Summer 2003

Estimation of Parameters in Replicated Time Series Regression Models

Genming Shi
Old Dominion University

ESTIMATION OF PARAMETERS IN REPLICATED TIME SERIES REGRESSION MODELS

by

Genming Shi
M.S. May 2001, Old Dominion University
B.S. July 1998, Beijing University

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

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
August 2003

Approved by:

Narasinga R. Chaganty (Director)

Dayanand N. Naik (Member)

Ram C. Dahiya (Member)

Larry Filer (Member)

ABSTRACT

ESTIMATION OF PARAMETERS IN REPLICATED
TIME SERIES REGRESSION MODELS

Genming Shi
Old Dominion University, 2003
Director: Dr. Narasinga R. Chaganty

The time series regression model was widely studied in the literature by several authors. However, statistical analysis of replicated time series regression models has received little attention. In this thesis, we study the application of quasi-least squares, a relatively new method, to estimate the parameters in replicated time series models with general ARMA(p, q) correlation structure. We also study several established methods for estimating the parameters in those models, including the maximum likelihood, method of moments, and the GEE method. Asymptotic comparisons of the methods are made by fixing the number of repeated measurements in each series, and letting the number of replications n go to infinity. Our theoretical as well as some simulation results show that the quasi-least squares estimates are undoubtedly better than the moment estimates, and are good competitors and more robust than the maximum likelihood estimates. Examples are presented to illustrate the application of the quasi-least squares method to analyze real life data situations.

I dedicate this thesis to my wife Xia and my parents.

ACKNOWLEDGMENTS

I wish to extend my warmest thanks to my advisor, Professor N. Rao Chaganty, for his many suggestions and guidance throughout the development of this thesis. I am grateful to Professor Chaganty for showing me his work in progress on the analysis of the growth curve model using quasi-least squares estimating method. His work generated a great deal of interest and I saw the potential and applicability of those ideas for analyzing replicated time series models which is the main topic of this thesis. Professor Chaganty's SAS/IML programs for the first order autoregressive model formed the basis of the many computer programs that I developed for this thesis.

I would like to thank Professors Dayanand Naik, Ram C. Dahiya and Larry Filer for serving on my committee. Specifically, I would like to acknowledge Professors Naik and Dahiya for many helpful discussions. Special thanks are due to Professor Filer for editing parts of my thesis.

Finally, I would like to thank my fellow graduate students, friends and staff members in the Department of Mathematics and Statistics at Old Dominion University for providing a pleasant, friendly and stimulating research environment.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One of the most useful and widely studied statistical model is the linear regression model where the outcomes are serially dependent and follow a time series pattern. In this chapter, we formally state this model, also known as the time series regression model, and present a review of the various methods of estimation. We then discuss a replicated version of the time series regression model and summarize the methods for estimating the parameters.

The organization of this chapter is as follows. In Section 1.1, we give a formal definition of the replicated time series regression model. In Section 1.2, we briefly summarize the traditional estimating methods: maximum likelihood and the moment method of estimation. We point out some drawbacks with those methods and then introduce the quasi-least squares as an alternative method of estimation. In Section 1.3, we present an overview of the organization of this thesis. In Section 1.4, we give a summary of the notation and basic definition used in this thesis.

1.1 Replicated time series regression model

A popular model for analyzing repeated measurement data that occurs in real life is the replicated time series regression model with a stationary autoregressive moving average ARMA(p, q) error term. Indeed, time series analysis is developed

The model for this thesis is *Journal of the American Statistical Association*.

mainly to study a sequence of observations that occur with time. But there are situations, especially in longitudinal data analysis, where we have independent sequences of repeated observations. Here the number of repeated observations is small, whereas the number of independent sequences is large. Therefore, it is only natural to find the limiting distributions as the number of independent sequences converges to infinity, unlike in time series where the limiting distributions of the estimates are obtained as the length of the series goes to infinity.

For simplicity, let us first consider a single time series regression model. Suppose that $\{y_j\}$ follows a linear regression model of the form

$$y_j = \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_j, \; j = 1, \ldots, t, \tag{1.1.1}$$

where $\mathbf{x}_j = (x_{j1}, \ldots, x_{jr})'$ is a r-dimensional vector of deterministic or stochastic covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$ is the vector of unknown regression parameters describing the relationship between y_j and $\mathbf{x}_j$. Following Box *et al.* (1994), suppose y_j's are dependent and the error term $\{\varepsilon_j\}$ follows a stationary ARMA (p, q) process with mean 0 and unknown variance σ_ε^2, that is,

$$\varepsilon_j = \phi_1\varepsilon_{j-1} + \cdots + \phi_p\varepsilon_{j-p} + a_j - \theta_1 a_{j-1} - \cdots - \theta_q a_{j-q}, \tag{1.1.2}$$

or in terms of a_k's

$$\varepsilon_j = \sum_{k=0}^{\infty} \psi_k a_{j-k},$$

where $\{a_j\}$ is a white noise process with mean 0 and unknown variance σ^2 and ψ_k's are unknown coefficients.

Let us denote the autoregressive parameters by $\boldsymbol{\phi} = (\phi_1, \ldots \phi_p)'$, and the moving average parameters by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ and denote both set of parameters by $\boldsymbol{\lambda} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$. $\boldsymbol{\lambda}$ is unknown and satisfies certain condition so that $\{\varepsilon_j\}$ is stationary and invertible. Multiplying both sides of (1.1.2) by ε_{j-k} and taking expectation, we see that the k^{th} lag autocovariance function $\gamma_k = \mathrm{Cov}(\varepsilon_j, \varepsilon_{j+k})$ of $\{\varepsilon_j\}$ satisfies the difference equation

$$\begin{aligned}\gamma_k &= \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} + \gamma_{\varepsilon a}(k) - \theta_1\gamma_{\varepsilon a}(k-1) - \cdots - \theta_q\gamma_{\varepsilon a}(k-q) \\ &= \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} - \sigma^2(\theta_k\psi_0 + \theta_{k+1}\psi_1 + \cdots + \theta_q\psi_{q-k}) \qquad (1.1.3)\end{aligned}$$

with the convention that $\theta_0 = -1$. Here $\gamma_{\varepsilon a}(k)$ is the cross covariance function between ε and a and is defined by $\gamma_{\varepsilon a}(k) = \mathrm{E}(\varepsilon_{j-k}\, a_j)$, which is ψ_{-k} if $k \leq 0$ and 0 otherwise, E denotes the expected value. Equation (1.1.3) implies

$$\begin{aligned}\gamma_k &= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}\,, \quad k \geq q+1, \\ \gamma_0 &= \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma^2(1 - \theta_1\psi_1 - \cdots - \theta_q\psi_q)\,, \qquad (1.1.4)\end{aligned}$$

where $\gamma_0 = \sigma_\varepsilon^2$ is the variance of $\{\varepsilon_j\}$ and (1.1.4) has to be solved along with the p equations (1.1.3) to obtain $\gamma_0, \gamma_1, \ldots, \gamma_p$. Suppose $\rho_k = \gamma_k/\gamma_0$ is the k^{th} autocorrelation function of $\{\varepsilon_j\}$. Then

$$\begin{aligned}\rho_k &= \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}, \qquad k \geq q+1, \ (1.1.5) \\ \rho_k &= \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} - \\ &\quad \frac{\sigma^2(\theta_k\psi_0 + \theta_{k+1}\psi_1 + \cdots + \theta_q\psi_{q-k})}{\phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma^2(1 - \theta_1\psi_1 - \cdots - \theta_q\psi_q)}, \qquad 0 \leq k \leq q.\end{aligned}$$

Now suppose we have n independent sequences of data. Given a sample of $t_i\,(> 2p)$ observations, let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{it_i})'$ be a $t_i \times 1$ response vector for replication $i = 1, 2, \ldots, n$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ contains all observations. The error vector $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{it_i})'$ has mean $\mathbf{0}$ and covariance $\boldsymbol{\Gamma}_i(\boldsymbol{\lambda}, \sigma^2)$, and hence $\mathbf{y}_i$ has mean $\mathbf{X}_i\boldsymbol{\beta}$ and covariance

$$\boldsymbol{\Gamma}_i(\boldsymbol{\lambda}, \sigma^2) = \sigma_\varepsilon^2\mathbf{P}_i(\boldsymbol{\lambda}) = \sigma^2\mathbf{V}_i(\boldsymbol{\lambda}), \qquad (1.1.6)$$

where $\mathbf{\Gamma}_i(\boldsymbol{\lambda}, \sigma^2)$ is the t_i-dimensional covariance matrix with the k^{th} diagonal elements γ_{k-1}, $\mathbf{P}_i(\boldsymbol{\lambda})$ is the t_i-dimensional autocorrelation matrix with the k^{th} diagonal elements ρ_{k-1} and $\mathbf{V}_i(\boldsymbol{\lambda}) = \sigma_\varepsilon^2 \, \mathbf{P}_i(\boldsymbol{\lambda})/\sigma^2$. The stationary condition implies that $\mathbf{\Gamma}_i, \mathbf{P}_i$ and $\mathbf{V}_i$ are positive definite. Define the $t_i \times r$ covariates matrix as $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it_i})'$, and assume that $\mathbf{X}_i$ is of full rank r and satisfies the Grenander conditions (see, e.g., Anderson 1971, P. 572). The model (1.1.1) may be expressed in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \; i = 1, \ldots, n. \tag{1.1.7}$$

1.2 Review of literature and research motivation

An important problem in the replicated time series regression model is the estimation of the $(r + p + q + 1)$ parameters $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)$ in the model (1.1.7). If we treat the model as a regression model without any distributional assumptions, we can use generalized estimating equation (GEE) approach. On the other hand, if we treat the model as a time series model, we could use Box-Jenkins unconditional least squares and Bayesian approaches (Box *et al.* (1994)). But the traditional and popular techniques have been the moment (MOM) and the maximum likelihood (ML) estimating methods. Recently, Cheang and Reinsel (2000) discussed restricted maximum likelihood (REML) method for the model with AR(p) errors. This research mainly focuses on deriving MOM and ML estimating methods and making relative efficiency comparisons with respect to those methods.

If the vector of autoregressive and moving average parameters $\boldsymbol{\lambda}$ is known, $\mathbf{V}$ can be computed from $\boldsymbol{\lambda}$. Given $\mathbf{V}$, the efficient estimates of regression parameters can be computed using generalized least squares (GLS). That is, we minimize the adjusted sum of square errors

$$\begin{aligned} S(\beta, \lambda) &= \sum_{i=1}^{n} \varepsilon_i' \, \mathbf{V}_i^{-1} \, \varepsilon_i \\ &= \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i\beta)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta) \\ &= \sum_{i=1}^{n} \left[\mathbf{y}_i' \mathbf{V}_i^{-1} \mathbf{y}_i - 2\beta' \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i + \beta' \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \beta \right] \end{aligned} \tag{1.2.8}$$

with respect to β. It can also be written as

$$S(\beta, \lambda) = \sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{V}_i^{-1} \mathbf{U}_i \right), \tag{1.2.9}$$

where 'tr' means the *trace*, and

$$\mathbf{U}_i(\beta) \;=\; \varepsilon_i \varepsilon_i' = (\mathbf{y}_i - \mathbf{X}_i\beta)(\mathbf{y}_i - \mathbf{X}_i\beta)'.$$

On taking the partial derivative of $S(\beta, \lambda)$ in (1.2.8) with respect to β, we obtain

$$\frac{\partial S}{\partial \beta} = -2 \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i + 2 \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right) \beta,$$

which gives

$$\hat{\beta}_g = \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \cdot \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i, \tag{1.2.10}$$

where g stands for *generalized least squares estimate* and

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1}. \tag{1.2.11}$$

An unbiased estimate of σ^2 is

$$\hat{\sigma}_g^2 = \frac{1}{n\bar{t}} S(\beta, \lambda) = \frac{1}{n\bar{t}} \sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{V}_i^{-1} \mathbf{U}_i \right), \tag{1.2.12}$$

where $\bar{t} = \sum_{i=1}^{n} t_i/n$. It is well known that the estimates of β and σ^2 are optimal in many ways. Hence, in this thesis, we will focus on the estimation of λ.

Moment estimating method

The simplest method to estimate $\boldsymbol{\lambda}$ is the moment estimating method. Given the "residuals" $\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$ from the regression model (1.1.7), we obtain the moment estimate of $\boldsymbol{\lambda}$ by setting $\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i') = \hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'$ and solving the equations, this is equivalent to setting $\boldsymbol{\Gamma}_i = \hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'$, since $\boldsymbol{\Gamma}_i = \mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i') - \mathrm{E}(\boldsymbol{\varepsilon}_i)\mathrm{E}(\boldsymbol{\varepsilon}_i)' = \mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$. There are $(t_i - k)$ estimates for the k^{th} lag autocovariance γ_k for the i^{th} replication, we estimate those by averaging them, hence γ_k is estimated by $\hat{c}_{k0}/(\bar{t} - k)$, where

$$c_{k0} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t_i-k} \varepsilon_{ij}\ \varepsilon_{i(j+k)}, \quad k = 0, 1, \ldots, p \tag{1.2.13}$$

and 'hat' indicates evaluating based on the residuals. Some authors have used $\hat{c}_{k0}/\bar{t}$ as the estimate for γ_k instead, this two estimates are close to each other when t_i's are large. The k^{th} autocorrelation ρ_k may thus be estimated as the k^{th} sample autocorrelation $r_k = \bar{t}\, c_{k0}/((\bar{t} - k)\, c_{00})$. A more general definition of c_{kl} may be found in Section 4.1. For autoregressive error model, the parameter $\boldsymbol{\lambda}$ reduces to $\boldsymbol{\phi}$, the moment estimate of $\boldsymbol{\phi}$ is the same as the *Yule-Walker* estimate and we will discuss the details in Section 4.1; for moving average error model and mixed autoregressive-moving average error model, the moment estimate of $\boldsymbol{\lambda}$'s are a little difficult (Box *et al.* (1994) Chap.6, p.221, quadratically convergent process), and we will discuss it in Section 4.2 only for MA(1) case. The moment estimates of $\boldsymbol{\beta}$ and σ^2 are same as the GLS estimates (1.2.10) and (1.2.12). Since $\hat{\boldsymbol{\beta}}_g$ depends on $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}_g$ depends on the "residuals" $\hat{\boldsymbol{\varepsilon}}_i$'s, which require the estimate of $\boldsymbol{\beta}$, we need to solve for $(\hat{\boldsymbol{\beta}}_m, \hat{\boldsymbol{\lambda}}_m)$ recursively by combining (1.2.10) with the estimating equation of $\boldsymbol{\lambda}$ and then obtain σ_m^2 using (1.2.12) plugging in $(\hat{\boldsymbol{\beta}}_m, \hat{\boldsymbol{\lambda}}_m)$, where m stands for *moment estimates*. One iterative method is the Newton-Raphson method (Ralston and Wilf 1967, Carnahan *et al.* 1969), see Appendix I for a detailed discussion on the method.

Maximum likelihood estimating method

Another important and frequently used method is the maximum likelihood. We will derive the maximum likelihood estimates (MLEs) from the likelihood function and the ML equations. The solutions of the equations are not always in a closed form, therefore we solve the equations only for simple cases, e.g. AR(1) and AR(2) error models, and use Newton-Raphson method solving the ML equations for AR(p) error model, although an approximate maximum likelihood estimates were suggested by Box *et al.* (1994, p.300). We will show these are no improvement over the moment estimates in terms of efficiency. For more complicated models, e.g. MA(1) error model, Newton-Raphson method is again used to obtain the MLEs.

Assuming that the error term ε_i is normal. The likelihood function of $\mathbf{y}$ is

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n\bar{t}}{2}} \prod_{i=1}^{n} \sqrt{|\mathbf{V}_i|}} \exp\left(-\frac{\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma^2} \right),$$

hence, the log-likelihood function is

$$l(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) = -\frac{n\bar{t}}{2} \log(2\pi) - \frac{n\bar{t}}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \log |\mathbf{V}_i| - \frac{1}{2\sigma^2} S(\boldsymbol{\beta}, \boldsymbol{\lambda}), \quad (1.2.14)$$

where $S(\boldsymbol{\beta}, \boldsymbol{\lambda})$ is as defined in (1.2.8). We will see that $|\mathbf{V}_i|$ is independent of t_i for the AR(p) error model. In a single time series analysis (only one replication), (1.2.14) is dominated by the term involving $S(\boldsymbol{\beta}, \boldsymbol{\lambda})$ for moderate or large t, we can ignore the term involving $|\mathbf{V}_i|$ and hence obtain the unconditional least squares estimates, which is also the first step quasi-least squares estimates as shown next. Furthermore, if we estimate some initial values for a_j, $j \leq 0$, we may obtain the conditional least squares estimates (Box *et al.* (1994), p.226-227). But in our case, t_i's may be small and n is involved in both terms, we cannot ignore the term involving $|\mathbf{V}_i|$ when n is large.

Taking the partial derivative of (1.2.14) with respect to σ^2 and $\boldsymbol{\beta}$ we obtain

$$\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma^2}\frac{\partial S(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} \\
&= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i + \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)\boldsymbol{\beta}\right], \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n\bar{t}}{2\sigma^2} + \frac{S(\boldsymbol{\beta},\boldsymbol{\lambda})}{2\sigma^4}.
\end{aligned}$$

These yield

$$\boldsymbol{\beta} = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i\,, \tag{1.2.15}$$

$$\sigma^2 = \frac{1}{n\bar{t}} S(\boldsymbol{\beta},\boldsymbol{\lambda}) = \frac{1}{n\bar{t}} \sum_{i=1}^{n} \mathrm{tr}(\mathbf{V}_i^{-1}\mathbf{U}_i). \tag{1.2.16}$$

Note that they have the same expressions as the GLS estimates shown in (1.2.10) and (1.2.12), respectively. Now taking the partial derivative of (1.2.14) with respect to $\boldsymbol{\lambda}$ we get

$$\frac{\partial l}{\partial \boldsymbol{\lambda}} = -\frac{1}{2}\sum_{i=1}^{n} \frac{\partial \log|\mathbf{V}_i|}{\partial \boldsymbol{\lambda}} - \frac{1}{2\sigma^2}\frac{\partial S(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}.$$

Thus, the ML equation of $\boldsymbol{\lambda}$ is

$$\sigma^2 \cdot \sum_{i=1}^{n} \frac{\partial \log|\mathbf{V}_i|}{\partial \boldsymbol{\lambda}} + \sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\lambda}}\mathbf{U}_i\right) = \mathbf{0}. \tag{1.2.17}$$

Let $\partial \mathbf{V}_i^{-1}/\partial \boldsymbol{\lambda}$ be the matrix of partial derivatives with respect to λ_k. Equation (1.2.17) can also be written as

$$\sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\lambda}}\mathbf{U}_i\right) - \sigma^2 \cdot \sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\lambda}}\mathbf{V}_i\right) = \mathbf{0}. \tag{1.2.18}$$

We write in this form only for the purpose of comparison with quasi-least squares method. Here we used the identities:

$$\frac{\partial |\mathbf{V}_i|}{\partial \boldsymbol{\lambda}} = |\mathbf{V}_i| \cdot \mathrm{tr}\left(\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\lambda}}\mathbf{V}_i^{-1}\right),$$

$$\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\lambda}} = -\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\lambda}} \mathbf{V}_i^{-1}.$$

After plugging in the estimate of σ^2 in (1.2.16), equation (1.2.17) becomes

$$\sum_{i=1}^{n} \frac{\partial \log |\mathbf{V}_i|}{\partial \boldsymbol{\lambda}} \cdot \sum_{i=1}^{n} \operatorname{tr}\left(\mathbf{V}_i^{-1} \mathbf{U}_i\right) + n\bar{t} \cdot \sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\lambda}} \mathbf{U}_i\right) = \mathbf{0}. \tag{1.2.19}$$

Note that the solution of (1.2.19) usually is not in a closed form. The ML estimates $(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\lambda}}_l)$, where l stands for *maximum likelihood estimate*, are the simultaneous solutions of the equations (1.2.15) and (1.2.19) subject to the set of feasible values of $\boldsymbol{\lambda}$, which ensures the stationary and invertibility of the process $\{\varepsilon_j\}$. Finally, the ML estimate of σ^2 is obtained by (1.2.16) plugging in $(\hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\lambda}}_l)$.

Motivation

We have seen that there are several methods of estimating the parameters in replicated time series regression model. However, each method has some limitations. The moment method estimate is not very efficient, while the ML method needs the normality assumption of the data. Also, the MLEs are hard to obtain numerically and may be highly biased even for moderately large samples. GEE approach is mainly developed for general correlated regression model and has its own drawbacks. It will be shown in Section 4.1.2 that the GEE methods may be reduced to either moment or the maximum likelihood method in some cases. Restricted ML estimating method is a modification of the regular ML estimating method and is even more complicated than the ML method. The Box-Jenkins approach is normally used when we have only one series consisting a large collection of repeated measurements. Bayesian method requires a prior and is not very popular in time series data analysis. Therefore, we introduce another method called quasi-least squares estimating (QLS) method.

The QLS method was introduced by Chaganty (1997) to the analysis of longitudinal data, and then developed and generalized by Shults and Chaganty (1998), Chaganty and Shults (1998) and Chaganty (2003) to the analysis of serially correlated data and growth curve models. It turns out to be a good competitor to maximum likelihood estimating method in the sense that the estimate is easy to obtain and has efficiency close to MLE. The QLS method does a great job especially when the errors are equicorrelated or follow a first order autoregressive process (Chaganty (2003)). More generally, when the errors follow an AR(p) process, QLS method leads to a closed form unlike the ML method. Furthermore, we can verify that, given the means and variances of the errors, the optimal unbiased estimating equation in the sense of Godambe (1960) for $\boldsymbol{\lambda}$ is (1.2.17) or (1.2.18) without making any distributional assumptions, see Chaganty and Naik (2002). But the solution of this equation may not always exist in the feasible region, this leads to the QLS method, which modifies the equation to obtain a feasible solution.

Quasi-least squares estimating method

The quasi-least squares estimating method is a two step process with regarding to the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. It does not require any assumptions concerning the distribution of the data, therefore, it can also be used even if $\mathbf{y}_i$'s are not normal. Technically, first, we set the first term of equation (1.2.18) to be $\mathbf{0}$ and get a first stage estimate of $\boldsymbol{\lambda}$; second, setting the second term of (1.2.18) to be $\mathbf{0}$, we solve this equation by plugging in the first stage estimate. This gives a way to find a solution of (1.2.18), which always exists in the feasible region.

The first step of the QLS method consists of minimizing the objective function $S(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. This first step is also known as the unconditional least squares (ULS). Equating to $\mathbf{0}$ the partial derivative of (1.2.8) with respect to $\boldsymbol{\lambda}$, we get

$$\sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\lambda}} \mathbf{U}_i\right) = \mathbf{0}. \tag{1.2.20}$$

Suppose $\hat{\boldsymbol{\lambda}}_u$ is the solution of (1.2.20), where u stands for *unconditional least squares estimate*. Since (1.2.20) is not an unbiased estimating equation, that is, the expected value of the left hand side of the equation is not zero, we need to modify $\hat{\boldsymbol{\lambda}}_u$ to be unbiased. This leads to the second step of the QLS method.

The second stage of QLS method consists of solving the equation

$$\sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\boldsymbol{\lambda}}_u)}{\partial \boldsymbol{\lambda}} \cdot \mathbf{V}_i\right) = \mathbf{0} \tag{1.2.21}$$

to get a consistent estimate of $\boldsymbol{\lambda}$. Equating to zero the partial derivative of (1.2.8) with respect to $\boldsymbol{\beta}$, we obtain (1.2.10). Given an estimate of $\boldsymbol{\beta}$, we obtain an estimate $\hat{\boldsymbol{\lambda}}$ from (1.2.21), and then modify the estimate of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1}(\hat{\boldsymbol{\lambda}}) \mathbf{X}_i\right)^{-1} \cdot \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1}(\hat{\boldsymbol{\lambda}}) \mathbf{y}_i\,. \tag{1.2.22}$$

This procedure has to be done recursively until $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ converge to $(\hat{\boldsymbol{\beta}}_q, \hat{\boldsymbol{\lambda}}_q)$, where q stands for *quasi-least squares estimate*. Finally, a consistent estimate of σ^2 based on the residuals $\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_q$ is given by

$$\hat{\sigma}_q^2 = \frac{1}{nt} S(\hat{\boldsymbol{\beta}}_q, \hat{\boldsymbol{\lambda}}_q) = \frac{1}{nt} \sum_{i=1}^{n} \operatorname{tr}\left(\mathbf{V}_i^{-1}(\hat{\boldsymbol{\lambda}}_q)\hat{\mathbf{U}}_i\right), \tag{1.2.23}$$

where $\hat{\mathbf{U}}_i = \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i'/n$. If we solve (1.2.10) and (1.2.20) recursively we can obtain the ULS estimates of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$, and the estimate of σ^2 is then given by (1.2.12) plugging in the estimates of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$.

As a summary, the three methods, i.e. MOM, ML and QLS, have the same expressions for the estimations of $\boldsymbol{\beta}$ and σ^2. The main difference lies in how they estimate $\boldsymbol{\lambda}$. Note that the three estimates of $\boldsymbol{\beta}$ will have the same efficiency asymptotically as $n \to \infty$.

1.3 Overview of thesis

This thesis consists of five chapters. Chapter I is the introduction. Here, we introduced the general replicated time series regression model. The model is described in a general setting including the notations. Several estimating methods regarding to the unknown parameters in the model are presented. Again, these methods are introduced in a general way without giving precise details. The common feature of the methods is that they share the same functional form for the estimates of $\boldsymbol{\beta}$ and σ^2, which are the generalized least squares estimates. The main difference between the methods is how they estimate $\boldsymbol{\lambda}$ ($\boldsymbol{\phi}$ and $\boldsymbol{\theta}$). The advantage of quasi-least squares estimating method is that no assumptions about the distribution of the data are required. In the last section of this chapter, the notation and basic definition used in this thesis are listed. They may serve as an index for quick reference to the notations and definitions if the reader prefers to read a separate chapter or a section without going through the whole thesis.

In Chapter II, we study the application of the estimating methods to the model with AR(1) errors, which is the most important model for many practical situations. The methods are described in detail for this particular model. The asymptotic properties are illustrated by theorems. The simulation results are presented in several tables and figures. We apply the estimating methods when the data has a normal, Student t-distribution and Beta distribution, respectively. For normally distributed data, we compare the methods by finding the asymptotic distributions; when the data has a Student t-distribution or a Beta distribution, we compare the methods by simulation. The asymptotic properties and the simulation results reveal that the quasi-least squares estimating method is better than the moment estimating method, and it is a good competitor to the maximum likelihood estimating method. It is better than the ML method when the data has a Beta distribution. Finally, a real data analysis is presented to illustrate the estimating methods.

In Chapter III, the application of the estimating methods to the model with AR(2) errors is studied. As in Chapter I, the methods are described in detail for this model, and we compare them by simulation, since explicit derivation of the asymptotic distributions is difficult in this case. The feasible region of the parameters is a little more complicated. Special care is needed when calculating the numerical value of the estimates, since they are much easier to go out of the boundary. A real data analysis is also presented to illustrate the estimating methods.

Chapter IV is the generalization of the results in the previous chapters. Here, we study the model with autoregressive of order p, AR (p) and moving average of order one, MA (1) errors. When the error is an AR(p) process, the feasible region of the parameters is complicated and difficult to illustrate geometrically. The asymptotic distributions are also extremely complicated. The exact MLEs are hard to obtain and only approximations are available. When the model has a MA (1) errors, at first sight it appears to be easy, but actually it is very much involved because the inverse of the correlation matrix is not in a simple form. There are problems with convergence and appropriate approximations are needed when obtaining the numerical values. Our final goal is to generalize the estimating methods to the model with an ARMA (p, q) error, we will give a brief discussion of these important research topics in Chapter V.

Chapter V gives details of future directions for further research. This includes generalization of the estimations methods to ARMA (p, q) model, and the model with different autoregressive-moving average parameters for each groups.

Finally, the Appendices contain the discussion of Newton-Raphson method and a SAS program example used in this thesis.

1.4 Notation and basic definition

Throughout this thesis, matrices are represented by upper case bold letters, vectors by lower case bold letters. Greek lower case letters are used for parameters and Greek boldfaced lower case letters are used for a collection of parameters. The following notation and definitions are used throughout the thesis.

1. We assume that the series data $\{y_j\}$ is dependent and follows a linear regression model of the form

$$y_j = \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_j, \;\; j = 1, \ldots, t,$$

where $t > 2p$ is the number of observations and $\mathbf{x}_j = (x_{j1}, \ldots, x_{tr})'$ is a covariates vector. The unknown regression parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$ describes the relationship between y_j and $\mathbf{x}_j$. Suppose there were n replications, but the response within replications are uncorrelated. For the i^{th} replication, writing the responses as a single vector $\mathbf{y}_i$, the model in matrix notation is

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \;\; i = 1, \ldots, n.$$

The unknown regression parameter vector $\boldsymbol{\beta}$ is the same for all the replications.

2. For the i^{th} replication, $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{it_i})'$ is the $t_i \times 1$ response vector and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{it_i})'$ is the error vector. The $t_i \times r$ covariates matrix $\mathbf{X}_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{it_i})'$ is assumed to be of full rank r and assumed to satisfy the Grenander conditions (see, e.g., Anderson (1971, p.572)).

3. AR(1) denotes an autoregressive process of degree 1. Similarly AR(2) denotes an autoregressive process of degree 2. More generally, AR(p) denotes an autoregressive process of degree p. Now MA(1) is an acronym for a moving average process of degree 1 and MA(q) for a moving average process of degree q. Finally, ARMA(p, q) stands for an autoregressive and moving average process of degrees p and q.

4. For any replication, the error series $\{\varepsilon_j\}$ is assumed to be a stationary process following an ARMA (p, q) (here p or q, but not both, could be 0) model with mean 0 and variance σ_ε^2, that is,

$$\varepsilon_j = \phi_1 \varepsilon_{j-1} + \cdots + \phi_p \varepsilon_{j-p} + a_j - \theta_1 a_{j-1} - \cdots - \theta_q a_{j-q},$$

   where $\{a_j\}$ is a white noise process with mean 0 and variance σ^2. Define $\mathbf{a}_i = (a_{i1}, a_{i2}, \ldots, a_{it})'$.

5. Denote the autoregressive parameters by $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, the moving average parameters by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ and $\boldsymbol{\lambda} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$. We assume that $\boldsymbol{\lambda}$ satisfies appropriate conditions so that $\{\varepsilon_j\}$ is stationary and invertible.

6. The k^{th} lag autocovariance function of $\{\varepsilon_j\}$ is $\gamma_k = \text{Cov}(\varepsilon_j, \varepsilon_{j+k})$ and the k^{th} lag autocorrelation function is $\rho_k = \gamma_k/\gamma_0$, where $\gamma_0 = \sigma_\varepsilon^2$ is the variance of the series $\{\varepsilon_j\}$. The $t \times t$ autocovariance matrix $\boldsymbol{\Gamma}$ made at t successive times is a Toeplitz matrix from $(\gamma_0, \gamma_1, \cdots, \gamma_{t-1})$ with the k^{th} diagonal elements γ_{k-1}. The $t \times t$ matrix $\mathbf{P}(\boldsymbol{\lambda})$ is the autocorrelation matrix Toeplitzed from $(1, \rho_1, \ldots, \rho_{t-1})$ with the k^{th} diagonal elements ρ_{k-1}. Also define the $t \times t$

matrix $\mathbf{V} = (\sigma_\varepsilon^2/\sigma^2)\,\mathbf{P}$. The matrices $\mathbf{\Gamma}, \mathbf{P}$ and $\mathbf{V}$ with the subscript p means the dimension is p instead of t. The stationarity conditions imply that $\mathbf{\Gamma}, \mathbf{P}$ and $\mathbf{V}$ are positive definite. The sample autocorrelation matrix is represented by the $t \times t$ matrix $\mathbf{R}$. The p-dimensional autocorrelation vector is defined as $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_p)'$ and the sample autocorrelation vector is defined as $\mathbf{r} = (r_1, r_2, \ldots, r_p)'$. The lower triangular $t \times t$ matrix $\mathbf{L}$ is the Cholesky decomposition of $\mathbf{V}^{-1}$. Define $\Delta(\boldsymbol{\phi}) = 1 - \boldsymbol{\phi}'\boldsymbol{\rho} = 1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p$.

7. The error $\boldsymbol{\varepsilon}_i$ has mean $\mathbf{0}$ and covariance $\mathbf{\Gamma}_i = \sigma_\varepsilon^2 \mathbf{P}_i = \sigma^2 \mathbf{V}_i$, that is,

$$\mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \mathbf{\Gamma}_i, \qquad \text{for } i = 1, \ldots, n,$$

where 'E' is the *expected value* and 'Cov' denotes the *covariance*.

8. Define

$$\begin{aligned} \mathbf{U}_i(\boldsymbol{\beta}) &= \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})' \\ \hat{\mathbf{U}}_i &= \mathbf{U}_i(\hat{\boldsymbol{\beta}}), \quad i = 1, \ldots, n. \end{aligned}$$

9. The error sum of square errors is denoted by

$$\begin{aligned} S(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \sum_{i=1}^{n} \boldsymbol{\varepsilon}_i' \mathbf{V}_i^{-1} \boldsymbol{\varepsilon}_i \\ &= \sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{V}_i^{-1}\mathbf{U}_i\right). \end{aligned}$$

10. In order to obtain MLE, we need to assume that the error $\boldsymbol{\varepsilon}_i$ is normal. The likelihood function of $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n\bar{t}}{2}} \Pi_{i=1}^{n} \sqrt{|\mathbf{V}_i|}} \exp\left(- \frac{\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma^2} \right),$$

where $\bar{t} = \sum_{i=1}^{n} t_i/n$. Hence, the log-likelihood function is

$$l(\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) = -\frac{n\bar{t}}{2} \log(2\pi) - \frac{n\bar{t}}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \log|\mathbf{V}_i| - \frac{1}{2\sigma^2} S(\boldsymbol{\beta}, \boldsymbol{\lambda}).$$

The information matrix based on $\mathbf{y}$ is $\mathbf{I}_n(\beta, \phi, \sigma^2)$. For other methods, we do not need to make any assumptions about the distribution.

11. When the autocovariance matrix $\mathbf{\Gamma} = \sigma^2 \mathbf{V}$, the estimates of the autoregressive and moving average parameters are denoted by the 'hat' symbol and this is default symbol for all estimates; If $\mathbf{\Gamma} = \sigma_\varepsilon^2 \mathbf{P}$, the estimates are denoted by the 'tilde' symbol. In this thesis, we only discuss the case when writing $\mathbf{\Gamma} = \sigma^2 \mathbf{V}$, except for the AR(1) case. The subscript of the estimates g stands for *generalized least squares estimate*, l stands for *maximum likelihood estimate*, m stands for *moment estimate*, q stands for *quasi-least squares estimate*, u stands for *unconditonal least squares estimate*, and al stands for *approximate maximum likelihood estimate*. The asymptotic variance of the estimates are denoted by v with corresponding subscripts.

12. ML is the abbreviation for *maximum likelihood*, AML is the abbreviation for *approximate maximum likelihood*, MOM is the abbreviation for *moment*, and QLS is the abbreviation for *quasi-least squares*. GLS is the abbreviation for *generalized least squares* and ULS is the abbreviation for *unconditional least squares*. MLE is the abbreviation for *maximum likelihood estimate*. GEE is the abbreviation for *generalized estimating equations*.

13. Define the $t_i \times t_i$ matrix $\mathbf{C}_{ikl}$ such that $\mathbf{C}_{ilk} = \mathbf{C}_{ikl}$ and, for $k > l$, $2\,\mathbf{C}_{ikl}$ has $(t_i - k - l)$ one's on the $(k-l)^{th}$ diagonals above and below the main diagonal, excluding the first and last l elements on these diagonals, and zero's elsewhere, $k = 1, \ldots, p$, $l = 0, 1, \ldots, p$. The matrix $\mathbf{C}_{ikk}$ has $(t_i - 2k)$ one's on the main diagonal, excluding the first and last k elements, $k = 0, 1, \ldots, p$. Note that $\mathbf{C}_{i00}$ is simply the $t_i \times t_i$ identity matrix (also denoted by $\mathbf{I}_i$). The matrix $\mathbf{C}_{k0}$ with the superscript p means the dimension is p.

14. The $p \times 1$ vector ϵ_k is a unit vector with one in the k^{th} position and zero's elsewhere, $k = 1, \ldots, p-1$. The identity matrix of order t_i is denoted by $\mathbf{I}_i$. The diagonal matrix $\mathbf{D}$ has $(1, 2, \ldots, p)$ on the main diagonal and 0 elsewhere.

15. Define

$$\begin{aligned} c_{kl} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{tr}(\mathbf{C}_{ikl}\mathbf{U}_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=l+1}^{t_i-k} \varepsilon_{ji}\varepsilon_{i(j+k-l)}, \qquad k = 0, 1, \ldots, p, \quad l = 0, 1, \ldots, p, \\ c^*_{kl} &= c_{kl}/(\bar{t} - k - l), \qquad k = 0, 1, \ldots, p, \quad l = 0, 1, \ldots, p. \end{aligned}$$

Note that $c_{kl} = c_{lk}$, $c^*_{kl} = c^*_{lk}$. Further define $\mathbf{c}_0 = (c_{10}, c_{20}, \ldots, c_{p0})'$ and $\mathbf{c}^*_0 = (c^*_{10}, c^*_{20}, \ldots, c^*_{p0})'$. $\mathbf{C}$ is the $p \times p$ matrix with the $(k,\ l)^{th}$ element c_{kl} and $\mathbf{C}^*$ is the $p \times p$ matrix with the $(k, l)^{th}$ element c^*_{kl}. We add a hat ('$\hat{\ }$') to these symbols to indicate evaluating at $\hat{\boldsymbol{\beta}}$.

16. 'E' denotes *expected value*, 'tr' is an abbreviation for *trace*, 'Var' means *variance*, and 'Cov' means *covariance*. The operator 'vec' forms a vector by stacking the columns of one matrix, and $\otimes$ denotes the *Kronecker product*. The symbol $\xrightarrow{d}$ stands for *converging in distribution*.

17. The partial derivation of a scalar function $l(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)$ is a $p \times 1$ vector

$$\frac{\partial l(\boldsymbol{\beta},\ \boldsymbol{\phi},\ \sigma^2)}{\partial \boldsymbol{\phi}} = \left(\frac{\partial l(\boldsymbol{\beta},\ \boldsymbol{\phi},\ \sigma^2)}{\partial \phi_1}, \ \ldots\ , \frac{\partial l(\boldsymbol{\beta},\ \boldsymbol{\phi},\ \sigma^2)}{\partial \phi_p}\right)',$$

where $\phi_1, \ldots, \phi_p$ are the components of the vector $\boldsymbol{\phi}$.

18. The second partial derivation of a scalar function $l(\boldsymbol{\beta},\ \boldsymbol{\phi},\ \sigma^2)$ is a $p \times r$ matrix

$$\frac{\partial^2 l(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)}{\partial \boldsymbol{\phi} \partial \boldsymbol{\beta}'} = \begin{pmatrix} \frac{\partial^2 l}{\partial \phi_1 \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \phi_1 \partial \beta_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \phi_p \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \phi_p \partial \beta_r} \end{pmatrix},$$

where $\phi_1, \ldots, \phi_p$ and $\beta_1, \ldots, \beta_r$ are the components of the vector $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$.

CHAPTER II

APPLICATION OF THE ESTIMATING METHODS TO THE MODEL WITH AR(1) ERRORS

In this chapter, we study in detail various methods of estimation and their asymptotic properties for the first order autoregressive time series regression model. In Section 2.1, we present the model with AR(1) and discuss the moment and maximum likelihood methods for estimating the parameters in Section 2.2. The quasi-least squares estimating method is presented in Section 2.3. Asymptotic properties with explicit expressions of all three estimating methods are obtained in Section 2.4, assuming the number of replications go to infinity. Section 2.5 summaries the results for a special case when all $t_i \equiv t$. Comparisons are made by examining the asymptotic relative efficiencies through simulation in Section 2.6. We contrast the different methods of estimation using a dental study data in Section 2.7.

2.1 Model

Suppose that the data come from a first order autoregressive (AR(1)) process. The maximum likelihood approach for this model was discussed by Hasza (1980) for a single replication. The model is given by (1.1.7) and while (1.1.2) reduces to

$$\varepsilon_j = \phi\varepsilon_{j-1} + a_j. \tag{2.1.1}$$

The process satisfies the invertible condition automatically, but the stationary condition requires that $|\phi| < 1$. Using (1.1.5), the autocorrelation function satisfies the first order difference equation

$$\rho_k = \phi\rho_{k-1}, \quad k \geq 1,$$

which, with $\rho_0 = 1$, has the solution

$$\rho_k = \phi^k, \; k \geq 0. \tag{2.1.2}$$

On dividing throughout (1.1.4) by $\gamma_0 = \sigma_\varepsilon^2$ and replace ρ_1 with ϕ, the variance σ_ε^2 can be written as

$$\sigma_\varepsilon^2 = \frac{\sigma^2}{1-\phi^2}. \tag{2.1.3}$$

Thus, the error $\boldsymbol{\varepsilon}_i$ has mean $\mathbf{0}$ and covariance

$$\boldsymbol{\Gamma}_i(\phi, \sigma^2) = \sigma_\varepsilon^2 \mathbf{P}_i(\phi) = \sigma^2 \mathbf{V}_i(\phi) \tag{2.1.4}$$

where $\boldsymbol{\Gamma}_i(\phi,\, \sigma^2)$ is the $t_i \times t_i$ covariance matrix, $\mathbf{P}_i(\phi)$ is the $t_i \times t_i$ correlation matrix with the k^{th} diagonal element equal to ϕ^{k-1} and

$$\mathbf{V}_i(\phi) = \frac{1}{1-\phi^2}\mathbf{P}_i(\phi). \tag{2.1.5}$$

Note that $\mathbf{V}_i(\phi)$ is a function of ϕ only. The inverse of $\mathbf{V}_i(\phi)$ is given by

$$\mathbf{V}_i^{-1}(\phi) = \mathbf{C}_{i00} - 2\phi\mathbf{C}_{i10} + \phi^2\mathbf{C}_{i11}, \tag{2.1.6}$$

where $\mathbf{C}_{i00}$ is simply the t_i-dimensional identity matrix, $2\,\mathbf{C}_{i10}$ is a $t_i \times t_i$ tridiagonal matrix with 0's on the main diagonal and 1's on the upper and lower diagonals, and $\mathbf{C}_{i11}$ is the identity matrix with the first and last elements zero. We can write $\mathbf{V}_i^{-1}(\phi) = \mathbf{L}_i\mathbf{L}_i'$, where $\mathbf{L}_i$ is the Cholesky decomposition of $\mathbf{V}_i^{-1}(\phi)$, and $\mathbf{L}_i$ is lower triangular with first diagonal element equals to $\sqrt{1-\phi^2}$, remaining diagonal elements equal to 1, elements in the first off diagonal is $-\phi$, and 0 elsewhere, that is

$$\mathbf{L}_i = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & \cdots & 0 \\ -\phi & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\phi & 1 \end{bmatrix}_{t_i \times t_i}.$$

The adjusted sum of square errors $S(\boldsymbol{\beta}, \boldsymbol{\lambda})$ shown in (1.2.9) becomes

$$\begin{aligned} S(\boldsymbol{\beta}, \phi) &= \sum_{i=1}^{n} \operatorname{tr}(\mathbf{V}_i^{-1}\mathbf{U}_i) \\ &= n \cdot (c_{00} - 2\phi c_{10} + \phi^2 c_{11}), \end{aligned} \tag{2.1.7}$$

where

$$\begin{aligned} c_{00} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \operatorname{tr}(\mathbf{U}_i) &&= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{t_i} \varepsilon_{ij}^2, \\ c_{10} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \operatorname{tr}(\mathbf{C}_{i10}\mathbf{U}_i) &&= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \varepsilon_{ij}\varepsilon_{i(j+1)}, \\ c_{11} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \operatorname{tr}(\mathbf{C}_{i11}\mathbf{U}_i) &&= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{j=2}^{t_i-1} \varepsilon_{ij}^2. \end{aligned}$$

A more general definition of c_{kl}'s is in Section 4.1.

2.2 Moment and maximum likelihood estimates

Recall that the GLS estimates of $\boldsymbol{\beta}$ and σ^2 as in (1.2.10) and (1.2.12) are given by

$$\hat{\boldsymbol{\beta}}_g = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1} \cdot \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i, \tag{2.2.8}$$

$$\hat{\sigma}_g^2 = \frac{1}{\bar{t}}(c_{00} - 2c_{10}\phi + c_{11}\phi^2), \tag{2.2.9}$$

where $\bar{t} = \sum_{i=1}^{n} t_i/n$. Given the "residuals" $\hat{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$, by setting $\boldsymbol{\Gamma} = \hat{\varepsilon}_i\hat{\varepsilon}_i'$ the variance of y_{ij} is estimated by $\hat{c}_{00}/\bar{t}$, and the first order autocovariance is estimated by $\hat{c}_{10}/(\bar{t}-1)$. Thus the moment estimate of $\rho_1 = \phi$ is

$$\hat{\phi} = r_1 = \frac{\bar{t}\hat{c}_{10}}{(\bar{t}-1)\hat{c}_{00}}. \tag{2.2.10}$$

Hence, the moment estimates $(\hat{\boldsymbol{\beta}}_m, \hat{\phi}_m)$ are the simultaneous solutions of (2.2.8) and (2.2.10) and $\hat{\sigma}_m^2$ is obtained by (2.2.9) plugging in $(\hat{\boldsymbol{\beta}}_m, \hat{\phi}_m)$. Note that some authors have used the following moment estimate

$$\hat{\phi}_m' = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t_i-1} \hat{\varepsilon}_{ij}\hat{\varepsilon}_{i(j+1)}/(t_i-1)}{\sum_{i=1}^{n}\sum_{j=1}^{t_i} \hat{\varepsilon}_{ij}^2/t_i}. \tag{2.2.11}$$

The exact expressions of MLEs are difficult to obtain, but not impossible in this case. Assuming the error ε_i's are normal, the likelihood function of $\mathbf{y}$ is given by

$$L(\boldsymbol{\beta}, \phi, \sigma^2) = \frac{1}{2\pi\sigma^{2(n\bar{t})/2}\Pi_{i=1}^{n}|\mathbf{V}_i|^{1/2}} \exp\left(-\frac{\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma^2}\right).$$

The log-likelihood function is given by

$$l(\boldsymbol{\beta}, \phi, \sigma^2) = -\frac{n\bar{t}}{2}\log(2\pi) - \frac{n\bar{t}}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{V}_i| - \frac{S(\boldsymbol{\beta}, \phi)}{2\sigma^2}, \qquad (2.2.12)$$

where $S(\boldsymbol{\beta}, \phi)$ is defined in (2.1.7). Equating to zero the partial derivative of the equation above with respect to $\boldsymbol{\beta}$ and σ^2 we obtain (2.2.8) and (2.2.9), respectively. The determinant of $\mathbf{V}_i$ is given by

$$|\mathbf{V}_i| = |\mathbf{L}_i|^{-2} = \frac{1}{1-\phi^2},$$

which yields,

$$\frac{\partial \log|\mathbf{V}_i|}{\partial\phi} = \frac{2\phi}{1-\phi^2}.$$

From (2.1.6) we obtain

$$\frac{\partial \mathbf{V}_i^{-1}(\phi)}{\partial\phi} = 2\left(\mathbf{C}_{i11}\phi - \mathbf{C}_{i10}\right). \qquad (2.2.13)$$

Thus the partial derivative of the log-likelihood function (2.2.12) with respect to ϕ is

$$\frac{\partial l}{\partial\phi} = -n\left(\frac{\phi}{1-\phi^2} + \frac{c_{11}\phi - c_{10}}{\sigma^2}\right). \qquad (2.2.14)$$

This gives the ML equation of ϕ as

$$c_{11}\phi^3 - c_{10}\phi^2 - (c_{11} + \sigma^2)\phi + c_{10} = 0. \qquad (2.2.15)$$

If σ^2 is known, it has been shown by Haiza(1980) that (2.2.15) has exactly one root in the interval (-1, 1) and can be obtained in a closed form; if σ^2 is unknown, we substitute in (2.2.15) the estimate of σ^2 given in (2.2.9) and obtain

$$(\bar{t}-1)c_{11}\phi^3 - (\bar{t}-2)c_{10}\phi^2 - (\bar{t}c_{11} + c_{00})\phi + \bar{t}c_{10} = 0. \qquad (2.2.16)$$

We can similarly show that there exists a unique closed solution in $(-1,1)$

$$\hat{\phi} = \frac{1}{3(\bar{t}-1)c_{11}} \left[(\bar{t}-2)c_{10} - 2a\cos(\alpha)\right], \tag{2.2.17}$$

where

$$\begin{aligned}
\alpha &= \frac{\pi}{3} + \frac{1}{3}\arccos\left(a^{-3}b\right), \\
a &= \sqrt{(\bar{t}-2)^2 c_{10}^2 + 3(\bar{t}-1)c_{00}c_{11} + 3\bar{t}(\bar{t}-1)c_{11}^2}, \\
b &= \frac{1}{2}c_{10}\left[2(\bar{t}-2)^3 c_{10}^2 + 9(\bar{t}-1)(\bar{t}-2)c_{00}c_{11}\right. \\
&\qquad \left. -9\bar{t}(\bar{t}-1)(2\bar{t}-1)c_{11}^2\right].
\end{aligned}$$

Thus, the ML estimates $(\hat{\boldsymbol{\beta}}_l, \hat{\phi}_l)$ are the simultaneous solutions of (2.2.8) and (2.2.17) and $\hat{\sigma}_l^2$ is then obtained plugging in $(\hat{\boldsymbol{\beta}}_l, \hat{\phi}_l)$ in (2.2.9).

2.3 Quasi-least squares estimates

We now derive the quasi-least squares estimating method. Equating to zero the partial derivative of $S(\boldsymbol{\beta}, \phi)$ with respect to $\boldsymbol{\beta}$ we get (2.2.8). The estimate of ϕ is obtained in two steps as follows. Equating to zero the partial derivative of $S(\boldsymbol{\beta}, \phi)$ with respect to ϕ and noting (1.2.20), we get

$$\sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi}\mathbf{U}_i\right) = 0.$$

Using the result of (2.2.13) we have

$$2(c_{11}\phi - c_{10}) = 0.$$

This gives the first step estimate, which is also the ULS estimate as

$$\hat{\phi}_u = \frac{c_{10}}{c_{11}}.$$

In step 2, we modify $\hat{\phi}_u$ to be consistent. Here we solve the equation

$$\begin{aligned}\sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\phi}_u)}{\partial \phi} \mathbf{V}_i\right) &= \frac{2}{1-\phi^2} \sum_{i=1}^{n} \left[\operatorname{tr}(\mathbf{C}_{i11}\mathbf{P}_i)\hat{\phi}_u - \operatorname{tr}(\mathbf{C}_{i10}\mathbf{P}_i)\right] \\ &= \frac{2}{1-\phi^2} \sum_{i=1}^{n} \left[(t_i - 2)\hat{\phi}_u - (t_i - 1)\phi)\right] = 0\end{aligned}$$

and the solution is

$$\hat{\phi} = \frac{\bar{t}-2}{\bar{t}-1}\hat{\phi}_u = \frac{(\bar{t}-2)c_{10}}{(\bar{t}-1)c_{11}}. \tag{2.3.18}$$

Thus the QLS estimates of $(\boldsymbol{\beta}, \phi)$ are the simultaneous solutions of (2.2.8) and (2.3.18), and the estimate of $\hat{\sigma}_q^2$ is obtained plugging $\hat{\phi}_q$ in (2.2.9). We can see that the QLS estimate of ϕ will always be less than the ULS estimate and it will always be greater than the moment estimate $\hat{\phi}_m$, provided that all t_i's are equal and the estimates of $\boldsymbol{\beta}$ are the same.

2.4 Asymptotic properties

In time series analysis, we usually consider the asymptotic properties when the number of observations t_i goes to infinity whereas here we consider it by fixing t_i and letting the number of replications n goes to infinity. For convenience, we first derive the asymptotic property of the MLEs by finding the information matrix.

The asymptotic properties of MLEs

We have the following results regarding to the first and second derivatives of the log-likelihood (2.2.12) with respect to $(\boldsymbol{\beta}, \phi, \sigma^2)$

$$\begin{aligned}\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}), \\ \frac{\partial l}{\partial \phi} &= -\frac{n\phi}{1-\phi^2} - \frac{n}{\sigma^2}(c_{11}\phi - c_{10}),\end{aligned}$$

$$
\begin{aligned}
\frac{\partial l}{\partial \sigma^2} &= -\frac{n\bar{t}}{2\sigma^2} + \frac{n}{2\sigma^4}\left(c_{00} - 2\phi\, c_{10} + \phi^2 c_{11}\right), \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta}^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i, \\
\frac{\partial^2 l}{\partial \phi^2} &= -\frac{n(1+\phi^2)}{(1-\phi^2)^2} - \frac{c_{11}}{\sigma^2}, \\
\frac{\partial^2 l}{\partial (\sigma^2)^2} &= \frac{n\bar{t}}{2\sigma^4} - \frac{n}{\sigma^6}\left(c_{00} - 2\phi\, c_{10} + \phi^2 c_{11}\right), \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \phi} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i' \frac{\partial \mathbf{V}_i^{-1}}{\partial \phi}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \\
\frac{\partial^2 l}{\partial \sigma^2 \partial \phi} &= \frac{n}{\sigma^4}\left(c_{11}\phi - c_{10}\right).
\end{aligned}
$$

Since

$$
\begin{aligned}
\mathrm{E}(c_{00}) &= \frac{\bar{t}\sigma^2}{1-\phi^2} \\
\mathrm{E}(c_{10}) &= \frac{(\bar{t}-1)\sigma^2\phi}{1-\phi^2} \\
\mathrm{E}(c_{11}) &= \frac{(\bar{t}-2)\sigma^2}{1-\phi^2},
\end{aligned}
$$

and using the formulas above we get

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}^2}\right) &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i, \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial (\sigma^2)^2}\right) &= -\frac{n\bar{t}}{2\sigma^4}, \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \phi^2}\right) &= -\frac{n}{(1-\phi^2)^2}\left[2\phi^2 + (\bar{t}-1)(1-\phi^2)\right], \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2}\right) &= \mathbf{0}, \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \phi}\right) &= \mathbf{0}, \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \phi \partial \sigma^2}\right) &= -\frac{n\phi}{(1-\phi^2)\sigma^2}.
\end{aligned}
$$

Thus the information matrix is given by

$$\mathbf{I}_n(\boldsymbol{\beta}, \phi, \sigma^2) = \begin{bmatrix} \frac{\sum_{i=1}^n \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i}{\sigma^2} & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \frac{n\left[2\phi^2 + (\bar{t}-1)(1-\phi^2)\right]}{(1-\phi^2)^2} & \frac{n\phi}{(1-\phi^2)\sigma^2} \\ \mathbf{0} & \frac{n\phi}{(1-\phi^2)\sigma^2} & \frac{n\bar{t}}{2\sigma^4} \end{bmatrix}. \tag{2.4.19}$$

Note that the determinant of the right bottom part $\mathbf{B}$ of $\mathbf{I}_n(\boldsymbol{\beta}, \phi, \sigma^2)$ is

$$\begin{aligned} |\mathbf{B}| &= \frac{n^2\bar{t}\left[2\phi^2 + (\bar{t}-1)(1-\phi^2)\right]}{2\sigma^4(1-\phi^2)^2} - \frac{n^2\phi^2}{\sigma^4(1-\phi^2)^2} \\ &= \frac{n^2(\bar{t}-1)\left[2\phi^2 + \bar{t}(1-\phi^2)\right]}{2\sigma^4(1-\phi^2)^2}. \end{aligned}$$

By finding the inverse of $\mathbf{B}$, we get

$$\mathbf{I}_n^{-1} = \frac{1}{n}\boldsymbol{\Sigma}_l, \tag{2.4.20}$$

where $\boldsymbol{\Sigma}_l = \text{diag}(v_1,\ \boldsymbol{\Sigma}_{2l})$, $\boldsymbol{\Sigma}_{2l} = \left[\begin{smallmatrix} v_{2l} & v_{23l} \\ v_{23l} & v_{3l} \end{smallmatrix}\right]$ and

$$v_1 = \sigma^2 \left[\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i'\,\mathbf{V}_i(\phi)\,\mathbf{X}_i\right]^{-1}, \tag{2.4.21}$$

$$v_{2l} = \frac{\bar{t}(1-\phi^2)^2}{(\bar{t}-1)[2\phi^2 + \bar{t}(1-\phi^2)]}, \tag{2.4.22}$$

$$v_{3l} = \frac{2\sigma^4\left[2\phi^2 + (\bar{t}-1)(1-\phi^2)\right]}{(\bar{t}-1)[2\phi^2 + \bar{t}(1-\phi^2)]}, \tag{2.4.23}$$

$$v_{23l} = -\frac{2\sigma^2\phi(1-\phi^2)}{(\bar{t}-1)[2\phi^2 + \bar{t}(1-\phi^2)]}.$$

Thus, we have the following theorem,

Theorem 2.1 *Consider the model (1.1.7). Assume that* $t_i \le t < \infty$ *for all* $i = 1, \ldots, n$, *and the errors* ε_i*'s are independent and normally distributed with mean zero and covariance* $\sigma^2\mathbf{V}_i$ *for* $i = 1, \ldots, n$. *Let* $\boldsymbol{\xi} = (\boldsymbol{\beta},\ \phi,\ \sigma^2)'$, *and* $\hat{\boldsymbol{\xi}}_l = (\hat{\boldsymbol{\beta}}_l,\ \hat{\phi}_l,\ \hat{\sigma}_l^2)'$ *be the ML estimates of* $\boldsymbol{\xi}$. *We have*

$$\hat{\boldsymbol{\xi}}_l \xrightarrow{d} N\left(\boldsymbol{\xi}_l,\ \ \mathbf{I}_n^{-1}\right), \quad as \quad n \to \infty,$$

where $\xrightarrow{d}$ *means converging in distribution and* $\mathbf{I}_n^{-1}$ *is given by (2.4.20).*

It is clear from Theorem 2.1 that the ML estimate of $\boldsymbol{\beta}$ is uncorrelated with $\hat{\phi}_l$ and $\hat{\sigma}_l^2$, but $\hat{\phi}_l$ and $\hat{\sigma}_l^2$ are negatively correlated.

Now we study the asymptotic properties of the moment and QLS estimates. Since ε_i's are independent, under simple conditions we can show that $\hat{\boldsymbol{\beta}}_m$ and $\hat{\boldsymbol{\beta}}_q$ are consistent and asymptotically efficient as $n \to \infty$. Further, $\hat{\phi}_q$ and $\hat{\phi}_m$ are consistent estimates of ϕ (see Theorem 2.2 and Remark 2.2). However, the asymptotic distributions of $\hat{\phi}_m$ and $\hat{\phi}_q$ depend on the higher order moments of the errors ε_i's. For comparison purpose, we assume normality for ε_i's. First we state Lemma 2.1 which is a simple extension of the result in Joe (1997, p. 301) and can be proved by Taylor's theorem. We will need this to establish Theorem 2.2.

Lemma 2.1 *Let $\mathbf{z}_i$ be independent random vectors of dimensions t_i, $1 \le i \le n$. Assume that $t_i \le t$ for all i. Let $\boldsymbol{\xi}$ be a parameter of fixed dimension, and the multivariate functions $h_i(\mathbf{z}_i, \boldsymbol{\xi})$ be such that*

$$\sum_{i=1}^{n} \mathrm{E}\left[h_i(\mathbf{z}_i,\, \boldsymbol{\xi})\right] = \mathbf{0}. \tag{2.4.24}$$

Define $\mathbf{M}_n(\boldsymbol{\xi}) = \frac{1}{n}\sum_{i=1}^{n} Cov\left(h_i(\mathbf{z}_i,\, \boldsymbol{\xi})\right)$ and $\mathbf{I}_n(\boldsymbol{\xi}) = -\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(\partial\, h_i(\boldsymbol{\xi})/\partial\, \boldsymbol{\xi}')$. Here $\mathbf{I}_n(\boldsymbol{\xi})$ may not necessarily be the information matrix. Suppose $\hat{\boldsymbol{\xi}}$ is the solution of the unbiased estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} h_i(\mathbf{z}_i,\, \boldsymbol{\xi}) = 0. \tag{2.4.25}$$

Then, under usual regularity conditions we have

$$\hat{\boldsymbol{\xi}} \xrightarrow{d} N\left(\boldsymbol{\xi},\quad \frac{1}{n}\,\mathbf{I}_n^{-1}\,\mathbf{M}_n\,\mathbf{I}_n^{-1}\right), \quad as \quad n \to \infty. \tag{2.4.26}$$

Theorem 2.2 *Consider the model (1.1.7). Assume that* $t_i \le t < \infty$ *for all* $i = 1, \ldots, n$, *and the errors* ε_i*'s are independent and normally distributed with mean zero and covariance matrix* $\sigma^2 \mathbf{V}_i$ *for* $i = 1, \ldots, n$. *Let* $\boldsymbol{\xi} = (\beta, \phi)'$, *and* $\hat{\boldsymbol{\xi}}_q = (\hat{\beta}_q, \hat{\phi}_q)'$ *and* $\hat{\boldsymbol{\xi}}_m = (\hat{\beta}_m, \hat{\phi}_m)'$ *be the moment and the QLS estimates of* $\boldsymbol{\xi}$ *respectively. Then we have*

$$\hat{\boldsymbol{\xi}}_i \xrightarrow{d} N\left(\boldsymbol{\xi}, \ \frac{1}{n}\boldsymbol{\Sigma}_{1i}\right), \quad as \quad n \to \infty, \quad i = m, \ q,$$

where $\boldsymbol{\Sigma}_{1i} = \mathrm{diag}(v_1, v_{2i})$, $i = m, q$, v_1 *is defined in (2.4.21) and*

$$\begin{aligned} v_{2m} &= \frac{1}{\bar{t}^2(\bar{t}-1)^2(1-\phi^2)^2}\Big[\bar{t}^2(\bar{t}-1) - \bar{t}(3\bar{t}^2 - 5\bar{t} + 6)\phi^2 + (\bar{t}-1)(3\bar{t}^2 - 4\bar{t} + 4)\phi^4 \\ &\quad -\bar{t}(\bar{t}-1)(\bar{t}-2)\phi^6 + \left(\bar{t} - (\bar{t}-1)\phi^2\right)^2 \frac{4}{n}\sum_{i=1}^{n}\phi^{2t_i}\Big], \qquad (2.4.27) \\ v_{2q} &= \frac{1}{(\bar{t}-1)^2(\bar{t}-2)^2(1-\phi^2)^2}\Big[(\bar{t}-1)(\bar{t}-2)^2 - \bar{t}(\bar{t}-2)(3\bar{t}-7)\phi^2 \\ &\quad +\bar{t}(\bar{t}-1)(3\bar{t}-8)\phi^4 - \bar{t}(\bar{t}-1)(\bar{t}-2)\phi^6 + \frac{4}{n}\sum_{i=1}^{n}\phi^{2t_i}\Big]. \qquad (2.4.28) \end{aligned}$$

Proof: Let $\mathbf{P}_i(\phi)$ and $\mathbf{C}_{i10}, \mathbf{C}_{i11}$ are as defined in (2.1.4) and (2.1.6), we have the following identities:

$$\begin{aligned}
&\mathrm{tr}(\mathbf{P}_i) && = t_i, \\
&\mathrm{tr}(\mathbf{C}_{i11}\,\mathbf{P}_i) && = t_i - 2, \\
&\mathrm{tr}(\mathbf{C}_{i10}\,\mathbf{P}_i) && = (t_i - 1)\phi, \\
&v_{i11} = (1-\phi^2)^2\,\mathrm{tr}(\mathbf{P}_i\,\mathbf{P}_i) && = t_i(1-\phi^4) - 2\phi^2(1-\phi^{2t_i}), \\
&v_{i12} = (1-\phi^2)^2\,\mathrm{tr}(\mathbf{P}_i\,\mathbf{C}_{i10}\,\mathbf{P}_i) && = 2\phi\left[(t_i - 1) - t_i\phi^2 + \phi^{2t_i}\right], \\
&v_{i13} = (1-\phi^2)^2\,\mathrm{tr}(\mathbf{P}_i\,\mathbf{C}_{i11}\,\mathbf{P}_i) && = t_i(1-\phi^4) - 2(1-\phi^{2t_i}), \\
&v_{i22} = (1-\phi^2)^2\,\mathrm{tr}(\mathbf{C}_{i10}\,\mathbf{P}_i\,\mathbf{C}_{i10}\,\mathbf{P}_i) && = \frac{1}{2}\Big[(t_i - 1)(1 - 5\phi^4 + \phi^6) \\
& && \qquad +(3t_i - 7)\phi^2 + 4\phi^{2t_i}\Big], \\
&v_{i23} = (1-\phi^2)^2\,\mathrm{tr}(\mathbf{C}_{i11}\,\mathbf{P}_i\,\mathbf{C}_{i10}\,\mathbf{P}_i) && = 2\left[(t_i - 2)\phi - (t_i - 1)\phi^3 + \phi^{2t_i - 1}\right], \\
&v_{i33} = (1-\phi^2)^2\,\mathrm{tr}(\mathbf{C}_{i11}\,\mathbf{P}_i\,\mathbf{C}_{i11}\,\mathbf{P}_i) && = (t_i - 2)(1-\phi^4) - 2\phi^2(1-\phi^{2t_i - 4}).
\end{aligned}$$

Now let us set up the notation needed in Lemma 2.1. Let $\mathbf{z}_i = \boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i\,\boldsymbol{\beta}$, note that $\mathrm{E}(\mathbf{z}_i) = \mathbf{0}$ and $Cov(\mathbf{z}_i) = \sigma^2\,\mathbf{V}_i(\phi)$. Let $\boldsymbol{\xi} = (\boldsymbol{\beta},\, \phi)'$ and

$$h_i(\boldsymbol{\xi}) = \begin{bmatrix} \mathbf{X}_i'\,\mathbf{V}_i^{-1}(\phi)\,\mathbf{z}_i \\ \mathbf{z}_i'\,\mathbf{A}_i(\phi)\,\mathbf{z}_i \end{bmatrix}$$

where the matrix $\mathbf{A}_i(\phi)$ is a function of ϕ such that the estimate $\boldsymbol{\xi}$ is the solution of (2.4.25). We have

$$\frac{\partial h_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}'} = \begin{bmatrix} -\mathbf{X}_i'\mathbf{V}^{-1}\mathbf{X}_i & \mathbf{X}_i'\,\dfrac{\partial \mathbf{V}^{-1}}{\partial \phi}\,\mathbf{z}_i \\ -2\mathbf{X}_i'\mathbf{A}_i\mathbf{z}_i & \mathbf{z}_i'\,\dfrac{\partial \mathbf{A}_i}{\partial \phi}\,\mathbf{z}_i \end{bmatrix},$$

which gives

$$\mathrm{E}\left(\frac{\partial h_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}'}\right) = \begin{bmatrix} -\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i & \mathbf{0} \\ \mathbf{0}' & \sigma^2 \mathrm{tr}\left(\dfrac{\partial \mathbf{A}_i}{\partial \phi}\,\mathbf{V}_i\right) \end{bmatrix}.$$

This yields

$$\mathbf{I}_n(\boldsymbol{\xi}) = \begin{bmatrix} \sigma^2 v_1^{-1} & \mathbf{0}' \\ \mathbf{0} & -\dfrac{\sigma^2}{(1-\phi^2)}\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathrm{tr}\left(\dfrac{\partial \mathbf{A}_i}{\partial \phi}\,\mathbf{P}_i\right) \end{bmatrix}, \tag{2.4.29}$$

where v_1 is defined in (2.4.21). The covariance of $h_i(\boldsymbol{\xi})$ is given by

$$\mathrm{Cov}(h_i(\boldsymbol{\xi})) = \begin{bmatrix} \sigma^2\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i & \mathbf{X}_i'\mathbf{V}_i^{-1}\mathrm{E}(\mathbf{z}_i\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i) \\ \mathrm{E}'(\mathbf{z}_i\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i)\mathbf{V}_i^{-1}\mathbf{X}_i & \mathrm{Var}(\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i) \end{bmatrix}.$$

Under the assumption that $\mathbf{z}_i$'s are normally distributed we have

$$\begin{aligned} \mathrm{E}(\mathbf{z}_i\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i) &= \mathbf{0} \\ \mathrm{Var}(\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i) &= \frac{2\sigma^4}{(1-\phi^2)^2}\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)^2. \end{aligned}$$

Hence

$$\mathbf{M}_n(\boldsymbol{\xi}) = \begin{bmatrix} \sigma^4 v_1^{-1} & \mathbf{0}' \\ \mathbf{0} & \dfrac{2\sigma^4}{(1-\phi^2)^2}\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathrm{tr}(\mathbf{A}_i\,\mathbf{P}_i)^2 \end{bmatrix}. \tag{2.4.30}$$

Now we just need to show that (2.4.24) is satisfied and find $\mathbf{I}_n^{-1}\mathbf{M}_n\mathbf{I}_n^{-1}$ for different choice of $\mathbf{A}_i(\phi)$. Clearly, $\mathrm{E}(\mathbf{x}_i'\mathbf{V}_i^{-1}\mathbf{z}_i) = \mathbf{0}$.

a) For moment estimates, select $\mathbf{A}_i(\phi) = \bar{t}\mathbf{C}_{i10} - (\bar{t}-1)\phi\,\mathbf{C}_{i00}$. We have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{E}(\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i) &= \frac{\sigma^2}{1-\phi^2}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)\\ &= \frac{\sigma^2}{1-\phi^2}\left[\bar{t}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i10}\mathbf{P}_i) - (\bar{t}-1)\phi\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i00}\mathbf{P}_i)\right] = 0.\end{aligned}$$

Thus equation (2.4.24) is satisfied. Since $\partial\mathbf{A}_i(\phi)/\partial\phi = -(\bar{t}-1)\mathbf{C}_{i00}$, which implies that $\mathrm{tr}\left[(\partial\mathbf{A}_i/\partial\phi)\mathbf{P}_i\right] = -(\bar{t}-1)t_i$, (2.4.29) becomes

$$\mathbf{I}_n(\boldsymbol{\xi}) = \begin{bmatrix}\sigma^2 v_1^{-1} & \mathbf{0}' \\ \mathbf{0} & \dfrac{\sigma^2\bar{t}(\bar{t}-1)}{1-\phi^2}\end{bmatrix}. \qquad (2.4.31)$$

Further, we have

$$\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)^2 = \bar{t}^2\mathrm{tr}(\mathbf{C}_{i10}\mathbf{P}_i\mathbf{C}_{i10}\mathbf{P}_i) - 2\bar{t}(\bar{t}-1)\phi\mathrm{tr}(\mathbf{C}_{i10}\mathbf{P}_i\mathbf{P}_i) + (\bar{t}-1)^2\phi^2\mathrm{tr}(\mathbf{P}_i\mathbf{P}_i),$$

which yields,

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)^2 = \frac{1}{2}\bar{t}^2(\bar{t}-1)^2\,v_{2m}.$$

Substituting the expression above in (2.4.30) we get

$$\mathbf{M}_n(\boldsymbol{\xi}) = \begin{bmatrix}\sigma^4 v_1^{-1} & \mathbf{0}' \\ \mathbf{0} & \dfrac{\sigma^4\bar{t}^2(\bar{t}-1)^2}{(1-\phi^2)^2}\,v_{2m}\end{bmatrix}. \qquad (2.4.32)$$

Combining with (2.4.31) we get $\mathbf{I}_n^{-1}\mathbf{M}_n\mathbf{I}_n^{-1} = \boldsymbol{\Sigma}_{1m}$.

b) For QLS estimates, select $\mathbf{A}_i(\phi) = (\bar{t}-2)\,\mathbf{C}_{i10} - (\bar{t}-1)\phi\,\mathbf{C}_{i11}$. We have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{E}(\mathbf{z}_i'\mathbf{A}_i\mathbf{z}_i) &= \frac{\sigma^2}{1-\phi^2}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)\\ &= \frac{\sigma^2}{1-\phi^2}\left[(\bar{t}-2)\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i10}\mathbf{P}_i) - (\bar{t}-1)\phi\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i11}\mathbf{P}_i)\right],\end{aligned}$$

which implies that (2.4.24) is satisfied. Since $\partial\mathbf{A}_i/\partial\phi = -(\bar{t}-1)\mathbf{C}_{i11}$, we get $\mathrm{tr}\left[(\partial\mathbf{A}_i/\partial\phi)\,\mathbf{P}_i\right] = -(\bar{t}-1)(t_i-2)$, and (2.4.29) reduces to

$$\mathbf{I}_n(\boldsymbol{\xi}) = \begin{bmatrix}\sigma^2 v_1^{-1} & \mathbf{0}' \\ \mathbf{0} & \dfrac{\sigma^2(\bar{t}-1)(\bar{t}-2)}{(1-\phi^2)}\end{bmatrix}. \qquad (2.4.33)$$

Further we have

$$\begin{aligned}\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)^2 &= (\bar{t}-1)^2\phi^2\,\mathrm{tr}(\mathbf{C}_{i11}\mathbf{P}_i\mathbf{C}_{i11}\mathbf{P}_i) - 2\phi(\bar{t}-1)(\bar{t}-2)\mathrm{tr}(\mathbf{C}_{i11}\mathbf{P}_i\mathbf{C}_{i10}\mathbf{P}_i)\\ &\quad +(\bar{t}-2)^2\mathrm{tr}(\mathbf{C}_{i10}\mathbf{P}_i\mathbf{C}_{i10}\mathbf{P}_i),\end{aligned}$$

which yields,

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{A}_i\mathbf{P}_i)^2 = \frac{1}{2}(\bar{t}-1)^2(\bar{t}-2)^2 v_{2q}.$$

Substituting the expression above in (2.4.30) we get

$$\mathbf{M}_n(\boldsymbol{\xi}) = \begin{bmatrix}\sigma^4 v_1^{-1} & \mathbf{0}' \\ \mathbf{0} & \dfrac{\sigma^4}{(1-\phi^2)^2}(\bar{t}-1)^2(\bar{t}-2)^2 v_{2q}\end{bmatrix}. \tag{2.4.34}$$

Combining with (2.4.33) we get $\mathbf{I}_n\mathbf{M}_n\mathbf{I}_n^{-1} = \boldsymbol{\Sigma}_{1q}$. This complete the proof. ◁

Remark 2.1 *If we use (2.2.11) as the moment estimate of ϕ, then the asymptotic variance of $\hat{\phi}_m$ is given by*

$$\begin{aligned} v'_{2m} &= \frac{1}{(1-\phi^2)^2}\Bigg[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(t_i-1)} - \frac{\phi^2}{n}\sum_{i=1}^{n}\frac{(6-5t_i+3t_i^2)}{t_i\,(t_i-1)^2}\\ &\quad +\frac{\phi^4}{n}\sum_{i=1}^{n}\frac{(3t_i^2-4t_i+4)}{t_i^2\,(t_i-1)} - \frac{\phi^6}{n}\sum_{i=1}^{n}\frac{(t_i-2)}{t_i(t_i-1)} + \frac{4}{n}\sum_{i=1}^{n}\frac{\phi^{2t_i+4}}{t_i^2}\\ &\quad -\frac{8}{n}\sum_{i=1}^{n}\frac{\phi^{2t_i+2}}{t_i(t_i-1)} + \frac{4}{n}\sum_{i=1}^{n}\frac{\phi^{2t_i}}{(t_i-1)^2}\Bigg].\end{aligned}$$

As we can see from Theorem 2.2, the moment and QLS estimates of $\boldsymbol{\beta}$ are uncorrelated with both the estimates of ϕ and σ^2. However, the estimates of ϕ and σ^2 are correlated. The following Theorem 2.3 gives the asymptotic covariance of the estimates of $(\phi,\, \sigma^2)$. In order to prove Theorem 2.3, we will need the following Lemma 2.2.

Lemma 2.2 *Consider the model (1.1.7). Assume that $t_i \le t < \infty$, and the errors ε_i's are independent and normally distributed with mean zero and covariance $\sigma^2\mathbf{V}_i$.*

Let $c_{i00} = \mathrm{tr}(\mathbf{C}_{i00}\mathbf{U}_i)$, $c_{i10} = \mathrm{tr}(\mathbf{C}_{i10}\mathbf{U}_i)$, $c_{i11} = \mathrm{tr}(\mathbf{C}_{i11}\mathbf{U}_i)$, $\mathbf{c}_i = (c_{i00},\ c_{i10},\ c_{i11})'$ *for* $i = 1, \ldots, n$ *and* $\hat{\mathbf{c}} = \sum_{i=1}^n \hat{\mathbf{c}}_i/n$, *where '*$\hat{\ }$*' indicates evaluating at some estimate of* $\boldsymbol{\beta}$. *Then we have*

$$\hat{\mathbf{c}} \xrightarrow{d} N\left(\boldsymbol{\mu}_c,\ \ \frac{1}{n}\boldsymbol{\Sigma}_c\right), \quad as \quad n \to \infty, \tag{2.4.35}$$

where $\boldsymbol{\mu}_c = \frac{\sigma^2}{1-\phi^2}(\bar{t},\ (\bar{t}-1)\phi,\ \bar{t}-2)'$, $\boldsymbol{\Sigma}_c = \frac{2\sigma^4}{(1-\phi^2)^4}\{v_{ij}\}$ *with* $v_{ji} = v_{ij}$ *and*

$$\begin{aligned}
v_{11} &= \bar{t}(1-\phi^4) - 2\phi^2 + \frac{2\phi^2}{n}\sum_{i=1}^n \phi^{2t_i}, \\
v_{12} &= 2\phi\left[(\bar{t}-1) - \bar{t}\phi^2 + \frac{1}{n}\sum_{i=1}^n \phi^{2t_i}\right], \\
v_{13} &= (\bar{t}-2) - \bar{t}\phi^4 + \frac{2}{n}\sum_{i=1}^n \phi^{2t_i}, \\
v_{22} &= \frac{1}{2}(\bar{t}-1)(1-5\phi^4+\phi^6) + \frac{1}{2}(3\bar{t}-7)\phi^2 + \frac{2}{n}\sum_{i=1}^n \phi^{2t_i}, \\
v_{23} &= 2\left[(\bar{t}-2)\phi - (\bar{t}-1)\phi^3 + \frac{1}{n}\sum_{i=1}^n \phi^{2t_i-1}\right], \\
v_{33} &= (\bar{t}-2)(1-\phi^4) - 2\phi^2 + \frac{4}{n}\sum_{i=1}^n \phi^{2t_i-2}.
\end{aligned}$$

Proof: Since that $\varepsilon_i \sim N(\mathbf{0}, \sigma^2\mathbf{V}_i)$, we have

$$\hat{\mathbf{c}}_i \xrightarrow{d} N(\boldsymbol{\mu}_{ic},\ \ \boldsymbol{\Sigma}_{ic}), \quad as \quad n \to \infty, \quad for \quad i = 1,\ \ldots,\ n,$$

where $\boldsymbol{\mu}_{ic} = \frac{\sigma^2}{1-\phi^2}(t_i,\ (t_i-1)\phi,\ t_i-2)'$ and $\boldsymbol{\Sigma}_{ic} = \frac{2\sigma^4}{(1-\phi^2)^2}\mathbf{P}\otimes\mathbf{P} = \frac{2\sigma^4}{(1-\phi^2)^4}\{v_{ikl}\}$ is a symmetric matrix such that, $v_{i11}, v_{i12}, v_{i13}, v_{i22}, v_{i23}, v_{i33}$ are the identities defined in the proof of Theorem 2.2. From the central limit theorem, we have

$$\hat{\mathbf{c}} \xrightarrow{d} N\left(\boldsymbol{\mu}_c,\ \ \frac{1}{n}\boldsymbol{\Sigma}_c\right), \quad as \quad n \to \infty,$$

where $\boldsymbol{\mu}_c = \sum_{i=1}^n \boldsymbol{\mu}_{ic}/n$ and $\boldsymbol{\Sigma}_c = \sum_{i=1}^n \boldsymbol{\Sigma}_{ic}/n$. Note that $\sum_{i=1}^n \phi^{2t_i}/n < \phi^{2t}$, thus $\boldsymbol{\Sigma}_c$ is bounded. The proof of the lemma is completed by simplifying $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$.

◁

Theorem 2.3 *Consider the model (1.1.7). Assume that $t_i \leq t < \infty$ for all $i = 1, \ldots, n$, and the errors ε_i's are independent and normally distributed with mean zero and covariance $\sigma^2 \mathbf{V}_i$ for $i = 1, \ldots, n$. Let $\boldsymbol{\xi} = (\phi, \sigma^2)'$, and $\hat{\boldsymbol{\xi}}_m = (\hat{\phi}_m, \hat{\sigma}^2_m)'$ and $\hat{\boldsymbol{\xi}}_q = (\hat{\phi}_q, \hat{\sigma}^2_q)'$ be the moment and the quasi-least squares estimates of $\boldsymbol{\xi}$ respectively. Then we have*

$$\hat{\boldsymbol{\xi}}_i \xrightarrow{d} N\left(\boldsymbol{\xi}, \ \frac{1}{n}\Sigma_{2i}\right), \quad as \quad n \to \infty, \quad i = m, q,$$

where $\Sigma_{2i} = \begin{bmatrix} v_{2i} & v_{23i} \\ v_{23i} & v_{3i} \end{bmatrix}$, $i = m, q$, v_{2m} *and* v_{2q} *are defined in (2.4.27) and (2.4.28), and*

$$v_{3m} = \frac{4\sigma^4\phi^2}{\bar{t}^2(1-\phi^2)^2} v_{2m} + \frac{2\sigma^4}{\bar{t}}, \tag{2.4.36}$$

$$v_{3q} = \frac{4\sigma^4\phi^2}{\bar{t}^2(1-\phi^2)^2} v_{2q} + \frac{2\sigma^4}{\bar{t}}\left[1 - \frac{1}{(1-\phi^2)^4}\right], \tag{2.4.37}$$

$$v_{23i} = -\frac{2\sigma^2\phi}{\bar{t}(1-\phi^2)} v_{2i}, \quad i = m, q.$$

Proof: We first consider the moment estimate. Since $\hat{\boldsymbol{\xi}}_m$ is a function of $\hat{\mathbf{c}}$ in Lemma 2.2, we can apply delta theorem and then use (2.4.35) to prove it. Obviously, $\hat{\boldsymbol{\xi}}_m(\boldsymbol{\mu}_c) = \boldsymbol{\xi}$. We just need to find the asymptotic covariance. From (2.2.9) and (2.2.10) we get

$$\frac{d\hat{\boldsymbol{\xi}}_m}{d\hat{\mathbf{c}}'} = \begin{pmatrix} -\frac{\bar{t}c_{10}}{(\bar{t}-1)c_{00}^2} & \frac{\bar{t}}{(\bar{t}-1)c_{00}} & 0 \\ \frac{1}{\bar{t}} + \frac{2\hat{c}_{10}\hat{\phi}_m}{\hat{c}_{00}} - \frac{2\hat{c}_{11}\hat{\phi}_m^2}{\hat{c}_{00}} & -\frac{2\hat{\phi}_m}{\bar{t}}\left(\frac{\hat{c}_{11}\phi}{\hat{c}_{10}} - 2\right) & \frac{\hat{\phi}_m^2}{\bar{t}} \end{pmatrix},$$

which yields

$$\frac{d\hat{\theta}_m(\boldsymbol{\mu}_c)}{d\hat{\mathbf{c}}'} = \begin{pmatrix} -\frac{\phi(1-\phi^2)}{\bar{t}\sigma^2} & \frac{1-\phi^2}{(\bar{t}-1)\sigma^2} & 0 \\ \frac{1}{\bar{t}} + \frac{2\phi^2}{\bar{t}^2} & -\frac{2\phi}{\bar{t}-1} & \frac{\phi^2}{\bar{t}} \end{pmatrix}.$$

Thus $\Sigma_{2m} = \left[d\hat{\boldsymbol{\xi}}_m(\boldsymbol{\mu}_c)/d\hat{\mathbf{c}}'\right] \Sigma_c \left[d\hat{\boldsymbol{\xi}}_m'(\boldsymbol{\mu}_c)/d\hat{\mathbf{c}}\right]$ and by simplifying it, we complete the proof for the moment estimate. Next we consider the QLS estimate. Again, $\hat{\boldsymbol{\xi}}_m$ is a function of $\hat{\mathbf{c}}$ also. And as before, we can apply delta theorem and then use (2.4.35) to establish the result. Obviously, $\hat{\boldsymbol{\xi}}_q(\boldsymbol{\mu}_c) = \boldsymbol{\xi}$ and we need to find the asymptotic covariance. From (2.2.9) and (2.3.18) we get

$$\frac{d\hat{\boldsymbol{\xi}}_q}{d\hat{\mathbf{c}}'} = \begin{pmatrix} 0 & \dfrac{\bar{t}-2}{(\bar{t}-1)\hat{c}_{11}} & -\dfrac{(\bar{t}-2)\hat{c}_{10}}{(\bar{t}-1)\hat{c}_{11}^2} \\ 1/\bar{t} & -2\hat{\phi}_q/(\bar{t}-1) & \hat{\phi}_q^2/(\bar{t}-2) \end{pmatrix},$$

which yields

$$\frac{d\hat{\boldsymbol{\xi}}_q(\boldsymbol{\mu}_c)}{d\hat{\mathbf{c}}'} = \begin{pmatrix} 0 & \dfrac{1-\phi^2}{(\bar{t}-1)\sigma^2} & -\dfrac{\phi(1-\phi^2)}{(\bar{t}-2)\sigma^2} \\ 1/\bar{t} & -2\phi_q/(\bar{t}-1) & \phi_q^2/(\bar{t}-2) \end{pmatrix}.$$

Thus $\Sigma_{2q} = \left[d\hat{\boldsymbol{\xi}}_q(\boldsymbol{\mu}_c)/d\hat{\mathbf{c}}'\right] \Sigma_c \left[d\hat{\boldsymbol{\xi}}_q'(\boldsymbol{\mu}_c)/d\hat{\mathbf{c}}\right]$ and after some simplification, we get the result for the QLS estimate. This completes the proof of the theorem. $\triangleleft$

Remark 2.2 *In proving Theorems 2.2 and 2.3, the normality assumption is used only to derive explicitly the asymptotic variances of the estimates of* $(\hat{\phi}_m, \sigma_m^2)$ *and* $(\hat{\phi}_q, \sigma_q^2)$. *The assumption is not needed to establish consistency of these estimates nor to show that the estimates* $\hat{\boldsymbol{\beta}}_m$ *and* $\hat{\boldsymbol{\beta}}_q$ *are asymptotically efficient.*

Combining Theorems 2.1, 2.2 and 2.3 we have the following general theorem.

Theorem 2.4 *Consider the model (1.1.7). Assume that* $t_i \leq t < \infty$ *for all* $i = 1, \ldots, n$, *and the errors* $\boldsymbol{\varepsilon}_i$*'s are independent and normally distributed with mean zero and covariance* $\sigma^2\mathbf{V}_i$ *for* $i = 1, \ldots, n$. *Let* $\boldsymbol{\xi} = (\boldsymbol{\beta}, \phi, \sigma^2)'$, *and* $\hat{\boldsymbol{\xi}}_m = (\hat{\boldsymbol{\beta}}_m, \hat{\phi}_m, \hat{\sigma}_m^2)'$ *and* $\hat{\boldsymbol{\xi}}_q = (\hat{\boldsymbol{\beta}}_q, \hat{\phi}_q, \hat{\sigma}_q^2)'$ *be the moment, ML and QLS estimates of* $\boldsymbol{\xi}$

respectively. Then we have

$$\hat{\boldsymbol{\xi}}_i \xrightarrow{d} N\left(\boldsymbol{\xi}, \ \frac{1}{n}\Sigma_i\right), \quad as \quad n \to \infty, \quad i = m, l, q,$$

where $\Sigma_i = \mathrm{diag}(v_1, \Sigma_{2i})$, $i = m, l, q$, v_1, Σ_{2m} *and* Σ_{2q} *are defined in (2.4.21) and Theorem 2.3, and* Σ_l *is defined in (2.4.20).*

Remark 2.3 *As we can see from Theorem 2.4, the covariance between the estimates of* ϕ *and* σ^2 *in all cases is*

$$\mathrm{Cov}(\,\hat{\phi}_i\,,\ \hat{\sigma}_i^2\,) = \frac{-2\sigma^2\,\phi}{\bar{t}\,(1-\phi^2)}\,\mathrm{Var}(\,\hat{\phi}_i\,), \qquad for \quad i = m,\ l,\ q. \tag{2.4.38}$$

2.5 A special case

Now, let us consider a special case by assuming that the data is balanced, that is $t_i \equiv t$ for all i. We drop the subscript i and write $\mathbf{V}, \mathbf{C}$ and $\mathbf{c}$. Since t_i's are all equal, $\mathbf{U}_i$'s have the same dimension. Let

$$\overline{\mathbf{U}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{U}_i.$$

We can check that

$$\begin{aligned} c_{00} &= \mathrm{tr}(\overline{\mathbf{U}}), \\ c_{10} &= \mathrm{tr}(\mathbf{C}_{10}\overline{\mathbf{U}}), \\ c_{11} &= \mathrm{tr}(\mathbf{C}_{11}\overline{\mathbf{U}}). \end{aligned}$$

The large sample property of the maximum likelihood estimate $\hat{\phi}_l$, was originally established by Fujikoshi *et al.* (1990). The following theorem summaries the results regarding the asymptotical properties of the moment, ML and QLS estimates.

Theorem 2.5 *Consider the model (1.1.7). Assume that* $t_i \equiv t$ *for all* $i = 1, \ldots, n$, *and the errors* ε_i*'s are independent and normally distributed with mean zero and* *covariance* $\sigma^2\mathbf{V}$ *for any* i. *Let* $\boldsymbol{\xi} = (\boldsymbol{\beta}, \phi, \sigma^2)'$, *and* $\hat{\boldsymbol{\xi}}_i = (\hat{\boldsymbol{\beta}}_i, \hat{\phi}_i, \hat{\sigma}_i^2)'$, $i = m, l, q$, *be the moment, ML and QLS estimates of* $\boldsymbol{\xi}$ *respectively. All three estimates of* $(\boldsymbol{\beta}, \sigma^2)$ *have the same expression*

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}(\hat{\phi})\mathbf{X}_i\right)^{-1}\cdot\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}(\hat{\phi})\mathbf{y}_i,\\ \hat{\sigma}^2 &= (\hat{c}_{00} + \hat{c}_{11}\hat{\phi}^2 - 2\hat{c}_{10}\hat{\phi})/t,\end{aligned}$$

but

$$\begin{aligned}\hat{\phi}_m &= \frac{t\hat{c}_{10}}{(t-1)\hat{c}_{00}},\\ \hat{\phi}_l &= = \frac{1}{3(t-1)\hat{c}_{11}}\left[(t-2)\hat{c}_{10} - 2\hat{a}\cos(\hat{\alpha})\right],\\ \hat{\phi}_q &= \frac{t-2}{t-1}\,\hat{\phi}_u = \frac{(t-2)\hat{c}_{10}}{(t-1)\hat{c}_{11}},\end{aligned}$$

where a *and* α *are as defined in (2.2.17). Also,*

$$\hat{\boldsymbol{\xi}}_i \xrightarrow{d} N\left(\boldsymbol{\xi},\ \frac{1}{n}\Sigma_i\right), \quad as \quad n\to\infty, \quad i = m, l, q,$$

where $\Sigma_i = \mathrm{diag}(v_1, \Sigma_{2i})$ *with* $\Sigma_{2i} = \begin{bmatrix} v_{2i} & v_{23i} \\ v_{23i} & v_{3i}\end{bmatrix}$, $i = m,\ l,\ q$, *and*

$$\begin{aligned}v_1 &= \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\,\mathbf{V}(\phi)\,\mathbf{X}_i\right)^{-1},\\ v_{2m} &= \frac{1}{t^2(t-1)^2(1-\phi^2)^2}\Big[t^2(t-1) - \bar{t}(3t^2-5t+6)\phi^2\\ &\quad +(t-1)(3t^2-4t+4)\phi^4 - t(t-1)(t-2)\phi^6 + 4\phi^{2t}\left(t-(t-1)\phi^2\right)^2\Big],\\ v_{3m} &= \frac{4\sigma^4\phi^2}{t^2(1-\phi^2)^2}\,v_{2m} + \frac{2\sigma^4}{t}\\ v_{2l} &= \frac{t}{t-1}\cdot\frac{(1-\phi^2)^2}{2\phi^2+t(1-\phi^2)}\\ v_{3l} &= \frac{2\sigma^4\left[2\phi^2+(t-1)(1-\phi^2)\right]}{(t-1)\left[2\phi^2+t(1-\phi^2)\right]}\\ v_{2q} &= \frac{1}{(t-1)^2(t-2)^2(1-\phi^2)^2}\Big[(t-1)(t-2)^2 - t(t-2)(3t-7)\phi^2\end{aligned}$$

$$
\begin{aligned}
& \quad +t(t-1)(3t-8)\phi^4 - t(t-1)(t-2)\phi^6 + 4\phi^{2t}\Big] \\
v_{3q} &= \frac{4\sigma^4\phi^2}{t^2(1-\phi^2)^2}\, v_{2q} + \frac{2\sigma^4}{t}\left[1 - \frac{1}{(1-\phi^2)^4}\right] \\
v_{23i} &= -\frac{2\sigma^2\phi}{t(1-\phi^2)}\, v_{2i}, \quad i = m, l, q.
\end{aligned}
$$

Proof: The proof follows from Theorem 2.4, setting $t_i = t$. The asymptotic distribution of $\hat{\boldsymbol{\xi}}_l$ is found by (2.4.20) with the replacement of $\mathbf{V}_i$ by $\mathbf{V}$, $\bar{t}$ by t and the asymptotic distribution of $(\hat{\phi}_m, \hat{\sigma}_m^2)$ and $(\hat{\phi}_q, \hat{\sigma}_q^2)$ are obtained from (2.4.27) and (2.4.28), respectively, replacing $\bar{t}$ by t. Since $\boldsymbol{\varepsilon}_i$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\sigma^2\, \mathbf{V}$, we have

$$
\mathrm{vec}(\hat{\bar{\mathbf{U}}}) \xrightarrow{d} N\left(\sigma^2 \mathrm{vec}(\mathbf{V}),\ \ \frac{2\sigma^4}{n}\, \mathbf{V} \otimes \mathbf{V}\right), \quad as \quad n \to \infty, \tag{2.5.39}
$$

where the operator vec$(\cdot)$ creates a column vector by stacking the columns of a matrix below one another. Define $\hat{\mathbf{c}} = (\hat{c}_{00},\ \hat{c}_{10},\ \hat{c}_{11})'$. From (2.5.39) and delta theorem we get

$$
\hat{\mathbf{c}} \xrightarrow{d} N\left(\boldsymbol{\mu}_c\,,\ \ \frac{1}{n}\,\Sigma_c\right), \quad as \quad n \to \infty, \tag{2.5.40}
$$

where '$\hat{\ }$' means evaluating at $\hat{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_c = \frac{\sigma^2}{1-\phi^2}(t,\ (t-1)\phi,\ t-2)'$ and $\Sigma_c = \frac{2\sigma^4}{(1-\phi^2)^4}\{v_{kl}\}$ is a symmetric matrix such that v_{kl}'s are the identities without the subscript i as defined in the proof of Theorem 2.2 with the replacement of t_i by t. The proof of moment and quasi-least squares cases just follow the proof in Theorem 2.3. $\triangleleft$

In the next section, we will show that the QLS estimates are more efficient than the moment estimates, and good competitors to the ML estimates. Moreover, the QLS estimates are more efficient than the ML estimates when there is a slight departure from normality of the data.

Another form of estimates

Depending on how you write the covariance matrix $\mathbf{\Gamma}$, the three estimating methods can be derived in two forms by choosing different parameters to estimate. One way is to write $\mathbf{\Gamma} = \sigma^2\,\mathbf{V}(\phi)$, which is the way we derived above, and hence obtain the estimates of $(\boldsymbol{\beta}, \phi, \sigma^2)$; another way is to write $\mathbf{\Gamma} = \sigma_\varepsilon^2\,\mathbf{P}(\phi)$ and hence obtain the estimates of $(\boldsymbol{\beta}, \phi, \sigma_\varepsilon^2)$. We will discuss the second form only for AR(1) case. The estimates of $\boldsymbol{\beta}$ have the same efficiency no matter which form and method you choose; while the estimates of σ^2 and σ_ε^2 are different in two forms, but have the same expression for all three methods in either form. Thus, we will concentrate on comparing the efficiencies of the estimates of ϕ. The ML estimates of ϕ are the same no matter which form you choose because of the invariance property of the MLE. The moment estimate of ϕ is less efficient than the QLS estimate. Thus we only need to compare the QLS estimates in the two forms and the ML estimate. According to Chaganty (2003), the QLS estimate of ϕ in the second form may be written as

$$\tilde{\phi}_q = \frac{2\,\tilde{\phi}_u}{1+\tilde{\phi}_u^2} = \frac{2\hat{c}_{10}}{\hat{c}_{00}+\hat{c}_{11}}, \tag{2.5.41}$$

where

$$\tilde{\phi}_u = \frac{(\hat{c}_{00}+\hat{c}_{11}) - \sqrt{(\hat{c}_{00}+\hat{c}_{11})^2 - 4\hat{c}_{10}^2}}{2\hat{c}_{10}}.$$

Note that $\tilde{\phi}_q$ seems to be a combination of $\hat{\phi}_m$ and $\hat{\phi}_u$ in the first form, if we approximate $\hat{\phi}_m$ by $\hat{c}_{10}/\hat{c}_{00}$. It was also shown that

$$\tilde{\phi}_q \xrightarrow{d} N(\phi,\ \frac{1}{n}\,\tilde{v}_{2q}), \quad as \quad n \to \infty,$$

where

$$\tilde{v}_{2q} = \frac{1}{(t-1)^2}\,[t(1-\phi^2) - (1-\phi^{2t})]. \tag{2.5.42}$$

One thing to note that is the first form estimate $\hat{\phi}_q$ may not be feasible for some data. This means the estimate of ϕ may lie outside the boundary. For instance, assuming that the error $\{\varepsilon_j\}$ has a normal distribution, if we choose $n = 1$ and $\varepsilon_1 = -1.3, \varepsilon_2 = 0.1, \varepsilon_3 = 0.2$ and $\varepsilon_4 = 1$, then $\hat{\phi}_q = 1.2 > 1$. While one may show that the second form estimate $\tilde{\phi}_q$ will always lie inside the boundary (Chaganty 1997, p.51, Appendix A). But these are extreme cases occurring with small probability.

2.6 Asymptotic relative efficiency

We can compare the moment, ML and QLS estimates by finding the asymptotic relative efficiencies (AREs). Assuming that the error $\{\varepsilon_j\}$ are normally distributed, we can obtain the AREs by taking the ratio of the asymptotic variances since all estimates are consistent. As we can see, the AREs of the estimates of ϕ are symmetric about $\phi = 0$ and always equal to 1 when $\phi = 0$. Theorem 2.4 shows that three estimates of β have the same efficiency asymptotically, and the asymptotic properties of the estimates of σ^2 are similar to the asymptotic properties of the estimates of ϕ since the estimates of σ^2 have the same expression. Thus, we will concentrate on comparing the estimates of ϕ. The AREs of the moment and QLS estimates of ϕ with respect to the ML estimate are defined as $e(\hat{\phi}_m;\ \hat{\phi}_l) = v_{2l}/v_{2m}$ and $e(\hat{\phi}_q;\ \hat{\phi}_l) = v_{2l}/v_{2q}$, where v_{2m}, v_{2l}, and v_{2q} are defined in (2.4.27), (2.4.22) and (2.4.28), respectively. For convenience, we assume $t_i \equiv t$. Table 2.1 contains the AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ vs $\hat{\phi}_l$ when $t = 5,\ 10$ and 30. The numbers in the parentheses are the AREs of $\hat{\phi}_q$ vs $\hat{\phi}_l$.

Figures 2.1 and 2.2 show the AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ vs $\hat{\phi}_l$ when $t = 10$ and 30, respectively. The 3D plot of the ARE of $\hat{\phi}_q$ vs $\hat{\phi}_l$ is shown in Figure 2.3, where

Table 2.1. AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ (in parentheses) vs $\hat{\phi}_l$ when the data is normal.

ϕ	t=5	t=10	t=30
0.1	0.9940 (0.9993)	0.9964 (0.9999)	0.9987 (1.0000)
0.2	0.9759 (0.9971)	0.9854 (0.9998)	0.9947 (1.0000)
0.3	0.9455 (0.9926)	0.9662 (0.9994)	0.9874 (1.0000)
0.4	0.9028 (0.9845)	0.9372 (0.9986)	0.9761 (1.0000)
0.5	0.8484 (0.9699)	0.8961 (0.9970)	0.9589 (0.9999)
0.6	0.7841 (0.9431)	0.8391 (0.9935)	0.9327 (0.9999)
0.7	0.7125 (0.8918)	0.7617 (0.9846)	0.8907 (0.9994)
0.8	0.6350 (0.7885)	0.6606 (0.9573)	0.8169 (0.9980)
0.9	0.5356 (0.5626)	0.5351 (0.8440)	0.6666 (0.9866)
0.95	0.4357 (0.3494)	0.4464 (0.6453)	0.5342 (0.9361)
0.98	0.2856 (0.1619)	0.3255 (0.3623)	0.4180 (0.7428)

the x-axis is the value of ϕ ranges from -0.96 to 0.96, the y-axis is the value of t ranges from 5 to 45, and the z-axis is the ARE. Also define the ARE of $\hat{\phi}_q$ vs $\hat{\phi}_m$ as $e(\hat{\phi}_q;\ \hat{\phi}_m) = v_{2m}/v_{2q}$ and the 3D plot is shown in Figure 2.4. It is clear from the plots that $\hat{\phi}_q$ is better than $\hat{\phi}_m$ and is as good as $\hat{\phi}_l$ for sufficiently large value of t over the entire range of the parameter ϕ. When $|\phi|$ approaches the boundary 1, $\{\varepsilon_j\}$ is highly correlated and the process turns to be non-stationary, the estimates including the MLEs are not reliable in this case. We may apply a non-stationary model instead, which is beyond the discussion of this thesis. One can take some degree of differences of the data so that the new process is stationary and then obtain the parameter estimates as usual. New, let's define the AREs of $\hat{\sigma}^2_m$ and $\hat{\sigma}^2_q$ with respect to $\hat{\sigma}^2_l$ as $e(\hat{\sigma}^2_m;\ \hat{\sigma}^2_l) = v_{3l}/v_{3m}$ and $e(\hat{\sigma}^2_q;\ \hat{\sigma}^2_l) = v_{3l}/v_{3q}$, where v_{3m}, v_{3l} and v_{3q} are defined in (2.4.36), (2.4.23) and (2.4.37), respectively. Figure 2.5 shows the AREs of $\hat{\sigma}^2_m$ and $\hat{\sigma}^2_q$ vs $\hat{\sigma}^2_l$ when $t = 10$. The 3D plot of the ARE of $\hat{\sigma}^2_q$ vs $\hat{\sigma}^2_l$ as a function of ϕ and t is shown in Figure 2.6, which is similar to Figure 2.3. Define the ARE of $\hat{\sigma}^2_q$ with respect to $\hat{\sigma}^2_m$ as $e(\hat{\sigma}^2_q;\ \hat{\sigma}^2_m) = v_{3m}/v_{3q}$, the 3D plot is shown in Figure 2.7, which is similar to Figure 2.4. Hence, the QLS estimate of σ^2 is also

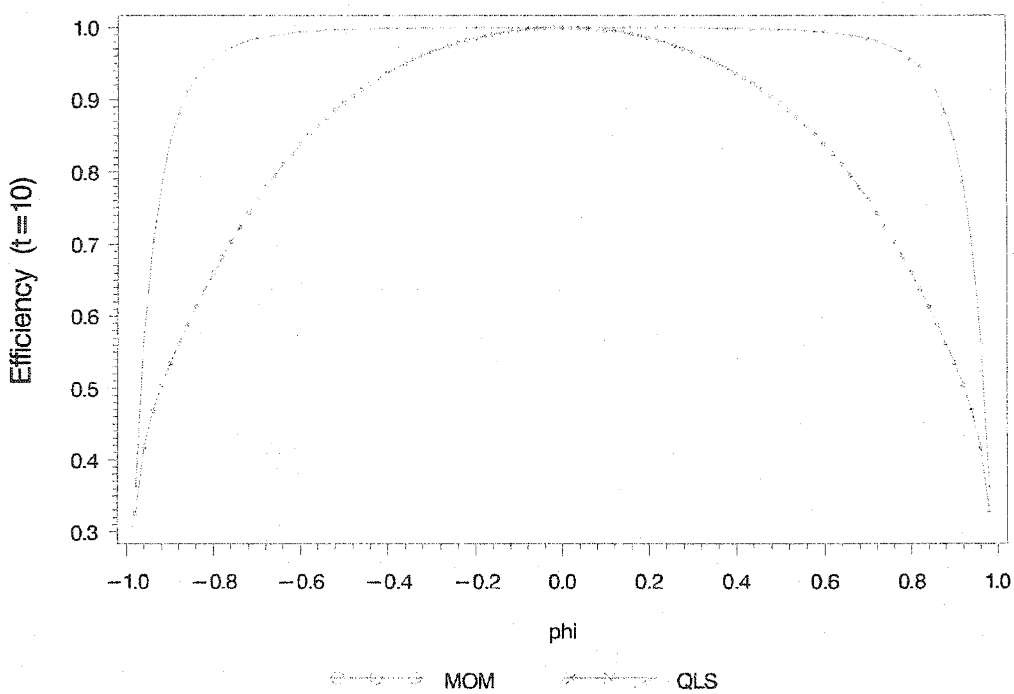

Figure 2.1. AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ vs $\hat{\phi}_l$ when the data is normal and $t = 10$.

better than the moment estimate and a good competitor to the ML estimate.

Now, let us compare the QLS estimates in the two forms. Define the ARE of $\hat{\phi}_q$ with respect to $\tilde{\phi}_q$ as $e(\hat{\phi}_q;\ \tilde{\phi}_q) = \tilde{v}_{2q}/v_{2q}$, where v_{2q} and $\tilde{v}_{2q}$ are given by (2.4.28) and (2.5.42) respectively. Figure 2.8 shows the plot of $e(\hat{\phi}_q;\ \tilde{\phi}_q)$ as a function of ϕ when $t = 10$. The 3D plot Figure 2.9 shows $e(\hat{\phi}_q;\ \tilde{\phi}_q)$ as a function of ϕ and t. It is clear from the plots that $\hat{\phi}_q$ is better than $\tilde{\phi}_q$ for almost the entire feasible region of ϕ except for the boundary. Further, let us define the ARE of $\tilde{\phi}_q$ with respect to $\hat{\phi}_l$ as $e(\tilde{\phi}_q; \hat{\phi}_l) = v_{2l}/\tilde{v}_{2q}$. Figure 2.10 shows the AREs of $\hat{\phi}_m$, $\hat{\phi}_q$ and $\tilde{\phi}_q$ with respect to $\hat{\phi}_l$ as a function of ϕ when $t = 10$. Based on the plot, the order of the goodness of the estimates is clear. Result of the ARE of $\hat{\sigma}_q^2$ with respect to $\tilde{\sigma}_q^2$ is similar.

Suppose now the data is from a distribution that is slightly different from the normal. We will compare the estimates through simulation. Without loss of

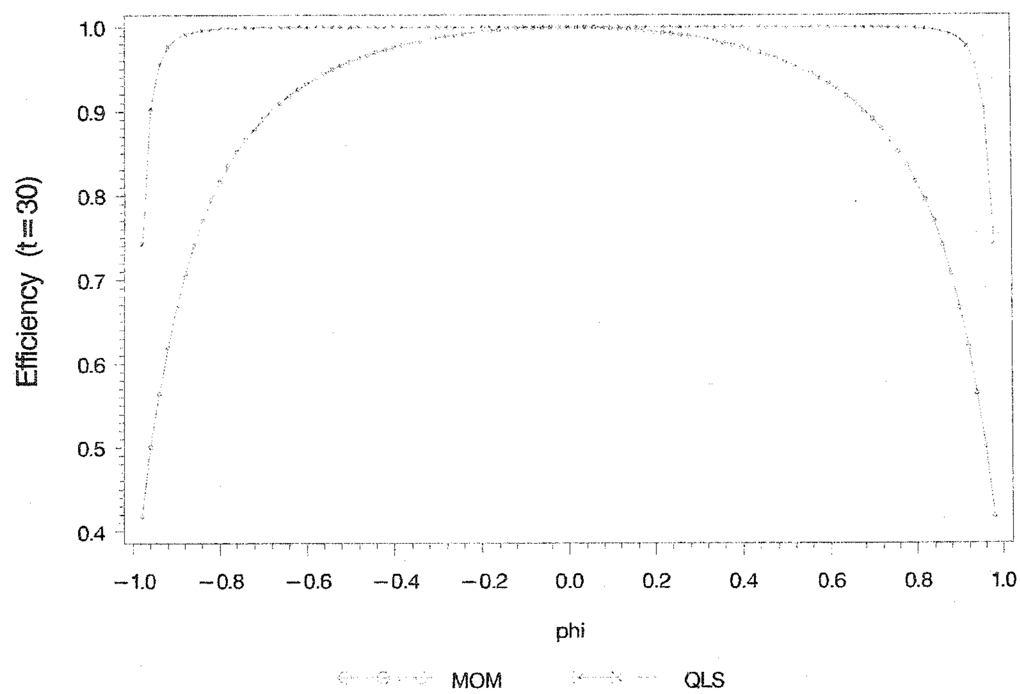

Figure 2.2. AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ vs $\hat{\phi}_l$ when the data is normal and $t = 30$.

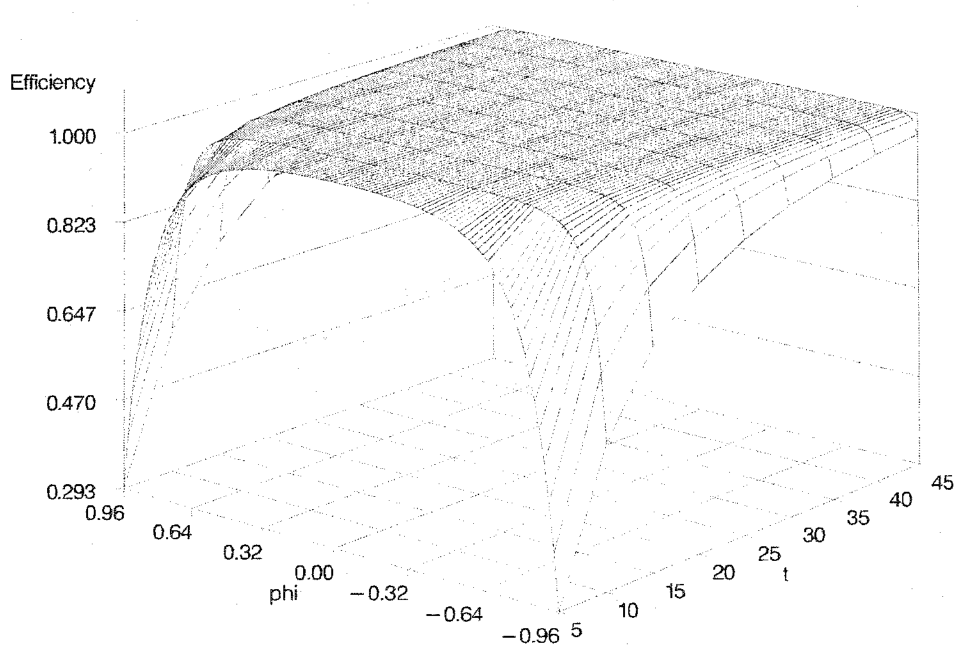

Figure 2.3. ARE of $\hat{\phi}_q$ vs $\hat{\phi}_l$ when the data is normal.

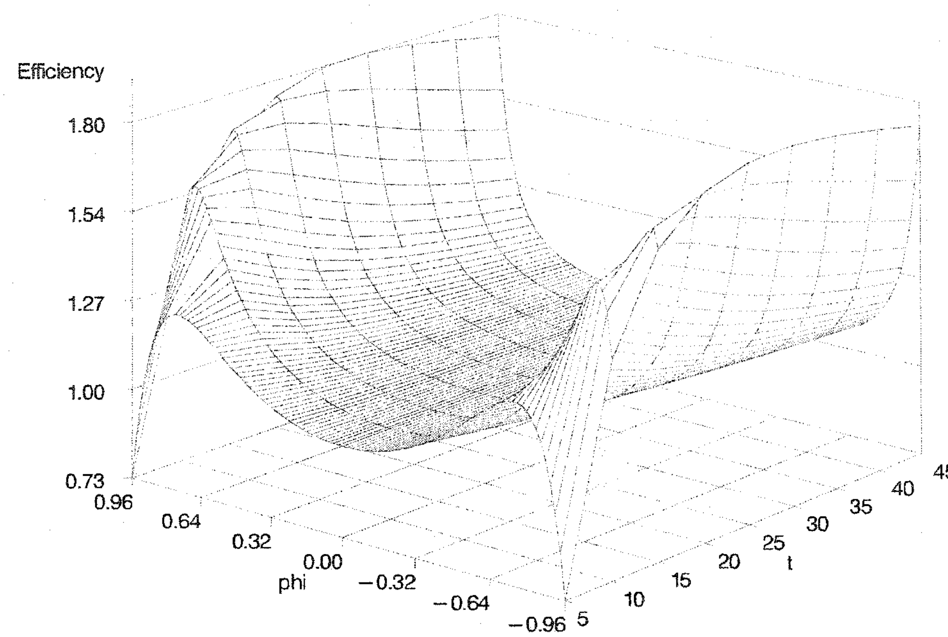

Figure 2.4. ARE of $\hat{\phi}_q$ vs $\hat{\phi}_m$ when the data is normal.

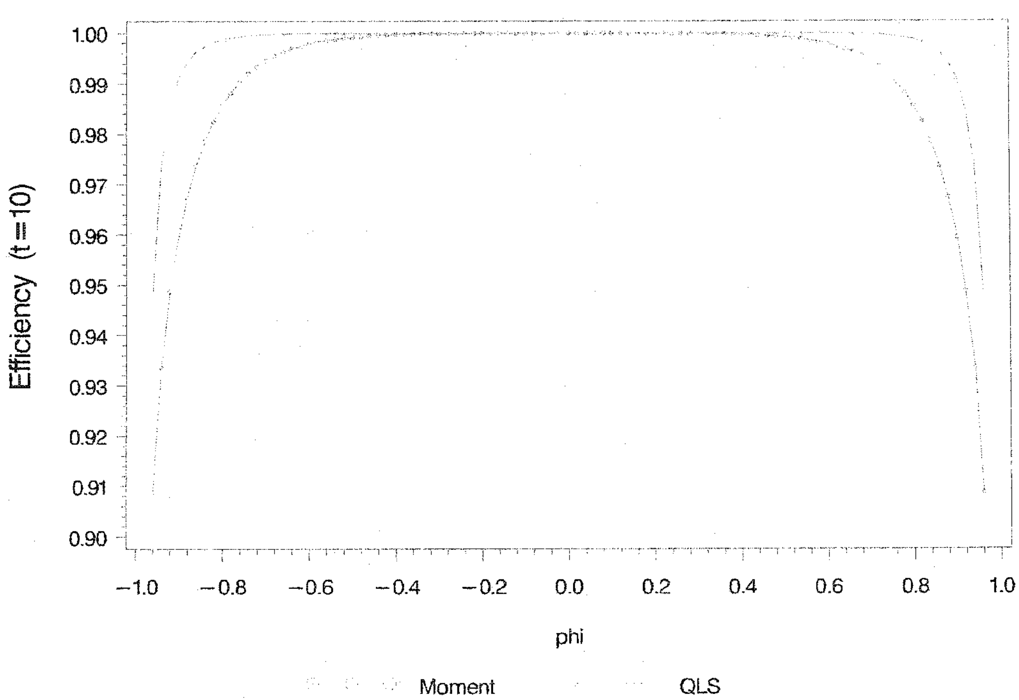

Figure 2.5. AREs of $\hat{\sigma}_m^2$ and $\hat{\sigma}_q^2$ vs $\hat{\sigma}_l^2$ when the data is normal and $t = 10$.

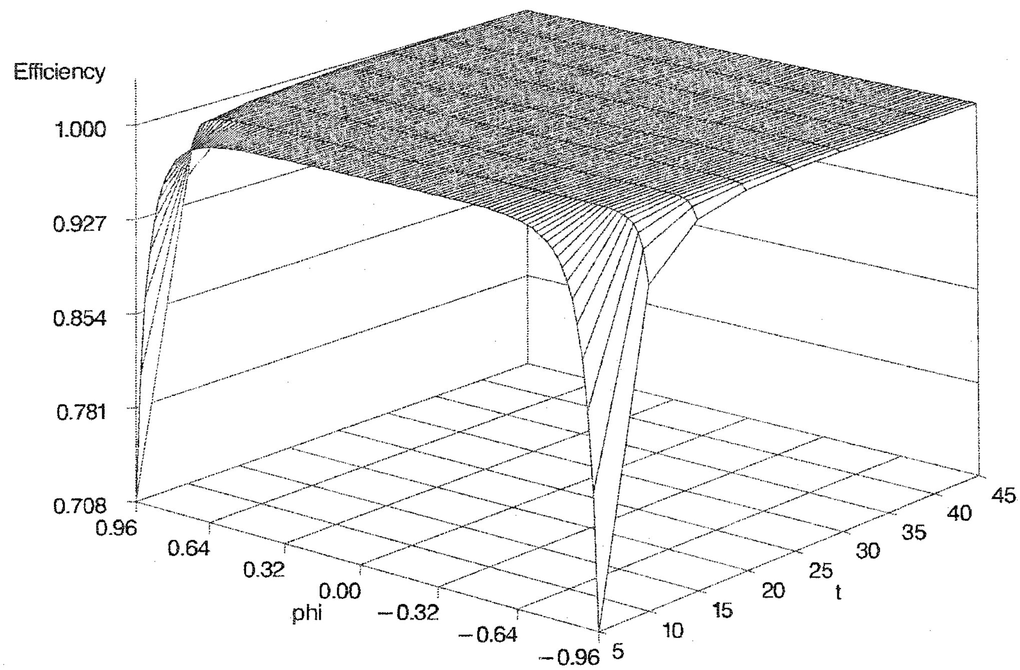

Figure 2.6. ARE of $\hat{\sigma}_q^2$ vs $\hat{\sigma}_l^2$ when the data is normal.

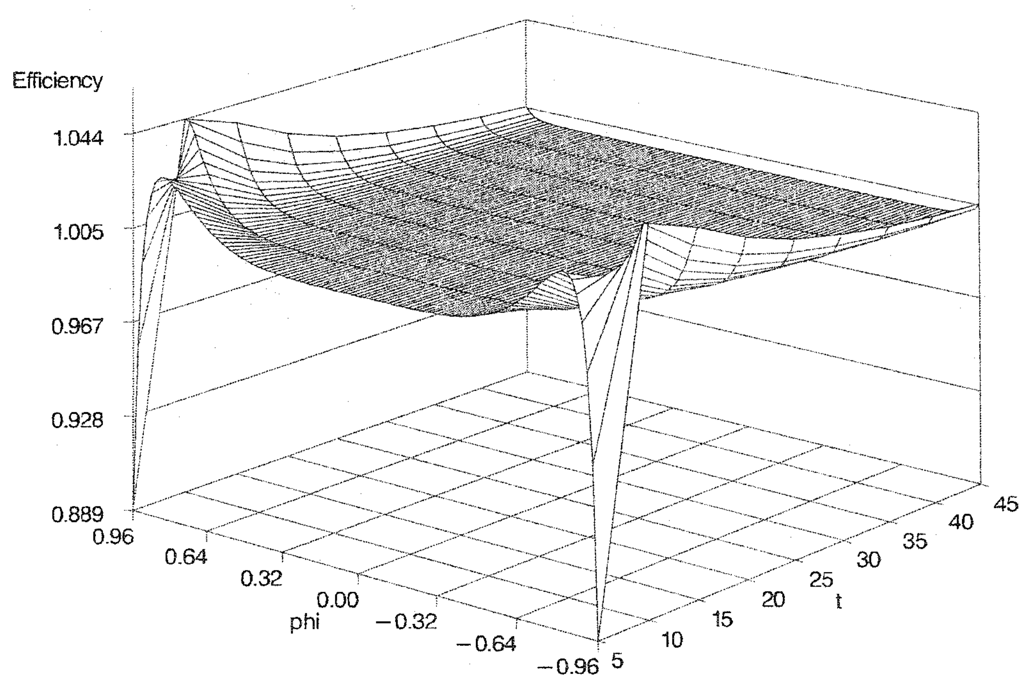

Figure 2.7. ARE of $\hat{\sigma}_q^2$ vs $\hat{\sigma}_m^2$ when the data is normal.

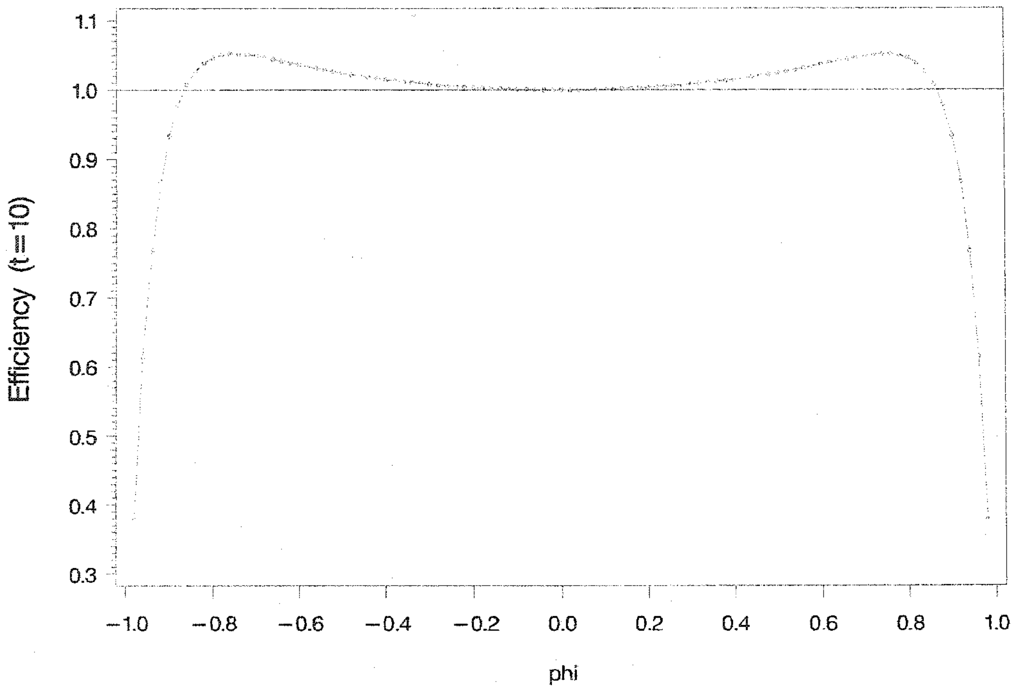

Figure 2.8. ARE of $\hat{\phi}_q$ vs $\tilde{\phi}_q$ when the data is normal and $t = 10$.

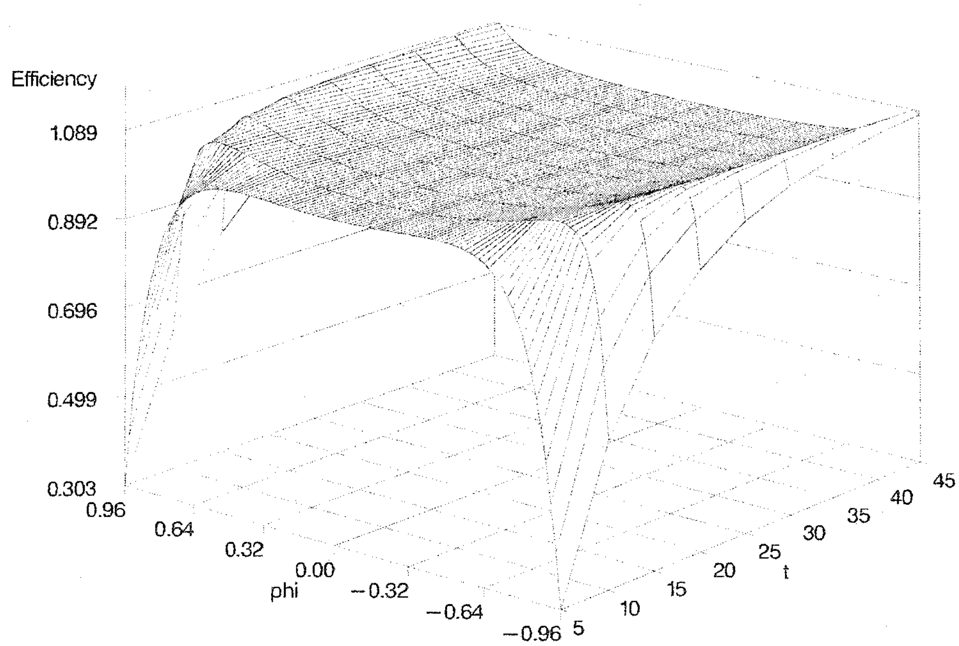

Figure 2.9. ARE of $\hat{\phi}_q$ vs $\tilde{\phi}_q$ when the data is normal.

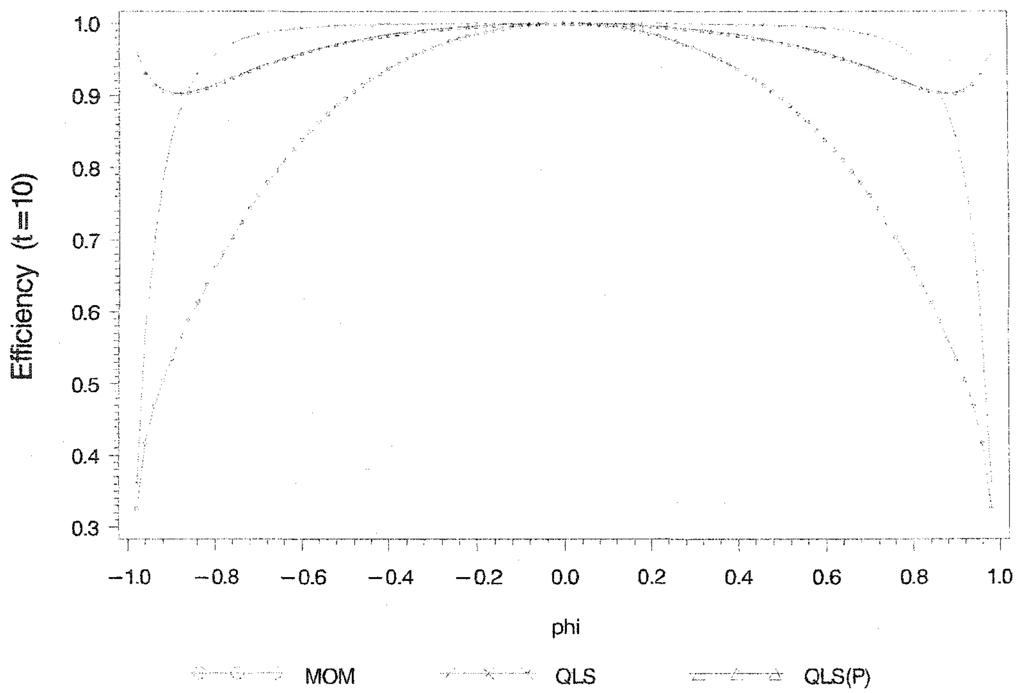

Figure 2.10. AREs of $\hat{\phi}_m$, $\hat{\phi}_q$ and $\tilde{\phi}_q$ when the data is normal and $t = 10$.

generality, we fix $\sigma^2 = 1$ and assume $\boldsymbol{\beta} = \mathbf{0}$. For fixed t and ϕ, we generate a t-dimensional vector $\boldsymbol{\varepsilon}$ whose elements form a random sample from the Student-t distribution with mean 0 and 5 degrees of freedom, and then let $\mathbf{y} = \sigma\sqrt{3/5}\mathbf{V}^{1/2}(\phi)\boldsymbol{\varepsilon}$ and generate 10 ($n = 10$, small sample) of them. The $\mathbf{y}$ generated this way has mean $\mathbf{0}$ and covariance $\sigma^2\mathbf{V}(\phi)$. Now the process is repeated 10000 times. For each replication we compute the moment, normal ML (assuming the data is normal) and QLS estimates of (ϕ, σ^2), and then compute the mean square errors (MSEs). Suppose $\hat{\phi}_i$ is the estimate of ϕ for the i^{th} replication, $i = 1, \ldots,\ 10000$, then $\text{MSE}(\hat{\phi}) = \sum_{i=1}^{10000}(\hat{\phi}_i - \phi)^2/10000$. The MSE calculated this way is an estimate of the asymptotic variance of the asymptotically unbiased estimator. The AREs are all defined with respect to the MLE this time, and are estimated by the ratio of the MSEs. Table 2.2 gives the efficiencies when the data is Student-t distribution for $t = 5$, 10 and 30. The numbers in the parentheses are the AREs of $\hat{\phi}_q$ and $\tilde{\phi}_q$.

Table 2.2. AREs of $\hat{\phi}_m$, $\hat{\phi}_q$ and $\tilde{\phi}_q$ (in parentheses) when the data is simulated from Student-t distribution.

ϕ	t=5	t=10	t=30
0.1	1.0257 (0.9186, 1.0372)	1.0066 (0.9912, 1.0128)	1.0020 (0.9994, 1.0036)
0.2	1.0237 (0.8961, 1.0405)	1.0057 (0.9811, 1.0196)	1.0040 (0.9986, 1.0077)
0.3	1.0011 (0.8646, 1.0387)	0.9885 (0.9800, 1.0190)	0.9885 (0.9987, 1.0035)
0.4	0.9574 (0.8519, 1.0226)	0.9787 (0.9552, 1.0269)	0.9917 (0.9967, 1.0116)
0.5	0.9218 (0.8239, 1.0150)	0.9426 (0.9489, 1.0208)	0.9852 (0.9934, 1.0195)
0.6	0.8510 (0.8129, 0.9840)	0.9188 (0.9288, 1.0237)	0.9615 (0.9901, 1.0178)
0.7	0.8072 (0.8058, 0.9668)	0.8484 (0.9032, 1.0033)	0.9455 (0.9761, 1.0297)
0.8	0.7482 (0.7980, 0.9408)	0.7734 (0.8782, 0.9730)	0.9098 (0.9622, 1.0314)
0.9	0.6795 (0.7514, 0.9259)	0.6590 (0.8716, 0.9242)	0.7970 (0.9306, 1.0039)
0.94	0.6338 (0.6841, 0.9248)	0.5905 (0.8415, 0.8977)	0.7191 (0.9018, 0.9780)
0.98	0.5365 (0.4628, 0.9535)	0.5085 (0.6939, 0.9196)	0.5779 (0.8596, 0.9261)

Figure 2.11 shows the plot of the AREs of $\hat{\phi}_m$, $\hat{\phi}_q$ and $\tilde{\phi}_q$ when $t = 30$. Figure 2.12 gives the 3D plot of the ARE of $\tilde{\phi}_q$. We found out that behavior of the ARE of $\hat{\phi}_q$ is similar to the normal case, whereas the ARE of $\tilde{\phi}_q$ is greater than 1 for most values of ϕ. Both of them are better than the moment estimate. The same property also holds for the estimates of σ^2.

Let us consider another case where the data is from Beta distribution, which is similar to a bell shaped curve. As before, we fix $\sigma^2 = 1$ and assume $\boldsymbol{\beta} = \mathbf{0}$, simulate t of ε_j from Beta(1.01, 1.01) instead of Student-t distribution. The density plot of Beta(1.01, 1.01) is shown in Figure 2.13. Let $\varepsilon_j = (\varepsilon_j - 1.01 \cdot 1.01)/\sqrt{1.01 \cdot 1.01^2}$, suppose $\boldsymbol{\varepsilon}$ is the vector contains the ε_j's, and then let $\mathbf{y} = \sigma \mathbf{V}^{1/2}(\phi)\boldsymbol{\varepsilon}$. We choose $n = 30$ and the process is replicated 10000 times. For each replication, we compute the MSEs. This time, we only consider the moment, QLS first form and ML estimates. Define the AREs as in the Student-t case. Table 2.3 gives the efficiencies when the data is from Beta(1.01, 1.01) distribution for $t = 5$, 10 and 30. The numbers in the parentheses are the ARE of $\hat{\phi}_q$. Figure 2.14 shows the plot of the

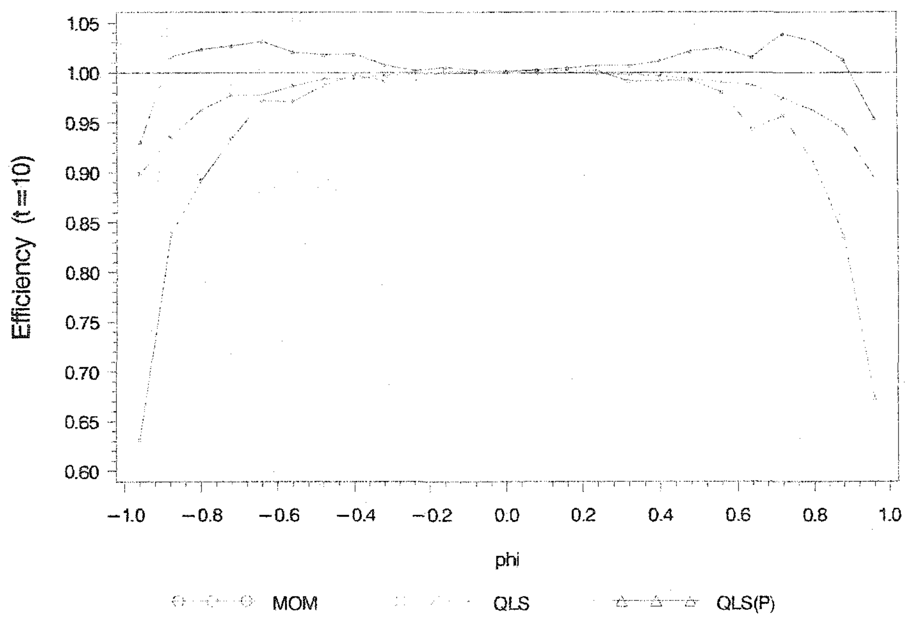

Figure 2.11. AREs of $\hat{\phi}_m$, $\hat{\phi}_q$ and $\tilde{\phi}_q$ when the data is simulated from Student-t distribution.

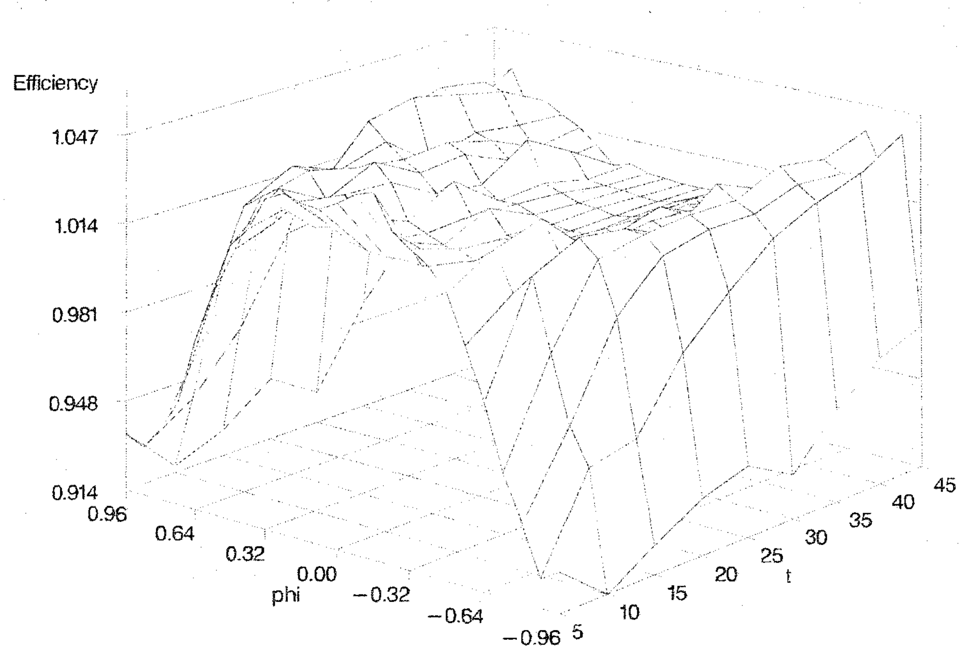

Figure 2.12. ARE of $\tilde{\phi}_q$ when the data is simulated from Student-t distribution.

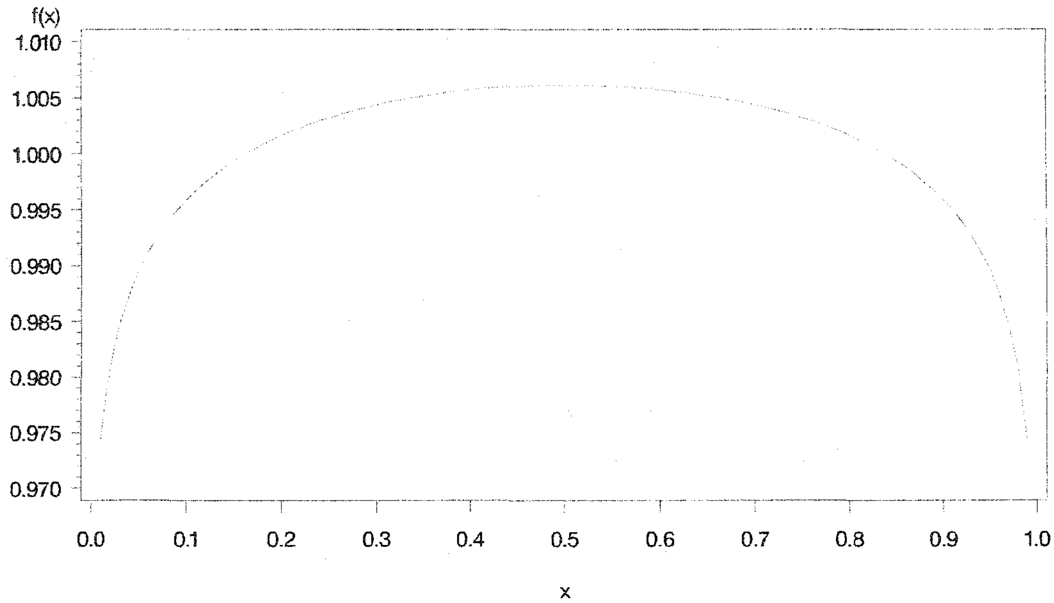

Figure 2.13. Density of Beta(1.01, 1.01) distribution.

AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ when $t = 30$. Figure 2.15 gives the 3D plot of the ARE of $\hat{\phi}_q$. We found out that the ARE of $\hat{\phi}_q$ is greater than 1 for most values of ϕ. That means the QLS estimates is better than the MLEs in this case. The same property also holds for the estimates of σ^2.

In practice, it is hard to decide whether the data is normally distributed. If the normality test fails, using ML estimates may not be the optimal method. Even if we know the exact marginal distribution of the data, it may be hard to write down the joint density of the data since the data is correlated within replication, e.g. Beta distribution. The QLS estimates are better than the moment estimates, and are good competitors to the MLEs and may be more robust than the MLEs. Thus, the QLS estimates are good candidates when the normality is not tenable.

Table 2.3. AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ (in parentheses) when the data is simulated from Beta(1.01, 1.01) distribution.

ϕ	t=5	t=10	t=30
0.1	0.9914 (0.9989)	0.9940 (1.0002)	1.0022 (0.9999)
0.2	0.9874 (0.9968)	0.9830 (1.0012)	0.9900 (1.0003)
0.3	0.9536 (1.0024)	0.9522 (1.0043)	0.9956 (0.9999)
0.4	0.9030 (1.0134)	0.9305 (1.0055)	0.9764 (1.0005)
0.5	0.8417 (1.0136)	0.8822 (1.0100)	0.9487 (1.0015)
0.6	0.7537 (1.0258)	0.8289 (1.0126)	0.9276 (1.0017)
0.7	0.6710 (1.0000)	0.6792 (1.0385)	0.8895 (1.0021)
0.8	0.5847 (0.8972)	0.5729 (1.0147)	0.7841 (1.0099)
0.9	0.4237 (0.5247)	0.4506 (0.9098)	0.6015 (1.0276)
0.96	0.2714 (0.2295)	0.2791 (0.5012)	0.4024 (0.9654)

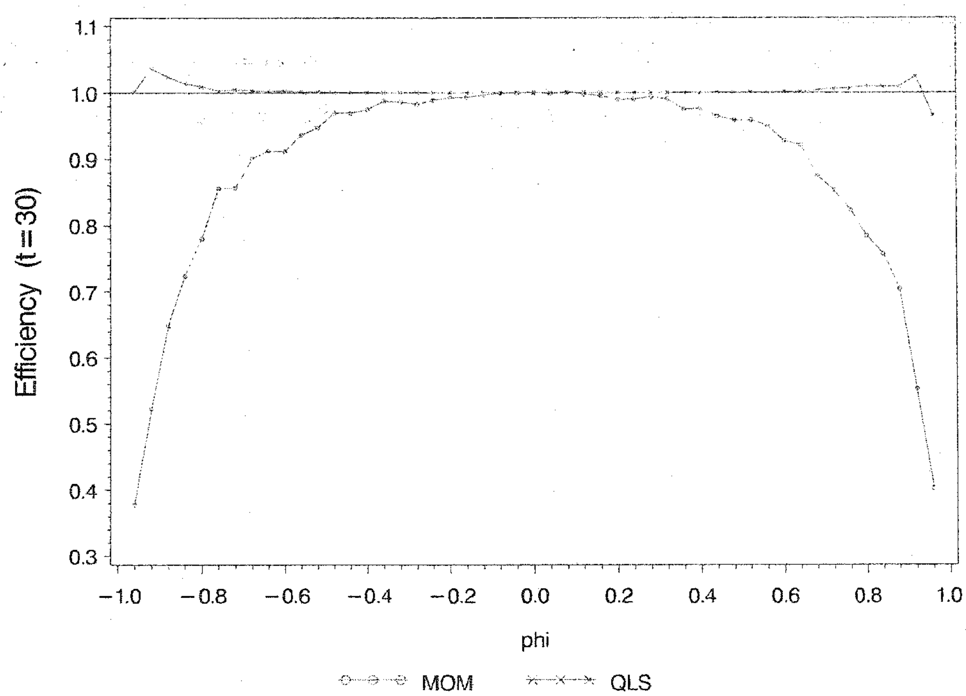

Figure 2.14. AREs of $\hat{\phi}_m$ and $\hat{\phi}_q$ when the data is simulated from Beta(1.01, 1.01) distribution and $t = 30$.

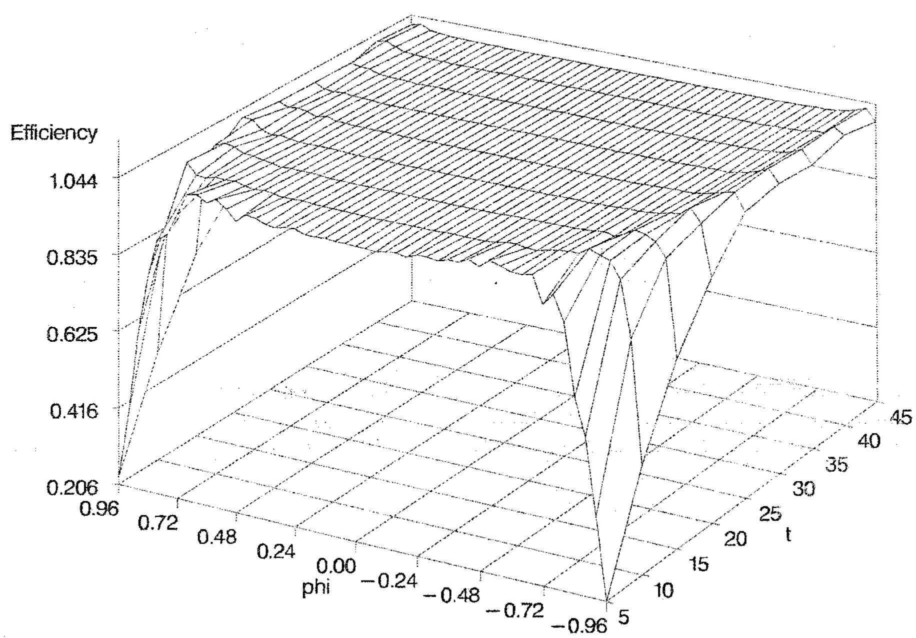

Figure 2.15. ARE of $\hat{\phi}_q$ when the data is simulated from Beta(1.01, 1.01) distribution.

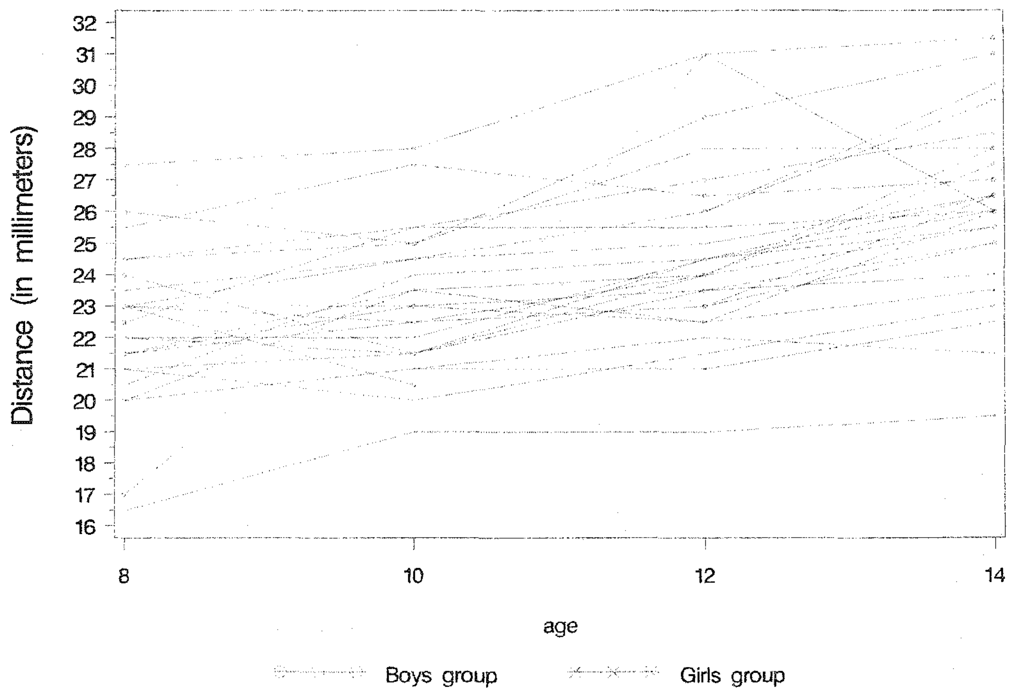

Figure 2.16. Measurements of distances on 11 girls and 16 boys, at 4 different ages.

2.7 Application to a dental study

We analyze the longitudinal data displayed in Table 1 of Potthoff and Roy (1964). The data were collected by investigators at the University of North Carolina Dental School in a dental study of 27 subjects (11 girls and 16 boys). Each measurement is the distance, in millimeters, from the center of each subjects pituitary to pteryomaxillary fissure recorded at 8, 10, 12, 14 years of age. The reason why there is an occasional instance where this distance decreases with age is that the distance represents the relative position of two points. We will assume that the (4×4) variance matrix of the 4 correlated observations is the same for all 27 individuals. The data is shown in Table 2.4 and the plot is given by Figure 2.16. We fit the following regression model:

Table 2.4. Measurements of distances (in millimeters) on 11 girls and 16 boys, at 4 different ages.

Girls group					Boys group				
	Age in years					Age in years			
Individual	8	10	12	14	Individual	8	10	12	14
1	21	20	21.5	23	1	26	25	29	31
2	21	21.5	24	25.5	2	21.5	22.5	23	26.5
3	20.5	24	24.5	26	3	23	22.5	24	27.5
4	23.5	24.5	25	26.5	4	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	5	20	23.5	22.5	26
6	20	21	21	22.5	6	24.5	25.5	27	28.5
7	21.5	22.5	23	25	7	22	22	24.5	26.5
8	23	23	23.5	24	8	24	21.5	24.5	25.5
9	20	21	22	21.5	9	23	20.5	31	26
10	16.5	19	19	19.5	10	27.5	28	31	31.5
11	24.5	25	28	28	11	23	23	23.5	25
					12	21.5	23.5	24	28
					13	17	24.5	26	29.5
					14	22.5	25.5	25.5	26
					15	23	24.5	26	30
					16	22	21.5	23.5	25
Mean	21.18	22.23	23.09	24.09	Mean	22.87	23.81	25.72	27.47

Table 2.5. Regression analysis of a dental study data using MOM, QLS and ML methods with AR(1) correlation structure.

Parameter	MOME	QLSE	MLE
β_g	17.3213 (1.6056)	17.3220 (1.6029)	17.3217 (1.6040)
β_b	16.5946 (1.3313)	16.5902 (1.3291)	16.5920 (1.3230)
γ_g	0.4838 (0.1384)	0.4837 (0.1383)	0.4837 (0.1384)
γ_b	0.7695 (0.1147)	0.7697 (0.1147)	0.7696 (0.1147)
ϕ	0.6135 (0.453)	0.6028 (0.432)	0.6071 (0.404)
σ^2	3.0787 (2.283)	3.0946 (2.278)	3.0881 (2.264)

$$\mu_{ij} = \beta_g x_{i1} + \beta_b x_{i2} + \gamma_g x_{i1} \cdot x_{j3} + \gamma_b x_{i2} \cdot x_{j3}, \quad 1 \leq j \leq 4, \;\; 1 \leq i \leq 27,$$

where x_{i1}, x_{i2} are indicators for the girl and boy, respectively, and x_{j3} is the subject's age at the jth measurement time. The model has also been studied by Jennrich and Schluchter (1986). Assuming the error is an AR(1) process, we apply the MOM and QLS methods and also apply the ML method assuming the data is normal. Table 2.5 contains estimates for the regression parameters, the correlation parameter and the scale parameter, computed using the MOM, QLS and ML methods. The numbers in the parenthesis are the corresponding standard errors. If we make the normality assumption, each measurement vector taken at four ages will have the same variance-covariance matrix but different mean for girls and boys group. The p-values of the tests based on the skewness for this two groups, which should be close to 1 if it is normal, turn out to be 0.9725 and 0.0025, respectively. Thus, the normality assumption of the girls group is satisfied, but not for the boys group. Hence the QLS estimates may be more appropriate. The residual plot based on the QLS estimates are given by Figure 2.17. We see that the residuals are almost randomly distributed around 0 at each age, regardless the sex group.

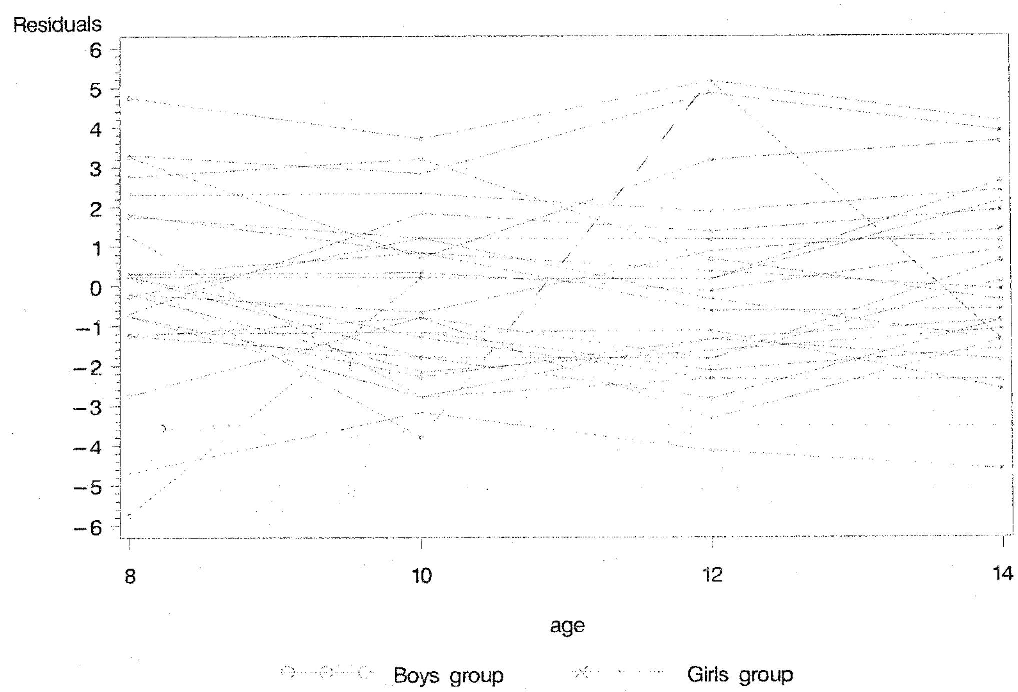

Figure 2.17. Residual plot based on the Quasi-least squares estimates, assuming the error is an AR(1) process.

2.8 Summary

In this chapter, we studied the moment, maximum likelihood, and two forms of the quasi-least squares estimates when the error is an AR(1) process. We focused on comparing the estimates of ϕ, since the efficiencies of the estimates of $\boldsymbol{\beta}$ and σ^2 will depend on the efficiency of the estimate of ϕ. Theocratical and simulation results showed that the maximum likelihood estimates are the best and the moment estimates are the worst for normal data. However, when the data departs from normal, the quasi-least squares methods may be better, especially the second form. A dental study was presented to illustrate the three estimating methods.

CHAPTER III

APPLICATION OF THE ESTIMATING METHODS TO THE MODEL WITH AR(2) ERRORS

In this chapter, we study the time series regression model with errors that follow an autoregressive process of order two (AR(2)). This is an important special case of the model with AR(p) errors, which we study in the next chapter. The organization of this chapter is as follows. In Section 3.1, we present the time series regression model with AR(2) errors. The moment and the maximum likelihood estimation methods will be discussed in Section 3.2, followed by the quasi-least squares estimating method in Section 3.3. The results of the previous sections are summarized for balanced data, that is, $t_i \equiv t$ in Section 3.4. In Section 3.5 we derive the asymptotic variance of the maximum likelihood estimates by finding the information matrix. In Section 3.6, simulation results are presented for comparisons of all the three methods. The robust properties of the quasi-least squares estimates are also studied in that section. Finally, an application to a dental study is presented in Section 3.7.

3.1 The model with AR(2) correlation structure

Now, let us assume that the error series $\{\varepsilon_j\}$ is a second order autoregressive process. We assume that the data consists of n independent replications and t_i observations in the i^{th} replication. The model is represented as (1.1.7), while (1.1.2) becomes

$$\varepsilon_j = \phi_1 \varepsilon_{j-1} + \phi_2 \varepsilon_{j-2} + a_j. \tag{3.1.1}$$

The stationary condition requires that $|\phi_2| < 1$ and $|\phi_1| < 1 - \phi_2$. Using (1.1.5), we can see that the autocorrelation functions satisfy the second order difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 1 \tag{3.1.2}$$

with $\rho_0 = 1$. Putting $k = 1, 2$, we get the *Yule-Walker* equations

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \rho_2. \end{aligned}$$

Let $\boldsymbol{\phi} = (\phi_1, \phi_2)'$, $\boldsymbol{\rho} = (\rho_1, \rho_2)'$, then $\boldsymbol{\phi}$ and $\boldsymbol{\rho}$ have the following relationship

$$\begin{aligned} \phi_1 &= \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \qquad |\phi_1| < 1 - \phi_2, \\ \phi_2 &= \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}, \qquad |\phi_2| < 1, \end{aligned} \tag{3.1.3}$$

and

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1-\phi_2}, \qquad |\rho_1| < 1, \;\; |\rho_2| < 1 \\ \rho_2 &= \phi_2 + \frac{\phi_1^2}{1-\phi_2}, \qquad \rho_1^2 < \frac{1}{2}(\rho_2 + 1). \end{aligned} \tag{3.1.4}$$

The plots of the feasible regions of $(\phi_1, \ \phi_2)$ and $(\rho_1, \ \rho_2)$ are given by Figure 3.1 and Figure 3.2, respectively. Let

$$\begin{aligned} \Delta &= 1 - \phi_1 \rho_1 - \phi_2 \rho_2 \\ &= \frac{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}{1-\phi_2}. \end{aligned} \tag{3.1.5}$$

Equation (1.1.4) implies that the variance of $\{\varepsilon_j\}$ is

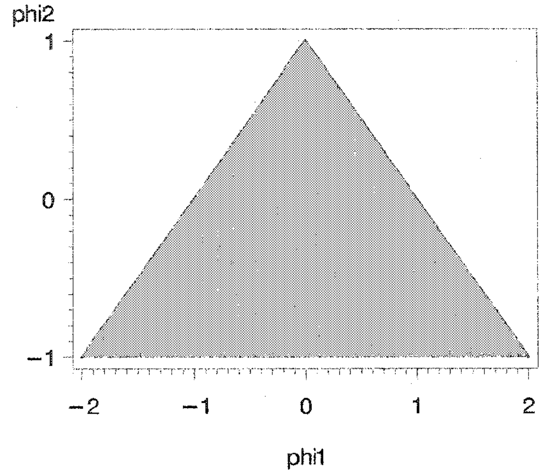

Figure 3.1. Feasible region of ϕ_1 *and* ϕ_2.

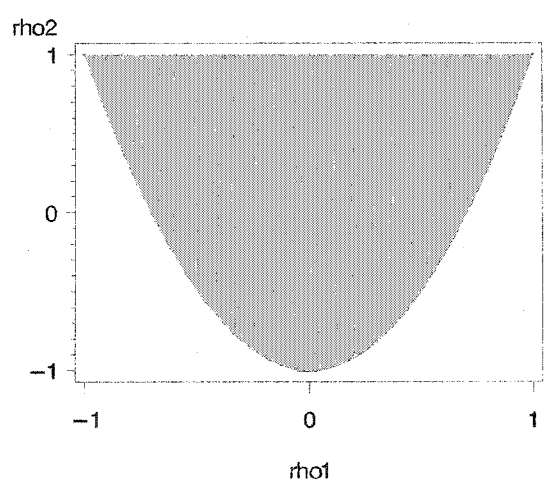

Figure 3.2. Feasible region of ρ_1 *and* ρ_2.

$$\sigma_\varepsilon^2 = \frac{1}{\Delta}\,\sigma^2. \tag{3.1.6}$$

According to (1.1.6), the covariance matrix of $\{\varepsilon_j\}$ made at t_i successive times is $\sigma^2\mathbf{V}_i(\boldsymbol{\phi})$ and

$$\mathbf{V}_i(\boldsymbol{\phi}) = \frac{1}{\Delta}\,\mathbf{P}_i(\boldsymbol{\phi}), \tag{3.1.7}$$

where $\mathbf{P}_i(\boldsymbol{\phi})$ is the autocorrelation matrix. Cheang and Reinsel (2000, p1176) showed that the inverse of $\mathbf{V}_i(\boldsymbol{\phi})$ is

$$\mathbf{V}_i^{-1}(\boldsymbol{\phi}) = \mathbf{C}_{i00} + \phi_1^2\mathbf{C}_{i11} + \phi_2^2\mathbf{C}_{i22} + 2\phi_1\phi_2\mathbf{C}_{i12} - 2\phi_1\mathbf{C}_{i10} - 2\phi_2\mathbf{C}_{i20}, \tag{3.1.8}$$

where $\mathbf{C}_{i00}$ is the $t_i \times t_i$ identity matrix, $\mathbf{C}_{i11}$ is the identity matrix with the first and last elements 0, and $\mathbf{C}_i22$ is the identity matrix with the first two and last two diagonal elements 0. The $t_i \times t_i$ matrix $2\mathbf{C}_{i0}$ has 1's on the first off diagonals and 0's elsewhere, $\mathbf{C}_{i12}$ is the same as $\mathbf{C}_{i10}$ except that the first and last nonzero elements are 0. The $t_i \times t_i$ matrix $2\mathbf{C}_{i20}$ has 1's on the second off diagonals and 0's elsewhere. Notice that $\mathbf{C}_{i00}$, $\mathbf{C}_{i10}$ and $\mathbf{C}_{i11}$ are already defined in (2.1.6). The Cholesky decomposition of $\mathbf{V}_i^{-1}(\boldsymbol{\phi})$ is $\mathbf{L}_i$ (see Cheang and Reinsel (2000, p1174), that is

$$\mathbf{V}_i^{-1}(\boldsymbol{\phi}) = \mathbf{L}_i'\mathbf{L}_i$$

and $\mathbf{L}_i$ is a lower triangular matrix with its first two main diagonal elements equal to $\sqrt{\Delta}$ and $\sqrt{1-\phi_2^2}$, respectively, its remaining main diagonal elements equal to 1, elements in the first off diagonal equal to $-\phi_1$ except that $\mathbf{L}_i[2,1] = -\phi_1\sqrt{1-\phi_2^2}/(1-\phi_2)$, elements in the second off diagonal equal to $-\phi_2$, and 0 elsewhere. The adjusted sum of square errors $S(\boldsymbol{\beta},\boldsymbol{\phi})$ defined in (1.2.8) simplifies to

$$\begin{aligned} S(\boldsymbol{\beta},\boldsymbol{\phi}) &= \sum_{i=1}^{n} \mathrm{tr}\left[\mathbf{V}_i^{-1}(\boldsymbol{\phi})\mathbf{U}_i(\boldsymbol{\beta})\right] \\ &= n\left(c_{00} + \phi_1^2 c_{11} + \phi_2^2 c_{22} + 2\phi_1\phi_2 c_{12} - 2\phi_1 c_{10} - 2\phi_2 c_{20}\right), \end{aligned} \tag{3.1.9}$$

where c_{00}, c_{10} and c_{11} are defined in (2.1.7), the remaining is defined as

$$\begin{aligned}
c_{20} &= \tfrac{1}{n}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i20}\mathbf{U}_i) &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t_i-2}\varepsilon_{ij}\varepsilon_{i(j+2)},\\
c_{22} &= \tfrac{1}{n}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i22}\mathbf{U}_i) &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j=3}^{t_i-2}\varepsilon_{ij}^2,\\
c_{12} = c_{21} &= \tfrac{1}{n}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{C}_{i12}\mathbf{U}_i) &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j=2}^{t_i-1}\varepsilon_{ij}\varepsilon_{i(j+1)}.
\end{aligned}$$

3.2 Moment and maximum likelihood estimates

Recall that the generalized least squares (GLS) estimates of $(\boldsymbol{\beta}, \sigma^2)$ as in (1.2.10) and (1.2.12) are given by

$$\hat{\boldsymbol{\beta}}_g = \left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\cdot\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i \quad (3.2.10)$$

$$\begin{aligned}
\hat{\sigma}_g^2 &= \frac{S(\boldsymbol{\beta}, \boldsymbol{\phi})}{n\bar{t}}\\
&= \frac{1}{\bar{t}}(c_{00} + \phi_1^2 c_{11} + \phi_2^2 c_{22} + 2\phi_1\phi_2 c_{12} - 2\phi_1 c_{10} - 2\phi_2 c_{20}). \quad (3.2.11)
\end{aligned}$$

Given the "residuals" $\hat{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i'\hat{\beta}$ from the model, the moment estimate $\hat{\boldsymbol{\phi}}_m$ may be obtained by setting $\boldsymbol{\Gamma}_i = \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'$. First we estimate the first and second lag autocorrelation ρ_1 and ρ_2 by the sample correlation r_1 and r_2, and then estimate ϕ_1 and ϕ_2 using the relationship between $\boldsymbol{\phi}$ and $\boldsymbol{\rho}$ in (3.1.3). We can estimate the k^{th} lag autocovariance γ_k by $\hat{c}_{k0}^* = \hat{c}_{k0}/(\bar{t}-k)$, $k = 0, 1, 2$. Thus r_1, r_2 are estimated by $\hat{c}_{10}^*/\hat{c}_{00}^*$ and $\hat{c}_{20}^*/\hat{c}_{00}^*$, respectively, hence

$$\begin{aligned}
\hat{\rho}_1 &= \frac{\hat{c}_{10}^*}{\hat{c}_{00}^*} = \frac{\hat{c}_{10}/(\bar{t}-1)}{\hat{c}_{00}/\bar{t}},\\
\hat{\rho}_2 &= \frac{\hat{c}_{20}^*}{\hat{c}_{00}^*} = \frac{\hat{c}_{20}/(\bar{t}-2)}{\hat{c}_{00}/\bar{t}}.
\end{aligned} \quad (3.2.12)$$

Therefore, the estimate of (ϕ_1, ϕ_2) is given by

$$\begin{aligned}
\hat{\phi}_1 &= \frac{\hat{\rho}_1(1-\hat{\rho}_2)}{1-\hat{\rho}_1^2} = \frac{\hat{c}_{10}^*(\hat{c}_{00}^* - \hat{c}_{20}^*)}{\hat{c}_{00}^{*2} - \hat{c}_{10}^{*2}},\\
\hat{\phi}_2 &= \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1-\hat{\rho}_1^2} = \frac{\hat{c}_{00}^*\hat{c}_{20}^* - \hat{c}_{10}^{*2}}{\hat{c}_{00}^{*2} - \hat{c}_{10}^{*2}}.
\end{aligned} \quad (3.2.13)$$

The moment estimates $(\hat{\phi}_m, \hat{\beta}_m)$ are then the simultaneous solutions of (3.2.13) and (3.2.10), and $\hat{\sigma}^2_m$ is given by (3.2.11) plugging in $(\hat{\phi}_m, \hat{\beta}_m)$.

Obtaining the maximum likelihood estimates is challenging. The log-likelihood function is given by (1.2.14). Equating to zero the partial derivative of the log-likelihood function with respect to β and σ^2, we obtain (3.2.10) and (3.2.11) respectively. Equating to zero the partial derivative of the log-likelihood function with respect to ϕ and recalling (1.2.17) we get

$$\sigma^2 \cdot \sum_{i=1}^{n} \frac{\partial \log |\mathbf{V}_i|}{\partial \phi} + \sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi} \mathbf{U}_i \right) = \mathbf{0}, \tag{3.2.14}$$

which is the ML equation for ϕ. Note that $|\mathbf{V}_i| = |\mathbf{V}_i^{-1}|^{-1}$, which implies

$$\frac{\partial \log |\mathbf{V}_i|}{\partial \phi} = -|\mathbf{V}_i^{-1}|^{-1} \frac{\partial |\mathbf{V}_i^{-1}|}{\partial \phi},$$

hence (3.2.14) becomes

$$\sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi} \mathbf{U}_i \right) - \sigma^2 \cdot \sum_{i=1}^{n} \left(|\mathbf{V}_i^{-1}|^{-1} \frac{\partial |\mathbf{V}_i^{-1}|}{\partial \phi} \right) = \mathbf{0}. \tag{3.2.15}$$

In order to simplify (3.2.15), we first note

$$\frac{\partial \mathbf{V}_i^{-1}(\phi)}{\partial \phi} = 2 \begin{bmatrix} \mathbf{C}_{i11} & \mathbf{C}_{i12} \\ \mathbf{C}_{i12} & \mathbf{C}_{i22} \end{bmatrix} [\phi \otimes \mathbf{I}_i] - 2 \begin{pmatrix} \mathbf{C}_{i10} \\ \mathbf{C}_{i20} \end{pmatrix},$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{I}_i$ is the identity matrix of order t_i. Thus the first term of (3.2.15) can be simplified as

$$\sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi} \mathbf{U}_i \right) = 2n \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} - 2n \begin{pmatrix} c_{10} \\ c_{20} \end{pmatrix}. \tag{3.2.16}$$

With some algebraic calculation, we can verify that

$$|\mathbf{V}_i^{-1}(\phi)| = |\mathbf{L}_i|^2 = (1+\phi_2)^2 \left[(1-\phi_2)^2 - \phi_1^2 \right], \tag{3.2.17}$$

which yields

$$\begin{aligned}
\frac{\partial|\mathbf{V}_i^{-1}(\phi)|}{\partial\phi_1} &= -2\phi_1(1+\phi_2)^2 \\
\frac{\partial|\mathbf{V}_i^{-1}(\phi)|}{\partial\phi_2} &= 2(1+\phi_2)\left[(1-\phi_2)^2-\phi_1^2\right]-2(1-\phi_2)(1+\phi_2)^2 \\
&= -2(1+\phi_2)(\phi_1^2-2\phi_2^2+2\phi_2). \qquad (3.2.18)
\end{aligned}$$

Substituting (3.2.16), (3.2.17) and (3.2.18) in (3.2.15) we get

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} - \begin{pmatrix} c_{10} \\ c_{20} \end{pmatrix} + \frac{\sigma^2}{(1+\phi_2)\left[(1-\phi_2)^2-\phi_1^2\right]}\begin{pmatrix} \phi_1(1+\phi_2) \\ \phi_1^2-2\phi_2^2+2\phi_2 \end{pmatrix} = \mathbf{0}. \qquad (3.2.19)$$

Furthermore, replacing σ^2 by (3.2.11) we get

$$\begin{aligned}
(\bar{t}-1)c_{11}\phi_1^3 \\
+(\bar{t}-2)(c_{12}\phi_2-c_{10})\phi_1^2 \\
-\left[\left(c_{00}-2c_{20}\phi_2+c_{22}\phi_2^2\right)+\bar{t}c_{11}(1-\phi_2)^2\right]\phi_1 \\
-\bar{t}(1-\phi_2)^2(c_{12}\phi_2-c_{10}) &= 0, \\
(\bar{t}-2)c_{22}\phi_2^4 \\
+\left[(\bar{t}-4)(c_{12}\phi_1-c_{20})-(\bar{t}-2)c_{22}\right]\phi_2^3 \\
-\left\{\left[(\bar{t}-1)c_{22}+2c_{11}\right]\phi_1^2+\left[(\bar{t}-4)c_{12}-4c_{10}\right]\phi_1\right. \\
\left.-\left[\bar{t}(c_{20}-c_{22})-2c_{00}-4c_{20}\right]\right\}\phi_2^2 \\
-\left\{(\bar{t}-2)c_{12}\phi_1^3-\left[2c_{11}-\bar{t}c_{22}+(\bar{t}-2)c_{20}\right]\phi_1^2\right. \\
\left.+(\bar{t}c_{12}+4c_{10})\phi_1-\left[2c_{00}+\bar{t}(c_{22}+c_{20})\right]\right\}\phi_2 \\
+\left[c_{11}\phi_1^4-(\bar{t}c_{12}+2c_{10})\phi_1^3+(c_{00}+\bar{t}c_{20})\phi_1^2+\bar{t}(c_{12}\phi_1-c_{20})\right] &= 0. \qquad (3.2.20)
\end{aligned}$$

The ML estimates $(\hat{\beta}_l, \hat{\phi}_l)$ are the simultaneous solutions of (3.2.10) and (3.2.20), and $\hat{\sigma}_l^2$ given by (3.2.11). Note that the exact MLEs are not in closed forms since it is hard to solve the "polynomial" equations (3.2.20). Approximate solutions could be obtained by solving (3.2.20) recursively. Box *et al.* (1994, p.300) suggest an approximate MLE for ϕ, by approximating the ML equation. This approximate MLE is given by

$$\hat{\phi}_{1al} = \frac{\hat{c}^*_{22}\hat{c}^*_{10} - \hat{c}^*_{12}\hat{c}^*_{20}}{\hat{c}^*_{11}\hat{c}^*_{22} - \hat{c}^{*2}_{12}}, \tag{3.2.21}$$

$$\hat{\phi}_{2al} = \frac{\hat{c}^*_{11}\hat{c}^*_{20} - \hat{c}^*_{12}\hat{c}^*_{10}}{\hat{c}^*_{11}\hat{c}^*_{22} - \hat{c}^{*2}_{12}}, \tag{3.2.22}$$

where 'al' stands for approximate MLE, and $\hat{c}^*_{kl} = \hat{c}_{kl}/(\bar{t}-k-l)$, $k = 1, 2$, $l = 0, 1, 2$ and $\hat{c}_{kl}$ is c_{kl} evaluating at $\hat{\boldsymbol{\beta}}$. We will discuss the details of this approximate MLE in Section 4.1.

3.3 Quasi-least squares estimates

In order to obtain the quasi-least squares estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, we need to minimize $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. Differentiating $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ with respect to $\boldsymbol{\beta}$ and equating to zero we get (3.2.10). The QLS estimate of $\boldsymbol{\phi}$ is obtained by two steps as follows.

The first step is to solve equation (1.2.20), given by

$$\sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\phi}} \mathbf{U}_i\right) = \mathbf{0}.$$

This is equivalent to setting (3.2.16) to be $\mathbf{0}$, which yields

$$\begin{aligned} \hat{\phi}_{1u} &= \frac{c_{22}c_{10} - c_{12}c_{20}}{c_{11}c_{22} - c^2_{12}}, \\ \hat{\phi}_{2u} &= \frac{c_{11}c_{20} - c_{12}c_{10}}{c_{11}c_{22} - c^2_{12}}. \end{aligned} \tag{3.3.23}$$

The above estimates are the ULS estimates, which are almost the same as the approximate MLEs in (3.2.22) when $\bar{t}$ is large. In the second step, we modify

$\hat{\boldsymbol{\phi}}_u = (\hat{\phi}_{1u}, \hat{\phi}_{2u})'$ to get consistent estimates. Recalling equation (1.2.21) we get

$$\begin{aligned}
\sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\boldsymbol{\phi}}_u)}{\partial \boldsymbol{\phi}} \cdot \mathbf{V}_i\right) &= \Delta^{-1} \cdot \sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\boldsymbol{\phi}}_u)}{\partial \boldsymbol{\phi}} \cdot \mathbf{P}_i\right) \\
&= 2\Delta^{-1} \sum_{i=1}^{n} \left\{ \begin{bmatrix} \operatorname{tr}(\mathbf{C}_{i11}\mathbf{P}_i) & \operatorname{tr}(\mathbf{C}_{i12}\mathbf{P}_i) \\ \operatorname{tr}(\mathbf{C}_{i12}\mathbf{P}_i) & \operatorname{tr}(\mathbf{C}_{i22}\mathbf{P}_i) \end{bmatrix} \hat{\boldsymbol{\phi}}_u \right. \\
&\qquad \left. - \begin{bmatrix} \operatorname{tr}(\mathbf{C}_{i10}\mathbf{P}_i) \\ \operatorname{tr}(\mathbf{C}_{i20}\mathbf{P}_i) \end{bmatrix} \right\} \\
&\overset{\text{set}}{=} \mathbf{0}.
\end{aligned}$$

Since

$$\begin{aligned}
\operatorname{tr}(\mathbf{C}_{ikk}\mathbf{P}_i(\boldsymbol{\rho})) &= (\bar{t} - 2k), \\
\operatorname{tr}(\mathbf{C}_{ik0}\mathbf{P}_i(\boldsymbol{\rho})) &= (\bar{t} - k)\rho_k, \quad k = 1, 2, \\
\operatorname{tr}(\mathbf{C}_{i12}\mathbf{P}_i(\boldsymbol{\rho})) &= (\bar{t} - 3)\rho_1,
\end{aligned}$$

the above equations can be written as

$$\begin{bmatrix} \bar{t} - 2 & (\bar{t} - 3)\rho_1 \\ (\bar{t} - 3)\rho_1 & \bar{t} - 4 \end{bmatrix} \hat{\boldsymbol{\phi}}_u - \begin{bmatrix} (\bar{t} - 1)\rho_1 \\ (\bar{t} - 2)\rho_2 \end{bmatrix} = \mathbf{0}.$$

Thus we have

$$\begin{aligned}
\hat{\rho}_1 &= \frac{(\bar{t} - 2)\hat{\phi}_{1u}}{(\bar{t} - 1) - (\bar{t} - 3)\hat{\phi}_{2u}}, \\
\hat{\rho}_2 &= \frac{(\bar{t} - 3)\hat{\phi}_{1u}^2}{(\bar{t} - 1) - (\bar{t} - 3)\hat{\phi}_{2u}} + \frac{(\bar{t} - 4)}{\bar{t} - 2}\hat{\phi}_{2u}.
\end{aligned}$$

From (3.1.3) we obtain

$$\begin{aligned}
\hat{\phi}_1 &= \frac{\hat{\rho}_1(1 - \hat{\rho}_2)}{1 - \hat{\rho}_1^2}, \\
\hat{\phi}_2 &= \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}.
\end{aligned} \tag{3.3.24}$$

The QLS estimates $(\hat{\boldsymbol{\beta}}_q, \hat{\boldsymbol{\phi}}_q)$ are the simultaneous solutions of (3.2.10) and (3.3.24), and $\hat{\sigma}_q^2$ is given by (3.2.11) plugging in $\hat{\boldsymbol{\beta}}_q$ and $\hat{\boldsymbol{\phi}}_q$.

3.4 A special case

An important special case is when the data is balanced. Assuming that $t_i \equiv t$, the notations of $\mathbf{V}, \mathbf{C}$ and $\mathbf{c}$ are the same as before, we just drop the subscript i. Since t_i's are all equal, $\mathbf{U}_i$'s have the same dimension. Define

$$\overline{\mathbf{U}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_i,$$

then we have

$$\begin{aligned} c_{00} &= \operatorname{tr}(\overline{\mathbf{U}}), \\ c_{10} &= \operatorname{tr}(\mathbf{C}_{10}\overline{\mathbf{U}}), \\ c_{11} &= \operatorname{tr}(\mathbf{C}_{11}\overline{\mathbf{U}}), \\ c_{20} &= \operatorname{tr}(\mathbf{C}_{20}\overline{\mathbf{U}}), \\ c_{22} &= \operatorname{tr}(\mathbf{C}_{22}\overline{\mathbf{U}}), \\ c_{12} = c_{21} &= \operatorname{tr}(\mathbf{C}_{12}\overline{\mathbf{U}}). \end{aligned}$$

The moment estimate of ϕ is given by (3.2.13), where $\hat{c}^*_{k0} = \hat{c}_{k0}/(t-k)$, $k = 0, 1, 2$, the MLE of ϕ is given by (3.2.20) with $\bar{t}$ replaced by t, and finally the QLS estimate of ϕ is given by (3.3.24) with $\bar{t}$ replaced by t.

3.5 Asymptotic properties

The asymptotic property of MLEs can be found by deriving the information matrix. First let us recall the following simple identities,

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 \\ \rho_2 &= \phi_1\rho_1 + \phi_2\rho_2 = \phi_2 + \phi_1\rho_1 \\ \Delta &= 1 - \phi_1\rho_1 - \phi_2\rho_2 = \frac{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}{1-\phi_2} \\ \mathbf{V}_2 &= \frac{1}{\sigma^2}\mathbf{\Gamma}_2 = \frac{1}{(1-\phi_2)\Delta(\phi)}\begin{bmatrix} 1-\phi_2 & \phi_1 \\ \phi_1 & 1-\phi_2 \end{bmatrix} \\ \mathbf{V}_2^{-1} &= (1+\phi_2)\begin{bmatrix} 1-\phi_2 & -\phi_1 \\ -\phi_1 & 1-\phi_2 \end{bmatrix} \end{aligned}$$

where $\mathbf{V}_2$ and $\mathbf{\Gamma}_2$ are $\mathbf{V}$ and $\mathbf{\Gamma}$ with dimension 2. We have the following results regarding to the partial derivative of the log-likelihood function with respect to $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, and σ^2:

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\
\frac{\partial l}{\partial \boldsymbol{\phi}} &= -\frac{n}{\sigma^2}\left[\mathbf{C}\boldsymbol{\phi} - \mathbf{c}_0 + \mathbf{\Gamma}_2\mathbf{D}\boldsymbol{\phi}\right] \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n\bar{t}}{2\sigma^2} + \frac{n}{2\sigma^4}\left(c_{00} - 2\boldsymbol{\phi}'\mathbf{c}_0 + \boldsymbol{\phi}'\mathbf{C}\boldsymbol{\phi}\right) \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i \\
\frac{\partial^2 l}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'} &= -\frac{n}{\sigma^2}\left\{2\mathbf{V}_2^2\begin{bmatrix} 2\phi_2(1+\phi_2) & 3\phi_1\phi_2 \\ \phi_1(1+\phi_2) & \phi_1^2 + 2\phi_2^2 \end{bmatrix} + \mathbf{\Gamma}_2\mathbf{D} + \mathbf{C}\right\} \\
\frac{\partial^2 l}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{n}{\sigma^6}\left(c_{00} - 2\boldsymbol{\phi}'\mathbf{c}_0 + \boldsymbol{\phi}'\mathbf{C}\boldsymbol{\phi}\right) \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta}'\partial \boldsymbol{\phi}} &= \frac{2}{\sigma^2}\sum_{i=1}^{n}\left\{[\mathbf{I}_2 \otimes (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})]\begin{pmatrix} \mathbf{C}_{i11} & \mathbf{C}_{i12} \\ \mathbf{C}_{i21} & \mathbf{C}_{i22} \end{pmatrix}[\boldsymbol{\phi}\otimes \mathbf{I}_i] - \begin{pmatrix} \mathbf{C}_{i10} \\ \mathbf{C}_{i20} \end{pmatrix}\right\}\mathbf{X}_i \\
\frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\
\frac{\partial^2 l}{\partial \boldsymbol{\phi}\partial \sigma^2} &= \frac{n}{\sigma^4}\left(\mathbf{C}\boldsymbol{\phi} - \mathbf{c}_0\right),
\end{aligned}
$$

where $\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$, $\mathbf{c}_0 = (c_{10}, c_{20})'$ and $\mathbf{D} = \mathrm{diag}(1, 2)$. The expected values of the second partial derivatives are given by

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right) &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'}\right) &= -\frac{n}{\sigma^2}\left[2\mathbf{V}_2^2\begin{pmatrix} 2\phi_2(1+\phi_2) & 3\phi_1\phi_2 \\ \phi_1(1+\phi_2) & \phi_1^2 + 2\phi_2^2 \end{pmatrix} + (\bar{t} - \mathbf{D})\mathbf{\Gamma}_2\right] \\
&\approx -n(\bar{t} - 2)\mathbf{V}_2 \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial (\sigma^2)^2}\right) &= -\frac{n\bar{t}}{2\sigma^4} \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}'\partial \boldsymbol{\phi}}\right) &= \mathbf{0}' \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}}\right) &= \mathbf{0} \\
\mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\phi}\partial \sigma^2}\right) &= -\frac{n}{\sigma^2}\,\mathbf{V}_2\,\mathbf{D}\boldsymbol{\phi}.
\end{aligned}
$$

Thus, the information matrix is given by

$$\mathbf{I}_n(\boldsymbol{\beta},\boldsymbol{\phi},\sigma^2)=\left[\begin{array}{ccc}\frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & n(\bar{t}-2)\mathbf{V}_2 & \frac{n}{\sigma^2}\mathbf{V}_2\mathbf{D}\boldsymbol{\phi} \\ \mathbf{0} & \frac{n}{\sigma^2}\boldsymbol{\phi}'\mathbf{D}\mathbf{V}_2 & \frac{n\bar{t}}{2\sigma^4}\end{array}\right].$$

Finally, the asymptotic variances of the ML estimates of $(\boldsymbol{\beta},\boldsymbol{\phi},\sigma^2)$ is given by

$$\mathbf{I}_n^{-1}(\boldsymbol{\beta},\boldsymbol{\phi},\sigma^2)=\frac{1}{n}\Sigma_l=\frac{1}{n}\operatorname{diag}(v_1,\Sigma_{2l}),$$

where $v_1=\sigma^2\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}$ and

$$\Sigma_{2l}=\frac{1}{k(\bar{t}-2)}\left[\begin{array}{cc} k\mathbf{V}_2^{-1}+\mathbf{D}\boldsymbol{\phi}\boldsymbol{\phi}'\mathbf{D} & -(\bar{t}-2)\sigma^2\mathbf{D}\boldsymbol{\phi} \\ -(\bar{t}-2)\sigma^2\boldsymbol{\phi}'\mathbf{D} & (\bar{t}-2)\sigma^4\end{array}\right],$$

where $k=\bar{t}(\bar{t}-2)/2-\boldsymbol{\phi}'\mathbf{D}\mathbf{V}_2\mathbf{D}\boldsymbol{\phi}$. Here we used the results

$$\left(\begin{array}{cc}\mathbf{A} & \mathbf{B}\\ \mathbf{B}' & \mathbf{D}\end{array}\right)^{-1}=\left(\begin{array}{cc}\mathbf{A}^{-1}+\mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1}\\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1}\end{array}\right),$$

where $\mathbf{E}=\mathbf{D}-\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{F}=\mathbf{A}^{-1}\mathbf{B}$. Other useful results regarding the inverse of a matrix include

$$\begin{aligned}(\mathbf{A}+\mathbf{u}\mathbf{v}')^{-1} &= \mathbf{A}^{-1}-\frac{(\mathbf{A}^{-1}\mathbf{u})(\mathbf{v}'\mathbf{A}^{-1})}{1+\mathbf{v}'\mathbf{A}^{-1}\mathbf{u}},\\ (\mathbf{A}+\mathbf{B}\mathbf{D}\mathbf{B}')^{-1} &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{D}^{-1}\right)^{-1}\mathbf{B}'\mathbf{A}^{-1}\\ &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\mathbf{E}\mathbf{B}'\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\mathbf{E}(\mathbf{E}+\mathbf{D})^{-1}\mathbf{E}\mathbf{B}\mathbf{A}'^{-1},\end{aligned}$$

where $\mathbf{E}=(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}$. The asymptotic distribution of the moment and quasi-least squares estimates could be established by finding the asymptotic distribution of c_{00}, c_{10}, c_{20}, c_{11}, c_{22} and c_{12}, and then applying the delta theorem. But the exact forms are too complicated. We omit the tedious computations and study the asymptotic properties using simulations.

3.6 Comparisons through simulation

Since the asymptotic properties of the estimates are so difficult to obtain theoretically, we will compare them by simulation. As we have studied previously, the asymptotic properties of $\boldsymbol{\beta}$ and σ^2 derived from three methods will be similar to the asymptotic properties of ϕ since they have the same functional forms. Therefore, we will concentrate on comparing the estimates of ϕ. By finding the asymptotic relative efficiencies (AREs), we will show that MLEs will work best if the error is normal, while QLS estimates may be better when the errors follow a contaminated normal distribution. Without loss of generality, we assume $\boldsymbol{\beta} = 0$ and $\sigma^2 = 1$ in our simulation. For convenient purpose, we let $t_i \equiv t$. Since we have two parameters to be estimated, it will be hard to simulate the data for all possible values of n and t. Therefore, we will generate samples only when $n = 30$ and $t = 20$. Generally speaking, the efficiencies of QLS and MOM estimates with respect to MLEs will increase when n and t becomes larger. We chose n to be 30 in order to make sure we have enough replications, and we chose t to be 20 to illustrate that the number of repeated measurements for each replication may be small.

We now study the asymptotic properties of the estimates when the data is normal. First we generate a sample of $n = 30$, $t-$dimensional ($t = 20$) vectors $\boldsymbol{\varepsilon}$ whose elements are from a standard normal distribution, and then we let $\mathbf{y} = \sigma^2\mathbf{V}^{1/2}(\phi)\boldsymbol{\varepsilon}$ ($\sigma^2 = 1$). The $\mathbf{y}$ will have mean $\mathbf{0}$ and variance $\sigma^2\mathbf{V}(\phi)$. The process is then repeated 10000 times. For each replication, we compute the MOME, MLE and QLSE of ϕ, and then compute the biases and mean square errors (MSEs). Obtaining the MLEs was hard since MLEs are not in closed forms. We have to solve equations (3.2.20) simultaneously. One could use Newton-Raphson method to solve the equations iteratively (see Appendix A), but we used the following method. Given the k^{th} step solution $\hat{\phi}^k$, the $(k+1)^{th}$ step solution $\hat{\phi}^{k+1}$ is obtained by solving equations (3.2.20) substituting the 'polynomial' coefficients with $\hat{\phi}^k$. We repeated this procedure until the difference between $\hat{\phi}^{k+1}$ and $\hat{\phi}^k$ is in the neighborhood of 10^{-6} or 10^{-7}. Our method turned out to be more efficient in the sense that the iterations converged faster. If the iteration procedure still does not converge after 500 steps (one can choose the maximum number of iteration differently, in our simulation, 500 is enough for the iteration to converge for most cases). The replications where the iterative process did not converge were excluded from the analysis when calculating the biases and MSEs. The QLS estimates and moment estimates are in closed forms. Given the estimate $\hat{\phi}_{1i}$, $i = 1, \ldots, 10000$, the bias of ϕ_1 is computed as $\text{Bias}(\hat{\phi}_1) = \sum_{i=1}^{10000}(\hat{\phi}_{1i} - \phi_1)/10000$, and the MSE of ϕ_1 is found by $\text{MSE}(\hat{\phi}_1) = \sum_{i=1}^{10000}(\hat{\phi}_{1i} - \phi_1)^2/10000$. The bias and MSE of ϕ_2 are obtained in the same way. The biases of the estimates obtained through the three methods are all close to 0 and do not differ much from each other, so the interest has been focused on comparing the MSEs.

Let us define the ARE of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ as $e(\hat{\phi}_{1q}, \hat{\phi}_{1l}) = MSE_{1l}/MSE_{1q}$. The subscript one is understandable to represent the notations associated with ϕ_1, and 'l', 'q' stand for 'Maximum likelihood estimate' and 'Quasi-least squares estimate', respectively. The ARE of $\hat{\phi}_{2q}$ with respect to $\hat{\phi}_{2l}$ is defined similarly. Table 3.1 and 3.2 contain the AREs of $\hat{\phi}_{1q}$ and $\hat{\phi}_{2q}$, respectively. We can see that the numbers are very close to one except when $\boldsymbol{\phi}$ is near the boundary. Note that the feasible region of ϕ_1 and ϕ_2 is triangular. Figures 3.3 and 3.4 are the 3D plots of the ARE of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$, and the ARE of $\hat{\phi}_{2q}$ with respect to $\hat{\phi}_{2l}$, respectively. The graphs are almost flat at one, except when $\boldsymbol{\phi}$ is near the boundary where the efficiencies are slightly below one. Therefore QLS estimate is as good as the MLE on a wide region. Tables 3.3 and 3.4 contain part of the AREs of $\hat{\phi}_{1m}$ with respect to $\hat{\phi}_{1q}$ and $\hat{\phi}_{2m}$ with respect to $\hat{\phi}_{2q}$, respectively. The contour plot of the ARE of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ (Figure 3.5) gives a more clear view. The contour plot of the ARE of $\hat{\phi}_{2q}$ with respect to $\hat{\phi}_{2l}$ is similar to Figure 3.5. The plots of the ARE of the moment estimate with respect to the MLE, which is the ratio of MSE_l

Table 3.1. ARE of $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is normal.

	ϕ_2						
ϕ_1	−0.9	−0.6	−0.3	0	0.3	0.6	0.9
−1.8	0.8935						
−1.5	0.9411	0.9961					
−1.2	0.9561	0.9931	0.9958				
−0.9	0.9646	0.9991	0.9990	0.9976			
−0.6	0.9984	0.9997	1.0020	0.9999	0.9963		
−0.3	0.9944	0.9993	0.9995	0.9997	1.0007	0.9941	
0	0.9939	0.9987	0.9996	0.9998	0.9972	0.9990	0.8343
0.3	0.9989	0.9978	0.9995	1.0000	1.0008	0.9946	
0.6	0.9972	0.9982	0.9986	1.0003	0.9964		
0.9	0.9716	1.0031	0.9996	0.9980			
1.2	0.9529	0.9986	0.9977				
1.5	0.9124	0.9932					
1.8	0.8693						

Table 3.2. ARE of $\hat{\phi}_{2q}$ vs $\hat{\phi}_{2l}$ when the data is normal.

	ϕ_2						
ϕ_1	−0.9	−0.6	−0.3	0	0.3	0.6	0.9
−1.8	0.8708						
−1.5	0.8695	0.9944					
−1.2	0.8540	0.9839	0.9948				
−0.9	0.8971	0.9979	0.9971	0.9904			
−0.6	0.8560	0.9933	1.0021	0.9993	0.9911		
−0.3	0.8859	0.9988	0.9987	1.0007	0.9983	0.9720	
0	0.8550	0.9844	0.9972	0.9990	0.9983	0.9987	0.8688
0.3	0.8416	0.9968	0.9959	1.0001	0.9979	0.9820	
0.6	0.8561	0.9904	1.0002	0.9976	0.9956		
0.9	0.8826	0.9972	1.0003	0.9877			
1.2	0.8635	1.0007	0.9989				
1.5	0.8498	0.9911					
1.8	0.8457						

Table 3.3. *ARE of* $\hat{\phi}_{1m}$ *vs* $\hat{\phi}_{1q}$ *when the data is normal.*

	ϕ_2						
ϕ_1	−0.9	−0.6	−0.3	0	0.3	0.6	0.9
−1.8	0.0093						
−1.5	0.1178	0.0758					
−1.2	0.3073	0.4386	0.3276				
−0.9	0.4622	0.6596	0.7629	0.7098			
−0.6	0.5528	0.8282	0.8595	0.9078	0.8540		
−0.3	0.6581	0.9092	0.9611	0.9840	0.9775	0.8989	
0	0.6815	0.9579	0.9950	1.0014	1.0073	0.9575	1.0966
0.3	0.6360	0.9419	0.9677	0.9812	0.9731	0.8966	
0.6	0.5577	0.8206	0.9126	0.9123	0.8389		
0.9	0.4577	0.6651	0.7403	0.6744			
1.2	0.3026	0.4582	0.3369				
1.5	0.1128	0.0784					
1.8	0.0078						

Table 3.4. *ARE of* $\hat{\phi}_{2m}$ *vs* $\hat{\phi}_{2q}$ *when the data is normal.*

	ϕ_2						
ϕ_1	−0.9	−0.6	−0.3	0	0.3	0.6	0.9
−1.8	0.0107						
−1.5	0.1511	0.0977					
−1.2	0.3515	0.5639	0.4113				
−0.9	0.4546	0.7076	0.8615	0.7485			
−0.6	0.5499	0.7953	0.8980	0.9339	0.7507		
−0.3	0.5917	0.8171	0.9516	0.9748	0.9479	0.7201	
0	0.6346	0.8744	0.9851	1.0058	0.9604	0.8345	0.6312
0.3	0.6813	0.8276	0.9754	0.9791	0.9459	0.7084	
0.6	0.5624	0.8169	0.9250	0.9642	0.6914		
0.9	0.4247	0.7395	0.8501	0.7204			
1.2	0.3265	0.5328	0.4223				
1.5	0.1381	0.1023					
1.8	0.0093						

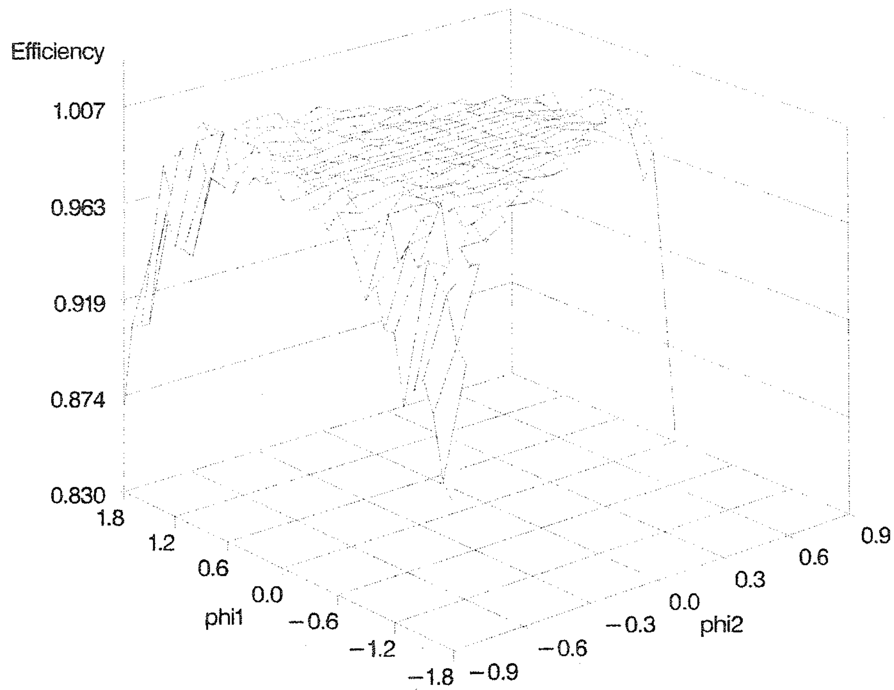

Figure 3.3. ARE of $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is normal.

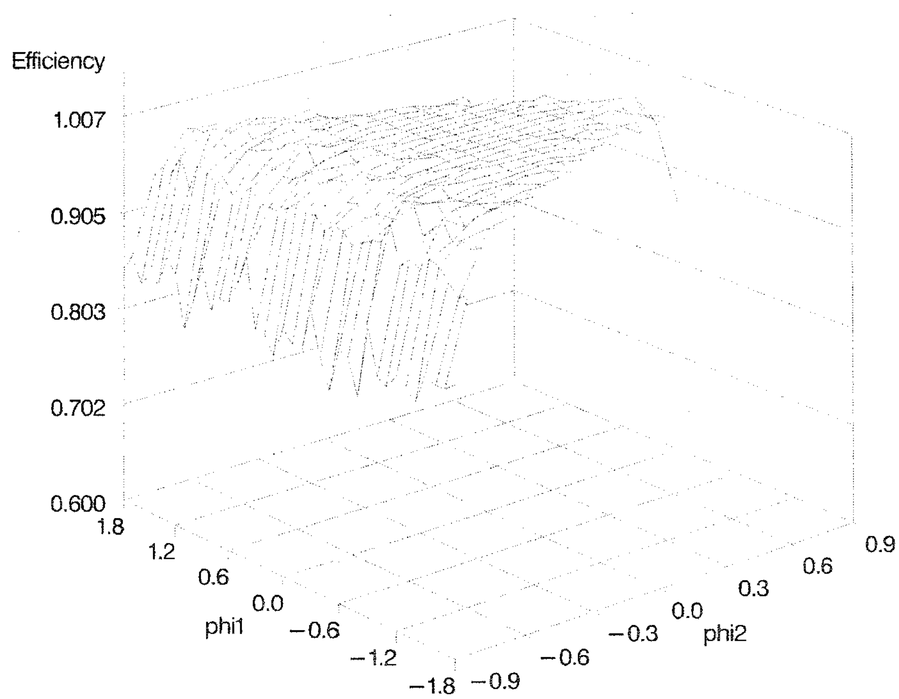

Figure 3.4. ARE of $\hat{\phi}_{2q}$ vs $\hat{\phi}_{2l}$ when the data is normal.

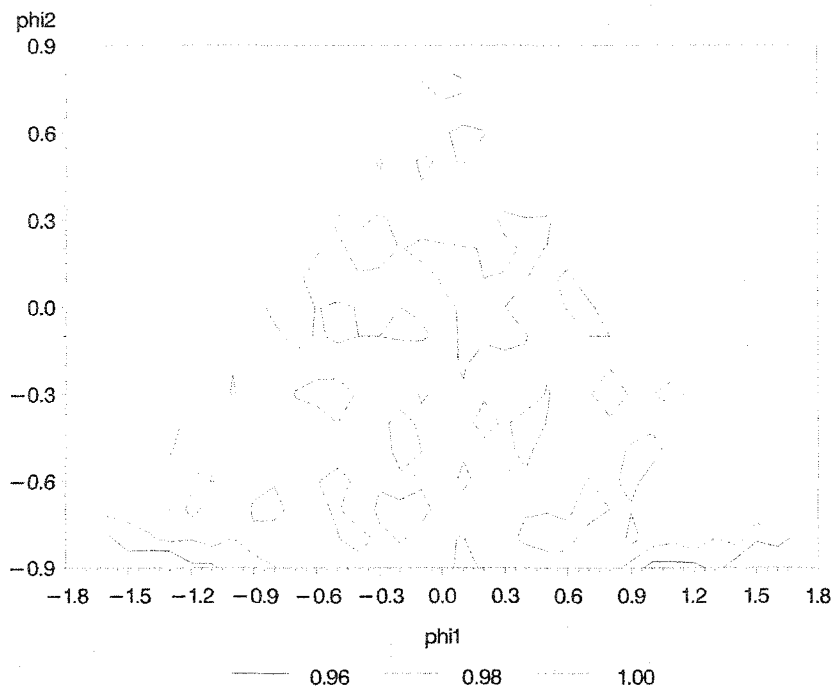

Figure 3.5. ARE of $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is normal.

and MSE_m, shows similar patterns.

Let us now compare the QLS estimates and the moment estimates. Define the ARE of $\hat{\phi}_{1m}$ with respect to $\hat{\phi}_{1q}$ as $e(\hat{\phi}_{1q},\ \hat{\phi}_{1m}) = MSE_{1q}/MSE_{1m}$ and define the ARE of $\hat{\phi}_{2m}$ with respect to $\hat{\phi}_{2q}$ similarly. Figures 3.6 and 3.7 give the 3D plots of the AREs. We can see that the efficiencies are far away from one, especially when ϕ is close to the boundary. This fact is demonstrated more clear in the contour plot of the ARE of $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1q}$ (Figure 3.8, the contour plot of the ARE of $\hat{\phi}_{2m}$ vs $\hat{\phi}_{2q}$ is similar). This shows that the QLS method is better than the moment method. Now consider a special case when $\phi_2 = 0$, the AR(2) process reduces to AR(1) process. Define the ARE of $\hat{\phi}_m$ with respect to $\hat{\phi}_l$ as the ratio of MSE_l and MSE_m, and the ARE of $\hat{\phi}_q$ with respect to $\hat{\phi}_1$ as the ratio of MSE_l and MSE_q. Since $\phi_2 \equiv 0$, we only need to study the estimates of ϕ_1. Figure 3.9 shows the AREs of $\hat{\phi}_{1m}$ and $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$, we can see that the curve representing the ARE

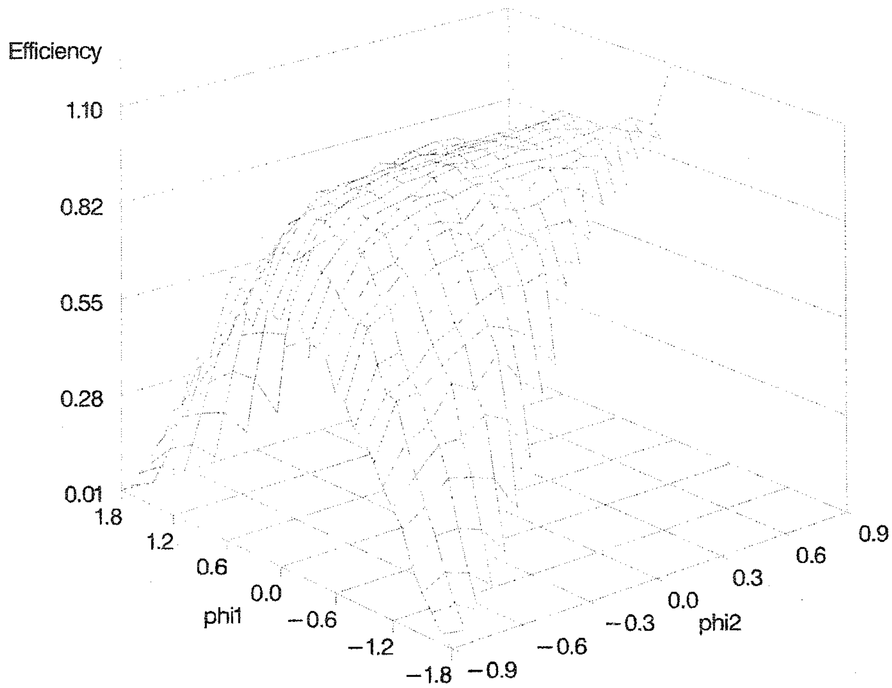

Figure 3.6. ARE of $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1q}$ when the data is normal.

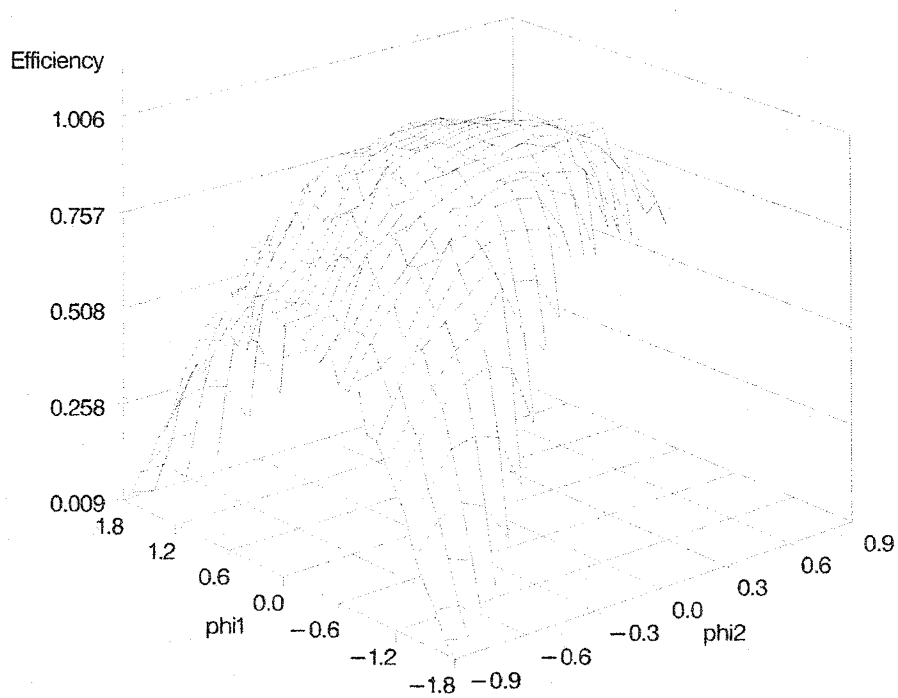

Figure 3.7. ARE of $\hat{\phi}_{2m}$ vs $\hat{\phi}_{2q}$ when the data is normal.

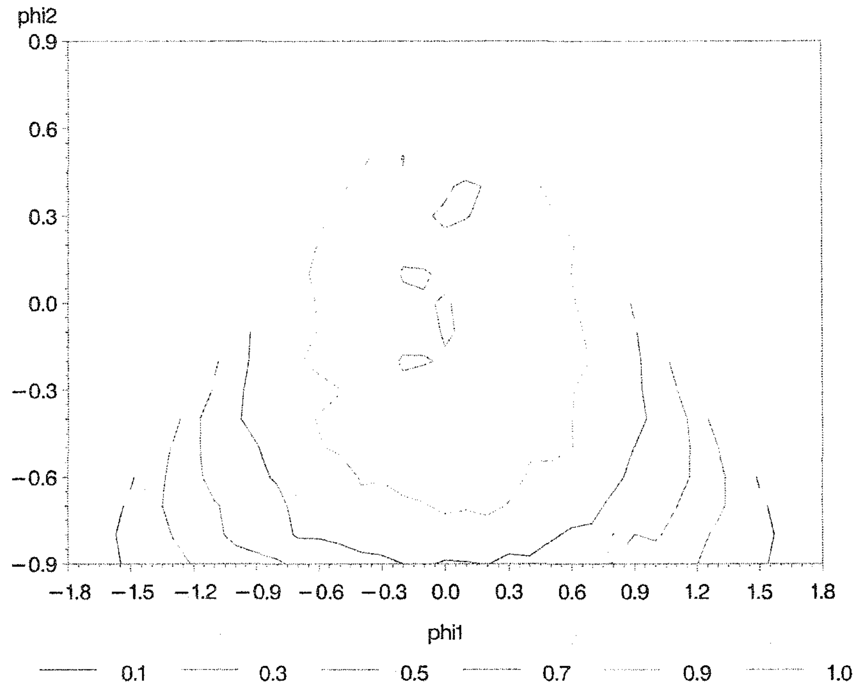

Figure 3.8. ARE of $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1q}$ when the data is normal.

of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ stays around one, while the curve representing the ARE of $\hat{\phi}_{1m}$ with respect to $\hat{\phi}_{1l}$ goes down quickly when ϕ_1 approaches the boundary. The result is consistent with the one in the previous chapter.

We now simulate the data from a contaminated normal distribution to study the robustness of the estimates. We chose the distribution to be $0.5\,N(0,4) + 0.5\,N(1,4)$. The mean and variance of this distribution are given by 0.5 and 4.25, respectively. In general, for any two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, if they are mixed evenly, the mean of the contaminated distribution is given by $(\mu_1 + \mu_2)/2$, and the variance is given by $(\sigma_1^2 + \sigma_2^2)/2 + (\mu_1 - \mu_2)^2/4$. The density plot is shown in Figure 3.10. From the plot, we can see that the distribution is very close to the normal distribution. In practice, it is easy to mistake the contaminated normal distribution with the normal distribution. Thus, it is necessary to derive an estimating method which is more reliable when the data may come from a

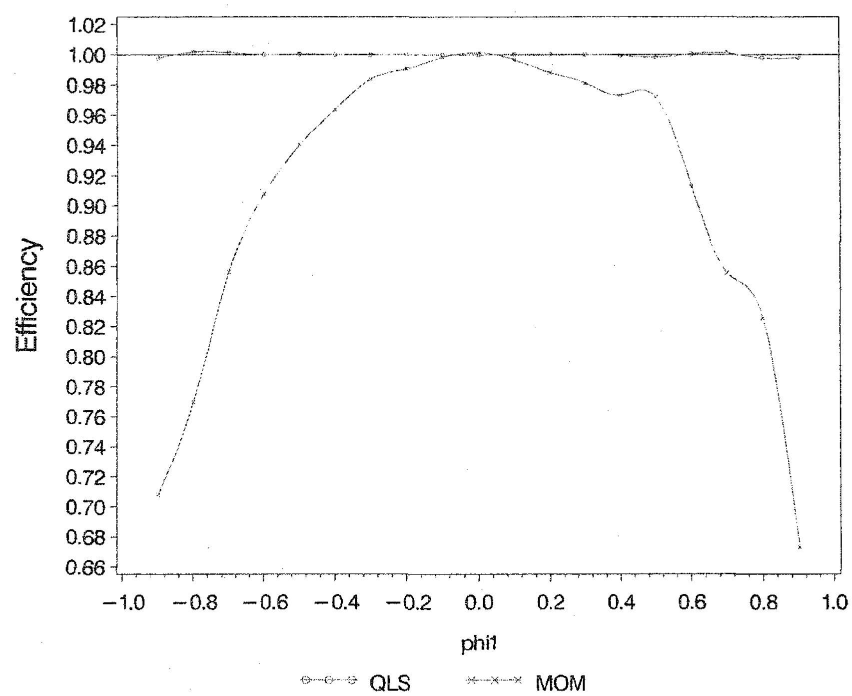

Figure 3.9. AREs of $\hat{\phi}_{1m}$ and $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is normal and $\phi_2 = 0$.

distribution that differs from or is similar to the normal distribution. We choose the distribution $0.5\,N(0,\,4) + 0.5\,N(1,\,4)$ just for illustrating purpose; the results should be similar for other contaminated normal distribution. Recall that in the previous chapter, we simulated the data from a Beta distribution when the error of the model is an AR(1) process, it should be noted that in AR(2) case, the results when simulating the data from a Beta distribution are similar to the results when simulating the data from a contaminated normal distribution, while the results are more noticeable in the latter case. We now describe how the data was generated. First we select a random number between 0 and 1, if the number is less than 0.5, we generate ε from $N(0, 4)$; otherwise, we generate ε from $N(1, 4)$. The ε generated this way will have mean 0.5 and variance 4.25, we modify ε to have mean 0 and variance 1 by subtracting it from 0.5 and then dividing by $\sqrt{4.25}$. We generated a vector of ε of size $t = 20$ consisting of ε's. Next, we generated a sample of $n = 30$

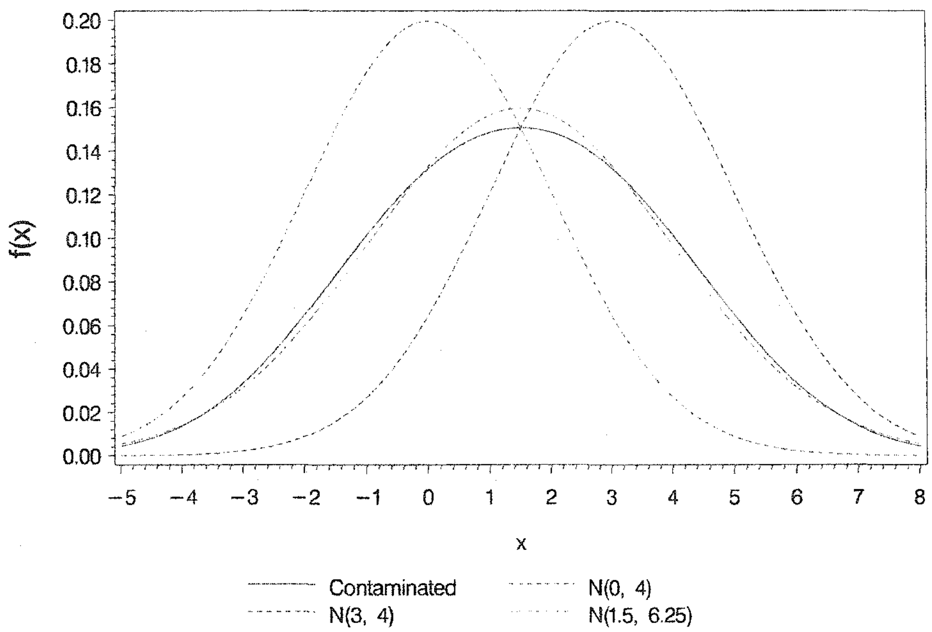

Figure 3.10. The density of the contaminated normal distribution $0.5N(0, 4) + 0.5N(1, 4)$.

vectors $\mathbf{y} = \sigma^2 \mathbf{V}^{1/2}(\boldsymbol{\phi})\varepsilon$'s ($\sigma^2 = 1$). The $\mathbf{y}$ will have mean $\mathbf{0}$ and variance $\sigma^2\mathbf{V}(\boldsymbol{\phi})$. The process is then repeated 10,000 times. For each replication, we compute the MOME, MLE and QLSE of $\boldsymbol{\phi}$, and then compute the biases and mean square errors (MSEs). When calculating the MLEs, we again use the method described before instead of Newton-Raphson method, and exclude the solutions when the MLEs do not converge. As before, let us define the ARE of $\hat{\boldsymbol{\phi}}_q$ with respect to $\hat{\boldsymbol{\phi}}_l$ as the ratio of MSE_l and MSE_q. Tables 3.5 and 3.6 contain the AREs of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ and $\hat{\phi}_{2q}$ with respect to $\hat{\phi}_{2l}$, respectively. Figures 3.11 and 3.12 give the 3D plots of the AREs of the QLS estimates with respect to the MLEs. We can see that the most of the numbers are greater than one, especially when ϕ_1 approaches the positive boundary. The contour plot of the ARE of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ (Figure 3.13) gives a more clear view. The contour plot of the ARE of $\hat{\phi}_{2q}$ with respect to $\hat{\phi}_{2l}$ is similar to Figure 3.13. Based on these plots, we can conclude that

Table 3.5. ARE of $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is simulated from contaminated normal $0.5\,N(0,4) + 0.5\,N(1,4)$.

	ϕ_2							
ϕ_1	−0.4	−0.2	−0.1	0.1	0.2	0.3	0.4	0.5
−1.2	1.0068							
−0.9	1.0044	1.0038	1.0052					
−0.6	1.0038	1.0022	1.0026	1.0030	1.0054	1.0088		
−0.3	1.0001	1.0008	1.0009	1.0011	1.0002	1.0027	1.0039	1.0091
0	1.0008	1.0001	1.0000	1.0008	1.0009	1.0024	1.0033	1.0065
0.3	0.9988	0.9997	1.0001	1.0016	1.0026	1.0077	1.0123	1.0581
0.4	0.9996	0.9991	1.0006	1.0027	1.0055	1.0609	1.0948	0.9938
0.5	1.0001	0.9990	1.0009	1.0816	1.0070	1.4403	0.9894	
0.6	0.9993	0.9997	1.0015	1.0036	1.4658	0.9931		
0.7	0.9960	1.0013	1.0011	1.5868	0.9960			
0.9	0.9975	1.0047	2.1031					
1.0	0.9962	1.4912	0.9929					
1.2	1.7925							

Table 3.6. ARE of $\hat{\phi}_{2q}$ vs $\hat{\phi}_{2l}$ when the data is simulated from contaminated normal $0.5N(0,4) + 0.5N(1,4)$.

	ϕ_2							
ϕ_1	−0.4	−0.2	−0.1	0.1	0.2	0.3	0.4	0.5
−1.2	0.9996							
−0.9	1.0035	1.0010	0.9919					
−0.6	1.0040	1.0002	0.9990	0.9958	0.9913	0.9929		
−0.3	1.0033	1.0009	0.9996	0.9980	0.9989	0.9975	0.9955	0.9916
0	1.0036	1.0011	0.9999	0.9986	1.0005	1.0020	1.0082	1.0077
0.3	1.0012	0.9997	0.9997	1.0012	1.0058	1.0111	1.0210	1.1753
0.4	1.0039	1.0001	1.0004	1.0053	1.0136	1.2112	1.2191	0.9888
0.5	1.0012	0.9987	1.0015	1.2821	1.0147	1.9590	0.9980	
0.6	1.0014	1.0015	1.0024	1.0170	1.9486	0.9932		
0.7	0.9993	1.0019	1.0079	2.0838	0.9972			
0.9	0.9997	1.0299	2.6084					
1.0	0.9998	1.7222	1.0563					
1.2	2.1639							

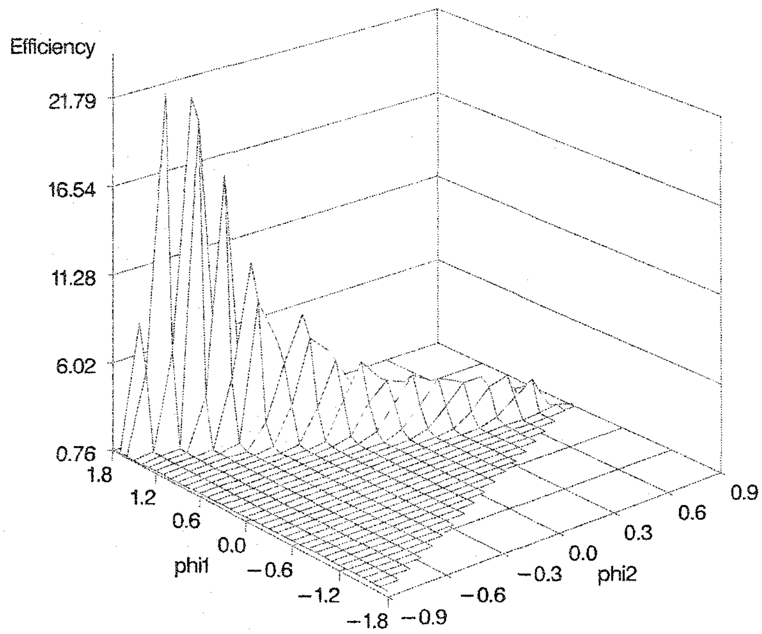

Figure 3.11. ARE of $\hat{\phi}_{1q}$ *vs* $\hat{\phi}_{1l}$ *when the data is contaminated normal* $0.5N(0,4) + 0.5N(1,4)$.

QLS estimates are better than the MLEs.

Now let us compare the QLS estimates and the moment estimates. Define the ARE of $\hat{\boldsymbol{\phi}}_m$ with respect to $\hat{\boldsymbol{\phi}}_q$ as the ratio of MSE_q and MSE_m. Tables 3.7 and 3.8 contain the AREs of $\hat{\phi}_{1m}$ with respect to $\hat{\phi}_{1q}$ and $\hat{\phi}_{2m}$ with respect to $\hat{\phi}_{2q}$, respectively. Figure 3.14 and Figure 3.15 give the 3D plots of the AREs of the moment estimates with respect to the QLS estimates. It is clear that the efficiencies are almost all less than 1, especially when $\boldsymbol{\phi}$ is close to the boundary. The contour plot of the ARE of $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1q}$ (Figure 3.16, the contour plot of the ARE of $\hat{\phi}_{2m}$ vs $\hat{\phi}_{2q}$ is similar) shows a much more clear view. Based on the plots, we again see that QLS method works much better than moment method. When $\phi_2 = 0$, the AR(2) process reduces to AR(1) process. As before, define the ARE of $\hat{\phi}_m$ with respect to $\hat{\phi}_l$ as the ratio of MSE_l and MSE_m, and the ARE of $\hat{\phi}_q$ with respect to $\hat{\phi}_l$ as the

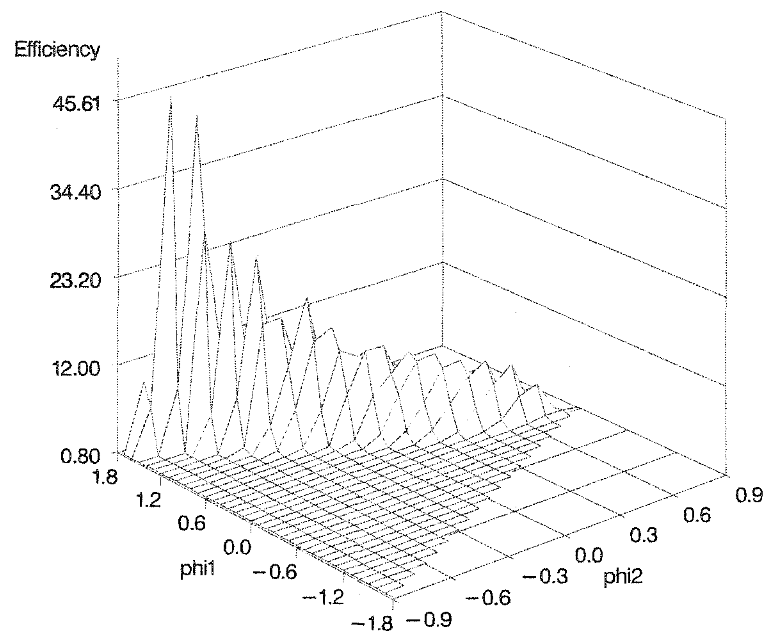

Figure 3.12. *ARE of* $\hat{\phi}_{2q}$ *vs* $\hat{\phi}_{2l}$ *when the data is contaminated normal* $0.5N(0,4) + 0.5N(1,4)$.

Table 3.7. AREs of $\hat{\phi}_{1m}$ *vs* $\hat{\phi}_{1q}$ *when the data is simulated from contaminated normal* $0.5N(0,4) + 0.5N(1,4)$.

	ϕ_1						
ϕ_2	−0.9	−0.6	−0.3	0	0.3	0.6	0.9
−1.8	0.0111						
−1.5	0.1291	0.1148					
−1.2	0.2869	0.4906	0.4587				
−0.9	0.3620	0.6957	0.7823	0.7422			
−0.6	0.5336	0.8063	0.9106	0.9390	0.8932		
−0.3	0.6163	0.9003	0.9675	0.9926	0.9798	0.8987	
0	0.6785	0.9592	0.9887	0.9995	0.9823	0.9543	1.0054
0.3	0.5674	0.9380	0.9641	0.9678	0.9607	0.7997	
0.6	0.5197	0.9133	0.9211	0.9126	0.8288		
0.9	0.4680	0.7315	0.7147	0.6101			
1.2	0.2929	0.4675	0.9390				
1.5	0.1262	0.0308					
1.8	0.0047						

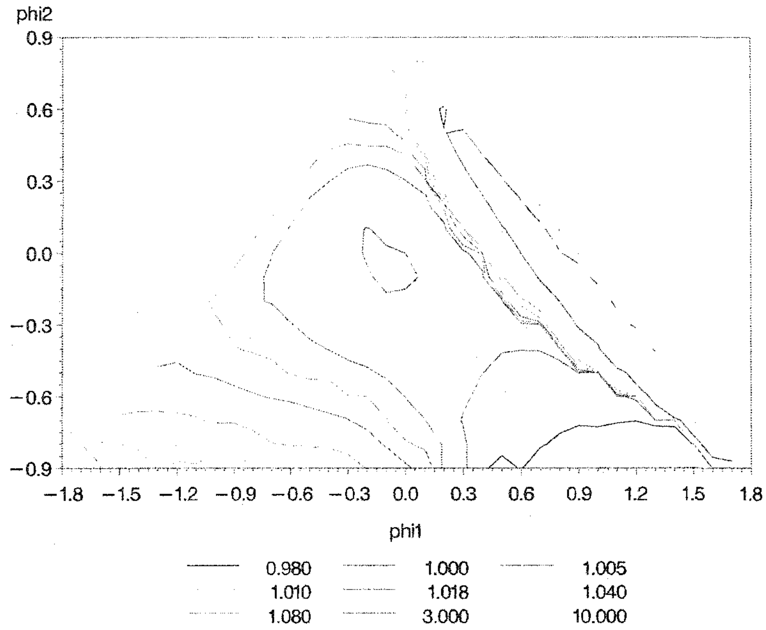

Figure 3.13. ARE of $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is simulated from contaminated normal $0.5N(0,4)+0.5N(1,4)$.

Table 3.8. AREs of $\hat{\phi}_{2m}$ vs $\hat{\phi}_{2q}$ when the data is simulated from contaminated normal $0.5N(0,4)+0.5N(1,4)$.

	ϕ_1						
ϕ_2	−0.9	−0.6	−0.3	0	0.3	0.6	0.9
−1.8	0.0125						
−1.5	0.1594	0.1475					
−1.2	0.3199	0.6004	0.5541				
−0.9	0.3952	0.7565	0.8776	0.7961			
−0.6	0.5371	0.7696	0.9335	0.9711	0.8622		
−0.3	0.5391	0.8023	0.9536	0.9840	0.9801	0.8414	
0	0.6336	0.8367	0.9802	0.9990	0.9532	0.8318	0.8052
0.3	0.6264	0.8272	0.9863	0.9771	0.8666	0.5637	
0.6	0.5688	0.8234	0.9705	0.8881	0.6521		
0.9	0.4547	0.7453	0.9243	0.5968			
1.2	0.3446	0.6368	0.2386				
1.5	0.1455	0.0417					
1.8	0.0077						

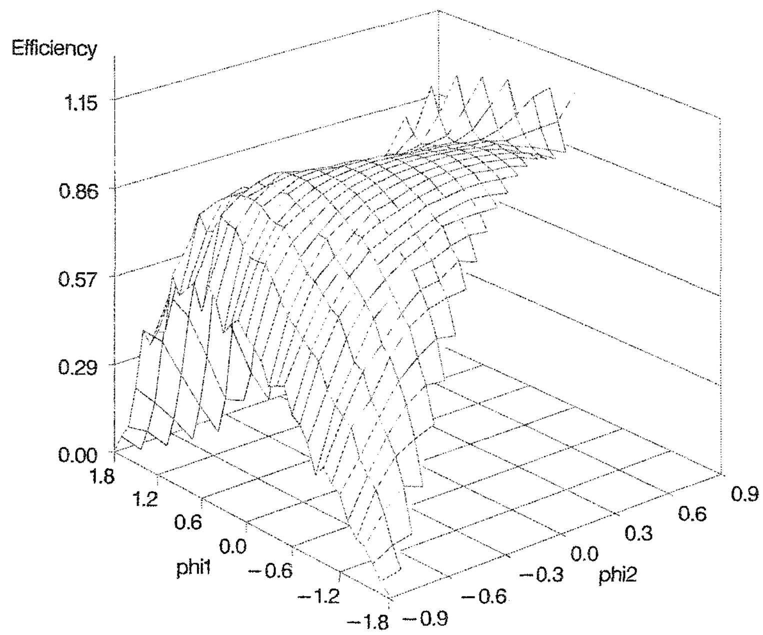

Figure 3.14. ARE of $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1q}$ when the data is simulated from contaminated normal $0.5N(0,4) + 0.5N(1,4)$.

ratio of MSE_l and MSE_q. Since $\phi_2 \equiv 0$, we only need to study the estimates of ϕ_1. Figure 3.17 shows the AREs of $\hat{\phi}_{1m}$ and $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$, we can see that the curve representing the ARE of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ stays above 1, especially when ϕ_1 is close to the boundary, while the one representing the ARE of $\hat{\phi}_{1m}$ with respect to $\hat{\phi}_{1l}$ goes down quickly when ϕ_1 approaches the boundary. The result is also consistent with the one in the previous chapter.

Another criteria to compare estimates is the bias. Given the estimate $\hat{\phi}_{1i}$, $i = 1, \ldots, 10000$, the bias of ϕ_1 is found by $\text{Bias}(\hat{\phi}_1) = \sum_{i=1}^{10000}(\hat{\phi}_{1i} - \phi_1)/10000$. In order to make comparisons, the bias has been changed to absolute value. Figure 3.18 shows the biases of all the three estimates when $\phi_2 = -0.5$, we see that the biases of QLS estimates and the biases of MLEs are very close, while the biases of moment estimate are far large than them. The bias of the estimate of ϕ_1 as a function of

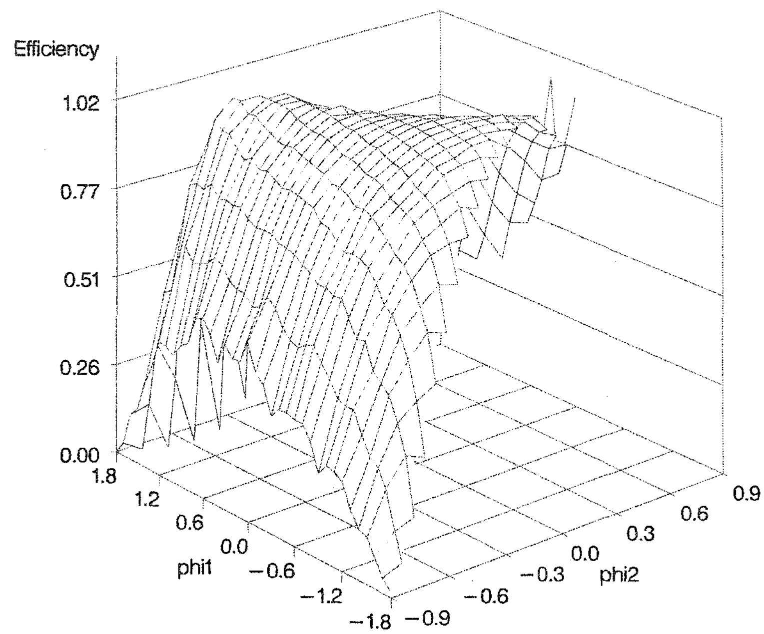

Figure 3.15. ARE of $\hat{\phi}_{2m}$ vs $\hat{\phi}_{2q}$ when the data is simulated from contaminated normal $0.5N(0,4) + 0.5N(1,4)$.

ϕ_2 and the bias of ϕ_2 as a function of ϕ_1 are similar. This further is a convincing evidence that the QLS estimates are slightly better than MLEs and much better than moment estimates, when the data is simulated from a contaminated normal distribution.

As a summary, we again demonstrate that, when the errors follow an AR(2) process, QLS estimates are better than moment estimates, and good competitors to MLEs. When the data is from a distribution which differs slightly from normal, QLS estimates may be better than the MLEs.

3.7 Application to a dental study

Let us recall the dental study that we discussed in the previous chapter. The

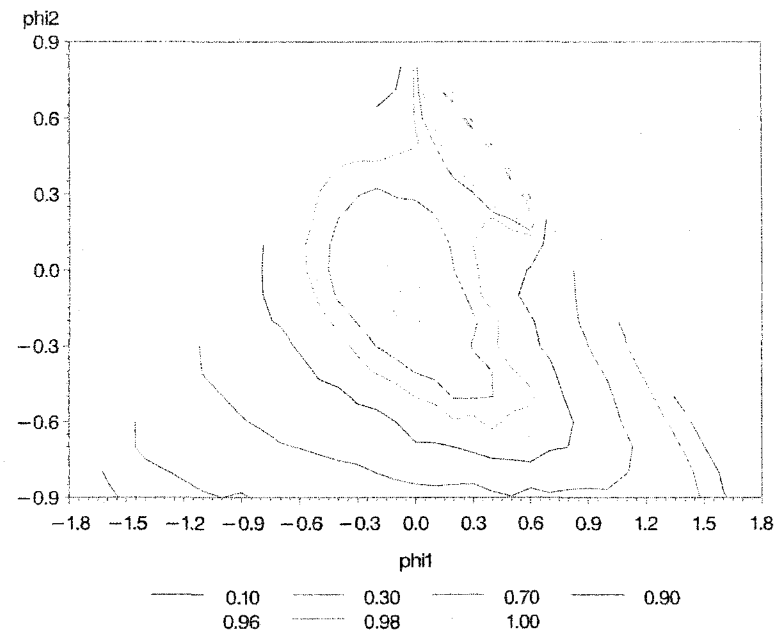

Figure 3.16. ARE of $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1q}$ when the data is simulated from contaminated normal $0.5N(0,4) + 0.5N(1,4)$.

data is shown in Table 2.4 and the plot is given by Figure 2.16. We fit the same regression model as before:

$$\mu_{ij} = \beta_g x_{i1} + \beta_b x_{i2} + \gamma_g x_{i1} \cdot x_{j3} + \gamma_b x_{i2} \cdot x_{j3}, \quad 1 \le j \le 4, \ \ 1 \le i \le 27,$$

where x_{i1}, x_{i2} are indicators for the girl and boy, respectively, and x_{j3} is the subject's age at the jth measurement time. Recall that the normality assumption of the girls group is satisfied, but not for the boys group. Assume the error is AR(2) process, we estimate the parameters using moment, maximum likelihood and quasi-least squares methods. Table 3.9 contains estimates for the regression parameters, the autoregressive parameters and the scale parameter. The residual plot based on the quasi-least estimates is in Figure 3.19. We see that the residuals are almost randomly distributed around 0 at different time, regardless of the patient group. Note that a more appropriate model would be is to fit different covariance matrix

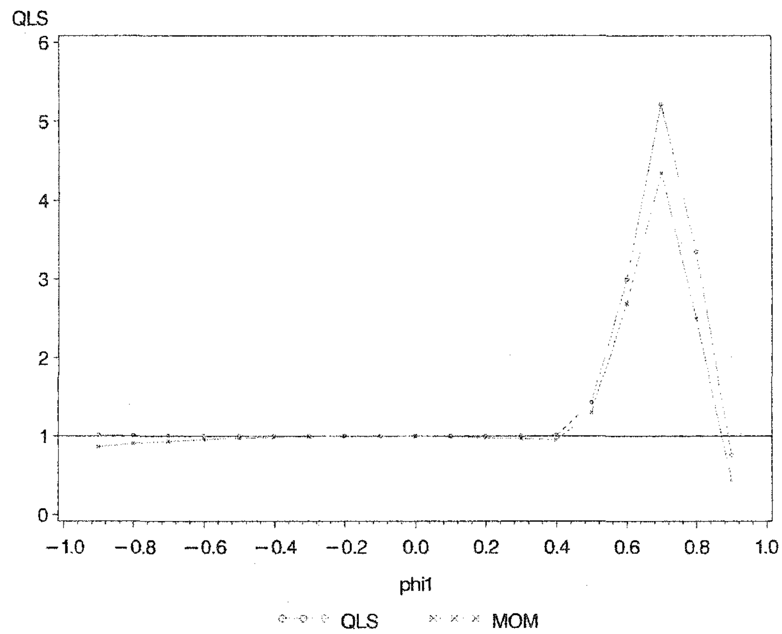

Figure 3.17. AREs of $\hat{\phi}_{1m}$ and $\hat{\phi}_{1q}$ vs $\hat{\phi}_{1l}$ when the data is simulated from contaminated normal $0.5N(0,4) + 0.5N(1,4)$ and $\phi_2 = 0$.

Table 3.9. Regression analysis of a dental study

Parameter	Moment	Quasi-least squares	Maximum likelihood
β_g	17.4041	17.4132	17.4046
β_b	16.2600	16.2216	16.2581
γ_g	0.4765	0.4757	0.4765
γ_b	0.7951	0.7979	0.7953
ϕ_1	0.3139	0.2569	0.3135
ϕ_2	0.4869	0.5474	0.4924
σ_1^2	2.3249	2.2824	2.3100
σ_2^2	2.3249	2.2824	2.3100

Estimates of the parameters using moment, maximum likelihood and quasi-least squares methods.

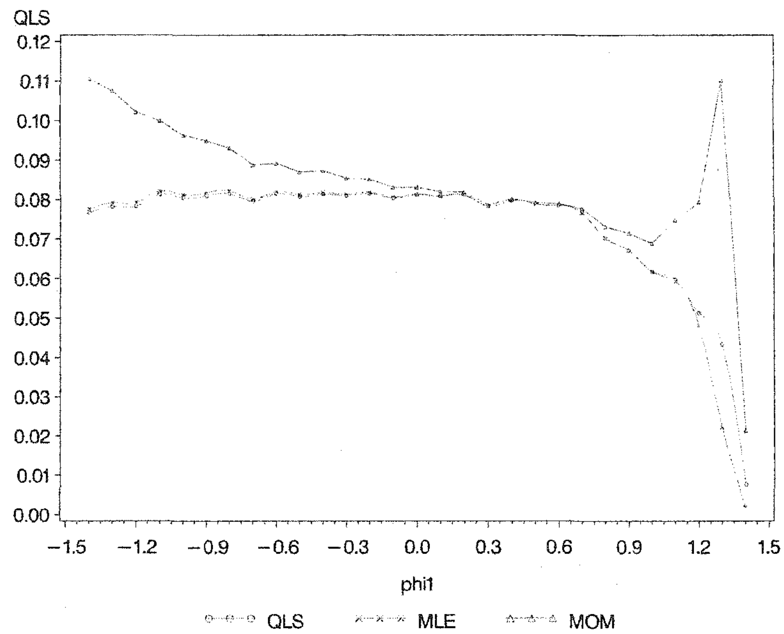

Figure 3.18. Bias of $\hat{\phi}_{1m}$, $\hat{\phi}_{1q}$ *and* $\hat{\phi}_{1l}$ *when the data is simulated from contaminated normal* $0.5N(0,4) + 0.5N(1,4)$ *and* $\phi_2 = -0.5$.

for each group, even if it is AR(1).

3.8 Summary

In this chapter, we studied the three methods of estimation, MOM, MLE and QLS when the errors in the time series regression model follow an AR(2) process. Through simulations, we showed that the quasi-least squares estimates may be better than the MLE's if the distribution is contaminated normal. The bias analysis also shows that the quasi-least squares estimates are good competitors to the maximum likelihood estimates. An application to a dental study is presented again but with an AR(2) error. The corresponding standard errors are not in a simple closed form, except for the maximum likelihood estimates. The asymptotic distribution of

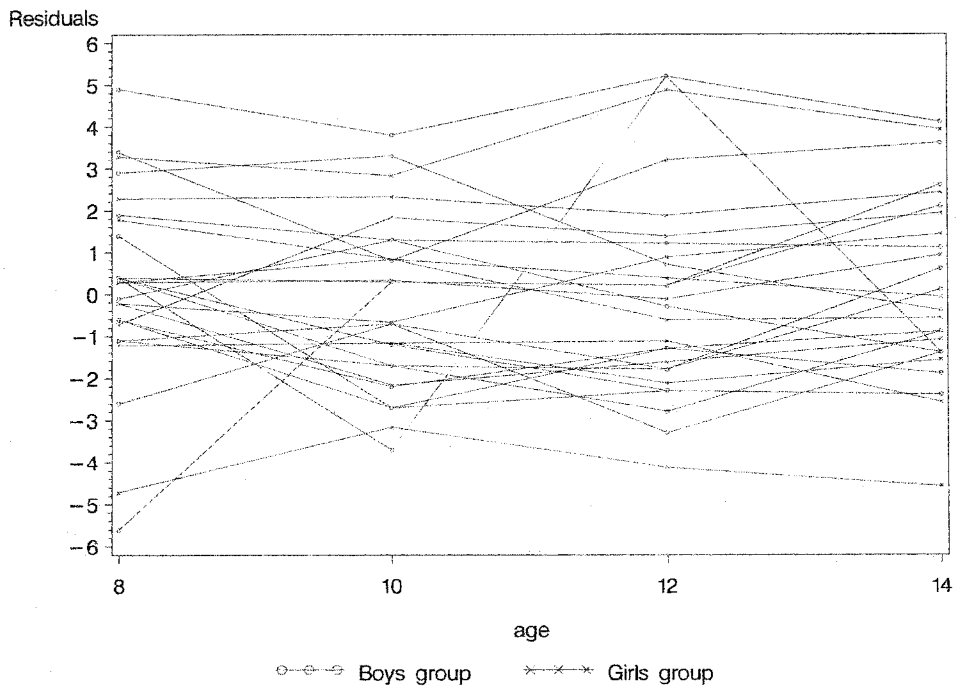

Figure 3.19. Residual plot based on the quasi-least estimates, assuming the error is an AR(2) process.

the maximum likelihood estimates are obtained by finding the information matrix.

CHAPTER IV

GENERALIZATION OF THE ESTIMATING METHODS

In Chapter IV, we extend the results in Chapters II and III to two important cases, the time series regression model with autoregressive of order p (AR(p)) and the moving average of order one (MA(1)) errors. The chapter is organized as follows: In Section 4.1, we present the model with AR(p) correlation structure and its properties. We discuss the GEE method and its variations, moment method, maximum likelihood method and quasi-least squares method, and the inter relationships between the estimating procedures. We show that the GEE method reduces to either the moment or maximum likelihood method in some cases. In this section, we also demonstrate the robustness of the quasi-least squares estimates when the distribution of the data is contaminated normal. In Section 4.2, we study the properties of the model with MA(1) errors. We discuss the moment, maximum likelihood and quasi-least squares methods. Simulation results are presented to show that the quasi-least squares estimates are good competitors to the maximum likelihood estimates.

4.1 The model with AR(p) errors

4.1.1 Model

In a seminal paper, Anderson (1978) introduced and studied some basic properties of the repeated measurements on autoregressive processes. The approximate maximum likelihood method for those processes was studied by Harvey and Phillips (1979), Laird, *et al.* (1987), Rochon and Helms (1989) and Rochon (1992). Laird, *et al.* (1987) also discussed the maximum likelihood and restricted maximum likelihood estimation procedures for the ARMA(p, q) model using the EM algorithm. Recently, Cheang and Reinsel (2000) discussed the bias reduction for the restricted maximum likelihood estimates for the AR(p) model.

Recall from (1.1.2) the autoregressive of order p errors satisfy the relation

$$\varepsilon_j = \phi_1 \varepsilon_{j-1} + \phi_2 \varepsilon_{j-2} + \cdots + \phi_p \varepsilon_{j-p} + a_j. \tag{4.1.1}$$

The stationary condition requires that the roots of $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = 0$ lie outside the unit circle, where B is the backward operator. Using (1.1.5), we can check that the autocorrelation function satisfies the p^{th} order difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k \geq 1 \tag{4.1.2}$$

with $\rho_0 = 1$. Note that this is analogous to the difference equation satisfied by the process $\{\varepsilon_j\}$ itself. By putting $k = 1, 2, \ldots, p$, we obtain the *Yule-Walker* equations

$$\begin{array}{ccccccccc} \rho_1 & = & \phi_1 & + & \phi_2 \rho_1 & + & \cdots & + & \phi_p \rho_{p-1} \\ \rho_2 & = & \phi_1 \rho_1 & + & \phi_2 & + & \cdots & + & \phi_p \rho_{p-2} \\ \vdots & & \vdots & & \vdots & & & & \vdots \\ \rho_p & = & \phi_1 \rho_{p-1} & + & \phi_2 \rho_{p-2} & + & \cdots & + & \phi_p \,. \end{array} \tag{4.1.3}$$

Say $\boldsymbol{\phi} = (\phi_1, \ \phi_2, \ \ldots, \ \phi_p)'$, $\boldsymbol{\rho} = (\rho_1, \ \rho_2, \ \ldots, \ \rho_p)'$, $\boldsymbol{\phi}$ and $\boldsymbol{\rho}$ have the following relationship

$$\boldsymbol{\phi} = \mathbf{P}_p^{-1} \boldsymbol{\rho}, \tag{4.1.4}$$

where $\mathbf{P}_p$ is the $p \times p$ autocorrelation matrix $\{\rho_{|i-j|}\}$ (Note the only difference between $\mathbf{P}_p$ and $\mathbf{P}_i(\boldsymbol{\phi})$ is the dimension) and

$$\mathbf{P}_p = \mathbf{I}_p + 2 \sum_{k=1}^{p-1} \mathbf{C}_{k0}^p \rho_k,$$

where $\mathbf{I}_p$ is the p-dimensional identity matrix, $2\,\mathbf{C}_{k0}^p$ is the $p \times p$ matrix with 1's on the k^{th} off diagonal and 0's elsewhere (Note the only difference between $\mathbf{C}_{k0}^p$ and $\mathbf{C}_{k0}$ is the dimension). Thus $\boldsymbol{\rho}$ can be also written in terms of $\boldsymbol{\phi}$ as

$$\boldsymbol{\rho} = (\mathbf{I}_p - 2\sum_{k=1}^{p-1} \mathbf{C}_{k0}^p \boldsymbol{\phi} \boldsymbol{\epsilon}_k')^{-1} \boldsymbol{\phi}, \tag{4.1.5}$$

where $\boldsymbol{\epsilon}_k$ is the $p \times 1$ unit vector with 1's in the k^{th} position and 0's elsewhere, $k = 1, 2, \ldots, p-1$. Expression (4.1.4) also implies that

$$\begin{aligned} \boldsymbol{\phi}'\boldsymbol{\rho} &= \boldsymbol{\phi}'\mathbf{P}_p\boldsymbol{\phi} = \boldsymbol{\rho}'\mathbf{P}_p^{-1}\boldsymbol{\rho}, \\ \text{and} \qquad \mathbf{r} &= \boldsymbol{\Gamma}\boldsymbol{\phi}, \end{aligned}$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_p)'$ is the sample autocorrelation vector, and $\boldsymbol{\Gamma}$ is the covariance matrix of $\boldsymbol{\varepsilon}_i$. According to equation (1.1.4), the variance of $\{\varepsilon_j\}$ is

$$\sigma_\varepsilon^2 = \frac{1}{\Delta(\boldsymbol{\phi})}\,\sigma^2, \tag{4.1.6}$$

where $\Delta(\boldsymbol{\phi}) = 1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p = 1 - \boldsymbol{\phi}'\boldsymbol{\rho}$. From (1.1.6), the covariance matrix of $\{\varepsilon_j\}$ made at t_i ($t_i > 2p$) successive times is $\sigma^2\mathbf{V}_i(\boldsymbol{\phi})$, and

$$\mathbf{V}_i(\boldsymbol{\phi}) = \frac{1}{\Delta(\boldsymbol{\phi})}\,\mathbf{P}_i(\boldsymbol{\phi}), \tag{4.1.7}$$

where $\mathbf{P}_i(\boldsymbol{\phi})$ is t_i-dimensional as usual. Following Cheang and Reinsel (2000), the model in matrix notation can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad \mathbf{L}_i\boldsymbol{\varepsilon}_i = \mathbf{a}_i, \quad i = 1, \ldots, n,$$

where $\mathbf{a}_i = (a_1, a_2, \ldots, a_{t_i})'$ and $\mathbf{L}_i$ is the Cholesky decomposition of $\mathbf{V}_i^{-1}$, that is $\mathbf{V}_i^{-1} = \mathbf{L}_i'\,\mathbf{L}_i$. Specifically, $\mathbf{L}_i$ is a lower triangular matrix with its first p diagonal elements equal to $\Delta^{1/2}, (\Delta/\Delta_1)^{1/2}, \ldots, (\Delta/\Delta_{p-1})^{1/2}$, its remaining diagonal elements equal to 1, elements in the $(l,j)^{th}$ position equal to $-\phi^*_{l-j,l-1}$ for $j = 1, \ldots, l-1$ and $l = 2, \ldots, p$, equal to $-\phi_{l-j}$ for $j = l-p, \ldots, l-1$ when $l > p$, and 0 elsewhere. In the first p rows of $\mathbf{L}_i$, the elements $\phi^*_{lk} = \phi_{lk}(\Delta/\Delta_k)^{1/2}, l = 1, \ldots, k$, where $\phi_{1k}, \ldots, \phi_{kk}$, for $k = 1, \ldots, p-1$, are solutions for coefficients $\phi_1, \cdots, \phi_k$ in the system of the first k *Yule-Walker* equations in (4.1.3) with p set equal to k, and $\Delta_k = 1 - \phi_{1k}\rho_1 - \phi_{2k}\rho_2 - \cdots - \phi_{kk}\rho_k$. For example, $\phi_{11} = r_1/r_0 = \rho_1$ and $\Delta_1 = 1 - \rho_1^2$. Thus, the determinate of $\mathbf{V}_i^{-1}$ is given by

$$|\mathbf{V}_i^{-1}| = |\mathbf{L}_i|^2,$$

which will not depend on t_i as shown in the next section. The inverse of $\mathbf{V}_i(\boldsymbol{\phi})$ can also be written

$$\mathbf{V}_i^{-1}(\boldsymbol{\phi}) = \mathbf{I} + \sum_{k=1}^{p}\sum_{l=1}^{p} \phi_k\phi_l \mathbf{C}_{ikl} - 2\sum_{k=1}^{p} \phi_k \mathbf{C}_{ik0}, \tag{4.1.8}$$

where $\mathbf{C}_{ilk} = \mathbf{C}_{ikl}$ is t_i-dimensional and for $k > l$, $2\mathbf{C}_{kl}$ has $(t_i - k - l)$ ones on the $(k-l)^{th}$ diagonals above and below the main diagonal, excluding the first and last l elements on these diagonals, and 0's elsewhere, $k = 1, \ldots, p$, $l = 0, 1, \ldots, p$, $\mathbf{C}_{ikk}$ has $(t_i - 2k)$ ones on the main diagonal, excluding the first and last k elements, $k = 1, \ldots, p$. This implies

$$\frac{\partial \mathbf{V}_i^{-1}(\boldsymbol{\phi})}{\partial \phi_k} = 2\left(\sum_{l=1}^{p} \phi_l \mathbf{C}_{ikl} - \mathbf{C}_{ik0}\right), \quad k = 1, \ldots, p. \tag{4.1.9}$$

Thus, the adjusted sum of square errors $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ defined in (1.2.8) simplifies to

$$\begin{aligned} S(\boldsymbol{\beta}, \boldsymbol{\phi}) &= \sum_{i=1}^{n} \mathbf{a}_i'\mathbf{a}_i = \sum_{i=1}^{n} \operatorname{tr}\left[\mathbf{V}_i^{-1}(\boldsymbol{\phi})\mathbf{U}_i(\boldsymbol{\beta})\right] \\ &= n \cdot \left(c_{00} - 2\boldsymbol{\phi}'\mathbf{c}_0 + \boldsymbol{\phi}'\mathbf{C}\boldsymbol{\phi}\right), \end{aligned} \tag{4.1.10}$$

where $\mathbf{c}_0 = \{c_{k0}\}$, $\mathbf{C} = \{c_{kl}\}$, $k, l = 1, \ldots, p$ and

$$c_{kl} = \frac{1}{n}\sum_{i=1}^{n} \operatorname{tr}(\mathbf{C}_{ikl}\mathbf{U}_i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=l+1}^{t_i-k} \varepsilon_{ij}\varepsilon_{i(j+k-l)}, \quad k = 0, \ldots, p, \; l = 0, \ldots, p.$$

4.1.2 Generalized estimation equation estimates

In recent years the most popular approach to the analysis of repeated measurements or clustered data has been the generalized estimating equations (GEE). Following Liang *et al.* (1992), the original approach by Liang and Zeger (1986) will be referred to as GEE1. A variation or an alternative approach (GEE2) has been suggested by Prentice and Zhao (1991). Hall and Severini (1998) proposed "extended generalized estimating equations" (EGEE). We will show that in some cases when the covariance follows an AR(p) model, the GEE methods are equivalent to either moment or maximum likelihood method. In fact, this is true for any ARMA (p, q) model.

GEE1

For the estimation of the nuisance parameters $\boldsymbol{\phi}$ and σ^2, Liang and Zeger proposed method of moment estimators based on the residuals $\hat{\varepsilon}_i = \mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$. Prentice and Zhao (1991) generalized the method of moment approach by suggesting *ad hoc* estimating equations for $\boldsymbol{\phi}$. Based on the estimate of $\boldsymbol{\phi}$, Liang and Zeger proposed the estimate of $\boldsymbol{\beta}$ by solving a quasi-score equation following a quasi-likelihood approach:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \mathbf{0},$$

which implies

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i\,. \tag{4.1.11}$$

Since we assume that the covariance matrix follows an AR(p) model, there is no "misspecification" issue. Under mild regularity conditions, the estimates of $(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)$ can be shown to be consistent, and the asymptotic covariance matrix of

$\hat{\boldsymbol{\beta}}$ is given by (1.2.11).

GEE2

In GEE2 and EGEE methods, the parameter σ^2 is included in $\boldsymbol{\phi}$ and all are estimated by a combined estimating equation. Based on the assumption that $\mathbf{Y}_i$ follows a quadratic exponential model (which implies forms for the third and fourth moments), the *score equation* for $(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)$ is

$$\sum_{i=1}^{n} \mathbf{D}_i' \Sigma_i^{-1} f_i = \mathbf{0}, \tag{4.1.12}$$

where

$$f_i = \begin{pmatrix} y_i - \mathbf{X}_i\boldsymbol{\beta} \\ \mathrm{s}_i - \boldsymbol{\sigma}_i \end{pmatrix}$$

with

$$\begin{aligned} \mathrm{s}_i &= \mathrm{vec}[(y_i - \mathbf{X}_i\boldsymbol{\beta})(y_i - \mathbf{X}_i\boldsymbol{\beta})'] = \mathrm{vec}(\mathbf{U}_i), \\ \boldsymbol{\sigma}_i &= \sigma^2 \mathrm{vec}(\mathbf{V}_i), \end{aligned}$$

and $\mathbf{D}_i$ is the derivative matrix respect to $(\boldsymbol{\beta}', \boldsymbol{\phi}', \sigma^2)$

$$\mathbf{D}_i = \frac{\partial \begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta} \\ \boldsymbol{\sigma}_i \end{pmatrix}}{\partial(\boldsymbol{\beta}', \boldsymbol{\phi}', \sigma^2)} = \begin{pmatrix} \mathbf{X}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \left(\partial \mathrm{vec}(\mathbf{V}_i)/\partial\boldsymbol{\phi}'\right) & \mathrm{vec}(\mathbf{V}_i) \end{pmatrix},$$

and Σ_i is the dispersion of $(y_i, s_i)'$

$$\Sigma_i = \mathrm{Var}\begin{pmatrix} y_i \\ \mathrm{s}_i \end{pmatrix} = \begin{pmatrix} \sigma^2\mathbf{V}_i & \mathrm{Cov}(y_i, \mathrm{s}_i) \\ \mathrm{Cov}(\mathrm{s}_i, y_i) & \mathrm{Var}(\mathrm{s}_i) \end{pmatrix}.$$

If the distribution of the data is symmetric (ex: normal data) then $\mathrm{Cov}(y_i, \mathrm{s}_i) = 0$. Equation (4.1.12) becomes

$$\sum_{i=1}^{n} \begin{pmatrix} \sigma^2\mathbf{X}_i\mathbf{V}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \sigma^2\left(\partial \mathrm{vec}(\mathbf{V}_i)/\partial\boldsymbol{\phi}'\right)' \\ \mathbf{0}' & \mathrm{vec}(\mathbf{V}_i)' \,\mathrm{Var}(\mathrm{s}_i)^{-1} \end{pmatrix} \begin{pmatrix} y_i - \mathbf{X}_i\boldsymbol{\beta} \\ \mathrm{vec}(\mathbf{U}_i) - \sigma^2\mathrm{vec}(\mathbf{V}_i) \end{pmatrix} = \mathbf{0},$$

which can be rewritten as,

$$\begin{aligned}
\sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta) &= \mathbf{0}, && (4.1.13)\\
\sum_{i=1}^{n} \left(\frac{\partial \mathrm{vec}(\mathbf{V}_i)}{\partial \phi'}\right)' \mathrm{Var}(\mathbf{s}_i)^{-1}\left[\mathrm{vec}(\mathbf{U}_i) - \sigma^2 \mathrm{vec}(\mathbf{V}_i)\right] &= \mathbf{0}, && (4.1.14)\\
\sum_{i=1}^{n} \mathrm{vec}(\mathbf{V}_i)' \,\mathrm{Var}(\mathbf{s}_i)^{-1}\left[\mathrm{vec}(\mathbf{U}_i) - \sigma^2 \mathrm{vec}(\mathbf{V}_i)\right] &= 0. && (4.1.15)
\end{aligned}$$

Equation (4.1.13) automatically implies (4.1.11), and equation (4.1.15) implies

$$\sigma^2 = \frac{\sum_{i=1}^n \mathrm{vec}(\mathbf{V}_i)'\mathrm{Var}(\mathbf{s}_i)^{-1}\mathrm{vec}(\mathbf{U}_i)}{\sum_{i=1}^n \mathrm{vec}(\mathbf{V}_i)'\mathrm{Var}(\mathbf{s}_i)^{-1}\mathrm{vec}(\mathbf{V}_i)} \,. \qquad (4.1.16)$$

Note that if we know the third and forth moments we should be able to find the specific form of $\mathrm{Var}(\mathbf{s}_i)$ involved in (4.1.14) and (4.1.16). The preceding expression (4.1.16) can be further simplified if we assume that the data is from a normal distribution, because then we have

$$\mathrm{Var}(\mathbf{s}_i) = 2\sigma^4\, \mathbf{V}_i \otimes \mathbf{V}_i \,,$$

where '$\otimes$' is the Kronecker product. Since

$$\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi_k} = -\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \phi_k}\mathbf{V}_i^{-1}, \qquad k = 1, \ldots, p,$$

which implies

$$\begin{aligned}
\frac{\partial \mathrm{vec}(\mathbf{V}_i^{-1})'}{\partial \phi_k} &= -\mathbf{V}_i^{-1}\frac{\partial \mathrm{vec}(\mathbf{V}_i)'}{\partial \phi_k}\mathbf{V}_i^{-1}\\
&= -\frac{\partial \mathrm{vec}(\mathbf{V}_i)'}{\partial \phi_k}\left(\mathbf{V}_i \otimes \mathbf{V}_i\right)^{-1}, \qquad k = 1, \ldots, p,
\end{aligned}$$

equation (4.1.14) becomes,

$$\sum_{i=1}^{n} \left(\frac{\partial \mathrm{vec}(\mathbf{V}_i)}{\partial \phi'}\right)' \left(2\sigma^4\, \mathbf{V}_i \otimes \mathbf{V}_i\right)^{-1}\left[\mathrm{vec}(\mathbf{U}_i) - \sigma^2 \mathrm{vec}(\mathbf{V}_i)\right] = \mathbf{0},$$

that is,

$$\frac{1}{2\sigma^4}\sum_{i=1}^{n} \left(\frac{\partial \mathrm{vec}(\mathbf{V}_i^{-1})}{\partial \phi'}\right)' \left[\mathrm{vec}(\mathbf{U}_i) - \sigma^2 \mathrm{vec}(\mathbf{V}_i)\right] = \mathbf{0},$$

and can be further rewritten as

$$\sum_{i=1}^{n} \frac{\partial \mathrm{tr}(\mathbf{V}_i^{-1}\mathbf{U}_i)}{\partial \boldsymbol{\phi}} - \sigma^2 \sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\phi}} \mathbf{V}_i\right) = \mathbf{0}, \tag{4.1.17}$$

where $\mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\phi}} \mathbf{V}_i\right) = \left[\mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi_1} \mathbf{V}_i\right), \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi_2} \mathbf{V}_i\right), \ldots, \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi_p} \mathbf{V}_i\right)\right]'$. Equation (4.1.17) can be shown in the next section to be the ML equation for $\boldsymbol{\phi}$ under normal assumption. Equation (4.1.16) can also be reduced to

$$\begin{aligned}
\sigma^2 &= \frac{\sum_{i=1}^n \mathrm{vec}(\mathbf{V}_i)'(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\mathrm{vec}(\mathbf{U}_i)}{\sum_{i=1}^n \mathrm{vec}(\mathbf{V}_i)'(\mathbf{V}_i^{-1} \otimes \mathbf{V}_i^{-1})\mathrm{vec}(\mathbf{V}_i)} \\
&= \frac{\sum_{i=1}^n \mathrm{tr}\left(\mathbf{V}_i\mathbf{V}_i^{-1}\mathbf{U}_i\mathbf{V}_i^{-1}\right)}{\sum_{i=1}^n \mathrm{tr}\left(\mathbf{V}_i\mathbf{V}_i^{-1}\mathbf{V}_i\mathbf{V}_i^{-1}\right)} \\
&= \frac{\sum_{i=1}^n \mathrm{tr}\left(\mathbf{U}_i\mathbf{V}_i^{-1}\right)}{\sum_{i=1}^n \mathrm{tr}(\mathbf{I}_{t_i})} \\
&= \frac{1}{t}\left(c_{00} - 2\boldsymbol{\phi}'\mathbf{c}_0 + \boldsymbol{\phi}'\mathbf{C}\boldsymbol{\phi}\right).
\end{aligned} \tag{4.1.18}$$

Here we used the result

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}).$$

In the next section we will show that equation (4.1.18) is equivalent to the ML equation for σ^2 under normal assumption. Thus equations (4.1.13), (4.1.17), and (4.1.18) which are the *quasi-score equations* for $GEE2$ under the normal assumption, are exactly the ML equations.

EGEE

The *score equation* for $EGEE$ model is illustrated as below. First we use *index notation* (see for example, McCullagh, 1987) where subscripts and superscripts indicate vector or matrix components and a summation is implied over any index which is repeated in a subscript and a superscript. In our notion v^{jk} is the $(j,k)^{th}$ element of $\mathbf{V}_i$, v_{jk} is the $(j,k)^{th}$ element of $\mathbf{V}_i^{-1}$. The partial derivative with respect to the u^{th} element of $\boldsymbol{\phi}$ will be denoted by ϕ^u. The partial derivatives of the resulting extended *quasi-likelihood* function with respect to the components of $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ and σ^2 give the following estimating functions:

$$\begin{aligned} U(\boldsymbol{\xi};\beta^b) &= (y^j - \mathbf{X}^j\boldsymbol{\beta})\, v_{ja}\, (\partial(\mathbf{X}^a\boldsymbol{\beta})/\partial\beta^b), \qquad b = 1,\ldots,r \\ U(\boldsymbol{\xi};\phi^u) &= -(y^j - \mathbf{X}^j\boldsymbol{\beta})(y^k - \mathbf{X}^k\boldsymbol{\beta})v^u_{jk} + v^{jk}v^u_{jk}, \quad u = 1,\ldots,p \\ U(\boldsymbol{\xi};\sigma^2) &= -(y^j - \mathbf{X}^j\boldsymbol{\beta})(y^k - \mathbf{X}^k\boldsymbol{\beta})(-v_{jk}/\sigma^4) + v^{jk}(-v_{jk}/\sigma^4), \end{aligned}$$

where $\boldsymbol{\xi} = (\boldsymbol{\beta}', \boldsymbol{\phi}', \sigma^2)'$. Stacking these estimating functions and summing over independent subjects yields the EGGE for $\boldsymbol{\xi}$:

$$\sum_{i=1}^{n} \begin{bmatrix} U_i(\boldsymbol{\xi};\boldsymbol{\beta}) \\ U_i(\boldsymbol{\xi};\boldsymbol{\phi}) \\ U_i(\boldsymbol{\xi};\sigma^2) \end{bmatrix} = \mathbf{0}.$$

In *matrix notation*, these can be written as:

$$\sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \mathbf{0}, \qquad (4.1.19)$$

$$\sum_{i=1}^{n} \left(\frac{\partial \mathrm{vec}(\mathbf{V}_i^{-1})}{\partial \boldsymbol{\phi}'}\right)' (\mathbf{s}_i - \boldsymbol{\sigma}_i) = \mathbf{0}, \qquad (4.1.20)$$

$$\sum_{i=1}^{n} \mathrm{vec}(\mathbf{V}_i^{-1})'(\mathbf{s}_i - \boldsymbol{\sigma}_i) = 0. \qquad (4.1.21)$$

Equation (4.1.19) is exactly the same as (4.1.13) and gives (4.1.11). Equation (4.1.20) and (4.1.21) can be further rewritten as

$$\begin{aligned} \sum_{i=1}^{n} \frac{\partial \mathrm{tr}(\mathbf{V}_i^{-1}\mathbf{U}_i)}{\partial \boldsymbol{\phi}} - \sigma^2 \sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\phi}} \mathbf{V}_i\right) &= \mathbf{0}, \\ \sum_{i=1}^{n} \mathrm{tr}(\mathbf{U}_i\mathbf{V}_i^{-1}) - \sigma^2 \sum_{i=1}^{n} \mathrm{tr}(\mathbf{I}_{t_i}) &= 0, \end{aligned}$$

which are exactly the same as (4.1.17) and (4.1.18).

Relationship with generalized estimating equations

It can easily be seen that in EGEE method, the estimating equation for $\boldsymbol{\beta}$ corresponding to equation (4.1.19) is exactly the same as equation (4.1.2). Therefore, as function of an estimator for $\boldsymbol{\phi}$, the GEE1 and EGEE estimators of $\boldsymbol{\beta}$ are the same; if the distribution of the data is symmetric, GEE2 methods also gives the same functional estimate of $\boldsymbol{\beta}$. Furthermore, if the data is normal, GEE2 method and EGEE method are equivalent, and both give maximum likelihood estimates. It is interesting to note that this relationship is always true regardless of the covariance structure. Thus the GEE1 method is equivalent to the moment method under the ARMA (p, q) covariance structure. Therefore, it is sufficient to study the moment and maximum likelihood methods.

4.1.3 Moment and maximum likelihood estimates

According to Section 1.2, the generalized least squares (GLS) estimates of $\boldsymbol{\beta}$ and σ^2 may be obtained as in (1.2.10) and (1.2.12), respectively, and can be simplified as

$$\hat{\boldsymbol{\beta}}_g = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}(\boldsymbol{\phi})\mathbf{X}_i\right)^{-1} \cdot \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}(\boldsymbol{\phi})\mathbf{y}_i \qquad (4.1.22)$$

$$\hat{\sigma}_g^2 = \frac{1}{t}\left(c_{00} - 2\boldsymbol{\phi}'\mathbf{c}_0 + \boldsymbol{\phi}'\mathbf{C}\boldsymbol{\phi}\right). \qquad (4.1.23)$$

Given the "residuals" from the model, the moment estimate of $\boldsymbol{\phi}$ may be obtained by setting $\boldsymbol{\Gamma} = \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'$. For the AR(p) errors case, the moment estimates are the same as *Yule-Walker* estimates. The *Yule-Walker* method starts by estimating $\boldsymbol{\phi}$ from the sample autocorrelation function of the GLS residuals using the *Yule-Walker* equations. Next, $\mathbf{V}_i(\boldsymbol{\phi})$ is estimated from the estimate of $\boldsymbol{\phi}$. The autocorrelation corrected estimate of the regression parameter $\boldsymbol{\beta}$ is then computed by (4.1.22) using the estimated matrix $\mathbf{V}_i(\hat{\boldsymbol{\phi}})$. Finally, the variance σ^2 is estimated by (4.1.23) plugging in the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$.

Define $\hat{c}^*_{k0} = \hat{c}_{k0}/(\bar{t}-k)$, $k = 0, 1, \ldots, p$, where '^' means evaluating based on the residuals, then $\hat{c}^*_{k0}$ is the k^{th} lag sample covariance and could be an estimate of the k^{th} lag autocovariance. We then obtain the following relations expressed in terms of the estimated autocorrelations $r_k = \hat{c}^*_{k0}/\hat{c}^*_{00}$, $k = 1, \ldots, p-1$,

$$\begin{array}{ccccccccc}
r_1 & = & \hat{\phi}_1 & + & \hat{\phi}_2 r_1 & + & \cdots & + & \hat{\phi}_p r_{p-1} \\
r_2 & = & \hat{\phi}_1 r_1 & + & \hat{\phi}_2 & + & \cdots & + & \hat{\phi}_p r_{p-2} \\
\vdots & & \vdots & & \vdots & & & & \vdots \\
r_p & = & \hat{\phi}_1 r_{p-1} & + & \hat{\phi}_2 r_{p-2} & + & \cdots & + & \hat{\phi}_p \, .
\end{array}$$

These are well-known *Yule-Walker* equations. In matrix notation, the estimate of $\boldsymbol{\phi}$ is obtained by

$$\hat{\boldsymbol{\phi}} = \mathbf{R}^{-1}\mathbf{r} = \hat{\mathbf{C}}_0^{*-1}\hat{\mathbf{c}}_0^*, \tag{4.1.24}$$

where $\mathbf{R}$ and $\hat{\mathbf{C}}_0^*$ are the $p \times p$ sample autocorrelation and sample covariance matrix with the $(k-1)^{th}$ diagonal elements r_k and $\hat{c}^*_{k0}$ respectively, $\mathbf{r} = (r_1, r_2, \ldots, r_p)'$, and $\hat{\mathbf{c}}_0^* = (\hat{c}^*_{10}, \hat{c}^*_{20}, \ldots, \hat{c}^*_{p0})'$. The moment estimates of $(\boldsymbol{\beta}, \boldsymbol{\phi})$ are the simultaneous solutions of (4.1.22) and (4.1.24) and the moment estimate of σ^2 is (4.1.23) plugging the estimates of $(\boldsymbol{\beta}, \boldsymbol{\phi})$. Notice that this is just an ad-hoc estimate, but it has been widely used since it is easy to apply and gives us a quick answer. It also serves as the preliminary estimates for other estimating methods.

If we approximate the sample variance σ^2_ε by $\hat{c}_{00}/\bar{t}$, the sample correlation vector $\mathbf{r}$ by $\hat{\mathbf{c}}_0/\hat{c}_{00}$ and the sample correlation matrix $\mathbf{R}$ by $\hat{\mathbf{C}}/\hat{c}_{00}$, where $\hat{\mathbf{c}}_0$ and $\hat{\mathbf{C}}$ are as defined in (4.1.10) evaluating based on residuals, this is true if $\bar{t}$ is large, the estimate of $\boldsymbol{\phi}$ reduces to $\hat{\boldsymbol{\phi}} = \mathbf{C}^{-1}\mathbf{c}_0$ and (4.1.23) becomes

$$\hat{\sigma}^2 \approx \hat{\sigma}^2_\varepsilon(1 - 2\hat{\boldsymbol{\phi}}'\mathbf{r} + \hat{\boldsymbol{\phi}}'\mathbf{R}\hat{\boldsymbol{\phi}})$$

$$= \hat{\sigma}_\varepsilon^2(1 - \hat{\boldsymbol{\phi}}'\mathbf{r}),$$

which agrees with the population result

$$\sigma^2 = \sigma_\varepsilon^2(1 - \boldsymbol{\phi}'\boldsymbol{\rho}).$$

In order to obtain the maximum likelihood estimates, we need to derive the ML equations. Recalling the log-likelihood function (1.2.14) and the estimate of σ^2 given by (4.1.23), we get

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\phi}) &= -\frac{n\bar{t}}{2}\left[\log(2\pi) - \log(n\bar{t}) + 1\right] - \frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{V}_i(\boldsymbol{\phi})| - \frac{n\bar{t}}{2}\log S(\boldsymbol{\beta}, \boldsymbol{\phi}) \\ &= \text{Const.} - \frac{n\bar{t}}{2}\log\left(\sum_{i=1}^{n}|\mathbf{L}_i|^{-1/\bar{t}}\mathbf{a}_i'\mathbf{a}_i|\mathbf{L}_i|^{-1/\bar{t}}\right). \end{aligned}$$

Thus, the ULS and QLS methods are to minimize $\sum_{i=1}^{n}\mathbf{a}_i'\mathbf{a}_i$, while the ML method is to minimize $\sum_{i=1}^{n}\tilde{\mathbf{a}}_i'\tilde{\mathbf{a}}_i$, where $\tilde{\mathbf{a}}_i = \mathbf{a}_i|\mathbf{L}_i|^{-1/\bar{t}}$. Now, equating the partial derivative of the log-likelihood function with respect to $(\boldsymbol{\beta}, \sigma^2)$ we get (4.1.22) and (4.1.23), respectively. Taking the partial derivative with respect to $\boldsymbol{\phi}$ we get

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\phi}} &= -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial \log|\mathbf{V}_i(\boldsymbol{\phi})|}{\partial \boldsymbol{\phi}} - \frac{1}{2\sigma^2}\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \\ &= -\frac{n}{\sigma^2}\left[\frac{\sigma^2}{2n}\sum_{i=1}^{n}\frac{\partial \log|\mathbf{V}_i(\boldsymbol{\phi})|}{\partial \boldsymbol{\phi}} + (\mathbf{C}\boldsymbol{\phi} - \mathbf{c}_0)\right]. \end{aligned} \qquad (4.1.25)$$

According to Cheang and Reinsel (2000) (p.1174 Eq. (8)),

$$\frac{\sigma^2}{2}\frac{\partial \log|\mathbf{V}_i(\boldsymbol{\phi})|}{\partial \boldsymbol{\phi}} = \boldsymbol{\Gamma}_p(\boldsymbol{\phi}, \sigma^2)\mathbf{D}\boldsymbol{\phi},$$

where $\boldsymbol{\Gamma}_p(\boldsymbol{\phi}, \sigma^2)$ is the $p \times p$ autocovariance matrix with the k^{th} diagonal elements γ_k, and $\mathbf{D}$ is a diagonal matrix with $(1, 2, \ldots, p)$ on the main diagonal and 0's elsewhere. Equating equation (4.1.25) to $\mathbf{0}$, we obtain the ML equations for $\boldsymbol{\phi}$

$$\boldsymbol{\Gamma}_p(\boldsymbol{\phi}, \sigma^2)\mathbf{D}\boldsymbol{\phi} + \mathbf{C}\boldsymbol{\phi} - \mathbf{c}_0 = \mathbf{0}. \qquad (4.1.26)$$

Thus, the MLEs of $(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)$ are the simultaneous solutions of (4.1.22), (4.1.26) and (4.1.23). It is noted that the exact MLEs are not in closed forms since (4.1.26) is hard to solve. Box *et al.* (1994, p.300) suggested an approximate MLE for the estimation of $\boldsymbol{\phi}$. By using $\hat{c}^*_{kl} = \hat{c}_{kl}/(\bar{t}-k-l)$ as an "estimate" of $\gamma_{k-l}, k, l = 1, \ldots, p$, that is $\boldsymbol{\Gamma}_p \approx \hat{\mathbf{C}}^* = \{\hat{c}^*_{kl}\}$, note that γ_{k-l} has several different estimates in this sense, the ML equation (4.1.26) becomes linear in $\boldsymbol{\phi}$ after plugging in the estimate of $\boldsymbol{\beta}$

$$\hat{\mathbf{C}}^* \boldsymbol{\phi} = \hat{\mathbf{c}}^*_0, \tag{4.1.27}$$

which yields

$$\hat{\boldsymbol{\phi}}_{al} = \hat{\mathbf{C}}^{*-1} \hat{\mathbf{c}}^*_0, \tag{4.1.28}$$

where '*al*' stands for approximate MLE. Note that this is close to the moment estimate (4.1.24). Actually, $\hat{\mathbf{C}}^*_0$ is a better approximation of $\boldsymbol{\Gamma}_p$ comparing to $\hat{\mathbf{C}}^*$, and since $E(\hat{\mathbf{C}}^*) = E(\hat{\mathbf{C}}^*_0)$, $\hat{\mathbf{C}}^* \approx \hat{\mathbf{C}}^*_0$; hence if this is true, the approximate MLE reduces to the moment estimate by replacing $\hat{\mathbf{C}}^*$ with $\hat{\mathbf{C}}^*_0$. In this sense, the approximate MLEs will not be better than the moment estimates.

We can obtain the asymptotic property of the exact MLEs by finding the information matrix. We have the following results regarding the partial derivatives of the log-likelihood function with respect to $(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)$:

$$\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \\
\frac{\partial l}{\partial \boldsymbol{\phi}} &= -\frac{n}{\sigma^2} (\boldsymbol{\Gamma}_p \mathbf{D} \boldsymbol{\phi} + \mathbf{C} \boldsymbol{\phi} - \mathbf{c}_0) \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n\bar{t}}{2\sigma^2} + \frac{n}{2\sigma^4} (c_{00} - 2\boldsymbol{\phi}' \mathbf{c}_0 + \boldsymbol{\phi}' \mathbf{C} \boldsymbol{\phi}) \\
\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \\
\frac{\partial^2 l}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} &= -\frac{n}{\sigma^2} \left[\frac{\partial (\boldsymbol{\Gamma}_p \mathbf{D} \boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} + \mathbf{C} \right] \\
\frac{\partial^2 l}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{n}{\sigma^6} (c_{00} - 2\boldsymbol{\phi}' \mathbf{c}_0 + \boldsymbol{\phi}' \mathbf{C} \boldsymbol{\phi})
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\phi}} &= -\frac{n}{\sigma^2} \sum_{l=1}^{p} \left\{ \frac{\partial c_{kl}}{\partial \boldsymbol{\beta}'} \right\} \phi_l + \frac{n}{\sigma^2} \left\{ \frac{\partial c_{k0}}{\partial \boldsymbol{\beta}'} \right\} \\
\frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}} &= -\frac{1}{\sigma^4} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \\
\frac{\partial^2 l}{\partial \boldsymbol{\phi} \partial \sigma^2} &= \frac{n}{\sigma^4} (\mathbf{C}\boldsymbol{\phi} - \mathbf{c}_0) .
\end{aligned}
$$

Note that

$$
\frac{\partial c_{kl}}{\partial \boldsymbol{\beta}'} = -\frac{2}{n} \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{C}_{ikl} \mathbf{X}_i ,
$$

which yields

$$
\mathrm{E}\left(\frac{\partial c_{kl}}{\partial \boldsymbol{\beta}'} \right) = \mathbf{0}'.
$$

Also we have

$$
\mathrm{E}(c_{kl}) = (\bar{t} - k - l)\gamma_{k-l} \quad k, l = 0, \ldots, p,
$$

which implies

$$
\begin{aligned}
\mathrm{E}(\mathbf{C}) = \{(\bar{t} - k - l)\gamma_{k-l}\} &= (\bar{t}\mathbf{I}_p - \mathbf{D})\boldsymbol{\Gamma}_p - \boldsymbol{\Gamma}_p \mathbf{D}, \\
\mathrm{E}(\mathbf{c}_0) = \{(\bar{t} - k)\gamma_k\} &= (\bar{t}\mathbf{I}_p - \mathbf{D})\boldsymbol{\Gamma}_p \, \boldsymbol{\phi}, \\
\mathrm{E}(\mathbf{c}_0) - \mathrm{E}(\mathbf{C})\boldsymbol{\phi} &= \boldsymbol{\Gamma}_p \mathbf{D} \boldsymbol{\phi}.
\end{aligned}
$$

Say $\boldsymbol{\Gamma}_p = \sigma^2 \mathbf{V}_p$, then according to Cheang and Reinsel (2000), the $(i, j)^{th}$ element of $\mathbf{V}_p^{-1}$ is

$$
v^{ij} = \sum_{k=1}^{min(i,j)} \left(\phi_{i-k}\phi_{j-k} - \phi_{p-(i-k)}\phi_{p-(j-k)} \right), \quad i, j = 1, \ldots, p,
$$

with the convention that $\phi_0 = -1$. For example,

$$
\begin{aligned}
\mathbf{V}_1^{-1} &= (1 - \phi^2)^{-1}, \\
\mathbf{V}_2^{-1} &= \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{bmatrix}.
\end{aligned}
$$

The derivation of $\mathbf{V}_p^{-1}$ with respect to ϕ_k is

$$\frac{\partial \mathbf{V}_p^{-1}}{\partial \phi_k} = 2\mathbf{M}_k = 2\left(\sum_{l=0}^{p-k-1} \phi_l \mathbf{C}_{kl}^p - \sum_{l=p-k+1}^{p} \phi_l \mathbf{C}_{p-k,p-l}^p\right), \quad k = 1, \ldots, p$$

with $\mathbf{C}_{kl}^p = \mathbf{0}$, if $k + l \geq p$. Since

$$\frac{\partial \mathbf{\Gamma}_p}{\partial \phi_k} = -\frac{1}{\sigma^2}\, \mathbf{\Gamma}_p \,\frac{\partial \mathbf{V}_p^{-1}}{\partial \phi_k}\, \mathbf{\Gamma}_p, \quad k = 1, \ldots, p,$$

we get

$$\begin{aligned} \frac{\partial(\mathbf{\Gamma}_p \mathbf{D}\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} &= \left\{\frac{\partial \mathbf{\Gamma}_p}{\partial \phi_k}\,(\mathbf{D}\boldsymbol{\phi})\right\} + \mathbf{\Gamma}_p\, \mathbf{D} \\ &= -2\, \mathbf{\Gamma}_p \{\mathbf{M}_k \mathbf{V}_p\, \mathbf{D}\boldsymbol{\phi}\} + \mathbf{\Gamma}_p\, \mathbf{D}. \end{aligned}$$

Thus, we have

$$\begin{aligned} \mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right) &= -\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \\ \mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}\right) &= -\frac{n}{\sigma^2}\,[(\bar{t}\mathbf{I}_p - \mathbf{D})\mathbf{\Gamma}_p - 2\, \mathbf{\Gamma}_p \{\mathbf{M}_k \mathbf{V}_p\, \mathbf{D}\boldsymbol{\phi}\}] \approx -n(\bar{t} - 2)\mathbf{V}_p \\ \mathrm{E}\left(\frac{\partial^2 l}{\partial (\sigma^2)^2}\right) &= -\frac{n\bar{t}}{2\sigma^4} \\ \mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\phi}}\right) &= \mathbf{0}' \\ \mathrm{E}\left(\frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}}\right) &= \mathbf{0} \\ \mathrm{E}\left(\frac{\partial^2 l}{\partial \boldsymbol{\phi} \partial \sigma^2}\right) &= -\frac{n}{\sigma^2}\, \mathbf{V}_p\, \mathbf{D}\boldsymbol{\phi}. \end{aligned}$$

Hence, the information matrix is given by

$$\mathbf{I}_n(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & n(\bar{t} - 2)\mathbf{V}_p & \frac{n}{\sigma^2}\mathbf{V}_p \mathbf{D}\boldsymbol{\phi} \\ \mathbf{0} & \frac{n}{\sigma^2}\boldsymbol{\phi}'\mathbf{D}\mathbf{V}_p & \frac{n\bar{t}}{2\sigma^4} \end{bmatrix}.$$

By finding the inverse of the bottom block matrix, we get

$$\mathbf{I}_n^{-1}(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2) = \frac{1}{n}\, \mathbf{\Sigma}_l, \tag{4.1.29}$$

where $\Sigma_l = \mathrm{diag}(v_1, \Sigma_{2l})$ with $v_1 = \sigma^2 \left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}$ and

$$\Sigma_{2l} = \frac{1}{k(\bar{t}-2)} \begin{bmatrix} k\mathbf{V}_p^{-1} + \mathbf{D}\phi\phi'\mathbf{D} & -(\bar{t}-2)\sigma^2\mathbf{D}\phi \\ -(\bar{t}-2)\sigma^2\phi'\mathbf{D} & (\bar{t}-2)\sigma^4 \end{bmatrix},$$

where $k = \bar{t}(\bar{t}-2)/2 - \phi'\mathbf{D}\mathbf{V}_p\mathbf{D}\phi$. This is the asymptotic variances of the ML estimates of (β, ϕ, σ^2).

4.1.4 Quasi-least squares estimates

In order to obtain quasi-least squares estimates of ϕ, we need to minimize $S(\beta, \phi)$ with respect to β and ϕ. Equating to zero the partial differentiate of $S(\beta, \phi)$ with respect to β we get (4.1.22). Estimate of ϕ is obtained by two steps as follows.

The first step is to solve equation (1.2.20). From (4.1.9) we get

$$\sum_{i=1}^n \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}}{\partial \phi}\mathbf{U}_i\right) = 2n(\mathbf{C}\phi - \mathbf{c}_0) = \mathbf{0},$$

which gives the ULS estimate

$$\hat{\phi}_u = \mathbf{C}^{-1}\mathbf{c}_0.$$

Note that if we estimate γ_{k-l} by $\hat{c}_{kl}/\bar{t}$ in stead of $\hat{c}_{kl}/(\bar{t}-k-l)$ and γ_k by $\hat{c}_k/\bar{t}$, then $\hat{\phi}_u$ is identical to $\hat{\phi}_{al}$. The second step needs to modify $\hat{\phi}_u$ to be consistent. From (1.2.21) we get

$$\begin{aligned} \sum_{i=1}^n \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\phi}_u)}{\partial \phi}\cdot \mathbf{V}_i\right) &= \Delta^{-1}\cdot \sum_{i=1}^n \mathrm{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\phi}_u)}{\partial \phi}\cdot \mathbf{P}_i\right) \\ &= 2\Delta^{-1}\cdot \left[\left\{\sum_{i=1}^n \mathrm{tr}(\mathbf{C}_{ikl}\mathbf{P}_i)\right\}\hat{\phi}_u - \left\{\sum_{i=1}^n \mathrm{tr}(\mathbf{C}_{ik0}\mathbf{P}_i)\right\}\right] \\ &\overset{\text{set}}{=} \mathbf{0}. \end{aligned}$$

Since that

$$\mathrm{tr}(\mathbf{C}_{ikl}\mathbf{P}_i) = (t_i - k - l)\rho_{k-l}, \quad k = 1,2,\dots,p, \;\; l = 0,1,\dots,p, \;\; i = 1,\dots,n,$$

the equation above simplifies to

$$\begin{bmatrix} \bar{t}-2 & (\bar{t}-3)\rho_1 & \cdots & [\bar{t}-(p+1)]\,\rho_{p-1} \\ (\bar{t}-3)\rho_1 & \bar{t}-4 & \cdots & [\bar{t}-(p+2)]\,\rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ [\bar{t}-(p+1)]\,\rho_{p-1} & [\bar{t}-(p+2)]\,\rho_{p-2} & \cdots & \bar{t}-2p \end{bmatrix} \hat{\boldsymbol{\phi}}_u - \begin{bmatrix} (\bar{t}-1)\rho_1 \\ (\bar{t}-2)\rho_2 \\ \vdots \\ (\bar{t}-p)\rho_p \end{bmatrix} = \mathbf{0}.$$

Let $\mathbf{D}_k$ be a $p \times p$ tridiagonal matrix with k^{th} diagonal elements in the order $\{\bar{t}-(k+2), \bar{t}-(k+2\cdot 2), \dots, \bar{t}-(2p-k)\}$, for $k = 0,1,2,\dots,p-1$. The equation above becomes

$$(\mathbf{D}_0 + \sum_{k=1}^{p-1} \mathbf{D}_k \rho_k)\hat{\boldsymbol{\phi}}_u - (\bar{t}\mathbf{I}_p - \mathbf{D})\boldsymbol{\rho} = \mathbf{0},$$

that is

$$\mathbf{D}_0\hat{\boldsymbol{\phi}}_u + \sum_{i=1}^{p-1} \mathbf{D}_k \hat{\boldsymbol{\phi}}_u \boldsymbol{\epsilon}_k' \boldsymbol{\rho} - (\bar{t}\mathbf{I}_p - \mathbf{D})\boldsymbol{\rho} = \mathbf{0}.$$

This gives

$$\hat{\boldsymbol{\rho}} = (\bar{t}\mathbf{I}_p - \mathbf{D} - \sum_{k=1}^{p-1} \mathbf{D}_k \hat{\boldsymbol{\phi}}_u \boldsymbol{\epsilon}_k')^{-1}\mathbf{D}_0\hat{\boldsymbol{\phi}}_u.$$

From (4.1.4) we get

$$\hat{\boldsymbol{\phi}} = \mathbf{P}_p^{-1}(\hat{\boldsymbol{\rho}})\hat{\boldsymbol{\rho}}. \tag{4.1.30}$$

The quasi-least squares estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ are the simultaneous solutions of (4.1.22) and (4.1.30), and $\hat{\sigma}_q^2$ is obtained plugging $\hat{\boldsymbol{\beta}}_q$ and $\hat{\boldsymbol{\phi}}_q$ in (4.1.23).

4.1.5 Comparisons through simulation

The asymptotic distribution of the moment and quasi-least squares estimates can be found by applying the delta theorem, but the formula will be very complicated. To avoid the tedious computations, we will compare the three estimating methods by simulation. The comparisons will be made with respect to the exact maximum likelihood estimates by solving the maximum likelihood equations using Newton-Raphson method. Our simulation results show that the approximate maximum likelihood estimates suggested by Box *et al.* (1994) are less efficient than the moment estimates. As before, we concentrate on comparing the estimates of ϕ. Particularly we choose $n = 30$ and $p = 4$. Without loss of generality, we assume $\boldsymbol{\beta} = \mathbf{0}$ and $\sigma^2 = 1$ in our simulation. The true parameters should be chosen in such a way that the series is stationary and invertible. It is hard to study all possible parameters, so we choose $\phi = (0.5, -0.2, 0.1, 0.2)$ for illustrating purposes, and then study the asymptotic properties of the estimates as t ranges from 10 to 55.

First we study the asymptotic properties of the estimates when the data is normal. As in previous chapters, we generate a t-dimensional vector ε whose elements are from a standard normal distribution. We then set $y = \sigma^2 \mathbf{V}^{1/2}(\phi)\varepsilon$ and generate a sample of size $n = 30$. The process is then repeated 10000 times. For each replication, we compute the moment, maximum likelihood and quasi-least estimates of ϕ, and then compute the mean square errors (MSEs). Since the MLE of ϕ is not in a closed form, we use the Newton-Raphson method to solve the ML equation. We set the precision to be $1e - 10$ when solving the ML equation. The QLS estimates and moment estimates are in closed forms. If the estimates (including moment, ML and quasi-least squares estimates) are not feasible, that sample was deleted and hence excluded from the analysis. It is interesting to note

Table 4.1. ARE of $\hat{\phi}_q$ and $\hat{\phi}_m$ (in parenthesis) vs $\hat{\phi}_l$ when the data is simulated from normal distribution

t	ϕ_1 (0.5)	ϕ_2 (−0.2)	ϕ_3 (0.1)	ϕ_4 (0.2)
10	0.7450 (0.7911)	0.7934 (0.9702)	0.8043 (0.8820)	0.8928 (0.9965)
15	0.9695 (0.8634)	1.0714 (0.9672)	0.9768 (0.9308)	1.0289 (0.9710)
20	0.9853 (0.9124)	1.0546 (0.9946)	0.9900 (0.9534)	1.0261 (0.9703)
25	0.9875 (0.9270)	1.0442 (0.9951)	0.9958 (0.9716)	1.0241 (0.9790)
30	0.9930 (0.9336)	1.0383 (0.9922)	0.9967 (0.9695)	1.0214 (0.9796)
35	0.9966 (0.9517)	1.0327 (0.9960)	0.9992 (0.9710)	1.0183 (0.9781)
40	0.9971 (0.9594)	1.0296 (0.9877)	0.9999 (0.9769)	1.0182 (0.9816)
45	0.9965 (0.9609)	1.0252 (0.9994)	0.9999 (0.9771)	1.0139 (0.9847)
50	0.9975 (0.9593)	1.0236 (0.9971)	1.0001 (0.9917)	1.0146 (0.9787)
55	0.9982 (0.9722)	1.0217 (1.0000)	1.0002 (0.9868)	1.0118 (0.9844)

that when $t = 10$, quasi-least squares estimates are more likely to be unfeasible (about 0.3%), while when t is larger than 10, all estimates seem to be feasible. The proportion of feasible estimates decreases as t increases.

Define the ARE of $\hat{\phi}_{1q}$ with respect to $\hat{\phi}_{1l}$ as $e(\hat{\phi}_{1q}; \hat{\phi}_{1l}) = MSE_{1l}/MSE_{1q}$, and define the ARE of $\hat{\phi}_{1m}$ with respect to $\hat{\phi}_{1l}$ similarly. The subscript '1' is understandable to represent the notations associated with ϕ_1. Define the ARE for ϕ_2, ϕ_3 and ϕ_4 similarly. Table 4.1 contains the AREs of ϕ. Figure 4.1 gives the plot of the AREs of the quasi-least squares and moment estimates of ϕ_1 with respect to the MLE. The ARE plots of ϕ_2, ϕ_3 and ϕ_4 are similar. Most of the efficiencies are close to but less than 1. Some quasi-least squares efficiencies are even larger than 1 (such as 1.03) due to the fact that the MLEs are obtained approximately by the Newton-Raphson method (exact solution is not possible, we set the precision to be $1e - 10$). It is clear from the plot that the efficiencies of the quasi-least squares estimates are much large than those of the moment estimates, and approach 1 as t increases.

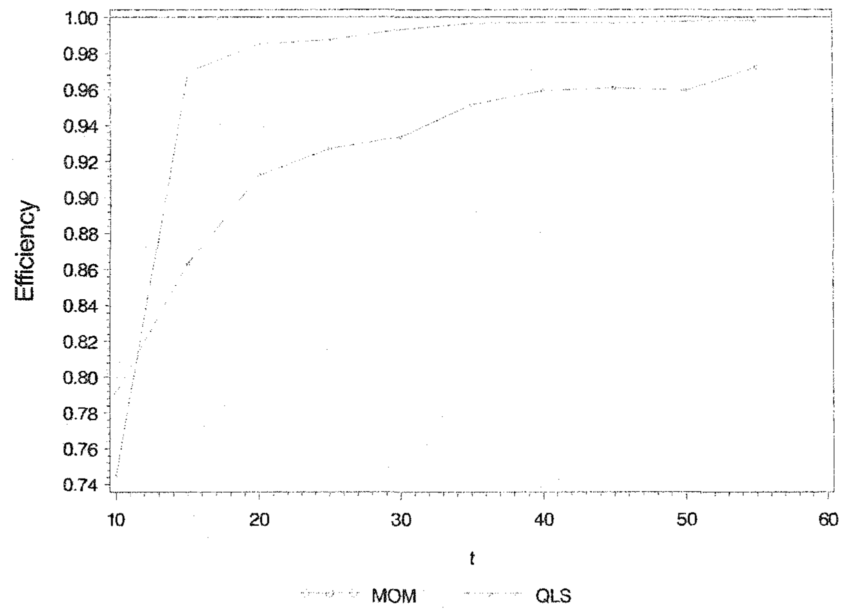

Figure 4.1. AREs of $\hat{\phi}_{1q}$ and $\hat{\phi}_{1m}$ vs $\hat{\phi}_{1l}$ when the data is simulated from normal distribution.

To study the robust property of the quasi-least squares estimates, we have simulated data from a Student-t distribution. We generate a t-dimensional (t ranges from 10 to 55) vector ε whose elements form a random sample from the Student-t distribution with mean 0 and 5 degrees of freedom, and then let $\mathbf{y} = \sigma\sqrt{3/5}\ \mathbf{V}^{1/2}(\boldsymbol{\phi})\,\varepsilon$ and generate 30 of them. Now the process is repeated 10000 times. For each replication we computed the moment, normal ML (assuming the data is normal) and quasi-least squares estimates of $\boldsymbol{\phi}$, and then compute the mean square errors. About 1% of the quasi-least squares estimates are not feasible when $t = 10$. All estimates are feasible when $t > 10$. The asymptotic relative efficiencies are defined as before. Table 4.2 gives the efficiencies. Figure 4.2 gives the plot of the AREs of the quasi-least squares and moment estimates of ϕ_2 with respect to the MLE. From the plot we see that most efficiencies of the quasi-least squares estimates are greater than 1, but the efficiencies of the moment estimates are not.

Table 4.2. ARE of $\hat{\phi}_q$ and $\hat{\phi}_m$ (in parenthesis) vs $\hat{\phi}_l$ when the data is simulated from Student-t distribution

t	ϕ_1 (0.5)	ϕ_2 (−0.2)	ϕ_3 (0.1)	ϕ_4 (0.2)
10	0.6238 (0.7664)	0.6157 (0.9175)	0.6694 (0.8729)	0.7373 (1.0427)
15	0.9618 (0.8639)	1.0476 (0.9536)	0.9615 (0.9316)	0.9978 (1.0019)
20	0.9826 (0.8845)	1.0465 (0.9606)	0.9798 (0.9468)	1.0135 (0.9800)
25	0.9903 (0.9151)	1.0404 (0.9739)	0.9901 (0.9692)	1.0131 (0.9971)
30	0.9930 (0.9294)	1.0378 (0.9772)	0.9933 (0.9786)	1.0153 (0.9855)
35	0.9974 (0.9331)	1.0340 (0.9706)	0.9968 (0.9811)	1.0143 (0.9810)
40	0.9966 (0.9414)	1.0273 (0.9832)	0.9970 (0.9756)	1.0117 (0.9850)
45	0.9968 (0.9473)	1.0243 (0.9768)	0.9986 (0.9764)	1.0126 (0.9965)
50	0.9977 (0.9595)	1.0229 (0.9890)	0.9991 (0.9849)	1.0114 (0.9921)
55	0.9976 (0.9626)	1.0206 (0.9854)	0.9990 (0.9834)	1.0109 (0.9843)

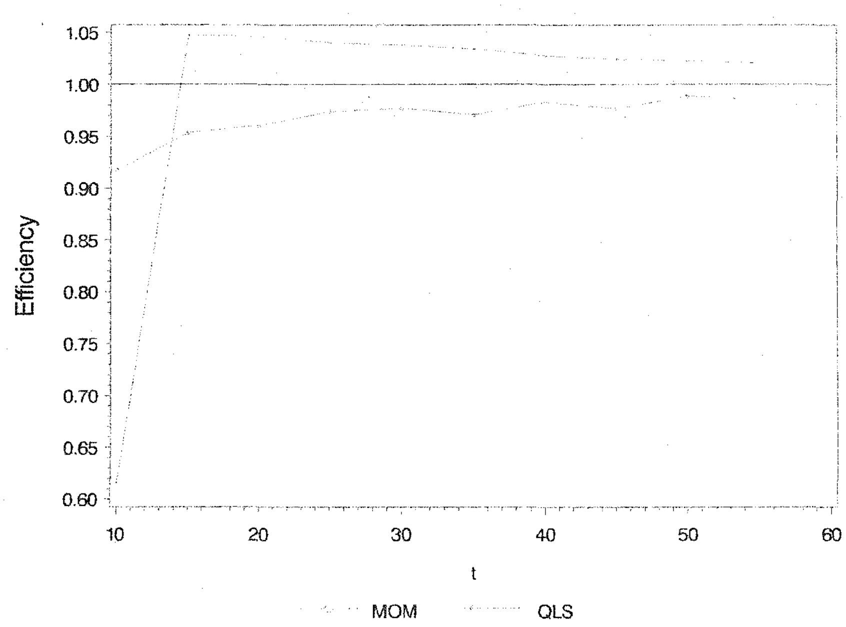

Figure 4.2. AREs of $\hat{\phi}_{2q}$ and $\hat{\phi}_{2m}$ vs $\hat{\phi}_{2l}$ when the data is simulated from Student-t distribution.

Table 4.3. ARE of $\hat{\phi}_q$ and $\hat{\phi}_m$ (in parenthesis) vs $\hat{\phi}_l$ when the data is simulated from $0.5N(0,4) + 0.5N(1,4)$

t	ϕ_1 (0.5)	ϕ_2 (-0.2)	ϕ_3 (0.1)	ϕ_4 (0.2)
10	1.4227 (1.1181)	1.0123 (0.9776)	1.6454 (1.3734)	7.7692 (8.0616)
15	1.4418 (1.2663)	1.1217 (1.0083)	1.6202 (1.3710)	5.0855 (4.0307)
20	1.4168 (1.3300)	1.0650 (0.9840)	1.5187 (1.3079)	3.1647 (2.4736)
25	1.3739 (1.3264)	1.0346 (0.9584)	1.4384 (1.2706)	2.4562 (1.9226)
30	1.3311 (1.3091)	1.0179 (0.9554)	1.3802 (1.2260)	2.0795 (1.6485)
35	1.2966 (1.2801)	1.0076 (0.9538)	1.3318 (1.1989)	1.8568 (1.4993)
40	1.2663 (1.2642)	1.0008 (0.9509)	1.2970 (1.1782)	1.7054 (1.3987)
45	1.2413 (1.2440)	0.9977 (0.9520)	1.2683 (1.1616)	1.6051 (1.3308)
50	1.2216 (1.2245)	0.9953 (0.9545)	1.2438 (1.1488)	1.5250 (1.2863)
55	1.2028 (1.2101)	0.9935 (0.9534)	1.2235 (1.1333)	1.4640 (1.2479)

As t goes larger, the difference becomes less prominent. The efficiency plot of ϕ_4 shows similar pattern, while the efficiencies of ϕ_1 and ϕ_3 are still less than but close to 1.

To further see the robust property of quasi-least squares estimates, we simulate the data from a contaminated normal distribution. First we generate a random number between 0 and 1, if the number is less than 0.5, we generate ε from $N(0,4)$; otherwise, we generate ε from $N(1,4)$. We then modify ε to have mean 0 and variance 1 by subtracting it from 0.5 and then dividing by $\sqrt{4.25}$. Let $\boldsymbol{\varepsilon}$ be a vector of a size t consisting of random sample of ε. Secondly, we let $\mathbf{y} = \sigma^2\mathbf{V}^{1/2}(\phi)\boldsymbol{\varepsilon}$ and generate a sample of size $n = 30$. The whole process is then repeated 10,000 times. For each replication, we compute the moment, maximum likelihood and quasi-least squares estimates of ϕ, and then compute the mean square errors. In this case about 15% of the maximum likelihood estimates are not feasible! But all moment and quasi-least squares estimates are feasible. Define the asymptotic relative efficiencies as before. Table 4.3 contains the AREs of ϕ. Figure 4.3 gives the plot

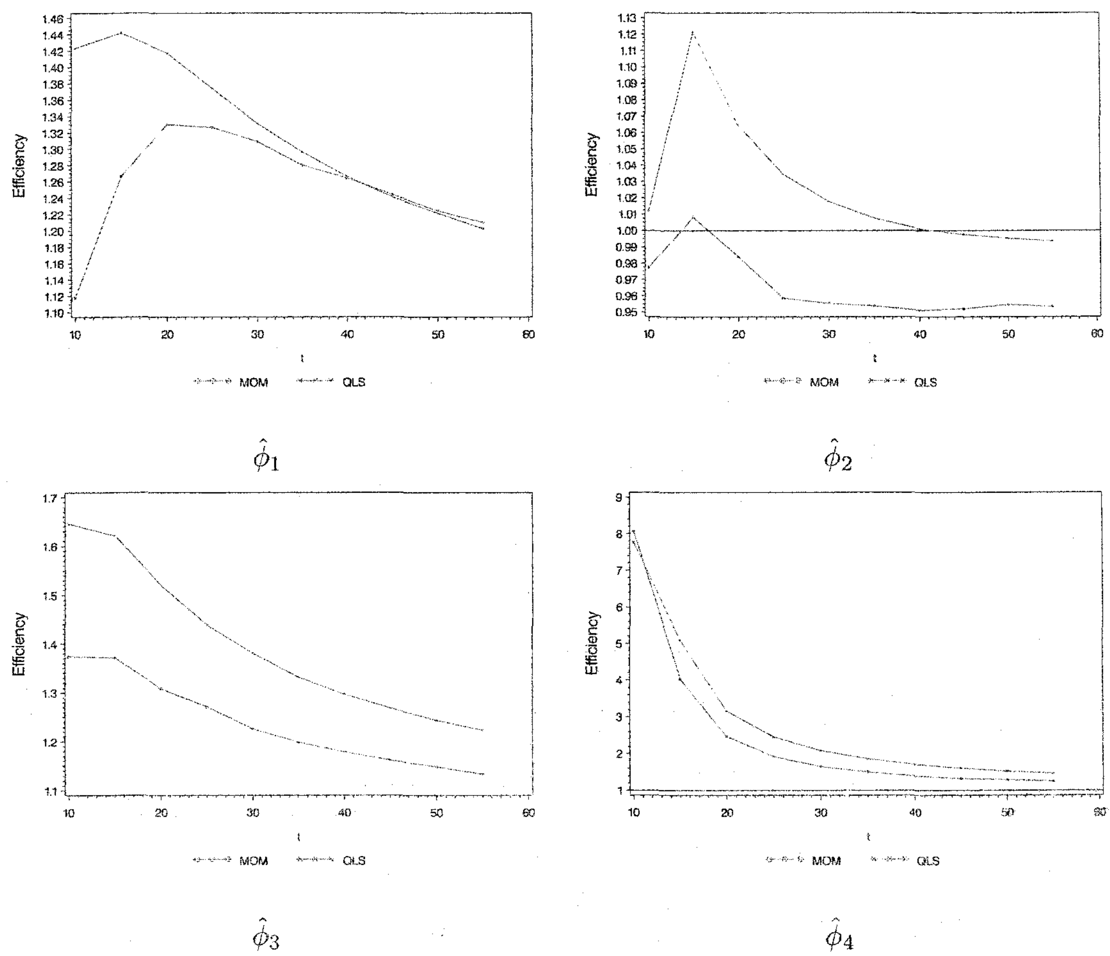

Figure 4.3. *AREs of* $\hat{\phi}_q$ *and* $\hat{\phi}_m$ *vs* $\hat{\phi}_l$ *when the data is simulated from* $0.5N(0,4) + 0.5N(1,4)$.

of the AREs of the quasi-least squares and moment estimates with respect to the maximum likelihood estimates of ϕ. From the plot we see that most efficiencies of the quasi-least squares estimates are much greater than 1, more so than the efficiencies of the moment estimates. It is clear that the quasi-least squares estimates are much better than the moment estimates. As t goes larger, the efficiencies all approach 1.

Based on the plots, we again demonstrate that, when the error of the model is an AR(p) process, quasi-least squares estimates are better than moment estimates, and good competitors to maximum likelihood estimates. When the data is from a distribution which differs slightly from normal, quasi-least squares estimates are more robust than maximum likelihood estimates.

The applications of the estimating methods for AR(p) ($p > 2$) model are limited, since most of the data could be analyzed by fitting a simple AR(1) or AR(2) model. For the simple model we have already studied the applications in the previous chapters. However, in this chapter we did comparisons of the estimating methods to provide guidelines if one is interested in fitting a more complicated model for real data.

4.2 The model with MA(1) errors

4.2.1 Model

A model that is complementary to the autoregressive process of order one (AR(1)) is the moving average process of order one (MA(1)). Anderson (1975) showed the basic properties of MA(1) process for time series regression model with replications. The maximum likelihood approach was studied by Haddad (1995) for a single series. Feigin, *et al.* (1996) studied the model using an alternative approach with positive innovations. The model is given by (1.1.7), and (1.1.2) reduces to

$$\varepsilon_j = a_j - \theta a_{j-1}, \quad j = 1, 2, \ldots, t. \tag{4.2.31}$$

The invertibility condition requires that $|\theta| < 1$. From (1.1.4), the variance of the process is given by

$$\gamma_0 = \sigma_\varepsilon^2 = \sigma^2(1 + \theta^2). \tag{4.2.32}$$

The autocorrelation satisfies

$$\begin{aligned} \rho_1 &= \frac{-\theta}{1+\theta^2}, \quad |\rho_1| < 0.5 \\ \rho_k &= 0, \quad k \geq 2 \end{aligned} \tag{4.2.33}$$

according to (1.1.5). This implies that

$$\theta = \begin{cases} \dfrac{-1 + \sqrt{1 - 4\rho^2}}{2\rho}, & \rho \neq 0 \\ 0, & \rho = 0, \end{cases} \tag{4.2.34}$$

where ρ is the abbreviation for ρ_1. Figure 4.4 shows the relationship between θ and ρ. The error $\boldsymbol{\varepsilon}_i$ has mean $\mathbf{0}$ and covariance

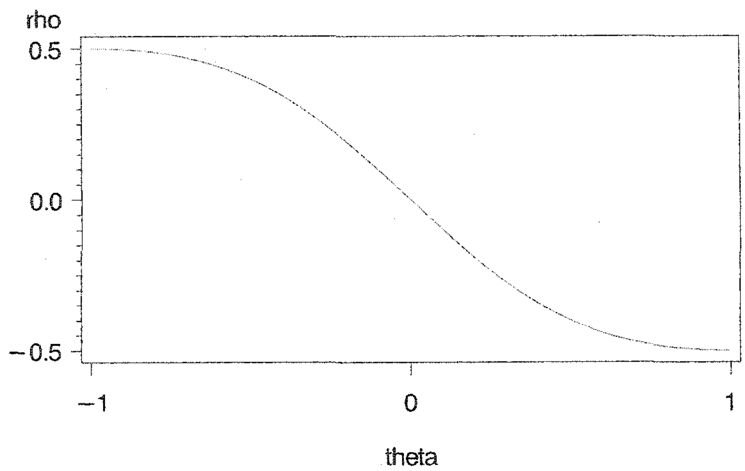

Figure 4.4. Relationship between θ and ρ

$$\mathbf{\Gamma}_i(\theta, \sigma^2) = \sigma_\varepsilon^2 \mathbf{P}_i(\rho) = \sigma^2 \mathbf{V}_i(\theta), \tag{4.2.35}$$

where $\mathbf{P}_i(\rho)$ is the $t_i \times t_i$ correlation matrix with ρ on the first off diagonal and zero's elsewhere, that is

$$\mathbf{P}_i(\rho) = \mathbf{I}_{t_i} + 2\,\rho\,\mathbf{C}_{i10},$$

and

$$\begin{aligned} \mathbf{V}_i(\theta) &= (1+\theta^2)\,\mathbf{P}_i(\rho) \\ &= (1+\theta^2)\,\mathbf{I}_{t_i} - 2\,\theta\,\mathbf{C}_{i10}, \end{aligned} \tag{4.2.36}$$

where $\mathbf{C}_{i10}$ is defined in (2.1.6) and $\mathbf{I}_{t_i}$ is the t_i−dimensional identity matrix. This gives

$$\frac{\partial \mathbf{V}_i}{\partial \theta} = 2\,(\theta\,\mathbf{I}_{t_i} - \mathbf{C}_{i10}). \tag{4.2.37}$$

Haddad (1995) has shown that

$$\begin{aligned} \mathbf{V}_i^{-1}(\theta) &= \frac{1}{1-\theta^2}\,(\mathbf{\Omega}_i(\theta) + \mathbf{\Omega}_i^*(\theta)) \\ &\approx \frac{1}{1-\theta^2}\,\mathbf{\Omega}_i(\theta) \end{aligned}$$

$$= \frac{1}{1-\theta^2}\left(\mathbf{I}_{t_i} + 2\sum_{k=1}^{t_i-1} \mathbf{C}_{ik0}\theta^k\right), \tag{4.2.38}$$

where the $t_i \times t_i$ matrix $\mathbf{\Omega}_i(\theta) = \{\theta^{|k-l|}\}$ is the t_i- dimensional first order autocorrelation matrix with parameter θ, $\mathbf{\Omega}_i^*(\theta) = (\omega_{kl})$ is given by

$$\omega_{kl} = \begin{cases} -\theta^{l+1}(1-\theta^{2(t_i-l+1)})/(1-\theta^{2t_i+2}) & \text{if } k=1,\ l=1,\ldots,t_i, \\ \omega_{1,(t_i-k+1)} & \text{if } k=t_i,\ l=1,\ldots,t_i, \\ 0 & \text{elsewhere,} \end{cases}$$

and $\mathbf{C}_{ik0}$'s are defined as in (4.1.8). Note that $\mathbf{\Omega}_i(\theta)$ is similar to $\mathbf{P}_i(\phi)$ in AR(1) model. Generally speaking, the matrix $\mathbf{\Omega}_i^*$ is close to 0 and negligible. But when n is large the matrix cannot be ignored to get efficient estimates. If we ignore $\mathbf{\Omega}_i^*$, equation (4.2.38) implies

$$\frac{\partial \mathbf{V}_i^{-1}(\theta)}{\partial\theta} = \frac{2}{(1-\theta^2)^2}\left[\theta\mathbf{I}_i + \sum_{k=1}^{t_i-1}\mathbf{C}_{ik0}\left(k\theta^{k-1} - (k-2)\theta^{k+1}\right)\right], \tag{4.2.39}$$

but by taking derivative of the exact inverse of $\mathbf{V}$ with respect to θ we get

$$\begin{aligned} \frac{\partial \mathbf{V}_i^{-1}(\theta)}{\partial\theta} &= -\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i(\theta)}{\partial\theta}\mathbf{V}_i^{-1} \\ &= -2\,\mathbf{V}_i^{-1}(\theta\mathbf{I}_{t_i} - \mathbf{C}_{i10})\mathbf{V}_i^{-1}. \end{aligned} \tag{4.2.40}$$

The adjusted sum of square errors $S(\boldsymbol{\beta},\theta)$ defined in (1.2.9) becomes

$$\begin{aligned} S(\boldsymbol{\beta},\theta) &= \sum_{i=1}^{n}\operatorname{tr}\left(\mathbf{V}_i^{-1}\mathbf{U}_i\right) \\ &\approx \frac{1}{1-\theta^2}\left(n\,c_{00} + 2\sum_{i=1}^{n}\sum_{k=1}^{t_i-1}\operatorname{tr}(\mathbf{C}_{ik0}\mathbf{U}_i)\theta^k\right), \end{aligned} \tag{4.2.41}$$

where c_{k0}'s are as defined in (4.1.10).

4.2.2 Moment and Maximum likelihood estimates

Recall that the generalized least squares (GLS) estimates of $(\boldsymbol{\beta},\sigma^2)$ as in (1.2.10) and (1.2.12) respectively are given by

$$\hat{\beta}_g = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1} \cdot \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i \tag{4.2.42}$$

$$\hat{\sigma}_g^2 = \frac{1}{n\bar{t}} S(\beta, \theta) = \frac{1}{n\bar{t}} \sum_{i=1}^{n} \mathrm{tr}(\mathbf{V}_i^{-1}\mathbf{U}_i)$$

$$\approx \frac{1}{n\bar{t}(1-\theta^2)} \left(n\, c_{00} + 2\sum_{i=1}^{n}\sum_{k=1}^{t_i-1} \mathrm{tr}(\mathbf{C}_{ik0}\mathbf{U}_i)\theta^k\right). \tag{4.2.43}$$

Given the "residuals" $\hat{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\beta}$, the variance of y_{ij} is estimated by $c_{00}/\bar{t}$ and the first order covariance is estimated by $c_{10}/(\bar{t}-1)$. This yields

$$\hat{\rho} = \frac{c_{10}/(\bar{t}-1)}{c_{00}/\bar{t}} = \frac{\bar{t}c_{10}}{(\bar{t}-1)c_{00}}.$$

By combining (4.2.42) and (4.2.34) after plugging in $\hat{\rho}$ and solving for θ in the feasible region $(-1,\ 1)$, we obtain the moment estimate $(\hat{\beta}_m,\ \hat{\theta}_m)$. The estimate of $\hat{\sigma}_m^2$ is obtained plugging in the estimates of β and θ in (4.2.43). More powerful algorithms may be used to get the estimates (Box *et al.* 1994, p.221). These methods avoid discussing the solution based on the value of $\hat{\rho}$.

The MLEs are derived as follows. Assuming the errors ε_i's are normal, and recalling (1.2.14), the log-likelihood function may be written

$$l(\beta, \theta, \sigma^2) = -\frac{n\bar{t}}{2}\log(2\pi) - \frac{n\bar{t}}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{V}_i(\theta)| - \frac{1}{2\sigma^2}S(\beta, \theta), \tag{4.2.44}$$

where $S(\beta,\ \theta)$ is defined in (4.2.41). Equating to zero the partial derivatives of (4.2.44) with respect to β and σ^2 we obtain (4.2.42) and (4.2.43), respectively. According to Haddad (1995), the determinate of $\mathbf{V}_i$ is given by

$$|\mathbf{V}_i| = \sum_{k=0}^{t_i} \theta^{2k} = \frac{1-\theta^{2t_i+2}}{1-\theta^2},$$

which yields

$$\frac{\partial \log|\mathbf{V}_i|}{\partial\theta} = \frac{2\theta}{1-\theta^{2(t_i+1)}}\left(\sum_{k=0}^{t_i-1}\theta^{2k} - t_i\theta^{2t_i}\right).$$

Using the approximate $\mathbf{V}_i$, we get the partial derivative of (4.2.44) with respect to θ as

$$\frac{\partial l}{\partial \theta} = -\theta \sum_{i=1}^{n} \frac{\sum_{k=0}^{t_i-1} \theta^{2k} - t_i \theta^{2t_i}}{1-\theta^{2(t_i+1)}} - \frac{n\theta c_{00} + \sum_{i=1}^{n} \sum_{k=1}^{t_i-1} \operatorname{tr}(\mathbf{C}_{ik0}\mathbf{U}_i)\left(k\theta^{k-1} - (k-2)\theta^{k+1}\right)}{\sigma^2(1-\theta^2)^2}.$$

The ML estimate is found by setting the equation above to zero and solving for θ. If we assume that $t_i \equiv t$, the equation above reduces to

$$\frac{\partial l}{\partial \theta} = -\frac{n\theta\left(\sum_{k=0}^{t-1}\theta^{2k} - t\theta^{2t}\right)}{1-\theta^{2(t+1)}} - \frac{n\left[c_{00}\theta + \sum_{k=1}^{t-1} c_{k0}\left(k\theta^{k-1} - (k-2)\theta^{k+1}\right)\right]}{\sigma^2(1-\theta^2)^2}.$$

Since $c_{k0} \approx 0,\ k \geq 2$, we obtain the approximate ML equation as

$$\sigma^2\theta(1-\theta^2)\left(\sum_{k=0}^{t-1}\theta^{2k} - t\theta^{2t}\right) + (c_{10} + c_{00}\theta + c_{10}\theta^2)\sum_{k=0}^{t}\theta^{2k} = 0. \qquad (4.2.45)$$

Substituting the estimate of $\sigma^2 \approx (c_{00} + 2c_{10}\theta)/[t(1-\theta^2)]$ and simplifying the equation above we obtain

$$\left[tc_{10} + (t+1)c_{00}\theta + (t+2)c_{10}\theta^2\right]\sum_{k=0}^{t-1}\theta^{2k} + t(1-\theta^2)c_{10}\theta^{2t} = 0. \qquad (4.2.46)$$

If t is large enough, θ^{2t} is close to 0, and the equation above becomes

$$tc_{10} + (t+1)c_{00}\theta + (t+2)c_{10}\theta^2 = 0,$$

which gives

$$\hat{\theta} = \begin{cases} \dfrac{-(t+1)c_{00} + \sqrt{(t+1)^2 c_{00}^2 - 4t(t+2)c_{10}^2}}{2(t+2)c_{10}}, & \text{if} \quad c_{10} \neq 0 \\ 0, & \text{if} \quad c_{10} = 0. \end{cases}$$

If we do not assume $c_{k0} \approx 0$, the ML equation is given by

$$\left(\frac{t+1}{t}\right)\left(\frac{1}{1-\theta^2} - \frac{1}{1-\theta^{2t+2}}\right)\left(c_{00} + 2\sum_{k=1}^{t-1} c_{k0}\theta^k\right) + \sum_{k=1}^{t-1} kc_{k0}\theta^k = 0.$$

Now, let us find the exact ML equation. From (4.2.44) we get

$$\begin{aligned}\frac{\partial l}{\partial \theta} &= -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial \log|\mathbf{V}_i(\theta)|}{\partial \theta} - \frac{1}{2\sigma^2}\frac{\partial S(\boldsymbol{\beta},\theta)}{\partial \theta} \\ &= -\frac{1}{2}\left\{\sum_{i=1}^{n}\frac{\partial|\mathbf{V}_i(\theta)|/\partial\theta}{|\mathbf{V}_i|} + \frac{n\bar{t}\,\partial S(\boldsymbol{\beta},\theta)/\partial\theta}{S(\boldsymbol{\beta},\theta)}\right\} \\ &= -\frac{1}{2}\left\{\sum_{i=1}^{n}\frac{\partial|\mathbf{V}_i(\theta)|/\partial\theta}{|\mathbf{V}_i|} + \frac{n\bar{t}\,\sum_{i=1}^{n}\mathrm{tr}\left[(\partial\mathbf{V}_i^{-1}(\theta)/\partial\theta)\mathbf{U}_i\right]}{\sum_{i=1}^{n}\mathrm{tr}(\mathbf{V}_i^{-1}\mathbf{U}_i)}\right\} \\ &\overset{\text{set}}{=} 0, \end{aligned} \tag{4.2.47}$$

after substituting the exact estimate of σ^2. If we assume $t_i \equiv t$ particularly, the exact ML equation can be simplified as

$$\frac{\partial|\mathbf{V}(\theta)|/\partial\theta}{|\mathbf{V}|} + \frac{t\,\mathrm{tr}\left[(\partial\mathbf{V}^{-1}(\theta)/\partial\theta)\overline{\mathbf{U}}\right]}{\mathrm{tr}(\mathbf{V}^{-1}\overline{\mathbf{U}})} = 0.$$

For MA(1) case, we have

$$\begin{aligned}|\mathbf{V}| &= \sum_{k=0}^{t}\theta^{2k}, \\ \frac{\partial|\mathbf{V}|}{\partial\theta} &= \sum_{k=1}^{t}2k\theta^{2k-1}, \\ \mathrm{tr}\left(\frac{\partial\mathbf{V}^{-1}(\theta)}{\partial\theta}\overline{\mathbf{U}}\right) &= -\mathbf{V}^{-1}\frac{\partial\mathbf{V}(\theta)}{\partial\theta}\mathbf{V}^{-1} \\ &= -2\,\mathrm{tr}\left[\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\mathbf{V}^{-1}\overline{\mathbf{U}}\right].\end{aligned}$$

The ML estimates $(\hat{\boldsymbol{\beta}}_l,\ \hat{\theta}_l)$ are the simultaneous solutions of (4.2.42) and (4.2.47), and $\hat{\sigma}_l^2$ is (4.2.43) plugging in $(\hat{\boldsymbol{\beta}}_l,\ \hat{\theta}_l)$. In Section 4.2.5, we will use the exact ML equation in the simulations instead of the approximate ML estimate.

4.2.3 Quasi-least squares estimates

To obtain QLS estimates, we need to minimize $S(\boldsymbol{\beta},\theta)$ with respect to $\boldsymbol{\beta}$ and θ. Equating to zero the partial derivative of $S(\boldsymbol{\beta},\theta)$ with respect to $\boldsymbol{\beta}$ we get (4.2.42). The estimate of θ is obtained by two steps as follows.

Equating to zero the partial derivative of $S(\boldsymbol{\beta}, \theta)$ with respect to θ we get

$$\sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\theta)}{\partial \theta} \cdot \mathbf{U}_i\right) \approx \frac{2}{(1-\theta^2)^2}\left[\sum_{i=1}^{n}\sum_{k=1}^{t_i-1} \operatorname{tr}(\mathbf{C}_{ik0}\mathbf{U}_i)\left(k\theta^{k-1} - (k-2)\theta^{k+1}\right) + n\theta\, c_{00}\right] \overset{\text{set}}{=} 0. \tag{4.2.48}$$

The solution involves solving the $\max(t_i)$-degree polynomial equation of θ, but since $c_{k0} \approx 0,\ k \geq 2$, an approximation to (4.2.48) may be

$$c_{10} + (c_{00} + 2c_{20})\theta + c_{10}\theta^2 = 0,$$

which yields

$$\frac{-\theta}{1+\theta^2} = \frac{c_{10}}{c_{00} + c_{20}}. \tag{4.2.49}$$

The ULS estimates of $(\boldsymbol{\beta},\ \theta)$ is then obtained by solving (4.2.42) and (4.2.49) recursively. The estimate of $\boldsymbol{\beta}$ obtained this way is also the QLS estimate. Suppose $\hat{\theta}_u$ is the ULS estimate, the second step of QLS method modifies $\hat{\theta}_u$ to be consistent. Recalling (1.2.21) in the second step we need to solve

$$\sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial \mathbf{V}_i^{-1}(\hat{\theta}_u)}{\partial \theta} \cdot \mathbf{V}_i\right) = \frac{2(1+\theta^2)}{(1-\hat{\theta}_u^2)^2}\left[\sum_{i=1}^{n}\sum_{k=1}^{t_i-1} \operatorname{tr}(\mathbf{C}_{ik0}\mathbf{P}_i)\left(k\hat{\theta}_u^{k-1} - (k-2)\hat{\theta}_u^{k+1}\right) + \hat{\theta}_u \sum_{i=1}^{n} \operatorname{tr}(\mathbf{P}_i)\right] \overset{\text{set}}{=} 0.$$

Since

$$\begin{aligned} \operatorname{tr}(\mathbf{P}_i) &= t_i \\ \operatorname{tr}(\mathbf{C}_{i10}\mathbf{P}_i) &= (t_i - 1)\,\rho \\ \operatorname{tr}(\mathbf{C}_{ik0}\mathbf{P}_i) &= 0, \qquad k \geq 2, \end{aligned}$$

the equation above becomes

$$\bar{t}\,\hat{\theta}_u + (\bar{t}-1)(1+\hat{\theta}_u^2)\,\rho = 0, \tag{4.2.50}$$

which yields

$$\hat{\rho}_q = \frac{\bar{t}}{\bar{t}-1}\cdot\frac{-\hat{\theta}_u}{1+\hat{\theta}_u^2} \approx \frac{\bar{t}c_{10}}{(\bar{t}-1)(c_{00}+c_{20})}.$$

From (4.2.34) we get

$$\hat{\theta}_q = \begin{cases} \dfrac{-(1+\hat{\theta}_u^2)+\sqrt{(1+\hat{\theta}_u^2)^2-4a^2\hat{\theta}_u^2}}{2a\hat{\theta}_u}, & \text{if} \quad \hat{\theta}_u \neq 0 \\ 0, & \text{if} \quad \hat{\theta}_u = 0, \end{cases} \tag{4.2.51}$$

where $a = -\bar{t}/(\bar{t}-1)$. The QLS estimate of σ^2 is obtained plugging in $\hat{\boldsymbol{\beta}}_q$, $\hat{\theta}_q$ in (4.2.43). The simulation result next is based on the solution without the approximation $c_{k0} \approx 0,\ k \geq 2$.

4.2.4 The asymptotic distribution

The asymptotic distribution is obtained when n goes to infinity while t_i's are held fixed. For simplicity, we assume that $t_i \equiv t$. The asymptotic distribution of the maximum likelihood estimates were obtained by finding the information matrix, while the distribution of the moment estimates and the quasi-least squares estimates were obtained by finding the distribution of $\overline{\mathbf{U}}$ and then applying the delta theorem. We will first find the asymptotic distribution of the maximum likelihood estimates. We will need the following results regarding the first and second derivatives of the log-likelihood (4.2.44) with respect to $\boldsymbol{\beta}, \theta$, and σ^2:

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \\ \frac{\partial l}{\partial \theta} &= -\frac{n}{2}\left[\frac{\partial|\mathbf{V}|/\partial\theta}{|\mathbf{V}|} + \frac{\partial \mathrm{tr}(\mathbf{V}^{-1}\overline{\mathbf{U}})/\partial\theta}{\sigma^2}\right], \end{aligned}$$

$$
\begin{aligned}
\frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2}\left[\frac{\bar{t}}{\sigma^2} - \frac{\mathrm{tr}(\mathbf{V}^{-1}\overline{\mathbf{U}})}{\sigma^4}\right],\\
\frac{\partial^2 l}{\partial \boldsymbol{\beta}^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}\mathbf{X}_i,\\
\frac{\partial^2 l}{\partial \theta^2} &= -\frac{n}{2}\left[\frac{\partial^2|\mathbf{V}|/\partial\theta^2}{|\mathbf{V}|} - \left(\frac{\partial|\mathbf{V}|/\partial\theta}{|\mathbf{V}|}\right)^2 + \frac{\partial^2\mathrm{tr}(\mathbf{V}^{-1}\overline{\mathbf{U}})/\partial\theta^2}{\sigma^2}\right],\\
\frac{\partial^2 l}{\partial (\sigma^2)^2} &= -\frac{n}{2}\left[\frac{2\mathrm{tr}(\mathbf{V}^{-1}\overline{\mathbf{U}})}{\sigma^6} - \frac{t}{\sigma^4}\right],\\
\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial\theta} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i'\frac{\partial\mathbf{V}^{-1}}{\partial\theta}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}),\\
\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial\sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}),\\
\frac{\partial^2 l}{\partial \sigma^2\partial\theta} &= \frac{n}{2\sigma^4}\frac{\partial\mathrm{tr}(\mathbf{V}^{-1}\overline{\mathbf{U}})}{\partial\theta},
\end{aligned}
$$

where

$$
\begin{aligned}
|\mathbf{V}| &= \sum_{k=0}^{t}\theta^{2k}\\
\frac{\partial|\mathbf{V}|}{\partial\theta} &= \sum_{k=1}^{t}2k\theta^{2k-1}\\
\frac{\partial^2|\mathbf{V}|}{\partial\theta^2} &= \sum_{k=1}^{t}2k(2k-1)\theta^{2k-2}\\
\mathrm{tr}\left(\frac{\partial\mathbf{V}^{-1}}{\partial\theta}\overline{\mathbf{U}}\right) &= -2\ \mathrm{tr}\left[\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\mathbf{V}^{-1}\overline{\mathbf{U}}\right]\\
\mathrm{tr}\left(\frac{\partial^2\mathbf{V}^{-1}}{\partial\theta^2}\overline{\mathbf{U}}\right) &= 8\ \mathrm{tr}\left[\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\mathbf{V}^{-1}\overline{\mathbf{U}}\right]\\
&\quad -2\mathrm{tr}(\mathbf{V}^{-1}\mathbf{V}^{-1}\overline{\mathbf{U}}).
\end{aligned}
$$

Since $\mathrm{E}(\overline{\mathbf{U}}) = \sigma^2\mathbf{V}$, we have

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 l}{\partial\boldsymbol{\beta}^2}\right) &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}\mathbf{X}_i,\\
\mathrm{E}\left(\frac{\partial^2 l}{\partial\theta^2}\right) &= -\frac{n}{2}\left[\frac{\partial^2|\mathbf{V}|/\partial\theta^2}{|\mathbf{V}|} - \left(\frac{\partial|\mathbf{V}|/\partial\theta}{|\mathbf{V}|}\right)^2\right.\\
&\quad \left. +8\ \mathrm{tr}\left[\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\right] - 2\ \mathrm{tr}(\mathbf{V}^{-1})\right],\\
\mathrm{E}\left(\frac{\partial^2 l}{\partial(\sigma^2)^2}\right) &= -\frac{nt}{2\sigma^4},\\
\mathrm{E}\left(\frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\sigma^2}\right) &= \mathrm{E}\left(\frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\theta}\right) = \mathbf{0},\\
\mathrm{E}\left(\frac{\partial^2 l}{\partial\theta\partial\sigma^2}\right) &= -\frac{n}{\sigma^2}\ \mathrm{tr}\left[\mathbf{V}^{-1}(\theta\mathbf{I}_t - \mathbf{C}_{10})\right].
\end{aligned}$$

Define $v_1 = \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{V}^{-1}\mathbf{X}_i\right)^{-1}$ and

$$\boldsymbol{\Sigma}_{2l}^{-1} = \frac{1}{n}\begin{bmatrix} -\mathrm{E}(\partial^2 l/\partial\theta^2) & -\mathrm{E}(\partial^2 l/\partial\theta\partial\sigma^2) \\ -\mathrm{E}(\partial^2 l/\partial\theta\partial\sigma^2) & -\mathrm{E}(\partial^2 l/\partial(\sigma^2)^2) \end{bmatrix},$$

then the asymptotic variance of $(\hat{\boldsymbol{\beta}}_l, \hat{\theta}_l, \hat{\sigma}_l^2)$ is given by

$$\mathbf{I}^{-1}(\boldsymbol{\beta}, \theta, \sigma^2) = \frac{1}{n}\ \boldsymbol{\Sigma}_l,$$

where $\boldsymbol{\Sigma}_l = \mathrm{diag}(v_1, \boldsymbol{\Sigma}_{2l})$.

In order to find the distribution of the moment and quasi-least squares estimates, we have to know the third and forth moments of the distribution. For convenience, we assume that the distribution is normal. The asymptotic variance of all three estimates of $\boldsymbol{\beta}$ is given by v_1. The efficiency of the estimates of σ^2 will be consistent with the efficiency of the estimates of θ. Thus we will only show the asymptotic distribution of the estimates of θ. Under normality assumption, we have

$$\varepsilon_i \sim N(\mathbf{0},\ \sigma^2\mathbf{V}),\quad i = 1, \ldots, n,$$

this implies

$$\begin{aligned}
\mathrm{E}(\mathrm{vec}(\mathbf{U}_i)) &= \sigma^2\,\mathrm{vec}(\mathbf{V}),\\
\mathrm{Cov}(\mathrm{vec}(\mathbf{U}_i)) &= 2\sigma^4\mathbf{V}\otimes\mathbf{V},\quad i = 1, \ldots, n.
\end{aligned}$$

Thus

$$\mathrm{vec}(\overline{\mathbf{U}}) \xrightarrow{d} N\left(\sigma^2(1+\theta^2)\mathrm{vec}(\mathbf{P}), \quad 2\sigma^4(1+\theta^2)^2\frac{\mathbf{P}\otimes\mathbf{P}}{n}\right), \quad as \quad n\to\infty.$$

Let $\mathbf{c} = (c_{00},\ c_{10},\ c_{20})' = \left(\mathrm{tr}(\overline{\mathbf{U}}),\ \mathrm{tr}(\mathbf{C}_{10}\overline{\mathbf{U}}), \mathrm{tr}(\mathbf{C}_{20}\overline{\mathbf{U}})\right)'$, then

$$\mathbf{c} \xrightarrow{d} N\left(\boldsymbol{\mu}_c,\ \frac{1}{n}\boldsymbol{\Sigma}_c\right), \quad as \quad n\to\infty, \tag{4.2.52}$$

where $\boldsymbol{\mu}_c = \sigma^2(1+\theta^2)(t,\ (t-1)\rho,\ 0)'$ and $\boldsymbol{\Sigma}_c = 2\sigma^4(1+\theta^2)^2\{v_{ij}\}$ is symmetric with

$$\begin{aligned}
v_{11} &= \mathrm{tr}(\mathbf{P}\cdot\mathbf{P}) = t + 2(t-1)\rho^2, \\
v_{12} &= \mathrm{tr}(\mathbf{P}\cdot\mathbf{C}_{10}\cdot\mathbf{P}) = 2(t-1)\rho, \\
v_{22} &= \mathrm{tr}(\mathbf{C}_{10}\mathbf{P}\cdot\mathbf{C}_{10}\mathbf{P}) = \frac{1}{2}(t-1) + (3t-5)\rho^2, \\
v_{13} &= \mathrm{tr}(\mathbf{P}\cdot\mathbf{C}_{20}\cdot\mathbf{P}) = (t-2)\rho^2, \\
v_{23} &= \mathrm{tr}(\mathbf{C}_{10}\mathbf{P}\cdot\mathbf{C}_{20}\mathbf{P}) = (t-2)\rho \\
v_{33} &= \mathrm{tr}(\mathbf{C}_{20}\mathbf{P}\cdot\mathbf{C}_{20}\mathbf{P}) = \frac{t-2}{2}(1+\rho) + \frac{\rho^2 t(t+2)}{8} - [\frac{3\rho^2}{8}, \quad \text{if } t \quad \text{is odd}].
\end{aligned}$$

Here we used the results

$$\begin{aligned}
\mathrm{tr}(\mathbf{C}_{10}) &= \mathrm{tr}(\mathbf{C}_{10}^3) = 0, \\
\mathrm{tr}(\mathbf{C}_{10}^2) &= \frac{1}{2}(t-1), \\
\mathrm{tr}(\mathbf{C}_{10}^4) &= \frac{1}{4}(3t-5), \\
\mathrm{tr}(\mathbf{C}_{20}) &= \mathrm{tr}(\mathbf{C}_{20}\mathbf{C}_{10}) = \mathrm{tr}(\mathbf{C}_{20}^3) = \mathrm{tr}(\mathbf{C}_{20}\mathbf{C}_{10}^3) = 0, \\
\mathrm{tr}(\mathbf{C}_{20}^2) &= \frac{t-2}{2}, \\
\mathrm{tr}(\mathbf{C}_{20}\mathbf{C}_{10})^2 &= \frac{1}{32}t(t+2) - [\frac{3}{32}, \quad \text{if } t \quad \text{is odd}], \\
\mathrm{tr}(\mathbf{C}_{10}\mathbf{C}_{20}\mathbf{C}_{10}) &= \frac{1}{4}(t-2).
\end{aligned}$$

Recall that

$$\begin{aligned}
\hat{\rho}_m &= \frac{\bar{t}c_{10}}{(\bar{t}-1)c_{00}}, \\
\hat{\rho}_q &\approx \frac{\bar{t}c_{10}}{(\bar{t}-1)(c_{00}+c_{20})}.
\end{aligned}$$

This gives

$$
\begin{aligned}
\frac{\partial \hat{\rho}_m}{\partial \mathbf{c}} &= -\frac{t}{(t-1)c_{00}^2}\begin{pmatrix} c_{10} \\ -c_{00} \\ 0 \end{pmatrix}, \\
\frac{\partial \hat{\rho}_q}{\partial \mathbf{c}} &= -\frac{t}{(t-1)(c_{00}+c_{20})^2}\begin{pmatrix} c_{10} \\ -(c_{00}+c_{20}) \\ c_{10} \end{pmatrix},
\end{aligned}
$$

which implies

$$
\begin{aligned}
\frac{\partial \hat{\rho}_m(\boldsymbol{\mu}_c)}{\partial \mathbf{c}} &= -\frac{1}{\sigma^2(1+\theta^2)t(t-1)}\begin{pmatrix} (t-1)\rho \\ -t \\ 0 \end{pmatrix} \\
\frac{\partial \hat{\rho}_q(\boldsymbol{\mu}_c)}{\partial \mathbf{c}} &= -\frac{1}{\sigma^2(1+\theta^2)t(t-1)}\begin{pmatrix} (t-1)\rho \\ -t \\ (t-1)\rho \end{pmatrix}.
\end{aligned}
$$

Using the delta theorem, we get

$$
\begin{aligned}
\hat{\rho}_m &\xrightarrow{d} N\left(\rho, \ \frac{1}{n}\boldsymbol{\Sigma}_{\rho m}\right), \\
\hat{\rho}_q &\xrightarrow{d} N\left(\rho, \ \frac{1}{n}\boldsymbol{\Sigma}_{\rho q}\right), \quad as \quad n \to \infty,
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\rho m} &= \frac{1}{t^2(t-1)^2}\left[t^2(t-1) + 2t(t-3)\rho^2 + 4(t-1)^3\rho^4\right], \\
\boldsymbol{\Sigma}_{\rho q} &= \frac{1}{t^2(t-1)^2}\Big[t^2(t-1) - (3t^3 - 10t^2 + 9t + 2)\rho^2 + (t-1)^2(t-2)\rho^3 \\
&\quad + \frac{1}{4}(t-1)^2\rho^4\left\{(t^2 + 34t - 48) - [\,3, \ \text{ if } t \ \text{ is odd}]\right\}\Big].
\end{aligned}
$$

Given the estimate $\hat{\rho}$ with asymptotic mean ρ and variance $\boldsymbol{\Sigma}_\rho/n$, the estimate of θ is given by $\hat{\theta} = (-1 + \sqrt{1 - 4\hat{\rho}^2})/(2\hat{\rho})$, $\hat{\rho} \neq 0$. This gives

$$
\frac{\partial \hat{\theta}(\boldsymbol{\mu}_{\hat{\rho}})}{\partial \hat{\rho}'} = \frac{1}{2\rho^2}\left(1 - \frac{1}{\sqrt{1-4\rho^2}}\right).
$$

Applying the delta theorem again, we get,

$$
\hat{\theta} \xrightarrow{d} N\left(\hat{\theta}(\boldsymbol{\mu}_{\hat{\rho}}), \ \frac{\partial \hat{\theta}(\boldsymbol{\mu}_{\hat{\rho}})}{\partial \hat{\rho}'}\left(\frac{\boldsymbol{\Sigma}_\rho}{n}\right)\frac{\partial \hat{\theta}(\boldsymbol{\mu}_{\hat{\rho}})}{\partial \hat{\rho}}\right), \quad as \quad n \to \infty,
$$

that is

$$
\hat{\theta} \xrightarrow{d} N\left(\theta, \ \frac{1}{n}\, a\, \boldsymbol{\Sigma}_\rho\right), \quad as \quad n \to \infty,
$$

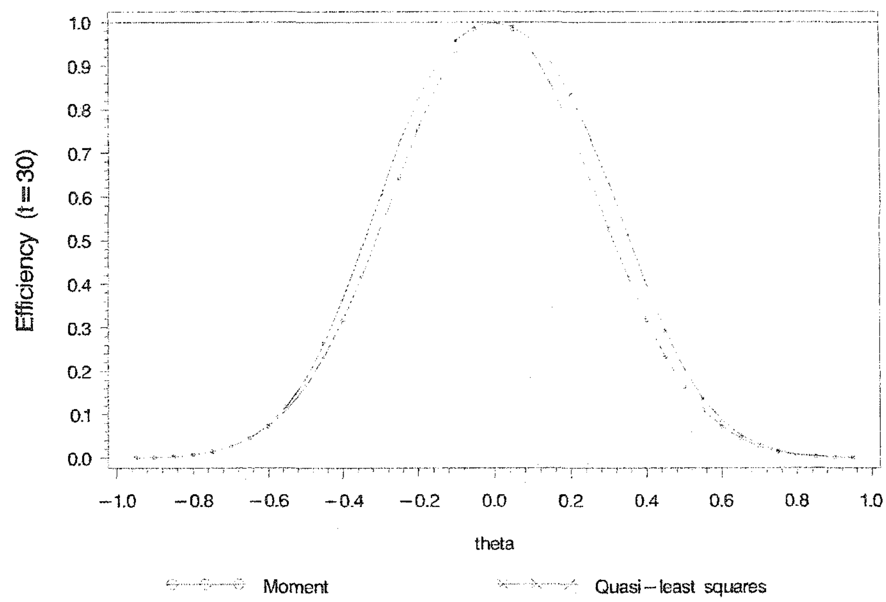

Figure 4.5. AREs of $\hat{\theta}_m$ and $\hat{\theta}_q$ when the data is normal.

where $a = \left(1 - 2\rho^2 - \sqrt{1-4\rho^2}\right)/(2\rho^4(1-4\rho^2))$. Thus the asymptotic distributions of $\hat{\theta}_m$ and $\hat{\theta}_q$ are given by

$$\begin{aligned} \hat{\theta}_m &\xrightarrow{d} N\left(\theta, \ \frac{1}{n}\Sigma_{2m}\right), \\ \hat{\theta}_q &\xrightarrow{d} N\left(\theta, \ \frac{1}{n}\Sigma_{2q}\right), \quad as \quad n \to \infty, \end{aligned}$$

where $\Sigma_{2m} = a\ \Sigma_{\rho m}$ and $\Sigma_{2q} = a\ \Sigma_{\rho q}$. The distribution of the exact quasi-least squares estimates is not easy to find. We present the distribution of an approximation here. In the next section we use simulations for comparisons with the exact quasi-least squares estimates. We first make theoretical comparisons of the moment, the maximum likelihood and the approximate quasi-least squares estimates. Figure 4.5 shows the plot of the AREs of $\hat{\theta}_q$ and $\hat{\theta}_m$ when $t = 30$, and Figure 4.6 gives the 3D plot of the ARE of $\hat{\theta}_q$. We see that the maximum likelihood estimates are the best, while the moment estimates are the worst. It is not surprising that the efficiency of the approximate quasi-least squares estimates are close to that of the moment estimates, since the estimates differ only by c_{20} in the denominator. But

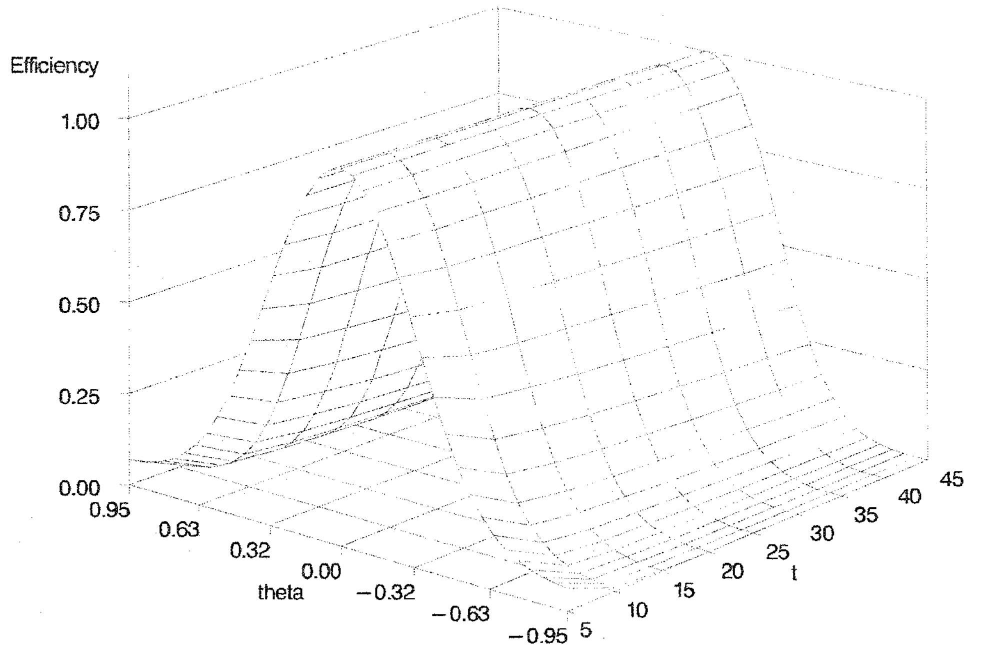

Figure 4.6. ARE of $\hat{\theta}_q$ when the data is normal.

as we will see the next section, the exact quasi-least squares estimates are much better than the moment estimates.

4.2.5 Simulation Results

In this section we use simulations to make efficiency comparisons between the exact maximum likelihood estimates and the quasi-least squares estimates for the MA(1) model. We will use Newton-Raphson method to obtain both of those estimates. As before, we assume $\boldsymbol{\beta} = 0$ and $\sigma^2 = 1$ in our simulation, and concentrate on comparing the estimates of θ. We choose t ranging from 5 to 45 and θ ranging from -0.95 to 0.95. We first study the asymptotic properties of the estimates when the data is normal. We generate a t-dimensional vector ε whose elements are from a standard normal distribution, and then we let $y = \sigma^2 \mathbf{V}^{1/2} \varepsilon$ and repeat the process

Table 4.4. AREs of $\hat{\theta}_q$, and $\hat{\theta}_m$ (in parentheses) when the data is simulated from normal distribution.

θ	t=5	t=10	t=30
0.0	1.0248 (1.0335)	1.0033 (1.0170)	1.0024 (1.0220)
0.1	1.0193 (0.9734)	1.0064 (0.9743)	0.9993 (0.9602)
0.2	0.9960 (0.8512)	0.9793 (0.8309)	0.9965 (0.8398)
0.3	0.9199 (0.7034)	0.9603 (0.6622)	0.9920 (0.6720)
0.4	0.9082 (0.5149)	0.9253 (0.4605)	0.9872 (0.4638)
0.5	0.9296 (0.5315)	0.8189 (0.2781)	0.9293 (0.2500)
0.6	0.9433 (0.5412)	0.7203 (0.2300)	0.8397 (0.1225)
0.7	1.0536 (0.5369)	0.5433 (0.2106)	0.6454 (0.0733)
0.8	2.9930 (1.1240)	0.5980 (0.1637)	0.3615 (0.0525)
0.9	0.2242 (0.1039)	0.6916 (0.3481)	0.2280 (0.0199)
0.95	0.0335 (0.0184)	0.0612 (0.0257)	0.0065 (0.0168)

$n = 30$ times. The whole process is then repeated 10000 times. For each replication, we compute the moment, maximum likelihood and quasi-least estimates of θ, and then compute the mean square errors (MSEs). We use Newton-Raphson method to solve the exact ML equation and quasi-least squares estimating equation by setting the precision to be $1e-10$. If any estimate is not feasible, the whole record was deleted and hence excluded from the analysis. Define the ARE of $\hat{\theta}_q$ with respect to $\hat{\theta}_l$ as $e(\hat{\theta}_q;\ \hat{\theta}_l) = MSE_l/MSE_q$, and define the ARE of $\hat{\theta}_m$ with respect to $\hat{\theta}_l$ similarly. Table 4.4 gives the efficiencies for $t = 5,\ 10$ and 30. The efficiencies are symmetric about $\theta = 0$, thus, only the efficiencies when θ is non-negative are shown. The numbers in the parentheses are the AREs of $\hat{\theta}_m$. Figure 4.7 shows the plot of the AREs of $\hat{\theta}_q$ and $\hat{\theta}_m$ when $t = 30$. Figure 4.8 gives the 3D plot of the ARE of $\hat{\theta}_q$. We found out that the ARE of $\hat{\theta}_q$ are larger than that of $\hat{\theta}_m$. Some efficiencies are larger than one due to the approximation of the Newton-Raphson iteration. Theoretically, the efficiencies should be one when θ is 0. The bias plot is given by Figure 4.9 when $t = 30$. It is clear from the plot that the maximum

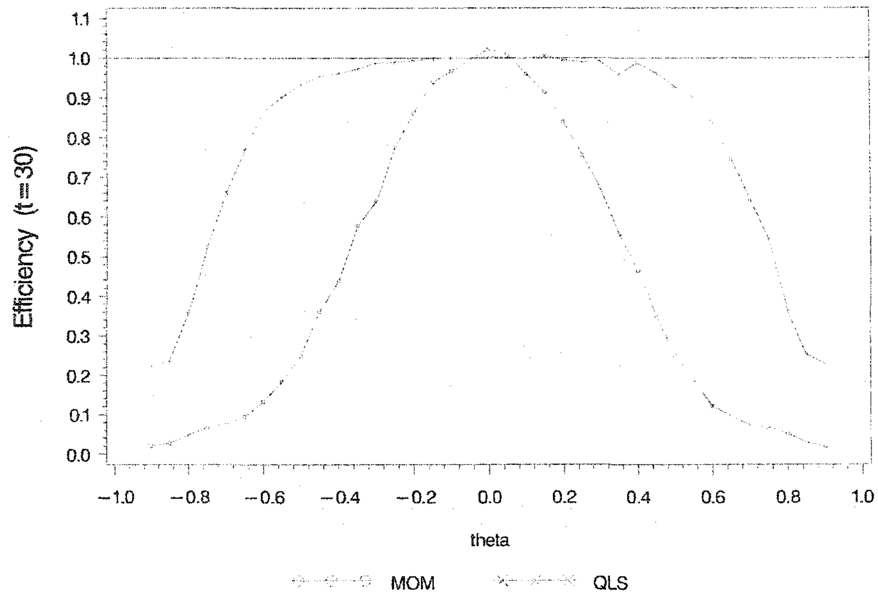

Figure 4.7. AREs of $\hat{\theta}_m$ and $\hat{\theta}_q$ when the data is simulated from normal distribution.

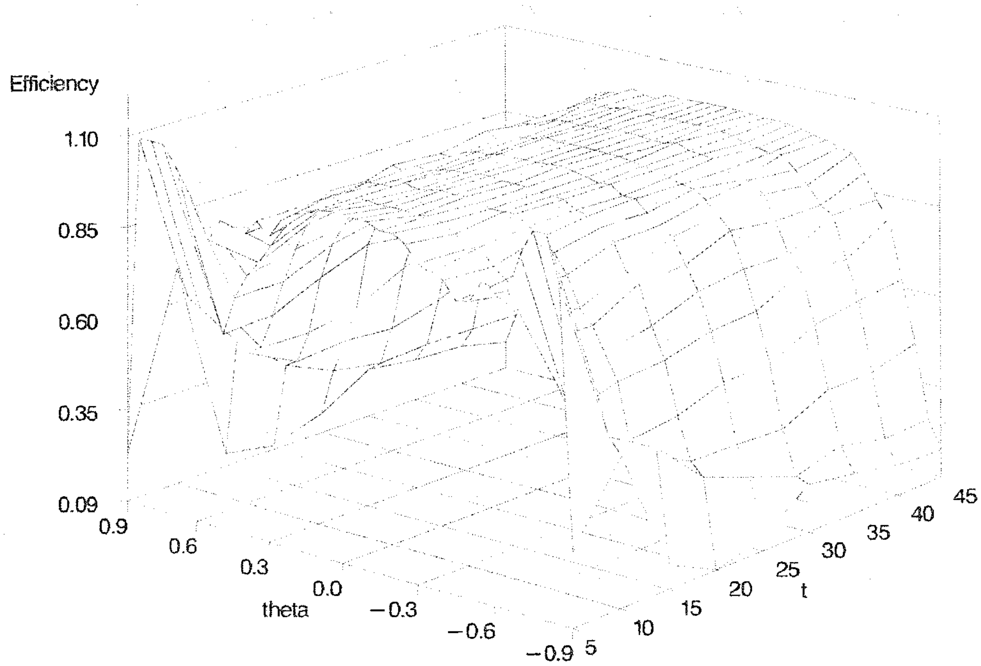

Figure 4.8. ARE of $\hat{\theta}_q$ when the data is simulated from normal distribution.

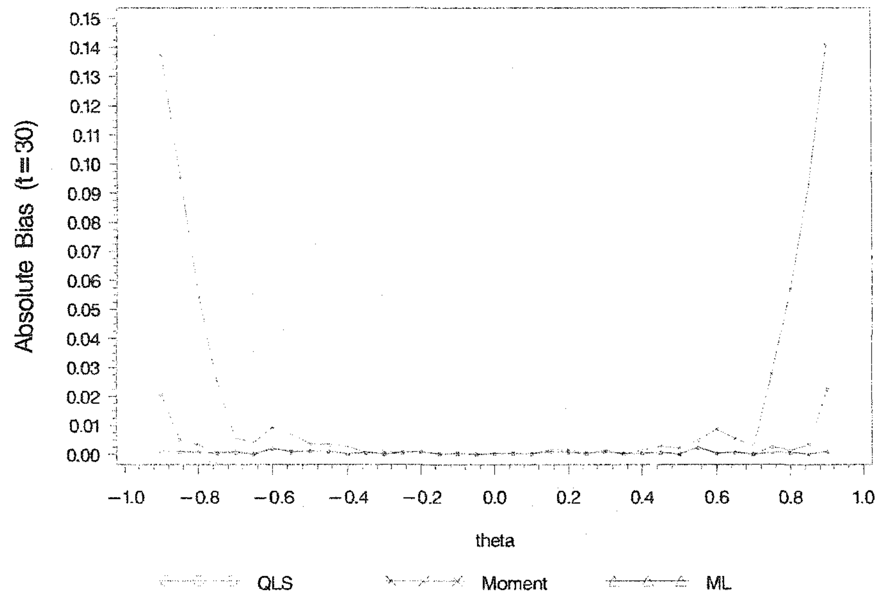

Figure 4.9. Biases of $\hat{\theta}_m$, $\hat{\theta}_q$ and $\hat{\theta}_l$ when the data is simulated from normal distribution.

likelihood estimates have the smallest bias, while the moment estimates have the most bias. Thus, the maximum likelihood estimates are the best based on either bias or mean square errors.

To further check the robust property of estimates, we simulate the data from the Student-t distribution with mean 0 and 5 degrees of freedom. We generate the samples as before but instead of normal we use the t-distribution. Table 4.5 gives the efficiencies for $t = 5,\ 10$ and 30. The efficiencies are symmetric about $\theta = 0$ and only the efficiencies when θ is non-negative are shown. The numbers in the parentheses are the AREs of $\hat{\theta}_m$. Figure 4.10 shows the plot of the AREs of $\hat{\theta}_q$ and $\hat{\theta}_m$ when $t = 30$. Figure 4.11 gives the 3D plot of the ARE of $\hat{\theta}_q$. We found out that the ARE of $\hat{\theta}_q$ are larger than that of $\hat{\theta}_m$, but still smaller than that of $\hat{\theta}_l$, although it is larger than 1 when θ is close to 1 and t is small. When θ is close to 1, the estimates become very unstable. We also tried contaminated

Table 4.5. AREs of $\hat{\theta}_q$, and $\hat{\theta}_m$ (in parentheses) when the data is simulated from Student-t distribution.

θ	t=5	t=10	t=30
0.0	1.0430 (1.0754)	1.0088 (1.0274)	1.0016 (1.0010)
0.1	1.0235 (0.9932)	1.0016 (0.9838)	1.0008 (0.9591)
0.2	0.9843 (0.8280)	0.9936 (0.8624)	0.9942 (0.8626)
0.3	0.9623 (0.7314)	0.9488 (0.6402)	0.9923 (0.6704)
0.4	0.9166 (0.5555)	0.9104 (0.4663)	0.9351 (0.4607)
0.5	0.8607 (0.5471)	0.7693 (0.2845)	0.8923 (0.2609)
0.6	0.9123 (0.6166)	0.6781 (0.2614)	0.7406 (0.1243)
0.7	1.3067 (0.7652)	0.5802 (0.2452)	0.5900 (0.0864)
0.8	1.7350 (0.8421)	0.5117 (0.1759)	0.2448 (0.0508)
0.9	0.2831 (0.1682)	0.3866 (0.1970)	0.0889 (0.0228)
0.95	0.1005 (0.0633)	0.1300 (0.0707)	0.0057 (0.0175)

normal distribution, but still observed the same pattern unfortunately. The bias plot is given by Figure 4.9 when $t = 30$. It is clear from the plot that the maximum likelihood estimates still have the smallest bias, while the moment estimates have the most bias. Thus, the maximum likelihood estimates are still the best based on either bias or mean square errors. But compared to moment estimates, the quasi-least squares estimates are much better.

Some discussion of the computational and programming issues is in order. The maximum likelihood estimates are not easy to compute numerically, even with a high-speed computer. For example, for the program to find the determinant of the matrix $\mathbf{V}$ and the determinant of the derivative matrix of $\mathbf{V}$ with respect to θ, the computer reported a "out of flow" error message. That is because the determinants are very large when t is large. I have to find the ratio of this two instead of finding each one separatively. Regarding the Newton-Raphson method, choosing the efficiency of convergence and the number of iteration is very important. Generally, if you want more efficiency, you have to let the program iterate more

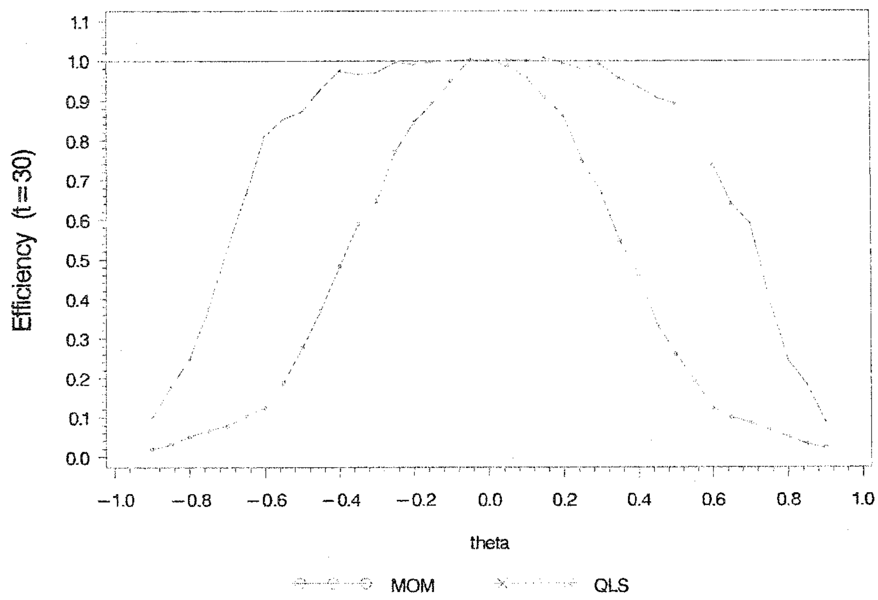

Figure 4.10. *AREs of* $\hat{\theta}_m$ *and* $\hat{\theta}_q$ *when the data is simulated from Student-t distribution.*

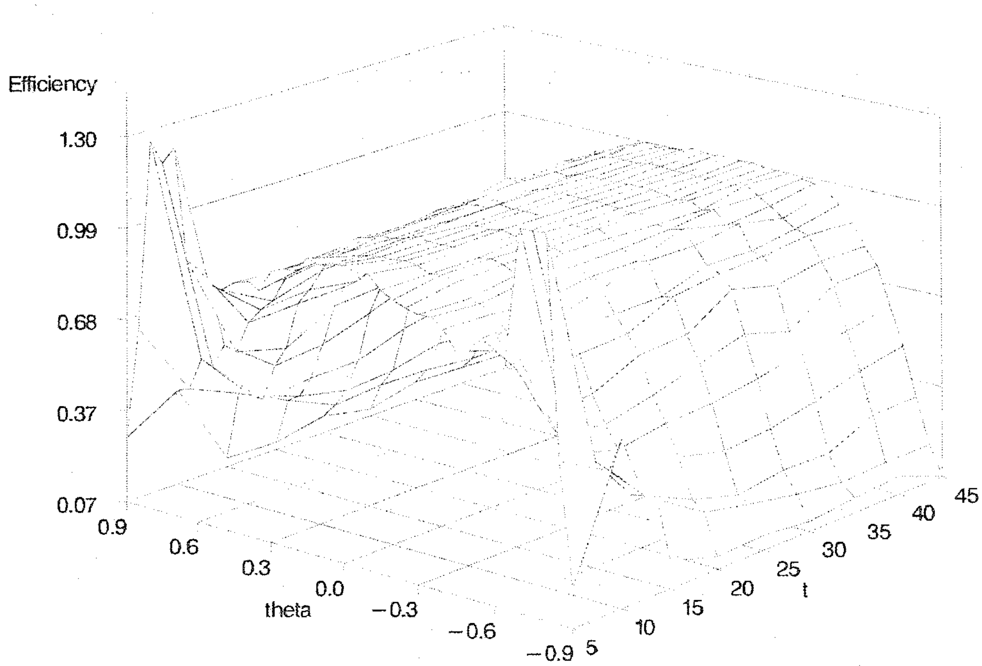

Figure 4.11. ARE of $\hat{\theta}_q$ *when the data is simulated from Student-t distribution.*

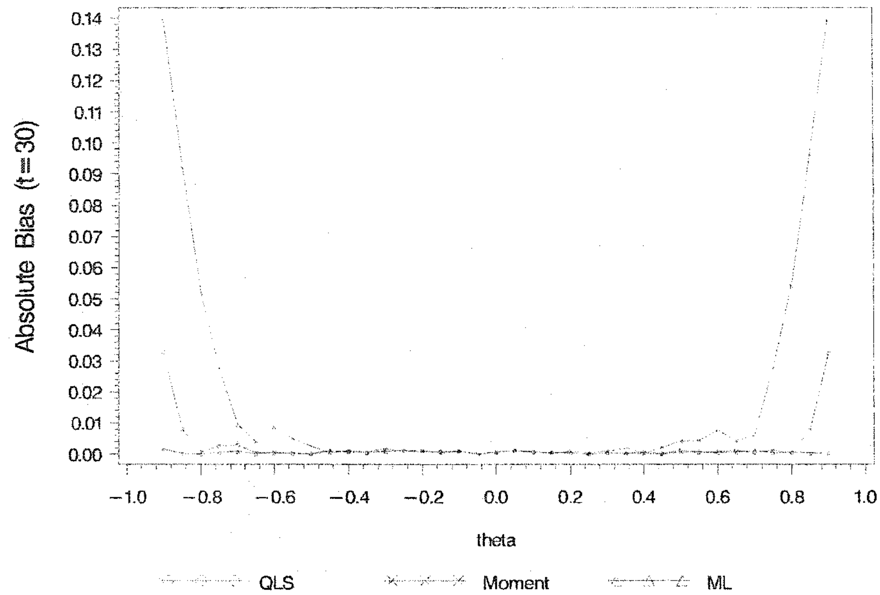

Figure 4.12. *Biases of $\hat{\theta}_m$, $\hat{\theta}_q$ and $\hat{\theta}_l$ when the data is simulated from Student-t distribution.*

times. For efficiency to be $1e-10$, the normal iteration is about 10. We set the maximum iteration to 30.

4.3 Summary

In this chapter, we studied the three estimating methods for the time series regression model with AR(p) and MA(1) errors. We also proved that the GEE methods including GEE1, GEE2 and EGEE reduce to either moment or maximum likelihood methods in this particular case. Specifically, GEE1 can be reduced to moment estimates, and GEE2 and EGEE can be reduced to maximum likelihood estimates. This is true for the generalized ARMA (p, q) model. Since we assume that the covariance matrix does not depend on the regression parameter, the issue of misspecification does not arise. Theoretical and simulation results show that the maximum likelihood estimates are the best when the data is normal. For the AR(p), the quasi-least squares estimates may be more robust, for example, when the data is from a contaminated distribution. While, for the MA(1) case, the pattern is different, but the quasi-least squares estimates are still good competitors.

CHAPTER V

CONCLUSIONS

Until now, we have studied the time series regression model with AR(p) and MA(1) error. It is essential to obtain an efficient estimate of the autoregressive and moving average parameters, since a more efficient estimate of the autoregressive and moving average parameter will result a more efficient regression (β) and scale (σ^2) parameter estimates. We found that the maximum likelihood estimates are best when the data is normal, and the quasi-least squares estimates are good competitors. We have not applied the methods to the model with MA(q) or mixed errors yet, though these parts are more complicated and challenging. We see that even in MA(1) case, the application of the moment, maximum likelihood and quasi-least squares methods are very complicated. The method we used in MA(1) case are actually generalized result for ARMA(p, q) model, except for moment method. A quadratically convergent process was suggested by Box *et al.* (1994) (Chap.6, p.221) for the moment method. Quasi-least squares methods second step also needs be modified a little bit according to different model. It's no doubt that the efficiencies of the estimates will behavior the same as the models we just studied.

Sometimes, the data may be grouped due to the nature of the study. It is desirable to fit the same order autoregressive-moving average model but with different parameters for each group. Suppose that there are K treatment group totally, and for the k^{th} group, $k = 1, \ldots, K$, there are n_k replications, for the i^{th} replication, there are t_{kn_ki} repeated measurements. The model may be expressed in matrix notation as

$$\mathbf{y}_{ki} = \mathbf{X}_{ki}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{ki}, \quad i = 1, \ldots, n_k, \;\; k = 1, \ldots, K,$$

where $\boldsymbol{\varepsilon}_{ki}$ is an ARMA(p, q) process with parameter $\boldsymbol{\lambda}_k$. We will reserve the notations introduced before. The adjusted sum of square errors become

$$\begin{aligned} S(\boldsymbol{\beta}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K) &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} \boldsymbol{\varepsilon}_{ki}'\mathbf{V}_{ki}^{-1}\boldsymbol{\varepsilon}_{ki} \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} (\mathbf{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta})'\mathbf{V}_{ki}^{-1}(\mathbf{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta}). \end{aligned}$$

On taking the partial derivative of $S(\boldsymbol{\beta}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K)$ with respect to $\boldsymbol{\beta}$ and equating to $\mathbf{0}$, we obtain

$$\hat{\boldsymbol{\beta}}_g = \left(\sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathbf{X}_{ki}'\mathbf{V}_{ki}^{-1}\mathbf{X}_{ki}\right)^{-1} \sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathbf{X}_{ki}'\mathbf{V}_{ki}^{-1}\mathbf{y}_{ki}.$$

The covariance is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 \left(\frac{1}{\sum_{k=1}^{K} n_k} \sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathbf{X}_{ik}'\mathbf{V}_{ik}^{-1}\mathbf{X}_{ik}\right)^{-1}.$$

The estimate of σ^2 is hence given by

$$\hat{\sigma}_g^2 = \frac{1}{\sum_{k=1}^{K}\sum_{i=1}^{n_k} t_{kn_k i}} \, S(\boldsymbol{\beta}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K).$$

The estimates of $\boldsymbol{\lambda}_k$'s need to modified accordingly. It is noted that this generalized model has more application than the regular one. For example, in the dental study as presented in Chapter II and III, we could have fitted the model with different autoregressive parameter for each group, since that is more acceptable.

When the series fitted turn out to be unstable, it is convenient to take the first or second order difference of the data and then fit the model again. Thus we could generalize the methods to the ARIMA(p, q) model. Not just generalizing to the time series model, we could generalize the methods to the growth curve model (see Chaganty (2003)) or nonlinear model. Basically, the methods are just like generalized least squares method, whenever you need to estimate of the parameters involved in a regression model, you can use all the methods. One important question remains: which one would you pick? Well, it depends on the situation. When you choose one method, always remember that there are some backups, and keep in mind that you can verify your results by applying other methods.

REFERENCES

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: Wiley.

Anderson, T. W. (1975), "Moving Average Processes," *The Statistician*, 24, 283–297.

Anderson, T. W. (1978), "Repeated Measurements on Autoregressive Processes," *Journal of the American Statistical Association*, 73, 371–378.

Box, George E. P., Jenkins, Gwilym M., and Reinsel, Gregory C. (1994), *Time Series Analysis: Forecasting and Control* (3rd ed.), Englewood Cliffs, New Jersey: Prentice-Hall, Inc..

Carnahan, B. Luther, H. A., and Wilks, J. O. (1969), *Applied Numerical Methods*, New York: Wiley.

Chaganty, N. Rao (1997), "An Alternative Approach to the Analysis of Logitudinal Data Via Generalized Estimating Equations," *Journal of Statistical Planning and Inference*, 63, 39–54.

Chaganty, N. Rao (2003), "Analysis of the Growth Curve Model Using Quasi-least Squares," *to appear.*

Chaganty, N. Rao, and Dayanand N. Naik (2002), "Analysis of Multivariate Longitudinal Data Using Quasi-least Squares," *Journal of Statistical Planning and Inference*, 103, 421–436.

Chaganty, N. Rao, and Shults, Justine (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.

Cheang, Wai-Kwong, and Reinsel, Gregory C. (2000), "Bias Reduction of Autoregressive Estimates in Time Series Regression Model Through Restricted Maximum Likelihood," *Journal of the American Statistical Association*, 95, 1173–1184.

Choudhury, Askar H., Hubara, Robert, and Louis, D. St.(1999), "Understanding Time-Series Regression Estimators," *The American Statistician*, 53(4), 342–348.

Crowder, M. J., and Hand, D. J. (1990), *Analysis of Repeated Measures*, London: Chapman and Hall.

Feigin, Paul D., Kratz, Marie F., and Resnick, Sidney I. (1996), "Parameter Estimation for Moving Averages With Positive Innovations," *The Annals of Applied Probability*, 6, 1157–1190.

Fujikoshi, Y., Kanda, T., and Tanimura, N. (1990), "The Growth Curve Model With an Autoregressive Covariance Structure," *Annals of the Institute of Statistical Mathematics*, 42(3), 533–542.

Godambe, V.P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208–1211.

Haddad, John N. (1995), "On the Closed Form of the Likelihood Function of the First Order Moving Average Model," *Biometrika*, 82, 232–4.

Hall, D. B., and Severini, T. A. (1998)", "Extended Generalized Estimating Equation for Clustered Data," *Journal of American Statistical Association*, 93, 1365–1375.

Harvey, A.C., and Phillips, G.D.A. (1995), "Maximum Likelihood Estimation of Regression Models With Autoregressive-moving Average Disturbances," *Biometrika*, 66, 49–58.

Hasza, David P. (1980), "A Note on Maxumum Likelihood Estimation for the First-Order Autogressive Process," *Communications in Statistics*, A9(13), 1411–1415.

Jennrich, R.I., Schluchter, M.D. (1986), "Unbalanced Repeated-Measures Models With Structured Covariance Matrices," *Biometrics*, 42, 805–820.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, London: Chapman and Hall.

Laird, Nan, Lange, Nicholas, and Stram, Daniel (1987), "Maximum Likelihood Computations With Reapted Measures: Application of the EM Algorithm," *Journal of the American Statistical Association*, 82, 97–105.

Lee, Elisa T. (1992), *Statistical Methods for Survival Data Analysis* (2nd ed.), New York: John Wiley & Sons.

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biomedical*, 73, 13–22.

Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992), "Multivariate Regression Analyses for Categorical Data (With Discussion)," *Journal of Royal Statistical Society*(B), 45, 3–40.

Lindsey, J. K. (1993), *Models for Repeated Measurements*, Oxford: Oxford University Press.

Potthoff, R.F., and Roy, S.N. (1964), "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems," *Biometrika*, 51, 665–680.

Prentice, R. L., and Zhao, L. P. (1991), "Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses," *Biometrics*, 47, 825–839.

Ralston, A., and Wilf, H. (1967), *Mathematical Methods for Digital Computers*, New York: Wiley.

Rochon, James (1992), "ARMA Covariance Structures With Time Heteroscedasticity for Repeated Measures Experiments," *Journal of the American Statistical Association*, 87, 777–783.

Rochon, James, and Helms, Ronals W. (1989), "Maximum Likelihood Estimation for Incomplete Repeated Measures Experiments Under an ARMA Covariance Structure," *Biometrics*, 45, 207–218.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statics*, New York: John Wiley.

Shults, Justine, and Chaganty, N. Rao (1998), "Analysis of Serially Correlated Data Using Quasi-Least Squares," *Biometrics*, 54, 1622–1630.

APPENDIX I

The Newton-Raphson Method

The Newton-Raphson method (Ralston and Wilf (1967), Carnahan *et al.* (1969)) is a numerical iterative procedure that can be used to solve nonlinear equations. It can only find the real roots of equations. We discuss and illustrate the use of the method by first considering a single nonlinear equation and then a set of nonlinear equations.

Let $f(\xi) = 0$ be the equation to be solved for ξ. The Newton-Raphson method requires an initial estimate of ξ, say $\hat{\xi}_0$, such that $f(\hat{\xi}_0)$ is close to zero, preferably, and then the first approximate iteration is given by

$$\hat{\xi}_1 = \hat{\xi}_0 - f(\hat{\xi}_0)/f'(\hat{\xi}_0)$$

where $f'(\hat{\xi}_0)$ is the first derivative of $f(\xi)$ evaluated at $\xi = \hat{\xi}_0$. In general, the $(k+1)^{th}$ iteration or approximation is given by

$$\hat{\xi}_{k+1} = \hat{\xi}_k - f(\xi_k)/f'(\xi_k)$$

where $f'(\hat{\xi}_k)$ is the first derivative of $f(\xi)$ evaluated at $\xi = \hat{\xi}_k$. The iteration terminates at the k^{th} iteration if $f(\hat{\xi}_k)$ is close enough to zero or the difference between $\hat{\xi}_k$ and $\hat{\xi}_{k-1}$ is negligible. The stopping rule is rather subjective. Acceptable rules are that $f(\hat{\xi}_k)$ or $D = \hat{\xi}_k - \hat{\xi}_{k-1}$ is in the neighborhood of 10^{-6} or 10^{-7}.

The Newton-Raphson method can be extended to solve a system of equations with more than one unknown. Suppose that we wish to find values of $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d)'$ such that

$$\mathbf{f}(\boldsymbol{\xi}) = \mathbf{0},$$

this section was written based on Lee, E. (1992) appendix A where $\mathbf{f}(\boldsymbol{\xi}) = (f_1(\boldsymbol{\xi}), f_2(\boldsymbol{\xi}), \ldots, f_d(\boldsymbol{\xi}))'$. Let a_{ij} be the partial derivative of f_i with respect to ξ_j, that is $a_{ij} = \partial f_i / \partial \xi_j$. The matrix

$$\mathbf{J} = \begin{bmatrix} a_{11} & \cdots & a_{1d} \\ a_{21} & \cdots & a_{2d} \\ \vdots & & \vdots \\ a_{d1} & \cdots & a_{dd} \end{bmatrix}$$

is called the *Jacobian matrix.* Let the inverse of $\mathbf{J}$, denoted by $\mathbf{J}^{-1}$, be

$$\mathbf{J}^{-1} = \begin{bmatrix} b_{11} & \cdots & b_{1d} \\ b_{21} & \cdots & b_{2d} \\ \vdots & & \vdots \\ b_{d1} & \cdots & b_{dd} \end{bmatrix}.$$

And let $\boldsymbol{\xi}^k = (\xi_1^k, \xi_2^k, \ldots, \xi_d^k)'$ be the approximate root at the k^{th} iteration and let $\mathbf{f}^k$ be the corresponding values of the functions $\mathbf{f}$, that is, $\mathbf{f}^k = \mathbf{f}(\boldsymbol{\xi}^k)$ and $\mathbf{J}^{-1^k}$ be $\mathbf{J}^{-1}$ evaluated at $\boldsymbol{\xi}^k$. Then the next approximation is given by

$$\boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k - \mathbf{J}^{-1^k}\mathbf{f}^k. \tag{A.1}$$

Specifically, that is

$$\begin{aligned} \xi_1^{k+1} &= \xi_1^k - (b_{11}^k f_1^k + b_{12}^k f_2^k + \cdots + b_{1d}^k f_d^k) \\ \xi_2^{k+1} &= \xi_2^k - (b_{21}^k f_1^k + b_{22}^k f_2^k + \cdots + b_{2d}^k f_d^k) \\ &\vdots \\ \xi_d^{k+1} &= \xi_d^k - (b_{d1}^k f_1^k + b_{d2}^k f_2^k + \cdots + b_{dd}^k f_d^k). \end{aligned}$$

The iterative procedure begins with a preselected initial approximate $\xi_1^0, \xi_2^0, \ldots, \xi_d^0$, proceeds following $(A.1)$, and terminates either when $f_1, f_2, \ldots, f_d$ are close enough to zero or when differences in the ξ values at two consecutive iterations are negligible.

APPENDIX II

A SAS Program

```
*****************************************************************
*    This program simulates the data from a Student-t distribution,   *
*    assuming that the data is balanced and the error is an AR(1)     *
*    process. First, It will compute the estimates of φ and σ²,       *
*    using moment, maximum likelihood and quasi-least squares         *
*    methods, respectively, and then calculate the mean square errors *
*    (MSE) of all three estimates for each simulation. Finally the    *
*    plots showing the comparisons of the efficiencies are generated. *
*****************************************************************;

proc iml;

sigsq = 1;                  * real value of σ²;
    n = 30;                 * no. of replications;
 nsim = 10000;              * no. of simulation;
 seed = 843623;             * random number seed;

* generate the data from a Student-t distribution with degrees of freedom
  specified by the main program;
start rant(df, seed);
  z = tinv(ranuni(seed), df) / sqrt( df/(df-2) );
return(z);            finish rant;

* calculate σ²;
start sigsq(phi, c00, c10, c11, t);
  sigsq = (c00 + c11*phi*phi - 2*c10*phi) / t;
return(sigsq);        finish sigsq;

* output the efficiencies to the file 'data.dat';
filename result 'data.dat';

do   t = 5     to 55   by 5;       * t: no. of repeated measurements;
do phi = -0.98 to 0.98 by 0.02;    * phi: the real value of φ;

   * keep tracking the program;
   file log; put 'Note!  Executing ...  ' t phi; closefile log;
```

```
* calculate the correlation matrix;
phii = 1;
do k = 1 to t-1;
      phii = phii // ( phii[k] * phi );
end;
corr = toeplitz(phii);

* define the matrices needed for calculating c10, c00, c11;
D11 = 0 || j(1, t-2, 1) || 0;   D11 = diag(D11);
D10 = toeplitz( 0 1||repeat(0, 1, t-2) ) / 2.0;

* initial MSEs, '1's mean the MSEs of σ^2 estimates;
mserl = 0;   mserl1 = 0;
mserq = 0;   mserq1 = 0;
mserm = 0;   mserm1 = 0;

do s = 1 to nsim;

   * create a sample from the t-distribution with degrees of freedom 5;
   data = j(n, t, 0);
   do i = 1 to n; do j = 1 to t;
        data[i, j] = rant(5, seed);
   end; end;

   * Cholesky decomposition of the covariance matrix, root function
     creates an upper triangular matrix;
   g = root( corr*sigsq / (1 - phi**2) );

   * 'error' simulated is an AR(1) process with parameter φ;
   error = data * g;

    ubar = error' * error / n;
     c00 = trace(ubar);
     c11 = trace( D11*ubar );
     c10 = trace( D10*ubar );


   a = sqrt((t-2)**2*c10**2+3*(t-1)*c00*c11+3*t*(t-1)*c11**2);
   b = c10 * (2*(t-2)**3*c10**2-9*t*(t-1)*(2*t-1)*c11**2
                                        + 9*(t-1)*(t-2)*c00*c11);
```

```
      alpha = arcos(1/2) + arcos( a**(-3)*b / 2 ) / 3;
      phil = (1/(3*(t-1)*c11)) * ((t-2)*c10 - 2*a*cos(alpha));
      * MLE of φ;

      phim = t*c10 / ( (t-1)*c00 );       * MOM estimate of φ;
      phiq = (t-2)*c10 / ( (t-1)*c11 ); * QLS estimate of φ;

      sigsql = sigsq(phil, c00, c10, c11, t); * MLE of σ^2;
      sigsqm = sigsq(phim, c00, c10, c11, t); * MOM estimate of σ^2;
      sigsqq = sigsq(phiq, c00, c10, c11, t); * QLS estimate of σ^2;

       mserl = mserl + ssq( phil-phi);
       mserq = mserq + ssq( phiq-phi);
       mserm = mserm + ssq( phim-phi);
      mserl1 = mserl1 + ssq( sigsql-sigsq );
      mserq1 = mserq1 + ssq( sigsqq-sigsq );
      mserm1 = mserm1 + ssq( sigsqm-sigsq );
   end;

   * asymptotic relative efficiencies;
   effq = mserl / mserq;     effq1 = mserl1 / mserq1;
   effm = mserl / mserm;     effm1 = mserl1 / mserm1;

   file result;
   put    t 5.0    phi 10.2    sigsq 8.1    effq 12.7    effm 12.7
          effq1 12.7    effm1 12.7;
end; end;

quit;

* input the asymptotic relative efficiencies;
data data;
      infile 'data.dat';
      input t phi sigsq effq effm effq1 effm1;
run;

goptions     reset = global gaccess = gsasfile gunit = pct
             htitle = 6        htext = 3        autofeed
            vorigin = 0in    horigin = 0in
              ftext = swiss    ftitle = swissb   cback = white
```

```
              hsize = 5.4in    vsize = 3.6in      dev = pslepsfc;

symbol1 i = join v = circle   c = red;
symbol2 i = join v = X        c = green;
symbol3 i = join v = triangle c = blue;
legend1 label = ( height = 1 " " );
  axis1 label = ( a = 90 h = 4.0 'Efficiency (t=10)' );

* ARE of QLS vs MLE;
filename gsasfile 'tq3d.ps';
proc g3d data = data;
     plot phi*t = effq/grid rotate = 50 tilt = 70
                         xticknum = 9 yticknum = 7 zticknum = 5;
run; quit;

data data1; set data;
     if t = 10;
     label effm = 'MOM' effq = 'QLS';
run;

* all AREs when t = 10;
filename gsasfile 't10.ps';
proc gplot data = data1;
   plot (effm effq)*phi/overlay legend=legend1 vaxis=axis1 vref=1;
run; quit;
```

VITA

Genming Shi

BAL 508, Department of Mathematics and Statistics,
Old Dominion University, Norfolk, Virginia 23529
Email: gshi@odu.edu

Education

PhD	*Computational Applied Math (Biostatistics)*, August 2003, Dept. of Math. and Stat., Old Dominion University
MS	*Computational Applied Math (Statistics)*, May 2001, Dept. of Math. and Statistics, ODU
BS	*Statistics*, July 1998, Department of Probability and Statistics, Beijing University, Beijing, China

Experience

Systems Analyst	NCI Information Systems Inc., Raytheon ConITS Team, Hampton, VA, May 2003
Teaching Assistant	Instructor, Dept. of Math. and Statistics, ODU, August 2001-May 2003
Teaching Assistant	Grader, Dept. of Math. and Statistics, ODU, 1998-2001
Statistician	Paragon Research Associates, Hampton, VA, 2000
Statistician	The Gallup Organization, Beijing, China, 1998

Publications

Chaganty, N. Rao and **Shi, G.** , "A note on the estimation of autocorrelation in repeated measurements" (under review)

Shi, G. and Chaganty, N. Rao, "Application of quasi-least squares to analyze replicated autoregressive time series model" (under review)

Professional Memberships

American Statistical Association   International Biometric Society (ENAR)

Typeset using LaTeX.