THE ANALYSIS OF REPEATED MEASUREMENT MODELS
UNDER NON-STANDARD DISTRIBUTIONAL ASSUMPTIONS

by

Hermí Boraine

Submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy in the subject Mathematical Statistics

in the Faculty of Science

University of Pretoria

Pretoria

May 1995

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am indebted to Professor Stephen Du Toit, my promoter, for the suggestion of the topic of this dissertation. It was indeed a privilege to work under his guidance and I have great respect for his commitment to research and his working capacity.

I would like to thank Professor Crowther and other colleagues for providing me with the equipment and other support conducive to the research and writing of this dissertation.

I am grateful to Ronel Schoeman and Rina Owen for providing me with technical assistance in preparing the manuscript, Professor De la Rey of the Psychology Department and Professor Lambrechts of the Graduate School of Management for providing me with data and Dr Trichardt of the Merensky Library for assistance in acquiring books and articles. I would also like to thank Chris Nagel who did the final editing.

This study would have been impossible without the continuous support and encouragement of my husband André and I gratefully acknowledge his contribution. I would also like to thank my son Bernard for his patience and my parents, parents in law, other family members and friends for their kind support.

Soli Deo Gloria

Notation

The following notation shall be adopted:

π : constant, π=3.14159

e : Euler's constant, e=2.71828

exp(x) : e^x, $-\infty < x < \infty$

*ln*x : natural logarithm of the real number x, x > 0

δ_{ij} : Kronecker's delta (=1 if i=j and =0 if $i \neq j$)

A:($p \times q$) : matrix of order $p \times q$

a:($p \times 1$) : column vector of order $p \times 1$

a : scalar

A' : transpose of A

a' : transpose of a (a row vector)

a_{ij} or $[A]_{ij}$: the element in the i-th row and j-th column of A

a_i or $[a]_{i1}$: the i-th element of a

A^{-1} : the inverse of A

$|A|$: determinant of A

$tr\ A$: trace of A

D_a : diagonal matrix with diagonal elements a_{11}, a_{22}, ...

$vecs[A]$: $\left[\dfrac{p(p+1)}{2} \times 1 \right]$ vector formed from the non-duplicate elements

   of the ($p \times p$) symmetric matrix A

$vec[A]$: ($pq \times 1$) vector formed from the q columns of the $p \times q$ matrix

   A

0 : null matrix

J_{ij} : matrix with all elements equal to zero with the exception of the
   element in the i-th row and j-th column which is equal to unity

J_{j1} : column vector with all elements equal to zero with the exception of the j-th element which is equal to unity

I_n : $n \times n$ identity matrix

$I_{n,r}$: $n \times r$ matrix with columns equal to the first r columns of I_n

$A \otimes B$: The right direct product or "Kronecker product" of matrices A and B defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdot & \cdot & \cdot & a_{1q}B \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{p1}B & a_{p2}B & \cdot & \cdot & \cdot & a_{pq}B \end{bmatrix}$$

$\dfrac{\partial A}{\partial x}$: matrix with typical element $\dfrac{\partial a_{ij}}{\partial x}$

$\dfrac{\partial a}{\partial x}$: column vector with typical element $\dfrac{\partial a_{i1}}{\partial x}$

$E(y):p \times 1$: expected value of the random vector y

$Cov(y,y'):(p \times p)$: covariance matrix of the random vector y with typical element

$$E\left[(y_i - E(y_i))(y_j - E(y_j))\right]$$

SUMMARY

TITLE:	THE ANALYSIS OF REPEATED MEASUREMENT MODELS UNDER NON-STANDARD DISTRIBUTIONAL ASSUMPTIONS
CANDIDATE:	HERMÍ BORAINE
PROMOTER:	PROFESSOR SHC DU TOIT
DEPARTMENT:	STATISTICS
DEGREE:	PHD (MATHEMATICAL STATISTICS)

In many experimental studies, repeated observations are made on each of a number of experimental units with the objective to fit a response curve to the data. Longitudinal data consist of repeated observations on many experimental units. It is reasonable to assume that although the response patterns of the different experimental units may differ, they can all be described by the same functional form. Differences in the response patterns between experimental units are modelled by allowing the parameters of the model to be stochastic. Linear as well as non-linear response functions are considered and it is assumed that the residuals of the models are generated by stationary autoregressive moving average (ARMA) processes.

The exact likelihood function of the observations of a random coefficient ARMA process is given as well as an approximation thereof based on numerical integration. It is shown that a Kalman recursive algorithm can be used in situations where the data is incomplete. The concept of marginal maximum likelihood estimation is discussed together with the use of the EM-algorithm to obtain maximum likelihood estimates. Bayes estimators of the coefficients of an ARMA process are given. It is shown how the Gibbs sampler can be used to calculate Bayes estimates.

Various models used to describe repeated measurement data are considered. It is

assumed that the error terms of these models are generated by an ARMA process with fixed or random coefficients. In repeated measurement experiments more than one related characteristic is often measured at each time point. Vector ARMA models can be used to analyze the change in the response vector over time. It is shown that results applying to the scalar case can be generalized to deal with vectors of measurements.

Two distributions in the elliptical class are considered as alternatives to the normal distribution as probability models for the white noise of an ARMA process. The results of two simulation studies are given.

OPSOMMING

ONDERWERP:	DIE ONTLEDING VAN HERHAALDE METINGSMODELLE ONDERHEWIG AAN NIE-STANDAARD VERDELINGSAANNAMES
KANDIDAAT:	HERMÍ BORAINE
PROMOTOR:	PROFESSOR SHC DU TOIT
DEPARTEMENT:	STATISTIEK
GRAAD:	PHD (WISKUNDIGE STATISTIEK)

Herhaalde metings word dikwels in eksperimente op elk van 'n aantal eksperimentele eenhede gemaak met die doel om 'n responsfunksie by die data te pas. Longitudinale data bestaan uit herhaalde metings wat van 'n aantal eksperimentele eenhede verkry is. Die aanname word gemaak dat, hoewel die responspatrone van individue mag verskil, almal van dieselfde funksionele vorm is. Verskille in die responspatrone tussen eksperimentele eenhede word voor voorsiening gemaak deurdat die parameters van die modelle stogasties kan wees. Lineêre sowel as nie-lineêre responsfunksies word beskou en die aanname word gemaak dat die residue van die modelle gegenereer word deur stasionêre outoregressiewe bewegende gemiddelde (ARMA) prosesse.

Die beraming van die parameters van 'n stogastiese koëffisiënt ARMA model word beskou. Die aanneemlikheidsfunksie, tesame met 'n benadering daarvan, gebaseer op numeriese integrasie, word gegee. Daar word aangetoon dat 'n Kalman rekursiewe algoritme gebruik kan word om die aanneemlikheidsfunksie te bereken wanneer die data onvolledig is. Marginale maksimum aanneemlikheidsberaming word bespreek tesame met die gebruik van die EM-algoritme vir die berekening van maksimum aanneemlikheidsberamings. Bayes beremers word gegee vir die koëffisiënte van 'n ARMA proses. Daar word getoon hoe die Gibbs steekproefgenereerder gebruik kan

word om Bayes beramings te verkry.

Verskeie modelle wat gebruik kan word om herhaalde meting data te beskryf word beskou. Daar word aangeneem dat die foutterme van die modelle deur ARMA prosesse met vaste of stogastiese koëffisiënte gegenereer word. In herhaalde meting eksperimente bestaan die waarnemings dikwels uit meer as een eienskap wat gemeet word. Daar word aangetoon dat resultate van toepassing op die skalaargeval uitgebrei kan word om voorsiening te maak vir vektor ARMA prosesse.

Twee verdelings in die elliptiese klas word as alternatiewe vir die normaalverdeling gebruik om te dien as waarskynlikheidsmodelle vir die witruis van 'n ARMA proses en die resultate van twee simulasiestudies word gegee.

CHAPTER 1 PROPOSED RESEARCH

Repeated observations are often made in experimental studies on each of a number of experimental units with the objective to fit a response curve to the data. Longitudinal data consist of repeated observations on many experimental units. Jones (1991) draws a distinction between time series data and longitudinal data. Time series data consist of repeated observations on a single experimental unit. Methods for the analysis of time series data often require that the number of repeated observations be at least 50. The number of experimental units in longitudinal studies is usually far greater than the number of time points, and the number of time points is often less than 10 (Reichardt (1991)). In this study the focus falls on the analysis of longitudinal data.

In many practical applications it is reasonable to assume that although the response patterns of the different experimental units may differ, they can all be described by the same functional form. Differences in the response patterns between experimental units are modelled by allowing the coefficients of the model to be stochastic.

Linear and non-linear response functions are considered and it is assumed that the residuals of the models are generated by stationary autoregressive moving average (ARMA) processes with either fixed or random coefficients.

In Chapter 2 results that are used in the estimation of the parameters of repeated measurement models are given. Maximum likelihood estimation is discussed and results pertaining to the multivariate normal distribution with a structured covariance matrix are given. An iterative procedure which can be used to maximize the likelihood function that can be used in situations where the maximum likelihood equations can not be expressed in closed form is reviewed. The likelihood functions derived in the subsequent chapters are often given in terms of a multidimensional integral which can not be calculated with standard integration results. It is shown how numerical integration can be used to calculate these likelihood functions. Basic results pertaining to the elliptical class of distributions are introduced and the generation of samples from

1-1

this class is discussed.

In Chapter 3 results reported on in literature concerning the estimation of fixed coefficient ARMA models are discussed. These results are extended in subsequent chapters, to allow for different sets of model assumptions.

The likelihood function is given for an ARMA model with Gaussian white noise. The state-space representation of an ARMA process is also given and it is shown how the Kalman recursive algorithm can be employed to estimate the coefficients of the ARMA process in the case of a complete, as well as an incomplete data set.

An expression is derived for the covariance matrix of a stationary ARMA process which takes into account the state of the process before any observations were made. It is also shown how the likelihood function of the repeated measurements can be calculated without direct calculation of the inverse or determinant of the covariance matrix. A reparameterization of the ARMA coefficients which is used in subsequent chapters is also included.

In Chapter 4 the estimation of the parameters of a random coefficient ARMA model is considered.

The exact likelihood function of the observations of an ARMA process is given as well as an approximation thereof based on numerical integration. It is shown that a Kalman recursive algorithm can be used in situations where the data is incomplete. The results of a simulation study, which was carried out to examine the properties of the maximum likelihood estimates for complete as well as incomplete data sets, are given.

The concept of marginal maximum likelihood estimation is discussed together with the use of the EM-algorithm to obtain maximum likelihood estimates. It is also shown how generalized least squares estimates of the parameters, of the random coefficient ARMA model, can be obtained by using numerical integration.

1-2

Since the assumptions made in random coefficient models are similar to Bayes assumptions, the Bayes estimates of the coefficients of an ARMA process are given. It is shown how the Gibbs sampler can be used to calculate Bayes estimates. An example of the implementation of the Gibbs sampler based on simulated data is provided.

In Chapter 5 various models used to describe repeated measurement data are considered. It is assumed that the error terms of these models are generated by an ARMA process with fixed or random coefficients. The results of previous chapters are now incorporated in more general models allowing for linear and non-linear response functions over time with fixed or random parameters. The results are applied to a South African unit trust data set.

In Chapter 6 two distributions in the elliptical class are considered as alternatives to the normal distribution as probability models for the white noise of an ARMA process. The results of two simulation studies are given. These studies were carried out to examine the properties of the maximum likelihood estimators in situations where the kurtosis of the data differs from that of a normal distribution. The exact likelihood functions based on the multivariate t-distribution and the Pearson Type II distribution are given and the corresponding maximum likelihood estimation results are reported on in the simulation study.

In repeated measurement experiments more than one related characteristic is often measured at each time point. Vector ARMA models can be used to analyze the change in the response vector over time.

It is shown in Chapter 7 that results applying to the scalar case can be generalized to deal with vectors of measurements. An expression is derived for the covariance matrix of the observation vector, that takes into account information regarding the process before any observations were made. The likelihood function of the observations are given under the assumption that the white noise terms are independent multivariate normal vector variates.

1-3

It is shown that in certain circumstances, the covariance matrix of the observations can be decomposed into a Kronecker product of the covariance matrix of the variables and the covariance matrix of the observations over time. Conclusions can then be drawn regarding the relationships between the variables on the one hand and a possible common pattern relating to change over time on the other hand. A practical application involving psychometric test results is given.

Most of the theory discussed in this dissertation has been implemented using FORTRAN computer programs.

CHAPTER 2 GENERAL RESULTS

2.1 INTRODUCTION

This chapter introduces results which will be used in subsequent chapters.

The emphasis of this dissertation is on the estimation of unknown parameters in a repeated measurements time series or regression model, under different sets of distributional and model assumptions.

Repeated measurement data sets typically consist of a response measured repeatedly on each of a number of experimental units. Changes in the pattern of responses may occur over time and it is frequently possible to describe these changes by means of a response function which may be linear or non-linear in its parameters. When measurements are made over time, the errors are usually correlated, and it will be assumed that the error terms are generated by an autoregressive moving average (ARMA) process. The assumption is made that the parameters of the response function and ARMA process, of the different subjects, are independent realisations of a random vector, γ, from a common multivariate population. The problem addressed in this dissertation is the estimation of the parameters of the multivariate distribution of γ. The parameters of the response function and ARMA process are therefore not fixed over the experimental units but are random variables.

In Section 2.2 maximum likelihood estimation is discussed. Specific attention is given to the estimation of the elements of the mean vector and covariance matrix of the multivariate normal distribution.

As it is often not possible to express the maximum likelihood estimators of the unknown parameters of a random coefficient model in closed form, an iterative optimization algorithm is required to obtain these estimates. Such an algorithm is discussed in Section 2.3.

2-1

When dealing with random coefficient models, calculation of the value of multidimensional integrals is frequently necessary. In many instances, these integrals cannot be solved analytically and we have to resort to numerical integration techniques which are discussed in Section 2.4.

Simulation studies are used throughout this dissertation as a practical means to evaluate estimation techniques based on the derivation of new theoretical results. Section 2.5 is concerned with multivariate data simulation techniques. The elliptical class of distributions is defined and some important results are given.

2.2 ESTIMATION PROCEDURES

This section concentrates on the estimation of a model's parameters by using maximum likelihood estimation. Ordinary least squares and generalized least squares discrepancy functions are given. General results pertaining to maximum likelihood estimation are given, after which the special case of the multivariate normal distribution is discussed. The likelihood ratio test is discussed as a means of testing the fit of a model under the assumption of multivariate normality.

Suppose $y_1, \ldots, y_N$ is a set of independent random vectors from a population of which the distribution depends on $\gamma \in \Omega$ where γ is a k-dimensional vector of parameters and Ω is the parameter space. Our objective is to estimate γ.

Ordinary least squares (OLS), generalized least squares (GLS) and maximum likelihood (ML) estimation are well known classical estimation techniques which are commonly used. In all three cases a non-negative scalar valued discrepancy function, denoted by $F(\gamma)$, can be defined which has to be minimized with respect to γ. The point in Ω where the discrepancy function is a minimum is an estimate of γ.

A standard procedure for the determination of a global maximum or minimum point

2-2

involves finding the first order partial derivatives of F with respect to the components of γ. The estimator $\hat{\gamma}$ is the point in the parameter space where

$$\frac{\partial F(\hat{\gamma},\gamma)}{\partial\gamma}\Big|_{\gamma=\hat{\gamma}} = 0. \tag{2.2.1}$$

It should be kept in mind that the first order partial derivatives are not necessarily defined in the neighbourhood of $\hat{\gamma}$. The value of the discrepancy function must also be examined in all points where (2.2.1) is not defined and on the boundary of the parameter space for possible minima.

Suppose that the change in the response pattern over time may be represented by the following set of regression equations:

$$y_i = f(\theta_i, t) + e_i \tag{2.2.2}$$

where $f(.)$ is the response function, θ_i is the vector of parameters, t represents points in time when observations were made and e_i is the vector of error terms of subject i, $i = 1, \dots, N$.

The discrepancy functions used in OLS and GLS estimation are given by the following expressions respectively:

$$F_{OLS} = \sum_{i=1}^{N} [y_i - E(y_i)]' [y_i - E(y_i)] \tag{2.2.3}$$

and

2-3

$$F_{GLS} = \sum_{i=1}^{N} [y_i - E(y_i)]'[Cov(y_i, y_i')]^{-1}[y_i - E(y_i)] \tag{2.2.4}$$

where $E(y_i)$ is the mean vector and $Cov(y_i, y_i')$ the covariance matrix of y_i subject to the model (2.2.2).

Maximum likelihood estimation (see e.g. Bickel and Doksum (1977)) will be discussed in more detail as it is the estimation procedure which will be largely used in the remaining chapters.

Maximum likelihood estimation

The joint density function of y_1, ..., y_N, as a function of γ, for a given vector of outcomes y_1, ..., y_N is called the likelihood or likelihood function and is denoted by $L(y_1, ..., y_N, \gamma)$.

The maximum likelihood estimate (MLE) is the value of γ in Ω where the likelihood is a maximum. A MLE does not always exist and is not always unique. It is often more convenient to work with lnL, the logarithm of the likelihood function, called the log-likelihood. The MLE's will minimize the function $-lnL(y_1, ..., y_N, \gamma)$ and this function (or any multiple thereof) can be seen as a discrepancy function.

Let F be any scalar function of γ. The k x 1 vector of first order partial derivatives of F with respect to the components of γ is known as the gradient of F,

$$g = \frac{\partial F}{\partial \gamma}. \tag{2.2.5}$$

The gradient of the log-likelihood is called the scoring vector (Longford (1993)). If the

2-4

likelihood is a maximum at an interior point $\hat{\gamma}$ of the parameter space and the scoring vector is defined in the neighbourhood of this point, then $\hat{\gamma}$ is the root of the scoring vector. The ML equations can be written as

$$\frac{\partial lnL(y_1,\ldots,y_N,\gamma)}{\partial \gamma}\big|_{\gamma=\hat{\gamma}} = 0. \tag{2.2.6}$$

ML estimation consists of finding the roots of the scoring vector and examining the boundaries of the parameter space and points in the parameter space where the scoring vector is not defined, for possible locations of maxima. The solution to the likelihood equations can often not be expressed in closed form. An iterative optimization procedure may then be used to maximize the likelihood function. Such an algorithm is discussed in the next section.

MLE's have very desirable properties when certain regularity conditions are satisfied. Different sets of regularity conditions are discussed by Wasan (1970) and Bickel and Doksum (1977). The most important of the regularity conditions is that γ is an interior point of the parameter space Ω and that γ is identified (unique in Ω) (Browne (1991)).

If the regularity conditions hold and the sample size N is large, the distribution of $\hat{\gamma}$ is approximately multivariate normal with mean vector γ and covariance matrix I^{-1} (see e.g. Bickel and Doksum (1977)) where I is the information matrix defined as

$$I = -E[H(\gamma)] \tag{2.2.7}$$

and the $k \times k$ matrix of second order derivatives, $H(\gamma)$, is the Hessian matrix of lnL with typical element

$$[H(\gamma)]_{ij} = \frac{\partial^2 lnL}{\partial \gamma_i \partial \gamma_j}. \tag{2.2.8}$$

2-5

Multivariate normal distribution with structured covariance matrix

Suppose $y_1, \ldots, y_N$ is a random sample from an n-dimensional $N(\mu, \Sigma(\eta))$ distribution where μ is unknown and $\Sigma(\eta)$ is a structured matrix with unknown parameters η. The vector of parameters, γ, is given by

$$\gamma' = (\mu', \eta').$$

(2.2.9)

The results given in this section were obtained or adapted from those given by Browne (1991). Browne extended Bargmann's results (Bargmann (1967)) for parametric structures of Σ to parametric structures of Σ and μ.

The likelihood is the joint density of $y_1, \ldots, y_N$, given by

$$L(y_1, \ldots, y_N, \gamma) = (2\pi)^{-\frac{1}{2}Nn} |\Sigma|^{-\frac{N}{2}} exp[-\frac{1}{2}\sum_{i=1}^{N} (y_i - \mu)' \Sigma^{-1}(y_i - \mu)]$$

(2.2.10)

and the log-likelihood is therefore

$$lnL(y_1, \ldots, y_N, \gamma) = -\frac{1}{2}N(n\,ln2\pi + ln|\Sigma| + tr[\Sigma^{-1}G])$$

(2.2.11)

where G (n x n) is given by

$$G = \frac{(N-1)}{N}S + (\bar{y} - \mu)(\bar{y} - \mu)',$$

(2.2.12)

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (2.2.13)$$

and S is the sample covariance matrix defined by

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})(y_i-\bar{y})' . \qquad (2.2.14)$$

It can be shown (see e.g. Browne and Du Toit (1992)) that a typical element of the gradient of lnL is given by

$$\frac{\partial lnL}{\partial \gamma_i} = \frac{N}{2}trP\frac{\partial \Sigma}{\partial \gamma_i} + NtrR\frac{\partial \mu}{\partial \gamma_i} \qquad (2.2.15)$$

where

$$P = \Sigma^{-1}(G-\Sigma)\Sigma^{-1} \qquad (2.2.16)$$

and

$$R = (\bar{y}-\mu)'\Sigma^{-1}. \qquad (2.2.17)$$

From (2.2.15) to (2.2.17) it follows that the MLE's for μ and η can be obtained by finding the roots of

$$\frac{\partial lnL}{\partial \mu_i} = Ntr[RJ_{il}] \qquad (2.2.18)$$

2-7

and

$$\frac{\partial lnL}{\partial \eta_i} = \frac{N}{2} tr \left[P \frac{\partial \Sigma}{\partial \eta_i} \right]$$ (2.2.19)

respectively.

The elements of the inverse *information matrix* can be used to construct confidence intervals and tests of hypotheses on the population parameters. For large values of N, the inverse of the information matrix is an approximation of the covariance matrix of the estimates, $\hat{\gamma}$ (see e.g. Bickel and Doksum (1977)). A typical element of the information matrix is given by

$$[\dot{I}]_{ij} = N \left\{ tr \left[\frac{\partial \mu'}{\partial \gamma_i} \Sigma^{-1} \frac{\partial \mu}{\partial \gamma_j} \right] + \frac{1}{2} tr \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \gamma_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \gamma_j} \right] \right\}.$$ (2.2.20)

Test of model fit: multivariate normal distribution

In the case of the multivariate normal distribution with mean μ and structured covariance matrix, $\Sigma(\eta)$, a likelihood ratio test can be used to test the fit of a model. The null hypothesis states that the mean vector and covariance matrix have a specified structure,

$$H_0: \quad \mu = \mu(\gamma); \quad \Sigma = \Sigma(\gamma)$$ (2.2.21)

where γ is identified, against the general alternative

$$H_1: \quad \mu \text{ is any } n \times 1 \text{ vector}; \quad \Sigma \text{ is any non-negative } n \times n \text{ matrix}.$$ (2.2.22)

2-8

Suppose the MLE of γ under H_0 is denoted by $\hat{\gamma}$ so that

$$\hat{\mu} = \mu(\hat{\gamma}) \qquad (2.2.23)$$

and

$$\hat{\Sigma} = \Sigma(\hat{\gamma}) \qquad (2.2.24)$$

are the MLE's of μ and Σ under H_0. Under H_1 the MLE's of μ and Σ are (2.2.13) and

$$W = \frac{N-1}{N}S \qquad (2.2.25)$$

respectively where S is given by (2.2.14). The likelihood ratio test statistic is defined as

$$\lambda = \frac{L(\hat{\mu}, \hat{\Sigma})}{L(\bar{y}, W)}. \qquad (2.2.26)$$

If H_0 is true and if certain regularity conditions are satisfied, the limiting distribution of

$$-2ln\lambda = N\{ln|\hat{\Sigma}| - ln|W| + tr[\hat{\Sigma}^{-1}\hat{G}] - n\} \qquad (2.2.27)$$

where (cf. (2.2.12))

$$\hat{G} = G(\hat{\mu}) \qquad (2.2.28)$$

is the central chi-square distribution with degrees of freedom q, where q is the difference between the number of parameters estimated under the alternative and null hypotheses. If γ is k-dimensional, then $q = n + \frac{1}{2}n(n+1) - k$.

2.3 OPTIMIZATION

In situations where it is not possible to express the ML equations (2.2.6) in closed form, an iterative procedure is required to minimize the discrepancy function, $F(\gamma)$.

In Section 2.3 the optimization method developed by Browne and Du Toit (1992) is discussed. The basic ideas behind the algorithm are introduced by explaining Newton's method which is conceptually simple and forms the basis of many methods which were developed later.

The optimization method that will be described was developed by Browne and Du Toit (1992) and implemented in the computer program AUFIT (Du Toit and Browne (1982)). Their algorithm is based on the Fisher scoring algorithm (see e.g. Gill *et al* (1981)). In the case of a structured mean and covariance matrix, the algorithm may be regarded as a sequence of Gauss-Newton steps.

Suppose a discrepancy function $F(\gamma)$ needs to be minimized with respect to γ. The assumption is made that F is twice differentiable and can be approximated by a quadratic function in the neighbourhood of its minimum. The *Newton* optimization method will converge in a finite number of steps when the discrepancy function is quadratic. If the discrepancy function is not quadratic, it is adapted to converge reliably. A trial solution, γ_0, is chosen and its function value, $F(\gamma_0)$, first derivatives, g, and second derivatives, A, are all calculated in this point. The function $F(\gamma)$ can be approximated by

2-10

$$F(\gamma) \simeq F(\gamma_0) + g'(\gamma - \gamma_0) + \frac{1}{2}(\gamma - \gamma_0)'A(\gamma - \gamma_0). \tag{2.3.1}$$

The assumption is made that $F(\gamma)$ is reasonably well approximated by a second order Taylor series expansion. Let the gradient vector of the approximating function be denoted by $g(\gamma)$. Differentiation of (2.3.1) with respect to γ gives

$$g(\gamma) = g + A(\gamma - \gamma_0). \tag{2.3.2}$$

To find the next estimate of γ, $g(\gamma)$ is set equal to 0 and the equation is solved for γ. From (2.3.2) it follows that

$$g + A(\gamma - \gamma_0) = 0 \tag{2.3.3}$$

and hence

$$\gamma = \gamma_0 - A^{-1}g. \tag{2.3.4}$$

Suppose γ_{i-1} is the $(i-1)$-th solution in the Newton algorithm. The gradient vector, g_{i-1}, and the Hessian matrix, H_{i-1}, are evaluated in γ_{i-1}, and the next solution is

$$\gamma_i = \gamma_{i-1} - A_{i-1}^{-1}g_{i-1}. \tag{2.3.5}$$

According to Beale (1988) there are two basic problems with (2.3.5). The first is that there is no guarantee that $F(\gamma_i) < F(\gamma_{i-1})$. This is due to the fact that the step, $(A_{i-1})^{-1}g_{i-1}$, is based entirely on the behaviour of $F(\gamma)$ in the neighbourhood of γ_{i-1}. The method can be modified by incorporating a line search where the next solution is obtained from

$$\gamma_i = \gamma_{i-1} - \alpha_i A_{i-1}^{-1}g_{i-1} \tag{2.3.6}$$

2-11

where α_i is a positive scalar (sometimes referred to as a step size parameter) that is initially made one and repeatedly reduced until $F(\gamma_i) < F(\gamma_{i-1})$.

The second problem is that A_{i-1} is not necessarily positive definite. To overcome this problem, a constant can be added to the diagonal elements of A to make it positive definite. The next solution is

$$\gamma_i = \gamma_{i-1} - \alpha_i (A_{i-1} + \lambda I)^{-1} g_{i-1}. \tag{2.3.7}$$

Newton's method requires the calculation of second order derivatives. The algorithm performs well when the initial estimate γ_0 is close to the true minimum of F. The disadvantages of using the Hessian are that the expressions for its elements can be complicated, and may require a substantial amount of computer time for each iteration. It also tends to converge slowly or diverge if the initial approximation is poor (Browne (1982)).

The basic idea of quasi Newton or variable metric methods is to iteratively build up, a good approximation to the inverse Hessian matrix H^{-1} (Press $et\ al$ (1986)). The Davidon-Fletcher-Powell (DFP) and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithms are frequently implemented as updating schemes of the inverse Hessian matrix. The Fisher scoring algorithm uses the information matrix instead of the Hessian matrix (Agresti (1990)). For the multivariate normal distribution, the ij-th element of the information matrix is given by (2.2.20). At convergence, when N is large, the inverse information matrix can be used as an approximation of the covariance matrix of the estimators.

Suppose the parameters of a multivariate normal distribution are estimated and $-lnL$ is used as discrepancy function. A typical element of the information matrix, $\dot{I}$, is

2-12

$$[\dot{I}]_{ij} = -E\left[\frac{\partial^2 lnL}{\partial\gamma_i\partial\gamma_j}\right].$$

(2.3.8)

For the $N(\mu,\Sigma)$ distribution, $\dot{I}$ is given by (2.2.20) and the gradient can be calculated using (2.2.15).

The optimization algorithm requires a subroutine to calculate the value of $F(\gamma_k)$. Although code can be supplied for the calculation of exact derivatives, practical experience has shown that the algorithm is not sensitive to the use of numerical derivatives (Du Toit (1993)). Both (2.2.15) and (2.2.20) require the partial derivative of Σ with respect to γ_i. The approximation used by AUFIT is

$$\frac{\partial\Sigma}{\partial\gamma_j} \approx \frac{\Sigma(\gamma+\epsilon_j J_{jl})-\Sigma(\gamma)}{\epsilon_j}$$

(2.3.9)

where J_{jl} is a column vector with the only non-zero element a one in the j-th position and

$$\epsilon_j = \epsilon \cdot max\{1; ln|1+\gamma_j|\}$$

(2.3.10)

where ϵ is a small positive scalar (e.g. $\epsilon = 10^{-5}$).

The estimation of standard errors may be affected by the choice of ϵ, especially when the model is highly nonlinear.

Iteration of (2.3.6) is continued until some convergence criterion is met. A convergence criterion that is not affected by rescaling the data which is based on the residual cosine was given by Dennis (1977). The residual cosine of parameter j at the i-th iteration is defined as

2-13

$$c_j = \frac{[-g_i]_{jl}}{[A_i]_{jj}(F(\gamma_i))^{\frac{1}{2}}} \quad , \quad j = 1, \ldots, k. \qquad (2.3.11)$$

Convergence is attained if all absolute residual cosines fall below a certain tolerance limit (e.g. 10^{-3}) or if the value of the discrepancy function is smaller than a prescribed limit (e.g. 10^{-6}) within a number (say three) of consecutive iterations.

2.4 NUMERICAL INTEGRATION

The basic principles of numerical integration are discussed for the one-dimensional case and then extended for the multidimensional case. Results required in Chapters 4 and 5 are stated.

One-dimensional integrals

Suppose it is not possible to find a closed form solution for an integral of the form

$$\int_a^b W(x)f(x) \, dx \qquad (2.4.1)$$

where $W(x)$ is a given non-negative weight function on the interval $[a;b]$. $W(x)$ must be positive and continuous on $[a;b]$. The aim of Gaussian integration rules or quadrature formulas is to find a set of n abscissas (or nodes), $x_1, \ldots, x_n$ and weights $w_1, \ldots, w_n$ such that the approximation

$$\int_a^b W(x)f(x) \, dx \approx \sum_{i=1}^n w_i f(x_i) \qquad (2.4.2)$$

2-14

is exact when $f(x)$ is a polynomial. An introduction to the theory of Gaussian quadrature is given by Krylov (1962) and Stroud and Secrest (1966).

The theory of orthogonal polynomials is used to solve for both the weights and abscissas. The scalar product of two functions f and g over a weight function W is defined as

$$\langle f|g \rangle = \int_a^b W(x)f(x)g(x) \, dx. \tag{2.4.3}$$

The functions f and g are orthogonal with respect to W if their scalar product is zero. A function is said to be normalized if its scalar product with itself is one.

Let

$$p_j(x) = x^j + a_1 x^{j-1} + \ldots + a_j \tag{2.4.4}$$

be a normalized real polynomial of degree j.

A set of orthogonal polynomials can be constructed by the Gram-Schmidt orthogonalization process. For each weight function these polynomials are unique and defined by the following recursions:

$$p_0(x) = 1 \tag{2.4.5}$$

$$p_1(x) = \left[x - \frac{\langle xp_0|p_0 \rangle}{\langle p_0|p_0 \rangle} \right] p_0(x) \tag{2.4.6}$$

2-15

$$p_{i+1}(x) = \left[x - \frac{\langle xp_i|p_i \rangle}{\langle p_i|p_i \rangle} \right] p_i(x) - \left[\frac{\langle p_i|p_i \rangle}{\langle p_{i-1}|p_{i-1} \rangle} \right] p_{i-1}(x) \quad , i \geq 1. \tag{2.4.7}$$

Stoer and Bulirsch (1980) have proved that the roots of the orthogonal polynomial $p_n(x)$ are real and lie in the interval (a,b). These roots are the abscissas $x_1, \ldots, x_n$ of the n-point Gaussian quadrature formula (2.4.2). The weights $w_1, \ldots, w_n$ are the solutions of the following non-singular system of equations:

$$\begin{bmatrix} p_0(x_1) & p_0(x_2) & \cdots & p_0(x_n) \\ p_1(x_1) & p_1(x_2) & \cdots & p_1(x_n) \\ \vdots & \vdots & \vdots & \vdots \\ p_{n-1}(x_1) & p_{n-1}(x_2) & \cdots & p_{n-1}(x_n) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} \langle p_0|p_0 \rangle \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.4.8}$$

The weights will all be positive and

$$\int_a^b W(x)p(x) \, dx = \sum_{i=1}^n w_i p(x_i) \tag{2.4.9}$$

will be exact for all polynomials of degree $2n - 1$ or less. The proof of this result can be found in Stoer and Bulirsch (1980).

A few of the commonly used weight functions listed in Press *et al* (1986) are given in Table 2.4.1.

2-16

Table 2.4.1 Quadrature weight functions

Interval (a,b)	Weight function $W(x)$	Gauss-
$(-1;1)$	1	Legendre
$(-1;1)$	$(1-x^2)^{-\frac{1}{2}}$	Chebyshev
$(0;\infty)$	$x^c exp(-x)$	Laguerre
$(-\infty;\infty)$	$exp(-x^2)$	Hermite

Tabulations of the abscissas and weights for various weight functions can be found in standard references, e.g. Abramowitz and Stegun (1964).

The accuracy of the Gaussian quadrature formulas depends on how well the function $f(x)$ in (2.4.1) is approximated by a polynomial of degree $2n - 1$.

For a normal distribution the Gauss-Hermite quadrature formula is very useful. Unless otherwise stated, the weight function $exp(-x^2)$ is used when referring to numerical integration or Gaussian quadrature.

The following example illustrates the numerical integration technique by applying it to a simple problem.

Example 2.4.1

Suppose the value of the following integral has to be calculated by means of Gaussian quadrature:

2-17

$$I = \int_{-\infty}^{\infty} exp\left[-3\left(x-\tfrac{1}{2}\right)^2\right]\left(x^4-sinx\right)\, dx.$$

<div align="right">(2.4.10)</div>

The following transformation of x enables us to use the Gauss-Hermite weight function:

$$y = \sqrt{3}\left(x-\tfrac{1}{2}\right).$$

<div align="right">(2.4.11)</div>

The integral can be written in terms of the transformed variable as

$$I = \int_{-\infty}^{\infty} exp[-y^2]\left[\left(\sqrt{3}\,y+\tfrac{1}{2}\right)^4 - sin\left(\sqrt{3}\,y+\tfrac{1}{2}\right)\right]\sqrt{3}\; dy.$$

<div align="right">(2.4.12)</div>

A three-point quadrature formula will be sufficient to approximate (2.4.10). The orthogonal polynomials with respect to the weight function $W(y) = exp(-y^2)$ are derived from (2.4.5) to (2.4.7) and $p_3(y)$ is given by

$$p_3(y) = y(y^2 - \tfrac{3}{2}).$$

The abscissas are the roots of $p_3(y)$ namely $y_1 = -\sqrt{\tfrac{3}{2}}$, $y_2 = 0$ and $y_3 = \sqrt{\tfrac{3}{2}}$.

To find the weights, the system of equations (2.4.8) is solved:

<div align="center">2-18</div>

$$\begin{bmatrix} 1 & 1 & 1 \\ y_1 & y_2 & y_3 \\ y_1^2-\frac{1}{2} & y_2^2-\frac{1}{2} & y_3^2-\frac{1}{2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \sqrt{\pi} \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -\sqrt{\frac{3}{2}} & 0 & \sqrt{\frac{3}{2}} \\ 1 & -\frac{1}{2} & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sqrt{\pi} \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.295409 \\ 1.181636 \\ 0.295409 \end{bmatrix}$$

The weights are given by $w_1 = w_3 = 0.295409$ and $w_2 = 1.181636$.

(These abscissas and weights correspond with the tabulated values. For known weight functions such as the above, it is not necessary to perform the calculations manually.)

The integral (2.4.10) is approximated by

$$\sum_{i=1}^{3} w_i \cdot \sqrt{3} \left[\left(\sqrt{3}\,y_i + \tfrac{1}{2}\right)^4 - sin\left(\sqrt{3}\,y_i + \tfrac{1}{2}\right) \right]$$
$$= 27.097$$

using (2.4.2).

Multidimensional integrals

In situations where it is not possible to calculate the value of a multidimensional integral

2-19

analytically, it is sometimes possible to solve the problem through repeated (nested) application of one-dimensional Gaussian quadrature formulas. It can be accomplished by writing the multidimensional integral as a product type integral. In the context of the multivariate normal distribution, the region of integration is not bounded and the problem is simplified. In Section 4.3 the following three-dimensional integral has to be solved:

$$I(e) = \int \int \int |\Sigma(\eta)|^{-\frac{1}{2}} exp\left[-\frac{1}{2}e'\Sigma(\eta)^{-1}e\right] \cdot exp\left[-\frac{1}{2}(\eta-\eta_0)'\Psi^{-1}(\eta-\eta_0)\right] d\eta \qquad (2.4.13)$$

where the matrix $\Sigma(\eta)$ is a function of $\eta{:}3 \times 1$, a set of parameters. The following transformation of the vector η to the vector u is carried out to simplify the quadratic form:

$$u = \sqrt{\frac{1}{2}}\,\Psi^{-\frac{1}{2}}(\eta-\eta_0). \qquad (2.4.14)$$

The integral can be written as

$$I(e) = \int \int \int exp[-u'u]|\Sigma(u)|^{-\frac{1}{2}} exp\left[-\frac{1}{2}e'\Sigma(u)^{-1}e\right]|\sqrt{2}\,\Psi^{\frac{1}{2}}|\ du$$

$$(2.4.15)$$

$$= \int exp(-u_1^2) \int exp(-u_2^2) \int exp(-u_3^2)f(u)\ du \qquad (2.4.16)$$

where

$$f(u) = |\Sigma(u)|^{-\frac{1}{2}}|\sqrt{2}\,\Psi^{\frac{1}{2}}|\ exp\left[-\frac{1}{2}e'\Sigma(u)^{-1}e\right]. \qquad (2.4.17)$$

The weight function is Gauss-Hermite and the one-dimensional weights and abscissas

2-20

can be used to approximate (2.4.12). The approximation is given by

$$I(e) \approx \sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \sum_{k=1}^{n} w_k f(x_i, x_j, x_k). \qquad (2.4.18)$$

In practice, four quadrature terms per integral yielded a good approximation for the true value of the integral.

In repeated measurement models where the coefficients are regarded as random, numerical integration is used to obtain maximum likelihood estimates of the parameters (see Section 4.3). The joint distribution of the observations and the random coefficients must be integrated to obtain the marginal distribution of the observations. This is required to obtain MLE's for the model parameters.

Gibbs sampling (see Section 4.7) is an alternative estimation procedure which may be used if numerical integration is to be avoided.

2.5 MULTIVARIATE DATA GENERATION

Simulation studies are used throughout this dissertation as a practical means to evaluate estimation techniques. A sample is generated from a population with known characteristics. The performance of different estimation procedures can be assessed by a comparison of estimated and theoretical results. In this dissertation the assumption of multivariate normality is often made. In Section 3.2, for instance, the joint distribution of the observations from a time series with fixed coefficients is assumed to be multivariate normal. When modelling repeated measurement data, where observations for a number of experimental units are available, it is reasonable to regard the coefficients of the model as random variables. The multivariate normal distribution is considered as probability model for the random coefficients in Chapters 4 and 5.

2-21

The section starts with a brief discussion of three different techniques used for multivariate statistical simulation. Methods for the generation of samples from multivariate normal and other members of the elliptical family of distributions are also discussed.

A useful book in this regard is that of Johnson (1987). He highlights three general methods used for multivariate generation, namely conditional distributions, transformation methods and rejection methods. The *rejection approach* is not widely used in multivariate data generation.

The *conditional distribution approach* requires knowledge of the conditional distributions. Instead of generating a vector with n components from a multivariate distribution, the problem is reduced to generating n observations from univariate distributions. Suppose an observation from $f(x)$ is required where $x' = (x_1, \ldots, x_n)$. The distribution of x can be factorised as follows

$$f(x) = f_1(x_1)f_2(x_2|x_1)f_3(x_3|x_1,x_2) \ldots f_n(x_n|x_1,\ldots,x_{n-1}).$$ (2.5.1)

The first observation, x_1, is generated from the marginal distribution of x_1, the second observation from the conditional distribution of $x_2|x_1$, and so on until the last observation is obtained from the conditional distribution of $x_n|x_1, \ldots, x_{n-1}$.

The basic idea behind the *transformation approach* is to write the variable under consideration, as a function of variables with known distributions which are easy to generate.

This approach is usually followed when sampling from the multivariate normal and other elliptical distributions.

2-22

Multivariate normal distribution

The generation of samples from a multivariate normal distribution is simplified through the availability of built-in routines for the generation of standard normal variables in computer software. Suppose an observation, x, is required from the n-dimensional $N(\mu, \Sigma)$ distribution. Let the n components of $z' = (z_1, \ldots, z_n)$ be independent $N(0,1)$ variables so that z is $N(0,I)$. The random vector x can be expressed as a function of z using the transformation

$$x = Lz + \mu \tag{2.5.2}$$

where L is the Choleski root of Σ, i.e. $\Sigma = LL'$. An effective algorithm to calculate the Choleski decomposition of a positive definite matrix is given by Martin, Peters and Wilkinson (1965).

Some basic distributional results will be given before discussing the generation of data from the elliptical class of distributions.

Elliptical class of distributions

The elliptical class of distributions (also referred to as elliptically contoured distributions) can be used in robustness studies as an alternative to the multivariate normal distribution. The best known member in this class is the multivariate normal distribution, but it also includes density functions whose contours of equal density have the same elliptical shape as that of the normal distribution, but with a longer or shorter tail relative to the normal distribution.

The properties of the elliptical distributions were obtained and studied by authors like Kelker (1970), Chmielewski (1981), Cambanis, Huang and Simons (1981) and Muirhead (1982). Gupta and Varga (1993) give results for matrix elliptical distributions.

2-23

The p x 1 random vector x has an elliptically contoured distribution if the characteristic function of $x - \mu$ is a function of the quadratic form $t'Vt$ where μ is a p x 1 vector in R^p and V is a p x p non-negative definite matrix. The characteristic function of $x - \mu$ is

$$\phi_{x-\mu}(t) = exp(it'\mu)\psi(t'Vt) \tag{2.5.3}$$

for some function ψ.

Not all elliptical distributions defined by (2.5.3) have density functions. The focus will be on the restricted class of elliptical distributions which have density functions and a positive definite V. The density function of this restricted class is of the form

$$f(x) = c_p|V|^{-\frac{1}{2}}h[(x-\mu)'V^{-1}(x-\mu)] \tag{2.5.4}$$

for some function h and constant c_p. The notation $E_p(\mu,V,h)$ will be used to denote an p-variate elliptical distribution with parameters μ, V and h.

A *spherically symmetric distribution* is an elliptical distribution with $\mu = 0$ and $V = I$. An important characteristic of a spherically symmetric variable x is that the distribution of x is the same as that of Px for all p x p orthogonal matrices P.

An $E_p(\mu,V,h)$ variable, y, can be obtained from a spherically symmetric variable x by means of the following transformation:

$$y = Cx+\mu \tag{2.5.5}$$

where $V = CC'$ and C is a non-singular p x p matrix.

2-24

The *moments of an elliptical distribution* are functions of the parameters μ and V, but they do not always exist. Suppose x is $E_p(\mu, V, h)$. Provided they exist,

$$E(x) = \mu \qquad (2.5.6)$$

and

$$\text{Cov}(x, x') = \alpha V \qquad (2.5.7)$$

for some constant α. The value of α is defined as

$$\alpha = -2\psi'(0) \qquad (2.5.8)$$

where the function ψ is determined by the characteristic function (2.5.3).

The *marginal distributions* of x are all elliptically symmetric, have the same functional form (denoted by h in (2.5.4)) and have the same kurtosis. The *kurtosis* is

$$\gamma_2 = \frac{3[\psi''(0) - \psi'(0)^2]}{\psi'(0)^2}. \qquad (2.5.9)$$

The kurtosis, γ_2, of a normal distribution is zero. Curves for which γ_2 is zero are called *mesokurtic*, those for which $\gamma_2 > 0$ are called *leptokurtic* and those for which $\gamma_2 < 0$ are called *platykurtic*. Relative to a normal distribution, leptokurtic distributions are sharply peaked and platykurtic distributions will be flat-topped (see Kendall and Stuart, Vol.1 (1958)). In Chapter 5 the Pearson Types VII and II distributions which both belong to the elliptical class and are leptokurtic and platykurtic respectively, are considered as alternatives to the normal distribution.

The above results are illustrated in the following example.

2-25

Example 2.5.1

Suppose x is p-variate $N(\mu, \Sigma)$. The characteristic function of x is

$$\phi_{x-\mu}(t) = exp(it'\mu)\,exp\left(-\tfrac{1}{2}t'\Sigma t\right). \qquad (2.5.10)$$

From (2.5.3) it follows that the function ψ is

$$\psi(y) = exp\left(-\tfrac{1}{2}y\right) \qquad (2.5.11)$$

so that

$$\psi'(y) = -\tfrac{1}{2}exp\left(-\tfrac{1}{2}y\right). \qquad (2.5.12)$$

The value of α is 1, using (2.5.8) and (2.5.12). It follows that V is the covariance matrix of x.

The density function of x is

$$f(x) = (2\pi)^{-\frac{p}{2}}\,|\Sigma|^{-\frac{1}{2}}exp\left[-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]. \qquad (2.5.13)$$

By comparing this result with (2.5.4) the function h is identified as

$$h(y) = exp\left(-\tfrac{1}{2}y\right) \qquad (2.5.14)$$

and the proportionality constant is

2-26

$$c_p = (2\pi)^{-\frac{p}{2}}.$$

<div align="right">(2.5.15)</div>

The kurtosis as defined by (2.5.9) is zero for the multivariate normal distribution.

∎

Another important result that relates to the dependence structure of the components of x is now given.

A basic result in multivariate normal theory is that if x is $N(\mu,D)$ where D is a diagonal matrix, the components of x are all independent. Within the class of elliptical distributions, the normal distribution is characterized by independence when V is diagonal. (Muirhead (1982), Theorem 1.5.3.) The normal distribution is therefore the only elliptical distribution with this property.

The following theorems are useful when observations are generated from distributions in the elliptical class. The results are proved in e.g. Muirhead (1982).

Theorem 2.5.1 (Johnson (1987), p.108)

If x is $E_p(\mu, V, h)$ and B is a $r \times p$ matrix of rank r $(r \leq p)$, then Bx is $E_r(B\mu, BVB', h)$.

This result is used, for instance, in the transformation of a spherically symmetric distribution to a general elliptical distribution as given by (2.5.2).

The following theorem gives the distribution of the quadratic form $(x - \mu)'V^{-1}(x - \mu)$.

<div align="center">2-27</div>

Theorem 2.5.2 (Muirhead (1982), Theorem 1.5.5)

If x is $E_p(\mu, V, h)$, the density function of the quadratic form $q = (x - \mu)'V^{-1}(x - \mu)$ is

$$g(q) = \frac{\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)} c_p q^{\frac{p}{2}-1} h(q), \quad q > 0. \tag{2.5.16}$$

In the case of the multivariate normal distribution the distribution of q is $\chi^2(p)$, and for some other elliptical distributions the density $g(q)$ is a standard form, e.g. the beta or Pearson Type VI.

The following theorem gives the distribution of a normalized (norm of variable is unity) spherically symmetric variable.

Theorem 2.5.3 (Muirhead (1982), Theorem 1.5.6)

If x is $E_p(0, I, h)$, the density function of $t(x) = (x'x)^{-\frac{1}{2}}x$ is uniform on the boundary of the p-dimensional hypersphere and $t(x)$ and $\|x\| = (x'x)^{\frac{1}{2}}$ are independent.

Cambanis, Huang and Simons (1981) provided a representation of elliptically contoured distributions which is useful when generating observations based on the transformation method.

Theorem 2.5.4 (Johnson (1987), p.110)

If x is $E_p(\mu, V, h)$, it can be generated from

$$x = rBu^{(p)} + \mu \tag{2.5.17}$$

2-28

where r is a positive random variable having the distribution of

$q^{\frac{1}{2}} = ((x - \mu)'V^{-1}(x - \mu))^{\frac{1}{2}};$

$u^{(p)}$ is a random vector that is uniformly distributed on the unit hypersphere;

r and $u^{(p)}$ are independent and

B is a $p \times p$ matrix such that $V = BB'$.

Proof

Suppose the random vector y is $E_p(0,I,h)$. Then y can be written as

$$y = (y'y)^{\frac{1}{2}}(y'y)^{-\frac{1}{2}}y \qquad (2.5.18)$$
$$= rt(y)$$

where $r^2 = y'y$ and $t(y) = (y'y)^{-\frac{1}{2}}y$. The distributions of r^2 and $t(y)$ are given by Theorems 2.5.2 and 2.5.3 respectively. The spherical variable y can therefore be generated as the product of the positive square root of r^2 and a random vector that is uniformly distributed on the unit hypersphere. The result (2.5.17) follows if x is set equal to $By + \mu$ where $BB' = V$. ∎

The representation (2.5.17) is referred to as the *Cambanis representation*.
The following transformation of a multivariate normal random vector (Muller (1959)) is usually used when generating a uniform variable on the p-dimensional hypersphere:

$$u = (z'z)^{-\frac{1}{2}}z \qquad (2.5.19)$$

where z is $N(0,I_p)$.

The Cambanis method is used to generate variables from the multivariate Pearson Type

2-29

II and Type VII distributions. For each of these distributions the functional form h is known and the distribution of r^2 is a standard form that is easy to generate.

2.6 SUMMARY

Results used in the estimation of the parameters of repeated measurement models are given. Maximum likelihood estimation is discussed and results pertaining to the multivariate normal distribution with a structured covariance matrix are given. An iterative procedure which can be used to maximize the likelihood function in situations where the maximum likelihood equations can not be expressed in closed form is reviewed. The likelihood functions derived in the subsequent chapters are often given in terms of a multidimensional integral which can not be calculated with standard integration results. It is shown how numerical integration can be used to calculate these likelihood functions. Basic results pertaining to the elliptical class of distributions are introduced and the generation of samples from this class is discussed.

2-30

CHAPTER 3 MAXIMUM LIKELIHOOD ESTIMATION OF FIXED COEFFICIENT ARMA MODELS

3.1 INTRODUCTION

Consider the situation where $y_1 = e_1, ..., y_n = e_n$ denote observations made over time on an experimental unit or let $e_1, ..., e_n$ denote deviations from a model fitted to repeated measurement data. In this chapter it is assumed that the set of observations $(e_1, ..., e_n)$ is the outcome of a process generated by a stationary autoregressive moving average (ARMA) model of order p,q.

Results reported in the literature on the estimation of fixed coefficient stationary ARMA models are discussed. These results are then extended in subsequent chapters to allow for repeated measurements on a number of experimental units (Chapter 4), random ARMA coefficients (Chapter 4), linear and non-linear models with random coefficients and time series error terms (Chapter 5), non-standard distributional assumptions on the white noise terms (Chapter 6) as well as vector ARMA models (Chapter 7).

In Section 3.2 the ARMA(p,q) model is defined and an expression for the likelihood function is given. A so- called state-space representation of an ARMA process is given in Section 3.3. It is shown in Section 3.4 how the Kalman filter can be used to calculate the likelihood function recursively. Recursive estimation of the ARMA coefficients in the case of missing observations is discussed.

The covariance matrix of $e_1, ..., e_n$ is a structured matrix. A factorization of the covariance matrix as proposed by Du Toit (1979, 1990a) and which leads to considerable saving of computational effort is given in Section 3.5.

In Section 3.6 an expression is derived for the observation vector of an ARMA(p,q) process in terms of the state vector.

3-1

To ensure that the estimated ARMA coefficients fall inside the admissible parameter space for stationary and invertible models, a reparameterization is performed on the coefficients. In Section 3.7 a possible reparameterization is discussed.

Detailed results will be given for the ARMA(2,1) model. This model is frequently explored as a special case of the ARMA(p,q) model in this dissertation, as it provides an adequate description for a large variety of time series situations encountered in practice (Pandit and Wu (1983)).

3.2 THE LIKELIHOOD FUNCTION OF AN ARMA PROCESS

Suppose a fixed number of observations, e_1, ..., e_n, are made over time on an experimental unit or subject. The case is considered where e_1, ..., e_n are the outcome of a stationary autoregressive moving average (ARMA) model of order p,q with zero mean. The ARMA(p,q) model is defined by

$$e_t - \alpha_1 e_{t-1} - \ldots - \alpha_p e_{t-p} = u_t - \beta_1 u_{t-1} - \ldots - \beta_q u_{t-q} \qquad (3.2.1)$$

where the u_t's are random shock or noise terms, assumed to be mutually independent $N(0,\sigma^2)$-variables. The variance of u_t is referred to as white noise variance.

Expression (3.2.1) can be written as

$$(1 - \alpha_1 B - \ldots - \alpha_p B^p)e_t = (1 - \beta_1 B - \ldots - \beta_q B^q)u_t \qquad (3.2.2)$$

or

$$\alpha(B)e_t = \beta(B)u_t \qquad (3.2.3)$$

where B is the backshift operator, so that, for instance,

3-2

$$B^s e_t = e_{t-s}.$$

<div align="right">(3.2.4)</div>

The subclass of stationary and invertible models (Box *et al* (1994), p.50-51) is studied throughout this dissertation. For stationarity and invertibility, the roots (which may be real or complex) of the characteristic equations

$$\alpha(B) = 0$$

<div align="right">(3.2.5)</div>

$$\beta(B) = 0$$

<div align="right">(3.2.6)</div>

must lie outside the unit circle.

Newbold (1974) has derived the exact likelihood function of the ARMA process. Box *et al* (1994) discuss least squares estimation, Bayesian estimation and maximum likelihood estimation of the unknown parameters of an ARMA model. Chib and Greenberg (1993) propose Gibbs sampling for Bayes inference in regression models with ARMA(p,q) errors.

The parameters of the model (3.2.1) are the ARMA coefficients α_1, ..., α_p, β_1, ..., β_q and the white noise variance, σ^2. Let $\gamma' = (\alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \sigma^2)$. Every observation, e_t, can be expressed as a linear combination of random noise terms, u_t, u_{t-1}, ... (see e.g. Box *et al* (1994) p.85). From (3.2.3) follows:

$$e_t = \frac{\beta(B)}{\alpha(B)} u_t$$

<div align="right">(3.2.7a)</div>

$$= \sum_{i=0}^{\infty} \psi_i B^i u_t \quad \text{where} \quad \psi_0 = 1$$

<div align="right">(3.2.7b)</div>

<div align="center">3-3</div>

$$= \sum_{i=0}^{\infty} \psi_i u_{t-i}.$$

(3.2.7c)

Under the assumption of Gaussian random noise, the joint distribution of e_1, ..., e_n is multivariate normal given by (cf. (2.2.10))

$$f(e) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} exp\left(-\frac{1}{2} e' \Sigma^{-1} e\right)$$

(3.2.8)

so that the log-likelihood is

$$lnL = -\frac{1}{2}\left[nln2\pi + nln\sigma^2 + ln|\Lambda| + \frac{1}{\sigma^2} e' \Lambda^{-1} e\right]$$

(3.2.9)

where $\Lambda = \frac{1}{\sigma^2}\Sigma$.

The covariance matrix of e_1, ..., e_n in Toeplitz form is (see e.g. Press *et al* (1986), p. 47)

$$\Sigma = \begin{bmatrix} c_0 & c_1 & \cdots & \cdots & c_{n-1} \\ c_1 & c_0 & c_1 & \cdots & c_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n-1} & c_{n-2} & \cdots & \cdots & c_0 \end{bmatrix}$$

(3.2.10)

where c_k, the autocovariance at lag k, is

3-4

$$c_k = E(e_t e_{t+k}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}. \tag{3.2.11}$$

Assuming that the other components of γ are known, then the maximum likelihood estimate of σ^2 can be obtained in closed form and is given by

$$\hat{\sigma}^2 = \frac{1}{n} e' \Lambda^{-1} e. \tag{3.2.12}$$

Let L' denote the likelihood function concentrated with respect to σ^2, which is obtained by substituting (3.2.12) into (3.2.8). The concentrated log-likelihood is

$$lnL' = c - \frac{1}{2}[nln(e' \Lambda^{-1} e) + ln|\Lambda|] \tag{3.2.13}$$

where c is a constant. It is possible to express the concentrated log-likelihood in terms of the error terms. Let $\Lambda = LL'$ be the Choleski decomposition of Λ where L is a lower triangular matrix. If we let $u = L^{-1}e$, the vector u contains a sequence of orthonormal variables, i.e. $u_1, ..., u_n$ are independent $N(0, \sigma^2)$-variables (see e.g. Wecker and Ansley (1983)). The concentrated log-likelihood is

$$lnL' = c - \frac{n}{2}ln(u'u). \tag{3.2.14}$$

Maximum likelihood estimates of the ARMA coefficients maximize (3.2.14).

If σ^2 is not concentrated then the log-likelihood is

$$lnL = -\frac{1}{2}\left[nln2\pi + nln\sigma^2 + \frac{1}{\sigma^2}u'u \right]. \tag{3.2.15}$$

3-5

Unconditional least squares estimates can be obtained by minimizing the discrepancy function $F = z'z$.

3.3 STATE-SPACE REPRESENTATION OF A STATIONARY ARMA PROCESS

There are several equivalent state-space representations of ARMA models referred to in literature. Akaike's (1973, 1974, 1975) Markovian representation provides a minimal state-space representation for recursive calculation of the likelihood function for a Gaussian ARMA process for given values of the parameters. Jones (1980) extended the procedure for the case of missing observations and provided an efficient procedure to compute the initial state covariance matrix. This representation is also found in some standard statistical packages, such as SAS/ETS® (1988), and will be given here. Results from Jones (1993) Chapters 4 and 6 are adapted and given in this section.

Gomez and Maravall (1994) summarized the results of Jones (1980) and extended the results to include non-stationary series.

The state-space representation consists of two equations, namely the equation of state and the observational equation. In Akaike's Markovian representation, the *state of the process* is defined by the following r x 1 vector where $r = \text{Max}(p, q+1)$:

$$z(t) = \begin{bmatrix} e(t|t) \\ e(t+1|t) \\ \vdots \\ e(t+r-1|t) \end{bmatrix} . \qquad (3.3.1)$$

The notation $e(t+j|t)$ denotes the j-step prediction of e_{t+j} at time t or $E[e(t+j)|e_s, s \leq t]$. The first component, $e(t|t)$, is the value of the process at time t, $e(t|t) = e_t$.

3-6

The state vector $z(t)$ contains the observation at time t and projections of the next r-1 terms based on the observations up to time t.

The *equation of state* is given in matrix notation by

$$z(t+1) = Fz(t)+gu_{t+1}.$$

(3.3.2)

The r x r matrix F, the *state transition matrix*, is

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ \alpha_r & \alpha_{r-1} & \dots & \alpha_2 & \alpha_1 \end{bmatrix}$$

(3.3.3)

and the r x 1 vector g is

$$g = \begin{bmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_{r-1} \end{bmatrix}$$

(3.3.4)

where the ψ_i -weights are given by (3.2.7).

The second equation in the state-space representation is the *observational equation*

3-7

$$y(t) = h'z(t) + v(t) \qquad (3.3.5)$$

where h is the $r \times 1$ vector

$$h' = (1 \quad 0 \quad \ldots \quad 0). \qquad (3.3.6)$$

Note that the operation $h'z(t)$ simply results in the first component of $z(t)$ and the operation $h'Ah$, where A is an $r \times r$ matrix, yields the element in the first row and first column of A.

The term $v(t)$ is a Gaussian random variable denoting observational error at time t. The observational error terms are uncorrelated at different times and uncorrelated with the random noise terms and have zero mean and variance

$$R = E[v(t)^2]. \qquad (3.3.7)$$

Jones (1993) remarks that observational error can only be included in models if $q < p$, since this results in an ARMA process of order (p,p). If $q \geq p$, observational error simply changes the values of the moving average coefficients. The consequence is that moving average coefficients and observational error variance are confounded and cannot be estimated separately.

Although the general case, where provision is made for observational error, is given here, the focus will be on models with no observational error, as given by Gomez and Maravall (1994). The results can easily be adapted for the special case by just ignoring $v(t)$ in the observational equation.

3-8

3.4 KALMAN RECURSIVE ESTIMATION

The *Kalman filter* is a set of equations which are evaluated repeatedly as new observations become available. In this way an estimator can be obtained and updated as more information is gained. The equations can be grouped into two sets, namely the *prediction* and *updating* equations. The first part of the process consists of forming the optimal predictor of the next observation, given all the past observations. This is obtained by means of the prediction equations. In the second part of the process the new observation is incorporated into the estimator of the state vector by employing the updating equations. As each new observation becomes available, a prediction error can be calculated by taking the difference between the observation and its prediction. These prediction errors can be used in maximum likelihood estimation, just as the likelihood function for a set of dependent observations may be decomposed in terms of the prediction errors.

Kalman recursive estimation is a well known method to calculate the exact likelihood of an ARMA process. (See e.g. Jones (1980) and Harvey (1981).)

Under the assumption of normally distributed error terms and for given values of the ARMA parameters, the likelihood function can be calculated using Kalman recursive estimation.

Consider the joint distribution of e_1, ..., e_n as given in (3.2.8). It can be written as

$$f(e) = (2\pi\sigma^2)^{-\frac{n}{2}} |\Lambda|^{-\frac{1}{2}} exp \left[-\frac{1}{2\sigma^2} e' \Lambda^{-1} e \right] \qquad (3.4.1)$$

where $\sigma^2 \Lambda = \Sigma$.

Let $u = Le$ where L is the Choleski root of Λ. Now u_1, ..., u_n are the error terms of

3-9

the ARMA model (3.2.1) and their joint distribution is

$$f(u) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}u'u\right].$$
(3.4.2)

Note that (3.2.8), the likelihood function in terms of e_1, ..., e_n and (3.4.2) are equivalent expressions which can be used to get MLE's of the unknown parameters.

The error terms are also sometimes called innovation terms or one-step ahead forecast errors. The purpose of the Kalman recursions is to calculate the innovation terms and their variance. For a stationary ARMA process, the variance of the innovation terms is fixed and equal to the white noise variance.

The initial state vector $z(0|0)$ is an estimate of the state at time zero before any data are collected. For a process with zero mean, this will be a vector of zeros. The covariance matrix of this vector, the initial state covariance matrix, is denoted by $P(0|0)$. The generation of this matrix is discussed in the next section.

The Kalman recursion can be performed in eight steps. The prediction equations for the state vector and its covariance matrix are given in the first two steps by (3.4.3) and (3.4.4) respectively. In steps 3, 4 and 5 the prediction of the next observation, prediction error and its variance are calculated. The sixth step consists of updating the likelihood function. The Kalman updating equations are given in (3.4.10) through (3.4.12) and are calculated in steps 7 and 8 of the recursion. These eight steps are as follows:

Step 1:
Calculate a one step prediction of the state vector, by taking expectations on both sides of (3.3.2). Note that $E[u(t+1)]$ is zero at time t.

3-10

$$z(t+1|t) = Fz(t|t) \tag{3.4.3}$$

Step 2:

Calculate the covariance matrix of $z(t+1|t)$, by determining the covariance of both sides of equation (3.3.2):

$$P(t+1|t) = FP(t|t)F' + \sigma^2 gg' . \tag{3.4.4}$$

Step 3:

Calculate the prediction of the next observation, $y(t+1|t)$, by taking expected values on both sides of (3.3.5) and noting that $E[v(t+1)] = 0$ at time t.

$$\begin{aligned} y(t+1|t) &= h'z(t+1|t) \\ &= e(t+1|t) \end{aligned} \tag{3.4.5}$$

Step 4:

Calculate the innovation (the difference between the observation $y(t+1)$ and the prediction of $y(t+1)$ at time t, $y(t+1|t)$.

$$\bar{y}(t+1) = y(t+1) - y(t+1|t) \tag{3.4.6}$$

The innovation is a central concept in Kalman filter theory. The innovation term contains the new information in the current observation, $y(t+1)$, that was unpredictable from the past observations. The innovation is uncorrelated (orthogonal) with the past observations. If the random noise terms are Gaussian, this implies that all the innovations are independent and Gaussian. The likelihood will consequently be the product of the likelihood of the innovations.

If $y(t+1)$ is missing, it is replaced with $y(t+1|t)$. Consequently, the innovation term will be zero.

3-11

Step 5:

Calculate the variance of the innovation term.

$$V(t+1) = h'P(t+1|t)h + R \qquad (3.4.7)$$

R is the observational variance and $h'P(t+1|t)h$ is the variance of $y(t+1|t)$ in (3.4.4), calculated in step 3. We also use the fact that the observational error is independent of past observations. For a stationary process with no observational error, the variance of all innovation terms is σ^2, the element in row one column one of $P(t+1|t)$.

Step 6:

Update the likelihood function. As noted in step 4, the innovations are independent normal with zero mean and variance defined in step 5. Let the variance of the i-th innovation be V_i. The likelihood can be written as

$$L = \prod_{i=1}^{n} (2\pi V_i)^{-\frac{1}{2}} exp\left[-\frac{\tilde{y}_i^2}{2V_i} \right] \quad \text{where} \quad V_i = V(i) \text{ and } \tilde{y}_i = \tilde{y}(i) \qquad (3.4.8)$$

and $-2lnL$ is

$$-2lnL = c + \sum_{i=1}^{n} \left[lnV_i + \frac{\tilde{y}_i^2}{V_i} \right]. \qquad (3.4.9)$$

Step 7:

Update the estimate of the state vector.

$$z(t+1|t+1) = z(t+1|t) + k(t+1)\tilde{y}(t+1) \qquad (3.4.10)$$

3-12

where

$$k(t+1) = P(t+1|t)hV_{t+1} \qquad (3.4.11)$$

is referred to as the Kalman gain. The Kalman gain vector is constructed in such a way as to ensure orthogonality of the innovation terms.

If $y(t+1)$ is missing, the innovation term is zero and $z(t+1|t+1) = z(t+1|t)$.

Step 8:

Calculate the covariance matrix of the updated state.

$$P(t+1|t+1) = P(t+1|t)-k(t+1)h'P(t+1|t) \qquad (3.4.12)$$

If stationarity is assumed, the state covariance matrix is constant for all values of t and only needs to be calculated once.

If $y(t+1)$ is missing, then $z(t+1|t+1) = z(t+1|t)$ so that $P(t+1|t+1) = P(t+1|t)$.

Initial state covariance matrix

For a stationary process the state covariance matrix, $P(t|t)$, is constant for all t. Let $P = P(t|t)$. From the state equation (3.3.2) it follows that

$$P = FPF' + \sigma^2 gg' \qquad (3.4.13)$$

so that

$$vec(P) = vec(FPF') + vec(\sigma^2 \cdot gg') \qquad (3.4.14a)$$

3-13

$$= F \otimes F vec(P) + vec(Q) \tag{3.4.14b}$$

and

$$(I - F \otimes F) vec(P) = vec(Q) \tag{3.4.15}$$

where $Q = \sigma^2 gg'$. (Result (3.4.14b) can be found in, Magnus and Neudecker (1988), p. 30.) From (3.4.15) it follows that

$$vec(P) = (I - F \otimes F)^{-1} vec(Q). \tag{3.4.16}$$

The dimension of the matrix that must be inverted is r^2. It is possible to get an alternative expression in terms of $vecs(P)$ (where $vecs(P)$ is the vector formed from the non-duplicate elements of the $r \times r$ matrix P) instead of $vec(P)$, which will reduce the dimension of the matrix to be inverted to $\frac{1}{2}r(r+1)$. The definition of matrices which changes $vec(P)$ into $vecs(P)$ and vice versa is rather tedious and is given in e.g. Browne (1974). In the case of the ARMA(2,1) model the possible saving in computer time does not justify the effort of reducing the dimension by one.

Missing observations

Assume that observations were made at times t_1, t_2, ..., t_m where $t_1 < t_2 < ... < t_m$. The observational equation (3.3.5) can be replaced by

$$y(t) = h'(t)z(t) + \alpha(t)W(t) + v(t) \tag{3.4.17}$$

where $h'(t) = (1, 0, ..., 0)$ and $\alpha(t) = 0$ if e_t is observed and $h'(t) = (0, ..., 0)$ and $\alpha(t) = 1$ if e_t is missing. The variable $W(t)$ is standard normal independent of e_{t1}, e_{t2}, ..., e_{tm}. When e_t is missing, then $z(t+1|t+1) = z(t+1|t)$ and $P(t+1|t+1) = P(t+1|t)$.

3-14

The innovation term will consequently be zero and have no effect on the discrepancy function (3.4.9).

There can be a number of consecutive missing observations. If a large block of data is missing, the information about the past observations decreases. The state vector will approach the initial state vector which is the null vector in most cases. The covariance matrix of the state vector will approach the initial state covariance matrix.

The results will now be applied to an ARMA(2,1) process.

Example 3.4.1

Suppose the observed series $(e_1, \ldots, e_n)$ is the outcome of a stationary ARMA(2,1) process

$$e_t - \alpha_1 e_{t-1} - \alpha_2 e_{t-2} = u_t - \beta u_{t-1} \tag{3.4.18}$$

where the u_t 's are independent $N(0,\sigma^2)$ variables.

The problem is to estimate γ where $\gamma' = (\alpha_1, \alpha_2, \beta, \sigma^2)$. It is possible to concentrate σ^2 from the likelihood and estimate it separately, using the estimates of the ARMA coefficients. The latter have to be estimated with an iterative optimization algorithm as discussed in Section 2.3. The Kalman recursive procedure can be used to calculate the discrepancy function for maximum likelihood estimation. The state-space equations for the ARMA(2,1) model and the recursive steps are given in order to calculate the discrepancy function.

The main advantage of the Kalman procedure is that matrices and vectors of order 2, instead of order n are manipulated, which would have been the case if the log-likelihood as given in (3.2.9) were to be calculated.

3-15

The observations can be expressed in terms of random noise (cf. (3.2.7a) through to (3.2.7c)):

$$
\begin{aligned}
e_t &= \frac{1-\beta B}{1-\alpha_1 B-\alpha_2 B^2} u_t \\
&= (1-\beta B)[1+\alpha_1 B+\alpha_2 B^2+(\alpha_1 B+\alpha_2 B^2)^2+\ldots] u_t \\
&= [1+(\alpha_1-\beta)B+(\alpha_1^2-\alpha_1\beta+\alpha_2)B^2+\ldots] u_t \\
&= [\psi_0+\psi_1 B+\psi_2 B^2+\ldots] u_t
\end{aligned}
\tag{3.4.19}
$$

The state vector of an ARMA(2,1) model is

$$
z(t) = \begin{bmatrix} e(t|t) \\ e(t+1|t) \end{bmatrix}
\tag{3.4.20}
$$

and the state equation is

$$
\begin{bmatrix} e(t+1|t+1) \\ e(t+2|t+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \alpha_2 & \alpha_1 \end{bmatrix} \begin{bmatrix} e(t|t) \\ e(t+1|t) \end{bmatrix} + \begin{bmatrix} 1 \\ \alpha_1-\beta \end{bmatrix} u_{t+1}.
\tag{3.4.21}
$$

Under the assumption of no observational error the observational equation is

$$
\begin{aligned}
y(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} e(t|t) \\ e(t+1|t) \end{bmatrix} \\
&= e(t|t)
\end{aligned}
\tag{3.4.22}
$$

3-16

To start the Kalman recursive procedure, the initial state vector is set to zero. The state covariance matrix, P, is constant in the case of a stationary model. From (3.4.16) it follows that

$$vec(P) = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -\alpha_2 & -\alpha_1 \\ 0 & -\alpha_2 & 1 & -\alpha_1 \\ -\alpha_2^2 & -\alpha_1\alpha_2 & -\alpha_1\alpha_2 & 1-\alpha_1^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \alpha_1-\beta \\ \alpha_1-\beta \\ (\alpha_1-\beta)^2 \end{bmatrix} . \tag{3.4.23}$$

Let the non-duplicate elements of P be denoted by p_{11}, p_{12} and p_{22}. Note that if $\sigma^2 = 1$ and the concentrated log-likelihood (3.2.14) is maximized, then P and $P(t+1|t)$ in the recursion are in fact P^*/σ^2 and $P^*(t+1|t)/\sigma^2$, where P^* and $P^*(t+1|t)$ are the state covariance matrices for the general case and σ^2 is not necessarily equal to one.

The aim of the recursive procedure is to calculate the value of the discrepancy function given in (3.4.9). Before the recursion is started, it is necessary to set two summation variables equal to zero. After the final recursion, the variable $SUMA$ will be the sum of $ln(V_i)$ and the variable $SUMB$ will be the sum of $\dfrac{\tilde{y}_i^2}{V_i}$, $i = 1, 2 \ldots, n$.

The following steps are repeated for $t = 0, 1, \ldots, n\text{-}1$:

Step 1: The one step prediction of the state vector (from (3.4.3)) is

$$z(t+1|t) = \begin{bmatrix} e(t+1|t) \\ \alpha_2 e(t|t)+\alpha_1 e(t+1|t) \end{bmatrix} . \tag{3.4.24}$$

3-17

Step 2: The covariance matrix of (3.4.24) (from (3.4.4)) is

$$P(t+1|t) = \begin{bmatrix} 1+p_{22} & \alpha_2 p_{12}+\alpha_1 p_{22}+\alpha_1-\beta \\ \alpha_2 p_{12}+\alpha_1 p_{22}+\alpha_1-\beta & \alpha_2^2 p_{11}+2\alpha_1\alpha_2 p_{12}+\alpha_1^2 p_{22}+\alpha_1-\beta \end{bmatrix}.$$

(3.4.25)

Step 3: The prediction of the next observation (from (3.4.5)) is the first element of the predicted state vector

$$y(t+1|t) = e(t+1|t).$$

(3.4.26)

Step 4: The innovation term at time $t+1$ (from (3.4.6)) is

$$\tilde{y}(t+1) = e_{t+1}-e(t+1|t).$$

(3.4.27)

Step 5: The variance of the innovation term (from (3.4.7)) is simply the element in row 1, column 1 of $P(t+1|t)$. In the case of no observational error, R is set equal to zero.

$$V(t+1) = 1+p_{22}$$

(3.4.28)

Step 6: Accumulate the variables $SUMA$ and $SUMB$.

$$SUMB \leftarrow SUMB+\frac{\tilde{y}(t+1)^2}{V(t+1)}$$

(3.4.29a)

$$SUMA \leftarrow SUMA+lnV(t+1)$$

(3.4.29b)

Step 7: The state vector is updated using (3.4.10) and (3.4.11).

Stationarity is assumed, so that the state covariance matrix remains constant as

3-18

calculated in (3.4.23).

After the last recursive step ($t = n\text{-}1$), the value of the discrepancy function (3.4.9) can be calculated by

$$Q = SUMA + SUMB.\qquad\qquad(3.4.30)$$

3.5 THE COVARIANCE MATRIX OF A STATIONARY ARMA PROCESS

Let $e' = (e_1, \ldots, e_n)$ denote n observations of an ARMA(p,q) process with zero mean. Du Toit (1979) proposed a decomposition of the covariance matrix of e which leads to a considerable saving in computing time and effort. The covariance matrix differs from the one given in (3.2.10) in that it takes into account the state of the process before any observations were made. The decomposition is given in this section.

Let $r = \max(p,q)$. In accordance with the notation used in the previous section, let $e(j|0)$ denote the prediction of e_j at time 0, $j = 1, \ldots, r$, so that

$$e(j|0) = E[e_j | e_s,\ s \le 0]\qquad\qquad(3.5.1)$$

$$= \sum_{i=j}^{r} (\alpha_i e_{j-i} - \beta_i u_{j-i})$$

where $\alpha_{p+1} = \alpha_{p+2} = \ldots = \alpha_r = 0$ if $p < r$ and $\beta_{q+1} = \beta_{q+2} = \ldots = \beta_r = 0$ if $q < r$.

Note that the j-step prediction at time 0 only contains terms that were not observed. If $j > r$, the j-step prediction at time 0 is zero, the mean of the process. The observation e_j, $j = 1, 2, \ldots, r$, is a function of terms that were observed ($e_1, \ldots, e_{j-1}$) as well as its

3-19

j-step prediction at time 0. The first observation, for instance, can be written as

$$e_1 = u_1 + \sum_{i=1}^{p} \alpha_i e_{1-i} - \sum_{i=1}^{q} \beta_i u_{1-i}$$

$$= u_1 + e(1|0) \,. \tag{3.5.2}$$

In matrix notation the equations defining $e_1, \ldots, e_n$ is

$$T_\alpha e = T_\beta u + I_{n,r} x(0) \tag{3.5.3}$$

where T_α and T_β are n x n lower triangular matrices, e and u are n x 1 vectors containing the observations and random noise terms respectively, $I_{n,r}$ is an n x r matrix with columns the first r columns of the n x n identity matrix and $x(0)$ an r x 1 vector containing one-step to r-step predictions at time 0.

The matrix T_α is defined by

$$T_\alpha = \begin{bmatrix} 1 & 0 & \ldots & & & & \\ -\alpha_1 & 1 & 0 & \ldots & & & \\ -\alpha_2 & -\alpha_1 & 1 & \ldots & & & \\ \vdots & & & & & & \\ -\alpha_p & -\alpha_{p-1} & \ldots & -\alpha_1 & 1 & \ldots & 0 \\ 0 & -\alpha_p & \ldots & & & & \\ \vdots & & & & & & \\ 0 & 0 & \ldots & -\alpha_p & -\alpha_{p-1} & \ldots & -\alpha_1 & 1 \end{bmatrix} \tag{3.5.4}$$

and, similarly, the matrix T_β is given by

$$T_\beta = \begin{bmatrix} 1 & 0 & \cdots & & & & & \\ -\beta_1 & 1 & 0 & \cdots & & & & \\ -\beta_2 & -\beta_1 & 1 & 0 & \cdots & & & \\ \vdots & & & & & & & \\ -\beta_q & -\beta_{q-1} & -\beta_1 & 1 & 0 & \cdots & & \\ 0 & -\beta_q & -\beta_{q-1} & \cdots & -\beta_1 & 1 & 0 & \cdots \\ \vdots & & & & & & & \\ 0 & 0 & \cdots & -\beta_q & -\beta_{q-1} & \cdots & -\beta_1 & 1 \end{bmatrix} . \tag{3.5.5}$$

The state vector, $x(0)$, is

$$x(0) = \begin{bmatrix} e(1|0) \\ e(2|0) \\ \vdots \\ e(r|0) \end{bmatrix} . \tag{3.5.6}$$

Equation (3.5.3) can be written as

$$e = T_\alpha^{-1}[T_\beta u + I_{n,r} x(0)] \tag{3.5.7}$$

since T_α is non-singular. The covariance matrix of e, denoted by Σ, is

$$\Sigma = T_\alpha^{-1}[\sigma^2 T_\beta T_\beta' + I_{n,r}\text{Cov}[x(0),x(0)']I_{n,r}']T_\alpha^{-1'} . \tag{3.5.8}$$

3-21

The vector $x(0)$ can be considered a state vector and its covariance matrix can be derived from the state equation using the results of Section 3.4. The derivation of the state equation will be given for an ARMA(2,1) process. The state transition matrix, F, and the vector g of the state equation (3.3.2) will be derived. The state equation can be written as

$$\begin{bmatrix} e(2|1) \\ e(3|1) \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} e(1|0) \\ e(2|0) \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} u_1 .$$

(3.5.9)

For an ARMA(2,1) process the two elements of the state vector $x(0)$ are

$$\begin{aligned} e(1|0) &= \alpha_1 e_0 + \alpha_2 e_{-1} - \beta u_0 \\ &= e_1 - u_1 \end{aligned}$$

(3.5.10)

and

$$e(2|0) = \alpha_2 e_0 .$$

(3.5.11)

By substitution of (3.5.10) and (3.5.11), the state equation (3.5.9) can be written as

$$\begin{bmatrix} \alpha_1 e_1 + \alpha_2 e_0 - \beta u_1 \\ \alpha_2 e_1 \end{bmatrix} = \begin{bmatrix} f_{11}(e_1 - u_1) + f_{12}\alpha_2 e_0 + g_1 u_1 \\ f_{21}(e_1 - u_1) + f_{22}\alpha_2 e_0 + g_2 u_1 \end{bmatrix}$$

$$= \begin{bmatrix} f_{11} e_1 + f_{12}\alpha_2 e_0 + (g_1 - f_{11})u_1 \\ f_{21} e_1 + f_{22}\alpha_2 e_0 + (g_2 - f_{21})u_1 \end{bmatrix} .$$

(3.5.12)

3-22

It can be readily seen that the state transition matrix is

$$F = \begin{bmatrix} \alpha_1 & 1 \\ \alpha_2 & 0 \end{bmatrix}$$
(3.5.13)

and the vector g is

$$g' = [\alpha_1 - \beta, \quad \alpha_2].$$
(3.5.14)

If stationarity is assumed, the covariance matrix can be obtained from (3.4.16). Let the covariance matrix of $x(0)$ be denoted by P. Then

$$vec(P) = \sigma^2 \begin{bmatrix} 1-\alpha_1^2 & -\alpha_1 & -\alpha_1 & -1 \\ -\alpha_1\alpha_2 & 1 & -\alpha_2 & 0 \\ -\alpha_1\alpha_2 & -\alpha_2 & 1 & 0 \\ -\alpha_2^2 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} (\alpha_1-\beta)^2 \\ \alpha_2(\alpha_1-\beta) \\ \alpha_2(\alpha_1-\beta) \\ \alpha_2^2 \end{bmatrix}.$$
(3.5.15)

For an ARMA(p,q) process the transition matrix is

(3.5.16)

$$F = \begin{bmatrix} \alpha_1 & \cdot & & & & \\ \alpha_2 & \cdot & & I_{r-1} & & \\ \vdots & \cdot & & & & \\ \cdots & \cdot & \cdots & \cdots & \cdots & \cdots \\ \alpha_r & \cdot & 0 & 0 & 0 & 0 \end{bmatrix}$$

3-23

and the vector g is

$$g = \begin{bmatrix} \alpha_1 - \beta_1 \\ \alpha_2 - \beta_2 \\ \vdots \\ \alpha_r - \beta_r \end{bmatrix} . \qquad (3.5.17)$$

From (3.4.16) P, the covariance matrix of $x(0)$, can be obtained. Equation (3.5.8) can be written as

$$\begin{aligned} \Sigma &= \sigma^2 T_\alpha^{-1} A T_\alpha^{-1\prime} \\ &= \sigma^2 \Lambda \end{aligned} \qquad (3.5.18)$$

where

$$A = T_\beta T_\beta' + I_{n,r} P^* I_{n,r} \qquad (3.5.19)$$

and

$$P^* = \frac{1}{\sigma^2} P . \qquad (3.5.20)$$

For an ARMA(2,1) process, the matrix A has a tridiagonal structure and can be factorized as the product of a lower and an upper triangular matrix, $A = LU$. The LU decomposition (see Press $et\ al$ (1986), p.31) leads to considerable simplifications in the calculation of $e'\Lambda^{-1}e$, and $|\Lambda|$ in the log-likelihood (cf. (3.2.9)). The quadratic form is calculated with forward and reverse substitutions instead of an explicit inversion of

3-24

Λ, which can be very time consuming for a long series. Since Λ can be factorized into triangular matrices, its determinant is simply the product of their diagonal elements. The factorization of A and the calculation of the quadratic form and determinant are now given for an ARMA(2,1) process.

For an ARMA(2,1) process, the matrix A defined by (3.5.19) is

$$A = \begin{bmatrix} 1+p_{11}^* & p_{21}^*-\beta & 0 & \ldots & \ldots & \ldots & 0 \\ p_{21}^*-\beta & 1+\beta^2+p_{22}^* & -\beta & 0 & \ldots & \ldots & 0 \\ 0 & -\beta & 1+\beta^2 & -\beta & 0 & \ldots & 0 \\ \vdots & & & & & & \\ 0 & 0 & & \ldots & \ldots & -\beta & 1+\beta^2 \end{bmatrix} \qquad (3.5.21)$$

where $p_{ij}^* = [P^*]_{ij}$.

Now A can be written as the product of a lower, L, and an upper, U, triangular matrix with

$$L = \begin{bmatrix} \zeta_1 & 0 & 0 & \ldots & \ldots & 0 \\ p_{21}^*-\beta & \zeta_2 & 0 & \ldots & \ldots & 0 \\ 0 & -\beta & \zeta_3 & 0 & \ldots & 0 \\ \vdots & & & & & \\ 0 & 0 & \ldots & \ldots & -\beta & \zeta_n \end{bmatrix}, \qquad (3.5.22)$$

3-25

$$U = \begin{bmatrix} 1 & \dfrac{p_{21}^* - \beta}{\zeta_1} & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & 1 & \dfrac{-\beta}{\zeta_2} & 0 & \ldots & \ldots & 0 \\ 0 & 0 & 1 & \dfrac{-\beta}{\zeta_3} & 0 & \ldots & 0 \\ \vdots & & & & & & \\ 0 & 0 & \ldots & \ldots & \ldots & \ldots & 1 \end{bmatrix}$$

(3.5.23)

and

$$\zeta_1 = 1 + p_{11}^*$$

(3.5.24a)

$$\zeta_2 = 1 + \beta^2 + p_{22}^* - \frac{1}{\zeta_1}(p_{21}^* - \beta)^2$$

(3.5.24b)

$$\zeta_i = 1 + \beta^2\left[1 - \frac{1}{\zeta_{i-1}}\right], \quad i = 3,\ldots, n.$$

(3.5.24c)

The determinant of Λ is

$$\begin{aligned} |\Lambda| &= |T_\alpha^{-1}LUT_\alpha^{-1'}| \\ &= |L| \\ &= \prod_{i=1}^{n} \zeta_i \end{aligned}$$

(3.5.25)

3-26

since the diagonal elements of T_α^{-1} and U are all unity.

For an ARMA(2,1) process, the quadratic form, $e'\Lambda^{-1}e$, can be calculated using the technique of forward and reverse substitution. From (3.5.18) it follows that

$$e'\Lambda^{-1}e = e'(T_\alpha^{-1}AT_\alpha^{-1'})^{-1}e \qquad\qquad (3.5.26a)$$

$$= e'T'_\alpha A^{-1}T_\alpha e \qquad\qquad (3.5.26b)$$

$$= y'U^{-1}L^{-1}y \quad\text{ where } y = T_\alpha e \qquad\qquad (3.5.26c)$$

$$= v'w \quad\text{ where } v' = y'U^{-1} \text{ and } w = L^{-1}y. \qquad\qquad (3.5.26d)$$

The n x 1 vectors v and w can be obtained by solving the sets of equations

$$y = U'v \qquad\qquad (3.5.27)$$

and

$$y = Lw. \qquad\qquad (3.5.28)$$

From (3.5.27) it follows that

$v_1 = y_1;$
$v_2 = y_2 - u_{12}v_1; \quad \dots$
$v_n = y_n - u_{n-1,n}v_{n-1},$

$$\qquad\qquad (3.5.29)$$

and from (3.5.28) it follows that

3-27

$$w_1 = \frac{y_1}{l_{11}} \ , \quad w_2 = \frac{(y_2 - l_{21}w_1)}{l_{22}} \ , \ \ldots \ , \ w_n = \frac{(y_n - l_{n,n-1}w_{n-1})}{l_{nn}} \ . \tag{3.5.30}$$

3.6 REPRESENTATION OF AN ARMA(p,q) OBSERVATION VECTOR IN TERMS OF THE STATE VECTOR

In this section it is shown that the vector of observations, e, can be expressed in terms of the state vector defined by (3.3.2). An expression for e equivalent to (3.5.7) is derived for complete data and then generalized for the case of missing observations.

Proposition 3.6.1

Let $e' = (e_1, \ldots, e_n)$ represent n observations from a stationary ARMA(p,q) process. Define r as $max(p,q)$. Then e can be written as

$$e = (I_n \otimes h')T_\alpha^{-1}[T_\beta u + I_{nr,r}Fz(0)] \tag{3.6.1}$$

where the n-dimensional vector u contains the random noise terms, F is the state transition matrix defined by (3.3.3) and $z(0)$ is the initial state vector (the state vector is given in (3.3.1)). Vector h is r-dimensional and defined by (3.3.6). The non-singular matrix T_α: nr x nr is

$$T_\alpha = \begin{bmatrix} I_r & 0 & \ldots & \ldots & \ldots & 0 \\ -F & I_r & 0 & \ldots & \ldots & \vdots \\ 0 & -F & I_r & 0 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & -F & I_r \end{bmatrix} \tag{3.6.2}$$

3-28

and T_β: nr x n is

$$T_\beta = I_n \otimes g \tag{3.6.3}$$

where g: r x 1 is defined by (3.3.4). The nr x r matrix $I_{nr,r}$ consists of the first r columns of I_{nr}.

Proof

From the state equation of an ARMA process (see (3.3.2)) the following system of equations can be constructed to define the state vector at times 1, 2 ..., n:

$$
\begin{aligned}
z(1) &= gu_1 + Fz(0) \\
-Fz(1) + z(2) &= gu_2 \\
- Fz(2) + z(3) &= gu_3 \\
&\vdots \\
- Fz(n-1) + z(n) &= gu_n
\end{aligned}
\tag{3.6.4}
$$

In matrix notation it can be written as

$$
\begin{bmatrix}
I_r & 0 & \dots & \dots & \dots & 0 \\
-F & I_r & 0 & \dots & \dots & \vdots \\
0 & -F & I_r & 0 & \dots & \vdots \\
0 & 0 & \dots & \dots & -F & I_r
\end{bmatrix}
\begin{bmatrix}
z(1) \\ z(2) \\ \vdots \\ \vdots \\ z(n)
\end{bmatrix}
=
\begin{bmatrix}
g & 0 & \dots & 0 \\
0 & g & 0 & \dots \\
\vdots & & & \\
\vdots & & & \\
0 & 0 & \dots & g
\end{bmatrix}
\begin{bmatrix}
u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_n
\end{bmatrix}
+
\begin{bmatrix}
Fz(0) \\ 0 \\ \vdots \\ \vdots \\ 0
\end{bmatrix}
\tag{3.6.5}
$$

or

$$T_\alpha z = T_\beta u + I_{nr,r} F z(0).$$ (3.6.6)

The matrix T_α is a lower triangular matrix with all diagonal elements equal to one, and is therefore non-singular so that the vector of state vectors is

$$z = T_\alpha^{-1}[T_\beta u + I_{nr,r} F z(0)].$$ (3.6.7)

The elements of e are z_1, z_{r+1}, z_{2r+1}, ..., $z_{(n-1)r+1}$ and can be obtained by left multiplication of (3.6.7) with the matrix $G = I_n \otimes h'$. ∎

Proposition 3.6.2

Suppose observations were made at times t_1, t_2, ..., t_m and let $l_1 = t_2 - t_1$, $l_2 = t_3 - t_2$, ..., $l_{m-1} = t_m - t_{m-1}$. The vector of observations, e: $m \times 1$ can be written as

$$e = (I_m \otimes h') T_\alpha^{-1}[T_\beta u + I_{rm,r} F z(t_1 - 1)]$$ (3.6.8)

where h and F have the same definition as in Proposition (3.6.1). The non-singular matrix T_α: $mr \times mr$ is

$$T_\alpha = \begin{bmatrix} I_r & 0 & \dots & \dots & \dots & 0 \\ -F^{l_1} & I_r & 0 & \dots & & \vdots \\ 0 & -F^{l_2} & I_r & 0 & \dots & \vdots \\ \vdots & & & & & \\ 0 & \dots & \dots & \dots & -F^{l_{m-1}} & I_r \end{bmatrix}.$$ (3.6.9)

3-30

The vector z: mr x 1 contains, amongst other information, the m observations and is given by

$$z = \begin{bmatrix} z(t_1) \\ z(t_2) \\ \vdots \\ z(t_m) \end{bmatrix} .$$
(3.6.10)

Without loss of generality, the time axis may be shifted by subtracting t_1 - 1 from each time point. We then have $t_1 = 1$, $t_2 = 1 + l_1$, $t_3 = 1 + l_1 + l_2$, ..., and $t_m = 1 + l_1$ $+ l_2 + ...$, $+ l_{m-1}$. Observations start at time 1 to time t_m with intervals equal to l_1, l_2, ..., l_{m-1}. The vector $z(0)$ denotes the state of the process before any observations were made. (In the complete data case all intervals are one and $t_m = n$.)

The simplest way to define T_β is by its rows. The matrix T_β is grouped by its rows into m submatrices each consisting of r rows and t_m columns. The total dimension of T_β is rm x t_m. The first row submatrix has vector g (defined in (3.3.4)) as its first column and zero vectors elsewhere. The j-th row submatrix, $j = 2, ..., m$, has zero vectors as columns except for column t_j and the preceding l_j - 1 columns. Column t_j is g, column t_j-1 is Fg, column t_j-2 is F^2g ..., column t_j-$(l_j$-1) is $F^{lj-1}g$. In matrix notation T_β can be expressed as

$$T_\beta = \begin{bmatrix} g & 0 & ... & ... & ... & ... & ... & ... & ... & ... & 0 \\ 0 & ... & 0 & F^{l_1-1}g & F^{l_1-2}g & ... & Fg & g & 0 & ... & 0 \\ \vdots \\ 0 & ... & ... & ... & ... & ... & ... & F^{l_{m-1}-1}g & ... & Fg & g \end{bmatrix} .$$
(3.6.11)

3-31

The t_m x 1 vector of random noise terms, u, is complete. $I_{rm,r}$ is defined as the first r columns of I_{rm}. After shifting the time axis, $z(t_1 - 1)$ is $z(0)$.

Proof

The state equation (3.3.2) is used to find the following expressions for $z(t_1)$, $z(t_2)$, ..., $z(t_m)$.

$$z(t_1) \quad = \quad Fz(t_1-1)+gu_{t_1} \tag{3.6.12}$$

$$z(t_1+1) \quad = \quad Fz(t_1)+gu_{t_1+1} \tag{3.6.13}$$

$$z(t_1+2) \quad = \quad F[Fz(t_1)+gu_{t_1+1}]+gu_{t_1+2}$$

$$\vdots$$

$$z(t_2) \quad = \quad z(t_1+l_1)$$

$$= \quad F^{l_1}z(t_1)+F^{l_1-1}gu_{t_1+1}+...+Fgu_{t_1+l_1-1}+gu_{t_2}$$

$$\vdots$$

$$z(t_m) \quad = \quad z(t_{m-1}+l_{m-1}) \tag{3.6.14}$$

$$= \quad F^{l_{m-1}}z(t_{m-1})+F^{l_{m-1}-1}gu_{t_{m-1}+1}+...+Fgu_{t_{m-1}+l_{m-1}-1}+gu_{t_m}$$

Equations (3.6.12) to (3.6.14) can be rearranged as

3-32

$$z(t_1) = gu_{t_1} + Fz(t_1-1) \qquad (3.6.15)$$

$$-F^{l_1}z(t_1) + z(t_2) = F^{l_1-1}gu_{t_1+1} + \ldots + Fgu_{t_1+l_1-1} + gu_{t_2}$$

$$-F^{l_2}z(t_2) + z(t_3) = F^{l_2-1}gu_{t_2+1} + \ldots + Fgu_{t_2+l_2-1} + gu_{t_3}$$

$$\vdots$$

$$-F^{l_{m-1}}z(t_{m-1}) + z(t_m) = F^{l_{m-1}-1}gu_{t_{m-1}+1} + \ldots + Fgu_{t_{m-1}+l_{m-1}-1} + gu_{t_m}.$$

In matrix notation (3.6.15) can be expressed as

$$T_\alpha z = T_\beta u + I_{rm,r} Fz(t_1-1). \qquad (3.6.16)$$

Since T_α is non-singular (triangular with all diagonal elements unity), z can be obtained by left multiplication of (3.6.16) by T_α^{-1} so that

$$z = T_\alpha^{-1}[T_\beta u + I_{rm,r} Fz(t_1-1)]. \qquad (3.6.17)$$

As in Proposition 3.6.1, the observations $e_{t_1}, e_{t_2}, \ldots e_{t_m}$ can be extracted from z by left multiplication of (3.6.17) by the matrix $G = I_m \otimes h'$. ∎

The following example illustrates this concept.

Example 3.6.1

Suppose observations at times 1, 3 , 6 and 7 of a stationary ARMA(2,1) process are available and are denoted by $e' = (e_1, e_3, e_6, e_7)$. It will be shown that e can be expressed in terms of the state vector of an ARMA process.

The state vector and state equation of an ARMA(2,1) process are given in (3.4.20) and

3-33

(3.4.21). For this example the intervals are $l_1 = 2$, $l_2 = 3$ and $l_3 = 1$, $m = 4$, $t_m = 7$ and the value of r is 2. From (3.6.12) through to (3.6.14) and the state equation (3.3.2) we get:

$$z(1) = gu_1 + Fz(0) \tag{3.6.18}$$

$$\begin{aligned} z(3) &= Fz(2) + gu_3 \\ &= F[Fz(1) + gu_2] + gu_3 \\ &= F^2z(1) + Fgu_2 + gu_3 \end{aligned} \tag{3.6.19}$$

$$z(6) = F^3z(3) + F^2gu_4 + Fgu_5 + gu_6 \tag{3.6.20}$$

$$z(7) = Fz(6) + gu_7 \tag{3.6.21}$$

The system of equations can be written in matrix notation as

$$\begin{bmatrix} I & 0 & 0 & 0 \\ -F^2 & I & 0 & 0 \\ 0 & -F^3 & I & 0 \\ 0 & 0 & -F & I \end{bmatrix} \begin{bmatrix} z(1) \\ z(3) \\ z(6) \\ z(7) \end{bmatrix} = \begin{bmatrix} g & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & Fg & g & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & F^2g & Fg & g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_7 \end{bmatrix} + \begin{bmatrix} Fz(0) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.6.22}$$

so that

$$z = T_\alpha^{-1}[T_\beta u + I_{8,2}Fz(0)] \tag{3.6.23}$$

3-34

Now e can be extracted from z by left multiplication of (3.6.23) by G where

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad (3.6.24)$$

$$= I_4 \otimes (1 \quad 0)$$

so that

$$e = GT_\alpha^{-1}[T_\beta u + I_{8,2} Fz(0)] \; . \qquad (3.6.25)$$

3.7 REPARAMETERIZATION OF ARMA COEFFICIENTS

The unknown ARMA coefficients are estimated by means of an iterative optimization algorithm. In the case of unconstrained optimization, new values of the coefficients may fall outside the parameter space of a stationary and invertible model. This problem may be solved in a number of ways.

One approach is to make use of a constrained optimization procedure where constraints are imposed on the parameters using Lagrange multipliers (see e.g. Magnus and Neudecker (1988), p. 130 and Browne and Du Toit (1992)).

A second approach is to estimate the roots of the characteristic equations, (3.2.5) and (3.2.6), instead of the ARMA coefficients. For stationarity and invertibility, the roots of the autoregressive and moving average portion must lie outside the unit circle. Suppose, for example, that the roots of equation (3.2.5) of an AR(2) model are λ_1^{-1} and

3-35

λ_2^{-1}. Then

$$[(1-\lambda_1 B)(1-\lambda_2 B)]e_t = u_t \qquad\qquad (3.7.1)$$

so that

$$[1-(\lambda_1+\lambda_2)B+\lambda_1\lambda_2 B^2]e_t = u_t. \qquad\qquad (3.7.2)$$

It follows that the AR coefficients are

$$\alpha_1 = \lambda_1+\lambda_2 \text{ and } \alpha_2 = -\lambda_1\lambda_2. \qquad\qquad (3.7.3)$$

For stationarity, $|\lambda_i| < 1$, $i = 1, 2$. It is therefore possible to put bounds on the values of λ_1 and λ_2 and carry out unconstrained optimization.

The procedure was implemented successfully in fitting an ARMA(2,1) model. An advantage is that it can be easily extended to models of higher order. A possible difficulty with this approach is that provision has to be made for complex roots and the programming language should therefore support complex algebra.

A third approach is to reparameterize in terms of the partial autocorrelations (Jones (1980 and 1993)). A necessary and sufficient condition for an AR(p) process to be stationary is that all p partial autocorrelations be between -1 and 1 (not inclusive). Instead of working with the autoregressive coefficients, the optimization is performed with respect to the partial autocorrelations.

Let α_{pj} denote the j-th coefficient of an AR(p) process. The partial autocorrelation at lag p is α_{pp}, the p-th autoregressive coefficient. The relationship between the partial autocorrelations and the AR coefficients is given by the Levinson-Durbin recursion (Box *et al* (1994), p.88). For processes of order 2, ..., p

3-36

$$\alpha_{kj} = \alpha_{k-1,j} - \alpha_{kk}\alpha_{k-1,k-j} \ , \quad j = 1,\ldots,k-1 . \tag{3.7.4}$$

As an example, expressions for α_1, α_2 and α_3, the AR coefficients of an AR(3) process, are derived in terms of the partial autocorrelations, α_{11}, α_{22} and α_{33}, using (3.7.4):

$$\alpha_{31} = \alpha_{21} - \alpha_{33}\alpha_{22} \tag{3.7.5a}$$

$$\alpha_{32} = \alpha_{22} - \alpha_{33}\alpha_{21} \tag{3.7.5b}$$

$$\alpha_{21} = \alpha_{11} - \alpha_{22}\alpha_{11} . \tag{3.7.5c}$$

By substitution of (3.7.5c) into (3.7.5a) and (3.7.5b) it follows that α_1 and α_2 are

$$\alpha_1 = \alpha_{11}(1 - \alpha_{22}) - \alpha_{33}\alpha_{22} \tag{3.7.6a}$$

$$\alpha_2 = \alpha_{22} - \alpha_{33}\alpha_{11}(1 - \alpha_{22}) \tag{3.7.6b}$$

The third AR coefficient is simply the third partial autocorrelation.

$$\alpha_3 = \alpha_{33} \tag{3.7.6c}$$

To reparameterize, equations (3.7.6a) through to (3.7.6c) must be expressed in terms of α_1, α_2 and α_3. The same procedure is performed on the moving average coefficients to ensure invertibility.

The MLE's of the coefficients of an ARMA(p,q) process can be obtained from expressions similar to (3.7.6a) through to (3.7.6c) and by using the invariance property of MLE's. The reparameterization was applied to the coefficients of an ARMA(2,1)

3-37

model. The difficulty with this approach, as with the root transformation, is finding the standard errors of the estimators of the ARMA coefficients. The optimization algorithm provides the inverse of the information matrix which is an approximation of the covariance matrix of the transformed parameters. Approximations of the standard errors of the ARMA coefficients can be obtained using a Taylor series expansion. The results for an AR(2) model are given as an example.

For notational simplicity, let the partial autocorrelations, α_{11} and α_{22}, be denoted by ρ_1 and ρ_2. From (3.7.4) it follows that

$$\alpha_1 = \rho_1(1-\rho_2) \qquad\qquad (3.7.7a)$$

and

$$\alpha_2 = \rho_2. \qquad\qquad (3.7.7b)$$

The optimization algorithm yields MLE's of ρ_1 and ρ_2, which are denoted by $\hat{\rho}' = (\hat{\rho}_1, \hat{\rho}_2)$, as well as an approximate covariance matrix, which will be denoted by S. By the invariance property of MLE's, the MLE's of α_1 and α_2 are respectively

$$\hat{\alpha}_1 = \hat{\rho}_1(1-\hat{\rho}_2) \qquad\qquad (3.7.8a)$$

and

$$\hat{\alpha}_2 = \hat{\rho}_2. \qquad\qquad (3.7.8b)$$

Let $r_0 = E(\hat{\rho})$ and let $f(\hat{\rho})$ be a function of $\hat{\rho}$. The second order Taylor series approximation of $f(\hat{\rho})$ in the point r_0 is

3-38

$$f(\hat{\rho}) = f(r_0) + g'(\hat{\rho} - r_0) + \frac{1}{2}(\hat{\rho} - r_0)'A(\hat{\rho} - r_0) \tag{3.7.9}$$

where g is the gradient vector and A is the Hessian matrix with respect to $f(\hat{\rho})$. The variance of $f(\hat{\rho})$ can be approximated by the first and second order moments of (3.7.9). By taking expected values on both sides of (3.7.9) we get

$$E[f(\hat{\rho})] = f(r_0) + \frac{1}{2}E[(\hat{\rho} - r_0)'A(\hat{\rho} - r_0)] . \tag{3.7.10}$$

The second moment of $f(\hat{\rho})$ is

$$E[f(\hat{\rho})^2] = [f(r_0)]^2 + \frac{1}{2}E[(\hat{\rho} - r_0)'B(\hat{\rho} - r_0)] \tag{3.7.11}$$

where

$$
\begin{aligned}
b_{ij} &= \frac{\partial^2 f(\hat{\rho})^2}{\partial \hat{\rho}_i \partial \hat{\rho}_j} \\
&= 2[g_i g_j + f(\hat{\rho})a_{ij}] ,
\end{aligned}
\tag{3.7.12}
$$

$$g_i = \frac{\partial f(\hat{\rho})}{\partial \hat{\rho}_i} \tag{3.7.13}$$

and

$$a_{ij} = \frac{\partial^2 f(\hat{\rho})}{\partial \hat{\rho}_i \partial \hat{\rho}_j} . \tag{3.7.14}$$

The covariance matrix of $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)'$ is required for inference regarding the distribution of $\alpha = (\alpha_1, \alpha_2)'$ and can be calculated from the first and second order

3-39

moments of the following functions:

$$\hat{\alpha}_1 = f_1(\hat{\rho}) = \hat{\rho}_1(1 - \hat{\rho}_2) \qquad (3.7.15a)$$

$$\hat{\alpha}_2 = f_2(\hat{\rho}) = \hat{\rho}_2 \qquad (3.7.15b)$$

$$\hat{\alpha}_1\hat{\alpha}_2 = f_3(\hat{\rho}) = \hat{\rho}_1\hat{\rho}_2(1 - \hat{\rho}_2) \qquad (3.7.15c)$$

The first and second order moments ((3.7.10) and (3.7.11)) of f_1 and f_2 are required for the variance of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ and the first order moment of f_3 is required for the covariance of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

3.8 SUMMARY

In this chapter results reported on in literature concerning the estimation of fixed coefficient ARMA models are discussed. These results are extended in subsequent chapters, to allow for different sets of model assumptions.

The likelihood function is given for an ARMA model with Gaussian white noise. The state-space representation of an ARMA process is given and it is shown how the Kalman recursive algorithm can be employed to estimate the coefficients of the ARMA process in the case of a complete as well as an incomplete data set.

An expression is derived for the covariance matrix of a stationary ARMA process which takes into account the state of the process before any observations were made. It is also shown how the likelihood function of the repeated measurements can be calculated without direct calculation of the inverse or determinant of the covariance matrix.

3-40

An expression is derived for the observation vector in terms of the state vector of an ARMA process for the case of a complete as well as an incomplete data set.

A reparameterization of the ARMA coefficients which is used in subsequent chapters is also included.

3-41

CHAPTER 4 RANDOM COEFFICIENT ARMA MODELS - NORMAL DISTRIBUTION

4.1 INTRODUCTION

In many experimental studies, repeated observations are made on each of several experimental units with the objective of fitting a response curve to the data. These are often called growth curve, repeated measure or longitudinal studies. This chapter focuses on responses recorded over time. Serial correlation may therefore exist between the measurements of an experimental unit.

Jones (1993) mentions different approaches to longitudinal data analysis. If the number of experimental units is large relative to the number of observations per unit, the multivariate analysis methods, which assume a specific covariance structure between the observations of an experimental unit, can be used. The idea of stochastic or random coefficients was introduced by Rao (1959) and developed by Potthoff and Roy (1964), Rao (1965) and Grizzle and Allen (1969).

Another approach is to analyze each subject (experimental unit) individually and then regard the summary measures (e.g. means and variances) as random samples from some distribution. The distribution of a summary measure depends on the number of observations in the sample. A problem therefore arises when there is an unequal number of observations per experimental unit. The distributions of these summary measures differ from experimental unit to experimental unit; consequently, these measures can not be regarded as a random sample from a common distribution.

A combination of the first two approaches is proposed. The basic assumption is that although the response pattern of the experimental units may differ, they can all be described by the same functional form. Differences between experimental units are modelled by allowing the coefficients of the model to be stochastic.

This chapter focuses on the residuals of some model (linear or non-linear) fitted to a data set consisting of repeated measurements taken on a sample of experimental units. It will be assumed that the residuals are generated by an autoregressive moving average (ARMA) process with random coefficients. Furthermore, a multivariate normal distribution is assumed as a probability model for the transformed coefficients. The ARMA coefficients will be transformed in order to make the assumption of multivariate normality more realistic.

Nicholls and Quinn (1982) discuss least squares and maximum likelihood estimation of random coefficient autoregressive models. Horvath (1991) derives maximum likelihood estimators for the parameters in linear discrete time systems, with random coefficients, under the assumption that the variances of the coefficients are equal and that the coefficients are mutually independent.

In Section 4.2 the model and assumptions are given. The transformation of the ARMA coefficients is also introduced.

Section 4.3 deals with the likelihood function and provides a useful approximation thereof. The Kalman filter is proposed to calculate the likelihood function when missing observations occur. The results of a simulation study, where the performance of the estimation procedure for a complete data set as well as a data set with missing observations is assessed, are given.

The purpose of Section 4.4 is to explain marginal maximum likelihood (MML) estimation and the use of the expected maximization (EM)-algorithm.

There is a strong correspondence between the random coefficient models and Bayesian models. Similarities and differences between these approaches are discussed in Section 4.5. In Section 4.6 Bayes estimates are derived for the transformed ARMA coefficients and white noise variance.

4-2

The Gibbs sampler can be used to estimate the posterior distributions of the coefficients. Section 4.7 deals with the Gibbs sampler and explains how it can be used to obtain Bayes estimates of the unknown parameters and their posterior distributions. Also, a numerical example based on simulated data is provided.

4.2 MODEL AND ASSUMPTIONS

Suppose a fixed number of observations, n, are taken over time on each of N experimental units. Let the vector $e_i = (e_1, \ldots, e_n)'$ denote the observations of the i-th experimental unit. These observations can, for instance, be residual terms of a model. The case is considered where the observations $e_1, \ldots, e_n$ are generated by a stationary ARMA(p,q) process. (See Section 3.2 for the definition and assumptions of an ARMA model.)

In many practical applications, it is reasonable to assume that the responses of the different experimental units are generated by ARMA processes, but that the coefficients of these processes may vary. The coefficients $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots \beta_q$ are assumed to be random variables which may be dependent.

The problem is to find a suitable distribution to describe the stochastic behaviour of the coefficients and simultaneously satisfy the stationarity and invertibility conditions of the process. Consider as an example the AR(2) process. For stationarity the coefficients α_1 and α_2 must lie within the triangular region

$$\alpha_1 + \alpha_2 < 1$$
$$\alpha_2 - \alpha_1 < 1$$
$$-1 < \alpha_2 < 1. \qquad (4.2.1)$$

The support set of the joint distribution of α_1 and α_2 must therefore be a subspace of (4.2.1). The higher the dimension of the model, the more complicated the stationarity

4-3

and invertibility regions become and the more difficult it will be to choose a suitable distribution for the coefficients.

Reparameterization of the ARMA coefficients is discussed in Section 3.7 for fixed coefficient models. When $p > 1$ or $q > 1$, the autoregressive or moving average coefficients are transformed to partial autocorrelations using (3.7.4). For stationarity and invertibility all partial autocorrelations, $\alpha_{11}, \ldots, \alpha_{pp}, \beta_{11}, \ldots, \beta_{qq}$, must lie between -1 and 1 (not inclusive). Note that $\alpha_{pp} = \alpha_p$ and that $\beta_{qq} = \beta_q$, and consequently only $\alpha_1, \ldots, \alpha_{p-1}$ and $\beta_1, \ldots, \beta_{q-1}$ are transformed. The following transformation can be used to transform the partial autocorrelations to variables $\eta_1, \ldots, \eta_{p+q}$ which all assume values over the real line:

$$\eta_k = ln\left[\frac{1+\alpha_{kk}}{1-\alpha_{kk}}\right]. \qquad (4.2.2)$$

The inverse transformation is

$$\alpha_{kk} = \frac{1-exp(-\eta_k)}{1+exp(-\eta_k)}. \qquad (4.2.3)$$

The vector $\eta = (\eta_1, \ldots, \eta_{p+q})'$ will be referred to as the vector of transformed ARMA coefficients.

In this chapter the following **assumptions** are made:

A1: The n x 1 vector $e_i = (e_1, \ldots, e_n)'$ represent observations at times $1, \ldots, n$ of a stationary ARMA(p,q) process. Let e_i, $i = 1, \ldots, N$, denote the vector of observations for experimental unit i. It is assumed that the vector of coefficients of the ARMA model are random and that the observed coefficients of each experimental unit are a random

4-4

sample from a common multivariate population.

A2: A transformation of the ARMA coefficients is given in (4.2.2). It is assumed that the vector η of transformed ARMA coefficients has a $N(\eta_0, \Psi)$ distribution where Ψ is non-singular and $\eta = (\eta_1, \ldots, \eta_{p+q})'$.

A3: The conditional distribution of e_i given η, is $N(0, \Sigma)$ where Σ is non-singular, $i = 1, \ldots, N$. The matrix Σ is defined in Section 3.5 (cf. (3.5.8)). This covariance matrix is derived under the assumption of a non-null initial state covariance matrix, in other words, information regarding the process prior to the first observation is taken into consideration.

A4: It is assumed that the vectors $e_1 | \eta$, $e_2 | \eta$, $\ldots$, $e_N | \eta$ are independent.

A5: The white noise variance, σ^2, is fixed and is the same for all experimental units.

Estimates are required for the components of η_0, $vecs(\Psi)$, the non-duplicate elements of Ψ and σ^2, the white noise variance. Let

$$\gamma = \begin{bmatrix} \eta_0 \\ vecs(\Psi) \\ \sigma^2 \end{bmatrix} \qquad (4.2.4)$$

denote the $k \times 1$ parameter vector where
$$k = p + q + \tfrac{1}{2}(p + q)(p + q + 1) + 1 .$$

Let $\hat{\eta}_0$ and $\hat{\Psi}$ denote the estimates of η_0 and Ψ. The $N(\hat{\eta}_0, \hat{\Psi})$ distribution may be used as an approximation of the distribution of η. The moments of the ARMA coefficients

4-5

can be expressed as functions of the components of η, and can be estimated by means of numerical integration using the approximate distribution of η. In the case of an ARMA(2,1) process, it follows from (3.7.7a), (3.7.7b) and (4.2.3) that

$$E(\alpha_1) = E[\alpha_{11}(1-\alpha_{22})]$$

$$= E\left\{\frac{2exp(-\eta_2)[1-exp(-\eta_1)]}{[1+exp(-\eta_1)][1+exp(-\eta_2)]}\right\} \qquad (4.2.5)$$

$$= E[g(\eta_1, \eta_2)]$$

$$\approx \int\int\int g(\eta_1, \eta_2) \cdot (2\pi)^{-\frac{3}{2}} |\hat{\Psi}|^{-\frac{1}{2}} exp[-\frac{1}{2}(\eta-\hat{\eta}_0)' \hat{\Psi}^{-1}(\eta-\hat{\eta}_0)] d\eta,$$

$$E(\alpha_2) = E(\alpha_{22})$$

$$= E\left[\frac{1-exp(-\eta_2)}{1+exp(-\eta_2)}\right] \qquad (4.2.6)$$

and

$$E(\beta) = E\left[\frac{1-exp(-\eta_3)}{1+exp(-\eta_3)}\right] \qquad (4.2.7)$$

Expressions for higher order moments are derived in the same way. Numerical integration can also be used to calculate estimates of the moments of the joint distribution of the coefficients.

In the next section the likelihood function of $e_1, \ldots, e_N$ for the ARMA(p,q) model with random coefficients will be derived.

4-6

4.3 MAXIMUM LIKELIHOOD ESTIMATION

Likelihood function

The following proposition gives the likelihood function of $e_1, \ldots, e_N$ for the ARMA(p,q) model with random coefficients defined as in the previous section.

Proposition 4.3.1

Let $e_i = (e_1, \ldots, e_n)'$, denote n repeated measurements made on the experimental unit i, $i = 1, \ldots, N$. Suppose that these measurements are generated by a stationary ARMA(p,q) process with random coefficients $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$, partial autocorrelations $\alpha_{11}, \ldots, \alpha_{pp}, \beta_{11}, \ldots, \beta_{qq}$ and white noise variance σ^2. Let η_j, the j-th component of the $(p+q)$ x 1 vector η, be defined by (4.2.2). If $e_i | \eta$ has a $N(0, \sigma^2 \Lambda)$ distribution, where Λ is defined by (3.5.18), and η has a $N(\eta_0, \Psi)$ distribution, the likelihood of a random sample $e_1, \ldots, e_N$ is given by

$$L(\gamma) = (2\pi)^{-\frac{1}{2}(nN+p+q)}(\sigma^2)^{-\frac{1}{2}nN}|\Psi|^{-\frac{1}{2}} \int \ldots \int |\Lambda|^{-\frac{1}{2}N} \cdot$$

$$exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} e_i' \Lambda^{-1} e_i\right] \cdot exp[-\tfrac{1}{2}(\eta-\eta_0)' \Psi^{-1}(\eta-\eta_0)]\, d\eta$$

$$(4.3.1)$$

Proof

The conditional joint density of $e_1, \ldots, e_N$ given η is (cf. (2.2.10))

4-7

$$f(e_1, \ldots, e_N | \eta) = \prod_{i=1}^{N} f(e_i | \eta)$$

$$= (2\pi\sigma^2)^{-\frac{1}{2}nN} |\Lambda|^{-\frac{1}{2}N} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} e_i' \Lambda^{-1} e_i \right] \qquad (4.3.2)$$

and the marginal density of η is

$$g(\eta) = (2\pi)^{-\frac{1}{2}(p+q)} |\Psi|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\eta-\eta_0)' \Psi^{-1}(\eta-\eta_0)] \ . \qquad (4.3.3)$$

The joint density of $e_1, \ldots, e_N$ and η is

$$h(e_1, \ldots, e_N, \eta) = f(e_1, \ldots, e_N | \eta) g(\eta) . \qquad (4.3.4)$$

The likelihood function is the marginal density of $e_1, \ldots, e_N$:

$$L(\gamma) = \int \ldots \int h(e_1, \ldots, e_N, \eta) \, d\eta \qquad (4.3.5)$$

so that (4.3.1) follows by substitution. ∎

Since the elements of Λ are non-linear functions of the components of η, it is not possible to find a closed form solution to the integral (4.3.1). Proposition 4.3.2 provides a useful approximation of the value of this integral.

4-8

Proposition 4.3.2

An approximation of the likelihood function (4.3.1) is

$$L(\gamma) \approx \pi^{-\frac{1}{2}(nN+p+q)}(2\sigma^2)^{-\frac{1}{2}nN}\sum_{i=1}^{m}w_i\sum_{j=1}^{m}w_j\ldots\sum_{l=1}^{m}w_l f(x_i,x_j,\ldots,x_l) \qquad (4.3.6)$$

where $w_1, \ldots, w_m$ and $x_1, \ldots, x_m$ are the weights and abscissas of an m-point Gauss-Hermite quadrature formula and the scalar function $f(x)$ is

$$f(x) = |\Lambda(x)|^{-\frac{1}{2}N} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}e_i'\Lambda(x)^{-1}e_i\right]. \qquad (4.3.7)$$

Proof

The likelihood function (4.3.1) can be approximated by numerical integration (see Section 2.4). A suitable transformation of η yields a product type integral with Gauss-Hermite weight functions. Let the $(p + q)$- component vector v be defined by (cf. (2.4.14))

$$v = \sqrt{\frac{1}{2}}\Psi^{-\frac{1}{2}}(\eta-\eta_0) \qquad (4.3.8)$$

where $\Psi^{-\frac{1}{2}\prime}\Psi^{-\frac{1}{2}} = \Psi^{-1}$.

The corresponding transformation of the integral in terms of v is

4-9

$$L(\gamma) = \pi^{-\frac{1}{2}(nN+p+q)}(2\sigma^2)^{-\frac{1}{2}nN} \int \cdots \int exp(-v'v)\,|\Lambda(v)|^{-\frac{1}{2}N}\cdot$$

$$exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} e_i'\Lambda^{-1}(v)e_i\right]\,dv \qquad\qquad (4.3.9)$$

where $\Lambda(v)$ denotes the matrix Λ in terms of v instead of η, as in (4.3.1). The integral in (4.3.9) can be written as a product type integral, so that

$$L(\gamma) = \pi^{-\frac{1}{2}(nN+p+q)}(2\sigma^2)^{-\frac{1}{2}nN} \int \cdots \int exp(-v_1^2-v_2^2- \cdots -v_{p+q}^2)f(v_1,v_2, \cdots v_{p+q})\,dv_1\cdots dv_{p+q}$$

$$= \pi^{-\frac{1}{2}(nN+p+q)}(2\sigma^2)^{-\frac{1}{2}nN} \int exp(-v_1^2) \int exp(-v_2^2)\cdots \int exp(-v_{p+q}^2)f(v)\,dv$$

$$(4.3.10)$$

where the function $f(.)$ is defined by (4.3.7). Note that $exp(-v_i^2)$ is the Gauss-Hermite weight function (see Table 2.4.1). Numerical integration can therefore be used to approximate (4.3.10), so that result (4.3.6) follows. ∎

Remarks

The discrepancy function $-2lnL$ which can be used to find maximum likelihood estimates is given by

4-10

$$-2lnL \approx (nN+p+q)ln\,\pi+nNln(2\sigma^2)-2ln\sum_{i=1}^{m}\sum_{j=1}^{m}\,...\,\sum_{l=1}^{m} w_i w_j...w_l\,f(x_i,x_j,...,x_l) \quad (4.3.11)$$

where $f(x_i,...,x_l)$ is given by (4.3.7).

It is clear that the derivative of $-2lnL$ with respect to any component of γ is rather complicated and can only be approximated with numerical techniques. The MLE's can be obtained by an iterative optimization algorithm.

Such an algorithm will require a large number of evaluations of (4.3.11) during each iteration. The number of computations required by (4.3.10) depends on the number of repeated measurements (n), the number of experimental units (N), the number of parameters (k) and the number of quadrature terms (m). Computation time is to a very large extent influenced by k and it is therefore, apart from other reasons, wise to not overparameterize.

Two algorithms were implemented to evaluate $-2lnL$. In the first, the quadratic form $e_i'\Lambda(x)^{-1}e_i$ and the determinant $|\Lambda(x)|$ are calculated using (3.5.26) and (3.5.25) respectively. The second algorithm uses Kalman recursions, a procedure that is easily adapted to handle data sets with missing observations.

The Kalman recursions, as given in Section 3.4, can be used to calculate the likelihood function of a single series with fixed coefficients. The function $f(x)$ in (4.3.7) can be written as

$$f(x) = \prod_{i=1}^{N} |\Lambda(x)|^{-\frac{1}{2}} exp\left[-\frac{1}{2\sigma^2}e_i'\Lambda(x)^{-1}e_i\right] \quad (4.3.12)$$

or, alternatively, in terms of the innovations as

$$f(x) = exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}u_i'u_i\right]$$

(4.3.13)

where $u_i = L^{-1}e_i$ and L is the Choleski root of $\Lambda(x)$. For every evaluation of the likelihood function, the quadratic form, $u_i'u_i$ or $e_i'\Lambda(x)^{-1}e_i$, is calculated Nm^{p+q} times. When an iterative optimization procedure is used to maximize the likelihood function and both the gradient vector and Hessian matrix are determined numerically, the likelihood function value is calculated many times per iteration. It is therefore important to use a computationally efficient algorithm to determine the quadratic form.

Missing observations

Incomplete data sets or data sets with unequally spaced observations are often encountered in practice, and even more so in repeated measurement experiments where a number of experimental units participate in the study.

In Chapter 3, two methods used to calculate the likelihood function of a fixed coefficient ARMA process in the case of missing values were discussed. Jones (1980) proposed the use of the Kalman recursive procedure to calculate the likelihood function of an ARMA model when there are missing observations (see Section 3.4). An alternative is to use (4.3.12) with rows and columns corresponding to the missing observations removed from the matrix Λ.

Suppose the m_i observations for experimental unit i were made at times $t_{i,1},\ldots t_{i,m_i}$, $i = 1, \ldots , N$. Let $e_1, \ldots , e_N$ denote the observation vectors of the N experimental units. Note that their dimensions can vary depending on the number of missing observations.

4-12

Maximum likelihood estimates can be obtained by minimization of (4.3.11). The Kalman filter can be used to calculate the innovation terms of a experimental unit recursively. When a missing observation occurs, the innovation term is zero and the process is adapted accordingly.

Maximum likelihood estimates can also be obtained by adapting the covariance matrix of each experimental unit by removing the rows and columns from the complete data covariance matrix that correspond to the missing observations. The function $f(x)$ is formulated as (cf. (4.3.7))

$$f(x) = \prod_{i=1}^{N} |\Lambda_i(x)|^{-\frac{1}{2}} exp\left[-\frac{1}{2\sigma^2}e'_i\Lambda_i(x)^{-1}e_i\right] \qquad (4.3.14)$$

where

$$\Lambda_i(x) = \frac{1}{\sigma^2}\Sigma_i(x) \qquad (4.3.15)$$

and $\Sigma_i(x)$ is the covariance matrix of the observations of the i-th experimental unit.

It is important to keep in mind that a basic assumption of the Kalman filter is that the innovation terms must be Gaussian. The Kalman filter is therefore not appropriate when the distribution of the white noise terms is not normal. This situation is explored in Chapter 6.

Computational considerations

In this subsection problems that arise during the iterative estimation of the unknown parameters (4.2.4) are discussed.

4-13

The first problem concerns Ψ, the covariance matrix of η. The optimization algorithm requires that Ψ be positive definite at every step. This can be accomplished by expressing Ψ in its Gaussian decomposition and reparameterizing its elements (Bock (1990)). The Gaussian decomposition of Ψ is given by

$$\Psi = LDL' \tag{4.3.16}$$

where L is a unit lower triangular matrix and D is a diagonal matrix parameterized as

$$D = diag(\exp(\tau_i)) \quad , i = 1, ..., p+q. \tag{4.3.17}$$

For Ψ to be positive definite, all diagonal elements of D must be positive. Note that the diagonal elements of D are forced to be positive, regardless of the sign of τ_i , $i = 1, ..., p+q$. Minimization of the discrepancy function is carried out with respect to τ_i , $i = 1, ..., p+q$ and the elements of L, instead of Ψ.

If the dimension of Ψ is 3, it can be written as

$$\Psi = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} \exp(\tau_1) & 0 & 0 \\ 0 & \exp(\tau_2) & 0 \\ 0 & 0 & \exp(\tau_3) \end{bmatrix} \begin{bmatrix} 1 & l_{21} & l_{31} \\ 0 & 1 & l_{32} \\ 0 & 0 & 1 \end{bmatrix} \tag{4.3.18}$$

$$= \begin{bmatrix} \exp(\tau_1) & l_{21}\exp(\tau_1) & l_{31}\exp(\tau_1) \\ l_{21}\exp(\tau_1) & l_{21}^2\exp(\tau_1)+\exp(\tau_2) & l_{21}l_{31}\exp(\tau_1)+l_{32}\exp\tau_2 \\ l_{31}\exp(\tau_1) & l_{21}l_{31}\exp(\tau_1)+l_{32}\exp(\tau_2) & l_{31}^2\exp(\tau_1)+l_{32}^2\exp(\tau_2)+\exp(\tau_3) \end{bmatrix} . \tag{4.3.19}$$

4-14

By setting $\Psi_{11} = \exp(\tau_1)$, it follows that $\tau_1 = ln\Psi_{11}$. Similarly, expressions can be obtained for τ_2, τ_3, l_{21}, l_{31} and l_{32} in terms of the elements of Ψ.

Another problem that may be encountered when using numerical integration, is that the exponential term

$$exp\left[-\frac{1}{2\sigma^2}e_i'\Lambda(x)^{-1}e_i\right] \tag{4.3.20}$$

may become zero since it is determined over a wide range of values of x (a vector of quadrature abscissas). The problem can be reduced by adding a positive constant in the exponent. The last term of the discrepancy function (4.3.11)

$$-2ln\sum_{i=1}^{m}\sum_{j=1}^{m}...\sum_{l=1}^{m} w_i w_j...w_l \prod_{k=1}^{N} |\Lambda(x)|^{-\frac{1}{2}}exp\left[-\frac{1}{2\sigma^2}e_k'\Lambda(x)^{-1}e_k\right] \tag{4.3.21}$$

can be written as

$$2cN - 2ln\sum_{i=1}^{m}\sum_{j=1}^{m}...\sum_{l=1}^{m} w_i w_j...w_l \prod_{k=1}^{N} |\Lambda(x)|^{-\frac{1}{2}}exp\left[-\frac{1}{2\sigma^2}e_k'\Lambda(x)^{-1}e_k + c\right] . \tag{4.3.22}$$

The iterative procedure used to find maximum likelihood estimates may not converge to the true minimum value of $-2lnL$ if the initial parameter estimates are not close to their real minimum values. The success of the estimation procedure therefore depends to a large extent on the initial estimates. A method by which to obtain initial estimates is discussed in the next paragraph.

4-15

Initial estimates of parameters

Box *et al* (1994) give a general method for obtaining initial estimates of the parameters of a mixed ARMA process, where the parameters are the ARMA coefficients and the white noise variance. Their procedure forms the basis of a procedure that can be used where there are more than one experimental unit in a study and where the parameters are elements of the mean vector and covariance matrix of the ARMA coefficients.

For every independent time series, initial estimates of the ARMA coefficients can be calculated following the three steps given by Box *et al*. The N initial estimates of every coefficient are examined carefully by means of a histogram in order to identify possible outliers or irregularities that may indicate departures from assumptions. The coefficients are transformed (the transformations for an ARMA(p,q) model are given in (3.7.4) and (4.2.2)) in order to assume values over the real line. An initial estimate of η_0, the mean of η, is the mean or median of the sample of transformed coefficients. The elements of the covariance matrix of η, Ψ, can be estimated initially by the sample covariance matrix of the transformed coefficients.

The three steps to be carried out in order to obtain initial estimates of the coefficients of a single series are explained thus by Box *et al*:

The *first step* consists of calculating the initial estimates of the autoregressive coefficients by using the following result:

$$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \ldots \alpha_p\gamma_{k-p} \ , \ k \geq q+1 \tag{4.3.23}$$

where $\gamma_0, \ldots , \gamma_{p+q}$ are the autocovariances at lags $0, \ldots , p+q$ of the process. Let c_0, $c_1, \ldots , c_{p+q}$ denote the sample autocovariances at lags $0, \ldots , p+q$. From (4.3.24) it follows that $\hat{\alpha}_1, \ldots , \hat{\alpha}_p$ may be obtained by solving the p linear equations

$$C_{q+1} = \hat{\alpha}_1 C_q + \hat{\alpha}_2 C_{q-1} + \ldots + \hat{\alpha}_p C_{q-p+1}$$

$$C_{q+2} = \hat{\alpha}_1 C_{q+1} + \hat{\alpha}_2 C_q + \ldots + \hat{\alpha}_p C_{q-p+2}$$

$$\vdots$$

$$C_{q+p} = \hat{\alpha}_1 C_{q+p-1} + \hat{\alpha}_2 C_{q+p-2} + \ldots + \hat{\alpha}_p C_q \ . \tag{4.3.24}$$

Using the estimates $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$ obtained in the first step, the *second step* consists of the calculation of the first $q+1$ autocovariances of the derived series $w_t = \hat{\alpha}(B)e_t$. The process $\{w_t, \ t \le n\}$ is treated as a moving average process of order q. Let the autocovariances of w_t be denoted by c_0', $\ldots$, c_q'. The following relationship with the autocovariances of e_t is useful:

$$c_j' = \sum_{i=0}^{p} (\alpha_0 \alpha_i + \alpha_1 \alpha_{i+1} + \ldots + \alpha_{p-i} \alpha_p)(c_{j+i} + c_{j-i}) \quad \text{for } j = 0, \ldots , q \tag{4.3.25}$$

where $\alpha_0 = -1$.

The autocovariances c_0', $\ldots$, c_q' are used in *step three* in an iterative procedure to compute initial estimates of the moving average parameters and σ^2 (Wilson (1969)). The autocovariance of an MA process is

$$c_i = Cov(e_t, e_{t+i}) = \sum_{j=0}^{k-|i|} \beta_j \beta_{j+|i|} \quad \text{for } |i| \le q \ . \tag{4.3.26}$$

Let $\tau' = (\tau_0, \ldots , \tau_q)$ where

$$\sigma^2 = \tau_0^2,$$
$$\beta_j = -\tau_j/\tau_0, \ j = 1, \ldots , q. \tag{4.3.27}$$

4-17

If successive approximations towards the final solution are denoted by attaching a subscript (i), the $i+1$-th solution obtained by a Newton-Raphson algorithm is

$$\tau_{(i+1)} = \tau_{(i)} - (T_{(i)})^{-1} f_{(i)} \qquad (4.3.28)$$

where T, the Jacobian matrix, is defined as

$$T = \begin{bmatrix} \tau_0 & \tau_1 & \cdots & \tau_{q-1} & \tau_q \\ \tau_1 & \tau_2 & \cdots & \tau_q & 0 \\ \tau_2 & \tau_3 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \tau_q & 0 & \cdots & 0 & 0 \end{bmatrix} + \begin{bmatrix} \tau_0 & \tau_1 & \tau_2 & \cdots & \tau_q \\ 0 & \tau_0 & \tau_1 & \cdots & \tau_{q-1} \\ 0 & 0 & \tau_0 & \cdots & \tau_{q-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \tau_0 \end{bmatrix} \qquad (4.3.29)$$

and $f = (f_0, \ldots, f_q)'$ where $f_j = \sum_{i=0}^{q-j} \tau_i \tau_{i+j} - c_j'$. The elements of f are minimized when

the covariances of w_t, c_0', $\ldots$, c_q', are equal to the covariances obtained when calculating (4.3.26) using the fitted parameter values $\hat{\beta}_1, \ldots, \hat{\beta}_q$.

Starting values for the iterative process can be taken as $\tau_0 = \sqrt{c_0'}$, $\tau_1 = \ldots , \tau_q = 0$. The iterative process is continued until $|f_i| < \epsilon$, $i = 0, \ldots , q$ for some prescribed value of ϵ. When convergence is attained, the values of the parameters can be obtained using (4.3.27).

In the case of a pure autoregressive process, an estimate of the white noise variance is

4-18

$$\hat{\sigma}^2 = c_0 - \sum_{i=1}^{p} \hat{\alpha}_i c_i . \tag{4.3.30}$$

Simulation study

A simulation study was carried out to examine the properties of the maximum likelihood estimates obtained by minimizing (4.3.11) for both a complete and an incomplete data set.

Estimates were obtained for thirty different data sets. These data sets were generated to satisfy assumptions A1 to A5 (Section 4.2). Each data set consists of the 50 repeated measurements on each of 50 experimental units. The repeated measurements are the output of a stationary ARMA(2,1) process with random coefficients.

Estimates were obtained for the complete data set as well as for the same data set with ten percent of its values missing. The location of the missing values in the data set were determined by a random number generator. Numerical integration was performed with four quadrature terms ($m = 4$).

The mean and standard deviation of the thirty estimates of each parameter are given in Table 4.3.1. The values used in generating the data sets are also reported.

4-19

Table 4.3.1: Simulation study results for a random coefficient ARMA(2,1) model.

Parameter	Population value	Complete data		Incomplete data	
		Mean	S.D.	Mean	S.D.
$E(\alpha_1)$	0.25	0.2267	0.0894	0.1977	0.0782
$E(\alpha_2)$	0.5	0.4525	0.0951	0.4985	0.0475
$E(\beta)$	0.6	0.5121	0.1215	0.5129	0.1171
σ^2	2.0	2.1765	0.0662	2.3198	0.0782
$V(\alpha_1)$	0.015	0.0132	0.0262	0.0070	0.0096
$V(\alpha_2)$	0.03	0.0314	0.0421	0.0174	0.0226
$V(\beta)$	0.04	0.0376	0.0406	0.0702	0.0798
$Cov(\alpha_1,\alpha_2)$	0.00	-0.0163	0.0322	-0.0057	0.0063
$Cov(\alpha_1,\beta)$	0.00	0.0033	0.0134	0.0001	0.0130
$Cov(\alpha_2,\beta)$	0.00	-0.0022	0.0244	0.0033	0.0215
$-2lnL$		9210	76.64	8282	77.07

A number of conclusions can be drawn from these results. In both the complete and incomplete cases, the means of the autoregressive coefficients are estimated more accurately (judged by their smaller estimated standard errors) than the mean of the moving average coefficient. The variances of the ARMA coefficients are estimated more accurately for the complete data sets (judged by their difference from the

4-20

population value). The white noise variance is over-estimated. An approximate 95% confidence interval (mean $\pm$ 2 standard deviations) does not include the population value of 2.0. The over-estimation of σ^2 is probably due to increased variability in the data as a result of the variability of the ARMA coefficients. An increase in the number of quadrature terms may yield an improved estimate of σ^2.

4.4 MARGINAL MAXIMUM LIKELIHOOD AND THE EM-ALGORITHM

In this section marginal maximum likelihood (MML) by applying the expected maximization (EM)-algorithm (Dempster, Laird and Rubin (1977)) is discussed as an alternative method of calculating maximum likelihood estimates.

Let the vector e_i represent n repeated measurements made on experimental unit i and assume that e_i can be described by a stationary ARMA(p,q) model with white noise variance σ^2. Let η_k, $k = 1, \ldots, p+q$, be transformations of the ARMA coefficients defined by (4.2.2). The assumption is made that $e_i|\eta$ has a $N(0,\sigma^2\Lambda)$ distribution and that η has a $N(\eta_0,\Psi)$ distribution. The parameters $\gamma' = (\eta_0', vecs(\Psi)', \sigma^2)$ can be estimated by means of MML estimation (Bock (1990)).

The likelihood of $e_1, \ldots, e_N$ is (cf. (4.3.5))

$$L(e_1, \ldots e_N) = \int \cdots \int f(e_1, \ldots, e_N|\eta)g(\eta)\,d\eta \qquad (4.4.1)$$

which can be written as

$$L(e_1, \ldots e_N) = \int \cdots \int \left[\prod_{i=1}^{N} f(e_i|\eta) \right] g(\eta)\,d\eta \qquad (4.4.2)$$

where $g(\eta)$ denotes the marginal density of η, which is assumed to be $N(\eta_0, \Psi)$ and $f(e_i|\eta)$ denotes the conditional distribution of e_i given η which is assumed to be $N(0, \sigma^2 \Lambda)$.

The log-likelihood is consequently given by

$$lnL = ln \int \dots \int f(e_1, \dots e_N | \eta) g(\eta) d\eta. \tag{4.4.3}$$

Let $\theta' = (\eta_0', vecs(\Psi)')$. The marginal density, $g(\eta)$, is a function of θ, and $f(e_1, \dots, e_N | \eta)$ is a function only of σ^2. The parameter vector, γ, can therefore be divided into two sets, namely $\gamma' = (\theta', \sigma^2)$. It follows that

$$\frac{\partial lnL}{\partial \theta_j} = \frac{\int \dots \int f(e_1, \dots, e_N | \eta) \frac{\partial g(\eta)}{\partial \theta_j} d\eta}{f(e_1, \dots, e_N)}. \tag{4.4.4}$$

By using the following well known result:

$$\frac{\partial lng(\eta)}{\partial \theta_j} = \frac{\partial g(\eta)}{\partial \theta_j} \cdot \frac{1}{g(\eta)} \tag{4.4.5}$$

expression (4.4.4) can be written as

$$\frac{\partial lnL}{\partial \theta_j} = \int \dots \int \frac{f(e_1, \dots, e_N | \eta) \cdot g(\eta)}{f(e_1, \dots, e_N)} \frac{\partial lng(\eta)}{\partial \theta_j} d\eta. \tag{4.4.6}$$

The function

4-22

$$p(\eta \,|\, e_1, \ldots, e_N) = \frac{f(e_1, \ldots, e_N \,|\, \eta) g(\eta)}{f(e_1, \ldots, e_N)} \qquad (4.4.7)$$

is the conditional distribution of η given e_i, $i = 1, \ldots, N$, so that

$$\frac{\partial lnL}{\partial \theta_j} = E_{\eta | e_1, \ldots, e_N} \left[\frac{\partial lng(\eta)}{\partial \theta_j} \right] . \qquad (4.4.8)$$

The corresponding equation for σ^2 is

$$\frac{\partial lnL}{\partial \sigma^2} = E_{\eta | e_1, \ldots, e_N} \left[\frac{\partial}{\partial \sigma^2} \sum_{i=1}^{N} \ln f(e_i \,|\, \eta) \right] . \qquad (4.4.9)$$

From the assumption that η is $N(\eta_0, \Psi)$ it follows that

$$lng(\eta) = -\tfrac{1}{2}(p + q) ln 2\pi - \tfrac{1}{2} ln |\Psi| - \tfrac{1}{2} tr \left[\Psi^{-1}(\eta - \eta_0)(\eta - \eta_0)' \right] \qquad (4.4.10)$$

where $tr(A)$ denotes the trace of the square matrix A.

Let η_{0j} denote the j-th element of η_0. It follows from (2.2.18) that

$$\frac{\partial lng(\eta)}{\partial \eta_{0j}} = tr \left[(\eta - \eta_0)' \Psi^{-1} J_{j1} \right]$$

$$= \left[\Psi^{-1}(\eta - \eta_0) \right]_{j1} \qquad (4.4.11)$$

4-23

(where J_{j1} is a column vector with all elements equal to zero with the exception of the j-th element which is equal to unity) and hence

$$\frac{\partial lng(\eta)}{\partial \eta_0} = \Psi^{-1}(\eta - \eta_0).$$ (4.4.12)

The MML estimate of η_0 is therefore obtained by solving

$$E_{\eta|e_1,\dots,e_N}\left[\Psi^{-1}(\eta - \hat{\eta}_0)\right] = 0$$ (4.4.13)

and hence

$$\hat{\eta}_0 = E_{\eta|e_1,\dots,e_N}(\eta).$$ (4.4.14)

It is clear that $\hat{\eta}_0$ is the mean of the conditional distribution of $\eta|e_1, \dots, e_N$.

Let ψ_{rs} denote a typical element of $vecs(\Psi)$. It follows that (cf. (2.2.19))

$$\begin{aligned}
\frac{\partial lng(\eta)}{\partial \psi_{rs}} &= \tfrac{1}{2}tr\left\{\Psi^{-1}\left[(\eta-\eta_0)(\eta-\eta_0)' - \Psi\right]\Psi^{-1}\frac{\partial \Psi}{\partial \psi_{rs}}\right\} \\
&= \tfrac{1}{2}tr\left\{\Psi^{-1}\left[(\eta-\eta_0(\eta-\eta_0)' - \Psi\right]\Psi^{-1}\left[J_{rs} + (1-\delta_{rs})J_{sr}\right]\right\}
\end{aligned}$$ (4.4.15)

where δ_{rs} is Kronecker's delta, and J_{rs} is the matrix with all elements equal to zero with the exception of the element in the r-th row and s-th column which is equal to unity.

4-24

The matrix $G = (\eta - \eta_0)(\eta - \eta_0)'$ can be written as

$$G = \left[\eta - \mathrm{E}_{\eta|e}(\eta) - (\eta_0 - \mathrm{E}_{\eta|e}(\eta))\right]\left[\eta - \mathrm{E}_{\eta|e}(\eta) - (\eta_0 - \mathrm{E}_{\eta|e}(\eta))\right]'$$

$$= \left[\eta - \mathrm{E}_{\eta|e}(\eta)\right]\left[\eta - \mathrm{E}_{\eta|e}(\eta)\right]' - \left[\eta_0 - \mathrm{E}_{\eta|e}(\eta)\right]\left[\eta - \mathrm{E}_{\eta|e}(\eta)\right]'$$

$$- \left[\eta - \mathrm{E}_{\eta|e}(\eta)\right]\left[\eta_0 - \mathrm{E}_{\eta|e}(\eta)\right]' + \left[\eta_0 - \mathrm{E}_{\eta|e}(\eta)\right]\left[\eta_0 - \mathrm{E}_{\eta|e}(\eta)\right]'. \qquad (4.4.16)$$

where the notation $\eta|e$ is used to denote $\eta|e_1, \ldots, e_N$.

The likelihood can be concentrated with respect to η_0 by substituting $\hat{\eta}_0$ so that an estimate for G is

$$\hat{G} = (\eta - \hat{\eta}_0)(\eta - \hat{\eta}_0)'. \qquad (4.4.17)$$

The expected value of $\hat{G}$ with respect to the conditional distribution of $\eta|e_1, \ldots, e_N$ is

$$\mathrm{E}_{\eta|e}(\hat{G}) = \mathrm{Cov}_{\eta|e}(\eta, \eta') + \left[\hat{\eta}_0 - \mathrm{E}_{\eta|e}(\eta)\right]\left[\hat{\eta}_0 - \mathrm{E}_{\eta|e}(\eta)\right]'. \qquad (4.4.18)$$

By using (4.4.8) and (4.4.15) it follows that

$$\frac{\partial lnL}{\partial \psi_{rs}} = \mathrm{E}_{\eta|e}\left\{\frac{1}{2}\mathrm{tr}\left[\Psi^{-1}(\hat{G} - \Psi)\Psi^{-1}[J_{rs} + (1 - \delta_{rs})J_{sr}]\right]\right\}. \qquad (4.4.19)$$

The solution of the likelihood equation with respect to Ψ yields

$$\frac{\partial lnL}{\partial \Psi} = 0 \Rightarrow \mathrm{E}_{\eta|e}(\hat{G} - \Psi) = 0 \qquad (4.4.20)$$

4-25

and hence

$$\hat{\Psi} = \mathrm{E}_{\eta|e}(\hat{G})$$

$$= \mathrm{Cov}_{\eta|e}(\eta,\eta') + [\hat{\eta}_0 - \mathrm{E}_{\eta|e}(\eta)][\hat{\eta}_0 - \mathrm{E}_{\eta|e}(\eta)]' \quad . \tag{4.4.21}$$

Under the assumption that $e_i | \eta$ is $N(0,\sigma^2\Lambda)$ the conditional log-likelihood of e_i given η is

$$lnf(e_i|\eta) = -\frac{1}{2}\left[nln(2\pi\sigma^2) + ln|\Lambda| + \frac{1}{\sigma^2}e_i'\Lambda^{-1}e_i \right] \tag{4.4.22}$$

and the derivative with respect to σ^2 is

$$\frac{\partial lnf(e_i|\eta)}{\partial\sigma^2} = -\frac{1}{2}\left[\frac{n}{\sigma^2} - \frac{1}{\sigma^4}e_i'\Lambda^{-1}e_i \right] . \tag{4.4.23}$$

An MML estimate for σ^2 is obtained by substituting (4.4.23) in (4.4.9) and solving the likelihood equation. The result is

$$\hat{\sigma}^2 = \frac{1}{nN}\sum_{i=1}^{N}\mathrm{E}_{\eta|e}(e_i'\Lambda^{-1}e_i) . \tag{4.4.24}$$

The MML estimates (4.4.14), (4.4.21) and (4.4.24) are all given in terms of the conditional distribution of $\eta|e_1, \ldots, e_N$ defined in (4.4.7), which is the posterior distribution of η in Bayes terminology. In this application $p(\eta|e_1, \ldots, e_N)$ is not a standard distribution and numerical integration is used to obtain the required expected values.

4-26

The same procedure can be used to find expressions for the elements of the Hessian matrix in terms of conditional expectations. The second derivative of lnL with respect to θ_i and θ_j can be written as

$$\frac{\partial lnL}{\partial \theta_i \partial \theta_j} = \frac{\int \cdots \int f(e_1, \ldots, e_N | \eta) \frac{\partial^2 g(\eta)}{\partial \theta_i \partial \theta_j} d\eta}{f(e_1, \ldots, e_N)}$$ (4.4.25)

where, by using (4.4.5)

$$\frac{\partial^2 g(\eta)}{\partial \theta_i \partial \theta_j} = g(\eta) \left[\frac{\partial lng(\eta)}{\partial \theta_i} \cdot \frac{\partial lng(\eta)}{\partial \theta_j} + \frac{\partial^2 lng(\eta)}{\partial \theta_i \partial \theta_j} \right]$$ (4.4.26)

and therefore (cf. (4.4.8))

$$\frac{\partial^2 lnL}{\partial \theta_i \partial \theta_j} = E_{\eta | e_1, \ldots, e_N} \left[\frac{\partial lng(\eta)}{\partial \theta_i} \cdot \frac{\partial lng(\eta)}{\partial \theta_j} + \frac{\partial^2 lng(\eta)}{\partial \theta_i \partial \theta_j} \right].$$ (4.4.27)

The first order derivatives of $ln\, g(\eta)$ are given by (4.4.11) and (4.4.15) and the second order derivatives can be derived using standard matrix differentiation results.

The elements of the Hessian matrix can be evaluated numerically with Gauss quadrature and the Gauss-Newton algorithm, which require the Hessian matrix, can be used to obtain maximum likelihood estimates.

The expected maximization (EM)-algorithm (Dempster, Laird and Rubin (1977)) can also be used as an optimization algorithm. Starting values for the parameters are used to calculate $p(\eta | e_1, \ldots, e_N)$, $E_{\eta | e_1, \ldots, e_N}(\eta)$, $Cov_{\eta | e_1, \ldots, e_N}(\eta, \eta')$ and $E_{\eta | e_1, \ldots, e_N}(e_i' \Lambda^{-1} e_i)$. This step

4-27

is the expectation step. The values are substituted into (4.4.14), (4.4.21) and (4.4.24) to find estimates for η_0, Ψ and σ^2, (the maximization step). This process is repeated until convergence is attained.

The EM-algorithm is robust to poor initial estimates, but may require many iterations to converge relative to algorithms that use first and second order derivative information. A good strategy is to switch to such an algorithm after a number (say 100) of EM-iterations.

Note that it is possible to obtain estimates for the ARMA coefficients of each experimental unit. The estimated distribution of η is $N(\hat{\eta}_0, \hat{\Psi})$. The mean of the conditional distribution of $\eta | e_i$ can be used as an estimate of the transformed coefficients for the i-th experimental unit:

$$[\hat{\eta}_i]_{jl} = [E(\eta | e_i)]_{jl} = \int \cdots \int [\eta]_{jl} \frac{f(e_i | \eta) h(\eta)}{f(e_i)} d\eta. \qquad (4.4.28)$$

Estimates of the ARMA coefficients for the i-th experimental unit can be obtained in a similar way, by first expressing the ARMA coefficients as functions of the elements of η. This concept was also discussed in Section 4.2.

4.5 RANDOM COEFFICIENTS AND BAYESIAN THEORY

There is a degree of similarity between the estimation approach used in random coefficient models and Bayes estimation (see e.g. DeGroot (1970)). The purpose of this section is to discuss the similarities in and differences between these two approaches.

Let $y_1, \ldots, y_n$ represent a univariate random sample from some population of which the probability distribution depends on unknown parameters, denoted by γ.

4-28

A fundamental assumption in Bayes theory is that all parameters are random variables. The random coefficient approach allows a mixture of random variables and fixed parameters. In this chapter the ARMA coefficients are assumed random while the white noise variance is fixed. The basis for the assumption of random coefficients is that in repeated measurement models with a number of experimental units, the same order model may be adequate to describe all experimental units, but the coefficients are allowed to vary across the experimental units.

Both approaches require specification of the distribution of the parameters. Denote this distribution by $h(\gamma, \gamma_0)$ where γ_0 denotes the parameters of h. In Bayesian theory the function $h(\gamma, \gamma_0)$ is termed the prior distribution. The parameters of the prior distribution (*hyperparameters*), γ_0, must be specified. In the random coefficient approach this distribution is seen as the marginal distribution of the coefficients and the aim is to use the data to estimate γ_0.

Let $f(y_1, \dots, y_n | \gamma)$ denote the joint distribution of the sample given fixed values of the parameters. Bayesians call this the likelihood of the data given the parameters. In this study it is referred to as the conditional distribution of the data given the parameter values.

The product $g(y_1, \dots, y_n, \gamma) = f(y_1, \dots, y_n | \gamma) \, h(\gamma, \gamma_0)$ is calculated in both approaches. Apart from a proportionality constant, g is the posterior distribution of the parameters, used in Bayesian inference. The marginal posterior distributions are derived and Bayes estimates are calculated for the components of γ.

In the random coefficient approach, the marginal density function of $y_1, \dots, y_n$ is determined by integration of g over the values of γ:

$$L = \int \dots \int g(y_1, \dots y_n, \gamma) d\gamma \ . \tag{4.5.1}$$

4-29

This function is the likelihood of the data with respect to given parameter values. Its maximum points with respect to the components of γ_0 are maximum likelihood estimates. The parameters of the marginal density of γ are estimated; such density can be used to find expected values of any function of γ. In Section 4.2 the moments of the ARMA coefficients are derived by writing them as functions of normal variables and then estimating the parameters as described above.

The main points of correspondence between the Bayesian approach and the random coefficient approach as applied in this dissertation, are therefore the common assumption of random coefficients and the use of conditional probability results.

4.6 BAYES ESTIMATION

Bayes estimation of the coefficients of an ARMA process were studied by a number of authors, amongst whom Broemeling (1985). Recent articles (Chib (1993) and Chib and Greenberg (1993)) deal with Bayes inference in regression models with ARMA(p,q) errors.

Bayes estimation of the parameters of an ARMA model is given in this section in order to highlight differences that exist between the Bayes approach and the random coefficients approach.

Let e_i denote n repeated measurements made on experimental unit i, $i = 1, \ldots, N$. It is assumed that the elements of e_i are generated by an ARMA(p,q) process with white noise variance σ^2. The Bayes estimates of the coefficients of the ARMA process generating e_i and the white noise variance have to be determined.

Instead of estimating the ARMA coefficients directly, they are transformed as explained in Section 4.2. Let η denote the $p+q$-dimensional vector of transformed coefficients.

4-30

The prior distributions of η and σ^2 are

$$\eta \sim N(\eta_p, \Psi_p) \tag{4.6.1}$$

and

$$f(\sigma^2) \propto \frac{1}{\sigma^2} \quad , \sigma^2 > 0 \tag{4.6.2}$$

where η_p and Ψ_p are the prior values of η and Ψ, which are assumed to be known. It is also assumed that η and σ^2 are independent.

As in Section 4.2, it is assumed that the conditional distribution of $e_i | \eta$ is $N(0, \sigma^2 \Lambda)$. The likelihood of $e_1, \ldots, e_N$ is consequently given by (4.3.2).

The joint posterior distribution of η and σ^2 is

$$p(\eta, \sigma^2 | e_1, \ldots, e_N) = \frac{f(e_1, \ldots, e_N | \eta, \sigma^2) p(\sigma^2) p(\eta)}{f(e_1, \ldots, e_N)} \tag{4.6.3}$$

where $p(\sigma^2)$ and $p(\eta)$ denote the prior distributions of σ^2 and η, $f(e_1, \ldots, e_N | \eta, \sigma^2)$ denotes the likelihood and the denominator is a normalizing constant. It follows that

$$p(\eta, \sigma^2 | e_1, \ldots e_N) \propto (2\pi)^{-\frac{1}{2}(nN+p+q)} (\sigma^2)^{-\frac{1}{2}(nN+2)} |\Psi_p|^{-\frac{1}{2}} |\Lambda|^{-\frac{1}{2}N} \cdot$$

$$exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} e_i' \Lambda^{-1} e_i \right] \cdot exp[-\tfrac{1}{2}(\eta - \eta_p)' \Psi_p^{-1} (\eta - \eta_p)] \tag{4.6.4}$$

4-31

where the proportionality constant is

$$\frac{1}{f(e_1, \dots, e_N)} .$$

(4.6.5)

The marginal posterior of η is obtained by integrating of (4.6.4) over σ^2 and is given by

$$p(\eta \mid e_1, \dots e_N) = \int p(\eta, \sigma^2 \mid e_1, \dots e_N) d\sigma^2$$

(4.6.6)

and hence

$$p(\eta \mid e_1, \dots e_N) \propto (2\pi)^{-\frac{1}{2}(nN+p+q)} \mid \Psi_p \mid^{-\frac{1}{2}} \mid \Lambda \mid^{-\frac{1}{2}N} \exp[-\tfrac{1}{2}(\eta - \eta_p)' \Psi_p^{-1}(\eta - \eta_p)] \cdot$$

$$\int (\sigma^2)^{-\frac{1}{2}(nN+2)} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} e_i' \Lambda^{-1} e_i \right] d\sigma^2$$

(4.6.7)

where the proportionality constant is given by (4.6.5).

The integral can be solved by using the *inverted gamma* distribution. The density function of an inverted gamma distribution (Kotz *et al* (1983), vol. 4) with parameters λ and β (denoted by $IG(\lambda, \beta)$) is

$$f(y) = \frac{\lambda^\beta}{\Gamma(\beta)} \exp\left[-\frac{\lambda}{y} \right] \left[\frac{1}{y} \right]^{\beta+1} \quad , y > 0 .$$

(4.6.8)

4-32

The mean and variance of y are

$$E(y) = \frac{\lambda}{\beta-1} \quad \text{for } \beta > 1 \tag{4.6.9}$$

and

$$\text{Var}(y) = \frac{\lambda^2}{(\beta-1)^2(\beta-2)} \quad \text{for } \beta > 2. \tag{4.6.10}$$

The integral in (4.6.7) can be solved by using the inverted gamma distribution so that the marginal posterior of η is

$$p(\eta|e_1,...e_N) \propto (2\pi)^{-\frac{1}{2}(nN+p+q)} |\Psi_p|^{-\frac{1}{2}} |\Lambda|^{-\frac{1}{2}N} \exp[-\tfrac{1}{2}(\eta-\eta_p)'\Psi_p^{-1}(\eta-\eta_p)] \cdot$$

$$\frac{\Gamma\left[\frac{1}{2}nN\right]}{\left[\frac{1}{2}\sum_{i=1}^{N} e_i'\Lambda^{-1}e_i\right]^{\frac{1}{2}nN}} \tag{4.6.11}$$

which is not a standard distributional form. When using the squared error loss function, the Bayes estimate of the i-th element of η, η_i, is the mean of η_i with respect to (4.6.11) given by

$$\hat{\eta}_i = \int ... \int \eta_i \, p(\eta|e_1,...,e_N)d\eta. \tag{4.6.12}$$

4-33

The integral can be determined numerically by means of a Gauss-Hermite quadrature formula. The covariance matrix of η, $\text{Cov}(\eta,\eta')$ is an estimate of Ψ, and can also be determined by means of numerical integration.

The marginal posterior distribution of σ^2 is

$$p(\sigma^2 | e_1, ... e_N) = \int ... \int p(\eta,\sigma^2 | e_1, ... e_N) d\eta \qquad (4.6.13)$$

and from (4.6.4) it follows that

$$p(\sigma^2 | e_1, ... e_N) \propto (2\pi)^{-\frac{1}{2}(nN+p+q)}(\sigma^2)^{-\frac{1}{2}(nN+2)}|\Psi_p|^{-\frac{1}{2}}\int ... \int |\Lambda|^{-\frac{1}{2}N} \cdot$$

$$\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} e_i'\Lambda^{-1}e_i\right] \cdot \exp[-\tfrac{1}{2}(\eta-\eta_p)'\Psi_p^{-1}(\eta-\eta_p)] \; \partial\eta \qquad (4.6.14)$$

where the proportionality constant is given by (4.6.5). Since the elements of Λ are non-linear functions of η, it is clear that the integral has to be determined numerically. The Bayes estimate of σ^2 is

$$\hat{\sigma}^2 = \int \sigma^2 \; p(\sigma^2 | e_1, ... , e_N) d\sigma^2. \qquad (4.6.15)$$

4.7 THE GIBBS SAMPLER

In Sections 4.3 and 4.4 maximum likelihood and marginal maximum likelihood were discussed as techniques for estimating the parameters of a random coefficient ARMA model. Bayes estimators of the unknown parameters were derived in Section 4.6. All

4-34

these methods depend on numerical integration. Gibbs sampling does not require any integration and can be used as an alternative to numerical integration. Gibbs sampling is a technique for generating random variables from a distribution without having to calculate the density (Casella and George (1992)). Suppose certain characteristics of the marginal distributions of a collection of random variables, x_1 , ... , x_k, have to be examined. The Gibbs sampler requires the full conditional distributions of x_1 , ... , x_k, denoted by $f(x_1|x_2$, ... , $x_k)$, $f(x_2|x_1$, x_3 ... , $x_k)$ ... , $f(x_k|x_1$, ... , $x_{k-1})$ to obtain characteristics of their marginal distributions. Note that it is not always easy or even possible to derive the marginal distributions from the full conditionals.

The straightforward approach to finding a marginal distribution involves integration, which in some cases can be difficult. By applying the Gibbs sampler these complex calculations are replaced with a sequence of simpler calculations. Gibbs sampling may be indispensable in situations where the joint distribution of x_1 , ... , x_k or the marginal distributions can not be calculated.

The most common use of the Gibbs sampler has been in Bayesian models in which features of the posterior distribution, such as the marginal posterior distribution of a parameter, are difficult to evaluate (MacEachern and Berliner (1994)).

The theory of Gibbs sampling is based on elementary properties of Markov chains. Given an arbitrary set of starting values denoted by $x_1^{(0)}$, ... , $x_k^{(0)}$, a random sample can be obtained by successive sampling from the following conditional distributions:

$x_1^{(1)}$ from $f(x_1|x_2^{(0)}$, ... , $x_k^{(0)})$,

$x_2^{(1)}$ from $f(x_2|x_1^{(1)}$, $x_3^{(0)}$, ... , $x_k^{(0)})$,

$\vdots$

$x_k^{(1)}$ from $f(x_k|x_1^{(1)}$, $x_2^{(1)}$, ... , $x_{k-1}^{(1)})$.

$$(4.7.1)$$

This cycle is repeated and after i such iterations the observations $(x_1^{(i)}, x_2^{(i)}, \ldots, x_k^{(i)})$ are recorded. The successive observations from each variable, $x_j^{(0)}, x_j^{(1)}, \ldots, x_j^{(i)}, j = 1, \ldots, k$, form a Markov chain which converge to the marginal distribution as i increases. Geman and Geman (1984) showed that under mild conditions the joint distribution of $(x_1^{(i)}, x_2^{(i)}, \ldots, x_k^{(i)})$ tends to that of $(x_1, \ldots, x_k)$ and hence the distribution of $x_j^{(i)}$ tends to that of x_j, $j = 1, \ldots, k$; as $i \rightarrow \infty$.

Casella and George (1992) state that a sufficient condition for convergence to occur is that the marginal density should be proper. The success of Gibbs sampling depends on the asymptotic behaviour of the induced Markov chain. For the general applicability of the method it is desirable that the Markov chain be ergodic. Chan (1993) obtained mild sufficient conditions for the ergodicity of the Gibbs sampler.

In order to estimate the marginal density or its characteristics a large sample must be generated. Several methods of generation have been suggested. The first method is to repeat the procedure for one observation a large number of times, say m. In other words, m independent Markov chains of length i are generated and the sample consists of the i-th observation of each chain. These m observations can be regarded as a random sample from the marginal density, on condition that the chains are independent with each chain using a different starting value $x_1^{(0)}, x_2^{(0)}, \ldots, x_k^{(0)}$ (Van der Merwe and Botha (1993)).

The second method is to generate one long Gibbs sequence and to include all the observations after a certain point, the so called *burn-in period* (MacEachern and Berliner (1994)) or to sample in a systematic way from the sequence.

Geyer (1992) argues against subsampling the output of a stationary Markov chain, preferring instead to use the entire chain, except possibly for a relatively short burn-in period. According to MacEachern and Berliner (1994) the variance of the Gibbs estimators increase when subsampling from the sequence.

4-36

However, Smith and Roberts (1993) state that it appears to be more efficient to generate one long Gibbs sequence and, after an initial transient phase, form a sample by collecting equally spaced outcomes. The gaps are chosen to render serial correlation negligible. When a new problem is analyzed, they suggest that several chains should be generated from a wide range of initial values to monitor the evolutionary behaviour of the chain.

Once the sample is obtained, the marginal density and its characteristics can be estimated to any desired degree of accuracy. The mean of x_j, $j = 1, \dots ,k$ can be estimated as the sample mean

$$\hat{\mu}_j = \frac{1}{m}\sum_{l=1}^{m} x_{jl}^{(i)} \tag{4.7.2}$$

where $x_{jl}^{(i)}$ is the l-th observation of x_j. For cases where the means of the conditional distributions are available, an improved estimate is

$$\hat{\mu}_j = \frac{1}{m}\sum_{l=1}^{m} \mathrm{E}(x_j \,|\, x_{1l}^{(i)}, \dots, x_{j-1,l}^{(i)}, x_{j+1,l}^{(i)}, \dots, x_{kl}^{(i)}). \tag{4.7.3}$$

Any other moment can be estimated similarly.

The average of the conditional density of x_j given the remaining x's over the observations is a good approximation of the marginal density of x_j. The estimate is

$$\hat{f}(x_j) = \frac{1}{m}\sum_{l=1}^{m} f(x_j \,|\, x_{1l}^{(i)}, \dots, x_{j-1,l}^{(i)}, x_{j+1,l}^{(i)}, \dots, x_{kl}^{(i)}). \tag{4.7.4}$$

4-37

The Gibbs sampler can be used to estimate the parameters of a stationary $\text{ARMA}(p,q)$ model with random coefficients.

Bayes estimation of the parameters of an $\text{ARMA}(p,q)$ model

In this subsection assumptions A1 to A4 given in Section 4.2 are adhered to. In contrast with assumption A5 the white noise variance is assumed to be a random variable. The Gibbs sampler requires that the prior distribution of each parameter be specified.

The prior distribution of the transformed coefficients is given in (4.6.1) and the prior distribution of σ^2 is taken proportional to $\dfrac{1}{\sigma^2}$ (cf. (4.6.2)). The joint distribution of e_1,

... , e_N, η and σ^2 is

$$
\begin{aligned}
f(e_1,\ldots,e_N,\eta,\sigma^2) &= f(e_1,\ldots,e_N | \eta,\sigma^2) p(\eta) p(\sigma^2) \\
&\propto (\sigma^2)^{-\frac{1}{2}(nN+2)} |\Lambda|^{-\frac{1}{2}N} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} e_i'\Lambda^{-1}e_i\right] exp\left[-\tfrac{1}{2}(\eta-\eta_p)'\Psi^{-1}(\eta-\eta_p)\right]
\end{aligned}
$$

$$(4.7.5)$$

where $p(\eta)$ and $p(\sigma^2)$ denote the prior densities of η and σ^2 respectively.

The full conditional distribution of σ^2 is

$$
f(\sigma^2 | e_1,\ldots,e_N,\eta) \propto (\sigma^2)^{-\frac{1}{2}(2+nN)} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} e_i'\Lambda^{-1}e_i\right].
$$

$$(4.7.6)$$

It is clear that (4.7.6) is proportional to an inverse gamma density (cf. (4.6.8)) with

4-38

parameters

$$\lambda = \frac{1}{2}\sum_{i=1}^{N} e_i' \Lambda^{-1} e_i \text{ and } \beta = \frac{nN}{2}. \qquad (4.7.7)$$

The conditional mean of σ^2 is (cf. (4.6.9))

$$E(\sigma^2 | e_1, ..., e_N, \eta) = \frac{\frac{1}{2}\sum_{i=1}^{N} e_i' \Lambda^{-1} e_i}{\frac{1}{2}nN} \qquad (4.7.8)$$

which is equal to the maximum likelihood estimate of σ^2 for the case where the ARMA coefficients are assumed fixed (cf. (6.1.10)).

Sampling from the $IG(\lambda, \beta)$ distribution can be accomplished by first generating an observation, z, from a $\chi^2(2\beta)$ distribution. Then $y = \frac{2\lambda}{z}$ is an observation from the $IG(\lambda, \beta)$ distribution.

Sampling from the other conditional distributions is not straightforward. The problem is that these distributions are not standard forms such as the normal or beta distribution. The conditional distribution of the j-th component of η is

$$f(\eta_j|e_1,...,e_N,\eta_{-j},\sigma^2) \propto |\Lambda_j|^{-\frac{1}{2}N} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}e_i'\Lambda_j^{-1}e_i\right] \cdot$$

$$exp\left[-\frac{1}{2}(\eta_j-\eta_p)'\Psi_p^{-1}(\eta_j-\eta_p)\right]$$

$$(4.7.9)$$

where the notation η_{-j} is used to denote η with its j-th component omitted, written as η_{-j} $= (\eta_1, \ldots, \eta_{j-1}, \eta_{j+1}, \ldots, \eta_{p+q})'$, Λ_j is Λ with all the η's, except η_j, fixed and similarly, η_j is η with all the η's, except η_j, fixed.

The inverted distribution function method (Johnson (1987), p.19) may be used to sample from (4.7.9). A random number, u, is generated from the uniform(0;1) distribution and the following equation is solved for η_j:

$$F(\eta_j|e_1,...,e_N,\eta_{-j},\sigma^2) = u$$

$$(4.7.10)$$

where $F( )$ is the cumulative distribution function (CDF) associated with (4.7.9). A closed form expression for the CDF is not available. Equation (4.7.10) is solved by evaluating the CDF for different values in the range of η_j and then choosing the value of η_j closest to u. The more values of η_j for which the CDF is evaluated, the closer will the generated values resemble sample values from the population. Calculating the CDF is, however, a time-consuming task and there is a trade-off between computation speed and accuracy.

Computation time drastically increases with every additional parameter. For every observation from a non-standard distribution, the CDF must be evaluated a number of times (say 100, for accuracy up to the second decimal). Furthermore, the Gibbs sampler requires large samples. Unless an efficient algorithm for sampling from non-standard distributions is employed, the Gibbs sampler is not efficient for models of a high order

under assumptions A1 to A4.

The following example shows how the Gibbs sampler is implemented to estimate the parameters of an AR(1) model.

Example 4.7.1

The Gibbs sampler is used in this example to calculate Bayes estimates of the parameters of an AR(1) model. A data set containing 20 repeated measurements on each of 35 experimental units is generated in accordance with assumptions A1 to A5 (Section 4.2). Each time series is generated by an AR(1) process with $E(\alpha) = -0.75$, $Var(\alpha) = 0.007$ and $\sigma^2 = 5.0$.

Instead of generating values directly from the distribution of α, values for η (cf. (4.2.2)) are generated from the $N(\eta_0, \Psi)$ distribution. The expected value of η, η_0, and the variance of η, Ψ, are approximated by means of a second order Taylor series expansion (cf. (3.7.9)).

The prior distributions of α and σ^2 are

$$\alpha \sim uniform(-1;1) \tag{4.7.11}$$

and

$$f(\sigma^2) \propto \frac{1}{\sigma^2} \quad , \sigma^2 > 0 . \tag{4.7.12}$$

The likelihood function of $e_1, \ldots , e_{35}$ is

4-41

$$f(e_1, \ldots, e_{35} | \alpha, \sigma^2) = (2\pi\sigma^2)^{-350} |\Lambda|^{-\frac{35}{2}} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{35} e_i' \Lambda^{-1} e_i \right] . \tag{4.7.13}$$

From (4.7.11) to (4.7.13) it follows that the joint distribution of $e_1, \ldots, e_{35}$, α and σ^2 is

$$f(e_1, \ldots, e_{35}, \alpha, \sigma^2) \propto (\sigma^2)^{-351} |\Lambda|^{-\frac{35}{2}} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{35} e_i' \Lambda^{-1} e_i \right] . \tag{4.7.14}$$

The full conditional distribution of σ^2 is

$$f(\sigma^2 | e_1, \ldots, e_{35}, \alpha) \propto (\sigma^2)^{-351} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{35} e_i' \Lambda^{-1} e_i \right] \tag{4.7.15}$$

which is an inverted gamma distribution (cf. (4.6.8)) with parameters

$$\lambda = \frac{1}{2} \sum_{i=1}^{35} e_i' \Lambda^{-1} e_i \quad \text{and} \quad \beta = 350.$$

The full conditional distribution of α is

$$f(\alpha | e_1, \ldots, e_{35}, \alpha) \propto |\Lambda|^{-\frac{35}{2}} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{35} e_i' \Lambda^{-1} e_i \right] \tag{4.7.16}$$

4-42

which is not a standard form since the elements of Λ are non-linear functions of α.

It was established empirically that the Markov chains obtained by generating values from (4.7.15) and (4.7.16) converge quite fast and it was decided to generate chains of length $i = 20$. A sample of $m = 100$ values was obtained for each of α and σ^2.

The autoregressive parameter and white noise variance are estimated as the sample means of the generated α's and σ^2's (cf.(4.7.2)) and are given by $\hat{\alpha} = -0.7601$ and $\hat{\sigma}^2 = 5.0396$. The sample standard deviations of α and σ^2 are $\hat{\sigma}_\alpha = 0.0273$ and $\hat{\sigma}_{\sigma^2} = 0.0241$.

The marginal densities of α and σ^2 are estimated using (4.7.4) and are presented graphically by Figures 4.7.1 and 4.7.2.
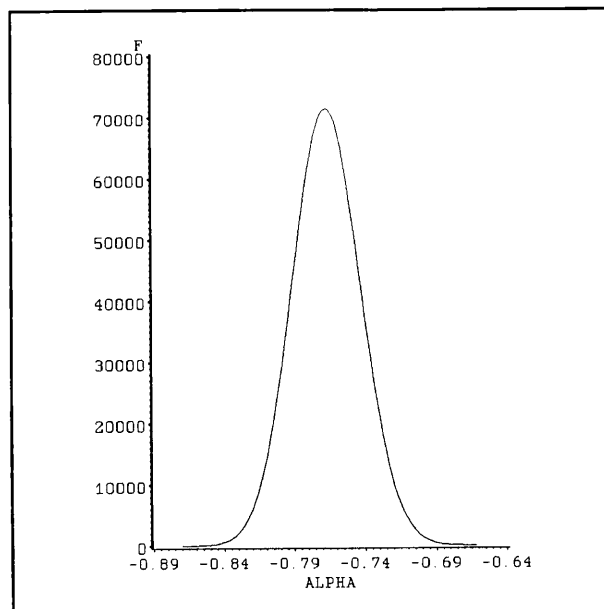
Figure 4.7.1 Estimated probability density of α

4-43
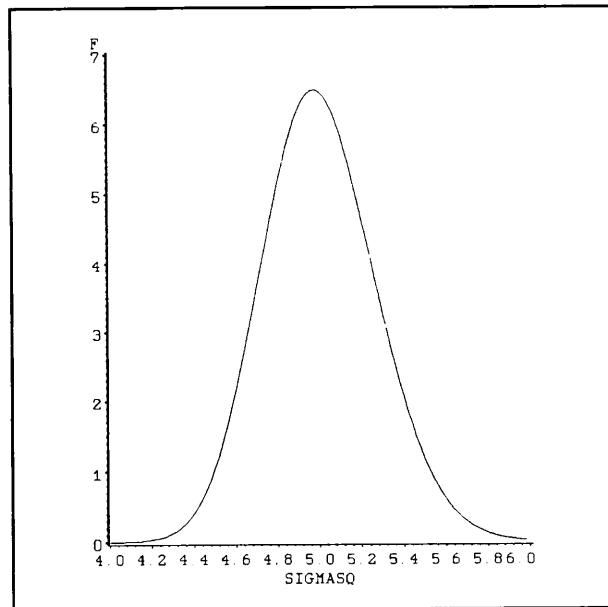
Figure   4.7.2   Estimated probability density of σ^2

Note that Figures 4.7.1 and 4.7.2 represent the density functions up to a scaling factor and that the scaling of the vertical axis is therefore not important.

The problems associated with the use of non-standard distributions in Gibbs sampling can be alleviated to some extent by choosing other prior distributions (this implies a replacement of assumptions A2 and A3). Chib and Greenberg (1993) developed a method for analysing ARMA(p,q) regression error models in a Bayesian framework by using Gibbs sampling and the Metropolis-Hastings algorithm. Their results can be adapted for the special case of no independent variables. A brief summary of their method is given.

Let $\alpha = (\alpha_1, \ldots, \alpha_p)'$ and $\beta = (\beta_1, \ldots, \beta_q)'$ denote the autoregressive and moving average parameters respectively and let $z(0)$ denote the initial state vector. The state-space representation of an ARMA process as given by Harvey (1981) is used. The prior distributions of α and β are assumed to be multivariate normal distributions truncated to their stationary and invertible regions respectively. An inverted gamma distribution

4-44

is used as the prior distribution of σ^2. The components of the initial state vector, $z(0)$, are introduced as parameters and their prior distribution is multivariate normal with mean zero. It is assumed that the parameters of the prior distributions (hyperparameters) are known.

Two transformations are performed on the data and the components of the initial state vector are introduced as additional parameters. The full conditional distributions that are used in the Gibbs sampling algorithm are a truncated multivariate normal distribution for α, an inverted gamma distribution for σ^2 and a multivariate normal distribution for $z(0)$. Kalman smoothing is used to obtain the parameters of the distribution of $z(0)$.

The full conditional distribution of β does not belong to a known family of distributions but can be simulated with the Metropolis-Hastings (MH) algorithm (Metropolis *et al* (1953) and Hastings (1970)). The MH algorithm is most effective when it is possible to draw from a density that resembles the target density. A truncated multivariate normal distribution whose parameters are revised at each iteration is proposed. The authors claim that at least 50% of the draws from this distribution should be accepted if the algorithm is implemented properly.

The method is justified by proving that the kernel density of the induced Markov chain converges to the true density. An advantage of the method of Chib *et al* is that the convergence of the Markov chain is improved by sampling blocks of parameters.

Simulated examples and an application on the U.S. real GNP data are given where samples as large as 6000 are taken beyond a burn-in period of 200 observations.

4.8 SUMMARY

In this chapter the estimation of the parameters of a random coefficient ARMA model is considered.

4-45

The exact likelihood function of the observations of an ARMA process is given as well as an approximation thereof based on numerical integration. It is shown that a Kalman recursive algorithm can be used in situations where the data is incomplete. The results of a simulation study which was carried out to examine the properties of the maximum likelihood estimates for complete as well as incomplete data sets, are given.

The concept of marginal maximum likelihood estimation is discussed together with the use of the EM-algorithm in order to obtain maximum likelihood estimates.

Since the assumptions made in random coefficient models are similar to Bayes assumptions, the Bayes estimates of the coefficients of an ARMA process are given. It is shown how the Gibbs sampler can be used to calculate Bayes estimates. An example of the implementation of the Gibbs sampler based on simulated data is given.

The reader should note that the application of Bayes techniques are not limited to stochastic parameter models. These techniques are generally applied to fixed as well as stochastic parameter models. A complete coverage of Bayes inference in all the models studied is, however, beyond the scope of this dissertation.

CHAPTER 5   REGRESSION MODELS WITH ARMA(p,q) ERROR TERMS

5.1 INTRODUCTION

In many practical applications, a response is measured on several occasions on each of a number of experimental units. To study the growth pattern of a specific tree species, for instance, the base circumference of a number of these trees can be measured periodically over a number of years. It is reasonable to expect that a functional relationship will exist between the circumference and the age of the tree.

In this chapter the assumption is made that the observed responses exhibit a linear or non-linear trend over time and additionally that the error terms of the model are generated by an ARMA(p,q) process.

Suppose n repeated measurements, denoted by $y_1, \ldots, y_n$, were made on an experimental unit. In the previous chapters the situation was considered where $y_j = e_j$, $j = 1, \ldots, n$, and where $e_1, \ldots, e_n$, were assumed to be generated by an ARMA(p,q) process with fixed or stochastic coefficients.

In this chapter the model is expanded to allow for a functional relationship between the dependent variable, y, and the time variable, t, which assumes the values $t_1, \ldots, t_n$, the points in time at which $y_1, \ldots, y_n$ are observed. If, for instance, there is a linear relationship over time, the measurement at time t_i can be written as

$$y_i = \beta_0 + \beta_1 t_i + e_i \, , \quad i = 1, \ldots n. \tag{5.1.1}$$

The estimation of the parameters of regression models with ARMA(p,q) error terms and for which the regression parameters and ARMA coefficients are assumed either fixed or stochastic are now discussed.

5-1

In Section 5.2 the maximum likelihood estimation of the parameters of a regression model with ARMA(p,q) error terms is considered. All parameters in the model are assumed fixed but unknown.

Section 5.3 deals with the situation where the parameters of the regression model are random and the error terms are generated by a fixed coefficient ARMA(p,q) process.

In Section 5.4, on the other hand, the parameters of the regression model are assumed fixed while the coefficients of the ARMA model describing the error terms are assumed random.

Both the regression parameters and the ARMA coefficients are assumed random in Section 5.5. It is shown that the method of marginal maximum likelihood can be used to obtain MLE's for the unknown parameters of such a model.

Considerable attention has been devoted in the literature to regression models with time series errors. Several models for analysing repeated measurement data exist in the literature. Harville (1977) and Jennrich and Schluchter (1986) give a review of the models and of approaches for estimating parameters. Linear random effects models for repeated measurement (longitudinal) data are discussed by, amongst others, Laird and Ware (1982), Jennrich and Schluchter (1986) and Lange and Laird (1989). The analysis of non-linear models with random coefficients are considered by authors such as Palmer *et al* (1991), Hirst *et al* (1991) and Herbst (1994). Racine *et al* (1986) discuss a Bayesian approach to non-linear random effects models.

5.2 FIXED PARAMETER REGRESSION MODEL WITH ARMA(p,q) ERRORS

Chib (1993) states that the regression model with autocorrelated errors is one of the most heavily analyzed models in econometrics.

Suppose that n repeated measurements, $y = (y_1, \ldots, y_n)'$, made on an experimental unit

5-2

can be adequately described by the following model:

$$y = f(\theta, t) + e \qquad (5.2.1)$$

where $f(\theta, t)$ is an n-dimensional linear or non-linear function in the parameters θ and $t = (t_1, \ldots, t_n)'$ is the vector of time points on which the responses were measured. The elements of e are assumed to be generated by a stationary and invertible ARMA(p,q) process with white noise variance σ^2. The vector e is assumed multivariate normal with zero mean and structured covariance matrix $\Sigma = \sigma^2 \Lambda$ which is defined by (3.5.8).

Since θ is regarded as fixed, it follows from (3.5.7) that y has a multivariate normal distribution with mean vector $f(\theta, t)$ and covariance matrix Σ.

If, for example, y has a linear trend over time with intercept θ_0 and slope θ_1, the model can be written as

$$y = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} + e. \qquad (5.2.2)$$

Growth can often be described by a non-linear function, such as a simple modified exponential, logistic or Gompertz curve. (See Herbst (1994) for a detailed discussion on the Richards family of growth curves.) If, for example, y follows a simple modified exponential trend over time, the j-th element of y can be written as

$$y_j = \theta_1(1 - \theta_2 \theta_3^{t_j}) \text{ where } \theta_1 > 0 \text{ and } 0 \leq \theta_3 \leq 1. \qquad (5.2.3)$$

5-3

Suppose n repeated measurements were made on each of a random sample of N experimental units. Let the n-dimensional vectors $y_1, \ldots, y_N$ denote the measurements on the different experimental units. The likelihood of $y_1, \ldots, y_N$ is

$$
\begin{aligned}
L &= (2\pi)^{-\frac{1}{2}nN} |\Sigma|^{-\frac{1}{2}N} exp\left[-\frac{1}{2}\sum_{i=1}^{N} [y_i - f(\theta,t)]' \Sigma^{-1}[y_i - f(\theta,t)] \right] \\
&= (2\pi\sigma^2)^{-\frac{1}{2}nN} |\Lambda|^{-\frac{1}{2}N} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} [y_i - f(\theta,t)]' \Lambda^{-1}[y_i - f(\theta,t)] \right]
\end{aligned}
\tag{5.2.4}
$$

where $\Sigma = \sigma^2\Lambda$. Note that (5.2.4) is similar to (6.1.8). The unknown parameters are the elements of θ, the ARMA coefficients and the white noise variance. The likelihood equations associated with all the parameters (with the exception of σ^2), obtained by setting the partial derivatives of lnL with respect to the parameters equal to zero, are not in a closed form and have to be solved iteratively. The maximum likelihood estimate for σ^2 is

$$
\hat{\sigma}^2 = \frac{1}{nN}\sum_{i=1}^{N} [y_i - f(\hat{\theta},t)]' \hat{\Lambda}^{-1}[y_i - f(\hat{\theta},t)]
\tag{5.2.5}
$$

and the corresponding concentrated likelihood is

$$
L' = \left[\frac{2\pi}{nN}\sum_{i=1}^{N} [y_i - f(\theta,t)]' \Lambda^{-1}[y_i - f(\theta,t)] \right]^{-\frac{1}{2}nN} |\Lambda|^{-\frac{1}{2}N} exp\left[-\frac{1}{2}nN \right] .
\tag{5.2.6}
$$

5-4

5.3 RANDOM PARAMETER REGRESSION MODELS WITH FIXED COEFFICIENT ARMA(p,q) ERRORS

Suppose a specific response function is suitable to describe repeated measurements on a number of experimental units from the same population. In many applications it is more realistic to allow the parameters of the response function to vary across experimental units. The assumption is made that the parameters of the regression model fitted to each experimental unit form a random sample from a specific population.

Random parameter growth curve models were studied by Du Toit (1979) and Herbst (1994). Other references are Beran and Hall (1992), Raj and Ullah (1981), Chow (1983) and Nicholls and Pagan (1985). In this section the maximum likelihood estimation of the parameters of linear and non-linear regression models with stochastic parameters and ARMA(p,q) errors is considered.

Let y_i be an n-dimensional vector denoting the repeated measurements on experimental unit i, $i = 1, ..., N$. Suppose y_i can be adequately described by the following model:

$$y_i = f(b_i, t) + e_i \qquad (5.3.1)$$

where $f(b_i, t)$ is an n-dimensional linear or non-linear function in the parameters, b_i. The parameter vector b_i is assumed to be an observation of an s-dimensional random vector. As in the previous section $t = (t_1, ..., t_n)'$ is the vector of time points at which the responses were measured. The elements of e_i are assumed to be generated by a stationary and invertible ARMA(p,q) process with white noise variance σ^2. Furthermore, the vectors e_i are assumed multivariate normal with zero mean and structured covariance matrix $\Sigma = \sigma^2 \Lambda$ which is defined by (3.5.8).

It is assumed that $b_1, ..., b_N$ is a random sample from a multivariate $N(\theta, \Phi)$ population and that b_i is independent of e_i. Note, that in many cases, it might be necessary to

5-5

reparameterize to make the assumption of normality of the regression parameters more realistic.

From the assumptions it follows that the conditional distribution of y_i given b is

$$y_i|b \sim N(f(b,t);\Sigma) .$$

(5.3.2)

Since the vectors $y_1|b$, $y_2|b$, ..., $y_N|b$ are independent, their joint distribution is

$$f(y_1,...,y_N|b) = \prod_{i=1}^{N} f(y_i|b)$$

$$= (2\pi)^{-\frac{1}{2}nN} |\Sigma|^{-\frac{1}{2}N} exp\left[-\frac{1}{2}\sum_{i=1}^{N} (y_i - f(b,t))' \Sigma^{-1}(y_i - f(b,t)) \right]$$

(5.3.3)

and the marginal distribution of y_1, ..., y_N is

$$f(y_1,...,y_N) = \int ... \int f(y_1,...,y_N|b)g(b)db$$

$$= (2\pi)^{-\frac{1}{2}(nN+s)} |\Sigma|^{-\frac{1}{2}N} |\Phi|^{-\frac{1}{2}} \int ... \int exp\left[-\frac{1}{2}\sum_{i=1}^{N} (y_i - f(b,t))' \Sigma^{-1}(y_i - f(b,t)) \right]$$

$$\cdot exp\left[-\frac{1}{2}(b-\theta)' \Phi^{-1}(b-\theta) \right] db .$$

(5.3.4)

The estimation of the parameters of both linear and non-linear regression models will

5-6

now be considered.

Linear regression models

Suppose the responses of the i-th experimental unit can be described by the model

$$y_i = Xb_i + e_i \qquad (5.3.5)$$

where X: $n \times s$ is a known design matrix and b_i is an s-dimensional vector of regression coefficients. Since y_i is a linear combination of independent normal variables b_i and e_i, the distribution of y_i is also multivariate normal with mean and covariance matrix given by

$$E(y_i) = X\theta \qquad (5.3.6)$$

and

$$Cov(y_i, y_i') = X\Phi X' + \Sigma . \qquad (5.3.7)$$

The vectors $y_1, \ldots, y_N$ are independent and their likelihood is

$$
L = \prod_{i=1}^{N} f(y_i)
$$

$$
= (2\pi)^{-\frac{1}{2}nN} |X\Phi X' + \Sigma|^{-\frac{1}{2}N} exp\left[-\frac{1}{2}\sum_{i=1}^{N} (y_i - X\theta)'(X\Phi X' + \Sigma)^{-1}(y_i - X\theta) \right] . \qquad (5.3.8)
$$

The determinant in (5.3.8) can be written as

5-7

$$|X\Phi X' + \Sigma| = |\Sigma||\Phi||\Phi^{-1} + X'\Sigma^{-1}X|. \tag{5.3.9}$$

For a proof of this result see Browne (1991). By using (5.3.9) the computational effort needed to calculate the determinant is reduced. This is because the order of both Φ and $X'\Sigma^{-1}X$ is s, which, in practice, is usually considerably smaller than n. An efficient procedure to calculate $|\Sigma|$ is discussed in Section 3.5 (cf. (3.5.25)).

It will now be shown that the quadratic form $X'\Sigma^{-1}X$ can be calculated without having to invert Σ. The matrix Σ can be written as LL' where L is a lower triangular matrix. It follows that $X'\Sigma^{-1}X = (L^{-1}X)'(L^{-1}X)$. If $V = L^{-1}X$, then $X'\Sigma^{-1}X = V'V$. The following proposition provides a method for calculating V efficiently.

Proposition 5.3.1

Suppose X is an n x s matrix and L is a non-singular lower triangular n x n matrix. Then $L^{-1}X$ can be calculated by backward substitution, thus eliminating the need to invert the n x n matrix L.

Proof

Let $V = L^{-1}X$. $\tag{5.3.10}$

Denote the i-th columns of X and V by x_i and v_i respectively, $i = 1, ..., s$. Then (5.3.10) can be written as

$$\begin{bmatrix} x_1 & x_2 & ... & x_s \end{bmatrix} = L\begin{bmatrix} v_1 & v_2 & ... & v_s \end{bmatrix} \tag{5.3.11}$$

so that

$$x_i = Lv_i, \quad i = 1,...,s \tag{5.3.12}$$

5-8

which can be written as

$$
\begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & & & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{ni} \end{bmatrix}
$$

(5.3.13)

The n equations are solved simultaneously to obtain the elements of v_i. It follows that

$$
v_{1i} = \frac{x_{1i}}{l_{11}}
$$

$$
v_{2i} = \frac{x_{2i} - l_{21} v_{1i}}{l_{22}}
$$

$$
\vdots
$$

(5.3.14)

$$
v_{ni} = \frac{x_{ni} - \sum_{k=1}^{n-1} l_{nk} v_{ki}}{l_{nn}}
$$

The columns of V are obtained by repeating this process for $i = 1, \ldots, s$. ■

The inverse of the covariance matrix of y_i can be written as (see Browne (1991))

$$
(X\Phi X' + \Sigma)^{-1} = \Sigma^{-1} - \Sigma^{-1} X (\Phi^{-1} + X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} .
$$

(5.3.15)

The quadratic form in (5.3.8) is therefore

5-9

$$y_i^{*'}\Sigma^{-1}y_i^* - y_i^{*'}\Sigma^{-1}X(\Phi^{-1}+X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y_i^* \qquad (5.3.16)$$

where

$$y_i^* = y_i - X\theta \; . \qquad (5.3.17)$$

Note that $y_i^{*'}\Sigma^{-1}X$ can be written as

$$\begin{aligned} y_i^{*'}\Sigma^{-1}X &= (L^{-1}y_i^*)'L^{-1}X \\ &= w'V \end{aligned} \qquad (5.3.18)$$

where $\Sigma = LL'$, with L a lower triangular matrix. The vector $w = L^{-1}y_i^*$ can be obtained by applying the procedure described in Proposition 5.3.1 and V is defined in (5.3.10). The quadratic form given in (5.3.16) can therefore be written as

$$w'w - w'V(\Phi^{-1}+V'V)^{-1}V'w \; . \qquad (5.3.19)$$

Maximum likelihood estimates for θ and Φ, the mean and covariance matrix of b, the ARMA coefficients and the white noise variance can be obtained by partial differentiation of lnL with respect to the unknown parameters. The log-likelihood function can be written as (cf. (5.3.8))

$$lnL = -\frac{1}{2}N\left[nln2\pi + ln|\Omega|\right] - \frac{1}{2}\sum_{i=1}^{N}(y_i - X\theta)'\Omega^{-1}(y_i - X\theta) \qquad (5.3.20)$$

where

5-10

$$\Omega = X\Phi X' + \Sigma. \tag{5.3.21}$$

From Section 2.2 it follows that lnL can be written as (cf. (2.2.11))

$$lnL = -\frac{1}{2}N\left[nln2\pi + ln|\Omega| + tr(\Omega^{-1}G)\right] \tag{5.3.22}$$

with

$$G = \frac{N-1}{N}S + (\bar{y} - X\theta)(\bar{y} - X\theta)' \tag{5.3.23}$$

where $\bar{y}$ and S are the sample mean vector and sample covariance matrix defined in (2.2.13) and (2.2.14).

Let γ denote the vector of unknown parameters. The partial derivative of lnL with respect to the elements of γ is (cf. (2.2.15))

$$\frac{\partial lnL}{\partial \gamma_j} = \frac{N}{2}trP\frac{\partial \Omega}{\partial \gamma_j} + NtrR\frac{\partial X\theta}{\partial \gamma_j} \tag{5.3.24}$$

where

$$P = \Omega^{-1}(G - \Omega)\Omega^{-1} \tag{5.3.25}$$

and

$$R = (\bar{y} - X\theta)'\Omega^{-1}. \tag{5.3.26}$$

5-11

It follows from (5.3.24) and (5.3.26) that the MLE of θ is

$$\hat{\theta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\overline{y} \qquad (5.3.27)$$

which is equal to the generalized least squares estimate.

The partial derivatives of Ω with respect to σ^2 and Φ_{rs} are

$$\frac{\partial\Omega}{\partial\sigma^2} = \frac{\partial}{\partial\sigma^2}(X\Phi X' + \sigma^2\Lambda)$$
$$= \Lambda \qquad (5.3.28)$$

where $\Lambda = \dfrac{1}{\sigma^2}\Sigma$ and

$$\frac{\partial\Omega}{\partial\Phi_{rs}} = X[J_{rs} + (1-\delta_{rs})J_{sr}]X' \ . \qquad (5.3.29)$$

Substitution of (5.3.28) and (5.3.29) into (5.3.24) yield

$$\frac{\partial lnL}{\partial\sigma^2} = \frac{N}{2}tr[\Omega^{-1}(G-\Omega)\Omega^{-1}\Lambda] \qquad (5.3.30)$$

and

$$\frac{\partial lnL}{\partial\Phi_{rs}} = \frac{N}{2}tr[\Omega^{-1}(G-\Omega)\Omega^{-1}X[J_{rs}+(1-\delta_{rs})J_{sr}]X']$$
$$= \frac{N}{2}(2-\delta_{rs})tr[X'\Omega^{-1}(G-\Omega)\Omega^{-1}XJ_{rs}] \qquad (5.3.31)$$

5-12

and hence

$$\frac{\partial lnL}{\partial \Phi_{rs}} = \frac{N}{2}(2-\delta_{rs})[X'\Omega^{-1}(G-\Omega)\Omega^{-1}X]_{sr} .$$

(5.3.32)

Note that it is not possible to express the MLE's of σ^2 and Φ in a closed form by setting (5.3.30) and (5.3.31) equal to zero. Furthermore, since the elements of Σ are non-linear functions of the ARMA coefficients, it is also not possible to obtain their MLE's in a closed form. All these estimates have to be determined iteratively. See Section 2.3 for a description of an optimization procedure to accomplish this.

Non-linear regression models

If $f(b,t)$ is a non-linear function of b, the integral in (5.3.4) can not be solved by applying analytical methods. Proposition 5.3.2 shows how Gauss quadrature can be used to obtain an approximation of the joint distribution of y_1, ..., y_N.

Proposition 5.3.2

An approximation of the joint probability density function (5.3.4) is

$$f(y_1,...,y_N) \approx \pi^{-\frac{1}{2}(nN+s)}(2\sigma^2)^{-\frac{1}{2}nN}|\Lambda|^{-\frac{1}{2}N}\sum_{j=1}^{m}\sum_{k=1}^{m}...\sum_{l=1}^{m}w_jw_k...w_l g(x_j,x_k,...,x_l)$$

(5.3.33)

with

$$g(x) = exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - f(x,t)\right)' \Lambda^{-1}\left(y_i - f(x,t)\right)\right] \quad (5.3.34)$$

where $w_1, \ldots, w_m$ and $x_1, \ldots, x_m$ are the weights and abscissas of an m-point Gauss-Hermite quadrature formula.

Proof

The joint density function (5.3.4) can be approximated by means of numerical integration (see Section 2.4). The integration variable is transformed in the following way:

$$v = \sqrt{\frac{1}{2}}\,\Phi^{-\frac{1}{2}}(b-\theta) \quad (5.3.35)$$

where

$$\Phi = \Phi^{\frac{1}{2}}\Phi^{\frac{1}{2}'} \quad (5.3.36)$$

so that

$$\Phi^{-1} = \Phi^{-\frac{1}{2}'} \Phi^{-\frac{1}{2}} . \quad (5.3.37)$$

The Jacobian of the transformation is

$$(2)^{\frac{1}{2}s}\,|\Phi|^{\frac{1}{2}} . \quad (5.3.38)$$

5-14

Transformation of the integral yields

$$f(y_1,\dots,y_N) = \pi^{-\frac{1}{2}(nN+s)}(2\sigma^2)^{-\frac{1}{2}nN}|\Lambda|^{-\frac{1}{2}N}\int\dots\int exp(-v'v)g(v)\,dv$$

$$= \pi^{-\frac{1}{2}(nN+s)}(2\sigma^2)^{-\frac{1}{2}nN}|\Lambda|^{-\frac{1}{2}N}\int exp(-v_1^2)\int exp(-v_2^2)\dots\int exp(-v_s^2)g(v)\,dv$$

$$(5.3.39)$$

where

$$g(v) = exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-f(v,t))'\Lambda^{-1}(y_i-f(v,t))\right].$$

$$(5.3.40)$$

Since $exp(-v_i^2)$ is the Gauss-Hermite weight function (see Table 2.4.1), (5.3.39) can be approximated by nested summations from 1 to m, where m is the number of quadrature terms, and $exp(-v_i^2)$ by w_i and v_i by x_i, $i = 1, \dots, m$, where $w_1, \dots, w_m$ and $x_1, \dots, x_m$ are the weights and abscissas of an m-point Gauss-Hermite quadrature formula. ∎

Maximum likelihood estimates for the unknown parameters, are found in the point in the parameter space where the discrepancy function

$$-2lnL \approx (nN+s)ln\pi+nNln2\sigma^2+Nln|\Lambda|-2ln\sum_{j=1}^{m}\sum_{k=1}^{m}\dots\sum_{l=1}^{m}w_jw_k\dots w_lg(x)$$

$$(5.3.41)$$

is a minimum. Minimization of (5.3.41) can be accomplished by means of an iterative optimization procedure, such as the one described in Section 2.3.

5-15

Note that it is also possible to obtain maximum likelihood estimates using the method of marginal maximum likelihood (see Section 4.4).

5.4 FIXED PARAMETER REGRESSION MODELS WITH RANDOM COEFFICIENT ARMA(p,q) ERRORS

In some applications it may be realistic to assume that the parameters of the linear or non-linear regression model are fixed, but the coefficients of the ARMA model describing the errors are random variables. Suppose the n repeated responses of experimental unit i, $i = 1, ..., N$, can be described by the following model:

$$y_i = f(\theta, t) + e_i \tag{5.4.1}$$

where the elements of e_i are the outcome of a stationary ARMA(p,q) process. The white noise terms are assumed independent normal variables with mean zero and variance σ^2. It is assumed that the coefficients of the ARMA process can vary across the different experimental units and that the coefficients of the different experimental units are a random sample from a common multivariate population. The ARMA coefficients are transformed to partial autocorrelations (cf. (3.7.4)) which are then transformed to coefficients which assume values from $-\infty$ to ∞ (cf. (4.2.2)). Let η denote the vector of transformed coefficients. The following distributional assumptions are made:

$$e_i | \eta \sim N(0, \sigma^2 \Lambda), \tag{5.4.2}$$

$$e_1 | \eta, \ e_2 | \eta, ..., e_N | \eta \ \text{are independent and} \tag{5.4.3}$$

$$\eta \sim N(\eta_0, \Psi). \tag{5.4.4}$$

5-16

These assumptions are exactly the same as those made in Section 4.2. The results of Chapter 4 can therefore be used to estimate the unknown parameters by letting

$$e_i = y_i - f(\theta, t).$$ (5.4.5)

5.5 RANDOM PARAMETER REGRESSION MODELS WITH RANDOM COEFFICIENT ARMA(p,q) ERRORS

In some situations where repeated measurements are made on a number of experimental units it may be reasonable to assume that the responses of the different experimental units can be described by the same model but that the parameters of the model may vary. In this section the case where both the regression parameters and the ARMA coefficients can vary, is considered.

Let y_i be an n-dimensional vector denoting the repeated measurements on experimental unit i, $i = 1, ..., N$. Suppose y_i can be adequately described by the following model:

$$y_i = f(b_i, t) + e_i$$ (5.5.1)

where $f(b_i, t)$ is an n-dimensional linear or non-linear function in the parameters, b_i. The parameter vector b_i is assumed to be an observation of an s-dimensional random vector. The vector $t = (t_1, ..., t_n)'$ represents the time points at which the responses were measured. The elements of e_i are assumed to be generated by a stationary and invertible ARMA(p,q) process with random coefficients and white noise variance σ^2.

The following distributional assumptions are made:

$$e_i | \eta \sim N(0, \sigma^2 \Lambda), \quad i=1,...,N$$ (5.5.2)

5-17

$$y_i | b, \eta \sim N(f(b,t), \sigma^2 \Lambda), \quad i=1,\dots,N \tag{5.5.3}$$

$$y_1 | b, \eta, \; y_2 | b, \eta, \dots, \; y_N | b, \eta \text{ are independent} \tag{5.5.4}$$

$$b \sim N(\theta, \Phi) \tag{5.5.5}$$

$$\eta \sim N(\eta_0, \Psi) \tag{5.5.6}$$

$$\mathrm{Cov}(b, \eta') = 0 \; . \tag{5.5.7}$$

From these assumptions it follows that the joint density of y_1, ..., y_N is

$$
\begin{aligned}
f(y_1,\dots,y_N) &= \int \dots \int f(y_1,\dots,y_N, b, \eta)\, db\, d\eta \\
&= \int \dots \int f(y_1,\dots,y_N | b, \eta) g(b) h(\eta)\, db\, d\eta \\
&= \int \dots \int \left[\prod_{i=1}^{N} f(y_i | b, \eta) \right] g(b) h(\eta)\, db\, d\eta \\
&= \int \dots \int (2\pi\sigma^2)^{-\frac{1}{2}nN} |\Lambda|^{-\frac{1}{2}N} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(y_i - f(b,t)\right)' \Lambda^{-1}\left(y_i - f(b,t)\right) \right] \\
&\quad \cdot (2\pi)^{-\frac{1}{2}s} |\Phi|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(b-\theta)' \Phi^{-1}(b-\theta) \right] \\
&\quad \cdot (2\pi)^{-\frac{1}{2}(p+q)} |\Psi|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\eta - \eta_0)' \Psi^{-1}(\eta - \eta_0) \right] db\, d\eta \; .
\end{aligned}
\tag{5.5.8}
$$

Note that the general case, where $f(b,t)$ is a non-linear function of b and where the

5-18

elements of Λ are non-linear functions of the elements of η, is considered. It is therefore not possible to solve (5.5.8) by applying analytical methods. The following proposition gives an approximation of (5.5.8) in terms of a Gaussian quadrature formula.

Proposition 5.5.1

An approximation of the joint density (5.5.8) is

$$f(y_1,\ldots,y_N) \approx \pi^{-\frac{1}{2}(nN+s+p+q)}(2\sigma^2)^{-\frac{1}{2}nN}\sum_{i1=1}^{m}\sum_{i2=1}^{m}\ldots\sum_{i(s+p+q)=1}^{m}w_{i1}\ldots w_{i(s+p+q)}g(x_1,x_2) \quad (5.5.9)$$

with

$$g(x_1,x_2) = |\Lambda(x_2)|^{-\frac{1}{2}N}exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-f(x_1,t))'\Lambda(x_2)^{-1}(y_i-f(x_1,t))\right] \quad (5.5.10)$$

and

$$\begin{aligned}x_1 &= (x_{i1} \ldots x_{is})' \\ x_2 &= (x_{i(s+1)} \ldots x_{i(s+p+q)})'\end{aligned} \quad (5.5.11)$$

and where $w_1, \ldots, w_m$ and $x_1, \ldots, x_m$ are the weights and abscissas of an m-point Gauss-Hermite quadrature formula.

5-19

Proof

The integral in (5.5.8) can be expressed as a product type integral by transforming the vectors $\boldsymbol{b}$ and $\boldsymbol{\eta}$ as follows:

$$\boldsymbol{v} = \sqrt{\tfrac{1}{2}}\, \boldsymbol{\Phi}^{-\frac{1}{2}}(\boldsymbol{b}-\boldsymbol{\theta}) \tag{5.5.12}$$

$$\boldsymbol{u} = \sqrt{\tfrac{1}{2}}\, \boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{\eta}-\boldsymbol{\eta}_0) \tag{5.5.13}$$

Hence (5.5.8) in terms of $\boldsymbol{v}$ and $\boldsymbol{u}$ is given by

$$f(y_1,\ldots,y_N) = \pi^{-\frac{1}{2}(nN+s+p+q)}(2\sigma^2)^{-\frac{1}{2}nN}\int \ldots \int exp(-\boldsymbol{v}'\boldsymbol{v})exp(-\boldsymbol{u}'\boldsymbol{u})g(\boldsymbol{v},\boldsymbol{u})\,d\boldsymbol{v}\,d\boldsymbol{u}$$

$$\tag{5.5.14}$$

where

$$g(\boldsymbol{v},\boldsymbol{u}) = |\Lambda(\boldsymbol{u})|^{-\frac{1}{2}N} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i-f(\boldsymbol{v},t)\right)'\Lambda(\boldsymbol{u})^{-1}\left(y_i-f(\boldsymbol{v},t)\right) \right]. \tag{5.5.15}$$

The $p+q+s$-dimensional integral in (5.5.14) can be expressed as a product type integral, and hence

$$f(y_1,\ldots,y_N) = \pi^{-\frac{1}{2}(nN+s+p+q)}(2\sigma^2)^{-\frac{1}{2}nN}\int exp(-u_1^2)\int exp(-u_2^2)\ldots\int exp(-u_{p+q}^2)$$

$$\cdot \int exp(-v_1^2)\int \ldots \int exp(-v_s^2)g(\boldsymbol{v},\boldsymbol{u})\,d\boldsymbol{v}\,d\boldsymbol{u} \ .$$

5-20

The function $exp(-v_i^2)$ is the weight function of a Gauss-Hermite quadrature formula. (See Section 2.4.) The result follows by using an m-point quadrature formula. ∎

The unknown parameters to be estimated are the elements of θ, Φ, η_0, Ψ, and σ^2. Maximum likelihood estimates can be obtained by minimizing $-2lnL$ which is given by

$$-2lnL = (nN+s+p+q)ln\pi+nNln(2\sigma^2)-2ln \sum_{i1=1}^{m} ... \sum_{i(s+p+q)}^{m} w_{i1}...w_{i(s+p+q)}g(x_1,x_2).$$ (5.5.17)

MLE's can also be obtained by using the method of marginal maximum likelihood. (See Section 4.4.)

The log-likelihood of y_1, ..., y_N is (cf. (5.5.8))

$$lnL = ln \int ... \int f(y_1,...,y_N|b,\eta)g(b)h(\eta)\,db\,d\eta$$ (5.5.18)

where $g(b)$ and $h(\eta)$ denote the marginal densities of b and η. The conditional density $f(y_1, ..., y_N|b,\eta)$ is a function of σ^2, $g(b)$ is a function of $\gamma' = (\theta', vecs(\Phi)')$ and $h(\eta)$ is a function of $\tau' = (\eta_0', vecs(\Psi)')$.

From (5.5.18) and (4.4.5) it follows that

$$\frac{\partial lnL}{\partial \gamma_i} = \int ... \int \frac{f(y_1,...,y_N|b,\eta)h(\eta)g(b)}{f(y_1,...,y_N)} \frac{\partial lng(b)}{\partial \gamma_i}\,db\,d\eta$$

$$= E_{b,\eta|y}\left[\frac{\partial lng(b)}{\partial \gamma_i}\right]$$

(5.5.19)

5-21

where

$$p(b,\eta \mid y_1,...,y_N) = \frac{f(y_1,...,y_N \mid b,\eta)h(\eta)g(b)}{f(y_1,...,y_N)} \tag{5.5.20}$$

is the conditional joint density of b and η, given y_1, ..., y_N. The notation $E_{b,\eta \mid y}(.)$ is used to denote the expected value relative to this conditional distribution.

It also follows that

$$\frac{\partial lnL}{\partial \tau_i} = E_{b,\eta \mid y}\left[\frac{\partial lnh(\eta)}{\partial \tau_i}\right] \tag{5.5.21}$$

and

$$\frac{\partial lnL}{\partial \sigma^2} = E_{b,\eta \mid y}\left[\frac{\partial lnf(y_1,...,y_N \mid b,\eta)}{\partial \sigma^2}\right]$$

$$= E_{b,\eta \mid y}\left[\sum_{i=1}^{N}\frac{\partial}{\partial \sigma^2}lnf(y_i \mid b,\eta)\right] . \tag{5.5.22}$$

From (2.2.18) it follows that

$$\frac{\partial lng(b)}{\partial \theta_j} = tr\left[(b-\theta)'\Psi^{-1}J_{j1}\right] \tag{5.5.23}$$

and hence

5-22

$$\frac{\partial lng(b)}{\partial \theta} = \mathbf{\Psi}^{-1}(b-\theta) \ .$$

(5.5.24)

The MLE of θ can be obtained by solving (cf. (5.5.19))

$$\frac{\partial lnL}{\partial \theta} = \mathrm{E}_{b,\eta|y}\left[\mathbf{\Psi}^{-1}(b-\theta)\right] = \mathbf{0}$$

(5.5.25)

so that

$$\begin{aligned}\hat{\theta} &= \mathrm{E}_{b,\eta|y}(b) \\ &= \mathrm{E}_{b|y}(b) \ .\end{aligned}$$

(5.5.26)

From (2.2.19) and (2.2.16) it follows that

$$\frac{\partial lng(b)}{\partial \phi_{rs}} = \frac{1}{2}tr\left\{\mathbf{\Phi}^{-1}\left[(b-\theta)(b-\theta)' - \mathbf{\Phi}\right]\mathbf{\Phi}^{-1}\left[J_{rs}+(1-\delta_{rs})J_{sr}\right]\right\} \ .$$

(5.5.27)

By following steps (4.4.16) to (4.4.21), the MLE of $\mathbf{\Phi}$ can be derived and is given by

$$\begin{aligned}\hat{\mathbf{\Phi}} &= \mathrm{Cov}_{b,\eta|y}(b,b') + \left[\hat{\theta}-\mathrm{E}_{b,\eta|y}(b)\right]\left[\hat{\theta}-\mathrm{E}_{b,\eta|y}(b)\right]' \\ &= \mathrm{Cov}_{b|y}(b,b') + \left[\hat{\theta}-\mathrm{E}_{b|y}(b)\right]\left[\hat{\theta}-\mathrm{E}_{b|y}(b)\right]' \ .\end{aligned}$$

(5.5.28)

Similarly, it can be shown that

$$\hat{\eta}_0 = \mathrm{E}_{\eta|y}(\eta) \ ,$$

(5.5.29)

$$\hat{\Psi} = \text{Cov}_{\eta|y}(\eta,\eta') + [\hat{\eta}_0 - \text{E}_{\eta|y}(\eta)][\hat{\eta}_0 - \text{E}_{\eta|y}(\eta)]' \qquad (5.5.30)$$

and that

$$\hat{\sigma}^2 = \frac{1}{nN}\sum_{i=1}^{N} \text{E}_{b,\eta|y}\left[(y_i - f(b,t))' \Lambda^{-1}(y_i - f(b,t))\right] . \qquad (5.5.31)$$

The EM-algorithm (see Section 4.4) can be used as an optimization algorithm. Numerical integration is used to calculate the conditional density $p(b,\eta|y_1, \ldots, y_N)$ and the conditional moments $\text{E}_{b,\eta|y}(b)$, $\text{E}_{b,\eta|y}(\eta)$, $\text{E}_{b,\eta|y}\left([y_i - f(b,t)]'\Lambda^{-1}[y_i - f(b,t)]\right)$, $\text{Cov}_{b,\eta|y}(b,b')$ and $\text{Cov}_{b,\eta|y}(\eta,\eta')$. This procedure requires values for the unknown parameters and is called the *expectation step*. The unknown parameters are estimated by substituting the values obtained in the expectation step into (5.5.26), (5.5.28), (5.5.29), (5.5.30) and (5.5.31) (*maximization step*). The process is repeated until convergence is attained.

The special case where the regression model is linear in the parameters is discussed next.

Linear regression model

Suppose the repeated measurements on the i-th experimental unit can be described by the following model:

$$y_i = Xb_i + e_i, \quad i = 1, \ldots, N \qquad (5.5.32)$$

where X is a known design matrix, $b_1, \ldots, b_N$ are a random sample from a common

5-24

population, and e_i is generated by an ARMA(p,q) process with random coefficients. From (5.5.32) and assumptions (5.5.2) to (5.5.7) it follows that the joint distribution of y_1, ..., y_N is

$$f(y_1,...,y_N) = \int ... \int f(y_1,...,y_N|b,\eta)g(b)h(\eta)\,db\,d\eta$$

$$= \int ... \int (2\pi\sigma^2)^{-\frac{1}{2}nN}|\Lambda|^{-\frac{1}{2}N}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-Xb)'\Lambda^{-1}(y_i-Xb)\right]$$

$$\cdot (2\pi)^{-\frac{1}{2}s}|\Phi|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(b-\theta)'\Phi^{-1}(b-\theta)\right]$$

$$\cdot (2\pi)^{-\frac{1}{2}(p+q)}|\Psi|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\eta-\eta_0)'\Psi^{-1}(\eta-\eta_0)\right]db\,d\eta \quad .$$

(5.5.33)

The conditional distribution of y_1, ..., y_N given η can be obtained in closed form by solving the following integral:

$$f(y_1,...,y_N|\eta) = \int ... \int f(y_1,...,y_N|b,\eta)g(b)\,db \quad .$$

(5.5.35)

Since the conditional distribution of $y_i|b,\eta$ is normal and the distribution of b is normal and the relationship between y_i and b is linear, it follows that $y_i|\eta$ is normal with mean vector and covariance matrix given by

$$E[y_i|\eta] = X\theta$$

(5.5.36)

and

5-25

$$\text{Cov}(y_i, y_i' | \eta) = X\Phi X' + \sigma^2 \Lambda$$
$$= \Omega \tag{5.5.37}$$

It also follows that $y_1 | \eta, \ldots, y_N | \eta$ are independent so that their joint distribution is given by

$$f(y_1, \ldots, y_N | \eta) = \prod_{i=1}^{N} f(y_i | \eta)$$

$$= (2\pi)^{-\frac{1}{2}nN} |\Omega|^{-\frac{1}{2}N} exp\left[-\frac{1}{2} \sum_{i=1}^{N} (y_i - X\theta)' \Omega^{-1} (y_i - X\theta) \right] . \tag{5.5.38}$$

Equation (5.5.33) simplifies to

$$f(y_1, \ldots, y_N) = \int \ldots \int (2\pi)^{-\frac{1}{2}nN} |\Omega|^{-\frac{1}{2}N} exp\left[-\frac{1}{2} \sum_{i=1}^{N} (y_i - X\theta)' \Omega^{-1} (y_i - X\theta) \right]$$

$$\cdot (2\pi)^{-\frac{1}{2}(p+q)} |\Psi|^{-\frac{1}{2}} exp\left[-\frac{1}{2} (\eta - \eta_0)' \Psi^{-1} (\eta - \eta_0) \right] d\eta . \tag{5.5.39}$$

The integral in (5.5.39) can be calculated by numerical integration. Proposition 4.3.2 can be used to find an approximation to the likelihood function. The unknown parameters may then be estimated by maximization of the likelihood by means of an iterative optimization algorithm.

5-26

5.6 PRACTICAL APPLICATION

Unit trusts are currently growing in popularity as a medium to long term investment instrument. Buying and selling prices of the different unit trusts are published in the daily press. The different funds vary in aspects such as portfolio structure (i.e. gold funds versus industrial funds versus general equity funds) and the risk involved for the investor, but their performance is affected in similar ways by economic factors such as a change in interest rates and the inflation rate. It is therefore realistic to assume that the growth of the different funds can be described by the same type of model, but that the parameters may vary across the different funds.

Thirteen comparable funds (general equity unit trusts) which have all been in existence for at least six years were selected. The month-end closing selling price for each unit trust is used for analysis purposes. The data set consists of 79 prices (from September 1988 to March 1995) for each of the 13 selected funds. Figure 5.6.1 provides a graphical representation of the data. The following is a list of the names of the different funds together with the alphabetical symbols used to represent them on the graph.

Symbol	Unit trust fund
A	Momentum Unit Trust (renamed to RMB Equities in February 1995)
B	Sage Fund
C	Sanlam Index Trust
D	The UAL Unit Trust
E	Sanlam Trust
F	Guardbank Growth Fund
G	Standard Bank Mutual Fund

5-27

H	Syfrets Growth Fund
I	Metboard Mutual Fund (Metfund)
J	NBS Equity Trust
K	Norwich Investors Fund
L	Southern Equities Fund
M	Old Mutual Investors' Fund

Let p_t represent the unit selling price at time t of a specific fund. Two different models were employed to explain the variation in p_t. The *first model* used to describe the selling price at time t is a non-linear model with stochastic parameters and ARMA(1,1) error terms, and can be written as

$$p_t = c_1 exp(c_2 t) + e_t \qquad\qquad (5.6.1)$$

where $c = (c_1, c_2)'$ is a random vector.

Let $p = (p_1, \ldots, p_n)'$ be the vector of repeated measurements on an experimental unit and e denote the corresponding vector of error terms. The following distributional assumptions are made:

$$e \sim N(0, \Sigma)$$

where Σ is an ARMA(1,1) structured matrix,

$p \mid c \sim N(f(c,t), \Sigma)$ where $f(c,t)$ is a vector representing the mean of p with j-th element $c_1 exp(c_2 t_j)$, and

5-28
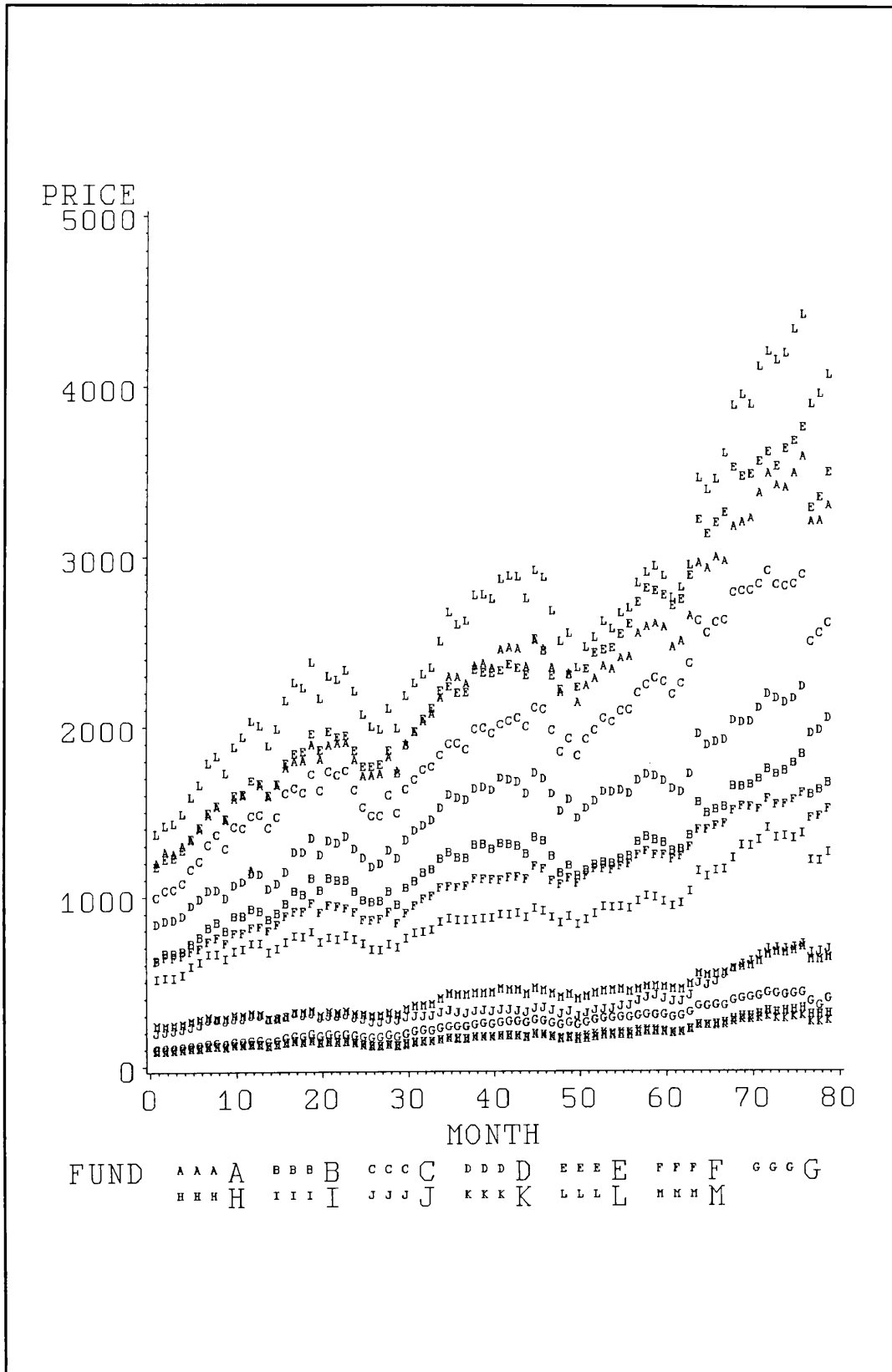
Figure 5.6.1 Selling prices of unit trust shares

5-29

$$c \sim N(\theta, \Phi).$$ (5.6.2)

Maximum likelihood estimates for the parameters were obtained by minimizing (5.3.41). The estimates together with their standard errors are given in Table 5.6.1.

Table 5.6.1 Estimation results obtained for the model (5.6.1)

Parameter	ML estimate	Standard error
θ_1	625.541	98.658
θ_2	0.01424	0.0011967
ϕ_{11}	233831.5	724.967
ϕ_{21}	-2.2229	0.184002
ϕ_{22}	$2.7153*10^{-5}$	$1.083*10^{-4}$
α	0.86631	0.00451
β	0.13618	0.03420
σ^2	13602.4	6.1160

Approximate 95% confidence intervals (parameter $\pm$ 2 standard errors) indicate that ϕ_{22} is not significant. The parameter c_2 can consequently be regarded as fixed.

The estimated mean of the selling prices at time $t = 0$ is $\hat{\theta}_1 = $ R625.54 and the estimated monthly inflation factor is $exp(\hat{\theta}_2) = 1.01434$, which is equivalent to an annual growth rate of 17.21%.

The following measure of fit can be used to compare the two models:

5-30

$$SSQ = \sum_{i=1}^{N}\sum_{j=1}^{n} [p_{ij} - E(\hat{p}_j)]^2 \tag{5.6.3}$$

where p_{ij} is the unit price associated with the i-th fund at time j and $E(\hat{p}_j)$ is the expected value of the fitted model at time j. An expression can be obtained for $E(\hat{p})$ using conditional distribution results.

$$E(\hat{p}) = E_{\hat{c}}[E(p|c)]$$
$$\tag{5.6.4}$$
$$= E_{\hat{c}}[f(c,t)]$$

where $E_{\hat{c}}$ is used to denote the expected value with respect to the estimated distribution of c, which is $N(\hat{\theta},\hat{\Phi})$. Results pertaining to the moment generating function of a normal distribution can be used to find an exact expression to calculate (5.6.4). The moment generating function of the estimated distribution of c is

$$E[exp(c_1 t_1 + c_2 t_2)] = exp(\hat{\theta}_1 t_1 + \hat{\theta}_2 t_2 + \tfrac{1}{2}\hat{\phi}_{11} t_1^2 + \hat{\phi}_{12} t_1 t_2 + \tfrac{1}{2}\hat{\phi}_{22} t_2^2) \tag{5.6.5}$$

The partial derivative of the left hand side of (5.6.5) with respect to t_1 is

$$\frac{\partial}{\partial t_1} E[exp(c_1 t_1 + c_2 t_2)] = E[c_1 exp(c_1 t_1 + c_2 t_2)] \tag{5.6.6}$$

which is equal to

$$E[c_1 exp(c_2 t_2)] \tag{5.6.7}$$

5-31

when setting $t_1 = 0$. The partial derivative of the right hand side of (5.6.5) with respect to t_1 is

$$exp(\hat{\theta}_1 t_1 + \hat{\theta}_2 t_2 + \frac{1}{2}\hat{\phi}_{11} t_1^2 + \hat{\phi}_{12} t_1 t_2 + \frac{1}{2}\hat{\phi}_{22} t_2^2)(\hat{\theta}_1 + \hat{\phi}_{11} t_1 + \hat{\phi}_{12} t_2) \qquad (5.6.8)$$

which is equal to

$$exp(\hat{\theta}_2 t_2 + \frac{1}{2}\hat{\phi}_{22} t_2^2)(\hat{\theta}_1 + \hat{\phi}_{12} t_2) \qquad (5.6.9)$$

when setting $t_1 = 0$. It follows from (5.6.7) and (5.6.9) that

$$E(\hat{p}_j) = E[c_1 exp(c_2 t_j)] = exp(\hat{\theta}_2 t_j + \frac{1}{2}\hat{\phi}_{22} t_j^2)(\hat{\theta}_1 + \hat{\phi}_{12} t_j) \qquad (5.6.10)$$

Expression (5.6.3) was calculated for model 1 and the result is $SSQ = 8.61*10^8$. Figure 5.6.2 provides a graphical representation of the expected values of the unit prices (cf.(5.6.10)) based on the estimation results given in Table 5.6.1 and the average of the 13 unit prices at each time point. It can be seen that the expected values of the estimated model do not follow the cyclical pattern in the data and the model therefore needs to be refined. Furthermore, the difference between the sample mean values (indicated by the symbol 'S' on the graph) and the expected values (indicated by the symbol 'E' on the graph) are consistently positive. It should, however, be kept in mind that the expected values are calculated with respect to the estimated distribution of c, whereas the average series is calculated by assigning equal weights to the 13 values at each time point. The model was not intended to provide a good fit of the average series.

Empirical Bayes estimates can be obtained for the parameters of the different funds by regarding the estimated distribution of c as the prior distribution of c. Let p_j denote the unit prices associated with the j-th fund, $j = 1, ..., 13$. Empirical Bayes estimates of

5-32
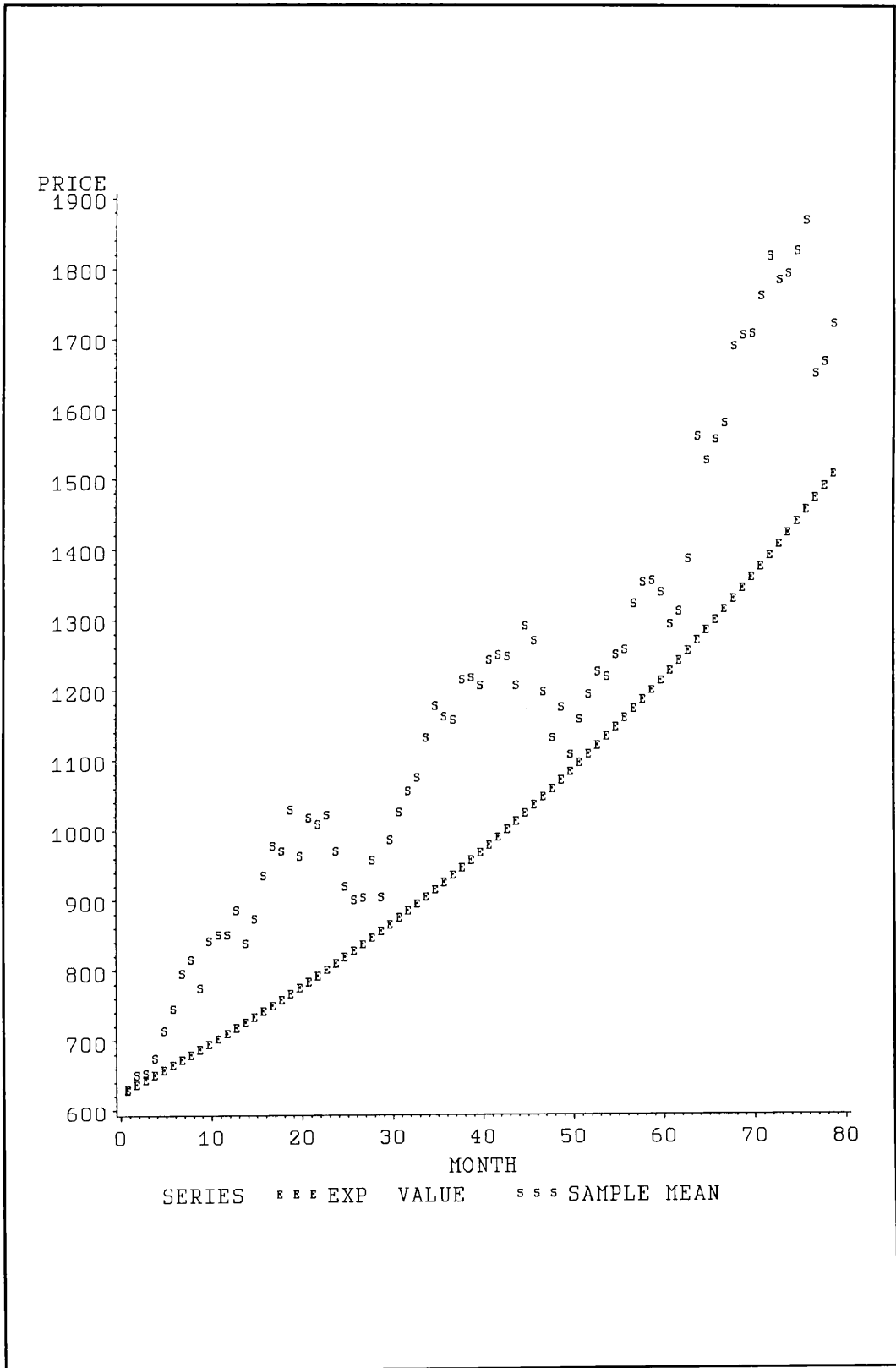
Figure 5.6.2 Sample means and expected values of unit prices

5-33

the parameters of the j-th fund are the elements of the posterior mean of c:

$$E(c_i|p_j) = \frac{\int \int c_i f(p_j|c)\hat{g}(c)dc}{\int \int f(p_j|c)\hat{g}(c)dc} \quad , \quad i = 1, 2 \quad\quad (5.6.11)$$

Table 5.6.2 give the empirical Bayes estimates of c_1 and c_2 for the 13 unit trust funds considered.

Table 5.6.2: Empirical Bayes estimates of c_1 and c_2.

Unit Trust Fund	$\hat{c}_1$	$\hat{c}_2$
A	715.56	0.01729
B	410.11	0.01513
C	642.94	0.01600
D	519.84	0.01490
E	710.47	0.01848
F	365.18	0.01562
G	74.79	0.02010
H	55.57	0.01984
I	297.07	0.01593
J	106.91	0.02072
K	55.28	0.01917
L	714.57	0.02208
M	130.54	0.01814

5-34

The *second model* considered to describe p_t is given by:

$$p_t = c_1 c_2^t a_t \tag{5.6.12}$$

where a_t is a multiplicative error term which is positive with a mean of one and both c_1 and c_2 are positive. A linear model is obtained by taking the logarithm on both sides of (5.6.12):

$$ln(p_t) = ln(c_1) + t\,ln(c_2) + ln(a_t) \tag{5.6.13}$$

Let $y_t = ln(p_t)$, $b_1 = ln(c_1)$, $b_2 = ln(c_2)$ and $e_t = ln(a_t)$. Equation (5.6.13) can be written as

$$y = Xb + e \tag{5.6.14}$$

where $y = (y_1, \ldots, y_n)'$, $b = (b_1, b_2)'$, $e = (e_1, \ldots, e_n)'$ and X is a design matrix given by

$$X = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}. \tag{5.6.15}$$

It is assumed that the error terms are generated by an ARMA(1,1) process with Gaussian white noise terms and fixed coefficients. The parameter vector b is assumed to be a bivariate normal random vector. The assumptions are:

5-35

$e \sim N(0,\Sigma)$

where Σ is an ARMA(1,1) structured matrix,

$y \mid b \sim N(Xb,\Sigma)$ and

$$b \sim N(\theta,\Phi). \tag{5.6.16}$$

Maximum likelihood estimates were obtained by minimizing $-lnL$ where lnL is given by (5.3.20). The elements of Φ, the covariance matrix of the regression parameters were reparameterized as described in Section 4.3 (cf. (4.3.16) to (4.3.19)). The transformed parameters are indicated by the notation ϕ_{11}^{*}, ϕ_{21}^{*} and ϕ_{22}^{*}. The parameter estimates and their standard errors are:

Table 5.6.3 Estimation results obtained for the parameters of model (5.6.12)

Parameter	ML estimate	Standard error
θ_1	6.15169	0.2547
θ_2	0.01254	0.0005
ϕ_{11}^{*}	-0.0605	0.3818
ϕ_{21}^{*}	-0.00076	0.00044
ϕ_{22}^{*}	-23.0424	35.3907
α	0.92043	0.0170
β	0.06525	0.0326
σ^2	0.00199	0.0000

5-36

Estimates for the elements of Φ were obtained by applying the inverse transformation discussed in Section 4.3 and are given by

$$\hat{\Phi} = \begin{bmatrix} 0.9008 & \\ -0.0065 & 0.000005 \end{bmatrix}.$$

Approximate 95% confidence intervals (parameter $\pm$ 2 standard errors) reveal that the elements of $\hat{\Phi}^*$ and the moving average parameter are not significantly different from zero.

Note that the estimated variance of the intercept, θ_1, is higher than the estimated variance of the slope, θ_2. This is expected since it can be seen on the graph that there is a greater variation in the size of the unit prices of the different funds than in their growth rates.

Since the variation of b_1 and b_2 are not significant, they are approximately equal to their means, and consequently

$$\hat{c}_1 \approx exp(\hat{\theta}_1) = 449.52 \quad \text{and}$$

$$\hat{c}_2 \approx exp(\hat{\theta}_2) = 1.01257 .$$

It follows that the mean unit price at the beginning of the series (September 1988) was R449.52 and the mean of the monthly growth rates is 1.257% (or 15.082% p.a.).

A measure of fit for model 2 (cf. (5.6.3) can be calculated by noting that

5-37

$$E(\hat{p}) = E_b[E(p \mid b)]$$

$$\text{(5.6.17)}$$

$$= E_b[exp(\boldsymbol{X}\boldsymbol{b})]$$

where E_b denote the expected value relative to the estimated distribution of $\boldsymbol{b}$, which is the $N(\hat{\theta}, \hat{\Phi})$ distribution.

An expression for the expected value of $\hat{p}_j$ can be obtained by using the moment generating function of a multivariate normal distribution:

$$E(\hat{p}_j) = E_b[exp(b_1 + b_2 t_j)]$$

$$= M_b(1, t_j) \qquad \text{(5.6.18)}$$

$$= exp(\hat{\theta}_1 + \hat{\theta}_2 t_j + \tfrac{1}{2}\hat{\phi}_{11} + \hat{\phi}_{12} t_j + \tfrac{1}{2}\hat{\phi}_{22} t_j^2) \ .$$

The value obtained for SSQ (cf. (5.6.3)) for model 2 is $SSQ = 9.20 * 10^8$, which is greater than the corresponding value for model 1.

5.7 SUMMARY

In this chapter various models used to analyze repeated measurement data are considered. It is assumed that the error terms of these models are generated by an ARMA process with fixed or random coefficients. The results of previous chapters are incorporated in more general models allowing for linear and non-linear response functions over time with fixed or random parameters. The results are applied to a South African unit trust data set.

5-38

CHAPTER 6 FIXED COEFFICIENT ARMA MODELS - NON-NORMAL ERRORS

6.1 INTRODUCTION

In Chapter 3 the estimation of the parameters of fixed coefficient ARMA models was considered. Results were given under the usual assumption of independent Gaussian white noise terms. This assumption is also applied throughout Chapter 4 and Chapter 5, as well as the assumption of random ARMA coefficients. Chapter 6 is concerned with white noise terms from the family of elliptical distributions, of which the normal distribution is a special case. It will be assumed that the ARMA coefficients are fixed, but unknown, and must therefore be estimated.

Bagchi and Guttman (1988) studied Bayesian regression analysis under non-normal errors and used a multivariate *t*-distribution to model the white noise terms of a first order autoregressive process. They state that departures from standard assumptions often occur in practice and that it is important to deal with them. Other contributions in the field of non-normal error terms were made by authors like Zellner (1976) and Broemeling (1985).

According to standard references on autoregressive moving average models (e.g. Box *et al* (1994) and Harvey (1981)) the white noise terms of an ARMA process are assumed to be *uncorrelated*. If the white noise terms are also assumed to be normally distributed, it follows immediately that they are *also independent*. Within the elliptical class of distributions the multivariate normal distribution is the only distribution where independence between two variables is implied when they are uncorrelated. In this chapter the assumption is made that the errors are uncorrelated, but not necessarily independent.

Distributions in the elliptical class are considered as alternatives to the normal distribution. These distributions are either *leptokurtic* or *platykurtic* depending on

6-1

whether their kurtosis is greater than or smaller than the kurtosis of a normal distribution.

A detailed discussion of the elliptical class of distributions is given in Section 2.5. The characteristic function and density function are given by (2.5.3) and (2.5.4) respectively. The mean and covariance matrix are given by (2.5.6) and (2.5.7). All marginal distributions are elliptically symmetric and their kurtosis is given by (2.5.9). Theoretical results required for the generation of observations from distributions in the elliptical class are given (see Theorems 2.5.1 to 2.5.4).

Two specific types of distributions in the elliptical class, namely the Pearson Type VII and the Pearson Type II, are considered in Sections 6.2 and 6.3 in order to illustrate leptokurtic and platykurtic types respectively. Maximum likelihood estimates are derived for the coefficients of the ARMA model and the white noise variance. A simulation study is used to compare the estimation results obtained from the exact distribution and the results obtained when multivariate normality is assumed. A conclusion is made regarding the performance of maximum likelihood estimators based on the normal distribution when the kurtosis of the data differs from that of a normal distribution.

General results that are used in the following sections are now given.

Distribution of observations

Let $e = (e_1, \ldots, e_n)'$ denote the outcome of a stationary ARMA(p,q) process. The observation at time t can be written as (cf. (3.2.1))

$$e_t = \alpha_1 e_{t-1} + \ldots + \alpha_p e_{t-p} + u_t - \beta_1 u_{t-1} - \ldots - \beta_q u_{t-q} \qquad (6.1.1)$$

where $u_t, u_{t-1}, \ldots$ denote the white noise terms which are assumed to be uncorrelated

6-2

with constant variance σ^2.

In Section 3.5 it was shown that the observation vector $e = (e_1, \ldots, e_n)'$ can be expressed as (cf. (3.5.7))

$$e = T_\alpha^{-1}\left[T_\beta u + I_{n,r}\, x(0)\right] \tag{6.1.2}$$

where $u = (u_1, \ldots, u_n)'$, $r = \max(p,q)$ and the components of $x(0)$ are forecasts of e_1, $\ldots$, e_r at $t=0$ (cf. (3.5.6)). Since the elements of $x(0)$ are linear combinations of u_0, $u_{-1}, \ldots$, the white noise terms prior to the first observation, and u contains white noise terms from the first observation onwards, it follows that the vectors u and $x(0)$ are uncorrelated. Matrices T_α and T_β are defined by (3.5.4) and (3.5.5) respectively and $I_{n,r}$ is a submatrix of I_n, formed by its first r columns.

The covariance matrices of u and $x(0)$ are $\sigma^2 I_n$ and P respectively where P can be obtained from (3.4.16), using (3.5.16) and (3.5.17). Both u and $x(0)$ have zero means. It will now be shown by making use of Theorem 2.5.1 that if the joint distribution of u and $x(0)$ is elliptical, then the distribution of e is also elliptical and of the same functional form as the joint distribution of u and $x(0)$.

Let the vector y be defined as follows:

$$y = \begin{bmatrix} u \\ x(0) \end{bmatrix}. \tag{6.1.3}$$

The assumption is made that y has an $(n+r)$-dimensional elliptical distribution. It follows that all marginal distributions are elliptically symmetric and have the same functional form and kurtosis (see Johnson (1987), p.109). The mean of y is zero and its covariance matrix is

6-3

$$\text{Cov}(y, y') = \begin{bmatrix} \sigma^2 I & 0 \\ 0 & P \end{bmatrix}.$$

(6.1.4)

The observation vector e can be written as a linear combination of the elements of y (cf. (6.1.2)):

$$e = By \text{ where } B = T_\alpha^{-1}[T_\beta \ I_{n,r}].$$

(6.1.5)

The rank of $B_{n \times (n+r)}$ is n.

It follows from Theorem 2.5.1 that e is elliptically contoured with zero mean and covariance matrix

$$\begin{aligned} \Sigma &= B\text{Cov}(y, y')B' \\ &= T_\alpha^{-1}[\sigma^2 T_\beta T_\beta' + I_{n,r} PI_{n,r}']T_\alpha^{-1'} \end{aligned}$$

(6.1.6)

which corresponds to (3.5.8). From Theorem 2.5.1 it also follows that the distributions of y and e have the same functional form. The vector e, for example, is distributed multivariate t, if y has a multivariate t-distribution. In Section 6.2 and 6.3 the vector y and hence the white noise terms u and observation vectors e, are assumed multivariate t and Pearson Type II respectively.

Generation of observations

The Cambanis approach (Cambanis, Huang and Simons (1981)) can be used to generate observations from both multivariate distributions under discussion. According to Theorem 2.5.4 the vector e with mean zero and covariance matrix Σ can be written as

6-4

$$e = rBu^{(n)} \tag{6.1.7}$$

where r is a positive random variable having the distribution of $(\alpha e' \Sigma^{-1} e)^{\frac{1}{2}}$;

$BB' = \alpha^{-1}\Sigma$, where α is defined by (2.5.8) and

$u^{(n)}$ is uniformly distributed on the unit hypersphere. (See (2.5.19) for a useful result on sampling from the unit hypersphere.)

Observations of r can be obtained for both the Pearson Type VII and Type II distributions by a suitable transformation of a beta variable.

To ensure stationarity and invertibility of the process during the estimation process, the parameters are transformed to partial autocorrelations when either the autoregressive or moving average order exceeds one (as discussed in Section 3.7). Let γ denote the parameter vector. The components of γ are the ARMA coefficients (or transformations where necessary), the white noise variance and other parameters of the white noise distribution.

Simulation study

Simulation studies are used in Sections 6.2 and 6.3 to compare estimation results based on the exact likelihood function and the likelihood function which is used under the assumption of independent $N(0,\sigma^2)$ white noise terms. Let e_i denote an n x 1 vector with typical element e_{ij} which represents the j-th response on individual i, $i = 1, \ldots, N, j = 1, \ldots, n$.

Under the assumption of independent $N(0,\sigma^2)$ white noise terms, the likelihood function of $e_1, \ldots, e_N$ is

$$L_N = (2\pi\sigma^2)^{-\frac{1}{2}nN} |\Lambda|^{-\frac{N}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} e_i' \Lambda^{-1} e_i \right] \tag{6.1.8}$$

6-5

where $\sigma^2\Lambda = \Sigma$, hence

$$-2lnL_N = nNln(2\pi\sigma^2)+Nln|\Lambda|+\frac{1}{\sigma^2}\sum_{i=1}^{N}(e_i'\Lambda^{-1}e_i)$$

$$\propto N\left(nln\sigma^2+ln|\Lambda|\right)+\frac{1}{\sigma^2}\sum_{i=1}^{N}e_i'\Lambda^{-1}e_i \; .$$

(6.1.9)

The MLE of σ^2 based on (6.1.8) is (cf. (3.2.12))

$$\hat\sigma^2 = \frac{1}{nN}\sum_{i=1}^{N}e_i'\Lambda^{-1}e_i$$

(6.1.10)

and the corresponding concentrated likelihood is

$$L_N' = (2\pi)^{-\frac{1}{2}nN}\left[\frac{1}{nN}\sum_{i=1}^{N}e_i'\Lambda^{-1}e_i\right]^{-\frac{1}{2}nN}|\Lambda|^{-\frac{1}{2}N}exp\left(-\frac{1}{2}nN\right)$$

(6.1.11)

and hence

$$-2lnL_N' = nNln\frac{2\pi}{nN}+nNln(\sum_{i=1}^{N}e_i'\Lambda^{-1}e_i)+Nln|\Lambda|+nN$$

$$\propto N\left[nln(\sum_{i=1}^{N}e_i'\Lambda^{-1}e_i)+ln|\Lambda|\right] \; .$$

(6.1.12)

6-6

In Sections 6.2 and 6.3 log-likelihood functions associated with probability models in the elliptical class other than the normal distribution are given. Models are fitted to the simulated data from these distributions. When estimating the parameters of the models, $-2lnL$ is used as the discrepancy function. The parameters are also estimated under the incorrect assumption of Gaussian white noise with (6.1.9) as the discrepancy function. The values of the discrepancy functions are compared to determine which model provides the best fit to the data.

The variances of the forecast errors at different lags may also be used to assess the fit of different models. In the simulation studies the variance of the forecast errors for 2 and 3 step ahead forecasts are given.

Let $a(l) = e_{t+l} - \hat{e}_t(l)$ denote the l-step forecast error at time t with $\hat{e}_t(l)$ the forecast of e_{t+l} at time t. The variance of the l step ahead forecast error is given by (cf. Box et al (1994), Section 4.2)

$$\text{Var}[a(l)] = \text{Var}[e_{t+l} - \hat{e}_t(l)] = \left[1 + \sum_{j=1}^{l-1} \psi_j^2\right]\sigma^2 \qquad (6.1.13)$$

where ψ_j is the weight of the random noise term u_{t-j} when e_t is written as a linear combination of random noise terms (cf. (3.2.7)).

6.2 THE ARMA(p,q) MODEL WITH PEARSON TYPE VII WHITE NOISE

In this section the white noise terms of an ARMA process are assumed to be multivariate Pearson Type VII. The observations e will consequently have the same distributional form. The general multivariate t-distribution is a particular parameterization of the Pearson Type VII distribution. This distribution may be used

6-7

to model situations where the kurtosis of the observations is greater than the kurtosis value for normally distributed observations. Note that the kurtosis of all marginal distributions are equal (a property of elliptically contoured distributions).

The density function, moments and likelihood function of the multivariate t-distribution are given. Sampling from this distribution is also discussed. The results of a simulation study and a discussion conclude this section.

Maximum likelihood estimation

The general n-variate t probability density function with v degrees of freedom is given by

$$f(y) = \frac{\Gamma\left[\frac{1}{2}(v+n)\right]}{\Gamma\left(\frac{1}{2}v\right)(\pi v)^{\frac{n}{2}}} |V|^{-\frac{1}{2}} [1+v^{-1}(y-\mu)'V^{-1}(y-\mu)]^{-\frac{1}{2}(v+n)}$$

(6.2.1)

with $v>0$ and $-\infty<y_i<\infty$, $i = 1, ..., n$.

The k-th moment of y_i exists only when $v>k$. The mean, covariance matrix and kurtosis of y are

$$E(y) = \mu \text{ if } v > 1$$

(6.2.2)

$$Cov(y,y') = \frac{v}{v-2}V \text{ if } v > 2$$

(6.2.3)

and

$$\gamma_2 = \frac{6}{\nu-4} \text{ if } \nu > 4 \tag{6.2.4}$$

respectively.

Note that the kurtosis of y is positive for $\upsilon > 4$ and approaches zero (the kurtosis of a normal distribution) for large values of υ.

Let e_i represent the n repeated measurements on subject i, $i = 1, \ldots, N$. If it is assumed that e_i has an n-variate $t(\upsilon)$ distribution with zero mean and covariance matrix Σ, the likelihood function of $e_1, \ldots, e_N$ is

$$L = \frac{\Gamma\left[\frac{1}{2}(\nu+n)\right]^N}{\Gamma\left(\frac{1}{2}\nu\right)^N \left(\pi(\nu-2)\sigma^2\right)^{\frac{1}{2}nN}} |\Lambda|^{-\frac{N}{2}} \prod_{i=1}^{N} \left[1+\frac{1}{(\nu-2)\sigma^2}e_i'\Lambda^{-1}e_i\right]^{-\frac{1}{2}(\nu+n)} \tag{6.2.5}$$

where $\sigma^2\Lambda = \Sigma$ (cf. (6.1.6)) and hence

$$-2lnL = N\left(-2ln\Gamma\left(\frac{1}{2}(\nu+n)\right)+2ln\Gamma\left(\frac{1}{2}\nu\right)+nln\left(\pi(\nu-2)\sigma^2\right)+ln|\Lambda|\right)$$
$$+(\nu+n)\sum_{i=1}^{N} ln\left[1+\frac{1}{(\nu-2)\sigma^2}e_i'\Lambda^{-1}e_i\right] . \tag{6.2.6}$$

The matrix V in (6.2.1) is $V = \frac{\nu-2}{\nu}\Sigma = \frac{(\nu-2)\sigma^2}{\nu}\Lambda$. Let $\gamma = (\eta_1, \ldots, \eta_{p+q}, \sigma^2, \upsilon)'$

denote the vector of unknown parameters where η_k, $k = 1, \ldots, p+q$, are the ARMA coefficients or transformations of the ARMA coefficients to partial autocorrelations when $p > 1$ or $q > 1$. The partial autocorrelations are obtained from the Levinson-

6-9

Durbin recursion given by (3.7.4).

The MLE of σ^2 can be expressed in closed form when $N = 1$ and is given by

$$\hat{\sigma}^2 = \frac{v}{n(v-2)} e' \Lambda^{-1} e$$

(6.2.7)

Note that it is possible to obtain an expression for $\frac{\partial \ln L}{\partial v}$ by using the so called

digamma function which is defined as the ratio

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

(6.2.8)

where $\Gamma(.)$ and $\Gamma'(.)$ denote the gamma function and its derivative respectively. For $f(v)$ sufficiently small, an approximation of $\Psi(f(v))$ is $log(f(v)-\frac{1}{2})$ (Kotz *et al* (1983), Vol.2). From (6.2.5) it follows that

$$\frac{\partial \ln L}{\partial v} = N \Psi\left[\frac{1}{2}(v+n)\right] - N\Psi\left(\frac{1}{2}v\right) - \frac{nN}{2(v-2)} - \frac{1}{2}\sum_{i=1}^{N} \ln\left[1 + \frac{q_i}{(v-2)\sigma^2}\right]$$

$$+ \frac{1}{2}(v+n)\sum_{i=1}^{N} \frac{q_i}{(v-2)((v-2)\sigma^2 + q_i)}$$

(6.2.9)

where $q_i = e_i' \Lambda^{-1} e_i$.

6-10

It can be readily seen that the MLE of v can not be expressed in a closed form. The MLE's of the transformed ARMA coefficients and σ^2 are also not in a closed form and must be determined numerically. The parameters are estimated by means of an iterative optimization algorithm (described in Section 2.3) using the discrepancy function (6.2.6).

Generation of observations

A simulation study is performed to examine the properties of the MLE's and to compare estimates under the assumption that the errors are from a multivariate t-distribution (obtained from (6.2.5) or (6.2.6)) to estimates obtained by assuming normality (obtained from (6.1.8) or (6.1.9)) when the actual distribution of the white noise is multivariate-t.

Observations from an n-variate $t(v)$ distribution with zero mean and covariance matrix Σ are required for the simulation study. The standard approach is to use the transformation

$$e = \left(\frac{s}{v}\right)^{-\frac{1}{2}} z \tag{6.2.10}$$

where z is $N(0, \frac{v-2}{v}\Sigma)$ and independent of s, which is $\chi^2(v)$.

Theorem 2.5.4 (the Cambanis approach) may also be used to generate observations from a multivariate t-distribution. The vector e with zero mean and covariance matrix Σ can be written as (6.1.7) where

$$BB' = \frac{v-2}{v}\Sigma. \tag{6.2.11}$$

6-11

The distribution of r^2 is Pearson Type VI. Sampling from the Pearson Type VI distribution can be accomplished by generating an observation, say w, from the BETA($\frac{1}{2}n, \frac{1}{2}v$) distribution and by applying the following transformation on w:

$$r^2 = \frac{w}{1-w} \; .$$

<div align="right">(6.2.12)</div>

The value of r is the positive square root of r^2.

Simulation Study

A simulation study was carried out to examine the effect of an increase in the kurtosis value of the white noise distribution on maximum likelihood estimates. Estimation results based on the exact likelihood function given by (6.2.5) and the likelihood function used under the assumption of multivariate normality given by (6.1.8) are compared. The estimates of the ARMA coefficients and white noise variance are compared to the values used in generating the data, which will be referred to as the *population values*. The value of the discrepancy function $-2lnL$ is a measure of fit. In the simulation study the values of (6.2.6) and (6.1.9) are compared.

An iterative optimization procedure (see Section 2.3) is used to minimize the discrepancy functions (6.2.6) and (6.1.9). Starting values for the parameters are required by the optimization program and the choice of good starting values is crucial for the algorithm to converge to the true minimum point. A procedure to obtain initial estimates of the unknown parameters is discussed in Section 4.3.

The results of the simulation study are reported in Tables 6.2.1 through to 6.2.12. Each table represents a different combination of the number of repeated measurements, n, and the autoregressive and moving average orders, p and q. Each data set consists of

50 observation vectors, $e_1, \ldots, e_{50}$, where e_i is n-dimensional, $i = 1, \ldots, N$, and is generated from a multivariate t-distribution with zero mean and covariance matrix Σ given by (6.1.6). Thirty data sets were generated for every combination studied. Both models were fitted to each data set. The means and standard deviations of the 30 estimates of the parameters, the variances of the step 2 and 3 forecast errors and the discrepancy function value are given in the tables. A count of the number of simulation runs is also given when a particular model's discrepancy function value is smaller than that of the other model (out of a total of 30). The following index can be used as a guide to the different cases studied:

Table	ARMA-process	n	White noise distribution
6.2.1	ARMA(2,1)	100	Normal
6.2.2	ARMA(2,1)	50	Normal
6.2.3	ARMA(2,1)	20	Normal
6.2.4	ARMA(2,1)	100	t
6.2.5	ARMA(2,1)	50	t
6.2.6	ARMA(2,1)	20	t
6.2.7	MA(1)	100	t
6.2.8	MA(1)	50	t
6.2.9	MA(1)	20	t
6.2.10	AR(1)	100	t
6.2.11	AR(1)	50	t
6.2.12	AR(1)	20	t

The first three tables give estimation results for the case where the white noise is normal and the standard assumptions are valid.

6-13

Simulation study results

Table 6.2.[1] Simulation study results for an ARMA(2,1) model with $n = 100$ and normal white noise.

Parameter	Population value	Normal likelihood		t-likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	0.2021	0.0244	0.2020	0.0244
α_2	0.5	0.4795	0.0128	0.4795	0.0128
β	0.6	0.5624	0.0242	0.5624	0.0241
σ^2	2.0	1.9900	0.0358	1.9902	0.0359
V[a(2)]		2.2486	0.0433	2.2489	0.0432
V[a(3)]		2.5784	0.0594	2.5787	0.0594
$-2lnL$		17630	89.92	17629	89.96
$-2lnL$ smaller		20		10	

6-14

Table 6.2.2: Simulation study results for an ARMA(2,1) model with $n = 50$ and normal white noise.

Parameter	Population value	Normal likelihood		t-likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	0.1380	0.0378	0.1377	0.0374
α_2	0.5	0.4463	0.0186	0.4462	0.0186
β	0.6	0.5078	0.0389	0.5076	0.0386
σ^2	2.0	1.9717	0.0412	1.9718	0.0412
V[a(2)]		2.2421	0.0572	2.2425	0.0571
V[a(3)]		2.5515	0.0780	2.5517	0.0776
$-2lnL$		8791	52.09	8791	52.20
$-2lnL$ smaller		17		13	

Table 6.2.3: Simulation study results for an ARMA(2,1) model with $n = 20$ and normal white noise.

Parameter	Population value	Normal likelihood		t-likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	-0.2703	0.3085	-0.1925	0.3039
α_2	0.5	0.2083	0.1760	0.2397	0.1515
β	0.6	0.1676	0.2955	0.2525	0.3193
σ^2	2.0	1.8833	0.0669	1.8827	0.0663
V[a(2)]		2.2601	0.1096	2.2489	0.1265
V[a(3)]		2.4594	0.1313	2.5787	0.1332
$-2lnL$		3470	35.47	3470	35.01
$-2lnL$ smaller		15		15	

6-16

Table 6.2.4: Simulation study results for an ARMA(2,1) model with $n = 100$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	0.2039	0.0249	0.2076	0.0311
α_2	0.5	0.4739	0.0141	0.4749	0.0247
β	0.6	0.5605	0.0277	0.5652	0.0399
σ^2	2.0	1.8799	0.2767	2.0059	0.4867
v	4.01	4.5647	0.7598		
V[a(2)]		2.1194	0.3129	2.2631	0.5493
V[a(3)]		2.4227	0.3615	2.5823	0.6091
$-2lnL$		15766	586.7	17535	1136
$-2lnL$ smaller		30		0	

6-17

Table 6.2.5: Simulation study results for an ARMA(2,1) model with $n = 50$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	0.1204	0.0432	0.1166	0.0755
α_2	0.5	0.4369	0.0257	0.4351	0.0360
β	0.6	0.4910	0.0495	0.4818	0.0893
σ^2	2.0	1.8343	0.2796	1.9570	0.5521
v	4.01	4.4615	0.6873		
V[a(2)]		2.0883	0.3270	2.2221	0.6357
V[a(3)]		2.3724	0.3797	2.5318	0.7646
$-2lnL$		7841	338.1	8690	619.1
$-2lnL$ smaller		30		0	

6-18

Table 6.2.6: Simulation study results for an ARMA(2,1) model with $n = 20$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	-0.2154	0.1346	-0.1983	0.1960
α_2	0.5	0.2410	0.0769	0.2509	0.1098
β	0.6	0.2239	0.1337	0.2233	0.1948
σ^2	2.0	1.6739	0.2177	1.6696	0.2113
v	4.01	4.4478	0.5461		
V[a(2)]		1.9976	0.2578	1.9687	0.2450
V[a(3)]		2.1877	0.2791	2.1606	0.2741
-2lnL		3083	118.1	3342	130.5
-2lnL smaller		30		0	

6-19

Table 6.2.7: Simulation study results for an MA(1) model with $n = 100$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
β	-0.3	-0.2973	0.0181	-0.3060	0.0292
σ^2	2.0	1.8718	0.2399	2.1673	1.0555
v	4.01	4.5086	0.8038		
V[a(2)]		2.0376	0.2595	2.3811	1.2045
-2lnL		15684	495.3	17706	1654
-2lnL smaller		30		0	

Table 6.2.8: Simulation study results for an MA(1) model with $n = 50$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
β	-0.3	-0.2923	0.0205	-0.3012	0.0386
σ^2	2.0	1.7654	0.2038	1.8494	0.4012
v	4.01	4.3891	0.5609		
V[a(2)]		1.9169	0.2211	2.0202	0.4420
-2lnL		7749	262.5	8578	507.4
-2lnL smaller		30		0	

6-20

Table 6.2.9: Simulation study results for an MA(1) model with n = 20 and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
β	-0.3	-0.2526	0.0376	-0.2570	0.0537
σ^2	2.0	1.7632	0.2433	1.8159	0.4523
v	4.01	4.7403	1.0025		
V[a(2)]		1.8785	0.2631	1.9405	0.4812
$-2lnL$		3151	116.2	3409	216.5
$-2lnL$ smaller		30		0	

Table 6.2.10: Simulation study results for an AR(1) model with n = 100 and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α	0.65	0.6296	0.0128	0.6346	0.0217
σ^2	2.0	1.8929	0.2108	2.0808	0.4417
v	4.01	4.4001	0.5220		
V[a(1)]		2.6434	0.2962	2.9214	0.6281
V[a(2)]		2.9414	0.3325	3.2627	0.7113
$-2lnL$		15785	485.9	17743	1049
$-2lnL$ smaller		30		0	

6-21

Table 6.2.11: Simulation study results for an AR(1) model with $n = 50$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α	0.65	0.6137	0.0126	0.6106	0.0216
σ^2	2.0	1.8709	0.2303	1.9794	0.4487
v	4.01	4.4567	0.5697		
V[a(1)]		2.5758	0.3162	2.7138	0.5843
V[a(2)]		2.8418	0.3501	2.9879	0.6276
-2lnL		7915	277.9	8746	508.9
-2lnL smaller		30		0	

Table 6.2.12: Simulation study results for an AR(1) model with $n = 20$ and white noise distribution $t(v=4.01)$.

Parameter	Population value	t-likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α	0.65	0.5520	0.0320	0.5490	0.0559
σ^2	2.0	1.8125	0.2686	1.8768	0.4780
v	4.01	4.6070	0.7495		
V[a(1)]		2.3697	0.3788	2.4472	0.6059
V[a(2)]		2.5433	0.4248	2.6284	0.6612
-2lnL		3181	125.3	3440	227.1
-2lnL smaller		30		0	

6-22

Discussion

It is common practice to assume that the error terms of an ARMA model are independent $N(0,\sigma^2)$ variables. The normal distribution belongs to a wider class of elliptical distributions. All elliptical distributions are symmetric and their kurtosis may be either greater or smaller than that of a normal distribution. The effect of an increased kurtosis on the error term distribution is examined by assuming that the error terms are uncorrelated t-variables.

A few interesting results emerge from Tables 6.2.4 through to 6.2.12. The white noise distribution in all these cases is n-variate t with parameter $v = 4.01$. As expected, the t-likelihood function value is consistently larger (or the discrepancy function value $-2lnL$ smaller) than the normal likelihood function value. The estimates are clearly biased but improve as n, the number of repeated measurements, increases. This is in line with asymptotic results regarding maximum likelihood estimates. The mean values of the estimates are not consistently closer to the population values for any of the models, but the standard errors of the estimates obtained from the t-likelihood are consistently less than the corresponding values obtained from the normal likelihood. The estimated variance of the forecast error at 2 and 3 steps is always less for the t-model than the normal model. Confidence intervals for the forecast errors at these steps will therefore be narrower for the t-distribution than the corresponding intervals based on the normal distribution. A further conclusion is that the ARMA(2,1) model can not be fitted successfully to data sets with a small number of repeated measurements.

The first three tables give estimation results for an ARMA(2,1) model with the white noise distributed normally. Both the normal and the t-likelihood functions are maximized, and the estimation results are almost identical. This is not surprising, as the t-distribution tends to a normal distribution when $v \rightarrow \infty$. The t-distribution is more general than the normal distribution, in the sense that it allows for both positive and zero kurtosis. The normal distribution with zero mean can be seen as a special case of

6-23

the *t*-distribution.

In view of the simulation study results, it would seem reasonable to assume a *t*-distribution rather than a normal distribution for the white noise of an ARMA model when the kurtosis of the white noise is non-negative.

The next section deals with negative kurtosis of the white noise distribution.

6.3 THE ARMA(*p*,*q*) MODEL WITH PEARSON TYPE II WHITE NOISE

This section deals with the Pearson Type II distribution as probability model for the white noise terms of a stationary ARMA(*p*,*q*) process. In contrast with the distribution considered in the previous section, the kurtosis of a Pearson Type II distribution is always less than that of the normal distribution.

In this section the density function of a Pearson Type II vector variable is given together with expressions for its mean, covariance matrix and kurtosis. The likelihood function of the observations $e_i = (e_1, ..., e_n)'$, $i = 1, ..., N$ is also given. Furthermore, the generation of observations from a Pearson Type II distribution is discussed. These observations are used in a simulation study to compare estimation results based on the exact likelihood function with those based on the likelihood function under the assumption of Gaussian white noise terms.

Maximum likelihood estimation

The *n*-dimensional Pearson Type II distribution is defined by Kotz (1975) as

$$f(y) = \frac{\Gamma\left(\frac{1}{2}n+m+1\right)}{\Gamma(m+1)\pi^{\frac{n}{2}}} |V|^{-\frac{1}{2}} [1-(y-\mu)' V^{-1}(y-\mu)]^m$$

(6.3.1)

6-24

with support set $(y - \mu)'V^{-1}(y - \mu) \leq 1$ and shape parameter $m > -1$. In the one dimensional case, $f(y)$ is U-shaped when $-1 < m < 0$, uniform when $m = 0$ and approaches the shape of the normal distribution for large values of m. Figures 6.3.1 to 6.3.4 are representations of $f(y)$ for different values of m and for a two-dimensional vector y.

The mean vector, covariance matrix and kurtosis of y are

$$E(y) = \mu, \tag{6.3.2}$$

$$\text{Cov}(y,y') = \frac{1}{2m+n+2}V, \tag{6.3.3}$$

and

$$\gamma_2 = \frac{-6}{2m+n+4}. \tag{6.3.4}$$

The kurtosis is negative and depends on the shape parameter, m, and the dimension of y. For large values of m and n, the kurtosis approaches zero.

In the present application, this implies that the larger the number of repeated measurements, the closer the multivariate Pearson Type II will resemble a multivariate normal distribution. The Pearson Type II distribution is a possible alternative to the normal distribution when the kurtosis of the white noise terms is less than that of a normal distribution and if the time series is relatively short.

A typical application of this theory is, for example, the analysis of psychometric test results obtained annually from a group of high school pupils.
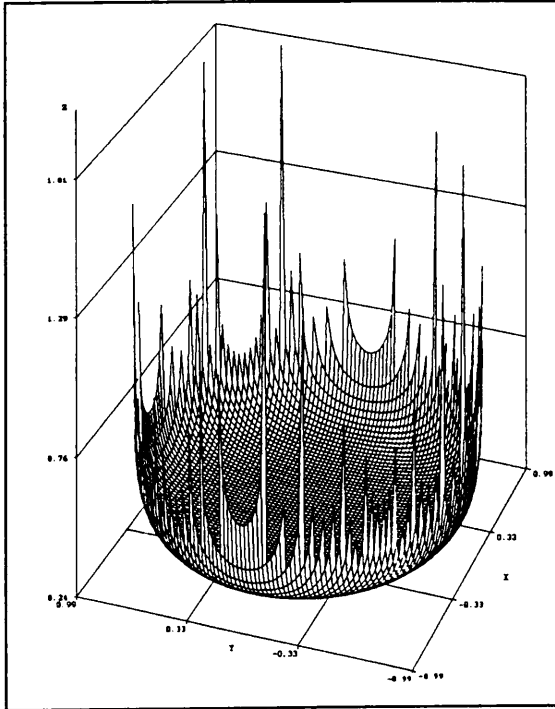
Figure 6.3.1: Pearson Type II
probability density function
with $m = -0.25$
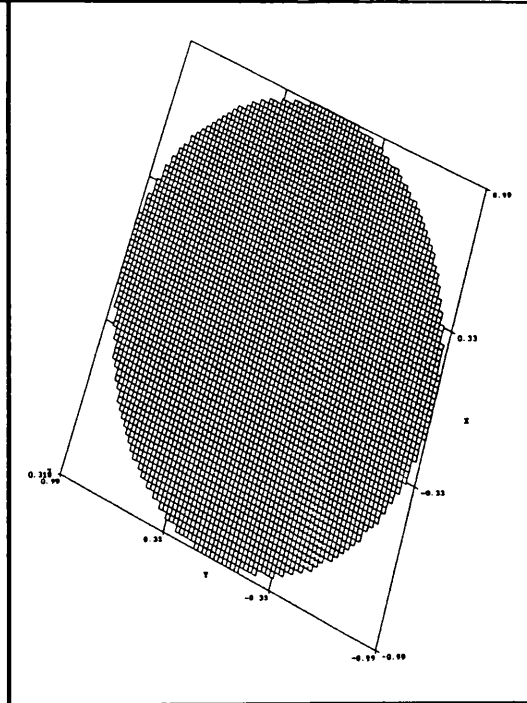
Figure 6.3.2 Pearson Type
II probability density
function with $m = 0$
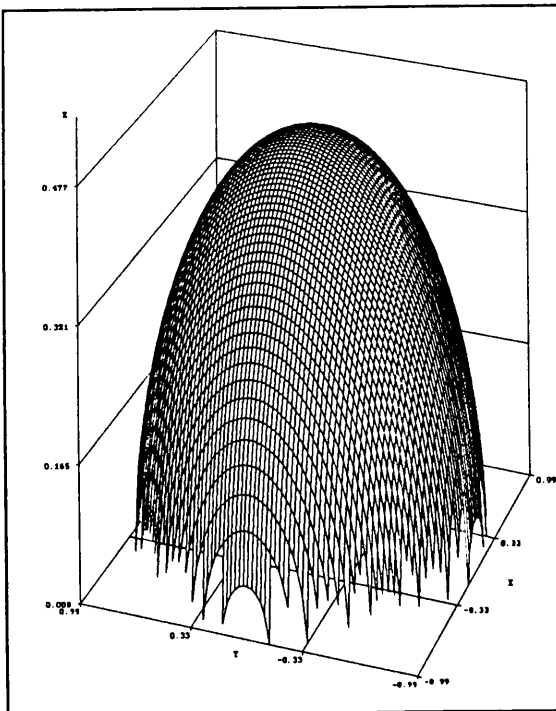
Figure 6.3.3 Pearson Type II
probability density function
with $m = 0.5$
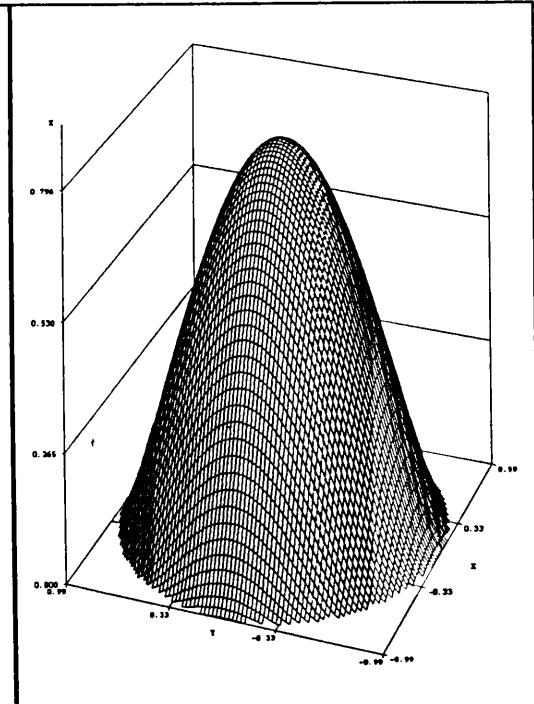
Figure 6.3.4 Pearson Type
II probability density
function with $m = 0.15$

6-26

Under the assumption that the white noise terms are uncorrelated Pearson Type II variables, it follows that the observation vector e has the same distributional form in n dimensions with zero mean and covariance matrix, Σ, given by (6.1.6). Let e_i represent n repeated measurements made on individual i, $i = 1, ..., N$. The likelihood function of $e_1, ..., e_N$ is given by

$$L = \left[\frac{\Gamma\left(\frac{1}{2}n+m+1\right)}{\Gamma(m+1)\pi^{\frac{n}{2}}} \right]^N \left[(2m+n+2)\sigma^2\right]^{-\frac{1}{2}nN} |\Lambda|^{-\frac{1}{2}N} \cdot$$

$$\prod_{i=1}^{N} \left[1 - \frac{1}{(2m+n+2)\sigma^2} e_i' \Lambda^{-1} e_i \right]^{m}$$

(6.3.5)

where $\Sigma = \sigma^2 \Lambda$. An expression for the MLE of σ^2 when $N = 1$ is

$$\hat{\sigma}^2 = \frac{2m+n}{n(2m+n+2)} e' \Lambda^{-1} e$$

(6.3.6)

As in the case of the shape parameter of a t-distribution, it is possible to obtain an expression for $\dfrac{\partial lnL}{\partial m}$, using the digamma function. The MLE of m can not, however, be expressed in a closed form. The function $-2lnL$ can be used as a discrepancy function and minimized with respect to the transformed ARMA coefficients and m. It is given by

$$-2lnL = N\left\{-2ln\Gamma(\tfrac{1}{2}n+m+1)+2ln\Gamma(m+1)+nln[\pi\sigma^2(2m+n+2)]+ln|\Lambda|\right\}$$

$$-2m\sum_{i=1}^{N}ln\left[1-\frac{1}{(2m+n+2)\sigma^2}e'_i\Lambda^{-1}e_i\right].$$

(6.3.7)

Partial derivatives of -2lnL with respect to the parameters are not expressible in a closed form and have to be determined numerically.

Generation of observations

The Cambanis approach (see Section 2.5) can be used to generate observations from the Pearson Type II distribution. The random variable r in (6.1.7) is the positive square root of r^2 where the distribution of r^2 is BETA($\tfrac{1}{2}n,m+1$). The matrix B in (6.1.7) can be obtained from

$$BB' = (2m+n+2)\Sigma.$$

(6.3.8)

Simulation study

Suppose that the white noise terms have a Pearson Type II distribution with negative kurtosis. In order to examine the properties of the MLE's of the unknown parameters based on the exact likelihood (6.3.5) and those of MLE's based on the normal distribution given by (6.1.8), use can be made of simulated data. The kurtosis of a Pearson Type II distribution (6.3.4) is a function of its shape parameter and its dimension. Positive values of the shape parameter, m, are considered. These values correspond to distributions with more mass around their means than around their extremes.

6-28

An iterative optimization procedure (see Section 2.3) is used to minimize the discrepancy functions (6.3.7) and (6.1.9). Starting values for the parameters are required by the optimization program and, as has already been noted, the selection of good starting values is crucial for the algorithm to converge to the true minimum point. A procedure to obtain initial estimates of the unknown parameters is discussed in Section 4.3.

The results of the simulation study are given in Tables 6.3.1 through to 6.3.9. Each table represents a different combination of the number of repeated measurements, n, and ARMA models. The means and standard deviations given in each table are based on 30 independent estimates obtained by minimizing the Pearson Type II discrepancy function (6.3.7) and the normal discrepancy function (6.1.9) on 30 simulated data sets. Each data set consists of a random sample of 50 experimental units.

The shape parameter, m, is fixed throughout the simulation study at a value of 2.0 and is therefore regarded as a known parameter. Estimation of the parameters of a Pearson Type II model by means of an iterative optimization procedure is more complex due to the fact that the support set of a Pearson Type II density is bounded. For each new set of parameter values obtained during the iterative procedure, the constraints

$$e_i' \Lambda^{-1} e_i \ < \ \sigma^2(2m+n+2) \ , \ i = 1, \ ..., \ N$$

must be enforced. This problem can be overcome by reparameterization. The term

$$-2m \sum_{i=1}^{N} ln \left[1 - \frac{1}{(2m+n+2)\sigma^2} e_i' \Lambda^{-1} e_i \right] \tag{6.3.9}$$

in (6.3.7) can be written as

$$-2m \sum_{i=1}^{N} \ln \left[\frac{\tau - e_i' \Lambda^{-1} e_i}{\tau} \right]$$

(6.3.10)

where $\tau = \sigma^2(2m+n+2)$ is the parameter instead of σ^2. The lower bound of τ is

$$\underset{i}{Max}(e_i' \Lambda^{-1} e_i) \ , \ i=1, \ \ldots \ N.$$

The following index can be used as a guide to the different cases studied:

Table	ARMA process	n
6.3.1	ARMA(2,1)	100
6.3.2	ARMA(2,1)	50
6.3.3	ARMA(2,1)	20
6.3.4	MA(1)	100
6.3.5	MA(1)	50
6.3.6	MA(1)	20
6.3.7	AR(1)	100
6.3.8	AR(1)	50
6.3.9	AR(1)	20

Simulation study results:

Table 6.3.1: Simulation study results for an ARMA(2,1) model with $n = 100$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	0.2238	0.5362	0.1952	0.0249
α_2	0.5	0.3195	0.3870	0.4769	0.0157
β	0.6	0.3968	0.4061	0.5542	0.0267
σ^2	2.0	3.2527	0.2299	1.9880	0.0100
V[a(2)]		5.8512	7.7517	2.2444	0.0182
V[a(3)]		12.140	29.324	2.5739	0.0373
-2lnL		19444	430.4	17624	25.26
-2lnL smaller		0		30	

Table 6.3.2: Simulation study results for an ARMA(2,1) model with $n = 50$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	0.1446	0.0927	0.1303	0.0409
α_2	0.5	0.4387	0.0438	0.4440	0.0215
β	0.6	0.5383	0.1299	0.5005	0.0440
σ^2	2.0	3.1727	0.3141	1.9694	0.0176
V[a(2)]		3.7073	0.6007	2.2399	0.0313
V[a(3)]		4.1872	0.6079	2.5490	0.0502
$-2lnL$		9567	232.5	8789	22.38
$-2lnL$ smaller		0		30	

6-32

Table 6.3.3: Simulation study results for an ARMA(2,1) model with $n = 20$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α_1	0.25	-0.2261	0.2569	-0.2201	0.2004
α_2	0.5	0.2286	0.1566	0.2375	0.1028
β	0.6	0.2304	0.2081	0.2260	0.1971
σ^2	2.0	2.7048	0.3041	1.8657	0.0489
V[a(2)]		3.2793	0.3853	2.2391	0.0722
V[a(3)]		3.6069	0.4434	2.4526	0.0893
$-2lnL$		3626	85.86	3461	26.32
$-2lnL$ smaller		0		30	

Table 6.3.4: Simulation study results for an MA(1) model with $n = 100$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
β	-0.3	-0.2943	0.0091	-0.2912	0.0096
σ^2	2.0	1.9936	0.0105	1.9791	0.0081
V[a(2)]		2.1664	0.0173	2.1471	0.0122
$-2lnL$		17510	15.03	17603	20.53
$-2lnL$ smaller		30		0	

6-33

Table 6.3.5: Simulation study results for an MA(1) model with $n = 50$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
β	-0.3	-0.2946	0.0174	-0.2895	0.0181
σ^2	2.0	1.9809	0.0206	1.9610	0.0175
V[a(2)]		2.1535	0.0306	2.1260	0.0281
-2*lnL*		8710	18.95	8778	22.32
-2*lnL* smaller		30		0	

Table 6.3.6: Simulation study results for an MA(1) model with $n = 20$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
β	-0.3	-0.2675	0.0274	-0.2573	0.0288
σ^2	2.0	1.9742	0.0593	1.9132	0.0349
V[a(2)]		2.1171	0.0729	2.0415	0.0429
-2*lnL*		3455	16.94	3487	18.11
-2*lnL* smaller		30		0	

6-34

Table 6.3.7: Simulation study results for an AR(1) model with $n = 100$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α	0.65	0.6366	0.0085	0.6319	0.0093
σ^2	2.0	2.0221	0.0271	1.9931	0.0119
V[a(2)]		2.8416	0.0450	2.7890	0.0310
V[a(3)]		3.1741	0.0596	3.1071	0.0488
-2*lnL*		17560	40.55	17638	29.79
-2*lnL* smaller		30		0	

Table 6.3.8: Simulation study results for an AR(1) model with $n = 50$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α	0.65	0.6176	0.0149	0.6128	0.0158
σ^2	2.0	2.0581	0.0571	1.9904	0.0196
V[a(2)]		2.8435	0.0832	2.7383	0.0466
V[a(3)]		3.1439	0.1033	3.0201	0.0734
-2*lnL*		8772	38.71	8815	24.65
-2*lnL* smaller		28		2	

6-35

Table 6.3.9: Simulation study results for an AR(1) model with $n = 20$ and Pearson Type II white noise.

Parameter	Population value	P.T. II likelihood		Normal likelihood	
		Mean	S.D.	Mean	S.D.
α	0.65	0.5728	0.0280	0.5558	0.0258
σ^2	2.0	2.2022	0.1424	1.9462	0.0535
$V[a(2)]$		2.9290	0.2327	2.5495	0.1082
$V[a(3)]$		3.1710	0.2804	2.7380	0.1409
$-2lnL$		3511	34.91	3503	27.70
$-2lnL$ smaller		13		17	

Conclusions

The Pearson Type II probability model is considered as an alternative to the normal probability model when the white noise of an ARMA process has a less peaked distribution than the normal distribution. Johnson (1987) states that the Pearson Types II and VII and the multivariate normal distributions include an ample number of elliptically contoured distributions for most Monte Carlo studies. Scale contaminated multivariate normal distributions are elliptically symmetric and may also be used to model situations where the white noise distribution has a negative kurtosis.

The overall conclusion arising from the simulation study (results summarized in Tables 6.3.4 to 6.3.9) is that the selection of the Pearson Type II distribution with shape parameter $m = 2$ is not an improvement to choosing the multivariate normal distribution. The standard errors of the estimates are not consistently better for any model. The fact that the discrepancy function value is generally lower for the Pearson Type II model is to be expected, since the data is simulated from a Pearson Type II

6-36

distribution.

A drawback when using the Pearson Type II model (as opposed to the t-distribution) is that it is not possible to control the degree of negative kurtosis through a parameter, because the kurtosis depends on the dimension of the vector variable. When the number of repeated measurements is increased, the kurtosis of the Pearson Type II distribution tends to zero.

6.4 SUMMARY

In this chapter two distributions in the elliptical class are considered as alternatives to the normal distribution as probability models for the white noise of an ARMA process. Simulation studies were carried out to examine the properties of the maximum likelihood estimators based on the likelihood function used under the assumption of normally distributed errors when the kurtosis of the error distribution is not normal. The exact likelihood functions based on the multivariate t-distribution and the Pearson Type II distribution are given and the corresponding maximum likelihood estimation results are reported on in the simulation study. The multivariate t-distribution was found to be an adequate model in situations where the kurtosis of the white noise is greater than that of a normal distribution. Difficulties were, however, experienced in the calculation of maximum likelihood estimates based on the multivariate Pearson Type II distribution and in the effective control of the kurtosis in order to cover the situation where the kurtosis of the white noise distribution is less than that of a normal distribution.

6-37

CHAPTER 7 VECTOR VARIATE REPEATED MEASUREMENT MODELS

7.1 INTRODUCTION

In many practical applications more than one characteristic is measured a number of times on each of several experimental units. Consider as an example a psychometric test written annually by a number of high school pupils. Such a test often consists of a number of subtests which are usually correlated. Instead of analysing the subtest scores separately by fitting a repeated measurement model to each subtest, more information can be obtained about interactions between the different subtest scores by extending the methodology for single responses to the vector case. It may for instance happen that the score in one subtest at time t depends to a greater degree on the score in a different subtest at time t-1 rather than on its own score at time t-1.

In this chapter the emphasis will be on the analysis of vector ARMA models. The identification, estimation and forecasting procedures for this type of model have been studied by e.g. Tunnicliffe-Wilson (1973), Hillmer and Tiao (1979), Tiao and Box (1981), Anderson (1978), Ansley and Kohn (1983), Spliid (1983) and Du Toit (1990b).

It will be shown in Section 7.2 how the ARMA(p,q) model for a single response can be extended to include the vector case. An expression for the covariance matrix of the observation vector is derived in Section 7.3. A multi-method multi-trait ARMA model is introduced in Section 7.4 and a practical application based on psychometric test results is given in Section 7.5.

7.2 VECTOR ARMA MODELS

Suppose that n repeated measurements of r characteristics are made on each of N experimental units. Let the vector y^* represent the n repeated measurements on an experimental unit. Then y^*:rn x 1 can be written as

$$y^* = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

(7.2.1)

where the subvector y_i:$r \times 1$ represents the i-th repeated measurement of the r characteristics, $i = 1, \ldots, n$. Suppose that the n repeated measurements can be adequately described by the following model:

$$y^* = \xi^* + e^*$$

(7.2.2)

where

$$E(y^*) = \xi^*$$

(7.2.3)

and

$$e^* = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

(7.2.4)

It is assumed that the error vectors $e_1, \ldots, e_n$ are generated by a vector ARMA(p,q) process defined by

$$e_t - \sum_{i=1}^{p} A_i e_{t-i} = u_t - \sum_{j=1}^{q} B_j e_{t-j} \quad , t = \ldots -1, 0, 1, \ldots$$

(7.2.5)

7-2

The diagonal elements of the $r \times r$ transition matrices A_i and B_j in (7.2.5) are the autoregressive and moving average coefficients of the error terms for the different characteristics. The interaction structure between the error terms of the different characteristics are reflected by the off-diagonal elements. The matrix A_i, $i = 1, ..., p$, can be written as

$$
A_i = \begin{bmatrix}
\alpha_{11,i} & \alpha_{12,i} & \cdots & \alpha_{1r,i} \\
\alpha_{21,i} & \alpha_{22,i} & \cdots & \alpha_{2r,i} \\
\vdots & & & \\
\alpha_{r1,i} & \alpha_{r2,i} & \cdots & \alpha_{rr,i}
\end{bmatrix}
\tag{7.2.6}
$$

and is not necessarily a symmetric matrix. A similar result holds for the matrix B_j in (7.2.5).

It is assumed that the white noise vectors $..., u_{-1}, u_0, u_1, ...$ are independent $N(0, \Omega)$ distributed variables. Let the $rn \times 1$ vector u^* be defined as

$$
u^* = \begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{bmatrix}.
\tag{7.2.7}
$$

The covariance matrix of u^* can be written as

7-3

$$\text{Cov}(\boldsymbol{u}^*, \boldsymbol{u}^{*'}) = \boldsymbol{I}_n \otimes \Omega$$

$$= \begin{bmatrix} \Omega & 0 & \dots & 0 \\ 0 & \Omega & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \Omega \end{bmatrix} \tag{7.2.8}$$

where Ω is an $r \times r$ covariance matrix.

For stationarity and invertibility, the roots of the characteristic equations

$$|\boldsymbol{I}_r - \boldsymbol{A}_1 B - \dots - \boldsymbol{A}_p B^p| = 0 \tag{7.2.9}$$

and

$$|\boldsymbol{I}_r - \boldsymbol{B}_1 B - \dots - \boldsymbol{B}_q B^q| = 0 \tag{7.2.10}$$

(where B is the backshift operator (cf. (3.2.4))) must fall outside the unit circle.

Example 7.2.1

Suppose that $r = 2$ characteristics measured at time points $1, \dots, n$ on an experimental unit can be described by a vector AR(1) process. Let $e_1, \dots, e_n$ denote the 2-dimensional observation vectors. The observation at time t can be written as

$$\boldsymbol{e}_t = \boldsymbol{A}\boldsymbol{e}_{t-1} + \boldsymbol{u}_t \tag{7.2.11}$$

or

$$\begin{bmatrix} e_{t1} \\ e_{t2} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} e_{t-1,1} \\ e_{t-1,2} \end{bmatrix} + \begin{bmatrix} u_{t1} \\ u_{t2} \end{bmatrix} \qquad (7.2.12)$$

$$= \begin{bmatrix} \alpha_{11} e_{t-1,1} + \alpha_{12} e_{t-1,2} + u_{t1} \\ \alpha_{21} e_{t-1,1} + \alpha_{22} e_{t-1,2} + u_{t2} \end{bmatrix} .$$

From (7.2.12) it can be readily seen that the matrix A is not necessarily symmetric. The coefficient α_{12} is a measure of the dependence of the first variable on the previous value of the second variable, while α_{21} is a measure of the dependence of the second variable on the previous value of the first variable.

From (7.2.9) it follows that the process is stationary if the roots of

$$|I_2 - AB| = 0$$

$$\qquad (7.2.13)$$

$$\Rightarrow (\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21})B^2 - (\alpha_{11} + \alpha_{22})B + 1 = 0$$

fall outside the unit circle.

It is assumed that $u_1, \ldots, u_n$ are independent $N(0, \Omega)$ vector variables. The unknown parameters to be estimated are the elements of A and the non-duplicate elements of Ω.

∎

7.3 THE COVARIANCE MATRIX OF A STATIONARY VECTOR ARMA PROCESS

In this section, expressions are derived for the vector e^* defined by (7.2.4) and its covariance matrix. It will be shown that results (3.5.7) and (3.5.8) can be extended to include the case where more than one characteristic is measured.

Let $s = \text{Max}(p,q)$. Expression (7.2.5) can be used to obtain the following set of equations:

$$e_1 \qquad\qquad\qquad = u_1 \qquad\qquad\qquad + e(1|0)$$

$$-A_1 e_1 + e_2 \qquad\qquad = -B_1 u_1 + u_2 \qquad\qquad + e(2|0)$$

$$\vdots$$

$$-A_{s-1} e_1 - A_{s-2} e_2 \ldots -A_1 e_{s-1} + e_s \quad = -B_{s-1} u_1 - B_{s-2} u_2 \ldots + u_s \qquad +e(s|0)$$

$$-A_s e_1 - A_{s-1} e_2 \ldots -A_1 e_s + e_{s+1} \quad = -B_s u_1 - B_{s-1} u_2 \ldots -B_1 u_s + u_{s+1}$$

$$\vdots$$

$$-A_s e_{n-s} - A_{s-1} e_{n-s+1} \ldots -A_1 e_{n-1} + e_n = -B_s u_{n-s} - B_{s-1} u_{n-s+1} \ldots -B_1 u_{n-1} + u_n$$

$$(7.3.1)$$

where $A_{p+1} = A_{p+2} = \ldots = A_s = 0$ if $p < s$ and $B_{q+1} = B_{q+2} = \ldots = B_s = 0$ if $q < s$, and the vectors $e(1|0)$, ..., $e(s|0)$ represent 1, ..., s step ahead forecasts of e_1, ..., e_s at time 0. The vectors $e(1|0)$, ..., $e(s|0)$ are functions of the observation and white noise vectors of the process which realised before any observations were made and are given by

7-6

$$e(1\,|\,0) \;=\; \sum_{i=1}^{p} A_i e_{1-i} \;-\; \sum_{j=1}^{q} B_j u_{1-j}$$

$$e(2\,|\,0) \;=\; \sum_{i=2}^{p} A_i e_{2-i} \;-\; \sum_{j=2}^{q} B_j u_{2-j}$$

$$\vdots$$

$$e(s\,|\,0) \;=\; \sum_{i=s}^{p} A_i e_{s-i} \;-\; \sum_{j=s}^{q} B_j u_{s-j} \;.$$
(7.3.2)

The vector $x(0)$ given by

$$x(0) : rs \times 1 \;=\; \begin{bmatrix} e(1\,|\,0) \\ e(2\,|\,0) \\ \vdots \\ e(s\,|\,0) \end{bmatrix},$$
(7.3.3)

can be considered to be a state vector and its covariance matrix can be derived from the state equation using the results of Section 3.4.

The set of equations (7.3.1) can be written in matrix form as

$$T_A e^* \;=\; T_B u^* \;+\; (I_{n,s} \otimes I_r)\,x(0)$$
(7.3.4)

where

$$T_A : nr \times nr = \begin{bmatrix} I_r & 0 & \cdots & & & & & \\ -A_1 & I_r & 0 & \cdots & & & & \\ -A_2 & -A_1 & I_r & \cdots & & & & \\ \vdots & & & & & & & \\ -A_p & -A_{p-1} & \cdots & -A_1 & I_r & \cdots & & 0 \\ 0 & -A_p & \cdots & & & & & \\ \vdots & & & & & & & \\ 0 & 0 & \cdots & -A_p & -A_{p-1} & \cdots & -A_1 & I_r \end{bmatrix}$$

(7.3.5)

and

$$T_B : nr \times nr = \begin{bmatrix} I_r & 0 & \cdots & & & & & \\ -B_1 & I_r & 0 & \cdots & & & & \\ -B_2 & -B_1 & I_r & 0 & \cdots & & & \\ \vdots & & & & & & & \\ -B_q & -B_{q-1} & -B_1 & I_r & 0 & \cdots & & \\ 0 & -B_q & -B_{q-1} & \cdots & -B_1 & I_r & 0 & \cdots \\ \vdots & & & & & & & \\ 0 & 0 & \cdots & -B_q & -B_{q-1} & \cdots & -B_1 & I_r \end{bmatrix} .$$

(7.3.6)

Since the matrix T_A is non-singular, (7.3.4) can be written as

$$e^* = T_A^{-1} [T_B u^* + (I_{n,s} \otimes I_r) x(0)] .$$

(7.3.7)

7-8

It follows that the covariance matrix of e^* is given by

$$\Sigma^* = \text{Cov}(e^*, e^{*\prime})$$

(7.3.8)

$$= T_A^{-1}[T_B \text{Cov}(u^*, u^{*\prime})T_B' + (I_{n,s} \otimes I_r)\text{Cov}[x(0), x(0)'](I_{n,s} \otimes I_r)']T_A^{-1\prime}$$

where $\text{Cov}(u^*, u^{*\prime})$ is given by (7.2.8) and

$$P = \text{Cov}[x(0), x(0)']$$

(7.3.9)

is defined by (7.3.16). Note that since $x(0)$ can be written in terms of u_0, u_{-1}, ... and u^* is a function of u_1, ..., u_n, it follows from the independence of the white noise vectors that $x(0)$ and u^* are independent.

The covariance matrix P can be derived from the state vector representation of $x(0)$. Let $x(t)$ be defined as (cf. (7.3.3))

$$x(t): rs \times 1 = \begin{bmatrix} e(t+1 \mid t) \\ e(t+2 \mid t) \\ \vdots \\ e(t+s \mid t) \end{bmatrix}.$$

(7.3.10)

By following the procedure described in Section 3.5 it can be shown that

$$x(t) = Fx(t-1) + Gu_t$$

(7.3.11)

where

$$F:rs \times rs = \begin{bmatrix} A_1 & \cdot & & & \\ A_2 & \cdot & & I_{r(s-1)} & \\ \vdots & \cdot & & & \\ \dots & \cdot & \dots & \dots & \dots & \dots \\ A_s & \cdot & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (7.3.12)$$

and

$$G:rs \times r = \begin{bmatrix} A_1 - B_1 \\ A_2 - B_2 \\ \vdots \\ A_s - B_s \end{bmatrix}. \qquad (7.3.13)$$

For a stationary process $\text{Cov}[x(t),x(t)'] = \text{Cov}[x(t\text{-}1),x(t\text{-}1)'] = P$ and from (7.3.11) it follows that

$$P = FPF' + G\Omega G'. \qquad (7.3.14)$$

A closed form expression can be derived for $vec(P)$ as follows (cf. (3.4.13) to (3.4.16)):

$$vec(P) = vec(FPF') + vec(G\Omega G')$$

$$= F \otimes F vec(P) + vec(G\Omega G')$$

$$\therefore \quad (I - F \otimes F) vec(P) = vec(G\Omega G') \qquad (7.3.15)$$

7-10

and hence

$$vec(P) = (I - F \otimes F)^{-1} vec(G \Omega G') \ .$$

(7.3.16)

Likelihood function

If it is assumed that both u^* and the initial state vector $x(0)$ have normal distributions, then e^* has a $N(0, \Sigma^*)$ distribution. The observations of the experimental units are assumed to be independent vector variables and the likelihood function of $e_1^*, \ldots, e_N^*$ is consequently given by

$$
\begin{aligned}
L &= \prod_{i=1}^{N} f(e_i^*) \\
&= \prod_{i=1}^{N} (2\pi)^{-\frac{1}{2}nr} |\Sigma^*|^{-\frac{1}{2}} exp\left[-\frac{1}{2} e_i^{*'} \Sigma^{*-1} e_i^* \right] \\
&= (2\pi)^{-\frac{1}{2}nrN} |\Sigma^*|^{-\frac{1}{2}N} exp\left[-\frac{1}{2} \sum_{i=1}^{N} e_i^{*'} \Sigma^{*-1} e_i^* \right]
\end{aligned}
$$

(7.3.17)

and hence

$$-2lnL = nrNln(2\pi) + Nln|\Sigma^*| + \sum_{i=1}^{N} e_i^{*'} \Sigma^{*-1} e_i^* \ .$$

(7.3.18)

From (7.2.2) and the distributional assumptions pertaining to e^* it follows that y^* has

7-11

a $N(\xi^*,\Sigma^*)$ distribution and the likelihood function of $y_1^*, \ldots, y_N^*$ is consequently given by

$$L = (2\pi)^{-\frac{1}{2}nrN}|\Sigma^*|^{-\frac{1}{2}N}exp\left[-\frac{1}{2}\sum_{i=1}^{N}(y_i^* - \xi_i^*)'\Sigma^{*-1}(y_i^* - \xi_i^*)\right]$$ (7.3.19)

and hence

$$-2lnL = nrNln(2\pi) + Nln|\Sigma^*| + \sum_{i=1}^{N}(y_i^* - \xi_i^*)'\Sigma^{*-1}(y_i^* - \xi_i^*) .$$ (7.3.20)

The discrepancy function (7.3.18) or (7.3.20) can be minimized to find maximum likelihood estimates of the unknown parameters. The calculation of initial estimates for the unknown parameters of a vector ARMA process may be difficult, depending on the complexity of the model. One approach that may be useful in finding initial estimates of the autoregressive matrices, is to regress the different variables on their own past values as well as on past values of the other variables. The regression coefficients can then be used as initial estimates for the elements of the matrices $A_1, \ldots, A_p$. To find initial estimates for the moving average matrices $B_1, \ldots, B_q$ is, however, more difficult. The author suggests starting with diagonal matrices, where the diagonal elements are obtained by considering the r variables separately as described in Section 4.3.

In applications where there are more variables than repeated measurements $(r > n)$, parameter identification may be impossible. In some cases it may not be realistic to assume that the correlation structure between the r variables is generated by an ARMA process. In the next section it will be shown that the covariance matrix Σ^* may be factored into the direct product of the covariance matrix of the time series process and the covariance matrix of the r variables.

7-12

7.4 MULTI-METHOD MULTI-TRAIT ARMA MODEL

Browne (1984) proposed a covariance structure which can be used to analyze the results of multi-method multi-trait experiments. The nr x nr matrix of observed scores is decomposed as follows

$$\Sigma = \Sigma_T \otimes \Sigma_M \qquad (7.4.1)$$

where Σ_T is an n x n time covariance matrix and Σ_M is an r x r method covariance matrix. Interactions between time on the one hand and variables on the other hand are then separated into two covariance matrices which may be investigated individually.

Let the nr x 1 observation vector y^* be defined as in (7.2.1) and suppose that y^* can be adequately described by the model (7.2.2). Since the elements of ξ^* are unknown fixed parameters of the model, the covariance matrix of y^* is given by Σ^* defined by expression (7.3.7). The matrix Σ^* can also be written as

$$\Sigma^* = \text{Cov}(y^*, y^{*\prime}) \qquad (7.4.2)$$

$$= \begin{bmatrix} \text{Cov}(y_1, y_1') & \text{Cov}(y_1, y_2') & \dots & \text{Cov}(y_1, y_n') \\ \text{Cov}(y_2, y_1') & & & \\ \vdots & & & \\ \text{Cov}(y_n, y_1') & & \dots & \text{Cov}(y_n, y_n') \end{bmatrix}$$

where $\text{Cov}(y_i, y_j')$ is the covariance matrix of the same set of r variables measured at times i and j. If the assumption is made that the covariance structure of the r variables stays the same (up to a scaling factor) over time and the scaling factor is determined by a time series process, then Σ^* can be written as

7-13

$$\Sigma^* = \begin{bmatrix} \sigma_{11}\Sigma_r & \sigma_{12}\Sigma_r & \cdots & \sigma_{1n}\Sigma_r \\ \sigma_{21}\Sigma_r & \sigma_{22}\Sigma_r & \cdots & \sigma_{2n}\Sigma_r \\ \vdots & & & \\ \sigma_{n1}\Sigma_r & \sigma_{n2}\Sigma_r & \cdots & \sigma_{nn}\Sigma_r \end{bmatrix}$$

$$= \Sigma_n \otimes \Sigma_r \tag{7.4.3}$$

where Σ_r is the covariance matrix of the r variables and Σ_n is the covariance matrix of the observations over time. If the assumption is made that the observations are generated by a stationary ARMA(p,q) process, an expression for Σ_n is given by (3.5.8).

Maximum likelihood estimates of the unknown parameters can be obtained by minimizing $-2lnL$ given by (7.3.18) or (7.3.20) depending on the model, where

$$|\Sigma^*| = |\Sigma_n \otimes \Sigma_r|$$

$$\tag{7.4.4}$$

$$= |\Sigma_n|^r |\Sigma_r|^n$$

and

$$\Sigma^{*-1} = \Sigma_n^{-1} \otimes \Sigma_r^{-1} . \tag{7.4.5}$$

Results (7.4.4) and (7.4.5) are given in Chapter 2 of Magnus and Neudecker (1988). These results lead to a considerable saving in computational effort. Instead of having to calculate the determinant and inverse of an nr x nr matrix Σ^*, the problem is reduced to manipulating n x n and r x r matrices.

A further advantage of the decomposition of Σ^* is a saving in the number of parameters.

7-14

The number of parameters in (7.3.8) is $r^2(p+q) + \frac{1}{2}r(r+1)$ which is made up of the r^2p

parameters of T_A, the r^2q parameters of T_B and the $\frac{1}{2}r(r+1)$ non-duplicate elements of

Ω, whereas the number of parameters in (7.4.3) is $p+q+1+\frac{1}{2}r(r+1)$ which consists of

p autoregressive coefficients, q moving average coefficients, the white noise variance

and the $\frac{1}{2}r(r+1)$ non-duplicate elements of Σ_r.

A practical application of the results obtained in this section follows.

7.5 PRACTICAL APPLICATION

As part of a selection process, the Szondi test (Szondi *et al* (1959)) was completed over
ten successive weeks by a group of postgraduate psychology students. The respondents
were shown six slides, each representing eight pathologies. Each respondent had to
indicate his preference by choosing the two persons they felt most attracted to and least
attracted to.

The number of times each pathology was selected by the respondent, was recorded.
These scores can be analyzed in order to make conclusions regarding aspects such as
introversion and extroversion, the need to receive and to give love, the tendency
towards finer and more aggressive emotions and dependence on other people.

The complete data set consists of $n = 10$ repeated measurements of $r = 16$
characteristics (preference for and aversion to the eight pathologies) on a sample of 111
respondents. It was decided to fit a linear model with ARMA error terms to the
preference scores associated with four of the pathologies.

Let y^*: 40 x 1 be given by (7.2.1) where the 4 x 1 vectors $y_1, \ldots, y_{10}$ represent the
measurements over the ten weeks for a respondent. The model describing y^* is

7-15

$$y^* = X^* \theta^* + e^* \qquad\qquad (7.5.1)$$

where

$$X^* = I_4 \otimes X, \qquad\qquad (7.5.2)$$

$$X = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots \\ 1 & t_{10} \end{bmatrix} \qquad\qquad (7.5.3)$$

and

$$\theta^* = (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \dots, \theta_{40}, \theta_{41})'. \qquad\qquad (7.5.4)$$

The parameters θ_{i0} and θ_{i1} denote the intercept and slope of the i-th variable, $i = 1$, ..., 4. It is assumed that the covariance matrix of y^* can be decomposed into the Kronecker product given in (7.4.3) where the covariance matrix of the time points, Σ_n, is a 10 x 10 ARMA(1,1) structured matrix and Σ_r is the covariance matrix of the four variables.

Expression (7.3.20) was minimized, using (7.4.4) and (7.4.5), to obtain maximum likelihood estimates of the unknown parameters, namely the elements of θ^*, the ARMA coefficients α and β, the white noise variance and a reparameterization of the non-duplicate elements of Σ_r.

The covariance matrix Σ_r can be factorized as follows:

7-16

$$\Sigma_r = D_\sigma P D_\sigma \qquad\qquad\qquad (7.5.5)$$

where P is the correlation matrix corresponding to Σ_r and D_σ is a diagonal matrix with the standard deviations of the variables as diagonal elements. Instead of optimizing with respect to the non-duplicate elements of Σ_r, the optimization was carried out with respect to the diagonal elements of D_σ and the off-diagonal non-duplicate elements of P. These parameters can be constrained to ensure a positive definite Σ_r throughout the optimization procedure. (An alternative reparameterization, which is also suitable, is given in (4.3.16).)

The following results were obtained from the computer output:

● The names used for the parameters are:

● Intercepts and slopes of the four variables:

```
A_PAR1    A_PAR2
B_PAR1    B_PAR2
C_PAR1    C_PAR2
D_PAR1    D_PAR2
```

● Diagonal elements of D_σ:

```
DGAM11    DGAM22    DGAM33    DGAM44
```

● Correlation matrix, P:

```
RHO11
RHO21    RHO22
RHO31    RHO32    RHO33
RHO41    RHO42    RHO43    RHO44
```

7-17

- Note that the diagonal elements are fixed parameters and therefore not estimated.

- White noise variance:

```
WHNOI1
```

- ARMA(1,1) coefficients and the initial state variance:

```
AR1         MA1         P11;
```

```
NUMBER OF OBSERVATIONS                      = 111
NUMBER OF VARIABLES                         =  40
NUMBER OF GROUPS                            =   1
NUMBER OF FREE PARAMETERS                   =  22
NUMBER OF FIXED PARAMETERS                  =   4
NUMBER OF EQUALITY CONSTRAINTS              =   0
NUMBER OF ACTIVE INEQUALITY CONSTRAINTS     =   0
NUMBER OF ACTIVE BOUNDS                     =   0
CONVERGENCE TOLERANCE FOR RESIDUAL COSINE   = 0.000100
```

- Parameter estimates and corresponding standard errors:

	A_PAR1	A_PAR2	B_PAR1	B_PAR2	C_PAR1	C_PAR2
ESTIM :	1.77456	-0.00443	1.71467	-0.00015	1.16070	0.02435
S.E. :	0.10579	0.01318	0.09271	0.01155	0.07702	0.00959

	D_PAR1	D_PAR2	DGAM11	DGAM22	DGAM33	DGAM44
ESTIM :	0.80683	-0.01916	0.89290	0.78244	0.65008	0.56104
S.E. :	0.06647	0.00828	0.02649	0.02312	0.00000	0.01676

	RHO11	RHO21	RHO22	RHO31	RHO32	RHO33
ESTIM :	1.00000	-0.20379	1.00000	-0.15073	-0.17507	1.00000
S.E. :	0.00000	0.02877	0.00000	0.02933	0.02910	0.00000

7-18

	RHO41	RHO42	RHO43	RHO44	WHNOI1	AR1
ESTIM :	-0.11116	-0.09761	-0.10000	1.00000	1.08239	-0.95295
S.E. :	0.02964	0.02973	0.02971	0.00000	0.04671	0.00957

	MA1	P11
ESTIM :	-0.60994	0.65290
S.E. :	0.02075	0.09798

● The slopes of the first and second variables are not significantly different from zero, indicating the absence of a linear relationship over time in the preference for the first two pathologies. The slopes of the remaining two variables are small, (0.0245 and -0.01916), indicating weak linear relationships over time.

```
┌─────────────────────────────────────┐
│  MEASURES OF FIT OF THE MODEL  │
└─────────────────────────────────────┘
```

● Discrepancy function : $\dfrac{-2lnL}{N}$ where $N = 111$:

```
SAMPLE DISCREPANCY FUNCTION VALUE            :10.097
POPULATION DISCREPANCY FUNCTION VALUE, Fo
BIAS CORRECTED POINT ESTIMATE                : 2.548
90 PERCENT CONFIDENCE INTERVAL               : ( 1.799 ; 3.370)


FIT INDEX (Modified McDonald : exp(-Fo)  )
POINT ESTIMATE                               : 0.078
90 PERCENT CONFIDENCE INTERVAL               : ( 0.034 ; 0.165)


ROOT MEAN SQUARE ERROR OF APPROXIMATION
Steiger-Lind : RMSEA = SQRT(Fo/DF)
POINT ESTIMATE                               : 0.055
90 PERCENT CONFIDENCE INTERVAL               : ( 0.046 ; 0.063)
```

7-19

```
EXPECTED CROSS-VALIDATION INDEX
POINT ESTIMATE (MODIFIED AIC)                    :10.494
90 PERCENT CONFIDENCE INTERVAL                   : ( 9.745 ;11.315)
CVI (MODIFIED AIC) FOR THE SATURATED MODEL       :15.495
```

● Discrepancy function value $-2lnL$:

```
┌─────────────────────┐
│   TEST STATISTIC    │                         :    1120.80
└─────────────────────┘
```

```
EXCEEDENCE PROBABILITIES:-
Ho: PERFECT FIT (RMSEA = 0.0)                    : 0.000
Ho: CLOSE FIT   (RMSEA .LE. 0.050)               : 0.162
MULTIPLIER FOR OBTAINING TEST STATISTIC =    111.0
DEGREES OF FREEDOM = 838
EFFECTIVE NUMBER OF PARAMETERS =   22
```

● Estimated correlation matrix of the four variables:

```
    CORRELATIONS (TESTS)

                VAR_A     VAR_B     VAR_C     VAR_D

VAR_A           1.000
VAR_B          -0.204     1.000
VAR_C          -0.151    -0.175     1.000
VAR_D          -0.111    -0.098    -0.100     1.000
```

● The correlation structure shows no strong linear relationships between the four variables. The four variables may therefore be analyzed separately without losing any information regarding possible interrelationships.

● Estimated covariance matrix $\hat{\Sigma}_n$:

COVARIANCE MATRIX (OCCASIONS)

	TIM1	TIM2	TIM3	TIM4	TIM5	TIM6
TIM1	1.735					
TIM2	0.993	1.803				
TIM3	0.947	1.058	1.864			
TIM4	0.902	1.008	1.116	1.919		
TIM5	0.860	0.960	1.063	1.169	1.970	
TIM6	0.819	0.915	1.013	1.114	1.217	2.016
TIM7	0.781	0.872	0.966	1.061	1.160	1.261
TIM8	0.744	0.831	0.920	1.011	1.105	1.201
TIM9	0.709	0.792	0.877	0.964	1.053	1.145
TI10	0.676	0.755	0.836	0.919	1.004	1.091

	TIM7	TIM8	TIM9	TI10
TIM7	2.057			
TIM8	1.300	2.095		
TIM9	1.239	1.336	2.129	
TI10	1.181	1.273	1.369	2.160

STANDARD DEVIATIONS (OCCASIONS)

	TIM1	TIM2	TIM3	TIM4	TIM5	TIM6
STD.DEV.	1.317	1.343	1.365	1.385	1.403	1.420

	TIM7	TIM8	TIM9	TI10
STD.DEV.	1.434	1.447	1.459	1.470

CORRELATIONS (OCCASIONS)

	TIM1	TIM2	TIM3	TIM4	TIM5	TIM6
TIM1	1.000					
TIM2	0.562	1.000				
TIM3	0.526	0.577	1.000			
TIM4	0.494	0.542	0.590	1.000		
TIM5	0.465	0.510	0.555	0.601	1.000	
TIM6	0.438	0.480	0.523	0.566	0.611	1.000

7-21

	TIM7	TIM8	TIM9	TI10		
TIM7	0.413	0.453	0.493	0.534	0.576	0.619
TIM8	0.390	0.428	0.466	0.504	0.544	0.585
TIM9	0.369	0.404	0.440	0.477	0.514	0.553
TI10	0.349	0.382	0.416	0.451	0.486	0.523

	TIM7	TIM8	TIM9	TI10
TIM7	1.000			
TIM8	0.626	1.000		
TIM9	0.592	0.633	1.000	
TI10	0.560	0.599	0.638	1.000

7.6 SUMMARY

In repeated measurement experiments, more than one related characteristic is often measured at each time point. Vector ARMA models can be used to analyze the change in the response vector over time.

It is shown that results applying to the scalar case can be generalized to deal with vectors of measurements. An expression is derived for the covariance matrix of the observation vector, that takes into account information regarding the process before any observations were made. The likelihood function of the observations are given under the assumption that the white noise terms are independent multivariate normal vector variates.

It is shown that there are practical situations where the covariance matrix of the observations can be decomposed into a Kronecker product of the covariance matrix of the variables and the covariance matrix of the observations over time. Conclusions can then be drawn regarding the relationships between the variables on the one hand and a possible common pattern relating to change over time on the other hand. A practical application involving psychometric test results is given.

7-22

CHAPTER 8 SUGGESTIONS FOR FURTHER RESEARCH

Various estimation techniques are considered for the parameters of random parameter regression models with error terms which are generated by a time series process. A detailed comparison of these techniques, using observed data, can be done to establish their strengths and weaknesses.

The multivariate normal distribution was used throughout this dissertation as a probability model for the parameters of the regression model, as well as for the transformed ARMA coefficients. These results can be extended to other probability models, such as other members of the elliptical class.

The multivariate t-distribution was found to be an adequate probability model when the kurtosis of the variables under consideration is greater than that of a normal distribution. A suitable distribution has to be obtained to serve as probability model in the case of negative kurtosis.

In many practical applications a series of observations are influenced by other factors, such as economic growth, resulting in a change in the ARMA coefficients over time. It might therefore be interesting to investigate models where the ARMA coefficients can be expressed as linear or non-linear functions over time.

The implementation of the Gibbs sampler in situations where the full conditional distributions are not standard distributions can be investigated further. This will include the development of conceptually simple algorithms for sampling from non-standard multivariate distributions.

The vector ARMA process with fixed transition matrices was considered. As in the scalar case, it might be reasonable to assume that these transition matrices can vary over the experimental units and should therefore be regarded as stochastic.

8-1

BIBLIOGRAPHY

ABRAMOWITZ, M and STEGUN, I.A. (1964). *Handbook of Mathematical Functions,* Applied Mathematics Series, vol. 55. Washington: National Bureau of Standards; reprinted in 1968 by Dover Publications, New York.

AGRESTI, A. (1990). *Categorical Data Analysis.* New York: John Wiley.

AKAIKE, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika, 60,* 255-265.

AKAIKE, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics, 26,* 363-387.

AKAIKE, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM J. Control, 13,* 163-173.

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

ANSLEY, C.F. and KOHN, R. (1983). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. *Biometrika, 70,* 275-278.

BAGCHI, P. and GUTTMAN, I. (1988). Bayesian Regression Analysis Under Non-Normal Errors. In *Bayesian Statistics 3* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith eds.), 543-552. Oxford University Press.

BARGMANN, R.E. (1967). Matrices and Determinants. In *Chemical Rubber Company Handbook of Tables for Mathematics,* ed. S.M. Selby, 144-166. Cleveland: Chemical Rubber Company.

BERAN, R. and HALL, P. (1992). Estimating coefficient distributions in random coefficient regressions. *The Annals of Statistics, 20, no. 4,* 1970-1984.

BICKEL, P.J. and DOKSUM, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* San Francisco: Holden-Day.

BOCK, R.D. (1990). Marginal maximum likelihood estimation for the two-stage random regressions model with second-stage covariates. *Technical Report,* Department of Psychology, University of Chicago, Chicago.

BOX, G.E.P., JENKINS, G.M. and REINSEL, G.C. (1994). *Time Series Analysis Forecasting and Control.* Third Edition. New Jersey: Prentice-Hall.

BROEMELING, L. (1985). *Bayesian Analysis of Linear Models.* New York: Marcel Dekker, Inc.

BROWNE, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8,* 1-24.

BROWNE, M.W. (1982). Covariance Structures. In: Hawkins, D.M. (Ed.). *Topics in Applied Multivariate Analysis.* Cambridge University Press.

BROWNE, M.W. (1984). The decomposition of multitrait-multimethod matrices. *The British Journal of Mathematical and Statistical Psychology, 31,* 1-21.

BROWNE, M.W. (1991). *STA 401 Matrix methods in Statistics.* Supplementary notes. University of South Africa, Pretoria.

BROWNE, M.W. and DU TOIT, S.H.C. (1992). Automated fitting of nonstandard models for mean vectors and covariance matrices. *Multivariate Behavioural Research, 27, No. 2,* 269-300.

CAMBANIS, S, HUANG, S. and SIMONS, G. (1981). On the Theory of Elliptically Contoured Distributions. *Journal of Multivariate Analysis, 11,* 368-385.

CASELLA, G. and GEORGE, E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician, 46,* 167-174.

CHAN, K.S. (1993). Asymptotic behaviour of the Gibbs sampler. *Journal of the American Statistical Association, 88, no. 421,* 320-326.

CHIB, S. (1993). Bayes regression with autoregressive errors: a Gibbs sampling approach. *Journal of Econometrics, 58, no. 3,* 275-294.

CHIB, S. and GREENBERG, E. (1993). Bayes Inference in Regression Models with ARMA(p,q) errors. *Washington University Working Paper, No. 172.* Economics Department. Washington University.

CHMIELEWSKI, M.A. (1981). Elliptically Symmetric Distributions: A Bibliography and Review. *International Statistical Review, 49,* 67-74.

CHOW, G.C. (1983). Random and changing coefficient models. In *Handbook of Econometrics.* (Z. Griliches and M.D. Intriligator, eds.) 2, 1213-1245. Amsterdam: North-Holland.

DEGROOT, M.H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

DEMPSTER. A.P., LAIRD, N.M. and RUBIN,D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39,* 1-38.

DENNIS, J.E. (1977). Non-linear least squares and equations. In: Jacobs, D. (Ed.). *The state of the Art in Numerical Analysis.* London: Academic Press.

DU TOIT, S.H.C. (1979). *The Analysis of Growth Curves.* Unpublished Ph.D. Thesis. University of South Africa.

DU TOIT, S.H.C. and BROWNE, M.W. (1982). *AUFIT (Version 1): A computer program for the automated fitting of non-standard models for means and covariances.* Research finding WS-27. Pretoria: Human Sciences Research Council.

DU TOIT, S.H.C. (1990a). Repeated Measurement ARMA processes I. *Technical Report No. 276,* Department of Statistics, Stanford University.

DU TOIT, S.H.C. (1990b). Repeated Measurement ARMA processes II. *Technical Report No. 281,* Department of Statistics, Stanford University.

DU TOIT, S.H.C. (1993). *Analysis of Multilevel Models. Part 1 Theoretical Aspects.* Pretoria: Human Sciences Research Council.

GELFAND, A.E. and SMITH, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 972-985.

GELFAND, A.E. and SMITH, A.F.M. (1991). Gibbs sampling for marginal posterior expectations. *Communications in Statistics, Theory and Methods, 20* (5 and 6), 1747-1766.

GEMAN, S. and GEMAN, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721-741.

GEYER, C.J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science, 7,* 473-483.

GILL, P.E. MURRAY, W. and WRIGHT, M.H. (1981). *Practical Optimization.* London: Academic Press.

GOMEZ, V. and MARAVALL, A. (1994). Estimation, Prediction, and Interpolation for Nonstationary Series With the Kalman Filter. *Journal of the American Statistical Association, 89,* 611-624.

GRIZZLE, J.E. and ALLEN, D.M. (1969). Analysis of growth and dose response curves. *Biometrics, 25,* 357-382.

GUPTA, A.K. and VARGA, T. (1993). *Elliptically Contoured Models in Statistics.* Dordrecht: Kluwer Academic Publishers.

HARVEY, A.C. (1981). *Time Series Models.* London: Philip Allan.

HARVILLE, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to related problems. *Journal of the American Statistical Association, 72,* 320-340.

HASTINGS, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika, 57,* 97-109.

HERBST, A. (1994). *Statistical modelling of growth.* Unpublished Ph. D. Thesis. University of Pretoria.

HILLMER, S.C. and TIAO, G.C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association, 74,* 652-661.

HIRST, K., BOYLE, D.W., ZERBE, G.O. and WILKENING, R.B. (1991). On nonlinear random effects models for repeated measurements. *Communications in*

9-5

Statistics B, Simulation and Computation, 20(2&3), 463-478.

HORVATH, M. (1991). Parameter estimation for Linear Discrete-Time Models with Random Coefficients. *IEEE Transactions on Automatic Control, 36,* 1213-1215.

JENNRICH, R.I. and SCHLUCHTER, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics, 42,* 805-820.

JOHNSON, M.E. (1987). *Multivariate Statistical Simulation.* New York: John Wiley & Sons.

JONES, K. (1991). The Application of Time Series Methods to Moderate Span Longitudinal Data. In: *Best methods for the analysis of change: recent advances, unanswered questions, future directions.* (L.M. Collins and J.L. Horn. (eds.)), 75-91. Washington DC: American Psychological Association.

JONES, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics, 22,* 389-395.

JONES, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach.* London: Chapman & Hall.

KELKER, D. (1970). Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization. *Sankyha,* 419-430.

KENDALL, M.G. and STUART, A. (1958). *The Advanced Theory of Statistics, Vol. 1,* London: Griffin.

KOTZ, S. (1975). Multivariate Distributions at a Cross Road. *Statistical Distributions in Scientific Work, 1,* 247-270.

KOTZ, S., JOHNSON, N.L. and READ, C.B. (1983). *Encyclopedia of Statistical Sciences, vol. 2.* John Wiley & Sons, Inc.

KOTZ, S., JOHNSON, N.L. and READ, C.B. (1983). *Encyclopedia of Statistical Sciences, vol. 4.* John Wiley & Sons, Inc.

KRYLOV, V.I. (1962). *Approximate calculation of integrals.* New York: Macmillan (Translated from Russian by Stroud, A.H.)

LAIRD, N.M. and WARE, J.H. (1982). Random effects models for longitudinal data. *Biometrics, 38,* 963-974.

LANGE, N. and LAIRD, N.M. (1989). The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Biometrics, 84,* 241-247.

LONGFORD, N.T. (1993). *Random Coefficient Models.* New York: Oxford University Press Inc.

MACEACHERN, S.N. and BERLINER, L.M. (1994). Subsampling the Gibbs Sampler. *The American Statistician, 48,* 188-190.

MAGNUS, J.R. and NEUDECKER, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* New York: John Wiley & Sons.

MARRIOTT, J.M. (1988). Reparameterisation for Bayesian Inference in ARMA Time Series. In: *Bayesian Statistics 3* (Bernardo, J.M., De Groot, M.H. and Smith, A.F.M. (Eds.)), *701-704.* Oxford University Press.

MARTIN, R.S., PETERS, G. and WILKINSON, J.H. (1965). Symmetric Decomposition of a Positive Definite Matrix. *Numerische Mathematik, 7,* 362-383.

9-7

METROPOLIS, N. ROSENBLUTH, A.W., TELLER, A.H. and TELLER, E. (1953).
Equations of State Calculations by fast Computing Machines. *Journal of Chemical Physics, 21,* 1087-1092.

MUIRHEAD, R.J. (1982). *Aspects of Multivariate Statistical Theory.* New York: John Wiley & Sons.

MULLER, M.E. (1959). A Note on a Method for Generating Points Uniformly on N-dimensional Spheres. *Communications of the ACM, 2,* 19-20.

NEWBOLD, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika, 61,* 423-426.

NICHOLLS, D.F. and PAGAN, A.R. (1985). Varying coefficient regression. In *Handbook of Statistics* (E.J. Hannan, P.R. Krishnaiah and M.M. Rao, eds.) *5,* 413-449. Amsterdam: North Holland.

NICHOLLS, D.F. and QUINN, B.G. (1982). *Random Coefficient Autoregressive Models: An Introduction.* New York: Springer-Verlag.

PALMER, M.J., PHILLIPS, B.F. and SMITH, G.T. (1991). Application of Nonlinear Models with Random Coefficients to Growth Data. *Biometrics, 47,* 623-635.

PANDIT, S.M. and WU, S.M. (1983). *Time series and system analysis with applications.* New York: John Wiley and Sons.

POTTHOFF, R.F. and ROY, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51,* 313-326.

PRESS, W.H., FLANNERY, B.P., TEUKOLSKY, S.A. and VETTERLING, W.T. (1986). *Numerical Recipes. The Art of Scientific Computing.* Cambridge University

Press.

RACINE, A., GRIEVE, A.P. and FLÜLER, H. (1986). Bayesian methods in practice: Experiences in the Pharmaceutical industry. *Applied Statistics, 35,* 93-150.

RAJ, B. and ULLAH, A. (1981). *Econometrics, A Varying Coefficients Approach.* London: Croom-Helm.

RAO, C.R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika, 46,* 49-58.

RAO, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika, 52,* 447-458.

REICHARDT, C.S. (1991). Comments: The Application of Time Series Methods to Moderate Span Longitudinal Data. In: *Best methods for the analysis of change: recent advances, unanswered questions, future directions.* (L.M. Collins and J.L. Horn. (eds.)), 75-91. Washington DC: American Psychological Association.

ROBERTS, G.O. and SMITH, A.F.M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications, 49, no. 2,* 207-216.

SAS INSTITUTE INC. (1988). *SAS/ETS® User's Guide, Version 6, First Edition, Cary,* NC: SAS Institute Inc.

SHUMWAY, R.H. (1988). *Applied Statistical Time Series Analysis.* New Jersey: Prentice-Hall.

SMITH, A.F.M. and ROBERTS, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal*

Statistical Society, Series B, 55, no. 1, 3-23.

SPLIID, H. (1983). A Fast Estimation Method for the Vector Autoregressive Moving Average Model With Exogenous Variables. *Journal of the American Statistical Association, 78,* 843-849.

STOER, J. and BULIRSCH, R. (1980). *Introduction to Numerical Analysis.* New York: Springer-Verlag.

STROUD, A.H. and SECREST, D. (1966). *Gaussian quadrature formulas.* New Jersey: Prentice-Hall.

TIAO, G. C. and BOX, G.E.P. (1981). Modelling multiple time series with applications. *Journal of the American Statistical Association, 76,* 802-816.

TUNNICLIFFE-WILSON, G. (1973). The estimation of parameters in multivariate time series analysis. *Journal of the Royal Statistical Society Series B, 35,* 76-85.

VAN DER MERWE, A.J. and BOTHA, T.J. (1993). Bayesian estimation in mixed linear models using the Gibbs sampler. *South African Statistical Journal, 27,* 149-180.

WASAN, M.T. (1970). *Parametric Estimation.* New York: McGraw-Hill.

WECKER, W. and ANSLEY, C.F. (1983). The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing. *Journal of the American Statistical Association, 78,* 81-89.

ZELLNER, A. (1976). Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms. *Journal of the American Statistical Association, 71,* 400-405.