422	0.7469	0.9484
	0.95	0.005	0.9207	0.7132	1	0.8932
		0.5	0.9422	0.7469	0.9422	0.9088
		0.95	0.9880	0.8184	0.8184	0.8868
0.95	0.005	0.005	0.6890	1	1	0.9033
		0.5	0.7132	1	0.9207	0.9572
		0.95	0.7627	1	0.7627	1
	0.5	0.005	0.7132	0.9207	1	0.8932
		0.5	0.7469	0.9422	0.9422	0.9484
		0.95	0.8184	0.9880	0.8184	0.9977
	0.95	0.005	0.7627	0.7627	1	0.8237
		0.5	0.8184	0.8184	0.9880	0.8868
		0.95	0.9583	0.9583	0.9583	0.9805

NOTE: The parameter value is $\phi = 3$.

Table 14. Small-Sample Efficiency of ρ_1 and ρ_2 for Antedependence

ρ_1	ρ_2	RE ρ_1	RE ρ_2	ρ_1	ρ_2	RE ρ_1	RE ρ_2
-0.95	-0.95	0.9765	0.9660	0	-0.3	1.0300	1.0297
	-0.7	0.8547	0.9848		0	1.0303	1.0501
	-0.3	0.7314	0.9984		0.1	1.0440	1.0425
	0	0.6926	1.0024		0.5	1.0357	0.9454
	0.1	0.7611	1.0058		0.8	1.0150	0.7697
	0.5	0.7616	0.9956	0.30	-0.95	1.0007	0.7831
	0.8	0.9140	0.9798		-0.7	1.0053	0.8961
	0.95	0.9814	0.9503		-0.3	1.0060	1.0205
-0.75	-0.95	0.9785	0.8689		0	1.0212	1.0504
	-0.7	0.9357	0.9400		0.1	1.0185	1.0560
	-0.3	0.8472	1.0055		0.5	1.0118	0.9227
	0	0.8054	1.0232		0.8	1.0080	0.7619
	0.1	0.8414	1.0113	0.60	-0.95	0.9906	0.8466
	0.5	0.8736	0.9859		-0.7	0.9621	0.9311
	0.8	0.9378	0.9155		-0.3	0.9107	1.0002
	0.95	0.9801	0.8455		0	0.9113	1.0339
-0.5	-0.95	0.9953	0.7181		0.1	0.8911	1.0388
	-0.7	0.9717	0.8563		0.5	0.9185	0.9606
	-0.3	0.9562	1.0350		0.8	0.9695	0.8207
	0	0.9720	1.0528	0.80	-0.95	0.9732	0.8892
	0.1	0.9609	1.0385		-0.7	0.8814	0.9601
	0.5	0.9671	0.9566		-0.3	0.8335	1.0210
	0.8	0.9876	0.8601		0	0.8047	1.0163
	0.95	0.9954	0.7651		0.1	0.8072	1.0134
0	-0.95	1.0033	0.7690		0.5	0.8260	0.9752
	-0.7	1.0128	0.8305		0.8	0.8831	0.8848

NOTE: The parameter values are $t = 3$ and $\phi = 3$; $\beta_0 = 22.5$, $\beta_1 = 2.5$ and $\beta_2 = 0.5$.

Table 15. Small-Sample Efficiency of ρ_1, ρ_2 and ρ_3 for Antedependence

ρ_1	ρ_2	ρ_3	RE ρ_1	RE ρ_2	RE ρ_3
0.005	0.005	0.005	1.0782	1.0691	1.0459
		0.5	0.9101	1.0654	1.0731
		0.95	0.6187	1.0475	1.0786
	0.5	0.005	1.0978	0.9304	1.058
		0.5	0.9054	0.9404	1.0614
		0.95	0.5186	0.9348	1.036
	0.95	0.005	1.0505	0.5545	1.0294
		0.5	0.9518	0.6355	1.0251
		0.95	0.6697	0.6245	1.0023
0.5	0.005	0.005	1.0734	1.0727	0.9031
		0.5	0.8997	1.0725	0.909
		0.95	0.6289	1.031	0.936
	0.5	0.005	1.0593	0.8975	0.9455
		0.5	0.9475	0.9244	0.9158
		0.95	0.6168	0.9356	0.9646
	0.95	0.005	1.0465	0.591	0.9343
		0.5	0.9395	0.6149	0.9436
		0.95	0.7026	0.7279	0.9934
0.95	0.005	0.005	1.0659	1.0702	0.5969
		0.5	0.9331	1.0423	0.6445
		0.95	0.6859	1.0064	0.7199
	0.5	0.005	1.0415	0.9569	0.6089
		0.5	0.9579	0.9718	0.607
		0.95	0.7553	0.9864	0.724
	0.95	0.005	1.0082	0.6864	0.6897
		0.5	0.9931	0.7022	0.7054
		0.95	0.8809	0.8811	0.922

NOTE: The parameter values are $t = 4$ and $\phi = 3$; $\beta_0 = 22.5$, $\beta_1 = 2.5$ and $\beta_2 = 0.5$.

Table 16. Parameter Estimates for the Cattle Growth Data

	QLS			MLE		
Parameter	Estimate	SE	p-val	Estimate	SE	p-val
Intercept(β_0)	-0.0838	1.1815	0.9435	-0.0834	1.1805	0.9437
Treatment-A(β_1)	0.1676	1.6709	0.9201	0.1668	1.6695	0.9204
ϕ	0.9801	1.0995	0.3727	0.9789	1.0297	0.3418
ρ_1	0.8377	0.2983	0.0050	0.8383	0.2750	0.0023
ρ_2	0.9212	0.1514	<0.0001	0.9219	0.1356	<0.0001
ρ_3	0.9263	0.1419	<0.0001	0.9270	0.1268	<0.0001
ρ_4	0.9416	0.1135	<0.0001	0.9422	0.1007	<0.0001
ρ_5	0.9584	0.0815	<0.0001	0.9589	0.0718	<0.0001
ρ_6	0.9462	0.1047	<0.0001	0.9470	0.0925	<0.0001
ρ_7	0.8946	0.1997	<0.0001	0.8936	0.1834	<0.0001
ρ_8	0.9656	0.0675	<0.0001	0.9645	0.0622	<0.0001
ρ_9	0.9066	0.1781	<0.0001	0.9041	0.1657	<0.0001
ρ_{10}	0.9314	0.1325	<0.0001	0.9308	0.1203	<0.0001
$-2\,\ell$	-657.2232			-657.1442		

CHAPTER 4

SUMMARY

In this dissertation we studied the alternative estimation procedures for various dependent models used to analyze continuous longitudinal data. In Chapter 2, we considered the ARMA (1, 1), AR(1), MA(1) and compound symmetry correlation structures to model the dependencies in the longitudinal measurements. We discussed restrictions on the parameter vector λ to guarantee positive definiteness of the associated correlation matrix $R(\lambda)$ in each case and stated few other properties in the context of time-series. We studied the pairwise likelihood approach to estimate correlation parameters for the ARMA (1, 1) correlation structure along with the regression coefficients and residual variance. Asymptotically, we showed that pairwise likelihood estimators of the correlation parameters are nearly as efficient as the maximum likelihood estimators when the data is normally distributed.

In Chapter 3, we considered the first-order antedependence correlation structure to model the dependency in the longitudinal data. We outlined the drawback of the pairwise likelihood method and studied another alternative method of estimation, known as quasi-least squares. We studied properties of the quasi-least squares estimates in the context of the antedependence model. It is relatively easy to implement because the method uses estimates that are in a closed form. Asymptotic relative efficiency calculations showed that the quasi-least squares is a good competitor to the maximum likelihood method to estimate the correlation parameters. Using simulations we showed that the quasi-least squares estimators could be more efficient than the maximum likelihood estimators in small samples for some values in the correlation parameter space.

In conclusion we showed that there are alternative methods of estimation for analyzing continuous longitudinal data using structured correlation matrices. These alternative methods are not only easy to implement and overcome some difficulties associated with the maximum likelihood method but they are also highly efficient in estimating the correlation parameters.

REFERENCES

Besag, J. (1977), "Efficiency of pseudolikelihood estimation for simple gaussian fields," *Biometrika*, 64, 616-618.

Chaganty, N. R. (1997), "An Alternative approach to the analysis of longitudinal data via generalized estimating equations," *Journal of Statistical Planning and Inference*, 63, 39-54.

Chaganty, N. R.(2003), "Analysis of growth curves with patterned correlation matrices using quasi-least squares," *Journal of Statistical Planning and Inference*, 117, 123-139.

Chaganty, N. R. and Shults J. (1999), "On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter," *Journal of Statistical Planning and Inference*, 76, 145-161.

Davis, C. S. (2002), *Statistical methods for the analysis of repeated measurements*, New York: Springer.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, Second Edition, Oxford University Press: Oxford.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, Wiley Series in probability and Statistics.

Gabriel, K. R. (1962), "Ante-dependence analysis of an ordered set of variables," *Annals of Mathematical Statistics*, 33(1), 201-212.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York: Springer.

Hasza, D. P. (1980), "A Note on Maximum Likelihood Estimation for the First-Order Autoregressive Process," *Communications in Statistics - Theory and Methods*, A9(13), 1411-1415.

Joe, H. (1997), *Multivariate models and dependence concepts*, London: Chapman and Hall/CRC.

Kenward, M. C. (1987), "A method for comparing profiles of repeated measurements," *Applied Statistics*, 36(3), 296308.

Kuk, A. Y. C., and Nott, D. J. (2000), "A pairwise likelihood approach to analyzing correlated binary data," *Statistics and Probability Letters*, 47, 329-335.

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

Lindsay, B. G. (1988), "Composite likelihood methods," *Contemporary Mathematics*, 80, 221-240.

Macchiavelli, R. E. and Arnold, S. F. (1994), "Variable-order antedependence models," *Communications in Statistics - Theory and Methods*, 23(9), 2683-2699.

Shults J. and Chaganty, N. R. (1999), "Analysis of Serially Correlated Data Using Quasi-Least Squares," *Biometrics*, 54, 1622-1630.

Varin, C., and Vidonib, P. (2006), "Pairwise likelihood inference for ordinal categorical time series," *Computational Statistics and Data Analysis*, 51, 2365-2373.

Weiss, R. E. (2005), *Modelling Longitudinal Data,* New York: Springer.

Zhao, Y. and Joe, H. (2005), "Composite Likelihood estimation in multivariate data analysis," *The Canadian Journal of Statistics*, 33, 335-356.

APPENDIX A

SOME FORMULAS AND DERIVATIVES

A.1 CLOSED FORM ROOT OF A CUBIC POLYNOMIAL

For the cubic equation $a_3\alpha^3 + a_2\alpha^2 + a_1\alpha^1 + a_0 = 0$, we have used the following closed form root

$$\widehat{\alpha} = \frac{2}{3a_3}\sqrt{(a_2^2 - 3a_1a_3)}\ \cos(\theta) - \frac{a_2}{3a_3}$$

where $\theta = \frac{1}{3}\arccos\left[-\frac{1}{2}\left(2a_2^3 - 9a_1a_2a_3 + 27a_0a_3^2\right)\left(a_2^2 - 3a_1a_3\right)^{-3/2}\right] + \frac{4\pi}{3}$.

The above formula can be found in Hasza (1980), who showed that the root falls in the interval $(-1, 1)$ for the cubic polynomials that arise in some of the estimation procedures discussed in this dissertation.

A.2 MATRIX DERIVATES

The following matrix derivatives can be found in Harville (1997). We have used these derivatives repeatedly in this dissertation. Suppose $R(\alpha)$ is a symmetric matrix whose elements are functions of the parameter α. Then

$$\frac{\partial \log|R(\alpha)|}{\partial\alpha} = \operatorname{tr}\left(R^{-1}(\alpha)\frac{\partial R(\alpha)}{\partial\alpha}\right)$$

$$\frac{\partial R^{-1}(\alpha)}{\partial\alpha} = -R^{-1}(\alpha)\frac{\partial R(\alpha)}{\partial\alpha}R^{-1}(\alpha)$$

$$\frac{\partial^2 R^{-1}(\alpha)}{\partial\alpha^2} = 2\,R^{-1}(\alpha)\frac{\partial R(\alpha)}{\partial\alpha}R^{-1}(\alpha)\frac{\partial R(\alpha)}{\partial\alpha}R^{-1}(\alpha)$$

A.3 DERIVATIVES FOR ANTEDEPENDENCE CORRELATION MATRIX

Let $R(\lambda)$ be the antedependence correlation matrix defined in (38), where $\lambda = (\rho_1, \rho_2, \ldots, \rho_{t-1})$. Note that $R^{-1}(\lambda)$ is given in (41). We denote the (i, j) th element of a matrix A by $[A]_{ij}$. Then we have

$$\left[\frac{\partial R^{-1}(\lambda)}{\partial \rho_k}\right]_{ij} = \begin{cases} \dfrac{2\rho_k}{1-\rho_k^2} & i = j = k,\ k+1 \\ \dfrac{(1+\rho_k^2)}{(1-\rho_k^2)^2} & i = k,\ j = k+1;\ i = k+1,\ j = k \\ 0 & \text{otherwise} \end{cases}$$

and

$$\left[R^{-1}(\lambda)\frac{\partial R(\lambda)}{\partial \rho_k}\right]_{ij} = \begin{cases} \dfrac{-\rho_k}{1-\rho_k^2} & i = j = k,\ k+1 \\ \dfrac{1}{1-\rho_k^2} & i = k,\ j = i+1; \\ & i = k+1,\ j = i-1 \\ \dfrac{-\rho_i \cdots \rho_j}{1-\rho_k^2} & i = k,\ j < i \\ \dfrac{\rho_{i+1} \cdots \rho_{j-1}}{1-\rho_k^2} & i = k,\ j > i+1 \\ \dfrac{\rho_{i-2} \cdots \rho_j}{1-\rho_k^2} & i = k+1,\ j < i-1 \\ \dfrac{-\rho_{i-1} \cdots \rho_{j-1}}{1-\rho_k^2} & i = k+1,\ j > i \\ 0 & \text{otherwise.} \end{cases}$$

If $\widetilde{\lambda} = (\widetilde{\rho}_1, \ldots, \widetilde{\rho}_t)$, then

$$\left[\frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_k} R(\lambda)\right]_{ij} = \begin{cases} \dfrac{1}{(1+\widetilde{\rho}_k^{\,2})} & i = k,\ j = i+1; \\ & i = k+1,\ j = i-1 \\ \dfrac{\rho_{i+1} \cdots \rho_{j-1}}{(1+\widetilde{\rho}_k^{\,2})} & i = k,\ j > i+1 \\ \dfrac{\rho_{i-2} \cdots \rho_{j}}{(1+\widetilde{\rho}_k^{\,2})} & i = k+1,\ j < i-1 \\ 0 & \text{otherwise.} \end{cases}$$

Further, for $l = k$, the diagonal elements are

$$\left[R^{-1}(\widetilde{\lambda})\, \frac{\partial R^{-1}(\widetilde{\lambda})}{\partial \rho_l}\, R^{-1}(\widetilde{\lambda})\, \frac{\partial R^{-1}(\lambda)}{\partial \rho_k}\right]_{ii} = \begin{cases} \dfrac{(1+\widetilde{\rho}_k^{\,2})}{(1-\widetilde{\rho}_k^{\,2})^2} & i = k,\ k+1 \\ 0 & \text{otherwise} \end{cases}$$

and for $l \neq k$, all the diagonal elements are zero.

APPENDIX B

R PROGRAMS

B.1 PAIRWISE LIKELIHOOD ESTIMATORS

Here we provide the R-program that computes the parameter estimates for pairwise likelihood estimators and estimated variances(asymptotic) for pairwise likelihood and maximum likelihood estimators with ARMA(1,1) correlation model.

Remarks: The following gives the notation for the observed data used in the program.
Y: Matrix of responses with each row consisting of the observations from each subject
X_orig: Matrix of covariates

Code:

```
library(psych)
library(MASS)
library(matrixcalc)

# Generates the correlation matrix R for ARMA(1,1) model
corrmatrixARMA <- function(ti,gamm,rho)
{
  H= powH(ti)
  R=gamm*(rho^H)
  diag(R)=1
  return(R)
}

# Function used in calculting the derivatives and the ARMA
   (1,1) correlation matrix
powH <- function(ti)
{ times <- 0:(ti-1)
  H <<- abs(outer(times, times, "-")) -1
```

```
  diag(H)=0
  return(H)
}

# Estimates beta
betaest <- function(i,R0,Y,X_data)
{ ti=ncol(Y)
  n= nrow(Y)
  if (i==1) R0=diag(ti)
  c1=matrix(0,p,p)
  c2= matrix(0,p,1)
  x=c(1:n)
  dim(x)=length(x)
  ca=apply(x,1, function(x,X_orig)
          {b1=cbind(rep(1,ti),matrix(rep(X_orig[x,],ti),ti,
                                         ncol(X_orig), byrow=TRUE))
           b2=as.matrix(Y[x,])
           c1<<-c1+t(b1)%*%solve(R0)%*%b1
           c2<<-c2+t(b1)%*%solve(R0)%*%b2
          },X_orig=X_orig)
  beta0=solve(c1)%*%c2
  return(beta0)
}

# Computes the residuals Z=Y-X'beta for a time-independent
   covariates
residZ <- function(beta0,Y,X_data)
{ n=nrow(X_data)
  ti=ncol(Y)
  j1=matrix(1,nrow=n,1)
  X=cbind(j1,X_data)
  X1=X%*%beta0
  X10=rep(X1,each=ti)
  return(Y-matrix(X10,n,ti,byrow=TRUE))
}
```

```
# Calculates the negative log-likelihood at a given set of
   parameters
mlogl <- function(theta,Z)
{ beta0= as.matrix(theta[1:p])
  phi=theta[p+1]
  gamm=theta[p+2]
  rho=theta[p+3]
  n=nrow(Y)
  ti=ncol(Y)
  R<-corrmatrixR(ti,gamm,rho)
  nlogl<-(n*ti/2)*log(2*pi)+(n*ti/2)*log(phi)
             +(n/2)*log(det(R)) +(1/(2*phi))*
            sum(diag(ginv(R)%*%t(Z)%*%Z))
  return(nlogl)
}

# Estimates gamma, rho
plest <- function(theta1,Z)
{
   beta0=theta1[1:p]
   phi0=theta1[p+1]

   #gamma etimate
   a3g=-n*(ti-1)*phi0
     c1g=c(1:(ti-1))
     c2g=c(2:ti)
   a2g=sum(rowSums(as.matrix(Z[,c1g]*Z[,c2g])))
   a1g=n*(ti-1)*phi0-sum(rowSums(as.matrix(Z[,c1g]^2
                                         + Z[,c2g]^2)))
   a0g=sum(rowSums(as.matrix(Z[,c1g]*Z[,c2g])))
   # roots using polyroot
   prootg <- polyroot(c(a0g,a1g,a2g,a3g))
     k=round(Im(prootg),5)
   gamma0= Re(prootg[k == 0])

   #rho estimate
```

```
    a3r=-n*(ti-2)*phi0*gamma0^4
      c1r=c(1:(ti-2))
      c2r=c(3:ti)
    a2r=sum(rowSums(as.matrix(Z[,c1r]*Z[,c2r])))*gamma0^3
    a1r=(n*(ti-2)*phi0-sum(rowSums(as.matrix(Z[,c1r]^2
                              + Z[,c2r]^2))))*gamma0^2
    a0r=sum(rowSums(as.matrix(Z[,c1r]*Z[,c2r])))*gamma0
    prootr <- polyroot(c(a0r,a1r,a2r,a3r))
      k=round(Im(prootr),5)
    rho0= Re(prootr[k == 0])
    theta2=c(gamma0,rho0)
     return(theta2)
}

# Calculates the asymptotic variances using the analytical
    expressions
AREPL_ARMA <- function(ti,n,theta_pl,X_orig)
{
  j <- cbind(rep(1:ti,each=ti),rep(1:ti))
  j <<- j[j[,1] < j[,2],]
      gpl=theta_pl[p+2]
      rpl=theta_pl[p+3]
      phipl=theta_pl[p+1]
      R=corrmatrixR(ti,gpl,rpl)

      In_pl=matrix(0,p+3,p+3)
        c1=matrix(0,p,p)
         x=c(1:n)
        dim(x)=length(x)
        ca=apply(x,1, function(x,X_orig,R)
        {b1=cbind(rep(1,ti),matrix(rep(X_orig[x,],ti),ti,
                              ncol(X_orig), byrow=TRUE))
          c1 <<- c1+t(b1)%*%solve(R)%*%b1
          return(c1)
        },X_orig=X_orig,R=R)
```

```
I11_pl= c1/(n*phipl)

  H=powH(ti)
  dRg= rpl^H                # derivative of R wrt gamma
  diag(dRg)=0

  H1=H-1
  H1[H1== -1] = 0
  dRr= H*(gpl*rpl^H1)       # derivative of R wrt rho

I22_pl= ti/(2*phipl^2)
I23_pl= tr(solve(R,tol=tol)%*%dRg)/(2*phipl)
I24_pl= tr(solve(R,tol=tol)%*%dRr)/(2*phipl)
I32_pl= -gpl*(ti-1)/(phipl*(1-gpl^2))
I33_pl=(ti-1)*(1+gpl^2)/(1-gpl^2)^2
  a=(1+gpl^2*rpl^2)/(1-gpl^2*rpl^2)^2
I42_pl=-gpl^2*rpl*(ti-2)/(phipl*(1-gpl^2*rpl^2))
I43_pl= (ti-2)*gpl*rpl*a
I44_pl= (ti-2)*gpl^2*a

In2_pl= matrix(c(I22_pl,I32_pl,I42_pl, I23_pl,I33_pl,
                        I43_pl, I24_pl,0,I44_pl),3,3)
In_pl[1:p,1:p]=I11_pl
In_pl[(p+1):(p+3),(p+1):(p+3)]= In2_pl

 Mn_pl=matrix(0,p+3,p+3)
 Mn_pl[1:p,1:p]=I11_pl
 M22_pl= I22_pl
 M23_pl= M32_pl= -gpl*(ti-1)/(phipl*(1-gpl^2))
 M24_pl= M42_pl= -gpl^2*rpl*(ti-2)/(phipl*(1-gpl^2*rpl^2))
  A1=matrix(c(gpl,1,gpl*rpl,gpl),2,2)
  R1star=matrix(c(-2*gpl,1+gpl^2,1+gpl^2,-2*gpl)
                             /(1-gpl^2)^2,2,2)
  j1 <- cbind(rep(1:(ti-1),each=ti-1),rep(1:(ti-1)))
  j1 <- j1[j1[,1] < j1[,2],]
  if (isTRUE(length(j1)!=0))
```

```
{  if (is.null(nrow(j1)))
    { x=1
      j1 <- matrix(c(j1),length(x),2)
    }
}
j1_m <- j1[j1[,2]-j1[,1] > 1, ]
if (isTRUE(length(j1_m)!=0))
{  if (is.null(nrow(j1_m))) x=1 else x=c(1:nrow(j1_m))
   dim(x)=length(x)
   j1_m1=matrix(c(j1_m),length(x),2)
   a1=apply(as.matrix(x),1,function(x,gpl,rpl,R1star)
             { j1=  j1_m1[x,1]
               j2=  j1_m1[x,2]
               A=R[c(j1,j1+1),c(j2,j2+1)]
               return(tr(R1star%*%A%*%R1star%*%t(A)))
             },gpl=gpl,rpl=rpl,R1star=R1star)
} else a1=0
M33_pl=((ti-1)*(1+gpl^2)/(1-gpl^2)^2)
      +(ti-2)*tr(R1star%*%A1%*%R1star%*%t(A1)) + sum(a1)

j2 <- cbind(rep(1:(ti-2),each=ti-2),rep(1:(ti-2)))
j2 <- j2[j2[,1] < j2[,2],]
if (isTRUE(length(j2)!=0))
{  if (is.null(nrow(j2))){ x=1
   j2 <- matrix(c(j2),length(x),2)}
}
j2_m <- j2[j2[,2]-j2[,1] > 2, ]
R2star=matrix(gpl*c(-2*gpl*rpl,1+gpl^2*rpl^2,
          1+gpl^2*rpl^2,-2*gpl*rpl)/(1-gpl^2*rpl^2)
             ^2,2,2)
A1=matrix(c(gpl,gpl,gpl*rpl^2,gpl),2,2)
A2=matrix(c(gpl*rpl,1,gpl*rpl^3,gpl*rpl),2,2)
if (isTRUE(length(j2_m)!=0))
{ if (is.null(nrow(j2_m))) x=1 else x=c(1:nrow(j2_m))
  dim(x)=length(x)
  a2=apply(as.matrix(x),1,function(x,gpl,rpl,R1star)
```

```
                 { j1=  j2_m[x,1]
                   j2=  j2_m[x,2]
                   j2_m=matrix(c(j2_m),length(x),2)
                   mp2=matrix(c(j2-j1-1, j2-j1+1,
                             j2-j1-3, j2-j1-1),2,2)
                   A=gpl*rpl^mp
                   return(tr(R1star%*%A%*%R1star%*%t(A)))
                 },gpl=gpl,rpl=rpl,R1star=R1star)
        } else a2=0

   M44_pl=((ti-2)*gpl^2*(1+gpl^2*rpl^2)/(1-gpl^2*rpl^2)^2)
          +(ti-3)*tr(R2star%*%A1%*%R2star%*%t(A1))
          +((ti-4)*tr(R2star%*%A2%*%R2star%*%t(A2)))

    j3_m <- cbind(rep(1:(ti-1),each=ti-2),rep(1:(ti-2)))
    R=corrmatrixR(ti,gpl,rpl)
    x=c(1:nrow(j3_m))
    dim(x)=length(x)
    a3=apply(x,1, function(x,gpl,rpl,j3_m,R)
            { j1= j3_m[x,1]
              j2= j3_m[x,2]
              Aj1j2=matrix(c(R[j1,j2] ,R[j1+1,j2],
                          R[j1,j2+2], R[j1+1,j2+2]),2,2)
              return(tr(R1star%*%Aj1j2%*%R2star%*%t(Aj1j2)))
            }, gpl=gpl,rpl=rpl,j3_m=j3_m,R=R)
   M34_pl = sum(a3)/2
   M43_pl= M34_pl
   Mn_pl[(p+1):(p+3),(p+1):(p+3)]=matrix(c(M22_pl, M32_pl,
                      M42_pl, M23_pl, M33_pl,M43_pl, M24_pl,
                      M34_pl,M44_pl),3,3)
   V_pl=  (solve(In_pl,tol=tol)%*% Mn_pl%*%
                  t(solve(In_pl,tol=tol)))/n
   vars_pl=as.matrix(diag(V_pl))
   return(round(vars_pl,5))
}
```

```
# Main program
tol <<- .Machine$double.eps^3
n <<- nrow(Y)
ti <<- ncol(Y)
p <<- ncol(X_orig)+1

i <- 1
ca=numeric(p+5)
thetapl_new=numeric(p+3)
repeat
{ thetapl_old=thetapl_new
  if(i==1) R0=diag(ti) else R0=corrmatrixARMA(ti,
                            thetapl_old[p+2],thetapl_old[p+3])
  beta0=betaest(i,R0,Y,X_orig)
  Z <<- residZ(beta0,Y,X_orig)
  phi0=tr(Z%*%solve(R0)%*%t(Z))/(n*ti)
  theta1_pl=c(beta0,phi0)
  theta2_pl= plest(theta1_pl,Z)
  thetapl_new=c(theta1_pl,theta2_pl)
  nlogl= mlogl(thetapl_new,Z)
  ca=rbind(ca,c(i,thetapl_new,nlogl))
  if (norm(as.matrix(thetapl_new-thetapl_old),"f")
               <.Machine$double.eps^.25) break()
  i=i+1
}
theta_pl=matrix(thetapl_new,1,)

#  The PL estimates for beta, phi, gamma, rho  #
betanames=paste("beta",0:(p-1))
colnames(theta_pl) = c(betanames,"Phi","Gamma","Rho")
print(round(theta_pl,4))

# The negative log-likelihood value
print(rbind("negative log-likelihood",nlogl))

# negtaive loglikelihood values at each iteration
```

```
colnames(ca)= c("iteration",colnames(theta_pl),"nlogl")
print(ca[-1,])

# Calculate the asymptotic variances for the pairwise
   likelihood estimators
pl_vars =  AREPL_ARMA (ti,n,theta_pl,X_orig)
```

B.2 QUASI-LEAST SQUARES AND MAXIMUM LIKELIHOOD ESTIMATORS

Here we provide the R-program that computes the parameter estimates using quasi-least squares and maximum likelihood method in case of first-order antedependence correlation structure. The code also includes the computations for obtaining the estimated variances(asymptotic) for quasi-least squares and maximum likelihood estimators.

As mentioned in previous section, Y represents the matrix of responses with each row consisting of the observations from each subject and X_orig is the matrix of covariates

```
# Generates the correlation matrix R for antedependence
   correlation matrix
corrmatrixANTE <- function(ti,rho)
{
  x <- list(1:ti)
  ind <- as.matrix(expand.grid(rep(x, 2)))
  r <- apply(ind, 1, function(ind,rho)
          { x1<-ind[1]
            x2<-ind[2]
            elem<-ifelse(x1>x2, prod(rho[x2:(x1-1)]),
                  ifelse(x1==x2,1,prod(rho[x1:(x2-1)])))
            return(elem)
          },rho=rho)
  RAnte<-matrix(r,ti,ti)
  return(RAnte)
```

```
}

# Estimates beta
betaest <- function(i,R0,Y,X_data)
{ ti=ncol(Y)
  n= nrow(Y)
  if (i==1) R0=diag(ti)
  c1=matrix(0,p,p)
  c2= matrix(0,p,1)
  x=c(1:n)
  dim(x)=length(x)
  ca=apply(x,1, function(x,X_orig)
          {b1=cbind(rep(1,ti),matrix(rep(X_orig[x,],ti),ti,
                                      ncol(X_orig), byrow=TRUE))
           b2=as.matrix(Y[x,])
           c1<<-c1+t(b1)%*%solve(R0)%*%b1
           c2<<-c2+t(b1)%*%solve(R0)%*%b2
          },X_orig=X_orig)
  beta0=solve(c1)%*%c2
  return(beta0)
}

# Computes the residuals Z=Y-X'beta for a time-independent
   covariates
residZ <- function(beta0,Y,X_data)
{ n<-nrow(X_data)
  ti<-ncol(Y)
  j1<-matrix(1,nrow=n,1)
  X<-cbind(j1,X_data)
  X1<-X%*%beta0
  X10<-rep(X1,each=ti)
  return(Y-matrix(X10,n,ti,byrow=TRUE))
}

# Calculates the negative loglikelihood for a given set of
   parameters
```

```
mlogl <- function(theta,Z)
{  beta0<- as.matrix(theta[1:p])
   phi<-theta[p+1]
   rho<-theta[(p+2):(length(theta))]
   n<-nrow(Z)
   ti<-ncol(Z)
   R <- corrmatrixANTE(ti,rho)
   nlogl<-(n*ti/2)*log(2*pi)+(n*ti/2)*log(phi)
            +(n/2)*log(det(R)) +(1/(2*phi))*
           sum(diag(ginv(R)%*%t(Z)%*%Z))
   return(nlogl)
}

# Calculates the Step-2 quasi-least squares estimates for
   correlation parameters
qlest <- function(Z)
{ c1g<-c(1:(ti-1))
  c2g<-c(2:ti)
  a<-colSums(as.matrix(Z[,c1g]^2 + Z[,c2g]^2))
  b<-colSums(as.matrix(Z[,c1g]*Z[,c2g]))
  rhos1<-(a-sqrt(a^2-4*b^2))/(2*b)
  rhos2<- 2*b/a
  return(rhos2)
}

# Calcultes final quasi-least squares estimates(iterative
   procedure)
thetaqlfunc <- function(Y,X_data,p)
{
  n <- nrow(Y)
  ti <- ncol(Y)
  i <- 1
  ca<-numeric(p+ti+2)
  thetaql_new<-numeric(p+ti)
  repeat
    { thetaql_old=thetaql_new
```

```
      beta1<-betaest(i,R0,Y,X_data)
      Z <- residZ(beta1,Y,X_data)
      rho1<- qlest(Z)
      R0<-corrmatrixANTE(ti,rho1)
      phi1 <- sum(diag(solve(R0,tol=tol) %*%t(Z)%*%Z))/(n*ti)
      thetaql_new<-c(beta1, phi1, rho1)
      nloglql<- round(mlogl(thetaql_new,Z),6)
      ca<-rbind(ca,c(i,thetaql_new,nloglql))
      if (norm(as.matrix(thetaql_new[-(p+1)]
          -thetaql_old[-(p+1)]),"f")<.Machine$double.eps^.25)
          break()
      i<-i+1
    }
  theta_ql<-matrix(round(thetaql_new,5),1,)
  return(theta_ql)
}

# Calcultes maximum likelihood estimates for correlation
   parameters
mlest <- function(Z,phi)
{ n<-nrow(Z)
  ti<-ncol(Z)
  c1g<-c(1:(ti-1))
  c2g<-c(2:ti)
  a3<-rep(-n*phi,ti-1)
  a2<-colSums(as.matrix(Z[,c1g]*Z[,c2g]))
  a1<--(colSums(as.matrix(Z[,c1g]^2 + Z[,c2g]^2))-n*phi)
  a0<-a2
  a<- matrix(c(a0,a1,a2,a3),ti-1,4)
  x<-c(1:nrow(a))
  dim(x)<-length(x)
  rhohatml<-apply(x,1,function(x,a)
      { cc<-a[x,]
        proot <- polyroot(cc)
        k<-round(Im(proot),5)
        if(length(proot[k == 0])>1)
```

```
            rhohat<-Re(proot[Re(proot)>0])
        else rhohat<- Re(proot[k == 0])
        return(rhohat)
      },a=a)
  return(rhohatml)
}

# Calcultes final maximum likelihood estimates(iteratve
   procedure)
thetamlfunc <-  function(Y,X_data,p)
{
  n <- nrow(Y)
  ti <- ncol(Y)
  i <- 1
  ca<-numeric(p+ti+2)
  thetaml_new<-numeric(p+ti)
  repeat
    { thetaml_old<-thetaml_new
      if(i==1) R0=diag(ti) else R0=corrmatrixANTE(ti,
                                  thetaml_old[(p+2):(p+ti)])
      beta1<-betaest(i,R0,Y,X_data)
      Z <- residZ(beta1,Y,X_data)
      phi1 <- sum(diag(solve(R0,tol=tol) %*%t(Z)%*%Z))/(n*ti)
      rho1<- mlest(Z,phi1)
      thetaml_new<-c(beta1, phi1, rho1)
      nloglml<- round(mlogl(thetaml_new,Z),6)
      ca<-rbind(ca,c(i,thetaml_new,nloglml))
      if (norm(as.matrix(thetaml_new[-(p+1)]
          -thetaml_old[-(p+1)]),"f")<.Machine$double.eps^.25)
          break()
      i<-i+1
    }
  theta_ml<-matrix(round(thetaml_new,5),1,)
  return(theta_ml)
}
```

```
# Sub-function to calculate elements in Godambe information
   matrix
elemM2ql <- function(x1,x2,phiql,rho,rhotild)
{
  den<-(1+rhotild[x1]^2)*(1+rhotild[x2]^2)
  if(abs(x1-x2)==0) elem<- 4*phiql^2/(1+rhotild[x1]^2)^2
  if(abs(x1-x2)==1) elem<-2*phiql^2*rho[x2]*rho[x1]/den
  if(abs(x1-x2)>1)
  {
   elem<-ifelse(x1>x2,2*phiql^2*rho[x2]
              *(prod(rho[(x2+1):(x1-1)])^2)*rho[x1]/den,
               2*phiql^2*rho[x1]*(prod(rho[(x1+1):(x2-1)])^2)
               *rho[x2]/den)
  }
  return(elem)
}

#AREs for QLS estimators
AREQL_ANTE <- function(ti,n,theta_est,X_orig)
{
    phiql<- theta_est[1]
    rhoql<-theta_est[2:ti]
    rhotild<- (1-sqrt(1-rhoql^2))/(rhoql)
    R<-corrmatrixANTE(ti,rhoql)

    In_ql<-matrix(0,p+ti,p+ti)
       c1=matrix(0,p,p)
       x=c(1:n)
       dim(x)=length(x)
       ca=apply(x,1, function(x,X_orig,R)
               {b1=cbind(rep(1,ti),matrix(rep(X_orig[x,],ti),
                                 ti,ncol(X_orig),byrow=TRUE))
                c1 <<- c1+t(b1)%*%solve(R)%*%b1
                return(c1)
               },X_orig=X_orig,R=R)
      In_ql[1:p,1:p] = c1/(n*phiql)
```

```
        In_ql[(p+1),(p+1)] <- ti/(2*phiql^2)
        In_ql[p+1,(p+2):(p+ti)]<-  -rhoql/(phiql*(1-rhoql^2))
        In_ql[(p+2):(p+ti),(p+2):(p+ti)]<-
                 diag(c(-2*phiql*(1+rhotild^2)/(1-rhotild^2)^2))

    Mn_ql<-matrix(0,p+ti,p+ti)
      Mn_ql[1:p,1:p] = c1/(n*phiql)
      Mn_ql[(p+1),(p+1)] <- ti/(2*phiql^2)
      x <- list(1:(ti-1))
      ind <- as.matrix(expand.grid(rep(x, 2)))
      r<- apply(ind, 1, function(ind,rho, rhotild,phiql)
                { x1<-ind[1]
                  x2<-ind[2]
                  elem<-elemM2ql(x1,x2,phiql,rho,rhotild)
                  return(elem)
                },rho=rhoql,rhotild=rhotild,phiql=phiql)
      Mn_ql[(p+2):(p+ti),(p+2):(p+ti)]<- matrix(r,ti-1,ti-1)
    G_ql<-(solve(In_ql,tol=tol)%*% Mn_ql%*%
                                        t(solve(In_ql,tol=tol)))/n
    vars_ql<-as.matrix(diag(G_ql))
    return(round(vars_ql,5))
}

#AREs for ML estimators
AREML_ANTE <- function(ti,n,theta_est,X_orig)
{
    phiml<- theta_est[1]
    rhoml<-theta_est[2:ti]
    rhotilm<- (1-sqrt(1-rhoml^2))/(rhoml)
    R<-corrmatrixANTE(ti,rhoml)

    In_ml<-matrix(0,p+ti,p+ti)
     c1=matrix(0,p,p)
     x=c(1:n)
     dim(x)=length(x)
     ca=apply(x,1, function(x,X_orig,R)
```

```
                {b1=cbind(rep(1,ti),matrix(rep(X_orig[x,],ti),
                                ti, ncol(X_orig),byrow=TRUE))
                 c1 <<- c1+t(b1)%*%solve(R)%*%b1
                 return(c1)
                },X_orig=X_orig,R=R)
      In_ml[1:p,1:p] <- c1/(n*phiql)
      In_ml[(p+1),(p+1)] <- ti/(2*phiml^2)
      In_ml[(p+2):(p+ti),p+1]<- In_ml[p+1,(p+2):(p+ti)] <-
                                -rhoml/(phiml*(1-rhoml^2))
      In_ml[(p+2):(p+ti),(p+2):(p+ti)] <-
                            diag(c((1+rhoml^2)/(1-rhoml^2)^2))
     V_ml <- solve(In_ml)/n
     vars_ml <- as.matrix(diag(round(V_ml,5)))
     return(round(vars_ml,4))
}

# MAIN PROGRAM

tol <<- .Machine$double.eps^3
n <<- nrow(Y)
ti <<- ncol(Y)
p <<- ncol(X_orig)+1
rhonames<-paste("Rho",1:(ti-1))
betanames=paste("beta",0:(p-1))

# QLS estimates of the parameters and their variances
theta_ql <- thetaqlfunc(Y,X_orig,p)
colnames(theta_ql) = c(betanames,"Phi",rhonames)
print(round(theta_ql,5))
vars_ql <- AREQL_ANTE(ti,n,theta_ql[(p+1):(p+ti)],X_orig)

# ML estimates of the parameters and their variances
theta_ml <- thetamlfunc(Y,X_orig,p)
colnames(theta_ml) = c(betanames,"Phi",rhonames)
print(round(theta_ml,5))
vars_ml <- AREML_ANTE(ti,n,theta_ml[(p+1):(p+ti)],X_orig)
```

VITA

Sirisha Mushti
Department of Mathematics and Statistics
Old Dominion Univeristy
Norfolk, VA 23529

Education

Ph.D. Old Dominion University, Norfolk, VA. (May 2013)
Major: Computational and Applied Mathematics (Statistics).

M.Sc University of Hyderabad, Hyderabad, AP, India. (June 2006)
Major: Statistics - Operations Research.

B.Sc St. Ann's Degree College, Hyderabad, AP, India. (June 2004)
Major: Mathematics, Statistics and Computer Science.

Experience

Research Assistant (08/2010 - 06/2012)
Epidemiology-Biostatistics Research Support Core,
Eastern Virginia Medical School, Norfolk, VA.

Teaching Assistant (08/2008 - 07/2010 & 08/2012 - 05/2013)
Old Dominion University, Norfolk, VA.

Executive Statistician (03/2007 - 06/2008)
Glaxo SmithKline Pharmaceuticals, Bangalore, India.

Typeset using LaTeX.