COVARIANCE MATRIX THEOREMS FOR ESTIMATORS OF TIME SERIES MODELS, WITH APPLICATIONS TO ACTIVE TRACKING PROBLEMS

by

Judith Mary Unwin
Royal Holloway College

Thesis submitted to the University of London for the degree of Doctor of Philosophy

June 1984

ABSTRACT

In this thesis, covariance matrices and generalised variances for maximum likelihood estimators of Gaussian autoregressive moving average time series models are derived. It is shown that estimators for pure moving average and pure autoregressive models have covariance matrices which are expressed in terms of two triangular matrices. Furthermore, the generalised variance is obtained from a factorisation of the determinant of the covariance matrix into four constituent parts. Examples of these theorems are given. The results are generalised for estimators of a mixed autoregressive moving average model in which there is either just one moving average parameter or just one autoregressive parameter. In particular the generalised variance is factorised into the determinants of the covariance matrices for efficient estimators of the parameters of the corresponding two pure models, and two other scalar terms. The submatrices of the covariance matrix for the efficient estimators of the parameters of the general mixed model are found by specifying four or five upper triangular matrices, whose non-zero elements are single parameters of the model, and then carrying out some matrix multiplications and additions. Provided the model is not too large, explicit expressions for the variances and covariances can be obtained. Examples, using mixed models, of these methods are given, and the adequacy of the fitted model is discussed in detail.

It is proposed that these theorems enable statistical tests to be applied to problems of active tracking, which, traditionally, are expressed in terms of polynomial-projecting dynamic linear models. The problem of testing for constant velocity is considered in detail. A test based on a generalisation of Student's t test is discussed. Several examples of this test procedure are given.

TABLE OF CONTENTS

ADDENDA

p.29 Define the autocorrelation of lag h,

$$\rho_h = \frac{\sum_{j=0}^{s-h} \beta_j \beta_{j+h}}{1 + \sum_{j=1}^{s} \beta_j^2} \qquad 1 \leq h \leq s$$

where s is the proposed order of the model.

p.75 Redundant or nearly redundant factors are not necessarily obvious, particularly as there may be large sampling errors in the parameter estimates. Thus factorizing $\alpha(z)$ & $\beta(z)$ using the obtained parameter estimates will not always help in identifying where the inadequacy of the fitted model lies. It would be useful instead to examine the variances and covariances of these estimates to see how large they are. If there are some unexpectedly large values in the covariance matrix then this suggests that the fitted model contains the wrong combination of autoregressive and moving average parameters. Another indication is provided by the generalised variance which will be large if the order of the model or the parameter values have been poorly determined.

p.132 In Section 8.2, we assume that the noise components b_t, r_t are independent and that they have approximately normal $N(0, \sigma_b^2)$, $N(0, \sigma_r^2)$ distributions respectively. The bearing can be measured very accurately in practical situations even at relatively long range which implies that the variance σ_b^2 will be small. In order to establish that $(\nabla X_t - \mu_X)$ has the same correlation structure as a first order moving average process, the analysis on pp.134-137 assumes $|b_t| < 1$. This assumption appears to be valid on the whole since σ_b^2 is small, but if $|b_t| > 1$ the result is less certain. The simulations in the next section attempt to demonstrate that the test still provides a valid method for detecting velocity changes using cartestian data converted from bearing and range measurements.

PREFACE

This thesis is not substantially the same as that submitted for any degree, diploma or other qualification at any other University, nor is any part of this thesis concurrently being submitted for any such degree, diploma or other qualification. Except where specific reference is made to the work of others, the material presented here is believed to be entirely original.

I acknowledge the financial support of an S.E.R.C. C.A.S.E. Award jointly with Ferranti Computer Systems Limited which enabled me to undertake this work.

Thanks are due to Dr. J.R. Moon of Ferranti for setting up the award and to Dr. D.W. Cruse for his continued support. I am grateful to his colleagues Dr. J.M. Stone and Mr. J. Wood for their help and guidance with the computer programming. I would especially like to thank my supervisor at Royal Holloway College, Mr. E.J. Godolphin, without whose support and encouragement this work would not have been possible.

Finally, I thank my mother for her care and patience in typing this thesis.

CHAPTER 1

INTRODUCTION

For many years the analysis of data collected over a period of time has involved scientific workers in numerous and varied fields of research. For instance, the data series published by Government departments are examples of time series of interest to economists. One of the longest such series is possibly the figures from the population census which began in the last century. Figures relating to sales of new or established products are needed by market researchers, and records of seismic activity are essential to the geophysicist in his efforts to predict future movements of the earth's crust. In general, observations from such time series are dependent, as in population series, for example, where the size of the population in any one year is dependent on population figures in previous years.

This dependence is usually due to some underlying process which may or may not be known to the analyst or control theorist. This is the case in the active tracking procedure in navigation. It is required to track moving objects in real-time, such as aircraft, and to register the general behaviour of these objects. The observed position of the object at time t is dependent on the position and velocity at previous times through the equations of motion, and this prior information is helpful in building up a model for the motion of the object. The model can be verified by the control theorist and the elements of the underlying processes, if known, are estimated from the available observations.

This situation is more useful than cases where the structure of the underlying process is not known, and a model is fitted from a consideration of the data only.

Most time series are not purely deterministic, hence accurate modelling requires the use of random processes. The classical statistical models, namely the autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) models are linear stationary models which employ random processes. It is frequently assumed that the elements of the sequence $\{\varepsilon_t\}$ have a common variance, and possibly that they are independent and have an identical normal distribution with mean zero. Many naturally occurring time series are not stationary, but this property can often by restored by a suitable non-parametric transformation, such as differencing.

In general it is one of the serious drawbacks of the employment of classical time series in model fitting procedures that the parameters have no straightforward interpretations. Also if the underlying process were to change, it is not clear how this would affect the model parameters. To circumvent this problem the control theorist exploits the underlying process in order to obtain the model. The components of the resulting 'state space formulation' have physical meanings, such as distance and velocity. The control theorist is more interested in the estimation of the state of the underlying process than the prediction of future observations. Several state estimation schemes, known as filters, have been proposed by Kalman (1960) and Kalman & Bucy (1961). One of the most meaningful and popular is a recursive filter, commonly referred to in the literature as the Kalman filter.

The state space model is considered in detail in a forecasting context by Harrison & Stevens (1976). By placing mild constraints such as time independence on the coefficient matrices, and normality on the noise components, the resulting state space model is seen to have some useful properties. For example, the predictors of the steady model are the same as those of certain low-order non-stationary time series models. For state space models of larger order, Godolphin & Stone (1980) have shown that by fixing values on the products of the coefficient matrices, these models can be interpreted as polynomial-projecting models of degree d, where the dimension, n, of the system vector satisfies $n \geq d+1$. In the equilibrium state the predictors of these models are identical to those of a class of non-stationary time series in which the degree of differencing required to restore stationarity is d or $d+1$. It follows that data generated by a state space model satisfying these conditions can also be described by stationary time series models after differencing the data a suitable number of times. This property then permits us to apply the well-established inference techniques of time series analysis to these data. The usefulness of this dual representation forms the basis of a test for constant velocity which will be described in detail in what follows.

A very general stationary time series is the ARMA model which has $p+q$ unknown parameters $\alpha_1,\ldots,\alpha_p,\beta_1,\ldots\beta_q$ and an unknown variance σ^2. Either p or q may be zero, yielding pure moving average or pure autoregressive models respectively. A problem that has occupied the attention of statisticians in recent years is the estimation of the parameters of the model, based on a realisation $\underline{X}$ of n consecutive observations. One of the forerunners in this field is Whittle (1953). He has shown that the maximum likelihood approach yields a consistent estimator which is asymptotically normal provided the random process $\{\varepsilon_t\}$ is Gaussian. Whittle has also shown that if $\{\varepsilon_t\}$ belongs to a wide class of non-Gaussian distributions then the least squares approach yields consistent estimators with asymptotically minimum variance among a certain class of estimators.

In common with Whittle, the maximum likelihood approaches of Durbin (1960) and Walker (1962) reduce the data $\underline{X}$ to a set of m sample serial correlations. These statistics are often used to examine standard problems in time series analysis. It appears to be widely accepted that most of the information on the model parameters is contained in sample serial correlations of relatively small lag. Typically m is of the order of 30 when n is greater than 100 with an expectation of little loss in estimator efficiency. In estimating the parameters of an autoregressive process, Mann & Wald (1943) showed that consistent estimators for $\alpha_1,\ldots,\alpha_p$ are obtained by replacing α_k with $\hat{\alpha}_k$ and the covariances γ_k by the sample serial covariances c_k in the Yule-Walker equations. Whittle later showed that these estimators were also efficient.

Several alternative approaches exist when moving average models are considered. Whittle's maximum likelihood estimator is not obtained in closed form even for $q=1$, as was emphasized by Durbin (1959). Instead, Durbin approximates the moving average process by a high order autoregression and then invokes the theorem of Mann & Wald. Walker (1961) also concludes that Whittle's approach would be rather cumbersome in practice. His alternative procedure is based on determining an asymptotically efficient estimator for the correlations and solving for $\beta_1,\ldots,\beta_q$ using the Cramér-Wold factorization. A comment by Whittle (1954,p.212) suggests a direct approach to maximising the likelihood. This has been considered further by Godolphin (1977, 1978a). He shows that the maximum likelihood estimators of the parameters are given asymptotically by a linear combination of the sample serial correlations. The calculations are straightforward for any value of q. A computer implementation has been published by Angell & Godolphin (1978).

Box & Jenkins (1970, Chapter 7) suggest a computational approach to maximising the likelihood function; their method is based on a search procedure for minimising the residual sum of squares $\Sigma\varepsilon_t^2$. Other related methods are given by Ansley (1979), Ljung & Box (1979) and Nicholls & Hall (1979). The approach of Anderson (1975) is based on the method of scoring. The methods of Box & Jenkins and Anderson employ the data as it stands without transforming the data to a set of sample serial correlations. Hannan's procedure (1969) is based on the fact that the periodogram approximately diagonalizes the covariance matrix of the observations, provided n is sufficiently large.

Despite the volume of material on inference for time series models, comparatively little interest has been shown in the literature in the computation of the covariance matrix for the efficient estimator $\hat{\underline{\theta}}$ of $\underline{\theta} = (\alpha_1 \dots \alpha_p \; \beta_1 \dots \beta_q)'$. This problem has been in existence for 30 years or more. The work of Whittle (1953) was added to by Durbin (1959) and Box & Jenkins (1970). It is shown that the maximum likelihood approach yields a covariance matrix $\underline{V}/n$ for $\hat{\underline{\theta}}$ which is smallest in the following sense. The difference $\underline{V}^* - \underline{V}$ is positive semi-definite when $\underline{V}^*/n$ is the covariance matrix for any alternative consistent estimator. This implies that the variances and covariances in $\underline{V}/n$ are each smaller than the corresponding variances and covariances in $\underline{V}^*/n$; also the generalised variance is smaller than that of $\underline{V}^*/n$, i.e. det $\underline{V} \leq$ det $\underline{V}^*$. Whittle (1953) also derived a formula for calculating the individual elements of $\underline{V}^{-1}$ where $n\underline{V}^{-1}$ is the information matrix of the model. The formula contains complex integrals and is rather awkward to apply. An alternative formulation exists, based on an algorithm of Quenouille (1947a). This expression eliminates the complex integrals, but is still very cumbersome to use except for the smallest of models. For models containing no moving average parameters, Durbin (1959) has produced a matrix expression for $\underline{V}$ as a whole. He noted that $\underline{V}^{-1} = \underline{\Gamma}_p$ where $\sigma^2\underline{\Gamma}_p$ is the covariance matrix for p consecutive observations of the process. However this result does not generalise to include ARMA models. Pagano (1973) has examined Durbin's result and formed an expression for $\underline{V}$ in terms of the products and differences of upper triangular matrices. Although Pagano's expression has been quoted by other workers, a proof does not appear to have been given in the literature. A simple adaptation yields an equivalent expression for purely moving average models as well, but no form for mixed models results from this duality.

Box & Jenkins (1970, §A7.5) suggest another approach for evaluating $\underline{V}/n$. They derive the information matrix for an autoregressive process of order $p+q$ whose parameters are the inverse zeros of $\alpha(z) = 1+\alpha_1z+\ldots+\alpha_pz^p$ and $\beta(z) = 1+\beta_1z+\ldots+\beta_qz^q$. In moderate or large samples this is approximately the information matrix for the ARMA(p,q) process with parameter vector $\underline{\Theta}$. This matrix then has to be inverted in order to obtain the covariance matrix for the efficient estimator of $\underline{\Theta}$. The procedure is demonstrated with the simplest case, the ARMA(1,1) model. However if p or q is strictly greater than unity the technique is complicated to put into practice, particularly if some of the zeros of $\alpha(z)$ or $\beta(z)$ are complex-valued. The accuracy of this method as the number of model parameters increases, is uncertain, and it is not clear whether this approach could be adapted to evaluate $\underline{V}$, or the information matrix directly.

In this thesis we present simple procedures for evaluating $\underline{V}$ for stationary linear time series models. Also of interest is the specification of the information matrix and the generalised variance. For pure models, a proof is given of an expression for $\underline{V}$ in terms of upper triangular matrices whose orders are equal to that of the model considered. The specification of $\underline{V}$ using these matrices simplifies the determinant of $\underline{V}$ into the product of four terms, each of which is easily calculated. These results are generalised for mixed models which contain either just one moving average or just one autoregressive parameter. By defining two further upper triangular matrices, the submatrices of the information matrix for the general autoregressive moving average process can be evaluated. The covariance matrix $\underline{V}/n$ is obtained by inverting the information matrix and preserving the same partitioning.

Under certain specified conditions, a subclass of general time series models have similar properties to polynomial-projecting dynamic linear models. Thus problems relating to active tracking can be examined using classical statistical tests as alternatives to the state estimation schemes usually associated with such problems. The testing for constant velocity is considered in detail. A test based on a generalisation of Student's t test is derived and the results of various simulations are given in detail.

In Chapter 2 we establish the basic statistical properties of three stationary time series, namely the autoregressive, moving average and mixed autoregressive moving average models. Some of the inference techniques mentioned previously are described, together with methods for testing for specification of models. State space models are also defined and the effect of restricting the components of the system is considered.

Chapter 3 examines in detail the purely autoregressive model. The model is treated separately not only for reasons of its simplicity but also because it appears to be widely used in practice. The techniques presented in this chapter form a basis of ideas which will be used or adapted later, when more complicated models are being considered. The elements of the information matrix are given by a simplified form of Whittle's result (1953) for the elements of $\underline{V}^{-1}$. However, if p is only moderately large it appears to be simpler to use an alternative method. A proof is given of an expression for $\underline{V}$ based on Durbin's result (1959). Hence the covariance matrix for the efficient estimator of the model parameters can easily be specified and inverting this matrix is the easiest way in which to form the information matrix. A simple formula is derived for the generalised variance which relies on the properties of the upper triangular matrices in the expression for $\underline{V}$.

In Chapter 4 we examine a subclass of general ARMA(p,q) models containing either just one moving average parameter or just one autoregressive parameter, i.e. we consider ARMA(p,1) and ARMA(1,q) models only. This subclass contains several simplifications compared to the general class of mixed models, and since in practice p and q are usually quite small, it seems likely that the chosen model may fall into this subclass. The information matrix is obtained in partitioned form with submatrices on the diagonal given by formulae from Chapter 3. The off-diagonal block is easily specified in this case, since it is a vector and not a matrix. Durbin's result is not applicable to the matrix $\underline{V}^{-1}$ as a whole, so the covariance matrix $\underline{V}/n$ can only be derived by inverting $\underline{V}^{-1}$ and preserving the same partitioning. The addition of one extra parameter makes the specification of the generalised variance also more complicated than for pure models. The problem is overcome by defining upper triangular matrices with similar properties to those in Chapter 3. The concise expression for the vector component of the information matrix also plays an important part. Proof is given of an elegant factorization of det $\underline{V}$, together with an example.

An algorithm is presented in Chapter 5 for evaluating the covariance matrix for the efficient estimators of the parameters of the general ARMA(p,q) model in which both p and q are strictly greater than unity. As in Chapter 4, the method is based on inverting $\underline{V}^{-1}$ written in a partitioned form. However the off-diagonal block of the information matrix $n\underline{V}^{-1}$ is no longer a simple vector. It would appear that the $p+q-1$ different elements of this $p \times q$ matrix can only be specified using Whittle's formula. But by defining two further upper triangular matrices and taking products and additions with the previously defined matrices, a $p \times p$ matrix is formed, whose inverse contains the off-diagonal block of $\underline{V}^{-1}$ in its first q columns. The result is proved assuming $p \geq q$; details are also given of the specification of the information matrix and the covariance matrix for $\hat{\underline{\theta}}$ if $p < q$.

In Chapter 6 we define the univariate state space model in the form given by Harrison & Stevens. By placing a mild constraint on the coefficient matrices a large subclass of these models can be interpreted as polynomial-projecting models. These models possess the property that the forecast function is a polynomial in the prediction lead-time. The same properties have been demonstrated to hold for certain non-stationary time series models in which the degree of differencing is equal to, or one greater than the degree of the polynomial-projecting model. In the former category, a deterministic term is also present in the model, and the number of moving average parameters is less than or equal to the degree of the polynomial-projecting model. In order that classical inference techniques of time series are applicable, the moving average model should be invertible. It seems that this criterion will usually be the deciding factor as to which time series model should be used to describe the given data.

Chapter 7 examines a problem of particular interest to control engineers engaged in the active tracking of marine craft. The aim is to quickly detect manoeuvres in the object which can be observed as velocity changes. By estimating the velocities before and after a suspected velocity change, their difference can be tested for significance. A test statistic is formulated which is a generalisation of Student's t test. The state space form for constant velocity is described. The appropriate time series model is given and relationships are derived between the variances of the noise components of each model. Estimates of the parameters of the time series model are required in the test statistic. These are obtained using the maximum likelihood principle. The test is applied to simulated data which represent the cartesian co-ordinates of the object relative to the observer. The two components of the data at each time point are taken to be independent, and the test is performed separately on each set of co-ordinate data. The sensitivity of the test is assessed by performing the test on sets of data containing a wide variety of velocity changes or no velocity change at all.

In Chapter 8 we consider the active tracking problem when the data assume a different form to that of Chapter 7. The range of the object is observed by noting how long the signal emitted by the observer takes to return to him. It is assumed that the bearing of the signal can also be measured. With these two pieces of information, the location of the object in the plane is known, within the accuracy of the measuring instruments. It is not feasible to apply the test described in the previous chapter to the bearing and range data independently, since either may remain constant even when the object is manoeuvring. However we show that the test appears to be valid if the bearing-range data are first converted into cartesian co-ordinate data. The resulting data sets relating to the X and Y co-ordinates are considered independent and the test proceeds as in Chapter 7. It appears that the test performs comparably well on the cartesian data converted from bearing-range data as on the genuine cartesian data of Chapter 7.

CHAPTER 2

TIME SERIES

2.1 Introduction

In this chapter the statistical properties of univariate time series models are considered. It is generally assumed that a realisation $X_1,\ldots,X_n$ of size n is available. Three basic time series models are defined, namely the autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) models. The estimation of a possibly non-zero mean is considered in some cases. Some methods are outlined for determining estimates of the parameters of the models. Various tests for establishing the order of the proposed model are also presented. In the final section state space models are described. These models attempt to describe time series data in a way that is more acceptable to practitioners in that the components of the system equation have intuitive interpretations.

2.2 Statistical Properties of Autoregressive Models

The autoregressive model of order p with zero mean is defined by

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \varepsilon_t \qquad (2.2.1)$$

where $\{\varepsilon_t\}$ is a sequence of uncorrelated Gaussian random variables with a common variance σ^2. i.e.

$$E(\varepsilon_t) = 0 \qquad E(\varepsilon_t\varepsilon_k) = \delta_{t,k}\sigma^2,$$

where $\delta_{t,k}$ is the Kronecker delta. The model (2.2.1) is stationary provided the inverse zeros of the autoregressive coefficient generating function

$$\alpha(z) = 1 + \alpha_1 z + \ldots + \alpha_p z^p$$

are less than one in modulus.

The autocovariances $\gamma_k = E(X_t X_{t+k})$ satisfy the Yule-Walker equations:

$$\gamma_k + \alpha_1\gamma_{k-1} + \ldots + \alpha_p\gamma_{k-p} = 0, \quad \gamma_k = \gamma_{-k} \qquad (2.2.2)$$

for $k \geq 1$. Multiplying (2.2.1) by X_{t+k} and taking expectations gives the Wold equations:

$$\gamma_k + \alpha_1\gamma_{k+1} + \ldots + \alpha_p\gamma_{k+p} = \sigma^2 b_k$$

where

$$B(z) = \sum_{k=o}^{\infty} b_k z^k = \frac{1}{\alpha(z)} ,$$

with $b_o \equiv 1$. The autocovariances can be obtained from the model parameters by taking the first Wold and the first p Yule-Walker equations and re-writing them in the form

$$\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} 1 & \underline{\alpha}' \\ \underline{\alpha} & \underline{A}'+\underline{W} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^2 \\ \underline{0}_p \end{bmatrix} \qquad (2.2.3)$$

where $\underline{\alpha} = (\alpha_1 \ . \ . \ . \ \alpha_p)'$ and

$$\underline{A} = \begin{bmatrix} 1 & \alpha_1 & \ldots & \alpha_{p-1} \\ 0 & 1 & \ldots & \alpha_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}, \quad \underline{W} = \begin{bmatrix} \alpha_2 & \alpha_3 & \ldots & \alpha_p & 0 \\ \alpha_3 & \alpha_4 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \alpha_p & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix}$$

and $\underline{0}_p$ is the zero vector of length p. It is interesting to note that the matrix $\underline{A}$ features again in the next chapter, where the covariance matrix for the efficient estimators of $(\alpha_1,\ldots,\alpha_p)$ is sought.

An alternative approach to evaluating the autocovariances is due to Quenouille (1947a). The autocovariance generating function is defined by

$$\Gamma(z) = \gamma_0 + \sum_{k=1}^{\infty} \gamma_k(z^k+z^{-k})$$

and satisfies

$$\Gamma(z) = \sigma^2 B(z)B(z^{-1}) \quad .$$

The expression $\Gamma(z)/\sigma^2$ is uniquely determined by

$$\frac{1}{\alpha(z)\alpha(z^{-1})} = K_0 + \frac{(K_1 z +\ldots+ K_p z^p)}{\alpha(z)} + \frac{(K_1 z^{-1} +\ldots+ K_p z^{-p})}{\alpha(z^{-1})} \quad ,$$

where $K_0,K_1,\ldots,K_p$ are found by equating coefficients of $z^0,z^1,\ldots z^p$. Then

$$\gamma_0 = K_0\sigma^2$$

and the covariances $\gamma_1,\gamma_2,\ldots$ are given by the relation

$$\alpha(z)\sum_{k=1}^{\infty} \gamma_k z^k = (K_1 z + K_2 z^2 + \ . \ . \ . + K_p z^p)\sigma^2 \ .$$

The converse problem, that of finding the model parameters from the autocovariances is straightforward for autoregressive models; by writing the Yule-Walker equations (2.2.2) in matrix notation it follows that

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = - \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix}$$

in which form the parameters can readily be obtained.

In considering the problem of estimating $\alpha_1,\ldots,\alpha_p$ and σ^2 we assume that the process (2.2.1) is stationary and that the sequence $\{\varepsilon_t\}$ consists of independent and identically distributed random variables. Mann & Wald (1943) showed that for large samples, the maximum likelihood solutions for $\hat{\alpha}_1,\ldots,\hat{\alpha}_p$ are given by the intuitively sensible approach of solving the Yule-Walker equations (2.2.2) with α_k replaced by $\hat{\alpha}_k$ and γ_k replaced by the sample serial covariance c_k of lag k defined by

$$c_k = \frac{1}{n-k}\sum_{t=1}^{n-k} X_t X_{t+k} \quad .$$

The variance is estimated by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(X_t + \sum_{k=1}^{p}\hat{\alpha}_k X_{t-k}\right)^2 \quad .$$

Mann & Wald showed that their approach yields consistent estimators for $\alpha_1,\ldots,\alpha_p$ & σ^2 and also that the joint limiting distribution of the statistics

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_1), \; . \; . \; . \; , \sqrt{n}(\hat{\alpha}_p - \alpha_p)$$

is multivariate normal, with mean zero and covariance matrix $\underline{V}$. The expression for $\underline{V}$ is rather complicated; one of their achievementswas to show that $\underline{V}$ is independent of n.

Whittle (1953) later showed that $\hat{\alpha}_1,\ldots,\hat{\alpha}_p$ were also efficient estimators of $\alpha_1,\ldots,\alpha_p$. From this paper we can also deduce that the corresponding covariance matrix is given by

$$\frac{\underline{V}}{n} = \frac{\underline{F}_{\alpha\alpha}^{-1}}{n}$$

where the (i,j)-th element of $\underset{\sim}{F}_{\alpha\alpha}$ is the constant term in the expansion of

$$\frac{1}{2}\frac{\partial}{\partial\alpha_i}\{\log B(z)B(z^{-1})\}\frac{\partial}{\partial\alpha_j}\{\log B(z)B(z^{-1})\} \qquad (2.2.4)$$

with $B(z) = \{\alpha(z)\}^{-1}$ for purely autoregressive models. A straightforward method for evaluating $\underset{\sim}{F}_{\alpha\alpha}^{-1}$ is presented in the next chapter, and the information matrix and generalised variance are also examined.

2.3 Statistical Properties of Moving Average Models

The model with unknown mean μ is defined by

$$X_t = \mu + \varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q} \qquad (2.3.1)$$

where μ is the deterministic term representing the mean of the process, and $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables as defined in the previous section. The invertibility condition is that the inverse zeros of the moving average coefficient generating function

$$B(z) = \beta(z) = 1 + \beta_1 z + \ldots + \beta_q z^q$$

are less than one in modulus.

Given a realisation $X_1,\ldots,X_n$, we require an estimate for μ. In general the observations are correlated and unless the process is completely stationary then their distributions may all be different. In these cases, general estimation methods such as maximum likelihood are not applicable. However, even if the observations have different distributions, they all have the same mean value μ, suggesting that the usual estimate, namely the sample mean, may still provide a reasonable estimate. It is easy to see that

$$E(\overline{X}) = \mu ,$$

so $\overline{X}$ is an unbiased estimate of μ. For large n,

$$\mathrm{Var}(\overline{X}) = (1 + \beta_1 + . . . + \beta_q)^2 \frac{\sigma^2}{n}$$

which tends to zero as n tends to infinity, so $\overline{X}$ is a consistent and unbiased estimator of μ. Depending on the actual values of the parameters, $\mathrm{Var}(\overline{X})$ may be larger than σ^2/n, the value applicable to n independent observations. It is interesting to note that the derivation of the result for $\mathrm{Var}(\overline{X})$ is given in a frequency domain context by Priestley (1981, pp318).

The estimation of the parameters of the model (2.3.1) has been considered by Durbin (1959). He approximates the model by a high order autoregressive process and then invokes the theorem of Mann & Wald (1943). Whittle (1951, 1953) considers the problem using the maximum likelihood principle. These estimates, $\hat{\beta}_1,\ldots,\hat{\beta}_q$ are functions of the sample serial correlations $r_k=c_k/c_o$ where

$$c_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t-\overline{X})(X_{t+k}-\overline{X}) .$$

Whittle's solution which is consistent and efficient is not in closed form, but can be found using an iterative process. A direct representation of the iterative solution in terms of the sample serial correlations has been derived by Godolphin (1977,1978a) and a computer implementation published by Angell & Godolphin (1978).

The joint limiting distribution of the statistics

$$\sqrt{n}(\hat{\beta}_1 - \beta_1), . . . , \sqrt{n}(\hat{\beta}_q - \beta_q)$$

is $N(0, \underset{\sim}{F}_{\beta\beta}^{-1})$ which is analogous to the result given in the previous section for purely autoregressive models. An interesting duality result was derived by Whittle. He showed that the covariance matrix for the efficient estimators of the parameters of an autoregressive process was the same as that of a moving average process provided the parameter sets were the same. This result follows since the (i,j)-th element of $\underset{\sim}{F}_{\beta\beta}$ is the constant term in the expansion of

$$\frac{1}{2}\frac{\partial}{\partial\beta_i}\{\log B(z)B(z^{-1})\}\frac{\partial}{\partial\beta_j}\{\log B(z)B(z^{-1})\}$$

which is equivalent to (2.2.3) with, of course, $\alpha(z)=\beta(z)$.

2.4 Statistical Properties of Mixed Models

The mixed autoregressive moving average model of order (p,q), allowing for a non-zero mean μ, is defined by

$$(X_t-\mu)+\alpha_1(X_{t-1}-\mu) +\ldots+ \alpha_p(X_{t-p}-\mu) = \varepsilon_t+\beta_1\varepsilon_{t-1} +\ldots+ \beta_q\varepsilon_{t-q} \quad (2.4.1)$$

and $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables as defined in Section 2.2. To ensure identifiability we impose the following condition :

The polynomials
$\alpha(z) = 1 + \alpha_1 z + \ldots + \alpha_p z^p$ & $\beta(z) = 1 + \beta_1 z + \ldots + \beta_q z^q$.
have no factors in common.

We also require that

(i) All the roots of $\alpha(z)$ & $\beta(z)$ lie outside the unit circle, for stationarity and invertibility respectively.

(ii) The parameters α_p and β_q are not both zero.

The mean is again estimated by $\overline{X}$ which is unbiased. For large n the variance of $\overline{X}$ is

$$\frac{(1 + \beta_1 + \ldots + \beta_q)^2}{(1 + \alpha_1 + \ldots + \alpha_p)^2} \frac{\sigma^2}{n} ,$$

which tends to zero as n increases; thus $\overline{X}$ is an unbiased and consistent estimator of μ.

The covariances can be found, given the parameter values of the model, by adapting Quenouille's algorithm as follows. The autocovariance generating function satisfies

$$\Gamma(z) = \sigma^2 B(z)B(z^{-1})$$

where $B(z) = \beta(z)/\alpha(z)$, and hence $\Gamma(z)/\sigma^2$ is uniquely determined by

$$\frac{\beta(z)\beta(z^{-1})}{\alpha(z)\alpha(z^{-1})} = K_0 + \frac{(K_1 z + \ldots + K_L z^L)}{\alpha(z)} + \frac{(K_1 z^{-1} + \ldots + K_L z^{-L})}{\alpha(z^{-1})} \qquad (2.4.2)$$

where $L = \max(p,q)$.

The converse problem, that of finding the values of the parameters given knowledge of the autocorrelations is perhaps more frequently encountered in practice. The Cramér-Wold factorization seeks a solution for $(\beta_1,\ldots,\beta_q)$ given the first q sample serial correlations and estimates of the autoregressive parameters; the technique has been considered in detail by Godolphin (1976).

The estimation of the parameters of models of the form (2.4.1) has been considered by several authors. Whittle (1953) has shown that the maximum likelihood approach yields consistent estimators $\hat{\alpha}_1,\ldots,\hat{\alpha}_p,\hat{\beta}_1,\ldots,\hat{\beta}_q$. He concludes that these estimators are efficient in the sense that the generalised variance, $\det(\underline{V}/n)$, is smaller than the corresponding generalised variance for any other set of consistent estimators. The methods of Walker (1961, 1962) and Durbin (1960) also reduce the data from the realisation $\underline{X}$ to a set of sample serial correlations r_j. Walker's approach is based on estimating the α's and ρ's, not the α's and β's, taking r_j as the initial estimate of ρ_j. The estimates of the moving average parameters are then found using the Cramér-Wold factorization. This approach requires the theory of the distribution of r_j due to Bartlett (1946) and Lomnicki & Zaremba (1957).

In recent years, a number of authors have examined the exact expression for the likelihood L, given by

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} (\det \underline{\Gamma})^{-\frac{1}{2}} \exp\{\frac{-1}{2\sigma^2} \underline{X}'\underline{\Gamma}^{-1}\underline{X}\}$$

where $\sigma^2\underline{\Gamma}$ is the covariance matrix of $\underline{X}$. Box & Jenkins (1970, §A7.4) derive the exact likelihood function for a moving average process, and this has been extended to the general case by Newbold (1974) and Galbraith & Galbraith (1974). Phadke & Kedem (1978) suggest the Cholesky decomposition $\underline{\Gamma} = \underline{E}\underline{E}'$ where $\underline{E}$ is lower triangular, so that $\underline{X}'\underline{\Gamma}^{-1}\underline{X} = \underline{Y}'\underline{Y}$ where $\underline{Y} = \underline{E}^{-1}\underline{X}$, and $\det \underline{\Gamma} = (\det \underline{E})^2$. Ansley (1979) extends the techniques of Phadke & Kedem to cover ARMA models and shows that his solution is more efficient than many other methods described in the literature. A closed form expression for $\underline{X}'\underline{\Gamma}^{-1}\underline{X}$ is presented by Ljung & Box (1979). They illustrate how $\det \underline{\Gamma}$ and $\underline{X}'\underline{\Gamma}^{-1}\underline{X}$ are evaluated, and, based on some numerical results, they claim that the method has similar efficiency to that of Ansley. Other methods have been given by Osborn (1976), Ali (1977), and Harvey & Phillips (1979) to name a few.

In common with Walker & Durbin, Pham-Dinh (1979) also employs the sample serial correlations. His approach is based on the spectral resolution of the likelihood function, and hence the generally simple form for the estimators is unfortunately obscured. An alternative approach yielding approximately maximum likelihood estimates in terms of the sample serial correlations is that of Godolphin (1980b, 1984), which we briefly outline. The log likelihood for the realisation $\underline{X} = (X_1 \ldots X_n)'$ of the model (2.4.1) is given by

$$\log L = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2}\log\det \underline{\Gamma}_n - \frac{1}{2\sigma^2}(\underline{X}-\mu\underline{1})'\underline{\Gamma}_n^{-1}(\underline{X}-\mu\underline{1}) \qquad (2.4.3)$$

where $\underline{1}$ is a vector of 1's of length n, and where $\sigma^2\underline{\Gamma}_n$ is the covariance matrix of $\underline{X}$. Differentiating (2.4.3) with respect to σ^2 and θ where $\theta \in \underline{\Theta} = (\alpha_1 \ldots \alpha_p\ \beta_1 \ldots \beta_q)'$ we obtain the following approximation to the likelihood equations:

$$\frac{\partial}{\partial\theta}\ (\underline{X} - \mu\underline{1})'\underline{\Gamma}_n^{-1}(\underline{X} - \mu\underline{1}) \ = \ 0 \qquad (2.4.4)$$

and

$$\hat{\sigma}^2 \ = \ \frac{1}{n}(\underline{X} - \bar{X}\underline{1})'\underline{\Gamma}_n^{-1}(\underline{X} - \bar{X}\underline{1})$$

where the contribution $\det \underline{\Gamma}_n$ has been ignored. To obtain the solution for $\underline{\Theta}$ in (2.4.4) it is possible to adopt a further approximation derived by Whittle (1954, §2.5) and considered later by Shaman (1976). We replace $\underline{\Gamma}_n^{-1}$ by $\underline{\Pi}_n$ where $\underline{\Pi}_n=((\pi_{|i-j|}))$ is the covariance matrix for n consecutive observations of the stationary ARMA(q,p) process

$$Y_t + \beta_1 Y_{t-1} + \ldots + \beta_q Y_{t-q} \ = \ \eta_t + \alpha_1\eta_{t-1} + \ldots + \alpha_p\eta_{t-p}$$

where η_t is a sequence of mutually uncorrelated random variables with expectation zero and variance unity. Using this approximation the likelihood equations (2.4.4) become

$$\frac{\partial}{\partial\theta} (\underline{X} - \mu\underline{1})'\underline{\Pi}_n (\underline{X} - \mu\underline{1}) = 0$$

which simplify to

$$\frac{\partial}{\partial\theta} \{\pi_o + 2 \sum_{j=1}^{m} \pi_j r_j\} = 0$$

with $r_j = c_j/c_o$ and

$$c_j = \frac{1}{n-j} \sum_{t=1}^{n-j} (X_t-\overline{X})(X_{t+j}-\overline{X}) .$$

Technically, the number, m, of sample serial correlations should be n-1, but it is generally accepted that m can be of the order of 30 even if n is large with an expectation of little loss of accuracy in the estimates. Solutions for the likelihood equations can be expressed as iterative equations for $\hat{\beta}_1,\ldots,\hat{\beta}_q$, whence the non-iterative solutions for $\hat{\alpha}_1,\ldots,\hat{\alpha}_p$, together with

$$\hat{\sigma}^2 = c_o(\pi_o + 2 \sum_{j=1}^{m} \pi_j r_j) .$$

2.5 Testing for Specification

The estimation of the parameters is an important part of any model fitting procedure. Very often it is necessary to first determine the order of the model. One technique used by many practitioners is to overfit models in the hope that estimates significantly different from zero will effectively determine the required number of parameters. However this practice has disadvantages which will be discussed later (see Section 4.5).

In this section we briefly discuss various tests for specification of stationary time series. One test which is reported to have good power properties is that of Whittle (1951, 1952). His procedure is derived from the likelihood ratio principle and assumes that the data are best fitted by an autoregressive model, but whose order is uncertain. Under the null hypothesis an AR(p) model is fitted and the maximum likelihood estimator of the variance is approximately

$$\hat{\sigma}_p^2 \simeq c_0 + \hat{\alpha}_1 c_1 + \ldots + \hat{\alpha}_p c_p \; .$$

An AR(p+k) model is fitted under H_1, where k is positive, and the variance is estimated by

$$\hat{\sigma}_{p+k}^2 \simeq c_0 + \hat{\alpha}_1 c_1 + \ldots + \hat{\alpha}_{p+k} c_{p+k} \; .$$

The test statistic

$$Q = (n-p-k)(\hat{\sigma}_p^2 - \hat{\sigma}_{p+k}^2)/\hat{\sigma}_{p+k}^2$$

is asymptotically distributed like χ^2 with k degrees of freedom if the null hypothesis is true. If $k=1$, then the asymptotic maximum likelihood estimator of α_{p+1} under H_1 is

$$\hat{a}_{p+1} = \frac{-\{c_{p+1} + \sum_{j=1}^{p} \hat{\alpha}_j c_{p+1-j}\}}{c_0 + \sum_{j=1}^{p} \hat{\alpha}_j c_j}$$

using a recursive procedure of Durbin (1960). Since

$$\hat{\sigma}_{p+1}^2 = (1 - \hat{a}_{p+1}^2)\, \hat{\sigma}_p^2$$

the test statistic simplifies to

$$Q' = (n-p-1)\frac{(\sigma_p^2 - \sigma_{p+1}^2)}{\hat{\sigma}_{p+1}^2} = (n-p-1)\frac{\hat{a}_{p+1}^2}{1-\hat{a}_{p+1}^2}$$

which under H_0 is asymptotically distributed like χ^2 with one degree of freedom.

Quenouille (1947b) provided a test of fit using partial autocorrelations. This was extended by Bartlett & Diananda (1950) and Walker (1952) compares the power of these two tests. He concludes that Quenouille's test is on the whole at least as powerful as that of Bartlett & Diananda, although this will depend on the form of the null hypothesis and the class of alternative hypotheses.

An interesting test for specification for moving average models is available, based on examining the correlation structure of the model. The hypotheses

H_0: observations are from an $MA(q)$

H_1: observations are from an $MA(q+k)$

are replaced by

$$H_0: \underset{\sim}{\rho}_1 \neq \underset{\sim}{0}, \quad \underset{\sim}{\rho}_2 = \underset{\sim}{0}, \quad \underset{\sim}{\rho}_3 = \underset{\sim}{0}$$

$$H_1: \underset{\sim}{\rho}_1 \neq \underset{\sim}{0}, \quad \underset{\sim}{\rho}_2 \neq \underset{\sim}{0}, \quad \underset{\sim}{\rho}_3 = \underset{\sim}{0}$$

where $\underset{\sim}{\rho}_1 = (\rho_1\ \rho_2 \ldots \rho_q)'$, $\underset{\sim}{\rho}_2 = (\rho_{q+1} \ldots \rho_{q+k})'$ and $\underset{\sim}{\rho}_3 = (\rho_{q+k+1} \ldots \rho_m)'$. The test statistic is

$$nQ_m = n\ \hat{\underset{\sim}{\rho}}_2'\ \underset{\sim}{R}_2\ \hat{\underset{\sim}{\rho}}_2$$

where $\underset{\sim}{R}_2$ is the covariance matrix for $\sqrt{n}\hat{\underset{\sim}{\rho}}_2$ under H_1. If the null hypothesis is true then nQ_m is asymptotically distributed like

χ^2 with k degrees of freedom. This method can be considered a modification of a goodness of fit test proposed by Wold (1949). The procedure is a straightforward adaptation of a test for detecting gaps in moving average processes of Godolphin (1978b). In this case the two hypotheses become

$$H_0: \underline{\rho}_1 \neq \underline{0}, \quad \underline{\rho}_2 \neq \underline{0}, \quad \underline{\rho}_3 = \underline{0}$$

$$H_1: \underline{\rho}_1 = \underline{0}, \quad \underline{\rho}_2 \neq \underline{0}, \quad \underline{\rho}_3 = \underline{0}$$

and the test statistic is based on the covariance matrix for $\sqrt{n}\underline{\rho}_1$ under H_1.

For mixed models, let $\underline{V}/n$ denote the covariance matrix for the efficient estimator of $\underline{\theta} = (\alpha_1 \ldots \alpha_p \; \beta_1 \ldots \beta_q)'$. Then an intuitively sensible test of $\underline{\theta} = \underline{0}$ is

$$n\hat{\underline{\theta}}'\underline{V}^{-1}\hat{\underline{\theta}}$$

which would be chi-squared with p+q degrees of freedom in large samples if the hypothesis were true. Now $\underline{V}^{-1}$ is n times the information matrix which in partitioned form is

$$n\underline{F} = n\begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & \underline{F}_{\beta\beta} \end{bmatrix}.$$

The derivations of $\underline{V}$ and $\underline{F}$ are considered in full in Chapter 5. At this point it is sufficient to say that $\underline{F}$ can be specified more easily than $\underline{V}^{-1}$ so the statistic becomes $n\hat{\underline{\theta}}'\underline{F}\hat{\underline{\theta}}$.

This test can easily be adapted in order to test say H_0: $\underline{\alpha} = \underline{0}$. Then

$$n\hat{\underline{\alpha}}'\underline{F}_{\alpha\alpha}\hat{\underline{\alpha}}$$

is asymptotically distributed like chi-squared with p degrees of freedom if H_0 is true. To test $\underline{\beta} = \underline{0}$ the test statistic is

$$n\underline{\hat{\beta}}'\underline{F}_{\beta\beta}\underline{\hat{\beta}} \ .$$

Here $n\underline{F}_{\alpha\alpha}$ & $n\underline{F}_{\beta\beta}$ are the information matrices for a pure autoregression and a pure moving average process respectively, and their specification is discussed in the next chapter.

A simple test which is often used by practitioners is the Box-Pierce test (1970) or its modification by Ljung & Box (1978). The Box-Pierce test requires the computation of

$$n \sum_{j=1}^{T} r_j^{\,2}$$

where r_j is the sample serial correlation and T is a sufficiently large number less than n. This statistic has a limiting χ^2 distribution on $T-p-q$ degrees of freedom. However, this simple test has the reputation of being unable to distinguish between several models which could be fitted to the data. The modification of Ljung & Box gives a test statistic

$$n(n+2) \sum_{k=1}^{T} (n-k)^{-1} r_k^{\,2}$$

which provides a closer approximation to χ^2 on $T-p-q$ degrees of freedom.

Another approach which contains the Box-Pierce test as a special case has been proposed by Godolphin (1980a). This method requires more computation but has greater power properties. It is based on Walker's idea that we should test the ρ's rather than the β's, using the Godolphin (1978a) estimation procedure. The set of sample serial correlations $(r_1,\ldots,r_T)$ are transformed to a set $\underline{w} = (w_1 \ \ldots \ w_{T-k})'$ which is partitioned into $(\underline{w}_* \ w_{m+1} \ \ldots \ w_{T-k})'$ with transformed covariance matrix

$$\begin{bmatrix} \underline{\Omega}_{11} & \underline{\Omega}_{12} \\ \underline{\Omega}_{12}' & \underline{\Omega}_{22} \end{bmatrix} .$$

$\underline{w}_* = (w_1 \ldots w_m)'$ is then estimated using Walker's iterative procedure. The test statistic

$$nQ_{T-k-m} = n\underline{w}_*'(\underline{\Omega}_{11} - \underline{\Omega}_{12}\underline{\Omega}_{22}^{-1}\underline{\Omega}_{12}')^{-1}\underline{w}_*$$

has a central χ^2 distribution on m degrees of freedom under the null hypothesis that the parameters of the ARMA(p,q) model have been correctly specified.

A comparison of the tests proposed by Whittle, Ljung & Box and Godolphin for autoregressive models has been made by Clarke & Godolphin (1982).

2.6 The State Space Model

The basic state space model is of the form

$$\underline{X}_t = \underline{F}_t\underline{\theta}_t + \underline{v}_t \qquad (2.6.1)$$

$$\underline{\theta}_t = \underline{G}_t\underline{\theta}_{t-1} + \underline{H}_t\underline{w}_t \qquad (2.6.2)$$

where $\underline{\theta}_t$ is the process vector varying in time, subject to the random term $\underline{H}_t\underline{w}_t$. The observations $\underline{X}_t$ of the function $\underline{F}_t\underline{\theta}_t$ are made at discrete, not necessarily regular, intervals of time and are subject to a random measurement error $\underline{v}_t$. The vectors $\underline{X}_t, \underline{v}_t$ are of order mx1, $\underline{\theta}_t$ is of order nx1 and $\underline{w}_t$ is of order rx1. The matrices $\underline{F}_t$, $\underline{G}_t$ & $\underline{H}_t$ are all matrices assumed known at time t, of dimension mxn, nxn and nxr respectively. The random vectors $\underline{v}_t$, $\underline{w}_t$ are taken to satisfy the following constraints:

$$\begin{array}{llll} E(\underline{v}_t) = \underline{0} & & E(\underline{w}_t) = \underline{0} & \\ E(\underline{v}_t\underline{v}_t') = \underline{V}_t & & E(\underline{w}_t\underline{w}_t') = \underline{W}_t & \\ E(\underline{v}_t\underline{v}_{t+k}') = \underline{0} & (k \neq 0) & E(\underline{w}_t\underline{w}_{t+k}') = \underline{0} & (k \neq 0) \ . \end{array} \qquad (2.6.3)$$

Also the noise components $\underline{v}_t$ & $\underline{w}_t$ are uncorrelated.

The estimator of the process vector $\underline{\theta}_t$ is given by

$$\hat{\underline{\theta}}_t = \underline{G}_t\hat{\underline{\theta}}_{t-1} + \underline{A}_t(\underline{X}_t - \underline{F}_t\underline{G}_t\hat{\underline{\theta}}_{t-1}) \qquad (2.6.4)$$

where $\underline{A}_t$ is the Kalman gain matrix. Kalman (1963) suggested that $\underline{A}_t$ should be chosen so as to minimise

$$\underline{C}_t = E\left[(\underline{\theta}_t - \hat{\underline{\theta}}_t)'(\underline{\theta}_t - \hat{\underline{\theta}}_t)\right] \ .$$

Various assumptions are frequently made concerning the forms of $\underline{V}_t$ and $\underline{W}_t$. If $\underline{V}_t$ is positive definite, then $\underline{A}_t$ can be expressed in a form more open to interpretation, namely

$$\underline{A}_t = \underline{C}_t\underline{F}_t'\underline{V}_t^{-1} \ .$$

It is interesting to see how $\underline{A}_t$ copes with various uncertainties in the model. For example, if $\underline{F}_t = \underline{I}$ and $\underline{V}_t$ is diagonal, then each element of $\underline{A}_t$ is proportional to the uncertainty of the estimate, and inversely proportional to the measurement noise. Thus if measurement noise is large and estimation errors are small, then $\underline{A}_t$ is small. Thus little attention is paid to the most recent observation because we have more confidence in the previous estimator. Conversely, if measurement noise is small and estimation errors are large, then $\underline{A}_t$ is large, demonstrating the need for more information.

The assumption of normality for $\underline{V}_t$ and $\underline{W}_t$ is frequently made; if $\underline{G}_t$ is independent of time, then $\underline{H}_t = \underline{I}$ and equations

(2.6.1), (2.6.2) define the state space model in the form given by Harrison & Stevens (1976). Under certain conditions, these models have similar properties to low order ARMA models. For example, if $\underline{v}_t$ is absent and the Kalman gain vector $\underline{A}_t$ has converged to $\underline{A}$ then the univariate state space model becomes

$$X_t = \underline{F}\hat{\underline{\theta}}_t$$
$$\hat{\underline{\theta}}_t = \underline{G}\hat{\underline{\theta}}_{t-1} + \underline{A}e_t$$

where $e_t = X_t - \underline{F}\underline{G}\hat{\underline{\theta}}_{t-1}$. The sequence $\{e_t\}$ which consists of one-step ahead prediction errors, replaces the random sequence $\{\varepsilon_t\}$ which is common to the time series models defined in §§2.2 - 2.5. The random term v_t can play an important part in these Harrison-Stevens models; the assumption that v_t is absent is restrictive in practice. However, if $\underline{V} \neq \underline{0}$ then it can still be shown that these models have the same forecast functions as a subclass of non-stationary time series models. This is considered in greater depth in Chapter 6.

In much of the literature it is stressed that the matrices $\underline{F}$, $\underline{G}$, $\underline{V}$ & $\underline{W}$ need to be specified with care. As suggested earlier, the coefficient matrices $\underline{F}$ & $\underline{G}$ are often known from the physical situation, but the specification and updating of $\underline{V}$ & $\underline{W}$ is more difficult, and would usually be carried out with confidence by practitioners or in consultation with them. In applications it will be necessary to specify $\underline{V}$ and the elements of the positive semi-definite matrix $\underline{W}$. The situation is eased if $\underline{W}$ is first diagonalised by an appropriate non-singular transformation $\underline{L}\underline{W}\underline{L}'$; however this may restrict the model so that the moving average parameters of the equivalent time series model no longer cover the entire stability region. The assumptions that are made about

$\underline{V}$ and $\underline{W}$ are equivalent to making assumptions about the moving average parameters. From a practical standpoint movements in the variances are more meaningful than the corresponding changes in the parameters. However, at present it appears to be much easier to test assumptions about the unknown parameters than to check the assumed values of the variances.

The Harrison-Stevens model is considered further in Chapter 6, where we place constraints on $\underline{F}$ and $\underline{G}$ so that comparisons can be drawn between these models and specific non-stationary time series. Data from such models can be rendered stationary by differencing a fixed number of times and then inference techniques associated with stationary time series are appropriate.

CHAPTER 3

THE PURE AUTOREGRESSIVE MODEL

3.1 Introduction

In this chapter the covariance matrix for the efficient estimators of the parameters of an autoregressive model is considered. The model is defined by

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \varepsilon_t \qquad (3.1.1)$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically **distributed Gaussian** random variables with expectation zero and a common variance σ^2.

The pure model (3.1.1) merits consideration in its own right, not only for reasons of its simplicity. The model has been discussed in the literature for many years and appears to be widely used in practice. Given a realisation of n consecutive observations, and that $\{\varepsilon_t\}$ is Gaussian, Whittle (1953) has shown that the maximum likelihood approach yields a consistent estimator of the vector $(\alpha_1 \ldots \alpha_p)'$ which is asymptotically normal. The covariance matrix $\underline{V}/n$ of this limiting distribution is smallest in the sense that $\underline{V}^* - \underline{V}$ is positive semi-definite when $\underline{V}^*/n$ is the covariance matrix for any alternative consistent estimator. Whittle has also shown that if $\{\varepsilon_t\}$ is non-Gaussian, then the least squares estimator has similar optimal properties when n is large.

Several interesting properties of $\underline{V}$ are derived by Whittle. In particular, he gives a formula involving complex integrals for the elements of the information matrix. The formula is, however, rather awkward to use in practice. Durbin (1959) derived a method for evaluating the covariance matrix of the efficient estimators of $(\alpha_1,\ldots,\alpha_p)$, by noting that the information matrix and the covariance matrix for p consecutive observations of the process (3.1.1) are identical, apart from a multiplicative constant. Box & Jenkins (1970, §A7.5) suggest another approach to the problem of evaluating $\underline{V}/n$. Their method involves treating the log likelihood as an approximately quadratic function. Provided the maximum is not close to a boundary, then the estimates of the elements of $\underline{V}$ are reasonable, even if the sample size is only moderate. However this method is rather complicated in practice, and its accuracy as the number of parameters increases is uncertain. Pagano (1973) suggested without proof that Durbin's result could be expressed in terms of the difference of two products of triangular matrices, whose non-zero elements are the parameters of the model.

A proof of this expression is given in Section 3.3. The generalised variance is also considered in this chapter. The two triangular matrices feature again in a factorization of the determinant of $\underline{V}$, this being an integral part of the generalised variance.

Analogous results exist on the whole for purely moving average models; for the sake of completeness, the model is considered separately in Section 3.6. Many of the ideas presented in this chapter are used or adapted later, when more complicated models are discussed.

3.2 The Information Matrix

In general $n\underline{F}$ will denote the information matrix, but in this chapter the notation $n\underline{F}_{\alpha\alpha}$ will be used. The (i,j)-th element of $\underline{F}_{\alpha\alpha}$ given by Whittle (1953) is the constant term in the expansion of

$$\frac{1}{2}\frac{\partial}{\partial\alpha_i}\log\{B(z)B(z^{-1})\}.\frac{\partial}{\partial\alpha_j}\log\{B(z)B(z^{-1})\}. \qquad (3.2.1)$$

This is equivalent to

Lemma 3.2.1

The (i,j)-th element of $\underline{F}_{\alpha\alpha}$ is the coefficient of z^{j-i} in the expansion of

$$\frac{1}{\alpha(z)\alpha(z^{-1})}$$

where $\alpha(z) = 1 + \alpha_1 z + \ldots + \alpha_p z^p$.

Proof

Since $B(z) = \{\alpha(z)\}^{-1}$,

$$\frac{\partial}{\partial\alpha_i}\log B(z)B(z^{-1}) = \frac{\partial}{\partial\alpha_i}\{-\log\alpha(z) - \log\alpha(z^{-1})\}$$

$$= \frac{-z^i}{\alpha(z)} - \frac{z^{-i}}{\alpha(z^{-1})} \quad . \quad (i=1,\ldots,p)$$

Therefore, expression (3.2.1) becomes

$$\frac{1}{2}\left[\frac{-z^i}{\alpha(z)} - \frac{z^{-i}}{\alpha(z^{-1})}\right]\left[\frac{-z^j}{\alpha(z)} - \frac{z^{-j}}{\alpha(z^{-1})}\right]$$

$$= \frac{1}{2}\left[\frac{z^{i+j}}{\alpha^2(z)} + \frac{z^{-(i+j)}}{\alpha^2(z^{-1})} + \frac{z^{i-j}+z^{j-i}}{\alpha(z)\alpha(z^{-1})}\right].$$

There are no constant terms in $z^{i+j}/\alpha^2(z)$ or $z^{-(i+j)}/\alpha^2(z^{-1})$ and the constant terms in $z^{i-j}/\{\alpha(z)\alpha(z^{-1})\}$ and $z^{j-i}/\{\alpha(z)\alpha(z^{-1})\}$ are the same owing to symmetry in the denominator. Hence the constant term in (3.2.1) is $z^{i-j}/\{\alpha(z)\alpha(z^{-1})\}$ and this is equivalent to the coefficient of z^{j-i} in $\{\alpha(z)\alpha(z^{-1})\}^{-1}$ as required.

By letting f_k denote the coefficient of z^k in $\{\alpha(z)\alpha(z^{-1})\}^{-1}$ so that $f_k = f_{-k}$ (k=1,2,...) by symmetry, it follows that

$$f_0 + \sum_{k=1}^{\infty} f_k(z^k + z^{-k}) = \frac{1}{\alpha(z)\alpha(z^{-1})} . \tag{3.2.2}$$

Hence the information matrix is given by n times the matrix $\underline{F}_{\alpha\alpha}$ where

$$\underline{F}_{\alpha\alpha} = \begin{bmatrix} f_0 & f_1 \cdot\cdot\cdot & f_{p-1} \\ f_1 & f_0 \cdot\cdot\cdot & f_{p-2} \\ \vdots & \vdots & \vdots \\ f_{p-1} & f_{p-2} \cdot\cdot & f_0 \end{bmatrix} .$$

To illustrate this result, the following example applies the Lemma directly to a simple model.

<u>Example</u> p=2.

The model is

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t .$$

Expressing $1/\{\alpha(z)\alpha(z^{-1})\}$ in partial fractions as given by Quenouille's algorithm (1947a):

$$1/\{\alpha(z)\alpha(z^{-1}) = K_0 + (K_1 z + K_2 z^2)/\alpha(z) + (K_1 z^{-1} + K_2 z^{-2})/\alpha(z^{-1})$$

Multiplying throughout by the lowest common denominator $\alpha(z)\alpha(z^{-1})$ yields

$$1 = K_0\{(1+\alpha_1 z+\alpha_2 z^2)(1+\alpha_1 z^{-1}+\alpha_2 z^{-2})\}$$
$$+(K_1 z+K_2 z^2)(1+\alpha_1 z^{-1}+\alpha_2 z^{-2}) + (K_1 z^{-1}+K_2 z^{-2})(1+\alpha_1 z+\alpha_2 z^2).$$

Equating coefficients of z^0, $z^{\pm 1}$, $z^{\pm 2}$,

$$1 = K_0(1+\alpha_1{}^2+\alpha_2{}^2) + 2K_1\alpha_1 + 2K_2\alpha_2$$
$$0 = K_0(\alpha_1+\alpha_1\alpha_2) + K_1(1+\alpha_2) + K_2\alpha_1$$
$$0 = K_0\alpha_2 + K_2 .$$

Solving for K_o gives

$$K_o = \frac{1+\alpha_2}{(1-\alpha_2)\{(1+\alpha_2)^2 - \alpha_1{}^2\}} \quad \text{and} \quad K_1 = \frac{-\alpha_1}{(1-\alpha_2)\{(1+\alpha_2)^2 - \alpha_1{}^2\}} .$$

In this case $f_o = K_o$ and $f_1 = K_1$, hence

$$n\underline{F}_{\alpha\alpha} = \frac{n}{(1-\alpha_2)\{(1+\alpha_2)^2 - \alpha_1{}^2\}} \begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix} .$$

When p is moderately large, it is preferable to apply a different approach to evaluate $\underline{F}_{\alpha\alpha}$. This will be described in the next section. The matrix $\underline{F}_{\alpha\alpha}$ is required in several hypothesis tests, details of which are given in Section 2.5.

3.3 The Covariance Matrix

Let $\underline{\alpha} = (\alpha_1 \dots \alpha_p)'$ denote the vector of parameters for the autoregressive model of order p with $\alpha_p \neq 0$

$$X_t + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} = \varepsilon_t . \qquad (3.3.1)$$

In the previous section a method was given for evaluating $\underline{F}_{\alpha\alpha}$. The covariance matrix for the efficient estimator of $\underline{\alpha}$ is the inverse of $n\underline{F}_{\alpha\alpha}$. However, it turns out to be unnecessary to find the individual elements of $\underline{F}_{\alpha\alpha}$ and invert this matrix. This was noted by Durbin (1959) who produced a simple method for obtaining $\underline{F}_{\alpha\alpha}^{-1}$ without the need for any awkward matrix inversions. His result is based on the fact that, apart from a multiplicative constant, $\underline{F}_{\alpha\alpha}$ is the covariance matrix for p consecutive observations of the process (3.3.1). By letting $\underline{X}_1 = (X_1 \ldots X_p)$, $\underline{X}_2 = (X_{p+1} \ldots X_{2p})$ denote $2p$ consecutive observations, he expressed the unconditional distribution of $(\underline{X}_1, \underline{X}_2)$ in two different ways. Equating the first p rows and columns of the matrices in the resultant quadratic forms, this yielded $\underline{F}_{\alpha\alpha}^{-1}$ directly.

Pagano (1973) suggested a neat expression for Durbin's formulation in terms of the difference of two products of triangular matrices. But although this formula has been quoted by other workers, a rigorous proof does not appear to have been given in the literature. In fact, the result follows from the commutative properties of upper triangular matrices and a proof of the result is given in Theorem 3.3.1 below.

Let $\underline{A}$ and $\underline{B}$ be upper triangular matrices of order $p \times p$ defined by

$$\underline{A} = \begin{bmatrix} 1 & \alpha_1 & \ldots & \alpha_{p-1} \\ 0 & 1 & \ldots & \alpha_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}, \quad \underline{B} = \begin{bmatrix} \alpha_p & \alpha_{p-1} & \ldots & \alpha_1 \\ 0 & \alpha_p & \ldots & \alpha_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \alpha_p \end{bmatrix}$$

Both $\underline{A}$ and $\underline{B}$ are non-singular and are symmetric about the minor diagonal. A useful property of $\underline{A}$ and $\underline{B}$ is provided by the following lemma:

Lemma The matrices $\underline{A}$ and $\underline{B}$ commute.

To show that $\underline{AB}=\underline{BA}$ we have

$$\underline{AB} = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-1} \\ 0 & 1 & \cdots & \alpha_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_p & \alpha_{p-1} & \cdots & \alpha_1 \\ 0 & \alpha_p & \cdots & \alpha_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \alpha_p \end{bmatrix} = \begin{bmatrix} \gamma_p & \gamma_{p-1} & \cdots & \gamma_1 \\ 0 & \gamma_p & \cdots & \gamma_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \gamma_p \end{bmatrix} \quad (3.3.2)$$

where $\gamma_k = \alpha_k + \alpha_1\alpha_{k+1} + \ldots + \alpha_{p-k}\alpha_p$ $(1 \leqq k \leqq p)$. Similarly $\underline{BA}$ is the right hand side of (3.3.2) and so the lemma follows.

Theorem 3.3.1 (Durbin, 1959)

$$\underline{F}_{\alpha\alpha}^{-1} = \underline{A}'\underline{A} - \underline{B}'\underline{B} .$$

Proof

Define

$$\underline{F}^* = \begin{bmatrix} f_p & f_{p+1} & \cdots & f_{2p-1} \\ f_{p-1} & f_p & \cdots & f_{2p-2} \\ \vdots & \vdots & & \vdots \\ f_1 & f_2 & \cdots & f_p \end{bmatrix}$$

where the elements of $\underline{F}^*$ and $\underline{F}_{\alpha\alpha}$ are defined by the relation (3.2.2). Equation (3.2.2) can be re-written as

$$\alpha(z^{-1}) \{f_0 + \sum_{k=1}^{\infty} f_k(z^k + z^{-k})\} = \frac{1}{\alpha(z)} = 1 + \sum_{k=1}^{\infty} a_k z^k \text{ say.}$$

Equating constants, and positive and negative powers of z respectively yields equations (3.3.3). Note that the last two equations require positive values of k in applications, and recall that $f_h = f_{-h}$.

$$\left.\begin{array}{lcl} f_0 + \alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_p f_p & = & 1 \\ f_k + \alpha_1 f_{k+1} + \alpha_2 f_{k+2} + \ldots + \alpha_p f_{k+p} & = & a_k \\ f_k + \alpha_1 f_{k-1} + \alpha_2 f_{k-2} + \ldots + \alpha_p f_{k-p} & = & 0 \end{array}\right\} \quad (3.3.3)$$

The last of equations (3.3.3) implies that

$$\begin{bmatrix} \underset{-}{F}_{\alpha\alpha} & \underset{-}{F}^{*} \end{bmatrix} \begin{bmatrix} \underset{-}{B}' \\ \underset{-}{A} \end{bmatrix} = \underset{-}{0}_p .$$

Hence

$$\underset{-}{F}^{*} = - \underset{-}{F}_{\alpha\alpha}\underset{-}{B}'\underset{-}{A}^{-1} . \tag{3.3.4}$$

The a's are given by the relation

$$\frac{1}{\alpha(z)} = 1 + \sum_{k=1}^{\infty} a_k z^k$$

which can be re-written as

$$a_k + \alpha_1 a_{k-1} + \cdot\cdot\cdot + \alpha_{k-1}a_1 + \alpha_k = 0 \qquad (1 \leqq k \leqq p)$$

where $a_o = \alpha_o = 1$. This last set of equations verifies the fact that

$$\underset{-}{A}^{-1} = \begin{bmatrix} 1 & a_1 & \cdots & a_{p-1} \\ 0 & 1 & \cdots & a_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} .$$

Consequently, equations (3.3.3) imply that

$$\begin{bmatrix} \underset{-}{F}_{\alpha\alpha} & \underset{-}{F}^{*} \end{bmatrix} \begin{bmatrix} \underset{-}{A}' \\ \underset{-}{B} \end{bmatrix} = \underset{-}{A}^{-1}$$

i.e.

$$\underset{-}{F}_{\alpha\alpha}\underset{-}{A}' + \underset{-}{F}^{*}\underset{-}{B} = \underset{-}{A}^{-1}. \tag{3.3.5}$$

Substituting for $\underset{-}{F}^{*}$ from (3.3.4) gives

$$\underset{-}{F}_{\alpha\alpha}(\underset{-}{A}' - \underset{-}{B}'\underset{-}{A}^{-1}\underset{-}{B}) = \underset{-}{A}^{-1}.$$

From the lemma, $\underset{-}{A}$ and $\underset{-}{B}$ commute; therefore $\underset{-}{A}^{-1}$ and $\underset{-}{B}$ commute. This implies

$$\underset{\sim}{F}_{\alpha\alpha}(\underset{\sim}{A}' - \underset{\sim}{B}'\underset{\sim}{B}\underset{\sim}{A}^{-1}) = \underset{\sim}{A}^{-1}$$

and post-multiplying by $\underset{\sim}{A}$

$$\underset{\sim}{F}_{\alpha\alpha}(\underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B}) = \underset{\sim}{I},$$

whence

$$\underset{\sim}{F}_{\alpha\alpha}^{-1} = \underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B}$$

and the proof is complete.

It is also worth noting that

<u>Corollary 3.3.2</u> $\quad \underset{\sim}{F}_{\alpha\alpha}^{-1} = \underset{\sim}{A}\underset{\sim}{A}' - \underset{\sim}{B}\underset{\sim}{B}'$

To show this, define a pxp matrix $\underset{\sim}{J}$ which is the "mirror image" of the identity matrix; i.e.

$$\underset{\sim}{J} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

The matrix $\underset{\sim}{J}$ is symmetric and non-singular. Also

$$\det \underset{\sim}{J} = (-1)^{p-1} \quad \text{and} \quad \underset{\sim}{J}^2 = \underset{\sim}{J}\underset{\sim}{J}' = \underset{\sim}{I} \ .$$

Pre- or post-multiplying $\underset{\sim}{A}$ or $\underset{\sim}{B}$ by $\underset{\sim}{J}$ produces some interesting results, as summarized below:-

The matrices $\underset{\sim}{A}\underset{\sim}{J}$ and $\underset{\sim}{J}\underset{\sim}{A}$ are symmetric since

$$\underset{\sim}{A}\underset{\sim}{J} = \begin{bmatrix} \alpha_{p-1} & \cdots & \alpha_1 & 1 \\ \alpha_{p-2} & \cdots & 1 & 0 \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}, \quad \underset{\sim}{J}\underset{\sim}{A} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & \alpha_{p-2} \\ 1 & \alpha_1 & \cdots & \alpha_{p-1} \end{bmatrix}$$

Hence $\underset{\sim}{A} = \underset{\sim}{J}\underset{\sim}{A}'\underset{\sim}{J}$ and $\underset{\sim}{A}' = \underset{\sim}{J}\underset{\sim}{A}\underset{\sim}{J}$. Thus pre-multiplying $\underset{\sim}{A}$ by $\underset{\sim}{J}$

reverses the order of the rows in $\underset{\sim}{A}$, and post-multiplying reverses the order of the columns. The two operations together produce a rotation of $\underset{\sim}{A}$ through 180^{0}. An analogous set of results exists for $\underset{\sim}{B}$. Since $\underset{\sim}{F}_{\alpha\alpha}^{-1}$ is symmetric about both diagonals, it is unaffected by a rotation through 180^{0}. Consequently Corollary 3.3.2 follows from

$$\begin{aligned} \underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B} &= \underset{\sim}{J}(\underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B})\underset{\sim}{J} \\ &= (\underset{\sim}{J}\underset{\sim}{A}')\underset{\sim}{A}\underset{\sim}{J} - (\underset{\sim}{J}\underset{\sim}{B}')\underset{\sim}{B}\underset{\sim}{J} \\ &= \underset{\sim}{A}(\underset{\sim}{J}\underset{\sim}{A}\underset{\sim}{J}) - \underset{\sim}{B}(\underset{\sim}{J}\underset{\sim}{B}\underset{\sim}{J}) \\ &= \underset{\sim}{A}\underset{\sim}{A}' - \underset{\sim}{B}\underset{\sim}{B}', \end{aligned}$$

as required.

To illustrate Theorem 3.3.1, the case $p=2$ is considered.

Example

The model is

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t$$

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \alpha_1 \\ 0 & 0 \end{bmatrix}, \quad \underset{\sim}{B} = \begin{bmatrix} \alpha_2 & \alpha_1 \\ 0 & \alpha_2 \end{bmatrix}$$

It follows immediately that

$$\frac{1}{n}\underset{\sim}{F}_{\alpha\alpha}^{-1} = \frac{1}{n}(\underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B}) = \frac{1}{n}\begin{bmatrix} 1-\alpha_2^{\,2} & \alpha_1(1-\alpha_2) \\ \alpha_1(1-\alpha_2) & 1-\alpha_2^{\,2} \end{bmatrix}.$$

Hence if $\hat{\alpha}_1$, $\hat{\alpha}_2$ denote the efficient estimators of α_1, α_2 respectively, then for suitably large n

$$\mathrm{Var}\,(\hat{\alpha}_1) = \mathrm{Var}\,(\hat{\alpha}_2) = \frac{1}{n}(1-\alpha_2^{\,2})$$

and

$$\mathrm{Cov}\,(\hat{\alpha}_1\hat{\alpha}_2) = \frac{1}{n}\alpha_1(1-\alpha_2).$$

Inverting $\underset{\sim}{F}_{\alpha\alpha}^{-1}/n$ gives the information matrix $n\underset{\sim}{F}_{\alpha\alpha}$ as found in the previous section using Quenouille's algorithm. Indeed, if p is moderately large, and if the information matrix only is required, it seems to be quicker to evaluate $\underset{\sim}{F}_{\alpha\alpha}^{-1}$ using Theorem 3.3.1 and then to invert $\underset{\sim}{F}_{\alpha\alpha}^{-1}/n$.

3.4 The Generalised Variance

The Generalised Variance (G.V) for the efficient estimator $\hat{\underset{\sim}{\alpha}}$ of $\underset{\sim}{\alpha}$ is defined to be the determinant of the covariance matrix for this estimator. That is,

$$\begin{aligned} \text{G.V.} &= \det\left(\frac{\underset{\sim}{F}_{\alpha\alpha}^{-1}}{n}\right) \\ &= \frac{1}{n^p} \det \underset{\sim}{F}_{\alpha\alpha}^{-1} . \end{aligned}$$

The aim of this section is to find an easy way to evaluate this determinant. Theorem 3.3.1 will prove useful in doing this. With the matrices $\underset{\sim}{A}$, $\underset{\sim}{B}$ and $\underset{\sim}{J}$ defined as in the previous section, $\det \underset{\sim}{F}_{\alpha\alpha}^{-1}$ factorizes as follows:

Theorem 3.4.1

$$\det \underset{\sim}{F}_{\alpha\alpha}^{-1} = \det(\underset{\sim}{A} - \underset{\sim}{J}\underset{\sim}{B}) \det(\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B}) .$$

Proof

Since $\underset{\sim}{A}$ and $\underset{\sim}{B}$ commute, it follows that

$$\underset{\sim}{B}'\underset{\sim}{J}\underset{\sim}{A} = \underset{\sim}{J}\underset{\sim}{B}\underset{\sim}{A} = \underset{\sim}{J}\underset{\sim}{A}\underset{\sim}{B} = \underset{\sim}{A}'\underset{\sim}{J}\underset{\sim}{B} .$$

So

$$(\underset{\sim}{A}' - \underset{\sim}{B}'\underset{\sim}{J})(\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B}) = \underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B} = \underset{\sim}{F}_{\alpha\alpha}^{-1} .$$

Hence

$$\det \underset{\sim}{F}_{\alpha\alpha}^{-1} = \det (\underset{\sim}{A}' - \underset{\sim}{B}'\underset{\sim}{J})(\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B})$$
$$= \det (\underset{\sim}{A}' - \underset{\sim}{B}'\underset{\sim}{J}) \det (\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B})$$
$$= \det (\underset{\sim}{A} - \underset{\sim}{J}\underset{\sim}{B}) \det (\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B}) \ .$$

By appealing to another result derived in a control theory context by Jury (1964, pp.87) it is possible to factorize this determinant still further. Firstly, define the matrix $\underset{\sim}{H}$, where

$$\underset{\sim}{H} = \underset{\sim}{A}^* - \underset{\sim}{J}\underset{\sim}{B}^*$$

and $\underset{\sim}{A}^*$, $\underset{\sim}{B}^*$ are derived from $\underset{\sim}{A}$ and $\underset{\sim}{B}$ respectively by deleting the p-th row and p-th column. Thus

$$\underset{\sim}{H} = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-2} \\ 0 & 1 & \cdots & \alpha_{p-3} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} - \underset{\sim}{J} \begin{bmatrix} \alpha_p & \alpha_{p-1} & \cdots & \alpha_2 \\ 0 & \alpha_p & \cdots & \alpha_3 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \alpha_p \end{bmatrix} \qquad (3.4.1)$$

$$= \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-2}-\alpha_p \\ 0 & 1 & \cdots & \alpha_{p-3}-\alpha_{p-1} \\ \vdots & \vdots & & \vdots \\ -\alpha_p & -\alpha_{p-1} & \cdots & 1-\alpha_2 \end{bmatrix} .$$

Then

<u>Theorem 3.4.2</u> (Jury, 1964)

$$\det (\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B}) = (1 + \alpha_1 + \alpha_2 + \cdots + \alpha_p) \det \underset{\sim}{H}$$
$$= \alpha(1) \det \underset{\sim}{H}$$
$$\det (\underset{\sim}{A} - \underset{\sim}{J}\underset{\sim}{B}) = (1 - \alpha_1 + \alpha_2 - \cdots + (-1)^p\alpha_p) \det \underset{\sim}{H}$$
$$= \alpha(-1) \det \underset{\sim}{H} \ .$$

By combining the results in these two theorems, the factorization of the determinant becomes

$$\det \underset{\sim}{F}_{\alpha\alpha}^{-1} = \alpha(1)\alpha(-1)(\det \underset{\sim}{H})^2 \ .$$

Finally, the generalised variance is given by

$$\text{G.V.} = \frac{1}{n^p}\,\alpha(1)\alpha(-1)(\det \underset{\sim}{H})^2 . \qquad (3.4.2)$$

Example $p = 3$

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 \\ 0 & 1 & \alpha_1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{so } \underset{\sim}{A}^* = \begin{bmatrix} 1 & \alpha_1 \\ 0 & 1 \end{bmatrix} .$$

$$\underset{\sim}{B} = \begin{bmatrix} \alpha_3 & \alpha_2 & \alpha_1 \\ 0 & \alpha_3 & \alpha_2 \\ 0 & 0 & \alpha_3 \end{bmatrix} \quad \text{so } \underset{\sim}{B}^* = \begin{bmatrix} \alpha_3 & \alpha_2 \\ 0 & \alpha_3 \end{bmatrix} .$$

$$\underset{\sim}{H} = \underset{\sim}{A}^* - \underset{\sim}{J}\underset{\sim}{B}^*$$

$$= \begin{bmatrix} 1 & \alpha_1 - \alpha_3 \\ -\alpha_3 & 1 - \alpha_2 \end{bmatrix} ,$$

so that

$$\det \underset{\sim}{H} = 1 - \alpha_2 + \alpha_3(\alpha_1 - \alpha_3) .$$

Thus using equation (3.4.2),

$$\text{G.V.} = (1+\alpha_1+\alpha_2+\alpha_3)(1-\alpha_1+\alpha_2-\alpha_3)\{1-\alpha_2 + \alpha_3(\alpha_1-\alpha_3)\}^2/n^3$$

$$= \{(1+\alpha_2)^2 - (\alpha_1+\alpha_3)^2\}\{1-\alpha_2 + \alpha_3(\alpha_1-\alpha_3)\}^2/n^3 \qquad (3.4.3)$$

The simplicity which results from the factorization (3.4.2) is made evident by considering the problem directly. In the case $p = 3$,

$$\frac{1}{n}\underset{\sim}{F}_{\alpha\alpha}^{-1} = \frac{1}{n}\begin{bmatrix} 1-\alpha_3^2 & \alpha_1 - \alpha_2\alpha_3 & \alpha_2 - \alpha_1\alpha_3 \\ \alpha_1 - \alpha_2\alpha_3 & 1+\alpha_1^2-\alpha_2^2-\alpha_3^2 & \alpha_1 - \alpha_2\alpha_3 \\ \alpha_2 - \alpha_1\alpha_3 & \alpha_1 - \alpha_2\alpha_3 & 1 - \alpha_3^2 \end{bmatrix}$$

It is not immediately obvious that the determinant of this matrix simplifies to the form (3.4.3), hence the factorization (3.4.2) clearly avoids some tedious algebraic manipulations. A simplified form for the stationarity conditions can also be deduced using (3.4.3). For this example the conditions are

$$\left.\begin{aligned} 1-\alpha_3^2 &> 0 \\ \det\begin{bmatrix} 1-\alpha_3^2 & \alpha_1-\alpha_2\alpha_3 \\ \alpha_1-\alpha_2\alpha_3 & 1+\alpha_1^2-\alpha_2^2-\alpha_3^2 \end{bmatrix} &> 0 \\ \text{and} \qquad \det \underset{-\alpha\alpha}{F}^{-1} &> 0. \end{aligned}\right\} \tag{3.4.4}$$

In their present form these inequalities involve fourth degree terms in the α's, but we now show how (3.4.4) can be simplified to

$$\left.\begin{aligned} 1+\alpha_1+\alpha_2+\alpha_3 &> 0 \\ 1-\alpha_1+\alpha_2-\alpha_3 &> 0 \\ 1-\alpha_2+\alpha_3(\alpha_1-\alpha_3) &> 0 \\ |\alpha_3| &< 1. \end{aligned}\right\} \tag{3.4.5}$$

Using the factorization (3.4.2) and after some algebraic manipulations, the inequalities (3.4.4) can be re-written as

$$1-\alpha_3^2 > 0$$

$$\{1+\alpha_2-\alpha_3(\alpha_1+\alpha_3)\}\{1-\alpha_2+\alpha_3(\alpha_1-\alpha_3)\} > 0 \tag{3.4.6}$$

$$(1+\alpha_1+\alpha_2+\alpha_3)(1-\alpha_1+\alpha_2-\alpha_3)\{1-\alpha_2+\alpha_3(\alpha_1-\alpha_3)\}^2 > 0 \tag{3.4.7}$$

If both brackets in (3.4.6) are negative then addition of these yields

$$2(1 - \alpha_3^2) < 0$$

which contradicts $|\alpha_3| < 1$. Thus both brackets are strictly positive and consequently

$$\{1 - \alpha_2 + \alpha_3(\alpha_1 - \alpha_3)\}^2 \nleq 0.$$

Hence the squared term can be omitted from (3.4.7) to give

$$(1+\alpha_1+\alpha_2+\alpha_3)(1-\alpha_1+\alpha_2-\alpha_3) > 0.$$

If both brackets are negative then

$$-\alpha_1 - \alpha_2 > 1 + \alpha_3 \quad \text{and} \quad \alpha_1 - \alpha_2 > 1 - \alpha_3$$

and hence

$$-(\alpha_1^2 - \alpha_2^2) > 1 - \alpha_3^2.$$

Also $1+\alpha_1^2-\alpha_2^2-\alpha_3^2$ is positive since it is a variance, thus both brackets are positive and the inequalities reduce to (3.4.5) together with

$$1 + \alpha_2 - \alpha_3(\alpha_1 + \alpha_3) > 0.$$

But this additional inequality is redundant since $1+\alpha_3>0$ & $1-\alpha_3>0$ gives

$$(1-\alpha_3)(1+\alpha_1+\alpha_2+\alpha_3) + (1+\alpha_3)(1-\alpha_1+\alpha_2-\alpha_3) > 0$$

or

$$2(1+\alpha_2 - \alpha_3(\alpha_1+\alpha_3)) > 0.$$

Hence the inequalities (3.4.4) simplify to (3.4.5).

It is interesting to note that for the AR(3) model, the stationarity conditions are

$$\alpha(1) > 0, \quad \alpha(-1) > 0, \quad \det \underline{H} > 0 \quad \text{and} \quad |\alpha_3| < 1.$$

These bear a remarkable resemblance to those of the AR(2) process, namely

$$\alpha(1) > 0, \quad \alpha(-1) > 0 \quad \text{and} \quad |\alpha_2| < 1.$$

In this case, det $\underline{H} = 1-\alpha_2 > 0$ is a redundant constraint. The stationarity conditions for models containing more parameters are necessarily more complicated, but can always be simplified to two first degree inequalities involving $\alpha(1)$ & $\alpha(-1)$, together with $|\alpha_p| < 1$ and at most $1+(p/2)$ or $1+(p+1)/2$ further constraints, depending on whether p is odd or even. A discussion given in a control theory context is that of Jury (1964, §3.5).

3.5 Derivation of the Moments of an Autoregressive Process

An interesting duality result exists between $\underline{F}_{\alpha\alpha}$, where $\underline{F}_{\alpha\alpha}^{-1}/n$ is the covariance matrix for the efficient estimator of $\underline{\alpha}$, and the pxp covariance matrix for p consecutive observations of the AR(p) process (3.1.1). The result was first noted by Siddiqui (1958). He observed that

$$\underline{F}_{\alpha\alpha} = \underline{\Gamma}_p$$

where $\sigma^2\underline{\Gamma}_p$ is the covariance matrix for the realization $X_1,\ldots,X_p$ of the process (3.1.1). This result holds because the autocovariance generating function $\Gamma(z)$ is defined by

$$\Gamma(z) = \sigma^2/\{A(z)A(z^{-1})\}$$

yielding

$$\frac{\Gamma(z)}{\sigma^2} = \frac{1}{A(z)A(z^{-1})} = \frac{1}{\alpha(z)\alpha(z^{-1})} = f_0 + \sum_{k=1}^{\infty} f_k(z^k+z^{-k}).$$

Hence the elements of $\Gamma(z)/\sigma^2$ and $\underline{F}_{\alpha\alpha}$ are given by the same formula, namely $\{\alpha(z)\alpha(z^{-1})\}^{-1}$, and so the matrices $\underline{\Gamma}_p$ and $\underline{F}_{\alpha\alpha}$

are identical.

There are several methods that can be employed for finding explicitly the individual elements of $\underline{\Gamma}_p$. One technique is to solve the Yule-Walker and Wold equations using equation (2.2.3); this method does not invoke Siddiqui's result, and may be tedious in practice. Alternatively, $\underline{F}_{\alpha\alpha}$ can be evaluated using Quenouille's algorithm as in Section 3.2. This method is also rather long, and it was concluded in Section 3.3 that even if $\underline{F}_{\alpha\alpha}$ is required, the quickest method is to first find $\underline{F}_{\alpha\alpha}^{-1}$ using Theorem 3.3.1 and then to invert this matrix.

The covariance matrix $\sigma^2\underline{\Gamma}_n$ for a realization $X_1,\ldots,X_n$ of size n of the process (3.1.1) can be determined relatively easily by expanding $\underline{\Gamma}_p$. To see this, consider again the example in Section 3.3. For the Yule process, the covariance matrix for the efficient estimator of α_1, α_2 was found to be

$$\frac{1}{n}\underline{F}_{\alpha\alpha}^{-1} = \frac{1}{n}\begin{bmatrix} 1-\alpha_2^{\,2} & \alpha_1-\alpha_1\alpha_2 \\ \alpha_1-\alpha_1\alpha_2 & 1-\alpha_2^{\,2} \end{bmatrix}.$$

Hence the covariance matrix for two consecutive observations of the same AR(2) process is $\sigma^2\underline{\Gamma}_2$ where

$$\underline{\Gamma}_2 = \underline{F}_{\alpha\alpha} = \frac{1}{(1-\alpha_2)\{(1+\alpha_2)^2-\alpha_1^{\,2}\}}\begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix} \quad \text{say.}$$

The Yule-Walker equations for this model are

$$\gamma_k + \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} = 0 \quad (k \geqq 1) \qquad \gamma_k = \gamma_{-k},$$

so for any $k>0$, γ_k can be expressed solely in terms of γ_o and γ_1. Therefore let

$$\sigma^2 \underline{\Gamma}_n = \sigma^2 \begin{bmatrix} \gamma_o & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_o & \cdots & \gamma_{n-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_o \end{bmatrix}$$

denote the covariance matrix for the realization $X_1,\ldots,X_n$ of the AR(2) process. Using the above algorithm, all the elements of $\underline{\Gamma}_n$ can be found in terms of the previously calculated γ_o and γ_1.

Clearly, this procedure generalises for any value of p to give an algorithm for deriving the autocovariances $\{\sigma^2\gamma_k\}$ of a general autoregressive process. Firstly $\underline{F}_{\alpha\alpha}^{-1}$ is found using Theorem 3.3.1, and then its inverse to give $\underline{\Gamma}_p$. This, together with σ^2, gives explicit values for the variance and first (p-1) covariances. The remaining moments $\gamma_p,\ldots,\gamma_{n-1}$ can then be derived from the Yule-Walker equations in terms of $\gamma_o,\gamma_1,\ldots,\gamma_{p-1}$.

3.6 The Pure Moving Average Model

The results presented so far in this chapter refer only to the pure autoregressive model of order p. However all of the results in Sections 3.2, 3.3 and 3.4 are applicable to moving average models as well. To clarify the situation the results will be discussed briefly for the moving average model of order q defined by

$$X_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \cdots + \beta_q\varepsilon_{t-q} \; . \qquad (3.6.1)$$

The basis of the fact that analogous results exist for a moving average model is a result due to Whittle (1953). He observed that the covariance matrix for the efficient estimators of the parameters of an autoregressive model is the same as that of a moving average model, provided the parameter sets are the same. Let $n\underset{\sim}{F}_{\beta\beta}$ denote the $q \times q$ information matrix for the process (3.6.1). The (i,j)-th element of $\underset{\sim}{F}_{\beta\beta}$ is the constant term in the expansion of

$$\frac{1}{2}\frac{\partial}{\partial\beta_i}\log\{B(z)B(z^{-1})\}.\frac{\partial}{\partial\beta_j}\log\{B(z)B(z^{-1})\}$$

which is equivalent to the coefficient of z^{j-i} in

$$\frac{1}{\beta(z)\beta(z^{-1})}$$

where $\beta(z) = 1 + \beta_1 z + \ldots + \beta_q z^q$.
Defining

$$\underset{\sim}{F}_{\beta\beta} = \begin{bmatrix} h_o & h_1 & \cdots & h_{q-1} \\ h_1 & h_o & \cdots & h_{q-2} \\ \vdots & \vdots & & \vdots \\ h_{q-1} & h_{q-2} & \cdots & h_o \end{bmatrix}$$

it follows that

$$h_o + \sum_{k=1}^{\infty} h_k(z^k + z^{-k}) = \frac{1}{\beta(z)\beta(z^{-1})} .$$

Clearly if $\beta(z) = \alpha(z)$ then the h's have the same defining equations as the f's in Section 3.2, whence $\underset{\sim}{F}_{\beta\beta} = \underset{\sim}{F}_{\alpha\alpha}$. For the same reasons as before, this is not necessarily the easiest way to evaluate the information matrix if q is moderately large. In general it seems to be quicker to first find the covariance matrix for the efficient estimators of $(\beta_1,\ldots,\beta_q)$ and then take its inverse, even if the information matrix only is required.

It is necessary to re-define $\underset{\sim}{A}$ and $\underset{\sim}{B}$ as matrices of order $q \times q$ with elements consisting of parameters of the model:

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \beta_1 & \cdots & \beta_{q-1} \\ 0 & 1 & \cdots & \beta_{q-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad \underset{\sim}{B} = \begin{bmatrix} \beta_q & \beta_{q-1} & \cdots & \beta_1 \\ 0 & \beta_q & \cdots & \beta_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \beta_q \end{bmatrix}.$$

Then, as before, the covariance matrix for the efficient estimators of $(\beta_1,\ldots,\beta_q)$ is

$$\frac{1}{n} \underset{\sim}{F}^{-1}_{\beta\beta} = \frac{1}{n} (\underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B}) = \frac{1}{n} (\underset{\sim}{A}\underset{\sim}{A}' - \underset{\sim}{B}\underset{\sim}{B}') .$$

With these new definitions of $\underset{\sim}{A}$ and $\underset{\sim}{B}$ the formula for the determinant of $\underset{\sim}{F}^{-1}_{\beta\beta}$ remains as before:

$$\det \underset{\sim}{F}^{-1}_{\beta\beta} = \det (\underset{\sim}{A} - \underset{\sim}{J}\underset{\sim}{B}) \det (\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B}) .$$

To complete the simplification of the generalised variance

$$\begin{aligned} \det (\underset{\sim}{A} + \underset{\sim}{J}\underset{\sim}{B}) &= (1 + \beta_1 + \cdots + \beta_q) \det \underset{\sim}{H} \\ &= \beta(1) \det \underset{\sim}{H} \\ \det (\underset{\sim}{A} - \underset{\sim}{J}\underset{\sim}{B}) &= (1 - \beta_1 + \cdots + (-1)^q \beta_q) \det \underset{\sim}{H} \\ &= \beta(-1) \det \underset{\sim}{H} , \end{aligned}$$

where

$$\underset{\sim}{H} = \begin{bmatrix} 1 & \beta_1 & \cdots & \beta_{q-2}-\beta_q \\ 0 & 1 & \cdots & \beta_{q-3}-\beta_{q-1} \\ \vdots & \vdots & & \vdots \\ -\beta_q & -\beta_{q-1} & \cdots & 1-\beta_2 \end{bmatrix} .$$

That is $\underline{H} = \underline{A}^* - \underline{J}\underline{B}^*$ where $\underline{A}^*$, $\underline{B}^*$ are derived from $\underline{A}$ and $\underline{B}$ respectively by deleting the q-th row and the q-th column. Thus the generalised variance is given by

$$G.V = \det(\underline{F}_{\beta\beta}^{-1}/n)$$

$$= \frac{1}{n^q}\beta(1)\beta(-1)(\det \underline{H})^2 .$$

CHAPTER 4

THE ARMA (p,1) AND ARMA (1,q) MODELS

4.1 Introduction

In Chapter 5, the autoregressive moving average (ARMA) model of order (p,q) for general p and q will be considered in detail. In the present chapter, a subclass of mixed models which contain either just one moving average or just one autoregressive parameter will be examined. These two categories are defined by

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \varepsilon_t + \beta\varepsilon_{t-1} \qquad (4.1.1)$$

and

$$X_t + \alpha X_{t-1} = \varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q} \,. \qquad (4.1.2)$$

In the previous chapter, the purely autoregressive model was considered. The covariance matrix for the efficient estimators of the parameters of the model was easily specified and the information matrix was given simply by its inverse. The introduction to the model of just one moving average parameter makes the evaluation of these two matrices and also the generalised variance significantly more complicated. The information matrix for the model (4.1.1) can be written in the partitioned form

$$n\underline{F} = n\begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & F_{\beta\beta} \end{bmatrix} .$$

The terms $n\underline{F}_{\alpha\alpha}$ and $nF_{\beta\beta}$ are just the information matrices for the AR(p) and MA(1) models respectively and can be specified using the techniques described in the previous chapter. However, with mixed models, there is the additional complication in the submatrix $\underset{\sim}{F}_{\alpha\beta}$, or in this chapter the vector $\underset{\sim}{F}_{\alpha\beta}$. The form of the vector $\underset{\sim}{F}_{\alpha\beta}$ for the models (4.1.1) and (4.1.2) is derived in the next section.

Durbin's result (Theorem 3.3.1) still applies for finding the inverses of $\underset{\sim}{F}_{\alpha\alpha}$ and $F_{\beta\beta}$ but the result does not generalise for mixed models, even though there is just one extra parameter. The covariance matrix can only be found in block form by inverting $n\underset{\sim}{F}$; some simplification is afforded by the fact that the matrix $\underset{\sim}{F}$ as a whole is still symmetric.

The evaluation of the determinant of the covariance matrix, to give the generalised variance, would appear to be intractable in all except a few simple models. However a neat expression for factorising det $\underset{\sim}{F}$ has been produced, and a proof of the result is given in Section 4.3.

In the rest of this chapter, only model (4.1.1) is considered in detail. However, if the model (4.1.2) is re-defined using the same parameter set as (4.1.1), i.e.,

$$X_t + \beta X_{t-1} = \varepsilon_t + \alpha_1\varepsilon_{t-1} + \ldots + \alpha_p\varepsilon_{t-p} \qquad (4.1.2a)$$

then it is possible to appeal to a generalisation of Whittle's result (1953). Clearly the generalised variance is the same for models (4.1.1) and (4.1.2a); the information matrix becomes

$$n\underset{\sim}{F} = n\begin{bmatrix} F_{\beta\beta} & \underset{\sim}{F}'_{\alpha\beta} \\ \underset{\sim}{F}_{\alpha\beta} & \underset{\sim}{F}_{\alpha\alpha} \end{bmatrix}$$

with the blocks defined as previously, and a similar transformation gives the covariance matrix from that of the model (4.1.1). In the final section, the adequacy of the chosen model will be considered.

4.2 The Information Matrix

The information matrix is defined in a partitioned form by

$$n\underset{\sim}{F} = n\begin{bmatrix} \underset{\sim}{F}_{\alpha\alpha} & \underset{\sim}{F}_{\alpha\beta} \\ \underset{\sim}{F}'_{\alpha\beta} & F_{\beta\beta} \end{bmatrix} \qquad (4.2.1)$$

where $\underset{\sim}{F}_{\alpha\alpha}$ is a pxp symmetric matrix

$\underset{\sim}{F}_{\alpha\beta}$ is a column vector of length p

and $F_{\beta\beta}$ is a scalar term.

The matrix $\underset{\sim}{F}_{\alpha\alpha}$ and the scalar $F_{\beta\beta}$ are independent of β and $(\alpha_1,\ldots,\alpha_p)$ respectively. In fact $n\underset{\sim}{F}_{\alpha\alpha}$ is simply the information matrix for a pure autoregression of order p as derived in the previous chapter. The matrix $\underset{\sim}{F}_{\alpha\alpha}$ can be evaluated either using Lemma 3.2.1 or else by inverting $\underset{\sim}{F}^{-1}_{\alpha\alpha}$ given by applying Theorem 3.3.1. The scalar term $F_{\beta\beta}$ is just the constant term in the expansion of

$$\frac{1}{\beta(z)\beta(z^{-1})}$$

where $\beta(z) = 1 + \beta z$, i.e., $F_{\beta\beta} = (1-\beta^2)^{-1}$.

Thus the only difficulty in specifying the complete information matrix lies in the derivation of the vector $\underset{\sim}{F}_{\alpha\beta}$. Whittle's result in Section 3.2 can be used to find the elements of $\underset{\sim}{F}_{\alpha\beta}$. In this case, by adapting Lemma 3.2.1, the (i,j)-th element of $\underset{\sim}{F}_{\alpha\beta}$ is the coefficient of z^{j-i} in

$$\frac{-1}{\alpha(z)\beta(z^{-1})} \; .$$

By defining

$$\underset{\sim}{F}_{\alpha\beta} = \begin{bmatrix} g_0 \\ g_{-1} \\ \vdots \\ g_{1-p} \end{bmatrix}$$

it follows that

$$\sum_{k=-\infty}^{\infty} g_k z^k = \frac{-1}{\alpha(z)\beta(z^{-1})} .$$

By adapting a Quenouille-type algorithm, the following lemma shows that the evaluation of the elements of $\underset{\sim}{F}_{\alpha\beta}$ is straightforward in the particular case where $\underset{\sim}{F}_{\alpha\beta}$ is a vector.

Lemma 4.2.1

$$\underset{\sim}{F}_{\alpha\beta} = \frac{-1(\underset{\sim}{\beta})}{S}$$

where the multiple S is the scalar quantity

$$S = \begin{bmatrix} \underset{\sim}{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} \underset{\sim}{\alpha} \\ \alpha_p \end{bmatrix}$$

and $\underset{\sim}{\beta} = (1 \; -\beta \; (-\beta)^2 \; . \; . \; . \; (-\beta)^{p-1})'$, $\underset{\sim}{\alpha} = (1 \; \alpha_1 \; \alpha_2 \; . \; . \; . \; \alpha_{p-1})'$.

Proof

The elements of $\underset{\sim}{F}_{\alpha\beta}$ are given by the coefficients of $z^0, z^{-1}, ... , z^{1-p}$ in $-1/\{\alpha(z)\beta(z^{-1})\}$. Expanding into partial fractions,

$$\frac{-1}{\alpha(z)\beta(z^{-1})} = K_0 + \frac{K_1 z + K_2 z^2 + \ldots + K_p z^p}{\alpha(z)} + \frac{Lz^{-1}}{\beta(z^{-1})} .$$

Multiplying throughout by the lowest common denominator gives

$$-1 = K_0\alpha(z)\beta(z^{-1}) + (K_1 z + K_2 z^2 + . . . + K_p z^p)\beta(z^{-1}) + Lz^{-1}\alpha(z).$$

Equating coefficients of z^{-1} and z^p gives

$$L = -K_o\beta \quad \text{and} \quad K_p = -K_o\alpha_p.$$

Consider next the coefficient of z^{p-1} :

$$0 = K_o(\alpha_{p-1} + \alpha_p\beta) + (K_{p-1} + K_p\beta) + L\alpha_p.$$

Substituting for L and K_p gives

$$0 = K_o(\alpha_{p-1} - \alpha_p\beta) + K_{p-1}.$$

On substituting in the equation for z^{p-2} for L and K_{p-1} it follows that

$$K_o(\alpha_{p-2} - \alpha_{p-1}\beta + \alpha_{p-2}\beta^2) + K_{p-2} = 0$$

By considering the equations for the coefficients of $z^{p-2}, z^{p-3}, \ldots, z^o$ in that order, it is possible at each stage to eliminate one further term K, and also L. The final substitution for K_1 and L in the equation relating the constants yields

$$K_o(1-\alpha_1\beta+\alpha_2\beta^2-\alpha_3\beta^3 + \ldots + \alpha_p(-\beta)^p) = -1$$

whence

$$\begin{aligned} K_o &= -(1-\alpha_1\beta+\alpha_2\beta^2-\alpha_3\beta^3 + \ldots + \alpha_p(-\beta)^p)^{-1} \\ &= \frac{-1}{S} . \end{aligned} \qquad (4.2.2)$$

The only terms in the expansion of $-1/\{\alpha(z)\beta(z^{-1})\}$ involving negative powers of z occur in the partial fraction term

$$\frac{Lz^{-1}}{\beta(z^{-1})} = Lz^{-1}(1 - \beta z^{-1} + \beta^2 z^{-2} - \beta^3 z^{-3} + \ldots) .$$

Thus the coefficient of z^{-h} for positive h is $(-\beta)^{h-1}L = (-\beta)^h K_o$, where K_o is given by (4.2.2).

Finally,

$$\underline{F}_{\alpha\beta} = \begin{bmatrix} g_0 \\ g_{-1} \\ g_{-2} \\ \vdots \\ g_{1-p} \end{bmatrix} = K_0 \begin{bmatrix} 1 \\ -\beta \\ \beta^2 \\ \vdots \\ (-\beta)^{p-1} \end{bmatrix} = \frac{-\underline{1}(\underline{\beta})}{S} .$$

Using this result for $\underline{F}_{\alpha\beta}$, together with the given expressions for $\underline{F}_{\alpha\alpha}$ and $F_{\beta\beta}$, it is possible to write the information matrix as given explicitly for an ARMA(p,1) model by the partitioned form (4.2.1).

<u>Example</u> ARMA(2,1) model

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t + \beta\varepsilon_{t-1}.$$

The information matrix is

$$n\underline{F} = n \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & F_{\beta\beta} \end{bmatrix}$$

and since $q = 1$, $F_{\beta\beta} = (1-\beta^2)^{-1}$. It appears that the easiest way to evaluate $\underline{F}_{\alpha\alpha}$ is to find $\underline{F}_{\alpha\alpha}^{-1}$ and then take its inverse. From Theorem 3.3.1,

$$\underline{F}_{\alpha\alpha}^{-1} = \begin{bmatrix} 1 - \alpha_2^{\;2} & \alpha_1(1 - \alpha_2) \\ \alpha_1(1-\alpha_2) & 1 - \alpha_2^{\;2} \end{bmatrix}$$

so that

$$\underline{F}_{\alpha\alpha} = \frac{1}{(1-\alpha_2)\{(1+\alpha_2)^2 - \alpha_1^{\;2}\}} \begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix}.$$

Using Lemma 4.2.1,

$$\underline{F}_{\alpha\beta} = \frac{-\underline{1}(\underline{\beta})}{S}$$

where $\underline{\beta} = \begin{bmatrix} -1 \\ \beta \end{bmatrix}$ and $S = 1 - \alpha_1\beta + \alpha_2\beta^2$.

Thus $\underline{F}_{\alpha\beta} = \dfrac{-1}{(1-\alpha_1\beta+\alpha_2\beta^2)} \begin{bmatrix} 1 \\ -\beta \end{bmatrix}$

and the information matrix is

$$n\underline{F} = n \begin{bmatrix} \dfrac{1}{(1-\alpha_2)\{(1+\alpha_2)^2-\alpha_1{}^2\}} \begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix} & \dfrac{-1}{1-\alpha_1\beta+\alpha_2\beta^2} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} \\ \dfrac{-1}{1-\alpha_1\beta+\alpha_2\beta^2} \begin{bmatrix} 1 & -\beta \end{bmatrix} & \dfrac{1}{1-\beta^2} \end{bmatrix}.$$

To complete this section, the expression for the information matrix for the ARMA(1,q) model (4.1.2) is derived. It is given by

$$n \begin{bmatrix} F^0_{\alpha\alpha} & \underline{F}^0_{\alpha\beta} \\ (\underline{F}^0_{\alpha\beta})' & \underline{F}^0_{\beta\beta} \end{bmatrix}.$$

The matrix $n\underline{F}^0_{\beta\beta}$ is now the qxq information matrix for the pure moving average process of order q, and $F^0_{\alpha\alpha} = (1-\alpha^2)^{-1}$. Also $\underline{F}^0_{\alpha\beta}$ is a row vector of length q given by

$$\underline{F}^0_{\alpha\beta} = \frac{-1}{S^0} (1 \quad -\alpha \quad \alpha^2 \quad . \; . \; . \quad (-\alpha)^{q-1})$$

where $S^0 = 1 - \alpha\beta_1 + \alpha^2\beta_2 - \alpha^3\beta_3 + \; . \; . \; . \; + (-\alpha)^q\beta_q$.

4.3 The Generalised Variance

The Generalised Variance (G.V.) is defined to be the determinant of the covariance matrix for the efficient estimators of $\{\alpha_1,\ldots,\alpha_p,\beta\}$.

$$\text{Generalised Variance} = \det(\underline{F}^{-1}/n) = \frac{1}{n^{p+1}\det \underline{F}},$$

where $\underline{F}$ is partitioned as in the previous section. In Section 3.4 a factorization was found for $\det \underline{F}_{\alpha\alpha}^{-1}$. In this case, a factorization of $\det \underline{F}$ is not so obvious. However, a simple factorization exists, and is presented in terms of $\det \underline{F}_{\alpha\alpha}^{-1}$ and $\det F_{\beta\beta} = F_{\beta\beta}$, together with two scalar terms which are easily evaluated. The details of the factorization are contained in the following theorem.

Theorem 4.3.1

$$\det \underline{F} = \det \underline{F}_{\alpha\alpha} \det F_{\beta\beta} R^2/S^2 \tag{4.3.1}$$

where $R = \begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} \underline{\alpha}^* \\ 1 \end{bmatrix}$, $S = \begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} \underline{\alpha} \\ \alpha_p \end{bmatrix}$,

with $\underline{\beta}$ and $\underline{\alpha}$ defined as before, and $\underline{\alpha}^* = (\alpha_p \ \alpha_{p-1} \ \cdots \ \alpha_1)'$.

Proof

$$\underline{F} = \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & F_{\beta\beta} \end{bmatrix}$$

Pre-multiplying $\underline{F}$ by a matrix whose determinant is unity will have no effect on the determinant of the product of these two matrices. Hence

$$\det \underline{F} = \det \begin{bmatrix} I & 0 \\ -\underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1} & 1 \end{bmatrix} \cdot \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & F_{\beta\beta} \end{bmatrix}$$

$$= \det \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{0} & F_{\beta\beta} - \underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta} \end{bmatrix}$$

$$= (F_{\beta\beta} - \underset{\sim}{F}'_{\alpha\beta}\, \underset{\sim}{F}^{-1}_{\alpha\alpha} \underset{\sim}{F}_{\alpha\beta}) \det \underset{\sim}{F}_{\alpha\alpha}$$

$$= (1 - \underset{\sim}{F}'_{\alpha\beta} \underset{\sim}{F}^{-1}_{\alpha\alpha} \underset{\sim}{F}_{\alpha\beta} F^{-1}_{\beta\beta})\, F_{\beta\beta} \det \underset{\sim}{F}_{\alpha\alpha}\ . \qquad (4.3.2)$$

Equating (4.3.1) and (4.3.2) it is required to prove that

$$1 - \underset{\sim}{F}'_{\alpha\beta} \underset{\sim}{F}^{-1}_{\alpha\alpha} \underset{\sim}{F}_{\alpha\beta} F^{-1}_{\beta\beta} = R^2/S^2\ . \qquad (4.3.3)$$

Using Theorem 3.3.1 and Lemma 4.2.1, equation (4.3.3) can be re-written as

$$1 - \frac{(-\underset{\sim}{\beta}')}{S}(\underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B})\frac{(-\underset{\sim}{\beta})}{S}(1-\beta^2) = \frac{R^2}{S^2} \qquad (4.3.4)$$

since $F^{-1}_{\beta\beta} = 1-\beta^2$. Multiplying throughout by the lowest common denominator S^2 gives

$$S^2 - R^2 = \underset{\sim}{\beta}'(\underset{\sim}{A}'\underset{\sim}{A} - \underset{\sim}{B}'\underset{\sim}{B})\underset{\sim}{\beta}(1-\beta^2)\ . \qquad (4.3.5)$$

To prove equation (4.3.5) it is necessary to derive some preliminary results involving the matrices $\underset{\sim}{A}, \underset{\sim}{B}$ and augmented $\underset{\sim}{A}$ & $\underset{\sim}{B}$ matrices. Firstly consider

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-1} \\ 0 & 1 & \cdots & \alpha_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad \text{and define} \quad \underset{\sim}{A}_+ = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p} \\ 0 & 1 & \cdots & \alpha_{p-1} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Then

$$\underset{\sim}{A}'_+\underset{\sim}{A}_+ = \begin{bmatrix} \underset{\sim}{\alpha} \\ \\ \alpha_p \end{bmatrix} \begin{bmatrix} \underset{\sim}{\alpha}' & \alpha_p \end{bmatrix} + \begin{bmatrix} 0 & \underset{\sim}{0}'_p \\ \underset{\sim}{0}_p & \underset{\sim}{A}'\underset{\sim}{A} \end{bmatrix} \qquad (4.3.6)$$

where $\underset{\sim}{0}_p$ is the zero vector of length p.

Also

$$\underline{A}_+'\underline{A}_+ = \begin{bmatrix} \underline{0} & \underline{\pi}_* \\ \underline{\pi}_*' & \pi_o \end{bmatrix} + \begin{bmatrix} \underline{A}'\underline{A} & \underline{0}_p \\ \underline{0}_p' & 0 \end{bmatrix} \qquad (4.3.7)$$

where

$$\underline{\pi}_* = (\pi_p \ \pi_{p-1} \ . \ . \ . \ \pi_1)'$$

and

$$\pi_k = \alpha_k + \alpha_1\alpha_{k+1} + . \ . \ . + \alpha_{p-k}\alpha_p \qquad (1 \leq k \leq p)$$

$$\pi_o = 1 + \underline{\alpha}_*'\underline{\alpha}_* = 1+\alpha_1{}^2+\alpha_2{}^2+ \ . \ . \ . \ +\alpha_p{}^2.$$

Again,

$$\underline{B} = \begin{bmatrix} \alpha_p & \alpha_{p-1} \cdot & \cdot & \cdot & \alpha_1 \\ 0 & \alpha_p & \cdot & \cdot & \cdot \ \alpha_2 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & \alpha_p \end{bmatrix} \text{ and define } \underline{B}_+ = \begin{bmatrix} \alpha_p & \alpha_{p-1} \cdot & \cdot & \cdot & 1 \\ 0 & \alpha_p & \cdot & \cdot & \cdot \ \alpha_1 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & \alpha_p \end{bmatrix}.$$

Then

$$\underline{B}_+'\underline{B}_+ = \begin{bmatrix} \underline{\alpha}_* \\ 1 \end{bmatrix} \begin{bmatrix} \underline{\alpha}_*' & 1 \end{bmatrix} + \begin{bmatrix} 0 & \underline{0}_p' \\ \underline{0}_p & \underline{B}'\underline{B} \end{bmatrix} \qquad (4.3.8)$$

and

$$\underline{B}_+'\underline{B}_+ = \begin{bmatrix} \underline{0} & \underline{\pi}_* \\ \underline{\pi}_*' & \pi_o \end{bmatrix} + \begin{bmatrix} \underline{B}'\underline{B} & \underline{0}_p \\ \underline{0}_p' & 0 \end{bmatrix}. \qquad (4.3.9)$$

It also follows from the definitions of S and R that

$$S^2 = \begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} \underline{\alpha} \\ \alpha_p \end{bmatrix} \begin{bmatrix} \underline{\alpha}' & \alpha_p \end{bmatrix} \begin{bmatrix} \underline{\beta} \\ (-\beta)^p \end{bmatrix}$$

$$R^2 = \begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} \underline{\alpha}_* \\ 1 \end{bmatrix} \begin{bmatrix} \underline{\alpha}_*' & 1 \end{bmatrix} \begin{bmatrix} \underline{\beta} \\ (-\beta)^p \end{bmatrix}.$$

In proving equation (4.3.5) consider first the product

$$\underline{\beta}'(\underline{A}'\underline{A} - \underline{B}'\underline{B})\underline{\beta} \quad = \quad \begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} \underline{A}_+'\underline{A}_+ - \underline{B}_+'\underline{B}_+ \end{bmatrix} \begin{bmatrix} \underline{\beta} \\ (-\beta)^p \end{bmatrix}$$

using (4.3.7) & (4.3.9)

$$= \begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \left\{ \begin{bmatrix} 0 & \underline{0}_p' \\ \underline{0}_p & \underline{A}'\underline{A}-\underline{B}'\underline{B} \end{bmatrix} + \begin{bmatrix} \underline{\alpha} \\ \alpha_p \end{bmatrix} \begin{bmatrix} \underline{\alpha}' & \alpha_p \end{bmatrix} - \begin{bmatrix} \underline{\alpha}_* \\ 1 \end{bmatrix} \begin{bmatrix} \underline{\alpha}_*' & 1 \end{bmatrix} \right\} \begin{bmatrix} \underline{\beta} \\ (-\beta)^p \end{bmatrix}$$

using (4.3.6) & (4.3.8).

Thus

$$\underline{\beta}'(\underline{A}'\underline{A} - \underline{B}'\underline{B})\underline{\beta} \quad = \quad S^2 - R^2 + \beta^2\underline{\beta}'(\underline{A}'\underline{A} - \underline{B}'\underline{B})\underline{\beta}$$

using the definitions of S^2 and R^2 together with the result that

$$\begin{bmatrix} \underline{\beta}' & (-\beta)^p \end{bmatrix} \begin{bmatrix} 0 & \underline{0}_p' \\ \underline{0}_p & \underline{A}'\underline{A}-\underline{B}'\underline{B} \end{bmatrix} \begin{bmatrix} \underline{\beta}' \\ (-\beta)^p \end{bmatrix} \quad = \quad \beta^2.\underline{\beta}'(\underline{A}'\underline{A} - \underline{B}'\underline{B})\underline{\beta}.$$

Substituting in equation (4.3.5) it follows that

$$\begin{aligned} \text{RHS}(4.3.5) &= \begin{bmatrix} S^2 - R^2 + \beta^2.\underline{\beta}'(\underline{A}'\underline{A} - \underline{B}'\underline{B})\underline{\beta} \end{bmatrix} (1-\beta^2) \\ &= (S^2 - R^2)(1-\beta^2) + \beta^2.\text{RHS}(4.3.5). \end{aligned}$$

That is,

$$\begin{aligned} (1-\beta^2).\text{RHS}(4.3.5) \quad &= \quad (S^2 - R^2)(1 - \beta^2) \\ &= \quad (1-\beta^2).\text{LHS}(4.3.5). \end{aligned}$$

Hence the identity (4.3.5) has been verified. The determinant has been neatly factorized into

$$\det \underset{\sim}{F} = \det \underset{\sim}{F}_{\alpha\alpha} \det F_{\beta\beta} R^2/S^2$$

$$= \frac{R^2}{\det \underset{\sim}{F}_{\alpha\alpha}^{-1} (1-\beta^2)S^2}$$

and the generalised variance is given by

$$\text{G.V.} = \frac{1}{n^{p+1}\det \underset{\sim}{F}}$$

$$= \frac{\det \underset{\sim}{F}_{\alpha\alpha}^{-1}(1-\beta^2)S^2}{n^{p+1}R^2} .$$

Example ARMA(2,1) model

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t + \beta\varepsilon_{t-1}.$$

In this case,

$$\underset{\sim}{F}_{\alpha\alpha}^{-1} = \begin{bmatrix} 1 - \alpha_2^2 & \alpha_1(1-\alpha_2) \\ \alpha_1(1-\alpha_2) & 1 - \alpha_2^2 \end{bmatrix}$$

so that directly, or using Theorem 3.4.2,

$$\det \underset{\sim}{F}_{\alpha\alpha}^{-1} = (1-\alpha_2)^2\{(1+\alpha_2)^2 - \alpha_1^2\}.$$

Also

$$R = (1 \ -\beta \ \beta^2)(\alpha_2 \ \alpha_1 \ 1)' = \alpha_2 - \alpha_1\beta + \beta^2$$

$$S = (1 \ -\beta \ \beta^2)(1 \ \alpha_1 \ \alpha_2)' = 1 - \alpha_1\beta + \alpha_2\beta^2$$

Therefore

$$\det \underset{\sim}{F} = \frac{R^2}{\det \underset{\sim}{F}_{\alpha\alpha}^{-1} (1-\beta^2)S^2}$$

$$= \frac{(\alpha_2 - \alpha_1\beta + \beta^2)^2}{(1-\alpha_2)^2\{(1+\alpha_2)^2-\alpha_1^2\}.(1-\beta^2)(1-\alpha_1\beta+\alpha_2\beta^2)^2}$$

so that

$$\text{G.V.} = \frac{1}{n^{p+1}\det \underline{F}}$$

$$= \frac{(1-\alpha_2)^2.\{(1+\alpha_2)^2 - \alpha_1^{\ 2}\}(1-\beta^2)(1-\alpha_1\beta+\alpha_2\beta^2)^2.}{n^3(\alpha_2 - \alpha_1\beta + \beta^2)^2}$$

The simplification that arises from the factorisation (4.3.1) becomes apparent when the problem is considered directly. From the example in the previous section,

$$\underline{F} = \begin{bmatrix} \dfrac{1}{(1-\alpha_2)\{(1+\alpha_2)^2-\alpha_1^{\ 2}\}}\begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix} & \dfrac{-1}{1-\alpha_1\beta+\alpha_2\beta^2}\begin{bmatrix} 1 \\ -\beta \end{bmatrix} \\ \dfrac{-1}{1-\alpha_1\beta+\alpha_2\beta^2}\begin{bmatrix} 1 & -\beta \end{bmatrix} & \dfrac{1}{1-\beta^2} \end{bmatrix}.$$

The factorization of $\det \underline{F}$ is in no way obvious from an examination of the information matrix. Clearly some very lengthy algebra can be avoided by using the formula (4.3.1).

In general if p is large it may be helpful to factorize $\det \underline{F}_{\alpha\alpha}^{-1}$ as in Section 3.4, i.e.

$$\det \underline{F}_{\alpha\alpha}^{-1} = \alpha(1)\alpha(-1)(\det \underline{H})^2$$

and

$$\text{G.V.} = \frac{\alpha(1)\alpha(-1)(\det \underline{H})^2(1-\beta^2)S^2}{n^{p+1}R^2}$$

where

$$\underline{H} = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-2}-\alpha_p \\ 0 & 1 & \cdots & \alpha_{p-3}-\alpha_{p-1} \\ \vdots & \vdots & & \vdots \\ -\alpha_p & -\alpha_{p-1} & \cdots & 1-\alpha_2 \end{bmatrix}.$$

4.4 The Covariance Matrix

Let the ARMA(p,1) model be defined by (4.1.1) and let $\underset{\sim}{\theta} = (\alpha_1 \dots \alpha_p\ \beta)'$ denote the vector of parameters for the model. The covariance matrix for the efficient estimator $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$ is given by the inverse of the information matrix.

$$\frac{\underset{\sim}{V}}{n} = \frac{1}{n}\underset{\sim}{F}^{-1}$$

where

$$\underset{\sim}{F} = \begin{bmatrix} \underset{\sim}{F}_{\alpha\alpha} & \underset{\sim}{F}_{\alpha\beta} \\ \underset{\sim}{F}'_{\alpha\beta} & F_{\beta\beta} \end{bmatrix}.$$

Since $\underset{\sim}{F}$ is symmetric, its inverse can be written down in a relatively straightforward manner.

Let

$$\underset{\sim}{V} = \begin{bmatrix} \underset{\sim}{P} & \underset{\sim}{Q} \\ \underset{\sim}{Q}' & T \end{bmatrix}.$$

Then since $\underset{\sim}{F}\underset{\sim}{V} = \underset{\sim}{I}$ it follows that

$$T = (F_{\beta\beta} - \underset{\sim}{F}'_{\alpha\beta}\underset{\sim}{F}^{-1}_{\alpha\alpha}\underset{\sim}{F}_{\alpha\beta})^{-1}$$

$$\underset{\sim}{Q} = -\underset{\sim}{F}^{-1}_{\alpha\alpha}\underset{\sim}{F}_{\alpha\beta}T$$

and

$$\underset{\sim}{P} = \underset{\sim}{F}^{-1}_{\alpha\alpha} + (\underset{\sim}{F}^{-1}_{\alpha\alpha}\underset{\sim}{F}_{\alpha\beta})T(\underset{\sim}{F}'_{\alpha\beta}\underset{\sim}{F}^{-1}_{\alpha\alpha}) .$$

In order to find the individual variances and covariances, the procedure is first to find T, then the product $(\underset{\sim}{F}^{-1}_{\alpha\alpha}\underset{\sim}{F}_{\alpha\beta})$ and finally to combine these two producing $\underset{\sim}{Q} = ((Q_i))$ and $\underset{\sim}{P} = ((P_{ij}))$.

Clearly

$$n \text{ var } (\hat{\beta}) \quad = \quad T$$

$$n \text{ cov } (\hat{\alpha}_i \hat{\alpha}_j) \quad = \quad P_{ij} \qquad 1 \leq i,j \leq p$$

and

$$n \text{ cov } (\hat{\alpha}_i \hat{\beta}) \quad = \quad Q_i \qquad 1 \leq i \leq p \; .$$

The expressions for $\underline{P}$, $\underline{Q}$ and T can all be simplified owing to the particular forms of $\underline{F}_{\alpha\beta}$ and T in the ARMA(p,1) model. Firstly, consider T.

$$\begin{aligned} T &= (F_{\beta\beta} - \underline{F}'_{\alpha\beta}\underline{F}^{-1}_{\alpha\alpha}\underline{F}_{\alpha\beta})^{-1} \\ &= F^{-1}_{\beta\beta}\,(1 - \underline{F}'_{\alpha\beta}\underline{F}^{-1}_{\alpha\alpha}\underline{F}_{\alpha\beta}F^{-1}_{\beta\beta})^{-1} \\ &= (1-\beta^2)S^2/R^2 \end{aligned}$$

using Theorem 4.3.1.

Recall

$$\underline{F}_{\alpha\beta} = \frac{(-1)\underline{\beta}}{S} \; .$$

Hence the vector $\underline{Q}$ can be simplified to

$$\underline{Q} = -\underline{F}^{-1}_{\alpha\alpha}\frac{(-1)}{S}\underline{\beta}(1-\beta^2)\frac{S^2}{R^2}$$

$$\underline{Q} = \frac{(1-\beta^2)S.}{R^2}\underline{F}^{-1}_{\alpha\alpha}\,\underline{\beta} \; .$$

Finally, the matrix $\underline{P}$ can be simplified to

$$\begin{aligned} \underline{P} &= \underline{F}^{-1}_{\alpha\alpha} + \underline{F}^{-1}_{\alpha\alpha}\frac{(-1)}{S}\underline{\beta}\;\frac{(1-\beta^2)S^2}{R^2}\;\frac{(-1)}{S}\underline{\beta}'\underline{F}^{-1}_{\alpha\alpha} \\ &= \underline{F}^{-1}_{\alpha\alpha} + \frac{(1-\beta^2)}{R^2}\left[\underline{F}^{-1}_{\alpha\alpha}\;\underline{\beta}\;\underline{\beta}'\underline{F}^{-1}_{\alpha\alpha}\right]. \end{aligned}$$

The only quantities requiring calculation are S, R, $\underline{\beta}$ and $\underline{F}_{\alpha\alpha}^{-1}$. These are all easy to evaluate and substitute into the expressions for $\underline{P}$, $\underline{Q}$ and T. No matrix inversions are required since $\underline{F}_{\alpha\alpha}^{-1}$ is calculated directly using Theorem 3.3.1. With the simplifications above, the matrix $\underline{F}_{\alpha\alpha}$ does not need to be calculated.

<u>Example</u> ARMA(3,1) model

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} = \varepsilon_t + \beta\varepsilon_{t-1} \,.$$

$$n \text{ var } (\hat{\beta}) = \frac{(1-\beta^2)S^2}{R^2} = \frac{(1-\beta^2)(1-\alpha_1\beta+\alpha_2\beta^2-\alpha_3\beta^3)^2}{(\alpha_3-\alpha_2\beta+\alpha_1\beta^2-\beta^3)^2} \,. \tag{4.4.1}$$

Using Theorem 3.3.1,

$$\underline{F}_{\alpha\alpha}^{-1} = \begin{bmatrix} 1-\alpha_3^2 & \alpha_1-\alpha_2\alpha_3 & \alpha_2-\alpha_1\alpha_3 \\ \alpha_1-\alpha_2\alpha_3 & 1+\alpha_1^2-\alpha_2^2-\alpha_3^2 & \alpha_1-\alpha_2\alpha_3 \\ \alpha_2-\alpha_1\alpha_3 & \alpha_1-\alpha_2\alpha_3 & 1-\alpha_3^2 \end{bmatrix} .$$

Hence

$$n \begin{bmatrix} \text{var}(\hat{\alpha}_1) & \text{cov}(\hat{\alpha}_1\hat{\alpha}_2) & \text{cov}(\hat{\alpha}_1\hat{\alpha}_3) \\ \text{cov}(\hat{\alpha}_1\hat{\alpha}_2) & \text{var}(\hat{\alpha}_2) & \text{cov}(\hat{\alpha}_2\hat{\alpha}_3) \\ \text{cov}(\hat{\alpha}_1\hat{\alpha}_3) & \text{cov}(\hat{\alpha}_2\hat{\alpha}_3) & \text{var}(\hat{\alpha}_3) \end{bmatrix} = \underline{F}_{\alpha\alpha}^{-1} + \left(\frac{1-\beta^2}{R^2}\right).\{\underline{F}_{\alpha\alpha}^{-1}\underline{\beta}\underline{\beta}'\underline{F}_{\alpha\alpha}^{-1}\}$$

where $\underline{\beta} = (1 \ (-\beta) \ \beta^2)'$,

whence

$$n \text{ var } (\hat{\alpha}_1) = 1-\alpha_3^2 + \frac{(1-\beta^2).\{(1-\alpha_3^2) - \beta(\alpha_1-\alpha_2\alpha_3) + \beta^2(\alpha_2-\alpha_1\alpha_3)\}^2}{(\alpha_3-\alpha_2\beta+\alpha_1\beta^2-\beta^3)^2} \tag{4.4.2}$$

$$n \text{ var } (\hat{\alpha}_2) = 1+\alpha_1{}^2-\alpha_2{}^2-\alpha_3{}^2 + \frac{(1-\beta^2)\{(\alpha_1-\alpha_2\alpha_3)(1+\beta^2)-(1+\alpha_1{}^2-\alpha_2{}^2-\alpha_3{}^2)\beta\}^2}{(\alpha_3-\alpha_2\beta+\alpha_1\beta^2-\beta^3)^2} \tag{4.4.3}$$

and

$$n \text{ var } (\hat{\alpha}_3) = 1-\alpha_3{}^2 + \frac{(1-\beta^2)\{(\alpha_2-\alpha_1\alpha_3) - \beta(\alpha_1-\alpha_2\alpha_3) + \beta^2(1-\alpha_3{}^2)\}^2}{(\alpha_3-\alpha_2\beta+\alpha_1\beta^2-\beta^3)^2} . \tag{4.4.4}$$

The covariances between the α's follow in a similar fashion.

Also

$$\begin{bmatrix} \text{cov}(\hat{\alpha}_1\hat{\beta}) \\ \text{cov}(\hat{\alpha}_2\hat{\beta}) \\ \text{cov}(\hat{\alpha}_3\hat{\beta}) \end{bmatrix} = \frac{(1-\beta^2)S.\underline{F}^{-1}_{\alpha\alpha}\underline{\beta}}{R^2}$$

$$= \frac{(1-\beta^2)(1-\alpha_1\beta+\alpha_2\beta^2-\alpha_3\beta^3)}{(\alpha_3-\alpha_2\beta+\alpha_1\beta^2-\beta^3)^2} \begin{bmatrix} 1-\alpha_3{}^2 - \beta(\alpha_1-\alpha_2\alpha_3) + \beta^2(\alpha_2-\alpha_1\alpha_3) \\ (1+\beta^2)(\alpha_1-\alpha_2\alpha_3) - \beta(1+\alpha_1{}^2-\alpha_2{}^2-\alpha_3{}^2) \\ \alpha_2-\alpha_1\alpha_3 - \beta(\alpha_1-\alpha_2\alpha_3) + \beta^2(1-\alpha_3{}^2) \end{bmatrix} .$$

4.5 Adequacy of the Fitted Model

One technique for checking the adequacy of the fitted model is suggested by Box & Jenkins (1970, Chapter 8). Their method involves over-fitting i.e. estimating the parameters of a more general and therefore larger model than the one which it is believed fits the data adequately. This assumes that the direction of the inadequacy can be guessed so as not to add factors simultaneously to both sides of an ARMA model.

The above results show that this approach needs to be treated with caution, since 'over-fitting' frequently leads to possibly much larger sample variances. This can be illustrated by considering again the example in Section 4.4. If an AR(3) process is believed to best fit the data then the sample variances are

$$n \operatorname{var}(\hat{\alpha}_1) = n \operatorname{var}(\hat{\alpha}_3) = 1 - \alpha_3^2$$

and

$$n \operatorname{var}(\hat{\alpha}_2) = 1 + \alpha_1^2 - \alpha_2^2 - \alpha_3^2 .$$

By over-fitting with one moving average parameter i.e. an ARMA(3,1) process, but whose underlying value is $\beta=0$, then these three variances become

$$n \operatorname{var}(\hat{\alpha}_1) = 1 - \alpha_3^2 + \frac{(1-\alpha_3^2)^2}{\alpha_3^2} = \frac{1-\alpha_3^2}{\alpha_3^2} \qquad (4.5.1)$$

$$n \operatorname{var}(\hat{\alpha}_2) = 1 + \alpha_1^2 - \alpha_2^2 - \alpha_3^2 + \frac{(\alpha_1-\alpha_2\alpha_3)^2}{\alpha_3^2}$$

$$n \operatorname{var}(\hat{\alpha}_3) = 1 - \alpha_3^2 + \frac{(\alpha_2-\alpha_1\alpha_3)^2}{\alpha_3^2}$$

by setting $\beta=0$ in equations (4.4.2), (4.4.3) & (4.4.4). For stationarity, the inverse zeros z_1, z_2, z_3 of $\alpha(z)$

$$1 + \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 = (1-z_1 z)(1-z_2 z)(1-z_3 z)$$

must all be less than one in modulus. If for example $z_1 = z_2 = z_3 = \frac{1}{2}$, then $\alpha_3 = 1/8$ and the variance (4.5.1) is increased by a factor of 64. Similar large increases in the sample variances $n \operatorname{var}(\hat{\alpha}_2)$ and $n \operatorname{var}(\hat{\alpha}_3)$ would be expected if an AR(3) is over-fitted with an ARMA(3,1) process whose moving average parameter is in fact zero.

If an AR(p) process is over-fitted with an ARMA(p,1) process the sample variances $\operatorname{var}(\hat{\alpha}_i)$ are increased by

$$\frac{1}{\alpha_p^2}(\alpha_{i-1} - \alpha_{p-(i-1)}\alpha_p)^2 \qquad 1 \leqq i \leqq p \qquad \alpha_0 \equiv 1$$

and the sample covariances, $\operatorname{cov}(\hat{\alpha}_i\hat{\alpha}_j)$ are increased by

$$\frac{1}{\alpha_p^2}(\alpha_{i-1} - \alpha_{p-(i-1)}\alpha_p)(\alpha_{j-1} - \alpha_{p-(j-1)}\alpha_p) \qquad 1 \leqq i,j \leqq p \quad \alpha_0 \equiv 1$$

provided the underlying value of β is zero.

Similarly if an ARMA(p,1) has been fitted, and is then over-fitted with an ARMA(p+m,1) model, then the variance

$$n \operatorname{var}(\hat{\beta}) = (1 - \beta^2)S^2/R^2$$

where

$$S = 1-\alpha_1\beta+\alpha_2\beta^2 + \ldots + \alpha_p(-\beta)^p , \quad R = \alpha_p-\alpha_{p-1}\beta+\alpha_{p-2}\beta^2 + \ldots + (-\beta)^p$$

becomes

$$\frac{1-\beta^2}{\beta^{2m}}\frac{S^2}{R^2}$$

provided the m extra autoregressive parameters were zero. For invertibility $|\beta| < 1$, so if $\beta = \frac{1}{2}$ say, then the variance is increased by a factor of 4 even if the second model has just one extra autoregressive parameter with underlying value zero.

Note that an essential condition for these moments to exist and not be overwhelmingly large is that the zeros of $\alpha(z)$ and $\beta(z)$ be quite distinct. Redundant or nearly redundant factors are not necessarily obvious, thus it is always advisable to factorize both polynomials and examine carefully for any equal or nearly equal factors. Thus if $\alpha_1=\beta+\delta$, $\alpha_2=\beta\delta+\delta$ and $\alpha_3=\beta\delta$ for δ real and $|\delta|<1$, then the denominator in each of the four moments (4.4.1), (4.4.2), (4.4.3), (4.4.4) will be zero.

If nearly equal factors exist, then this can often be rectified by modifying the autoregressive parameters and compensating with a suitable change to the moving average parameters. By writing an ARMA process in its infinite moving average or infinite autoregressive representation it may be found that the relevant pure model yields greater stability in its parameter estimates.

Clearly if there is any resemblance between the zeros of $\alpha(z)$ and $\beta(z)$, then the sampling moments will be very large, which explains why the estimation problem is reportedly difficult in these cases. See for example Box & Jenkins (1970 §7.3.5).

CHAPTER 5

THE GENERAL ARMA (p,q) MODEL

5.1 Introduction

In this chapter, the covariance matrix for the efficient estimators of the parameters of an autoregressive moving average model of order (p,q) is considered. The model is defined by

$$X_t + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \qquad (5.1.1)$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed Gaussian random variables with expectation zero and variance σ^2. A realisation of n consecutive observations is available.

The computation of the covariance matrix for the efficient estimator of $\underline{\theta} = (\alpha_1 \dots \alpha_p\ \beta_1 \dots \beta_q)'$ is obviously of considerable interest to practitioners engaged in fitting models of the form (5.1.1) to data. However the formulation of the matrix is a general problem which seems to have attracted rather little comment in the literature. The pioneering work of Whittle (1953) & Durbin (1959) has not resolved the problem completely: Whittle gave a formula for the elements of the information matrix but, in particular for the mixed model, this method is rather cumbersome: Durbin's method for evaluating the covariance matrix for the efficient estimator of the parameters is valid for pure models only.

In the previous chapter the mixed model in which either p or q is unity was considered. It was shown that the information matrix could no longer be given simply by the inverse of a centro-symmetric covariance matrix, as is the case with pure models. The partitioned form for $\underline{F}$ has an off-diagonal block $\underline{F}_{\alpha\beta}$. In the previous chapter this block was a vector which could be evaluated easily by adapting a Quenouille-type algorithm (cf Lemma 4.2.1), but the specification of $\underline{F}_{\alpha\beta}$ in the more general case is considerably more complicated. Since p & q are both greater than unity, $\underline{F}_{\alpha\beta}$ is a pxq submatrix of the partitioned form

$$\underline{F} = \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & \underline{F}_{\beta\beta} \end{bmatrix} .$$

Quenouille's algorithm affords little assistance in giving $\underline{F}_{\alpha\beta}$ even when p & q are both quite small and it is necessary to find an alternative approach. In the next section a method is given for obtaining $\underline{F}_{\alpha\beta}$; it involves the specification of four upper triangular matrices whose elements are the parameters of the model, and the carrying out of some matrix products and additions.

Durbin's result (Theorem 3.3.1) is still valid for finding the inverses of $\underline{F}_{\alpha\alpha}$ and $\underline{F}_{\beta\beta}$, but the covariance matrix as a whole can only be specified by inverting $n\underline{F}$. Some help in deriving the covariance matrix for the efficient estimator of $\underline{\theta}$ is given by writing it in a form partitioned conformably with $\underline{F}$. Explicit expressions for the variances and covariances of the estimators can be found in cases where these moments are not too complicated.

It is assumed throughout this chapter that both p and q are greater than unity, since the theory of ARMA(p,1) & ARMA(1,q) models has already been discussed fully in Chapter 4. Also without loss of generality, the inequality $p \geq q$ is understood. The symmetry inherent in Whittle's formula for the information matrix implies that nothing is lost by this assumption, but for the sake of completeness, the case $p<q$ will be considered briefly at the end of Section 5.3.

5.2 The Information Matrix

The information matrix in partitioned form is defined by

$$n\underline{F} = n\begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & \underline{F}_{\beta\beta} \end{bmatrix} \tag{5.2.1}$$

where $n\underline{F}_{\alpha\alpha}$ is the pxp information matrix for the pure autoregression (3.1.1) and $n\underline{F}_{\beta\beta}$ is the qxq information matrix for the pure moving average model (3.6.1). These two submatrices can be evaluated using either of the methods described in Chapter 3; if p and q are both small then Lemma 3.2.1 is appropriate, otherwise Theorem 3.3.1 can be employed to yield $\underline{F}_{\alpha\alpha}^{-1}$ & $\underline{F}_{\beta\beta}^{-1}$, and these matrices then have to be inverted. The computation is reduced in the case $p=q$ owing to a duality result of Whittle (1953). Having obtained $\underline{F}_{\alpha\alpha}^{-1}$, $\underline{F}_{\beta\beta}^{-1}$ is given immediately by replacing the α's in $\underline{F}_{\alpha\alpha}^{-1}$ with β's, since the covariance matrices for the efficient estimators of the parameter sets are the same.

Let $\alpha(z)$, $\beta(z)$ be polynomials of degree p,q respectively defined by

$$\alpha(z) = 1+\alpha_1 z+\ldots+\alpha_p z^p \ , \quad \beta(z) = 1+\beta_1 z+\ldots+\beta_q z^q$$

and it is assumed that $\alpha(z)$, $\beta(z)$ have no zeros on the unit circle and no factors in common. Defining $B(z)=\beta(z)/\alpha(z)$ and letting $\theta \varepsilon \underline{\Theta}$, then the simplified form of Whittle's result states that the (i,j)-th element of $\underline{F}_{\alpha\beta}$ is the constant term in the expansion of

$$\frac{1}{2}\frac{\partial}{\partial\theta_i}\{\log B(z)B(z^{-1})\}\frac{\partial}{\partial\theta_j}\{\log B(z)B(z^{-1})\} . \tag{5.2.2}$$

This is equivalent to

Lemma 5.2.1

The (i,j)-th element of $\underset{\sim}{F}_{\alpha\beta}$ is the coefficient of z^{j-i} in the expansion of

$$\frac{-1}{\alpha(z)\beta(z^{-1})} \quad .$$

Examination of a Quenouille-type approach on the term $-1/\{\alpha(z)\beta(z^{-1})\}$ yields no simplification for specifying $\underset{\sim}{F}_{\alpha\beta}$ in the general class of mixed models. This is because $\underset{\sim}{F}_{\alpha\beta}$ is a matrix and not simply a vector as was the case in Lemma 4.2.1. The complexities of such an approach are illustrated by considering the smallest model in this category, namely the ARMA(2,2) process.

Example

The model is

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2}$$

Expressing $-1/\{\alpha(z)\beta(z^{-1})\}$ in partial fractions

$$\frac{-1}{(1+\alpha_1 z+\alpha_2 z^2)(1+\beta_1 z^{-1}+\beta_2 z^{-2})} = K_0 + \frac{K_1 z+K_2 z^2}{1+\alpha_1 z+\alpha_2 z^2} + \frac{L_1 z^{-1}+L_2 z^{-2}}{1+\beta_1 z^{-1}+\beta_2 z^{-2}} \qquad (5.2.3)$$

Multiplying throughout by the lowest common denominator $\alpha(z)\beta(z^{-1})$ yields

$$-1 = K_0(1+\alpha_1 z+\alpha_2 z^2)(1+\beta_1 z^{-1}+\beta_2 z^{-2})$$
$$+ (K_1 z+K_2 z^2)(1+\beta_1 z^{-1}+\beta_2 z^{-2}) + (L_1 z^{-1}+L_2 z^{-2})(1+\alpha_1 z+\alpha_2 z^2)$$

Equating coefficients of z^k for k=0, 1, 2, -1, -2, respectively yields

$$-1 = K_0(1+\alpha_1\beta_1+\alpha_2\beta_2) + K_1\beta_1 + K_2\beta_2 + L_1\alpha_1 + L_2\alpha_2$$

$$0 = K_0(\alpha_1+\alpha_2\beta_1) + K_1 + K_2\beta_1 + L_1\alpha_2$$

$$0 = K_0\alpha_2 + K_2$$

$$0 = K_0(\beta_1+\alpha_1\beta_2) + K_1\beta_2 + L_1 + L_2\alpha_1$$

$$0 = K_0\beta_2 + L_2 \ .$$

In this example

$$\underline{F}_{\alpha\beta} = \begin{bmatrix} K_0 & K_1 \\ L_1 & K_0 \end{bmatrix}$$

so the above five equations need to be solved for K_0, K_1 and L_1. After some lengthy and tedious algebra the solutions are

$$K_0 = \frac{-(1-\alpha_2\beta_2)}{(1-\alpha_2\beta_2)^2 - (\alpha_1-\alpha_2\beta_1)(\beta_1-\alpha_1\beta_2)} = \frac{-(1-\alpha_2\beta_2)}{\Delta_{\alpha\beta}} \quad \text{say}$$

$$K_1 = \frac{\alpha_1-\alpha_2\beta_1}{\Delta_{\alpha\beta}} \qquad \text{and} \qquad L_1 = \frac{\beta_1-\alpha_1\beta_2}{\Delta_{\alpha\beta}} \ .$$

For general ARMA(p,q) models equation (5.2.3) becomes

$$-1/\{\alpha(z)\beta(z^{-1})\} = K_0 + (K_1z + K_2z^2 + \ldots + K_pz^p)/\alpha(z)$$
$$+ (L_1z^{-1} + L_2z^{-2} + \ldots + L_qz^{-q})/\beta(z^{-1}) \ .$$

In this case, multiplying throughout by $\alpha(z)\beta(z^{-1})$ and equating coefficients of z^k yields a total of $p+q+1$ equations which have to be solved simultaneously to give explicit values for $K_0,K_1,\ldots,K_p,L_1,\ldots,L_q$. This alone is a very lengthy procedure, but further calculations are then required to give the elements of $\underline{F}_{\alpha\beta}$ which, typically, are linear combinations of the K's and L's.

It is clear that we must reject this algorithm and seek an alternative method.

Define the pxp matrix

$$\underset{\sim}{G} = \begin{bmatrix} g_0 & g_1 & \cdots & g_{p-1} \\ g_{-1} & g_0 & \cdots & g_{p-2} \\ \vdots & \vdots & & \vdots \\ g_{1-p} & g_{2-p} & \cdots & g_0 \end{bmatrix}$$

where the sequence $\{g_k : k = 0,\pm1,\pm2,\ldots\}$ is given by

$$\sum_{k=-\infty}^{\infty} g_k z^k = \frac{-1}{\alpha(z)\beta(z^{-1})} \; . \qquad (5.2.4)$$

Comparison with Lemma 5.2.1 implies that $\underset{\sim}{F}_{\alpha\beta}$ is just the first q columns of $\underset{\sim}{G}$, and if p=q then $\underset{\sim}{G} = \underset{\sim}{F}_{\alpha\beta}$. It would appear that the problem of finding an easy expression for $\underset{\sim}{F}_{\alpha\beta}$ is intractable, since Whittle's result and Quenouille's algorithm afford little assistance in this general case. However, a straightforward method exists for specifying the whole of the matrix $\underset{\sim}{G}$, and $\underset{\sim}{F}_{\alpha\beta}$ can be extrapolated from this result. The method involves some additions and multiplications of four upper triangular matrices of order p. All of these matrices assume the form

$$\begin{bmatrix} x_0 & x_1 & \cdots & x_{p-1} \\ 0 & x_0 & \cdots & x_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_0 \end{bmatrix} .$$

The matrices $\underset{\sim}{A}$, $\underset{\sim}{B}$, $\underset{\sim}{C}$ & $\underset{\sim}{D}$ are defined as follows:

For $\underline{A}$: $x_o = 1$ and $x_k = \alpha_k$ $(1 \leqq k \leqq p-1)$

For $\underline{B}$: $x_k = \alpha_{p-k}$ $(0 \leqq k \leqq p-1)$

For $\underline{C}$: $x_o = 1$ and $x_k = \begin{cases} \beta_k & (1 \leqq k \leqq \min(q,p-1)) \\ 0 & \text{otherwise} \end{cases}$

For $\underline{D}$: $x_k = \begin{cases} \beta_{p-k} & (p-q \leqq k \leqq p-1) \\ 0 & \text{otherwise .} \end{cases}$

It is interesting to note that $\underline{A}$ & $\underline{B}$ also feature in the evaluation of the covariance matrix for the efficient estimators of the parameters of the pure autoregression of order p (cf Chapter 3). The result for $\underline{G}$ is contained in the following theorem:

Theorem 5.2.1

$$\underline{G} = (\underline{D}\underline{B}' - \underline{A}\underline{C}')^{-1}$$

Proof

Equation (5.2.4) can be re-written in two ways. Firstly, on multiplying by $\alpha(z)$ it becomes

$$\alpha(z) \sum_{k=-\infty}^{\infty} g_k z^k = \frac{-1}{\beta(z^{-1})} = -1 - \sum_{k=1}^{\infty} b_k z^{-k} . \tag{5.2.5}$$

Equating powers of z^k for $k=0$, $k>0$ & $k<0$ respectively yields

$$\left.\begin{aligned} g_0 + \alpha_1 g_{-1} + \alpha_2 g_{-2} + \ldots + \alpha_p g_{-p} &= -1 \\ g_k + \alpha_1 g_{k-1} + \alpha_2 g_{k-2} + \ldots + \alpha_p g_{k-p} &= 0 \\ g_k + \alpha_1 g_{k-1} + \alpha_2 g_{k-2} + \ldots + \alpha_p g_{k-p} &= -b_k . \end{aligned}\right\} \tag{5.2.6}$$

From Equation (5.2.5),

$$\beta(z^{-1}) \{1 + \sum_{k=1}^{\infty} b_k z^{-k}\} = 1$$

so that

$$b_k + \beta_1 b_{k-1} + \beta_2 b_{k-2} + \ldots + \beta_k \qquad = 0 \tag{5.2.7}$$

for $k = 1,2,\ldots,q$, and by taking $\beta_{q+1} = \ldots = \beta_p = 0$, equation (5.2.7) also holds for $k = q+1,\ldots,p$. This verifies the form for $\underset{\sim}{C}^{-1}$ which will be required later:

$$\underset{\sim}{C}^{-1} = \begin{bmatrix} 1 & b_1 & \ldots & b_{p-1} \\ 0 & 1 & \ldots & b_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} .$$

Secondly, on multiplying equation (5.2.4) by $\beta(z^{-1})$

$$\beta(z^{-1}) \sum_{k=-\infty}^{\infty} g_k z^k = \frac{-1}{\alpha(z)} = -1 - \sum_{k=1}^{\infty} a_k z^k .$$

In this case, equating powers of z^k for $k=0$, $k>0$ & $k<0$ respectively yields

$$\left.\begin{aligned} g_0 + \beta_1 g_1 \quad + \ldots + \beta_q g_q \quad &= -1 \\ g_k + \beta_1 g_{k+1} + \ldots + \beta_q g_{k+q} &= -a_k \\ g_k + \beta_1 g_{k+1} + \ldots + \beta_q g_{k+q} &= \ \ 0 \end{aligned}\right\} . \tag{5.2.8}$$

Define the pxp matrix

$$\underset{\sim}{G}^{*} = \begin{bmatrix} g_{-p} & g_{1-p} & \ldots & g_{-1} \\ g_{-1-p} & g_{-p} & \ldots & g_{-2} \\ \vdots & \vdots & & \vdots \\ g_{1-2p} & g_{2-2p} & \ldots & g_{-p} \end{bmatrix} .$$

Equations (5.2.6) & (5.2.8) together imply that

$$\begin{bmatrix} \underline{G}^* & \underline{G} \end{bmatrix} \begin{bmatrix} \underline{C}' & \underline{B}' \\ \underline{D} & \underline{A} \end{bmatrix} = \begin{bmatrix} \underline{0}_p & -(\underline{C}')^{-1} \end{bmatrix} \qquad (5.2.9)$$

where $\underline{0}_p$ is a pxp matrix consisting entirely of zeros.
From equation (5.2.9)

$$\underline{G}^*\underline{C}' + \underline{G}\underline{D} = \underline{0}$$

giving

$$\underline{G}^* = -\underline{G}\underline{D}(\underline{C}')^{-1} . \qquad (5.2.10)$$

Further,

$$\underline{G}^*\underline{B}' + \underline{G}\underline{A} = -(\underline{C}')^{-1} .$$

Substituting for $\underline{G}^*$ from equation (5.2.10) gives

$$\underline{G}(-\underline{D}(\underline{C}')^{-1}\underline{B}' + \underline{A}) = -(\underline{C}')^{-1} .$$

Now $\underline{B}$ & $\underline{C}^{-1}$ commute, since $\underline{C}^{-1}$ is of the same form as $\underline{A}$, and hence the matrices satisfy the lemma preceding Theorem 3.3.1. It follows therefore that $\underline{B}'$ and $(\underline{C}')^{-1}$ commute, so that

$$\underline{G}(\underline{A} - \underline{D}\underline{B}'(\underline{C}')^{-1}) = -(\underline{C}')^{-1} .$$

Post-multiplying by $\underline{C}'$ gives

$$\underline{G}(\underline{A}\underline{C}' - \underline{D}\underline{B}') = -\underline{I}$$

whence

$$\underline{G} = (\underline{D}\underline{B}' - \underline{A}\underline{C}')^{-1}$$

as required.

It is also worth noting that

Corollary 5.2.2

$$\underline{G} = (\underline{B}'\underline{D} - \underline{C}'\underline{A})^{-1}.$$

To see this, it is helpful to refer to the matrix $\underline{J}$ defined in Corollary 3.3.2. Then

$$\begin{aligned}\underline{G}^{-1} &= \underline{D}\underline{B}' - \underline{A}\underline{C}' \\ &= \underline{J}(\underline{D}\underline{B}' - \underline{A}\underline{C}')'\underline{J} \\ &= \underline{J}(\underline{B}\underline{D}' - \underline{C}\underline{A}')\underline{J} \\ &= (\underline{J}\underline{B})\underline{D}'\underline{J} - (\underline{J}\underline{C})\underline{A}'\underline{J} \\ &= \underline{B}'(\underline{J}\underline{D}'\underline{J}) - \underline{C}'(\underline{J}\underline{A}'\underline{J}) \\ &= \underline{B}'\underline{D} - \underline{C}'\underline{A} \ .\end{aligned}$$

Theorem 5.2.1 gives a straightforward method for evaluating $\underline{F}_{\alpha\beta}$. First $\underline{G}^{-1}$ is obtained by using the theorem. Then the inverse of $\underline{G}^{-1}$ is derived. Finally, $\underline{F}_{\alpha\beta}$ is the first q columns of $\underline{G}$.

To illustrate how this theory is applied, consider the following example.

<u>Example ARMA (2,2)</u>

The model is

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} \ .$$

The information matrix is

$$n\underline{F} = n \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & \underline{F}_{\beta\beta} \end{bmatrix} \ .$$

Using Theorem 3.3.1,

$$\underline{F}_{\alpha\alpha}^{-1} = \begin{bmatrix} 1-\alpha_2{}^2 & \alpha_1(1-\alpha_2) \\ \alpha_1(1-\alpha_2) & 1-\alpha_2{}^2 \end{bmatrix}$$

whence

$$\underline{F}_{\alpha\alpha} = \frac{1}{(1-\alpha_2)\{(1+\alpha_2)^2 - \alpha_1{}^2\}} \begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix} \ .$$

Similarly

$$\underset{\sim}{F}_{\beta\beta} = \frac{1}{(1-\beta_2)\{(1+\beta_2)^2 - \beta_1^2\}} \begin{bmatrix} 1+\beta_2 & -\beta_1 \\ -\beta_1 & 1+\beta_2 \end{bmatrix} .$$

For this model, $\underset{\sim}{G} = \underset{\sim}{F}_{\alpha\beta}$, and

$$\underset{\sim}{A} = \begin{bmatrix} 1 & \alpha_1 \\ 0 & 1 \end{bmatrix}, \underset{\sim}{B} = \begin{bmatrix} \alpha_2 & \alpha_1 \\ 0 & \alpha_2 \end{bmatrix}, \underset{\sim}{C} = \begin{bmatrix} 1 & \beta_1 \\ 0 & 1 \end{bmatrix}, \underset{\sim}{D} = \begin{bmatrix} \beta_2 & \beta_1 \\ 0 & \beta_2 \end{bmatrix} .$$

Thus $\underset{\sim}{G}^{-1} = \underset{\sim}{D}\underset{\sim}{B}' - \underset{\sim}{A}\underset{\sim}{C}'$

$$= \begin{bmatrix} \alpha_2\beta_2-1 & \alpha_2\beta_1-\alpha_1 \\ \alpha_1\beta_2-\beta_1 & \alpha_2\beta_2-1 \end{bmatrix}$$

and

$$G = \frac{1}{\{(1-\alpha_2\beta_2)^2 - (\alpha_1-\alpha_2\beta_1)(\beta_1-\alpha_1\beta_2)\}} \begin{bmatrix} -(1-\alpha_2\beta_2) & \alpha_1-\beta_1\alpha_2 \\ \beta_1-\alpha_1\beta_2 & -(1-\alpha_2\beta_2) \end{bmatrix} .$$

Thus the information matrix is

$$n\underset{\sim}{F} = n \begin{bmatrix} \frac{1}{\Delta_\alpha} \begin{bmatrix} 1+\alpha_2 & -\alpha_1 \\ -\alpha_1 & 1+\alpha_2 \end{bmatrix} & \frac{1}{\Delta_{\alpha\beta}} \begin{bmatrix} -(1-\alpha_2\beta_2) & \alpha_1-\alpha_2\beta_1 \\ \beta_1-\alpha_1\beta_2 & -(1-\alpha_2\beta_2) \end{bmatrix} \\ \frac{1}{\Delta_{\alpha\beta}} \begin{bmatrix} -(1-\alpha_2\beta_2) & \beta_1-\alpha_1\beta_2 \\ \alpha_1-\alpha_2\beta_1 & -(1-\alpha_2\beta_2) \end{bmatrix} & \frac{1}{\Delta_\beta} \begin{bmatrix} 1+\beta_2 & -\beta_1 \\ -\beta_1 & 1+\beta_2 \end{bmatrix} \end{bmatrix}$$

where

$$\Delta_\alpha = (1-\alpha_2)\{(1+\alpha_2)^2 - \alpha_1^2\}$$

$$\Delta_\beta = (1-\beta_2)\{(1+\beta_2)^2 - \beta_1^2\}$$

and

$$\Delta_{\alpha\beta} = (1-\alpha_2\beta_2)^2 - (\alpha_1-\alpha_2\beta_1)(\beta_1-\alpha_1\beta_2) .$$

The specification of $\underset{\sim}{F}$ in partitioned form is a useful preliminary to deriving the covariance matrix for the efficient estimator of the vector of parameters, $\underset{\sim}{\Theta}$. It is also very helpful in testing hypotheses. For example, in testing $\underset{\sim}{\Theta} = \underset{\sim}{0}$, the test statistic

$$n\, \hat{\underset{\sim}{\Theta}}' \underset{\sim}{V}^{-1} \hat{\underset{\sim}{\Theta}}$$

is asymptotically distributed like χ^2 with $p+q$ degrees of freedom if $\underset{\sim}{\Theta} = \underset{\sim}{0}$. But $n\underset{\sim}{V}^{-1}$ is simply the inverse of the covariance matrix, which is the information matrix. Clearly in order to perform this test we would calculate

$$n\, \hat{\underset{\sim}{\Theta}}' \underset{\sim}{F} \hat{\underset{\sim}{\Theta}}$$

thus avoiding the lengthy procedure of evaluating the covariance matrix, and taking its inverse. The specification of the submatrices of $\underset{\sim}{F}$ is required in other tests described in Section 2.5.

5.3 The Covariance Matrix

In the previous section a method was outlined for obtaining an explicit expression for the information matrix in the partitioned form

$$n\underset{\sim}{F} = n \begin{bmatrix} \underset{\sim}{F}_{\alpha\alpha} & \underset{\sim}{F}_{\alpha\beta} \\ \underset{\sim}{F}'_{\alpha\beta} & \underset{\sim}{F}_{\beta\beta} \end{bmatrix} . \qquad (5.3.1)$$

The covariance matrix $\underset{\sim}{V}/n$ for the efficient estimator of $\underset{\sim}{\Theta}$ is given by the inverse of $n\underset{\sim}{F}$, and can be written in a form partitioned conformably with (5.3.1), say

$$\frac{1}{n}\underline{V} = \frac{1}{n}\begin{bmatrix} \underline{F}_{\alpha\alpha}^{-1} + \underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta}\underline{W}\underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1} & -\underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta}\underline{W} \\ -\underline{W}\underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1} & \underline{W} \end{bmatrix} \tag{5.3.2}$$

where

$$\underline{W} = (\underline{F}_{\beta\beta} - \underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta})^{-1} = \underline{F}_{\beta\beta}^{-1}(\underline{I} - \underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta}\underline{F}_{\beta\beta}^{-1})^{-1} .$$

Only two matrix inversions are required; one of order pxp, the other of order qxq. The pxq matrix $\underline{F}_{\alpha\beta}$ is derived by evaluating the pxp matrix $\underline{G}^{-1}$ and finding its inverse; $\underline{F}_{\alpha\beta}$ is then given by the first q columns of G. The other inversion involves $\underline{W}$.

Let

$$\underline{W} = \underline{F}_{\beta\beta}^{-1}\,\underline{U}^{-1}$$

denote the submatrix of $\underline{V}$ containing the variances and covariances of the β's. The second inversion is of the qxq matrix $\underline{U}$. This is somewhat cumbersome to evaluate explicitly in practice, but would be straightforward if the numerical values of the parameters were known. Thus no matrix inversions of order p+q are required using the partitioned form (5.3.2) for $\underline{V}/n$.

It is worth noting that a simplification of $\underline{V}$ occurs when p=q, by appealing again to Whittle's result, cited in the previous section. His result can be extended to mixed models by noting that the first diagonal submatrix of $\underline{V}$ can be expressed in the form

$$\underline{Y} = \underline{F}_{\alpha\alpha}^{-1}\,\underline{T}^{-1}$$

where $\underline{T} = \underline{I} - \underline{F}_{\alpha\beta}\underline{F}_{\beta\beta}^{-1}\underline{F}'_{\alpha\beta}\underline{F}_{\alpha\alpha}^{-1}$. By replacing α's with β's and vice versa in $\underline{Y}$, the matrix $\underline{W}$ is produced. Thus in the case p=q, the off-diagonal submatrix and only one diagonal block need to be calculated.

For the sake of completeness, we consider the case where the order of the moving average component in (5.1.1) exceeds that of the autoregressive component. Retaining all previous definitions and also $0 < q \leq p$, let the model be defined by

$$X_t + \beta_1 X_{t-1} + \dots + \beta_q X_{t-q} = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_p \varepsilon_{t-p} \ . \tag{5.3.3}$$

The information matrix in partitioned form is n times the matrix

$$\begin{bmatrix} \underline{F}_{\beta\beta} & \underline{F}'_{\alpha\beta} \\ \underline{F}_{\alpha\beta} & \underline{F}_{\alpha\alpha} \end{bmatrix}$$

whilst the covariance matrix for the maximum likelihood estimator of the vector of parameters $(\beta_1 \dots \beta_q\ \alpha_1 \dots \alpha_p)'$ for the model (5.3.3) in the Gaussian case is n^{-1} times the matrix

$$\begin{bmatrix} \underline{W} & -\underline{W}\underline{F}'_{\alpha\beta}\underline{F}^{-1}_{\alpha\alpha} \\ -\underline{F}^{-1}_{\alpha\alpha}\underline{F}_{\alpha\beta}\underline{W} & \underline{F}^{-1}_{\alpha\alpha} + \underline{F}^{-1}_{\alpha\alpha}\underline{F}_{\alpha\beta}\underline{W}\underline{F}'_{\alpha\beta}\underline{F}^{-1}_{\alpha\alpha} \end{bmatrix}$$

These two partitioned forms correspond to (5.3.1) and (5.3.2) respectively, and the evaluation of the submatrices proceeds in exactly the same way as previously.

5.4 Concluding Remarks

Although the ARMA(p,1) and ARMA(1,q) models were discussed separately in the previous chapter, the partitioned forms for the information matrix and the covariance matrix for the efficient estimators of the parameters of the model are compatable with those described in this chapter. For ARMA(p,1) models, the matrix $\underline{W}$ becomes a scalar quantity and $\underline{F}_{\alpha\beta}$ is a column vector of length p.

Theorem 5.2.1 involves the inversion of a matrix of order p, so for q=1, $\underset{\sim}{F}_{\alpha\beta}$ can be specified more easily using Lemma 4.2.1, irrespective of the value of p. The details for the ARMA(2,1) process are given in Godolphin & Unwin (1983).

The specification of $\underset{\sim}{V}$ using equation (5.3.2) has many advantages over a method based on inverting the information matrix which has not been partitioned. In this general case, no simplified form for the determinant of $\underset{\sim}{F}$ exists, with the result that the evaluation of the inverse of $\underset{\sim}{F}$ using a method of cofactors would be very cumbersome in all but the smallest of mixed models. The individual components of $\underset{\sim}{V}$ in the partitioned form, namely $\underset{\sim}{F}_{\alpha\alpha}^{-1}$, $\underset{\sim}{F}_{\beta\beta}^{-1}$ & $\underset{\sim}{F}_{\alpha\beta}$, are easily evaluated, but their combinations and products which constitute the submatrices of $\underset{\sim}{V}$ itself are rather complicated. Thus explicit expressions for the variances and covariances of the estimators can only be found for models containing a small number of parameters. The ARMA(3,1) model was treated in full in the example in Section 4.4, but the details of the ARMA(2,2) process are considerably more complicated, despite the fact that the models have the same number of parameters.

By way of a summary of the theory contained in Chapters 3, 4 & 5, the following algorithm gives the salient details for specifying the information matrix and the covariance matrix for the efficient estimators of the parameters of univariate stationary time series models.

Algorithm

Step 1

For $p>q$ define the $p \times p$ matrices

$$\underline{A} = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-1} \\ 0 & 1 & \cdots & \alpha_{p-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad \underline{B} = \begin{bmatrix} \alpha_p & \alpha_{p-1} & \cdots & \alpha_1 \\ 0 & \alpha_p & \cdots & \alpha_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \alpha_p \end{bmatrix}, \quad \underline{C} = \begin{bmatrix} 1 & \beta_1 & \cdots & \beta_q & 0 & \cdots & 0 \\ 0 & 1 & \cdots & & \beta_q & \cdots & 0 \\ \vdots & \vdots & & & \vdots & & \beta_q \\ 0 & 0 & \cdots & & \cdots & & 1 \end{bmatrix}$$

$$\text{and } \underline{D} = \begin{bmatrix} 0 & \cdots & 0 & \beta_q & \cdots & \beta_2 & \beta_1 \\ 0 & \cdots & 0 & 0 & \cdots & \beta_3 & \beta_2 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \cdot & & \cdot & \cdot & & \cdot & \beta_q \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & & 0 & 0 & & 0 & 0 \end{bmatrix}.$$

If $p=q$ then $\underline{C}$ & $\underline{D}$ are equivalent to $\underline{A}$ & $\underline{B}$ respectively, with β's replaced by α's.

Step 2

Evaluate $\underline{G}^{-1} = \underline{B}'\underline{D} - \underline{C}'\underline{A}$ and take its inverse. With the information matrix for an ARMA(p,q) process defined in the partitioned form

$$n\underline{F} = n \begin{bmatrix} \underline{F}_{\alpha\alpha} & \underline{F}_{\alpha\beta} \\ \underline{F}'_{\alpha\beta} & \underline{F}_{\beta\beta} \end{bmatrix} \qquad (1)$$

then the off-diagonal submatrix $\underline{F}_{\alpha\beta}$ is given by the first q columns of $\underline{G}$.

Step 3

Evaluate $\underline{F}_{\alpha\alpha}^{-1} = \underline{A}'\underline{A} - \underline{B}'\underline{B}$. This matrix is n times the covariance matrix for the efficient estimators of the parameters of the pure autoregression of order p. Inverting $\underline{F}_{\alpha\alpha}^{-1}/n$ gives the information matrix for the AR(p) process and also the first diagonal submatrix of the information matrix of the ARMA(p,q) process, given by (1).

Step 4

Re-define $\underline{A}$ & $\underline{B}$ as matrices of order q

$$\underline{A}_\beta = \begin{bmatrix} 1 & \beta_1 & \cdots & \beta_{q-1} \\ 0 & 1 & \cdots & \beta_{q-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \qquad \underline{B}_\beta = \begin{bmatrix} \beta_q & \beta_{q-1} & \cdots & \beta_1 \\ 0 & \beta_q & \cdots & \beta_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \beta_q \end{bmatrix}.$$

Evaluate $\underline{F}_{\beta\beta}^{-1} = \underline{A}_\beta'\underline{A}_\beta - \underline{B}_\beta'\underline{B}_\beta$ which is n times the covariance matrix for the efficient estimators of $\{\beta_1,\ldots,\beta_q\}$. Inverting $\underline{F}_{\beta\beta}^{-1}/n$ gives the information matrix for the pure MA(q) process and the second diagonal submatrix in (1).

Step 5

The covariance matrix for the efficient estimators of the parameters of the ARMA(p,q) process is obtained using the partitioned form

$$\frac{1}{n}\underline{V} = \frac{1}{n}\begin{bmatrix} \underline{F}_{\alpha\alpha}^{-1} + \underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta}\underline{W}\underline{F}_{\alpha\beta}'\underline{F}_{\alpha\alpha}^{-1} & -\underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta}\underline{W} \\ -\underline{W}\underline{F}_{\alpha\beta}'\underline{F}_{\alpha\alpha}^{-1} & \underline{W} \end{bmatrix}$$

where $\underline{W} = (\underline{F}_{\beta\beta} - \underline{F}_{\alpha\beta}'\underline{F}_{\alpha\alpha}^{-1}\underline{F}_{\alpha\beta})^{-1}$.

CHAPTER 6

POLYNOMIAL-PROJECTING MODELS

6.1 Introduction

The study of data collected from processes which evolve in time has occupied the attention of scientific analysts for several decades. One difficulty that occurs when studying such data is that in many practical situations it is customary to encounter time series that are not stationary in themselves. However in many cases it may be possible to transform these data to stationary processes with properties similar to those of low-order autoregressive moving average models. Such transformations would preferably be free from unknown parameters. For example models of interest to the control engineer are typically based on physical considerations and may be expected to contain certain low-order polynomial trends. These models are frequently expressed in state-space formulation, where each component of the state space is intended to have some physical meaning. Thus it is usual to model the movement of a ship, for example, in terms of distance and velocity variables in two dimensions.

Furthermore, by suitably differencing the observed series it is possible, in principle, to obtain transformed time series with properties similar to those of low-order ARMA models. By definition these ARMA models contain a number of unknown parameters which can be estimated and these estimates examined for significance. Alternative forms of non-parametric transformations are sometimes used, such as the taking of logarithms, but in the rest of this thesis the only type of transformation considered is that of differencing. This provides one way in which to apply principles of statistical inference to state-space models so as to examine hypotheses of particular interest to the control engineer.

6.2 Univariate State Space Models

In Section 2.6 we defined the general form for the multivariate state space model, and briefly described the practicalities of restricting the components of the model equations. The models considered in this chapter are in the form given by Harrison & Stevens (1976). Thus the system matrix $\underline{G}_t$ is independent of time and $\underline{H}_t = \underline{I}$. The vector $\underline{F}_t$ is also independent of time but we relax the condition $\underline{F} = (1\ 0 \ldots 0)'$ which Harrison & Stevens require with all of their polynomial models. Thus the univariate model is defined by

$$X_t = \underline{F}\underline{\theta}_t + v_t \tag{6.2.1}$$

$$\underline{\theta}_t = \underline{G}\underline{\theta}_{t-1} + \underline{w}_t . \tag{6.2.2}$$

The random components v_t and $\underline{w}_t$ are independent and have $N(0,V)$, $N(\underline{0},\underline{W})$ distributions respectively and represent additive noise. The vectors $\underline{\theta}_t$ and $\underline{w}_t$ are of order nx1 with $\underline{G}$ of order nxn. In this univariate case, $\underline{F}$ is a vector of order 1xn.

In applying the model defined by (6.2.1),(6.2.2), the control engineer bases the components of the state space on intuitively sensible considerations. Each component is intended to have a physical meaning, such as distance, velocity and acceleration.

These concepts are useful when modelling the movement of objects such as aircraft or submarines.

The models considered in this chapter also satisfy the criterion

$$\underline{F}\underline{G}(\underline{G}-\underline{I})^{d+1} = \underline{0} \tag{6.2.3}$$

for some non-negative integer d. Godolphin & Stone (1980) have shown that these models can be interpreted as polynomial-projecting models of degree d; they possess the property that the minimum mean-squared error forecast for the model at a fixed time origin describes a polynomial in the lead time variable of degree d. Box & Jenkins (1970, Chapter 5) demonstrate that the same is true for models of the form

$$\nabla^{d+1} X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \tag{6.2.4}$$

where ∇ denotes the backward difference operator, and provided $q \leqq d+1$. The same equivalence holds for models of the form

$$\nabla^{d} X_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \tag{6.2.5}$$

provided that $q \leqq d$, and where $\mu \neq 0$ is a deterministic parameter.

It may occur that a model satisfying (6.2.3) will be represented by a system vector of $n = d+1$ components. Thus the system equation is

$$X_t = \underline{F}(\theta_{ot}\ \theta_{1t} \cdots \theta_{dt})' + v_t \tag{6.2.1)'$$

and the measurement equation is given by

$$\begin{bmatrix} \theta_{ot} \\ \theta_{1t} \\ \vdots \\ \theta_{dt} \end{bmatrix} = \underline{G} \begin{bmatrix} \theta_{ot-1} \\ \theta_{1t-1} \\ \vdots \\ \theta_{dt-1} \end{bmatrix} + \begin{bmatrix} w_{ot} \\ w_{1t} \\ \vdots \\ w_{dt} \end{bmatrix} . \tag{6.2.2)'$$

$$\nabla X_t = \underline{F}(\underline{G}-\underline{I})\underline{\theta}_{t-1} + \underline{F}\underline{w}_t + \nabla v_t$$

$$= \underline{F}\underline{G}(\underline{G}-\underline{I})\underline{\theta}_{t-2} + \underline{F}(\underline{G}-\underline{I})\underline{w}_{t-1} + \underline{F}\underline{w}_t + \nabla v_t \;.$$

Similarly

$$\nabla^2 X_t = \underline{F}\underline{G}(\underline{G}-\underline{I})^2\underline{\theta}_{t-3} + \underline{F}(\underline{G}-\underline{I})^2\underline{w}_{t-2} + \underline{F}(\underline{G}-\underline{I})\underline{w}_{t-1} + \underline{F}\nabla\underline{w}_t + \nabla^2 v_t$$

and in general, for $d \geqq 1$,

$$\nabla^{d+1}X_t = \underline{F}\underline{G}(\underline{G}-\underline{I})^{d+1}\underline{\theta}_{t-d-2} + \sum_{i=d}^{d+1}\underline{F}(\underline{G}-\underline{I})^i\underline{w}_{t-i} + \sum_{i=1}^{d-1}\underline{F}(\underline{G}-\underline{I})^{d-i}\nabla^i\underline{w}_{t-d+i} + \underline{F}\nabla^d\underline{w}_t + \nabla^{d+1}v_t \;. \tag{6.2.6}$$

From equation (6.2.3), the first term on the right hand side of (6.2.6) disappears. The remaining terms are linear combinations of white noise random variables $w_{j,t-i}$ $(0 \leqq i,j \leqq d)$ and also $\nabla^{d+1}v_t$. Thus the right hand side of (6.2.6) has covariances which are identical to those of a stationary moving average process of order $d+1$.

However it does not follow that this process will be invertible. This would appear to depend on the structure of the state space model, and in particular on the form of the system matrix, $\underline{G}$. In fact there is no polynomial-projecting model of degree d with $\underline{G}$ non-singular which yields values for $\beta_1,\ldots,\beta_q$ which cover the entire stability region. To achieve this result it appears that the system matrix would need to be of order greater than $(d+1)x(d+1)$. The implication of this constraint is that the description of a polynomial model of degree d would need more than $d+1$ system 'parameters'. In practical situations the criterion of invertibility may be the main factor in deciding which of the two models (6.2.4) or (6.2.5) is the more appropriate for describing data generated by the Harrison-Stevens model (6.2.1),(6.2.2).

6.3 Multivariate State Space Models

The output data X_t encountered by control engineers is, in practice, usually multivariate, rather than univariate. That is, for each time point t, there are s distinct observations available

$$X_t = (X_{1t},\dots,X_{st})'$$

where $s > 1$. For example, in tracking a moving target, such as a ship, its motion is in two dimensions. This motion might be described in cartesian co-ordinates relative to some fixed origin, in which case $s=2$. A question arises as to the inter-relation of the s components. In some cases of practical interest, such as the motion of the ship described above, it is permissible to assume that the components $X_{1t},\dots,X_{st}$ of X_t are independent of each other. In these cases the treatment of the corresponding multivariate state space models follows as a straightforward generalisation of that for the univariate state space models. Each component X_{it} $(1 \leqq i \leqq s)$ is assumed to have a state space representation which is unaffected by the univariate representation for components X_{jt} $(j \neq i)$. Hence the elements of X_t can all be treated separately. This is assumed in the simulations described in Chapter 7.

6.4 Inference for Polynomial-Projecting Models

In Section 6.2 it was shown that the Harrison-Stevens state space form defined by equations (6.2.1),(6.2.2) can be interpreted as one of two non-stationary time series models defined by (6.2.4) and (6.2.5). There exist many variants on the Harrison-Stevens representation, but these models still yield polynomial-projecting predictors. Also in many cases these models have fewer restrictions than the Harrison-Stevens models without additional complications in the structure of the system. Recall

$$X_t = \underline{F}\underline{\theta}_t + v_t$$

$$\underline{\theta}_t = \underline{G}\underline{\theta}_{t-1} + \underline{w}_t \ .$$

The system vector of parameters, $\underline{\theta}_t$, is estimated using the Kalman updating equation:

$$\hat{\underline{\theta}}_t = \underline{G}\hat{\underline{\theta}}_{t-1} + \underline{A}_t(X_t - \underline{F}\underline{G}\hat{\underline{\theta}}_{t-1}) \qquad (6.4.1)$$

where $\underline{A}_t$ is the Kalman Gain vector

$$\underline{A}_t = \underline{P}_t\underline{F}'(\underline{F}\underline{P}_t\underline{F}' + V_t)^{-1}$$

and where

$$\underline{P}_t = \underline{G}\underline{C}_{t-1}\underline{G}' + \underline{W}_t \ .$$

The matrix

$$\underline{C}_t = E\left[(\underline{\theta}_t - \hat{\underline{\theta}}_t)(\underline{\theta}_t - \hat{\underline{\theta}}_t)'\right]$$

denotes the covariance matrix for the difference between the system vector and its estimate. It is easily seen that

$$\underline{C}_t = (\underline{I} - \underline{A}_t\underline{F})\underline{P}_t$$

$$= \underline{P}_t - \underline{A}_t(\underline{F}\underline{P}_t\underline{F}' + V_t)\underline{A}_t' \ .$$

Godolphin & Stone (1980) have shown that the predictors of polynomial projecting models of degree d and the least squares predictors of models (6.2.4) and (6.2.5) are in a certain sense equivalent. Define

$$m_{it} = \underline{F}\underline{G}(\underline{G} - \underline{I})^{i-1}\hat{\underline{\theta}}_t \qquad i=1,\ldots, d+1$$

$$\alpha_{it} = \underline{F}\underline{G}(\underline{G} - \underline{I})^{i-1}\underline{A}_t \qquad i=1,\ldots, d+1 \ .$$

Multiplying the Kalman updating equation (6.4.1) by $\underline{F}\underline{G}$, $\underline{F}\underline{G}(\underline{G}-\underline{I}),\ldots,\ \underline{F}\underline{G}(\underline{G}-\underline{I})^d$ we obtain

$$m_{i,t} = m_{i+1,t-1} + m_{i,t-1} + \alpha_{it}(X_t - m_{1,t-1}) \qquad i=1,\ldots,d$$

$$m_{d+1,t} = m_{d+1,t-1} + \alpha_{d+1,t}(X_t - m_{1,t-1})$$

so that

$$\underline{M}_t = \underline{K}\underline{M}_{t-1} + \underline{\alpha}_t(X_t - m_{1,t-1}) \tag{6.4.2}$$

where

$$\underline{M}_t = (m_{1,t},\ldots,\ m_{d+1,t})'$$

$$\underline{\alpha}_t = (\alpha_{1,t},\ldots,\ \alpha_{d+1,t})'$$

and

$$\underline{K} = ((k_{ij})) = \begin{cases} 1 & i=j \text{ or } i+1 = j \\ 0 & \text{otherwise} . \end{cases}$$

Godolphin & Stone (1980) apply z-transforms to equation (6.4.2) and show that the stability region is given by the space in which the zeros of

$$\beta(z) = 1 + \beta_1 z + \ldots + \beta_{d+1} z^{d+1}$$

are greater than one in modulus. In addition, this analysis yields explicit values for the moving average parameters, namely

$$\beta_i = (-1)^i \binom{d+1}{i} + \sum_{j=0}^{i-1} (-1)^j \binom{d+1}{j} \underline{F}\underline{G}^{i-j}\underline{A} \tag{6.4.3}$$

where $\underline{A}$ denotes the limiting steady state value of the Kalman gain vector $\underline{A}_t$.

The form of the state space model is determined by the nature of the practical situation being examined. If equation (6.2.3) holds, then a time series of the form (6.2.4) or (6.2.5) may be suitable for describing the problem. The invertibility criterion will usually decide which model to use, as was discussed in Section 6.2.

Thus the maximum number of moving average parameters present is known, taking account of the fact that some of these may be zero. Then instead of adopting (6.4.3), the moving average parameters can be estimated using the well-established inference techniques of time series analysis. It follows that the estimation procedures described here can apply to a realisation $X_1, \ldots, X_n$ of the model

$$X_t = \nabla^d Z_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q} \qquad (6.4.4)$$

where $q \leqq d$ and where the mean, μ, may be an unknown parameter. If μ is unknown it is estimated by

$$\hat{\mu} = \overline{X} = \begin{cases} (Z_1 + Z_2 + \ldots + Z_k)/k & \text{if } d=0 \\ (\nabla^{d-1} Z_k - \nabla^{d-1} Z_d)/k\text{-}d & \text{if } d \geqq 1 \end{cases}$$

where $k=n+d$ is the size of the sample set before the differencing is carried out. This estimator of μ is both unbiased and consistent, as was shown in Chapter 2.

The estimation of the moving average parameters based on the principle of maximum likelihood has been considered by Godolphin (1977). Let $\sigma^2 \underline{\Gamma}_n$ denote the covariance matrix of $\underline{X}$ for the model (6.4.4) and $\underline{\Pi}_n = ((\pi_{|i-j|}))$ be the covariance matrix for n consecutive values of the autoregressive process $\{Y_t\}$ defined by

$$Y_t + \beta_1 Y_{t-1} + \ldots + \beta_q Y_{t-q} = \eta_t$$

where $\text{var}(\eta_t) = 1$. Approximate maximum likelihood equations are obtained for large n by ignoring the contribution to the log likelihood of $\det \underline{\Gamma}_n$ and by replacing $\underline{\Gamma}_n^{-1}$ by $\underline{\Pi}_n$. Thus for the moving average parameters,

$$\frac{\partial\pi_0}{\partial\beta_j} + 2\sum_{k=1}^{m}\frac{\partial\pi_k}{\partial\beta_j}r_k = 0 \qquad (6.4.5)$$

where $r_k = \dfrac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})/n-k}{\sum_{t=1}^{n}(X_t - \bar{X})^2/n}$.

The upper limit on the summation denotes a suitably large number $m>q$ for which the partial derivatives are negligible. This estimation procedure is unbiased in the sense that the likelihood equations are satisfied if the r_k's are replaced by their asymptotic expectations $r_k = \rho_k$ $(1\leq k\leq q)$, $r_k = 0$ $(k\geq q+1)$. i.e.

$$d_{j,o} + 2\sum_{k=1}^{q} d_{j,k}\rho_k = 0 \qquad (j=1, \ldots, q)$$

where $d_{j,k}$ denotes the partial derivative $\dfrac{\partial\pi_k}{\partial\beta_j}$. In matrix notation,

$$\underline{D}_0 + 2\underline{D}_1\underline{\rho} = \underline{0} \qquad (6.4.6)$$

where

$$\underline{D}_0 = \begin{bmatrix} d_{1,0} \\ d_{2,0} \\ \vdots \\ d_{q,0} \end{bmatrix}, \qquad \underline{D}_1 = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,q} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,q} \\ \vdots & \vdots & & \vdots \\ d_{q,1} & d_{q,2} & \cdots & d_{q,q} \end{bmatrix} .$$

The likelihood equations (6.4.5) can be written in a comparable form:

$$\underline{D}_0 + 2\underline{D}_1\underline{R}_1 + 2\underline{D}_2\underline{R}_2 = \underline{0} \qquad (6.4.7)$$

where $\underline{D}_2$ is a matrix of order $q \times (m-q)$ with elements $d_{i,q+j}$

and $\underline{R}_1$, $\underline{R}_2$ are vectors of sample serial correlations,

$$\underline{R}_1 = (r_1 r_2 \ldots r_q)', \quad \underline{R}_2 = (r_{q+1} r_{q+2} \ldots r_m)' .$$

Combining (6.4.6) and (6.4.7) gives the estimation equation for $\underline{\rho}$ as

$$\underline{\rho} = \underline{R}_1 + \underline{E}\underline{R}_2 \tag{6.4.8}$$

where $\underline{E} = \underline{D}_1^{-1}\underline{D}_2$. It is interesting to note that the elements of $\underline{D}_2$ can be expressed in terms of the elements of $\underline{D}_0$ & $\underline{D}_1$ by virtue of the fact that

$$x_k = x_{-k}, \quad x_k + \sum_{i=1}^{2q} \beta_i^* x_{k-i} = 0 \qquad k \geq q+1$$

where $x_k = d_{j,k}$, $1 \leq j \leq q$ and where the β^*'s satisfy

$$\{\beta(z)\}^2 = 1 + \sum_{h=1}^{2q} \beta_h^* z^h .$$

Thus the elements of $\underline{E} = \underline{D}_1^{-1}\underline{D}_2$ are given in terms of the corresponding elements of $\underline{D}_1^{-1}\underline{D}_0 = -2\underline{\rho}$ and $\underline{D}_1^{-1}\underline{D}_1 = \underline{I}$. In view of this result, the elements of $\underline{E}$ and hence the estimation of $\underline{\rho}$ requires no matrix inversions which is an advantage over some alternative estimation procedures such as Walker's (1961).

It is quite common in estimation problems that q takes the value 1 or 2. In such cases equation (6.4.8) is used to obtain $\underline{\rho}$ with initial estimate $\underline{\rho} = \underline{R}_1$ and the β's are then obtained using the Cramér-Wold factorization. However for $q>2$ it is preferable to transform equation (6.4.8) to a system in terms of $\hat{\underline{\beta}}$. Such a procedure would require more iterations to arrive at the solution, but avoids the need for the Cramér-Wold factorisation. To establish the β's it is required to find a $q \times q$ matrix $\underline{H}_1$ such that

$$\underline{H}_1 \underline{\rho} = \hat{\underline{\beta}} .$$

Substituting for $\underline{\rho}$ from (6.4.8) gives

$$\hat{\underline{\beta}} = \underline{H}_1\underline{R}_1 + \underline{H}_2\underline{R}_2 \tag{6.4.9}$$

where $\underline{H}_2 = \underline{H}_1\underline{E}$. From the form of the elements of $\underline{E}$ it follows that the elements of $\underline{H}_2$ are given in terms of the corresponding elements of $\underline{H}_1(-2\underline{\rho}) = -2\hat{\underline{\beta}}$ and $\underline{H}_1\underline{I} = \underline{H}_1$. Thus equation (6.4.9) gives an iterative solution for $\hat{\underline{\beta}}$ in terms of $\underline{H}_1$, $\underline{R}_1$ & $\underline{R}_2$ and requires no matrix inversions. To apply the procedure an initial estimate of $\underline{\beta}$ from within the invertibility region is required. The estimate $\underline{\beta} = \underline{0}$ is feasible, or alternatively a closer initial estimate can be provided by a single application of an alternative procedure, such as Walker's. A modified system for the estimator which is pseudoquadratically convergent has been derived by Godolphin (1978a) and a computer implementation by Angell & Godolphin (1978). Other tests for specification involving moving average parameters are described in Section 2.5.

To test hypotheses about the mean of the sample, we recall from Chapter 2 that the distribution of the estimate of μ is

$$\overline{X} \sim N(\mu, (1 + \beta_1 + \dots + \beta_q)^2 \frac{\sigma^2}{n}) .$$

To test the hypothesis that $\mu = 0$ we formulate the statistic

$$Q = \frac{(\overline{X})^2 n}{\hat{\sigma}^2 (1 + \hat{\beta}_1 + \dots + \hat{\beta}_q)^2} .$$

In large samples, Q is distributed like χ^2 with one degree of freedom if the hypothesis is true. The parameters $\beta_1,\ldots,\beta_q$ and the variance σ^2 are estimated using maximum likelihood. This test is adapted in the next chapter so as to test for different means within a given sample set.

CHAPTER 7

THE TESTING FOR DEVIATIONS IN STATE SPACE MODELS

7.1 Introduction

In this chapter the application of statistical test procedures to state space models is considered. Of particular interest is a problem which concerns control engineers who are engaged in the active tracking of marine craft. When active tracking is being carried out, a signal is emitted by the own-ship which registers any objects within its range. By determining also the direction of the signal it is thus possible to note the position in cartesian co-ordinates of the object relative to some fixed origin, in this case the own-ship. The aim of any test procedure is to detect quickly and accurately any manoeuvres in the object. It seems intuitively sensible to base any such test on detecting velocity changes. Clearly the object can change direction and still maintain a constant speed, but this manoeuvre is noted as a velocity change, since velocity is defined as speed with direction.

A test based on detecting velocity changes is presented. Given a set of data values, one point is chosen at which it is suspected that a velocity change took place. Then a statistic is formulated which is related to the difference between the sample means before and after the suspected velocity change. Thus we obtain a generalisation of Student's t test. The derivation of the test statistic is given in Section 7.3. In Section 7.4 the results of various simulations using the test are presented.

7.2 The Constant Velocity Model

In tracking an object such as a ship, the output data available to the control engineer will be multivariate. In this case, the cartesian co-ordinates of the object form two data sets, and for this practical situation, we assume that the X and Y co-ordinates are independent of each other. Thus each set of data has a state space representation which is unaffected by the other set, but the analysis of both sets of output is the same.

The state space model for the X co-ordinate is given by

$$X_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_{1t} \\ \theta_{3t} \end{bmatrix} + v_{1t} \qquad (7.2.1)$$

$$\begin{bmatrix} \theta_{1t} \\ \theta_{3t} \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_{1t-1} \\ \theta_{3t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ 0 \end{bmatrix} . \qquad (7.2.2)$$

The components of the system equation have physical interpretations in that θ_{1t} represents the position in the x direction relative to some fixed origin at time t, and $\tau\theta_{3t}$ is the x speed at time t, where τ is the measurement interval. The random component w_{3t} in (7.2.2) takes the value zero. This is because constant velocity means no deviation in velocity whatsoever, not even random deviation. The random components v_{1t} & w_{1t} are assumed to be $N(0,\underline{V})$ & $N(0,\underline{W})$ respectively.

Comparison with equations (6.2.1)' & (6.2.2)' shows that $d=1$ and thus the model defined by equations (7.2.1) & (7.2.2) satisfies the criterion

$$\underline{F}\underline{G}(\underline{G} - \underline{I})^2 = 0$$

where

$$\underset{\sim}{F} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \underset{\sim}{G} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} .$$

The question arises as to which of models (6.2.4) and (6.2.5) is the more appropriate for describing the given system. We shall show that twice differencing leads to a non-invertible moving average process of order 2.

Now from equation (7.2.2),

$$\nabla\theta_{1t} = \tau\theta_{3t-1} + w_{1t}$$

so that

$$\begin{aligned} \nabla^2 X_t &= \nabla^2\theta_{1t} + \nabla^2 v_{1t} \\ &= \nabla(\tau\theta_{3t-1} + w_{1t}) + \nabla^2 v_{1t} \\ &= 0 + w_{1t} - w_{1t-1} + v_{1t} - 2v_{1t-1} + v_{1t-2} . \end{aligned}$$

The covariances are given by

$$\begin{aligned} \gamma_0 &= 2w + 6v \\ \gamma_1 &= -w - 4v \\ \gamma_2 &= v \end{aligned}$$

where $w = \mathrm{var}(w_1)$, $v = \mathrm{var}(v_1)$ and thus

$$\rho_1 = \frac{-(w + 4v)}{2w + 6v} , \quad \rho_2 = \frac{v}{2w + 6v} .$$

We derive the corresponding parameters for the associated MA(2) process. If β_1, β_2 denote the moving average parameters, then

$$\begin{aligned} B(z) &= 1 + \beta_1 z + \beta_2 z^2 \\ &= (1 - z_1 z)(1 - z_2 z) \end{aligned}$$

and

$$C(z) = 1 + \rho_1(z + z^{-1}) + \rho_2(z^2 + z^{-2})$$

$$= (1+\beta_1^{\,2}+\beta_2^{\,2})^{-1}(1+\beta_1 z+\beta_2 z^2)(1+\beta_1 z^{-1}+\beta_2 z^{-2})$$

We see that $z=1$ is a double root of $C(z)=0$ since

$$C(1) = 1 - \frac{(w+4v)}{2w+6v}.2 + \frac{v}{2w+6v}.2 = 0 .$$

Furthermore we see that

$$z + z^{-1} = 2 + \frac{w}{v}$$

provides the other roots of $C(z)=0$. In this case,

$$z^2 + z^{-2} = (z + z^{-1})^2 - 2 = 2 + \frac{w^2}{v^2} + \frac{4w}{v}$$

and on substituting in $C(z)$ we have

$$1 - \frac{(w+4v)}{2w+6v}\left(2 + \frac{w}{v}\right) + \frac{v}{2w+6v}\left(2 + \frac{w^2}{v^2} + \frac{4w}{v}\right) = 0 .$$

Consequently

$$B(z) = (1 - z)(1 - \beta_2 z)$$

where β_2 is the solution of

$$\beta_2 + \frac{1}{\beta_2} = 2 + \frac{w}{v}$$

which has modulus less than unity. This solution for β_2 is easily seen to be

$$\beta_2 = \frac{w+2v}{2v} - \frac{1}{2v}\{w(w + 4v)\}^{\frac{1}{2}}$$

whence

$$\beta_1 = -(1 + \beta_2)$$

$$= -\frac{(w+4v)}{2v} + \frac{1}{2v}\{w(w + 4v)\}^{\frac{1}{2}} .$$

Since $B(z)$ has an inverse zero on the unit circle, the associated MA(2) process is not invertible. Thus the state space model defined by equations (7.2.1) & (7.2.2) is more appropriately described by a time series model of the form

$$\nabla X_t = \mu + \varepsilon_t + \beta\varepsilon_{t-1} . \qquad (7.2.3)$$

Differencing the system equation (7.2.1) once gives

$$\nabla X_t = \tau\theta_{3t-1} + w_{1t} + v_{1t} - v_{1t-1} . \qquad (7.2.4)$$

By definition of the model, θ_{3t} is free from error and so can be regarded as a deterministic parameter. Thus the mean, μ, in (7.2.3) can be represented entirely by the term $\tau\theta_{3t} = \tau\theta_{3t-1}$, in (7.2.4). Thus equation (7.2.4) can be re-written as

$$\nabla X_t - \mu = w_{1t} + v_{1t} - v_{1t-1} . \qquad (7.2.5)$$

Multiplying both sides of (7.2.5) by $\nabla X_t-\mu$ and $\nabla X_{t-1}-\mu$ in turn, and taking expectations gives

$$E\left[(\nabla X_t - \mu)^2\right] = var(w_1) + 2\, var(v_1)$$

$$E\left[(\nabla X_t - \mu)(\nabla X_{t-1} - \mu)\right] = - var(v_1) .$$

Thus the theoretical correlation is

$$\rho = \frac{- var(v_1)}{var(w_1) + 2var(v_1)} = \frac{-1}{2 + R} \quad \text{say,}$$

where $R = \text{var}(w_1)/\text{var}(v_1)$. Hence $\beta/(1+\beta^2) = -(2+R)^{-1}$, yielding

$$\beta = -2\left[2 + R + (R^2 + 4R)^{\frac{1}{2}}\right]^{-1}$$

as the invertible solution of the quadratic equation for β.

In the simulations described in Section 7.4, the standard deviations of v_1 and w_1 are varied. Thus the quantity of interest in the analysis of the results is the ratio of these standard deviations, namely $R^{\frac{1}{2}}$, and not R itself. It is interesting to know what theoretical value of β corresponds to a particular value for $R^{\frac{1}{2}}$. The table at the end of this section gives the complete range of $R^{\frac{1}{2}}$, together with the resultant theoretical values for β and the correlation ρ. The theoretical value for the variance $\text{var}(\varepsilon) = \sigma^2$ can be expressed in several equivalent ways, namely

$$\sigma^2 = -\frac{\text{var}(v_1)}{\beta} = \frac{\text{var}(w_1)}{(1+\beta)^2}$$

and on substituting in either form for β, it is possible to express σ^2 wholly in terms of $\text{var}(v_1)$ and $\text{var}(w_1)$.

To complete this section the constant velocity model for the Y co-ordinate is given by

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_{2t} \\ \theta_{4t} \end{bmatrix} + v_{2t}$$

$$\begin{bmatrix} \theta_{2t} \\ \theta_{4t} \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_{2t-1} \\ \theta_{4t-1} \end{bmatrix} + \begin{bmatrix} w_{2t} \\ 0 \end{bmatrix} .$$

The analysis of the Y output proceeds in exactly the same way as that described previously for the X output, with R denoting the ratio $\text{var}(w_2)/\text{var}(v_2)$.

$R^{\frac{1}{2}} = \dfrac{sd(w_1)}{sd(v_1)}$	$R = \dfrac{var(w_1)}{var(v_1)}$	β	ρ
0.00	0.0000	-1.0000	-0.5000
0.20	0.0400	-0.8190	-0.4902
0.25	0.0625	-0.7793	-0.4848
0.30	0.0900	-0.7416	-0.4785
0.35	0.1225	-0.7059	-0.4711
0.40	0.1600	-0.6721	-0.4630
0.50	0.2500	-0.6096	-0.4444
0.60	0.3600	-0.5536	-0.4237
0.70	0.4900	-0.5034	-0.4016
0.80	0.6400	-0.4584	-0.3788
0.90	0.8100	-0.4181	-0.3559
1.00	1.0000	-0.3820	-0.3333
1.20	1.4400	-0.3206	-0.2907
1.40	1.9600	-0.2711	-0.2525
1.60	2.5600	-0.2310	-0.2193
1.80	3.2400	-0.1983	-0.1908
2.00	4.0000	-0.1716	-0.1667
3.00	9.0000	-0.0917	-0.0909
∞	∞	0.0000	0.0000

7.3 Mean-Difference Test for Constant Velocity

If a velocity change has taken place, then the mean values of the process before and after this change point will be different. It seems intuitively sensible therefore to divide the output at the point of the suspected velocity change and to base a test on the estimated mean values for each subset of data.

Let $\underline{Z} = (z_1\ z_2\ \ldots\ z_{n+2})$ denote $n+2$ observations of the process. After differencing, the data is divided into two uncorrelated sets, not necessarily of equal size

$$\underline{X}_1 = (\nabla z_2\ \nabla z_3\ \ldots\ \nabla z_{n_1+1}) = (X_1\ X_2\ \ldots\ X_{n_1})$$

$$\underline{X}_2 = (\nabla z_{n_1+3}\ \nabla z_{n_1+4}\ \ldots\ \nabla z_{n+2}) = (X_{n_1+1}\ X_{n_1+2}\ \ldots\ X_n)$$

with $n_2 = n-n_1$ points in $\underline{X}_2$ and where the interval τ between observations remains constant throughout. The model is defined by

$$X_t = \nabla Z_t = \mu + \varepsilon_t + \beta\varepsilon_{t-1}$$

and the test is based on the assumption that

$$E(X_t) = \begin{cases} \mu_1 & t=1,\ldots,n_1 \\ \mu_2 & t=n_1+1,\ldots,n \end{cases} .$$

It is assumed that the parameter β and variance σ^2 remain the same for both sets. This assumption seems reasonable since β and σ^2 are given solely in terms of the variances of the random components v_{1t} and w_{1t}. Although the components themselves may vary with time, their variances are fixed and are independent of time.

Thus using the result in Chapter 2, the means $\bar{X}_1, \bar{X}_2$ for the two sets are distributed **approximately** as

$$\bar{X}_1 \sim N(\mu_1,\ (1+\beta)^2\sigma^2/n_1)$$

$$\bar{X}_2 \sim N(\mu_2,\ (1+\beta)^2\sigma^2/n_2) .$$

The situation is very similar to that of the classical two sample t test and for n_1 and n_2 large, the random variable

$$\frac{\bar{X}_1-\bar{X}_2}{\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2(1+\beta)^2\right]^{\frac{1}{2}}}$$

is distributed approximately $N\left(\frac{\mu_1-\mu_2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma(1+\beta)}, 1\right)$. Equivalently the random variable

$$K = \frac{(\bar{X}_1-\bar{X}_2)^2 n_1 n_2}{n\,\sigma^2(1+\beta)^2}$$

has asymptotically a (non-central) chi-squared distribution with one degree of freedom. To use this test it is required to estimate β and σ^2 using maximum likelihood. The likelihood takes the form

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}n_1} (\det \underline{\Gamma}_1)^{-\frac{1}{2}} \exp\{\frac{-1}{2\sigma^2}(\underline{X}_1 - \mu_1\underline{1})'\underline{\Gamma}_1^{-1}(\underline{X}_1 - \mu_1\underline{1})\}$$
$$\times\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}n_2} (\det \underline{\Gamma}_2)^{-\frac{1}{2}} \exp\{\frac{-1}{2\sigma^2}(\underline{X}_2 - \mu_2\underline{1})'\underline{\Gamma}_2^{-1}(\underline{X}_2 - \mu_2\underline{1})\} \tag{7.3.2}$$

where $\sigma^2\underline{\Gamma}_1$, $\sigma^2\underline{\Gamma}_2$ are the covariance matrices for the first and second set respectively and $\underline{1}$ is a vector of 1's of length compatable with $\underline{X}_1$ and $\underline{X}_2$. The matrix $\underline{\Gamma}_1$ is of order $n_1 x n_1$ and $\underline{\Gamma}_2$ is of order $n_2 x n_2$; the matrices are of the same form. Thus the joint likelihood is

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}n} (\det \underline{\Gamma}_1 \det \underline{\Gamma}_2)^{-\frac{1}{2}} \exp(-Q_L)$$

where

$$Q_L = \frac{1}{2\sigma^2}\{(\underline{X}_1 - \mu_1\underline{1})'\underline{\Gamma}_1^{-1}(\underline{X}_1 - \mu_1\underline{1}) + (\underline{X}_2 - \mu_2\underline{1})'\underline{\Gamma}_2^{-1}(\underline{X}_2 - \mu_2\underline{1})\} .$$

To maximise the likelihood we make two large sample approximations.

The first is to ignore the contribution to log L of $(\det \underset{\sim}{\Gamma}_1 \det \underset{\sim}{\Gamma}_2)^{-\frac{1}{2}}$ and the second approximation, derived by Whittle (1954, §2.5) and examined further by Shaman (1976), provides approximate inverses for $\underset{\sim}{\Gamma}_1$ and $\underset{\sim}{\Gamma}_2$. Whittle noted that $\underset{\sim}{\Gamma}_1^{-1}$ can be approximated for large n_1 by $\underset{\sim}{\Pi}_1$, where $\underset{\sim}{\Pi}_1 = E(\underset{\sim}{Y}\underset{\sim}{Y}')$ is the covariance matrix for the first order autoregression generated by

$$Y_t + \beta Y_{t-1} = \eta_t$$

where η_t is a sequence of uncorrelated random variables with expectation zero and variance unity. Thus if the covariance matrix for the first set is

$$\sigma^2\underset{\sim}{\Gamma}_1 = \sigma^2 \begin{bmatrix} 1+\beta^2 & \beta & \dots & 0 \\ \beta & 1+\beta^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1+\beta^2 \end{bmatrix}$$

then its inverse is approximated by

$$\sigma^2\underset{\sim}{\Pi}_1 = \frac{\sigma^2}{1-\beta^2} \begin{bmatrix} 1 & -\beta & \dots & (-\beta)^{n_1-1} \\ -\beta & 1 & \dots & (-\beta)^{n_1-2} \\ \vdots & \vdots & & \vdots \\ (-\beta)^{n_1-1} & (-\beta)^{n_1-2} & \dots & 1 \end{bmatrix} .$$

The maximum likelihood estimators of μ_1 and μ_2 are $\bar{X}_1$, $\bar{X}_2$ respectively, where

$$\bar{X}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} X_t \quad , \quad \bar{X}_2 = \frac{1}{n_2} \sum_{t=n_1+1}^{n} X_t \ .$$

Thus the first exponent in the likelihood (7.3.2) is

$$(\underline{X}_1 - \bar{X}_1\underline{1})'\underline{\Pi}_1(\underline{X}_1 - \bar{X}_1\underline{1}) = \frac{n_1}{1-\beta^2}\left(c_{1o} + 2\sum_{k=1}^{n_1-1}(-\beta)^k c_{1k}\right)$$

where

$$c_{1o} = \frac{1}{n_1}\sum_{t=1}^{n_1}(X_t - \bar{X}_1)^2, \qquad c_{1k} = \frac{1}{n_1-k}\sum_{t=1}^{n_1-k}(X_t - \bar{X}_1)(X_{t+k} - \bar{X}_1). \tag{7.3.3}$$

With exactly analagous results holding for the second set, it follows that the maximum likelihood estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\left\{\frac{n_1}{1-\hat{\beta}^2}\left(c_{1o} + 2\sum_{k=1}^{n_1-1}(-\hat{\beta})^k c_{1k}\right) + \frac{n_2}{1-\hat{\beta}^2}\left(c_{2o} + 2\sum_{k=1}^{n_2-1}(-\hat{\beta})^k c_{2k}\right)\right\} \tag{7.3.4}$$

with

$$c_{2o} = \frac{1}{n_2}\sum_{t=n_1+1}^{n}(X_t - \bar{X}_2)^2, \quad c_{2k} = \frac{1}{n_2-k}\sum_{t=n_1+1}^{n}(X_t - \bar{X}_2)(X_{t+k} - \bar{X}_2). \tag{7.3.5}$$

Equivalently,

$$\hat{\sigma}^2 = \frac{1}{n}(n_1\hat{\sigma}_1^{\,2} + n_2\hat{\sigma}_2^{\,2})$$

where

$$\hat{\sigma}_i^{\,2} = \frac{c_{io}}{1-\hat{\beta}^2}\left(1 + 2\sum_{k=1}^{n_i-1}(-\hat{\beta})^k r_{ik}\right) \qquad i = 1,2\ .$$

The maximum likelihood estimator for β is derived from

$$(\underline{X}_1 - \bar{X}_1\underline{1})'\frac{\partial\underline{\Pi}_1}{\partial\beta}(\underline{X}_1 - \bar{X}_1\underline{1}) + (\underline{X}_2 - \bar{X}_2\underline{1})'\frac{\partial\underline{\Pi}_2}{\partial\beta}(\underline{X}_2 - \bar{X}_2\underline{1}) = 0$$

and is given by the recursive equation

$$n_1\left[\beta c_{1o} - \{(1-\beta^2)\sum_{k=1}^{n_1-1}(-\beta)^{k-1}kc_{1k} - 2\beta\sum_{k=1}^{n_1-1}(-\beta)^k c_{1k}\}\right]$$
$$+ n_2\left[\beta c_{2o} - \{(1-\beta^2)\sum_{k=1}^{n_2-1}(-\beta)^{k-1}kc_{2k} - 2\beta\sum_{k=1}^{n_2-1}(-\beta)^k c_{2k}\}\right] = 0\ . \qquad (7.3.6)$$

This yields the estimating equation

$$\beta(n_1c_{1o}+n_2c_{2o})-\{(1-\beta^2)\sum_{k=1}^{m}(-\beta)^{k-1}k(n_1c_{1k}+n_2c_{2K})-2\beta\sum_{k=1}^{m}(-\beta)^k(n_1c_{1k}+n_2c_{2k})\}= 0. \qquad (7.3.7)$$

According to equation (7.3.6) the number, m, of sample serial covariances in (7.3.7) should be n_1-1 or n_2-1. However it appears to be generally accepted that little loss of estimator efficiency occurs if m is of the order of 30, even when n_1, n_2 are quite large; see, for example, Whittle (1954, p.212). The sample serial covariances are present in the estimation equations in products of the form $c_k(-\hat{\beta})^k$ and $c_k k(-\hat{\beta})^{k-1}$. Since $|\beta|<1$ for invertible models, both products rapidly approach zero as k increases, and so their contributions become negligible very quickly.

It is interesting to note that a "weighted" form or "overall" sample serial covariance is obtained by writing them in the form

$$\left.\begin{aligned} c_o &= (n_1c_{1o} + n_2c_{2o})/n \\ c_k &= (n_1c_{1k} + n_2c_{2k})/n\ . \end{aligned}\right\} \qquad (7.3.8)$$

Consequently equation (7.3.4) can be re-written as

$$\hat{\sigma}^2 = \frac{1}{1-\hat{\beta}^2}\left(1 + 2\sum_{k=1}^{m}(-\hat{\beta})^k c_k\right) \qquad (7.3.9)$$

and the estimating equation for β becomes

$$\hat{\beta}c_0 = (1-\hat{\beta}^2)\sum_{k=1}^{m}(-\hat{\beta})^{k-1}kc_k - 2\hat{\beta}\sum_{k=1}^{m}(-\hat{\beta})^k c_k . \qquad (7.3.10)$$

These are the familiar forms for the estimation equations for a single sample from a first order moving average process (cf Whittle 1951, p.82) and described in more detail for a general MA(q) process by Godolphin (1977, 1978a).

Having evaluated the covariances using (7.3.8) and formed estimates of β & σ^2 using (7.3.9) & (7.3.10) the statistic K can be formed and the test is ready for use. The results from various simulations are given in the next section.

7.4 Empirical Results

In this section, the results using the test described in the last section on various sets of simulated data are presented. Each data set either represents constant velocity, or it contains just one velocity change, and by altering the initial and new velocities of the object, it is possible to examine the sensitivity of the test to different magnitudes of velocity change. A comparison is made of the ability of the test to detect velocity increases and decreases of the same order. The remaining factors which will affect the performance of the test are the variances on the noise terms associated with the system and measurement equations. Recall from Section 7.2,

$$(1+\beta^2)\sigma^2 = \mathrm{var}(w_1) + 2\,\mathrm{var}(v_1)$$

$$\beta\sigma^2 = -\,\mathrm{var}(v_1),$$

so that

$$(1+\beta)^2\sigma^2 = \mathrm{var}(w_1).$$

The statistic for the proposed test is

$$K = \frac{(\overline{X}_1 - \overline{X}_2)^2 n_1 n_2}{n\sigma^2(1+\beta)^2}$$

where the parameters β & σ^2 are estimated using maximum likelihood. Note that the random variable K is inversely proportional to the variance of system noise, $var(w_1)$, even though the theoretical value of β depends only on $R^{\frac{1}{2}}$, the ratio of the standard deviations (cf §7.2). Thus the system noise will have a direct bearing on the size of velocity change that we would expect the test to detect. In practice we would expect the variances of system and measurement noise to be approximately equal, since they both represent deviation in metres on distance measurements. Thus in the first half of the results, the ratio, $R^{\frac{1}{2}}$, of standard deviations is fixed at 5/5 = 1.

To illustrate how the test procedure works in practice, the analysis of one pair of X & Y co-ordinate data is given in full. The data file is

6000	5000	Initial object position (X,Y in metres)
10	10	Initial object velocities (X,Y in $m.s^{-1}$)
51		Time of velocity change (in seconds)
10	14	New object velocities (X,Y in $m.s^{-1}$)
1		Measurement interval (in seconds)
5	5	Standard deviations of system noise (w_1, w_2)
5	5	Standard deviations of measurement noise (v_1,v_2)
1		No. of moving average parameters (q)
30		No. of sample serial correlations (m)
100		No. of points in sample (n)

The standard deviations of w_2 & v_2 and also the velocity change in the Y data have no effect on the analysis of the X data, since the two data sets are taken to be independent from the outset.

Consider first the analysis of the X co-ordinate data. Under the null hypothesis it is believed that the data represent constant velocity. Using all n data values, the estimated velocity is 9.578 and this is subtracted from the once-differenced data values, thus yielding the required form to the data for the proposed test. The sample variance and the first thirty sample covariances are given by

$$c_o = \frac{1}{n}\sum_{t=1}^{n}(X_t - 9.578)^2, \quad c_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - 9.578)(X_{t+k} - 9.578).$$

These are then substituted in (7.3.9) & (7.3.10) to give the maximum likelihood estimators of β and σ^2. In this case the estimates are -0.526 & 67.426 respectively, and the first fifteen sample serial correlations $r_k=c_k/c_o$ are given in Table 1 below.

Under the alternative hypothesis it is believed that the first n_1=50 data points have a fixed velocity and the remaining n_2=50 points also assume constant velocity, but of a different magnitude to the first set. The two velocities are estimated by 9.230 & 9.878 respectively. For the first subset

$$c_o = \frac{1}{n_1}\sum_{t=1}^{n_1}(X_t - 9.230)^2, \quad c_k = \frac{1}{n_1-k}\sum_{t=1}^{n_1}(X_t - 9.230)(X_{t+k} - 9.230)$$

whence the sample serial correlations c_k/c_o can be formed based on the data available before the suspected velocity change. The first fifteen sample serial correlations are given in Table 2. Similarly for the subset of data after the suspected velocity change

$$c_0 = \frac{1}{n_2}\sum_{t=n_1+1}^{n}(X_t - 9.878)^2, \quad c_k = \frac{1}{n_2-k}\sum_{t=n_1+1}^{n}(X_t - 9.878)(X_{t+k} - 9.878)$$

with the first fifteen values for r_k given in Table 2. The test requires an "overall" value for each c_k; these are given by equation (7.3.8) which, since $n_1=n_2=50$, is simply the average of the sample serial covariances previously calculated for each subset. The first fifteen "overall" values for r_k are given in the second row of Table 1.

Table 1

Estimated/Modified Sample Serial Correlations r_k

lag	1	2	3	4	5	6	7	8
Under H_0	-0.32	-0.11	-0.05	0.07	0.08	-0.21	0.08	0.06
Under H_1	-0.32	-0.12	-0.06	0.09	0.06	-0.20	0.06	0.10

	9	10	11	12	13	14	15
	-0.10	-0.04	0.02	0.07	0.01	-0.20	0.16
	-0.12	-0.03	0.00	0.09	0.03	-0.25	0.17

A graphical representation of these sample serial correlations under H_0 is provided by Figure 7.1. The r_k's in Table 1 are virtually identical at each lag under the two hypotheses, suggesting that no velocity change has occurred. This belief is supported by the test statistic.

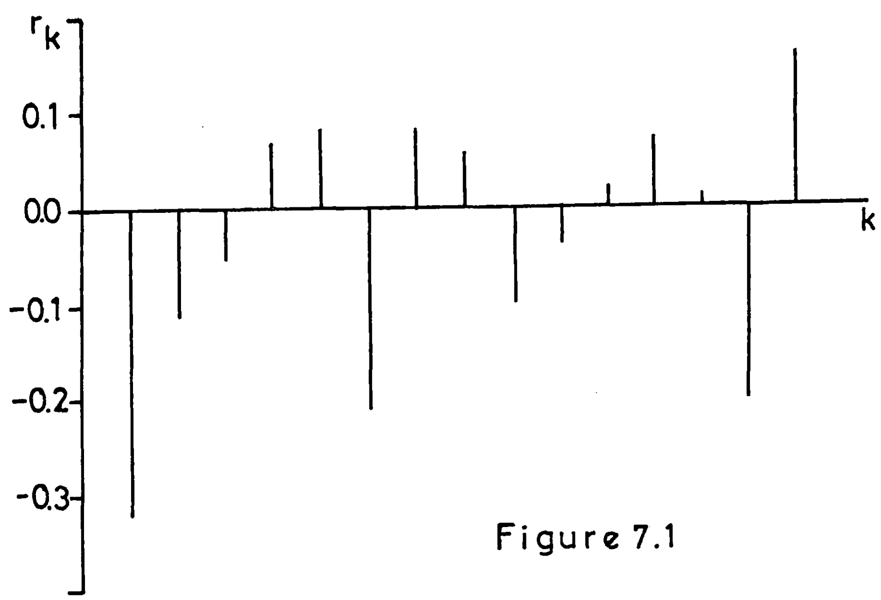


Figure 7.1

Table 2

Sample serial correlations using divided sample

	1	2	3	4	5	6	7	8
Before suspected velocity change	-0.33	-0.18	0.03	0.18	-0.03	-0.34	0.14	0.22
After suspected velocity change	-0.32	-0.05	-0.14	0.00	0.16	-0.05	-0.02	-0.02

9	10	11	12	13	14	15
0.20	-0.03	-0.06	0.13	0.01	-0.24	0.15
-0.05	-0.04	0.06	0.06	0.04	-0.25	0.20

The estimates of β & σ^2 using the "overall" values for c_k are -0.566 and 66.609 respectively. The proposed test statistic is

$$K = \frac{(\bar{X}_1 - \bar{X}_2)^2 \, n_1 n_2}{n\hat{\sigma}^2(1+\hat{\beta})^2}$$

$$= \frac{(9.230-9.878)^2 \; 50 \; 50}{100 \;\; 66.609 \;\; (1-0.566)^2} = 0.84$$

which is not significant. The test concludes correctly that there is no evidence to support the idea of a velocity change.

The analysis of the Y co-ordinate data follows in the same way. Under the null hypothesis the maximum likelihood estimators of β and σ^2 are -0.251 and 83.909 respectively; under the alternative hypothesis the estimates are -0.527 & 70.236. Using all n data values, the velocity estimate is 11.276, whereas for the subsets of n_1 and n_2 points the estimates are 8.774 and 13.637 respectively. Figures 7.2 and 7.3 give the estimated first fifteen terms of the correlogram under the null and alternative hypotheses respectively;

the values are also given in tabular form in Table 3. The sample serial correlations based on the subsets before and after the suspected velocity change are presented in Table 4.

Table 3

Estimated/Modified Sample Serial Correlations

lag	1	2	3	4	5	6	7	8
Under H_0	-0.32	0.21	-0.16	0.24	-0.12	0.10	-0.16	0.18
Under H_1	-0.40	0.12	-0.20	0.13	-0.14	0.00	-0.18	0.13

lag	9	10	11	12	13	14	15
Under H_0	0.08	0.08	-0.06	0.22	-0.01	0.11	-0.08
Under H_1	-0.01	0.00	0.00	0.15	-0.03	0.05	-0.14

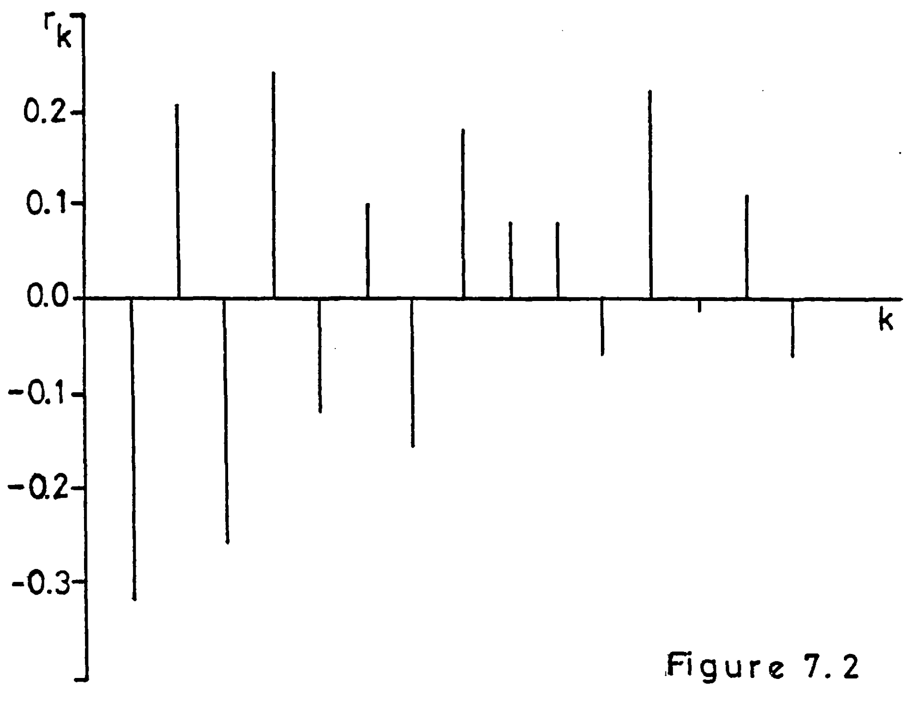


Figure 7.2

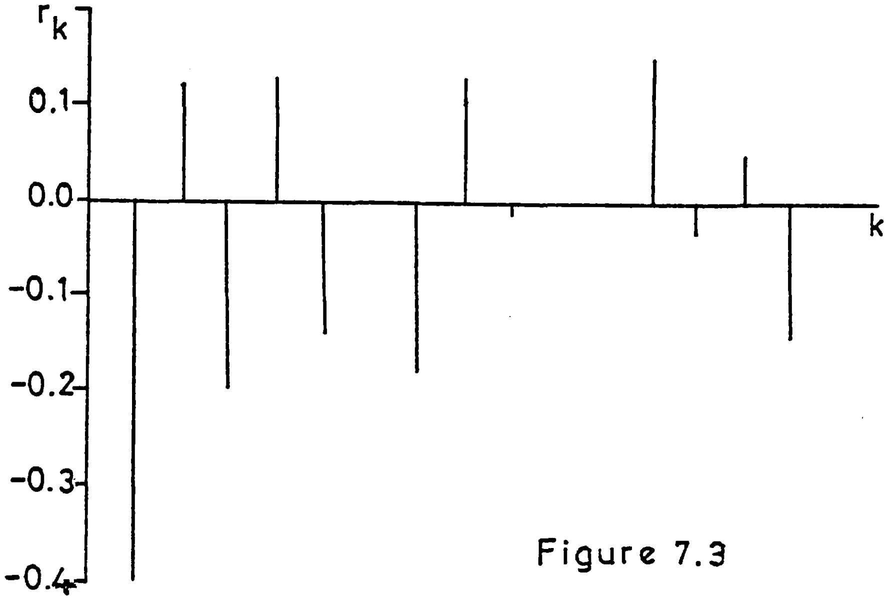


Figure 7.3

Table 4

Sample serial correlations assuming H_1

lag	1	2	3	4	5	6	7	8
Before suspected velocity change	-0.47	0.09	-0.28	0.29	-0.08	-0.04	-0.12	0.11
After suspected velocity change	-0.33	0.14	-0.12	-0.03	-0.19	0.04	-0.24	0.16

lag	9	10	11	12	13	14	15
Before suspected velocity change	0.06	-0.09	-0.08	0.12	-0.05	0.12	-0.05
After suspected velocity change	-0.08	0.09	0.07	0.18	-0.01	-0.02	-0.22

The test statistic for the Y co-ordinate data is

$$K = \frac{(8.774-13.637)^2 \quad 50 \quad 50}{100 \quad 70.236 \quad (1-0.527)^2} = 37.62 \; .$$

In this case the velocity change is detected with a large probability.

It is sensible to use several simulations to assess the value of the test procedure for detecting constant velocity or a velocity increase from 10 $m.s^{-1}$ to 14 $m.s^{-1}$ as described above. It may be that the test can behave unexpectedly on occasions owing to some untypical random disturbances and that this behaviour may not be reproduced on the majority of occasions. An atypical random number seed can produce unusual noise sequences and unfortunately the test appears to fail each time that particular random number seed is used. To balance out such effects, ten different random number seeds are chosen and thus ten sets of data containing the same velocity change are created. The test is then performed on each set in turn and its success rate over the ten runs is noted. A more accurate assessment of the sensitivity of the test can thus be made.

Nine further simulations were constructed of the data set presented earlier in this section. The results are given in Table 5, together with the results described in full for random number seed 19.

Table 5

Random No. seed	Test statistic	
	x data	y data
11	1.31	27.71***
12	1.52	21.97***
13	0.00	4.97**
14	0.08	33.04***
15	0.44	3.14*
16	0.01	22.80***
17	0.42	16.59***
18	2.50	21.25***
19	0.84	37.62***
20	0.16	11.06***

*denotes significant at 10%; **denotes significant at 5%;
***denotes significant at 1%.

Thus the test detects both a constant velocity of 10 m.s^{-1} and the velocity change from 10 m.s^{-1} to 14 m.s^{-1} with a large probability in the majority of simulations, the only possible exception on the Y data being the result obtained using random number seed 15.

Further tests were carried out on data containing different velocity changes. Each test was repeated ten times using random number seeds 11-20. As in the previous example each data set contained 100 points, and the velocity change occurred after 51 seconds. Denoting the X and Y data already discussed as Series I and II, the following six series were also examined:

Series III	Velocity	increase	from	10 m.s^{-1}	to	12 m.s^{-1}
Series IV	"	decrease	"	10 m.s^{-1}	"	5 m.s^{-1}
Series V	"	increase	"	10 m.s^{-1}	"	16 m.s^{-1}
Series VI	"	"	"	5 m.s^{-1}	"	8 m.s^{-1}
Series VII	"	"	"	5 m.s^{-1}	"	10 m.s^{-1}
Series VIII	"	decrease	"	16 m.s^{-1}	"	10 m.s^{-1}

The values obtained for the statistic K are presented in Table 6 below. The results in the table indicate that the velocity change is detected at the 99% significance level in every simulation of Series IV, V, VII & VIII. The random variable K is proportional to the square of the velocity change and the statistics on the whole indicate this. For this reason, the results for Series III & VI are less conclusive than those corresponding to large velocity changes. In fact the Series III statistics detect the velocity change at the 90% significance level in only six simulations out of ten, and that for Series VI in eight cases out of ten. The smaller than average value for K in all the series using random number seed 15 can be attributed to the unusual sequence of random numbers generated by that particular seed. Indeed, this may account for the failure to detect the velocity changes in Series III & VI using seed 15.

Table 6

Seed	Series III	Series IV	Series V	Series VI	Series VII	Series VIII
11	8.43***	10.59***	41.25***	17.59***	30.73***	29.53***
12	5.72**	106.11***	93.04***	13.14***	61.34***	35.20***
13	6.77***	40.14***	59.42***	0.42	41.37***	185.44***
14	3.87*	34.52***	41.58***	19.46***	28.39***	58.11***
15	0.90	22.05***	17.40***	0.52	11.32***	75.61***
16	5.20**	28.59***	44.07***	14.42***	30.79***	25.00***
17	2.42	37.95***	35.56***	9.30***	23.63***	38.04***
18	0.82	60.83***	34.59***	14.30***	21.51***	13.44***
19	13.98***	37.79***	88.14***	23.77***	63.62***	42.04***
20	2.05	24.72***	25.93***	6.07**	17.45***	28.08***

In practice there are two types of velocity change that we wish to detect. The first type results from the object maintaining the same speed but changing its direction. The second type is caused by the object changing its speed and possibly its direction as well. A change in direction cannot be instantaneous; the resultant increase or decrease in velocity, depending on whether the object is moving nearer or further away (relative to the own-ship) will take several seconds in reality, and the test described here is not designed for this type of lag. It is clear that the time taken over the velocity change will have an effect on the ability and sensitivity of the test to detect increases and decreases. The nature of the object (ship, aircraft or submarine) will determine its ability to increase or decrease its speed quickly. It seems that neither type of velocity change will be as discrete as in the simulations described here, but it is reasonable to suppose that the test would still be useful for these problems.

Further examination of the sensitivity of the test is made by considering different values for the standard deviations and examining the effect on detecting the same velocity changes as in Series I to VIII inclusive. By increasing the standard deviation of system noise from 5 to 8, the random variable K is significantly reduced, and this may result in the test not performing as well as previously. With $R^{\frac{1}{2}}=8/5$, the theoretical value for β is -0.2310. This is well within the unit circle, so no problems occur with lack of convergence, and hence a value for K is produced at each simulation. The following results, to be compared with Tables 5 & 6, are obtained with these values.

Table 7a

Seed	Series I	Series II	Series III	Series IV
11	1.27	11.43***	5.26**	3.23*
12	1.34	9.25***	0.95	42.09***
13	0.00	0.06	1.88	13.10***
14	0.11	12.13***	1.05	13.76***
15	0.55	0.03	0.11	11.86***
16	0.01	9.40***	1.81	8.69***
17	0.39	5.39**	0.42	14.39***
18	2.46	12.96***	0.01	27.17***
19	0.63	12.69***	5.96**	11.11***
20	0.22	4.06**	0.51	11.70***

Table 7b

Seed	Series V	Series VI	Series VII	Series VIII
11	21.46***	7.57***	16.39***	8.51***
12	27.43***	5.67**	17.40***	12.62***
13	17.83***	1.46	12.31***	96.28***
14	13.89***	7.48***	9.31***	15.91***
15	6.23**	0.23	3.83*	38.94***
16	14.41***	6.18**	10.13***	7.45***
17	10.14***	3.04*	6.51**	11.97***
18	7.88***	9.37***	4.32**	3.21*
19	32.96***	8.44***	24.17***	9.04***
20	9.47***	2.08	6.18**	13.91***

The effect of increasing the standard deviation of the system noise has been to reduce the sensitivity of the test. The statistics in Tables 7a & 7b are nearly all less than their counterparts in Tables 5 & 6. In a few cases only, this means that the velocity change is no longer detected, even at the 90% significance level. Such cases are apparent from Tables 7a & 7b; it is clear that the problem is more marked when detecting smaller velocity changes.

One problem that might arise using the test is if the subroutine for evaluating β fails to converge. Such difficulties may arise if $R^{\frac{1}{2}}$ becomes small, since this corresponds to theoretical values of β near the boundary of the unit circle. Naturally if β cannot be determined, then σ^2 cannot be estimated either by our procedure, and so in such cases nothing can be said about the object's motion using this test, as the statistic K cannot be formed. It is possible that indirect estimation procedures such as that of Walker (1961), which concentrates on estimating the correlogram efficiently, may circumvent this problem. However, it seems unlikely that this problem would occur frequently in practice.

The proposed test is appropriate for testing a wide range of magnitude of velocity changes, both increases and decreases. The statistic can always be formed, provided the ratio of the standard deviations $(R^{\frac{1}{2}})$ is within the limits of about $\left[\frac{1}{2},2\right]$. If the actual values of these standard deviations are not too high, then the sensitivity of the test appears to be very reasonable.

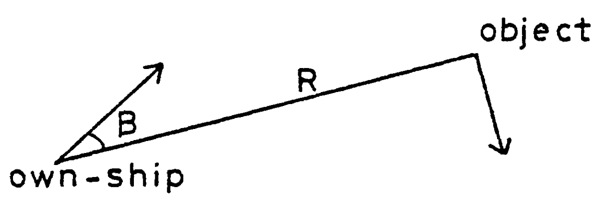
The scope of the test is wider than has been considered here. For instance one would expect, in practice, to have fewer observations available after the velocity change has happened, say $n_1=70$ and $n_2=30$. The procedure is still quite valid although one would expect the power of the test to be reduced in such cases. The requirement that n_2 be reasonably large may be a limiting factor for the test, and the assumption that a velocity change be discrete and immediate may not be realistic, however it is hoped that the test has the scope to cover these possibilities.

CHAPTER 8

THE TESTING FOR DEVIATIONS IN STATE SPACE MODELS USING DATA RELATING TO BEARING AND RANGE

8.1 Introduction

In practice it is often not possible to approach the problem of detecting velocity changes using the method described in the previous chapter. This is because it is not easy for marine navigational devices to measure the position of a neighbouring object in terms of cartesian co-ordinates. It is more natural in active tracking to measure the bearing, B, and the range, R, of the object relative to



the observer. With the own-ship's position assumed known, the range of the object can be deduced by noting how long the signal emitted by the own-ship takes to return there. If the bearing of the signal is also measured, then two pieces of information concerning the object's position are known, and thus, within the accuracy of the measurements, the location of the object in the plane can be deduced.

It is necessary to modify the state space equations to take account of the fact that the data are given in a different form. We derive a test to detect constant velocity and velocity changes by making assumptions about the magnitude and distribution of the noise on the bearing and range measurements. The properties of this test are considered by performing ten simulations and noting on how many occasions the test determines correctly whether or not a velocity change occurred. These results are presented in the final section together with a comparison of the performance of the test using bearing-range data and cartesian co-ordinate data.

8.2 The Constant Velocity Model

The constant velocity model is defined in terms of a set of system and observation equations. The set of system equations is the same as that of Section 7.2; thus the co-ordinate state position (θ_1, θ_2) and velocity (θ_3, θ_4) variables have the state-space representation

$$\begin{bmatrix} \theta_1 \\ \theta_3 \end{bmatrix}_t = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_3 \end{bmatrix}_{t-1} + \begin{bmatrix} w_1 \\ 0 \end{bmatrix}_t$$
$$\begin{bmatrix} \theta_2 \\ \theta_4 \end{bmatrix}_t = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_2 \\ \theta_4 \end{bmatrix}_{t-1} + \begin{bmatrix} w_2 \\ 0 \end{bmatrix}_t \quad . \tag{8.2.1}$$

The additive noise terms w_1, w_2 are present as before so that allowance can be made for the possibility that the position of the own-ship is not known exactly.

The observation equations cannot be the same as previously, however, since we do not observe directly the X and Y measurements. We propose the following model for bearing, B, and range, R :

$$\begin{bmatrix} B \\ R \end{bmatrix}_t = \begin{bmatrix} \tan^{-1}(\theta_1/\theta_2) \\ (\theta_1^2 + \theta_2^2)^{\frac{1}{2}} \end{bmatrix}_t + \begin{bmatrix} b \\ r \end{bmatrix}_t \tag{8.2.2}$$

The random components b_t, r_t represent the additive noise on the bearing and range measurements and are assumed to be independent and to have $N(0, \sigma_b^2)$, $N(0, \sigma_r^2)$ distributions respectively. Thus the data received by the observer consists of two pieces of information at each point in time, at a constant interval, τ. However, in their present forms, neither data set can be used singly to detect velocity changes. This is because either the range or the bearing of the object can remain constant while the object is in fact manoeuvring. If the constant velocity test described in Section 7.3 were applied separately to each set of data then the results would be inconclusive. A natural way of tackling this problem is to transform the bearing and range measurements into estimates of the cartesian co-ordinates by the formulae

$$X_t = R_t \sin B_t \tag{8.2.3}$$

$$Y_t = R_t \cos B_t \quad . \tag{8.2.4}$$

We now consider under what circumstances the first differences ∇X_t and ∇Y_t have the same correlation structure as an invertible first order moving average model with a non-vanishing mean, i.e.

$$\nabla X_t = \mu_x + \varepsilon_{xt} + \beta_x \varepsilon_{xt-1} \tag{8.2.5}$$

$$\nabla Y_t = \mu_y + \varepsilon_{yt} + \beta_y \varepsilon_{yt-1} \tag{8.2.6}$$

where μ_x, μ_y are deterministic terms representing the mean of each process and $\{\varepsilon_{xt}\}$, $\{\varepsilon_{yt}\}$ are independent random processes which are distributed $N(0, \sigma_x^2)$, $N(0, \sigma_y^2)$ respectively.

Consider first the structure of the X co-ordinate data derived from the bearing and range measurements using equation (8.2.3). The Taylor series expansion for $\sin B_t$ is

$$\sin B_t \simeq \sin(B_t-b_t) + b_t\cos(B_t-b_t) - \frac{b_t^2\sin(B_t-b_t)}{2!}$$
$$- \frac{b_t^3\cos(B_t-b_t)}{3!} + \ldots + \frac{b_t^k f^k(\sin(B_t-b_t))}{k!} + \ldots$$

We make the assumption that $|b_t|<1$ so that the expansion provides a valid approximation for $\sin B_t$ using relatively few terms. In examining the correlation structure of $(\nabla X_t - \mu_X)$ we shall assume that the fourth and higher powers of the noise terms b_t and r_t make negligible contributions to the variances and covariances and are thus discarded. By assumption the bearing noise has an $N(0, \sigma_b^2)$ distribution and thus the third moment of b_t is zero. For these reasons we approximate $\sin B_t$ by the first three terms in Taylor's series. From the observation equations,

$$\tan(B_t - b_t) = \frac{\theta_{1t}}{\theta_{2t}}$$

so that

$$\sin B_t \simeq \frac{\theta_{1t}}{R_t^*} + \frac{b_t\theta_{2t}}{R_t^*} - \frac{b_t^2\theta_{1t}}{2R_t^*}$$

where R_t^* is the true, but unobserved range measurement $(\theta_{1t}^2 + \theta_{2t}^2)^{\frac{1}{2}}$ from the equation

$$R_t = R_t^* + r_t \ .$$

Thus

$$X_t = \theta_{1t} + b_t\theta_{2t} - \frac{b_t^2\theta_{1t}}{2} + \frac{r_t\theta_{1t}}{R_t^*} + \frac{r_tb_t\theta_{2t}}{R_t^*} - \frac{r_tb_t^2\theta_{1t}}{2R_t^*} \ .$$

From the system equations,

$$\nabla\theta_{1t} = \tau\theta_{3t-1} + w_{1t}$$

and on substituting $\nabla\theta_{1t}$ in ∇X_t it is seen that $\tau\theta_{3t-1}$ is the only purely deterministic term. Thus the mean, μ_x, in equation (8.2.5) is represented by the term $\tau\theta_{3t-1}$. Then

$$\nabla X_t - \mu_x = w_{1t} + \nabla b_t\theta_{2t} - \frac{\nabla b_t^2\theta_{1t}}{2} + \frac{\nabla r_t\theta_{1t}}{R_t^*} + \frac{\nabla r_t b_t\theta_{2t}}{R_t^*} - \frac{\nabla r_t b_t^2\theta_{1t}}{2R_t^*} .$$

In the definition of the model, b_t and r_t are $N(0, \sigma_b^2)$ & $N(0, \sigma_r^2)$ independent random variables. Thus

$$E(r_t) = E(b_t) = 0$$

$$E(r_t^2) = E(r_{t-k}^2) = \sigma_r^2 \qquad k=1,2,\ldots$$

$$E(b_t^2) = E(b_{t-k}^2) = \sigma_b^2 \qquad k=1,2,\ldots$$

and

$$E(b_t^a r_t^c) = 0 \qquad \text{if } a \text{ or } c \text{ is odd.}$$

It follows that

$$E\left[(\nabla X_t - \mu_x)^2\right] = \text{var}(w_1) + q_t + q_{t-1}$$

where

$$q_t = E\left[\theta_{2t}^2 b_t^2\right] + E\left[\frac{\theta_{1t}^2 r_t^2}{(R_t^*)^2}\right] - E\left[\frac{\theta_{1t}^2 r_t^2 b_t^2}{(R_t^*)^2}\right] + E\left[\frac{\theta_{2t}^2 r_t^2 b_t^2}{(R_t^*)^2}\right]$$

This expression for q_t is unhelpful in this form since it involves θ_1 and θ_2 from the system equations. However it is possible to eliminate the θ's and to write q_t in terms of quantities from the observation equations. Thus

$$q_t = E\left[(R_t-r_t)^2 b_t^{\,2}\cos^2(B_t-b_t)\right] + E\left[r_t^{\,2}\sin^2(B_t-b_t)\right]$$
$$- E\left[r_t^{\,2}b_t^{\,2}\sin^2(B_t-b_t)\right] + E\left[r_t^{\,2}b_t^{\,2}\cos^2(B_t-b_t)\right]. \tag{8.2.7}$$

In order to establish that $(\nabla X_t-\mu_x)$ has the same correlation structure as a first order moving average process with the first covariance having the same parity as in Section 7.2 we need to show that $|\rho| \leqq \frac{1}{2}$ and

$$\begin{aligned} &E\left[(\nabla X_t - \mu_x)^2\right] > 0 \\ &E\left[(\nabla X_t - \mu_x)(\nabla X_{t-1} - \mu_x)\right] < 0 \\ &E\left[(\nabla X_t - \mu_x)(\nabla X_{t-k} - \mu_x)\right] = 0 \qquad k>1 . \end{aligned} \tag{8.2.8}$$

The third of conditions (8.2.8) is satisfied since the k-th covariance

$$E\left[(\nabla X_t - \mu_x)(\nabla X_{t-k} - \mu_x)\right]$$

consists of terms

$$E\left[(\quad)_{t-i}(\quad)_{t-k-j}\right] \qquad i,j=0,1 \quad ; \quad k>1$$

i.e. $(\nabla X_t - \mu_x)$ involves terms with lag t & t-1 and $(\nabla X_{t-k} - \mu_x)$ involves terms with lag t-k & t-k-1. Thus all of the terms of the k-th covariance have zero expectation, so

$$\gamma_k = 0 \qquad k>1 .$$

The first covariance is

$$\gamma_1 = E\left[(\nabla X_t - \mu_x)(\nabla X_{t-1} - \mu_x)\right] = -q_{t-1} .$$

In order that the second of conditions (8.2.8) is satisfied, we require that q_{t-1} be positive. Since q_t and q_{t-1} have the same form it seems likely that they would have the same parity. Re-writing equation (8.2.7) gives

$$q_t = E\left[b_t^2\{(R_t-r_t)^2+r_t^2\}\cos^2(B_t-b_t)\right] + E\left[r_t^2(1-b_t^2)\sin^2(B_t-b_t)\right]$$

which is always positive, since $|b_t|<1$ by assumption. Similarly q_{t-1} is positive, and so all of conditions (8.2.8) are satisfied.

The theoretical correlation is

$$\rho = \frac{-q_{t-1}}{\mathrm{var}(w_1) + q_t + q_{t-1}}$$

$$= \frac{-1}{2+\eta} \quad .$$

Thus $|\rho| \leqq \frac{1}{2}$ as required and hence $(\nabla X_t - \mu_x)$ has the same correlation structure as a first order moving average process. Note however that $\eta = (\mathrm{var}(w_1) + \nabla q_t)/q_t$ may be dependent on t. However, the theoretical values of β

$$\beta = -2\left[2 + \eta + \{(\eta)^2 + 4\eta\}^{\frac{1}{2}}\right]^{-1}$$

and $\sigma^2 = \mathrm{var}(w_1)/(1+\beta)^2$ may contain errors owing to the possible dependence of η on time. These errors may be reflected in the estimates of β and σ^2 required for the test statistic

$$K = \frac{(\bar{X}_1 - \bar{X}_2)^2 n_1 n_2}{n\sigma^2 (1+\beta)^2} \quad .$$

It is hoped that the dependence on t will not affect K unduly. The applicability of this test is considered in the next section.

A similar argument holds for the Y co-ordinate data converted from the bearing and range measurements using equation (8.2.4). In this case the correlation structure of $\nabla Y_t - \mu_y$ is

$$E\left[(\nabla Y_t - \mu_y)^2\right] = \mathrm{var}(w_2) + q_t^* + q_{t-1}^*$$

$$E\left[(\nabla Y_t - \mu_y)(\nabla Y_{t-1} - \mu_y)\right] = - q_{t-1}^*$$

$$E\left[(\nabla Y_t - \mu_y)(\nabla Y_{t-k} - \mu_y)\right] = 0 \qquad k>1$$

where

$$q_t^* = E\left[b_t^2\{(R_t - r_t)^2 + r_t^2\}\sin^2(B_t - b_t)\right] + E\left[r_t^2(1 - b_t^2)\cos^2(B_t - b_t)\right] .$$

The validity of the use of the constant velocity test is difficult to establish because of the dependence of ρ on time. The empirical work in the next section attempts to investigate whether this is a significant problem compared to the effect on the test statistic of a possible change in μ.

8.3 Empirical Results

In order to draw comparisons between the performance of the test using the cartesian data as in Chapter 7 and the cartesian data converted from bearing-range measurements, the velocity changes inherent in Series I-VIII are employed as before. The total number of data points available remains at 100 with the velocity change, if present, occurring at t=51 seconds. There is no reason to suspect that the variances would be different on the X and Y components of the system noise, so the standard deviations of both are fixed at 5, which corresponds with the value in the first half of the results in the previous chapter. The bearing and range components of the observation equations (8.2.2) also have additive noise; for the purpose of these simulations, their standard deviations are fixed at intuitively sensible values such that the variances $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ relating to equations (8.2.5) and (8.2.6) are not too large. For the first set of results on Series I-VIII the standard deviations of the bearing and range noise are 0.07 and 5.00 respectively. In practice this means that $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are of the order of 70, which compares favourably with the corresponding variances in Chapter 7; see, for example, the full analysis of one pair of X & Y co-ordinate data given at the beginning of Section 7.4.

The data file for Series I is

3000	3000	Initial object position (X,Y in metres)
10	10	Initial object velocities (X,Y in $m.s^{-1}$)
51		Time of velocity change (in seconds)
10	14	New object velocities (X,Y in $m.s^{-1}$)
1		Measurement interval (in seconds)
5	5	Standard deviations of system noise (w_1, w_2)
0.07	5	Standard deviations of bearing & range noise (b,r)
1		No. of moving average parameters (q)
30		No. of sample serial correlations (m)
100		Total no. of points in sample (n)

Consider first the analysis of the X co-ordinate data of Series I which represents a constant velocity of 10 $m.s^{-1}$. Under the null hypothesis it is believed that no velocity change takes place. The velocity is estimated to be 9.547 using all n data points, and to produce the required form for the data, the estimate is subtracted from the once differenced data values. The sample variance and covariances are then given by

$$c_o = \frac{1}{n}\sum_{t=1}^{n}(X_t - 9.547)^2, \quad c_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - 9.547)(X_{t+k} - 9.547)$$

and the first fifteen sample serial correlations c_k/c_o are given in Table 1 below. Using equations (7.3.9), (7.3.10), the maximum likelihood estimates of β and σ^2 are -0.557 and 77.074 respectively. These parameter estimates are not required for the test, but it is interesting to compare the values with those obtained under the alternative hypothesis.

It is believed under H_1 that a velocity change occurs and this is tested for at the point $t = 51$. The two subsets of data are treated separately in an identical fashion to the whole data set under the null hypothesis. The velocity estimates are 9.177 and 9.790 respectively and the first fifteen sample serial correlations for each subset are given in Table 2. Since there is an equal number of data points before and after the suspected change point, the "overall" sample serial correlations are simply given by the average over the two subsets. The first fifteen values are given in Table 1. These overall values are then employed in the estimation formulae for β and σ^2. The maximum likelihood estimates are -0.617 and 74.983 respectively.

Table 1

Estimated/Modified Sample Serial Correlations r_k

lag	1	2	3	4	5	6	7	8
Under H_0	-0.40	-0.02	-0.01	0.06	-0.06	-0.02	-0.06	0.11
Under H_1	-0.41	-0.04	0.00	0.04	-0.04	-0.02	-0.04	0.11

lag	9	10	11	12	13	14	15
Under H_0	-0.15	0.11	-0.04	0.01	-0.04	-0.08	0.12
Under H_1	-0.15	0.10	-0.04	0.02	0.00	-0.16	0.14

A graphical representation of these sample serial correlations under H_0 is provided by Figure 8.1.

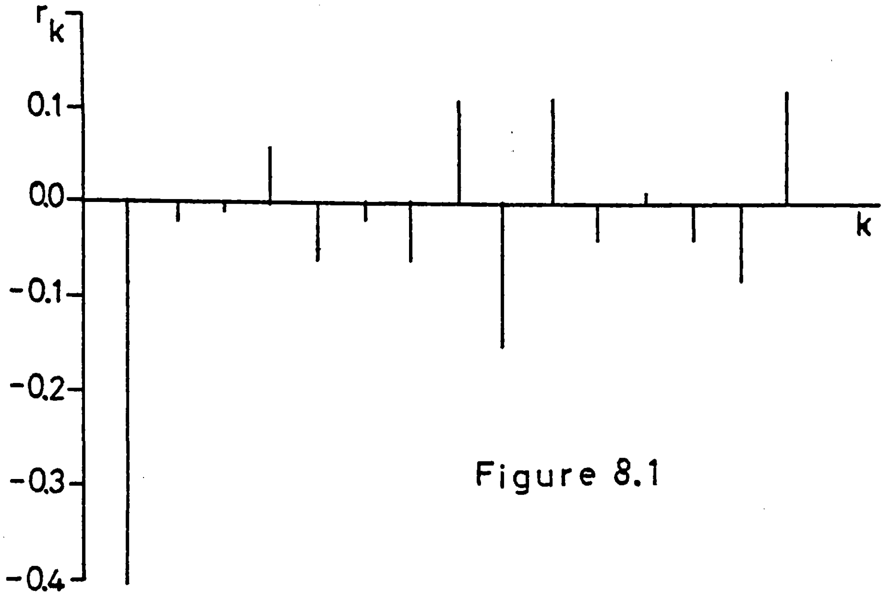


Figure 8.1

Table 2

Sample serial correlations using divided sample

	1	2	3	4	5	6	7	8
Before suspected velocity change	-0.47	0.05	-0.03	0.09	-0.08	-0.13	0.11	0.12
After suspected velocity change	-0.35	-0.13	0.03	-0.01	0.00	0.08	-0.19	0.10

	9	10	11	12	13	14	15
Before suspected velocity change	-0.30	0.23	-0.13	-0.01	0.06	-0.27	0.17
After suspected velocity change	-0.02	-0.03	0.05	0.04	-0.06	-0.04	0.10

Not only are the estimates of β and σ^2 very similar under the two hypotheses, but the sample serial correlations in Table 1 are virtually identical at each lag. This suggests that no velocity change has taken place, and this is clarified by examining the test statistic;

$$K = \frac{(\overline{X}_1 - \overline{X}_2)^2 \; n_1 n_2}{n\hat{\sigma}^2(1+\hat{\beta})^2}$$

$$= \frac{(9.177-9.790)^2 \quad 50 \quad 50}{100 \quad 74.983 \quad (1-0.617)^2} = 0.85$$

which is not significant. The test concludes correctly that there is no evidence to support the belief that a velocity change took place.

The Y co-ordinate data forming Series II are analysed in the same way. The parameters of interest under the null hypothesis are estimated by

$$\overline{Y} = 11.177, \quad \hat{\beta} = -0.288, \quad \hat{\sigma}^2 = 101.887$$

and the first fifteen of the thirty calculated sample serial correlations are given in Table 3 and in diagrammatic form in Figure 8.2.

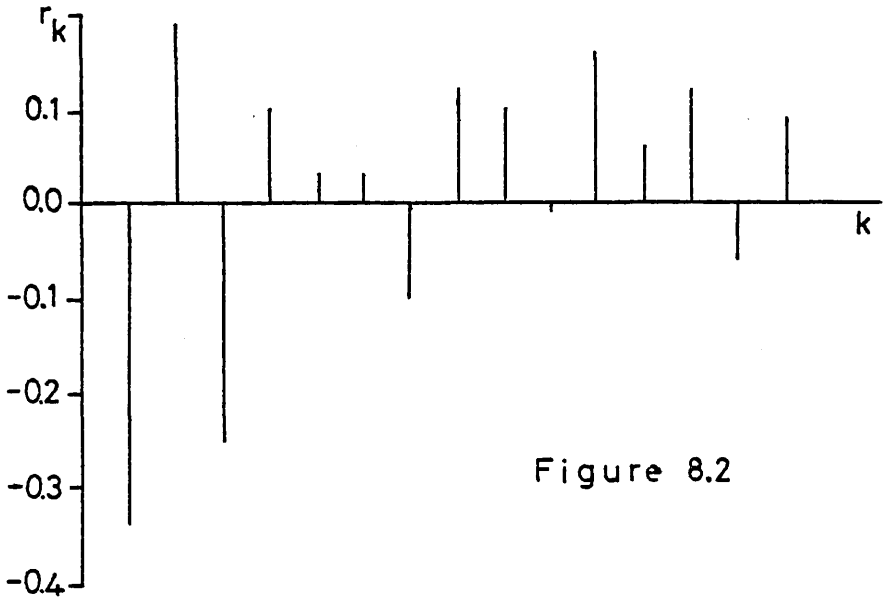


Figure 8.2

Table 3

Estimated/Modified Sample Serial Correlations

lag	1	2	3	4	5	6	7	8
Under H_0	-0.34	0.19	-0.15	0.10	0.03	0.03	-0.10	0.12
Under H_1	-0.41	0.12	-0.20	0.02	-0.01	-0.04	-0.13	0.12

lag	9	10	11	12	13	14	15
Under H_0	0.10	-0.01	0.16	0.06	0.12	-0.06	0.09
Under H_1	0.02	-0.03	0.07	0.01	0.06	-0.07	0.00

Under the alternative hypothesis

$$\bar{Y}_1 = 8.717, \quad \bar{Y}_2 = 13.464, \quad \hat{\beta} = -0.630, \quad \hat{\sigma}^2 = 86.335.$$

The overall sample serial correlations are given in Table 3 and Figure 8.3; those relating to the subsets before and after the suspected velocity change are given in Table 4 overleaf.

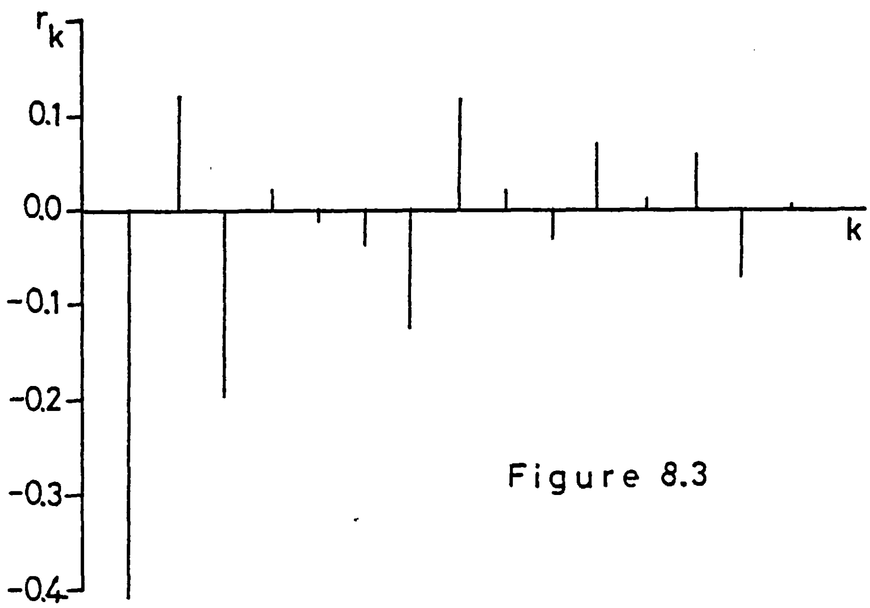


Figure 8.3

Table 4

Sample serial correlations using divided sample

	1	2	3	4	5	6	7	8
Before suspected velocity change	-0.34	-0.11	-0.11	0.15	0.03	-0.08	-0.26	0.29
After suspected velocity change	-0.46	0.32	-0.27	-0.09	-0.03	0.00	-0.01	-0.03

	9	10	11	12	13	14	15
Before suspected velocity change	0.05	-0.10	-0.07	0.05	-0.07	0.21	-0.06
After suspected velocity change	0.01	0.03	0.19	-0.03	0.18	-0.31	0.06

The test statistic for the Y co-ordinate data is

$$K = \frac{(\overline{Y}_1 - \overline{Y}_2)^2 \; n_1 n_2}{n\hat{\sigma}^2(1+\hat{\beta})^2}$$

$$= \frac{(8.717-13.464)^2 \quad 50 \quad 50}{100 \quad 86.335 \quad (1-0.630)^2} = 47.74$$

which is highly significant, as is to be expected on comparing the correlogram in Figure 8.3 to that of Figure 8.2.

In order to draw accurate conclusions about the ability of the test with Series I & II data, nine further simulations were constructed. The results are given in Table 5, together with the results already obtained in the above analysis using random number seed 19.

Table 5

Random No. seed	Test Statistic	
	x data	y data
11	1.60	24.18***
12	0.81	22.47***
13	0.04	3.98**
14	0.23	26.48***
15	0.39	2.83*
16	0.06	36.38***
17	0.46	20.49***
18	3.21*	28.40***
19	0.85	47.74***
20	0.02	9.56***

*denotes significant at 10%; **denotes significant at 5%;

***denotes significant at 1%.

As in the previous chapter, the test detects both a constant velocity of 10 $m.s^{-1}$ and the velocity change from 10 $m.s^{-1}$ to 14 $m.s^{-1}$ with a high probability in most cases. It appears that the constant velocity result for the X co-ordinate data is unperturbed by the velocity change in the Y co-ordinate, even though the data sets are not strictly independent.

The difficulty in drawing a direct comparison between these results and the corresponding results in Chapter 7 is that for cartesian data converted from bearing-range measurements, we cannot establish the theoretical value of σ^2 based on the four standard deviations in the data file. In view of the more complicated nature of the data, we would not expect the test to perform better using converted cartesian data than with pure cartesian co-ordinate measurements. However the results in Table 5 appear to be more conclusive on average in this chapter than in Chapter 7. This can be attributed to the fact that since the random variable K is inversely proportional to σ^2, the standard deviations on the noise components in the data file have a lesser effect as a whole than those in Chapter 7, thus yielding a smaller overall variance. These standard deviations are believed to be set at sensible values, and the results indicate that the test is still suitable for cartesian data converted from bearing-range measurements.

With Series III-VIII defining the same velocity changes as in the previous chapter, i.e.

Series III	Velocity	increase	from	10 $m.s^{-1}$	to	12 $m.s^{-1}$
Series IV	"	decrease	"	10 $m.s^{-1}$	"	5 $m.s^{-1}$
Series V	"	increase	"	10 $m.s^{-1}$	"	16 $m.s^{-1}$
Series VI	"	"	"	5 $m.s^{-1}$	"	8 $m.s^{-1}$
Series VII	"	"	"	5 $m.s^{-1}$	"	10 $m.s^{-1}$
Series VIII	"	decrease	"	16 $m.s^{-1}$	"	10 $m.s^{-1}$

the results using the test are given in Table 6 below. The remaining components in the data file presented earlier in this section are applicable to all the data sets. The results indicate that the velocity change is detected at the 99% significance level in Series IV, V, VII and VIII with every simulation. The results of Series III and VI are slightly less conclusive, as is to be expected, since a smaller velocity change is being examined in these cases.

Table 6

Seed	Series III	Series IV	Series V	Series VI	Series VII	Series VIII
11	9.62***	21.32***	45.71***	14.81***	34.01***	31.87***
12	3.74*	24.41***	55.18***	13.31***	40.23***	36.37***
13	8.14***	182.05***	66.47***	0.07	48.79***	234.17***
14	2.33	33.21***	30.21***	15.66***	21.84***	47.83***
15	0.96	61.46***	18.05***	0.37	11.31***	80.61***
16	5.42**	25.19***	41.46***	23.01***	30.83***	38.47***
17	2.05	34.07***	31.21***	11.29***	22.21***	48.68***
18	0.35	11.72***	28.26***	19.37***	18.33***	18.86***
19	15.43***	38.20***	100.25***	28.85***	70.68***	55.86***
20	2.96*	22.20***	29.37***	4.93**	19.87***	30.23***

The sensitivity of the test is examined further by considering different standard deviations for the noise components in the basic data file. By increasing the standard deviation of the bearing noise from 0.07 to 0.1, decreasing that of the range noise from 0.5 to 0.1 and leaving the system noise variance unaltered, the overall effect is that of an increase in the variance σ^2. Typically the increase is from 70 to 85 in practice. Using Series I-VIII as before, the following values for the statistic K were obtained.

Table 7a

Seed	Series I	Series II	Series III	Series IV
11	1.64	24.36***	9.44***	24.61***
12	1.32	21.24***	5.29**	23.51***
13	0.06	3.51*	9.59***	214.65***
14	0.13	25.56***	3.24*	31.53***
15	0.34	2.37	0.86	56.94***
16	0.02	42.94***	5.96**	30.47***
17	0.48	20.27***	2.71*	34.00***
18	1.10	34.00***	0.99	13.98***
19	1.07	65.37***	15.83***	54.16***
20	0.03	9.47***	2.65	26.19***

Table 7b

Seed	Series V	Series VI	Series VII	Series VIII
11	44.10***	15.02***	33.28***	34.80***
12	77.47***	12.64***	63.38***	34.77***
13	74.78***	0.00	61.11***	268.24***
14	36.73***	15.29***	26.67***	44.33***
15	15.93***	0.26	10.46***	73.15***
16	47.88***	27.22***	36.26***	46.51***
17	38.34***	11.16***	29.20***	48.78***
18	41.94***	23.35***	27.38***	22.21***
19	97.25***	39.63***	73.05***	74.08***
20	27.73***	4.70**	18.46***	35.25***

Although the variance σ^2 is increased, the value of the statistic K has remained about the same, comparing the values in Tables 5 and 6 with their counterparts in Tables 7a and 7b. This is because the only terms in the statistic K which depend on the standard deviations of the noise components are σ^2 and β in the product

$$\hat{\sigma}^2(1 + \hat{\beta})^2 .$$

Not only has σ^2 increased, but so has $|\beta|$, so the product $\hat{\sigma}^2(1 + \hat{\beta})^2$ remains about constant. Thus the sensitivity of the test has been unaffected by the change in the standard deviations. The results in Tables 7a & 7b further support the applicability of this test for cartesian data converted from bearing-range measurements.

One possible effect of different combinations of values for the standard deviations on the noise components is that the subroutine for evaluating β might fail to converge on every simulation. This problem can be overcome by increasing the number, m, of sample serial correlations in the estimation equation for β, equation (7.3.10). In practical situations it may be of prime importance to keep the computer load to a minimum, and so restrict m and maybe also the total number of observations. In such cases it is necessary to balance the risk of the test failing to detect a manoeuvre owing to insufficient data against minimising the calculation time of the computer. The procedure is still quite valid if there are fewer observations after than before the velocity change, but one would expect the power to be reduced. As in the previous chapter, the velocity change will not be as discrete in practical situations as the simulations suggest, but it is expected that the test would have the scope to deal with non-discrete velocity changes in the data.

CHAPTER 9

SUMMARY

The main objective of this work is to present simple procedures for evaluating $\underset{\sim}{V}/n$, the covariance matrix for the efficient estimators of the parameters of stationary linear time series models. The formulation is obviously of interest to practitioners engaged in fitting such models to data, but it is a general problem which seems to have attracted rather little comment in the literature. Two related problems which are also of interest are the specification of the information matrix and the generalised variance. Furthermore a subclass of non-stationary time series models has been shown to have similar properties to polynomial-projecting dynamic linear models under certain conditions. This enables classical statistical tests to be employed as alternatives to state estimation schemes; the usefulness of this result is considered later in relation to the active tracking problems encountered by control engineers.

A very general stationary time series model is the autoregressive moving average (ARMA) model of order (p,q) defined by

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q}$$

where $\{\varepsilon_t\}$ is a sequence of uncorrelated Gaussian random variables with expectation zero and a common variance σ^2. Either p or q may be zero, yielding pure moving average (MA) and pure autoregressive (AR) models respectively. In Chapter 2 the basic statistical properties of these three classes of time series models are established. The estimation of the parameters of these models is briefly considered together with methods for testing for specification. State space models are also defined.

In Chapter 3 the pure autoregressive model is considered. The model is treated separately not only for reasons of its simplicity. The pure MA and AR models are used widely in practice, and the algorithms presented here provide the basis for several of the results given in later chapters where models containing more parameters are considered. Based on a result of Durbin (1959) the covariance matrix $\underline{V}/n$ is expressed in terms of two triangular matrices whose elements are the parameters of the model. The example demonstrates the ease with which $\underline{V}$ can be evaluated; it appears to be quicker to find the information matrix by inverting the result for $\underline{V}$ than to use a Quenouille-type algorithm. The generalised variance is obtained from a factorization of det $\underline{V}$ into four constituent parts. This factorization is also seen to greatly simplify the stationarity conditions of these models. Analogous results exist on the whole for pure moving average models, but for the sake of completeness the results are stated in the final section for an MA(q) process.

The addition of just one moving average parameter to the models of Chapter 3 makes the specification of $\underline{V}$, its inverse and also the generalised variance considerably more complicated. In Chapter 4, we restrict attention to ARMA(p,1) and ARMA(1,q) models. Since p and q are generally quite small in practice, it seems likely that if a pure model is not suitable to the given situation, then this subclass of mixed models may contain the appropriate model. The information matrix is given in a partitioned form with the blocks on the diagonal given by results in the previous chapter. A proof is given of an elegant expression for the generalised variance which is based on augmenting the triangular matrices defined in the previous chapter. This result, together with the concise expression for the off-diagonal block of the information matrix eases the specification of $\underline{V}$. The ability to write $\underline{V}$ explicitly is seen to be of great assistance in examining the adequacy of the fitted model.

The general ARMA(p,q) model is considered in Chapter 5. The off-diagonal block of the information matrix is no longer simply a vector, but a $p \times q$ matrix. The $p+q-1$ different elements can be found individually by applying a Quenouille-type algorithm to Whittle's formula, but this is a very lengthy procedure. By defining two further upper triangular matrices and carrying out some simple products and additions, a $p \times p$ matrix is formed, the inverse of which contains the off-diagonal block of the information matrix in its first q columns. An example shows how easily the formula can be applied. An algorithm is presented for evaluating $\underline{V}$ in a form partitioned conformably with the information matrix; the variances and covariances of the estimators can thus be found in cases where these moments are not too complicated.

In Chapter 6 a class of state space models are examined whose forecast functions are polynomials in the prediction lead time. In the steady state comparisons are made between these models and a class of non-stationary time series models which possess the same property. The degree of differencing required to restore stationarity is equal to or one more than the degree of the polynomial-projecting model. The former model also has a deterministic term, representing the mean of the process. In order to apply inference techniques of time series we require that the model be invertible and this condition will usually mean that only one of the models is appropriate. Procedures for estimating the mean and the moving average parameters of the chosen model are described.

The accurate tracking of manoeuvring objects in the sea or air is of interest to control engineers. One way to detect manoeuvres is to estimate the velocity before and after a suspected manoeuvre in the object and testing the difference for significance. A test statistic is formulated in Chapter 7, based on Student's t test. The test is applied to simulated cartesian co-ordinate data of the object's position relative to the observer, and the ability of the test to detect different manoeuvres is assessed using a wide variety of velocity changes, or no velocity change at all. The simulations give encouraging results, showing that the test is capable of detecting a wide range of manoeuvres.

Chapter 8 examines further the active tracking problem of the previous chapter but using data representing the bearing and range of the object instead of its cartesian co-ordinates. The data in their present form cannot be used to detect velocity changes since either bearing or range may remain constant while the object is in fact manoeuvring. However the bearing and range data can easily be converted to cartesian co-ordinates and then the test can be applied as before. Unfortunately it appears to be difficult to show precisely that the test is still valid for such data, but the results give a strong indication that the test is still an effective method for detecting manoeuvres.

REFERENCES

ALI, M.M. (1977). Analysis of autoregressive-moving average models: Estimation and prediction. Biometrika, 64, 535-545.

ANDERSON, T.W. (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. Ann. Statist., 3, 1283-1304.

ANGELL, I.O. & GODOLPHIN, E.J. (1978). Implementation of the direct representation for the maximum likelihood estimator of a Gaussian moving average process. J. Statist. Comput. Simul., 8, 145-160.

ANSLEY, C.F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process. Biometrika, 66, 59-65.

BARTLETT, M.S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. J.R. Statist. Soc., B, 8, 27-41.

BARTLETT, M.S. & DIANANDA, P.H. (1950). Extensions of Quenouille's test for autoregressive schemes. J.R. Statist. Soc., B, 12, 108-115.

BOX, G.E.P. & JENKINS, G.M. (1970). Time Series Analysis, Forecasting and Control. San Francisco: Holden Day.

BOX, G.E.P. & PIERCE, D.A. (1970). Distribution of autocorrelations in autoregressive integrated moving average time series models. J. Amer. Statist. Ass., 65, 1509-1526.

CLARKE, B.R. & GODOLPHIN, E.J. (1982). Comparative power studies for goodness of fit tests of time series models. J. Time Series Analysis, 3, 141-151.

DURBIN, J. (1959). Efficient estimation of parameters in moving average models. Biometrika, 46, 306-316.

DURBIN, J. (1960). The fitting of time series models. Rev. Inst. Int. Statist., 28, 233-244.

GALBRAITH, R.F. & GALBRAITH, J.I. (1974). On the inverse of some patterned matrices arising in the theory of stationary time series. J. Appl. Prob., 11, 63-71.

GODOLPHIN, E.J. (1976). On the Cramér-Wold factorization. Biometrika, 63, 367-380.

GODOLPHIN, E.J. (1977). A direct representation for the maximum likelihood estimator of a Gaussian moving average process. Biometrika, 64, 375-384.

GODOLPHIN, E.J. (1978a). Modified maximum likelihood estimation of Gaussian moving averages using a pseudo-quadratic convergence criterion. Biometrika, 65, 203-206.

GODOLPHIN, E.J. (1978b). A large-sample test for detecting gaps in moving average models. J.R. Statist. Soc., B, 40, 290-295.

GODOLPHIN, E.J. (1980a). A method for testing the order of an autoregressive moving average process. Biometrika, 67, 699-703.

GODOLPHIN, E.J. (1980b). Estimation of Gaussian linear models. Cahiers du Cero, 243-254.

GODOLPHIN, E.J. (1984). A direct representation for the maximum likelihood estimator of a Gaussian autoregressive moving average process. Biometrika, 71, To appear.

GODOLPHIN, E.J. & STONE, J.M. (1980). On the structural representation for polynomial projecting predictor models based on the Kalman filter. J.R. Statist. Soc., B, 42, 35-46.

GODOLPHIN, E.J. & UNWIN, J.M. (1983). Evaluation of the covariance matrix for the maximum likelihood estimator of a Gaussian autoregressive moving average process. Biometrika, 70, 279-284.

HANNAN, E.J. (1969). The estimation of mixed moving average autoregressive systems. Biometrika, 56, 579-593.

HARRISON, P.J. & STEVENS, C.F. (1976). Bayesian Forecasting (with Discussion). J.R. Statist. Soc., B, 38, 205-247.

HARVEY, A.C. & PHILLIPS, G.D.A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika, 66, 49-58.

JURY, E.I. (1964). Theory and Application of the z-transform method. Wiley.

KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems. Trans. A.S.M.E. Ser. D.J. Basic. Eng. 82, 35-45.

KALMAN, R.E. (1963). New Methods in Wiener Filtering Theory. Proc. 1st. Symp. on Eng. Applications of Random Function Theory and Probability Theory. (J.L. Bogdanoff and F. Kozin, Eds.) Wiley.

KALMAN, R.E. & BUCY, R.C. (1961). New results in linear filtering and prediction theory. Jour. Basic. Eng. (ASME translation). 83D, 95-108.

LJUNG, G.M. & BOX, G.E.P. (1978). On a measure of lack of fit in time series models. Biometrika, 65, 297-303.

LJUNG, G.M. & BOX, G.E.P. (1979). The likelihood function of stationary autoregressive moving average models. Biometrika, 66, 265-272.

LOMNICKI, Z.A. & ZAREMBA, S.K. (1957). On the estimation of autocorrelation in time series. Ann. Math. Statist., 28, 140-158.

MANN, H.B. & WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. Econometrica, 11, 173-220.

NEWBOLD, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process. Biometrika, 61, 423-426.

NICHOLLS, D.F. & HALL, A.D. (1979). The exact likelihood function of multivariate autoregressive moving average models. Biometrika, 66, 259-264.

OSBORN, D.R. (1976). Maximum likelihood estimation of moving average processes. Ann. Econ. Soc. Meas., 5, 75-87.

PAGANO, M. (1973). When is an autoregressive scheme stationary? Comm Statist., 1, 533-544.

PHADKE, M.S. & KEDEM, G. (1978). Computation of the exact likelihood function of multivariate moving average models. Biometrika, 65, 511-519.

PHAM-DINH, T. (1979). The estimation of parameters for autoregressive moving average models from sample autocovariances. Biometrika, 66, 555-560.

PRIESTLEY, M.B. (1981). Spectral Analysis and Time Series, Vol. I. London: Academic Press.

QUENOUILLE, M.H. (1947a). Notes on the calculation of autocorrelations of linear autoregressive schemes. Biometrika, 34, 365-367.

QUENOUILLE, M.H. (1947b). A large sample test for goodness of fit of autoregressive schemes. J.R. Statist. Soc., A, 110, 123-129.

SHAMAN, P. (1976). Approximations for stationary covariance matrices and their inverses, with application to ARMA models. Ann. Statist., 4, 292-301.

SIDDIQUI, M.M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. Ann. Math. Statist. 29, 585-588.

WALKER, A.M. (1952). Some properties of the asymptotic power functions of goodness of fit tests for linear autoregressive schemes. J.R. Statist. Soc., B, 14, 117-134.

WALKER, A.M. (1961). Large sample estimation of parameters for moving average models. Biometrika, 48, 343-357.

WALKER, A.M. (1962). Large sample estimation of parameters for autoregressive processes with moving average residuals. Biometrika, 49, 117-132.

WHITTLE, P. (1951). Hypothesis testing in Time Series Analysis. Uppsala: Almquist and Wiksell.

WHITTLE, P. (1952). Tests of fit in time series. Biometrika, 39, 309-318.

WHITTLE, P. (1953). Estimation and information in stationary time series. Ark. Mat. Fys. Astr., 2, 423-434.

WHITTLE, P. (1954). Some recent contributions to the theory of stationary processes. Appendix 2 in Wold (1954).

WOLD, H.O.A. (1949). A large sample test for moving averages. J.R. Statist. Soc., B, 11, 297-305.

WOLD, H.O.A. (1954). A study in the Analysis of Stationary Time Series. (Second Edition). Stockholm Academic Press.