

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

REFERENCE ONLY

THIS BOOK MAY NOT BE
TAKEN OUT OF THE LIBRARY

University of Southampton, 1987

FACULTY OF MATHEMATICAL STUDIES

DEPARTMENT OF MATHEMATICS

TIME SERIES ANALYSIS OF COMPOSITIONAL DATA

by

Teresa Maria Brunson

Thesis submitted for the degree of Doctor of Philosophy

UNIVERSITY OF SOUTHAMPTON, 1987

SCHOOL OF MATHEMATICAL SCIENCES

DEPARTMENT OF STATISTICS

TIME SERIES ANALYSIS OF COMPOSITIONAL DATA

BY

DR. J. H. J. VAN DER VAART



THANKS TO THE UNIVERSITY OF SOUTHAMPTON FOR THE USE OF THE LIBRARY

UNIVERSITY OF SOUTHAMPTON
FACULTY OF MATHEMATICAL STUDIES
DEPARTMENT OF MATHEMATICS

ABSTRACT

Doctor of Philosophy

TIME SERIES ANALYSIS OF COMPOSITIONAL DATA

by Teresa Maria Brunsdon

In recent years various methods have been developed for modelling multivariate (or vector) time series. However if each vector consists of proportions so that elements must sum to unity these methods break down. Data with this sum-constraint are termed compositional data. It is the aim of this thesis to propose a possible approach to such data.

The method applied is to find a function that will map the sum-constrained data onto an unconstrained space, that is to map the spherical simplex onto the real plane. Two specific mappings are investigated. These turn out to be multivariate generalizations of the well known logistic transformation. However, both of these functions are asymmetrical. For the first this asymmetry is induced by the choice of one of the variables in the vector series, as a reference variable. It is shown that the model is invariant under this choice. For the second, a specific order to the variables must be imposed. However, this is seen to be useful in examining a type of compositional independence known as neutrality.

Methods for using the resulting two models for forecasting are discussed. There are two main problems that occur. The first is that the moments of the underlying distributions corresponding to these models cannot be evaluated algebraically. This means that the minimum mean square error forecast cannot be evaluated. The second is that these distributions are not necessarily uni-modal, which may make the use of the minimum mean square error forecast nonsensical. Various solutions are suggested, and these are compared in a short numerical study.

The final part of the thesis examines the relationships between the components of the proportions. This utilizes time series methods for examining Wiener-Granger causality, and combines them with various concepts of compositional independence. These latter concepts include neutrality as mentioned above, and have been developed to deal with the sum-constraint.

ACKNOWLEDGEMENTS

No piece of work this size can be accomplished without the aid and support of others. I have greatly benefited from discussions with other colleagues. In particular I would like to express my deep appreciation to Fred Smith and Chris Skinner, for their help and encouragement throughout the course of this project.

Also I would like to thank those who have helped in the final stages of this piece of work. James Hepworth for his aid in spelling, Kelvin Bampfield for kindly producing some diagrams, Anne Owens for helping with the typing and showing me how to use (and abuse?) the word processor; and Alan Wood for binding it all together.

Finally I would like to thank my family and friends for their continual encouragement.

"To Him who is able to do immeasurably more
than all we could ask or imagine"

-many thanks.

CONTENTS

	<u>Page Number</u>
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
CONTENTS	iv
CHAPTER 1 INTRODUCTION	1
1.0 Introduction	1
1.1 Notational Conventions	4
CHAPTER 2 MULTIVARIATE TIME SERIES MODELS	5
2.0 Introduction	5
2.1 The ARMA Model	5
2.2 Identification	9
2.3 Estimation	18
2.4 Diagnostic Checks	21
2.5 Forecasting	22
2.6 Independence Between Time Series	27
2.6.1 Some Definitions of Causality	28
2.6.1 Tests and Measures for Causality, Feedback and Independence	31
2.7 Summary	38
CHAPTER 3 COMPOSITIONAL DATA	39
3.0 Introduction	39
3.1 A Formal Description of Compositional Data	40
3.2 The Nature of Compositional Data	44
3.3 Compositional Time Series	50
3.4 The Proposed Approach	56
3.5 Independence of Compositional Data	62
3.6 Summary	74
CHAPTER 4 TRANSFORMED MULTIVARIATE TIME SERIES MODELS	75

4.0	Introduction	75
4.1	Compositional Time Series Models	75
4.2	A Linear Approximation to the Model	82
4.3	Invariance Properties	93
4.4	An Example of Analysis on Two Compositional Time Series	113
4.5	Summary	132
CHAPTER 5	DISTRIBUTIONAL PROPERTIES OF LOGISTIC TIME SERIES MODELS IN RELATION TO FORECASTING	133
5.0	Introduction	133
5.1	Some Basic Properties of the Logistic-Normal Distribution	134
5.1.1	"Marginal" Distribution Property	134
5.1.2	"Conditional" Distribution Property	139
5.2	Location Parameters	140
5.2.1	The Median	141
5.2.2	The Mode	141
5.2.3	The Mean	148
5.2.4	A Numerical Study of the $L_m(\underline{\mu}, \underline{\Sigma})$ Distribution	154
5.3	Some Examples of the $M_m(\underline{\mu}, \underline{\Sigma})$ Distribution	172
5.4	Confidence Region Based on the Logistic-Normal Distribution	176
5.5	A Simulated Example	178
5.6	Summary	187
CHAPTER 6	DEPENDENCE AND INDEPENDENCE IN COMPOSITIONAL TIME SERIES	188
6.0	Introduction	188
6.1	Extrinsic Analysis	188
6.2	Intrinsic Analysis	207
6.2.1	Complete Subcompositional Independence	208
6.2.2	Partition Independence Properties	215

6.2.3 Progressing Partitions	237
6.2.4 Some Numerical Examples Of Partition Independence	240
6.5 Summary and Conclusion	249
CHAPTER 7 CONCLUSION	250
7.0 Conclusion	250
7.1 Further Work	252
APPENDIX A	254
REFERENCES	278

"In the beginning..."

John 1:1

CHAPTER 1

Introduction

1.0 Introduction

This thesis is concerned with the analysis of "compositional time series". A compositional time series is a sequence of observations at discrete time points; $\dots -1, 0, 1, \dots$ where each observation consists of a vector of proportions. For example consider a repeated survey in which at each repetition a categorical variable is recorded, such as in an attitudinal survey where a number of people are questioned on their attitude to some issue. Their reply might be coded as one of:-

for; indifferent; against; don't know.

For each survey repetition this gives a vector of proportions of the respondents preferring different categories. Two specific examples where such data arise are public opinion polls on preference for a political party and market research questionnaires on preference for particular brands.

There are many other examples of such data. The term "compositional" comes from the field of geology, in which constituents of the soil are examined. When soil samples are taken the amount of each constituent (e.g. silt, sand or clay) contained in the sample must necessarily be presented as a vector of proportions, i.e. the composition of the sample. Often several such samples are taken along a line segment and the resulting data is then a compositional "time" series, where "time" is now represented by distance. Another example is that of a chemical process in which the amount of each compound is

measured as the process develops. In a closed system we may obtain a compositional time series which represents the chemical process. A final example occurs in economics in a study of expenditure and income. In a household the expenditure on certain commodities may be examined over several time periods. When expenditure on say "food", "housing", "clothing", etc. is presented as a proportion of total expenditure we have a compositional time series.

The interest of the analyst in such data will vary from application to application. One common requirement is to be able to forecast such series. In the example of the repeated survey a forecast using past surveys may be incorporated into the estimation of the composition from the current survey thus augmenting the accuracy of the estimates. Such an application has already been developed for univariate time series by Blight and Scott(1973), Scott and Smith(1974), Scott et al(1977) and Smith(1978). Forecasting is useful in its own right since it helps predict the future of the system.

Other interests refer more to the structure of the compositional time series. In a study of expenditure one may wish to see how income affects expenditure. If the main wage earner receives a pay rise how does expenditure change (if at all)? Or, in the public opinion poll, does the proportion who prefer a minor party affect the proportion preferring one of the major parties over its main rival?

It is the intention of this thesis to develop models for compositional time series and methods to answer some of these questions. This involves building on the techniques that are available to handle compositional data, together with those for general multivariate time series. The literature on the analysis of compositional data is summarised in Aitchison(1986) and, has to our knowledge been confined only to the case of independent observations. Time series analysis is a vast subject for which there is a large quantity of literature which reveals several possible approaches. We will restrict ourselves to the multivariate ARMA models since they represent a well known and useful class of models for our purposes.

We begin in chapter 2 with a brief description of the multivariate ARMA model and some of its properties. We also consider the concept of dependence between time series, such as Wiener-Granger causality, and discuss methods to test for such dependencies.

In chapter 3 we examine the properties of compositional data. Many problems are found to occur, especially in understanding the relationships between different variables. The main reason behind these difficulties lies in the "sum-constraint", i.e. the fact that the data consists of proportions and hence must sum to one. This makes it difficult to examine such well used statistics as the correlation coefficient. It is also shown how a multivariate ARMA model breaks down when it is applied directly to a compositional data set. We then examine some of the procedures developed to overcome the problems encountered in the non-time-series context, and in particular, the approach of Aitchison(1982) who defines several types of independence between the variables in a composition. His approach is seen to consist of transforming the data so as to remove the sum-constraint. It is this approach that we shall pursue in later chapters.

In chapter 4 some models for compositional time series are presented. They employ the transformation suggested by Aitchison(1982), which requires the use of a reference variable. We show that the models are invariant to the choice of reference variable, the model being affected by a linear transformation of its parameters. A linear approximation to the model is examined and shown to be a standard multivariate ARMA model. Finally the model is applied to two data sets; the GALLUP(c) political opinion poll and the National Opinion Poll (NOP).

The transformations used in chapter 4 lead to a particular transformed normal class. This so called "logistic normal distribution" was first highlighted by Aitchison and Shen(1980). In chapter 5 we examine the mean, median and mode of this distribution. It is seen that numerical methods must be applied to calculate these parameters and consequently an approximation to the mean is derived. Various other properties of this distribution are discussed and a number of numerical examples are given. The applications of the properties of this distribution to the forecasting problem are then discussed.

In chapter 6 the concepts of Wiener-Granger causality discussed earlier in chapter 2, and the compositional independence properties of Aitchison(1982) are combined to form some new concepts of causality and dependence between compositional time series variables. The application of these new types of independence are discussed, and various means for testing them are developed. Finally a numerical

example is given, again using the political opinion poll data.

The final chapter summarizes the main ideas and indicates further areas of possible development.

1.1 Notational Conventions

To make comprehension easier, certain conventions are adhered to wherever possible. No distinction is made between a random variable and the values it takes. Vectors and scalars are denoted by small letters and matrices by capitals. Hence:-

ϕ is a scalar;

\underline{v}_t is a vector with i^{th} element v_{it} ;

\underline{A}_k is a matrix with $(i, j)^{\text{th}}$ element $a_{ij}(k)$, equivalently we will write $\{a_{ij}\} = \underline{A}$;

$\text{dg}(g_1, \dots, g_m)$ will denote the $m \times m$ diagonal matrix with elements g_1, \dots, g_m ;

\underline{U}_m will denote the $m \times m$ matrix of 1's; and

\underline{e}_m will denote the $m \times 1$ vector of 1's.

"For everything there is a season and a time for
every matter under heaven...."

Ecclesiastes 3:1

CHAPTER 2

Multivariate Time Series Models

2.0 Introduction

A comprehensive review of multivariate time series models is beyond the scope of this thesis. We restrict ourselves to only the results needed in subsequent chapters, and in general present these results in as brief a way as possible. A more detailed description may be found in for example, Harvey(1981), Fuller(1976) and Hannan(1970).

2.1 The ARMA Model

We begin by defining the models of interest, which are multivariate generalizations of the univariate models made popular by Box and Jenkins(1976). We will denote by $\{X_t\}$ a vector time series of stochastic variables for $t=\dots,-1,0,1,\dots$

Definition 2.1

A $m \times 1$ vector time series $\{\underline{x}_t\}$ is said to be an m -dimensional white-noise process if the \underline{x}_t are independently and identically distributed with mean $\underline{0}$, and non-singular covariance matrix $\underline{\Sigma}$.

Definition 2.2

A $m \times 1$ vector time series $\{\underline{v}_t\}$ is said to be a multivariate autoregressive moving-average process of dimension m and order (p,q) if it may be written:

$$\underline{v}_t + \underline{\phi}_1 \underline{v}_{t-1} + \dots + \underline{\phi}_p \underline{v}_{t-p} = \underline{\varepsilon}_t + \underline{\theta}_1 \underline{\varepsilon}_{t-1} + \dots + \underline{\theta}_q \underline{\varepsilon}_{t-q} ;$$

where $\underline{\phi}_1, \dots, \underline{\phi}_p, \underline{\theta}_1, \dots, \underline{\theta}_q$ are $m \times m$ constant matrices,

$\Phi_p \neq 0$, $\Theta_q \neq 0$, and $\{\underline{\varepsilon}_t\}$ is an m -dimensional white-noise process. We will denote this type of process by $ARMA_m(p,q)$ and note that we may rewrite it as:

$$\Phi(B)\underline{v}_t = \Theta(B)\underline{\varepsilon}_t, \quad (2.1.1)$$

where
$$\Phi(B) = \underline{I}_m + \Phi_1 B + \dots + \Phi_p B^p, \quad (2.1.2)$$

$$\Theta(B) = \underline{I}_m + \Theta_1 B + \dots + \Theta_q B^q, \quad (2.1.3)$$

and B is the backshift operator:

$$B^k X_t = X_{t-k}.$$

Definition 2.3

An $ARMA_m(p,0)$ process is said to be an autoregressive process of order p and dimension m . We will denote this by $AR_m(p)$.

Definition 2.4

A moving-average process of dimension m and order q is an $ARMA_m(0,q)$ process. We may write this as $MA_m(q)$.

An important property of time series is that of stationarity:-

Definition 2.5

A stochastic process $\{\underline{Y}_t\}$ is weakly (or second-order) stationary if

$$i) E[\underline{Y}_t] = \underline{\mu},$$

$$t = \dots, -1, 0, 1, \dots$$

and
$$ii) \text{Cov}(\underline{Y}_t, \underline{Y}'_{t-k}) = \underline{\Gamma}_k; k = \dots, -1, 0, 1, \dots$$

where $\underline{\mu}$ and $\underline{\Gamma}_k$ are independent of t .

Lemma 2.6

A necessary and sufficient condition for an $ARMA_m(p,q)$ process to be (weakly) stationary is that the roots of:

$$|\Phi(z^{-1})| = 0,$$

lie inside the unit circle.

Proof

e.g. Hannan(1970) page 14.

If we have a nonstationary vector process we may transform the data. For example, as in the univariate case, we could difference the data:-

$$\underline{W}_t = (\underline{I}_m - \underline{B})^d \underline{v}_t,$$

where \underline{B} is the $m \times m$ matrix difference operator :- $\underline{B}^s \underline{x}_t = \underline{x}_{t-s}$. With obvious notation we may then proceed to define the $ARIMA_m(p,d,q)$ process. However differencing has to be applied with great care, if at all, especially when handling multivariate time series. This fact has been noted by among others Hillmer and Tiao(1979), Tiao and Box(1981) and in more detail by Lütkepohl(1982). Tjøstheim and Paulsen(1982) suggest using the concept of, "Almost Non-Stationary" (ANS). Since such transformations are available we will without loss of generality assume stationarity.

Definition 2.7

If an $ARMA_m(p,q)$ process may be written as an infinite autoregressive process it is said to be invertible.

Lemma 2.8

A necessary and sufficient condition for an $ARMA_m(p,q)$ process to be invertible is that the roots of:

$$|\underline{\Theta}(z^{-1})| = 0,$$

lie inside the unit circle.

Proof

e.g. Hannan(1970).

In our definitions we have not included a term to represent the mean of the series. There are two ways of incorporating it into the model. Consider the series $\{\underline{v}_t\}$ and let

$$\underline{v}_t = \underline{v}_t - E[\underline{v}_t] \tag{2.1.4}$$

where \underline{v}_t follows an $ARMA_m(p,q)$ process. Then \underline{v}_t also follows an $ARMA_m(p,q)$ process, with mean $E[\underline{v}_t]$.

Alternatively we may include the mean in the model:-

$$\underline{\Phi}(B)\underline{v}_t = \underline{\mu} + \underline{\Theta}(B)\underline{\varepsilon}_t, \tag{2.1.5}$$

where in general $E[\underline{v}_t] \neq \underline{\mu}$.

Some packages assume this latter form of model, unless the series was first differenced.

Another extension to the $ARMA_m(p,q)$ model is the seasonal model. These models are required because data often comes in the form of, for example, monthly or quarterly figures. A seasonal $AR_m(p)$ model, with seasonality s and autoregressive component P may be written as

$$\underline{\Phi}^{(s)}(B^s)\underline{\Phi}(B)\underline{v}_t = \underline{\varepsilon}_t, \quad (2.1.6)$$

where $\underline{\Phi}(B^s)$ is a polynomial of degree P in B^s .

One may expand the left hand side of (2.1.6) into one polynomial in B of order $sP+p$, i.e. as an $AR_m(sP+p)$ model, but where some of the resulting $\underline{\Phi}_i$ matrices will be zero. Consequently we will without loss of generality consider only non-seasonal models.

Before considering the practicalities of the ARMA model we need to know the conditions under which the model is identified. Here we are addressing identification in the economic sense of parameter redundancy, as opposed to the identification stage of fitting an $ARMA_m(p,q)$ process, which will be discussed in section 2.2. It has been shown by Hamman (1969) that an $ARMA_m(p,q)$ process is identified if the following conditions hold:-

Conditions 2.9

- i) $\underline{\Phi}_0 = \underline{\Theta}_0 = \underline{I}_m$.
- ii) The roots of $|\underline{\Phi}(z^{-1})|=0$, and $|\underline{\Theta}(z^{-1})|=0$ must lie within the unit circle. (i.e. the model must be both stationary and invertible.)
- iii) $\underline{\lambda}(\underline{\Phi}_p - \underline{\Theta}_q) = \underline{0}$ if and only if $\underline{\lambda} = \underline{0}$. i.e. the matrix $[\underline{\Phi}_p | \underline{\Theta}_q]$ is of full rank.

Box and Jenkins(1976) consider their application of ARMA models as consisting of four stages:-

- i) Identification - before estimating any parameters of the model we must first obtain estimates of the order p,d,q .
- ii) Estimation - estimation of the parameters themselves.
- iii) Diagnostic Checks - test the goodness of fit for the estimated model.

iv) Forecasting - use model to estimate future values.

We now consider each of these stages in turn. For the later models we develop we will primarily be interested in i) and iv); much of what we require in estimation and diagnostic checking being determined by the computer software that we have had access to.

2.2 Identification

The process of identification consists of examining the data and comparing their properties to those of ARMA models of various orders. This is usually done by considering a variety of summary statistics which are estimates of some function of the time series. The simplest of these is the autocorrelation function.

Definition 2.10

The k^{th} process autocovariance function matrix of a time series $\{y_t\}$ is $\underline{\Gamma}(k) = \{y_{ij}^{(k)}\}$,

where $\underline{\Gamma}(k) = E[\underline{Y}_t \underline{Y}'_{t-k}]$,

and $\underline{Y}_t = y_t - E[y_t]$; $k = \dots, -1, 0, 1, \dots$.

As a function of k the $\underline{\Gamma}(k)$, $k = \dots, -1, 0, 1, \dots$ are known as the autocovariance function (ACVF).

From this we may define the correlation in two ways:-

Definition 2.11

The k^{th} process autocorrelation matrices $\underline{P}(k)$ and the matrix normalized crosscovariance $\underline{\Omega}(k)$ are defined by:-

i) $\underline{P}(k) = \{\rho_{ij}(k)\}$,

$$\text{where } \rho_{ij}(k) = \frac{y_{ij}(k)}{\sqrt{y_{ii}(0)y_{jj}(0)}}$$

$$\text{ii) } \underline{\Omega}(k) = \{\omega_{ij}(k)\},$$

$$= \underline{\Gamma}(k) \underline{\Gamma}(0)^{-1}.$$

As a function of k $\underline{P}(k)$ and $\underline{\Omega}(k)$ are known as the autocorrelation function (ACF) and, the matrix normalized crosscovariance function (MNC).

We note that if we let $\underline{\Delta}_{m \times m} = \text{dg}(y_{11}(0), y_{22}(0), \dots, y_{mm}(0))$, then we may express $\underline{P}(k)$ as :-

$$\underline{P}(k) = \underline{\Delta}^{-1/2} \underline{\Gamma}(k) \underline{\Delta}^{-1/2}$$

In practice $\underline{P}(k)$ is the most commonly used, since as shown below it can be used to identify a $\text{MA}_m(q)$ process. However $\underline{\Omega}(k)$ is employed by for example Hosking(1980a) in residual diagnostic checking and by Tjøstheim and Paulsen(1982) for identification.

The parameters of an $\text{ARMA}_m(p,q)$ process are related to the ACVF:-

Lemma 2.12

For a stationary invertible $\text{ARMA}_m(p,q)$ process,

$$\underline{\Phi}(B) \underline{v}_t = \underline{\Theta}(B) \underline{\varepsilon}_t,$$

then
$$\underline{\Phi}(z) \underline{\Gamma}(z) \underline{\Phi}'(z^{-1}) = \underline{\Theta}(z) \underline{\Sigma} \underline{\Theta}'(z^{-1}),$$

where
$$\underline{\Gamma}(z) = \sum_{i=-\infty}^{\infty} \underline{\Gamma}(i) z^i,$$

$\underline{\Phi}(z)$ and $\underline{\Theta}(z)$ are defined by (2.1.2) and (2.1.3), and $\underline{\Sigma}$ is the error covariance matrix of $\underline{\varepsilon}_t$.

Proof

See e.g. Hosking(1980a)

Corollary 2.12.1

For an $\text{AR}_m(p)$ process :-

$$\sum_{i=0}^p \underline{\Phi}_i \underline{\Gamma}(j-i) = \begin{cases} \underline{0} & j > 0, \\ \underline{\Sigma} & j = 0, \\ \underline{\Sigma} \underline{\Psi}'_j & j < 0. \end{cases} \quad (\text{Yule-Walker equations})$$

Corollary 2.12.2 (Yule-Walker equations)

For an $\text{ARMA}_m(p,q)$ process

$$\sum_{i=0}^p \underline{\Phi}_i \underline{\Gamma}(j-i) = \underline{0} \quad \text{if } j > q.$$

Corollary 2.12.3

For an MA_m(q) process

$$\underline{\Gamma}(j) = \begin{cases} \sum_{i=0}^q \theta_i \underline{\Sigma} \theta'_{i-j} & |j| \leq q, \\ \underline{0} & |j| > q. \end{cases}$$

The consequences of this lemma and its corollaries are that the ACVF, the ACF, and functions of them may be used to identify p and q for a stationary time series. In order to do this we first require some estimates of $\underline{\Gamma}(k)$, $\underline{P}(k)$, and $\underline{\Omega}(k)$; $k = \dots, -1, 0, 1, \dots$.

Definition 2.13

The kth sample autocovariance function matrix $\underline{C} = \{c_{ij}(k)\}$ is given by:-

$$\underline{C}(k) = \frac{1}{n} \sum_{t=1}^n y_t y'_{t-k},$$

where

$$y_t = \underline{Y}_t - \bar{\underline{Y}}_t,$$

$$\bar{\underline{Y}}_t = \frac{1}{n} \sum_{t=1}^n \underline{Y}_t,$$

and we assume that $\underline{Y}_1, \dots, \underline{Y}_n$ are sampled from a time series. As a function of k, $\underline{C}(k)$ is the sample autocovariance function.

Definition 2.14

The kth process sample autocorrelation matrices $\underline{R}(k) = \{r_{ij}(k)\}$ of a time series $\{\underline{Y}_t; t=1, \dots, n\}$ are:-

$$r_{ij}(k) = \frac{\sum_{t=1}^n y_{ti} y_{t-k, j}}{\sqrt{\left(\sum_{t=1}^n y_{ti}^2 \right) \left(\sum_{t=1}^n y_{tj}^2 \right)}} = \frac{c_{ij}(k)}{\sqrt{c_{ii}(0)c_{jj}(0)}}.$$

And kth process sample matrix normalized crosscovariance matrices are:-

$$\underline{S}(k) = \underline{C}(k) \underline{C}(0)^{-1}.$$

Similarly as a function of k , $\underline{R}(k)$ and $\underline{S}(k)$ form the sample auto-correlation function and sample matrix normalized cross covariance function of $\{y_t\}$. If $\underline{D}_{m \times m} = \text{dg}(c_{11}(0), \dots, c_{mm}(0))$, then we have

$$\underline{R}(k) = \underline{D}^{-1/2} \underline{C}(k) \underline{D}^{-1/2}.$$

We note that for $\underline{\Gamma}(k)$, $\underline{P}(k)$ and their estimated values that:-

$$\begin{aligned} \underline{\Gamma}'(k) &= \underline{\Gamma}(k), \\ \underline{P}'(k) &= \underline{P}(k), \\ \underline{C}'(k) &= \underline{C}(k), \end{aligned} \tag{2.2.1}$$

and $\underline{R}'(k) = \underline{R}(k)$.

It is known that the joint distribution of $\underline{C}(k)$ is asymptotically normal with mean $\underline{\Gamma}(k)$ and variance $O(1/n)$. (e.g. Hannan(1970), Fuller(1976)). Similarly $\underline{R}(k)$ (see Hannan(1970) p229) and $\underline{S}(k)$ (Tjøstheim and Paulsen(1982)) follow asymptotic normal distributions. In particular if we consider the autocorrelation function of two white-noise processes then for large n the $r_{ij}(k)$'s would have a normal distribution with mean 0 and variance n^{-1} .

Recall from corollary 2.12.3 for an $MA_m(q)$ process that:-

$$\underline{\Gamma}(j) = \underline{0} \quad |j| > q.$$

We may examine the $\underline{C}(j)$ function, and will be able to detect if we have a $MA_q(q)$ process if $\underline{C}(j)$ is sufficiently "small" for $|j| > q$. To define "small" we would need to know $\underline{\Gamma}(j)$ exactly. Instead we standardize $\underline{C}(j)$ and examine $\underline{R}(k)$. For univariate processes the procedure is usually carried out by plotting $r(k)$. This type of graph is known as the correlogram. For multivariate time series we could plot values of $r_{ij}(k)$ for $i \leq j = 1, \dots, m$ against k ($k=0, 1, \dots$ if $i=j$ and $k=\dots, -1, 0, 1, \dots$ if $i \neq j$). This would require a total of $m(m+1)/2$ plots, which is not too bad for $m=2$ or possibly $m=3$, however beyond this it is not easy to compare $m(m+1)/2$ such graphs simultaneously. Alternatively Tiao and Box(1981) have suggested a schematic way of presenting the auto-correlation function $\underline{R}(k)$. The elements of $\underline{R}(k)$ are replaced by a "+" if $r_{ij}(k)$ is greater than 2 s.e.'s, a "-" if it is less than 2 s.e.'s, and a "." otherwise.

If we have an autoregressive component the $\underline{R}(k)$ will not quickly die away. For various orders of p in an $AR_m(p)$ process the $\underline{R}(k)$ are known to follow a certain pattern. (for the univariate case see for example Box and Jenkins(1976)). The easiest solution would be to look for some function which had similar properties for the $AR_m(p)$ as does the $\underline{\Gamma}(k)$ function for the $MA_m(q)$ process. Such a function has been found in the partial autocorrelation function.

Definition 2.15

The partial autocorrelation function $\underline{F}(k)$ for an $AR_m(p)$ process is:-

$$\underline{F}(k) = \begin{cases} \underline{\phi}_p & k = p, \\ \underline{0} & k > p. \end{cases}$$

To formulate how we may actually compute $\underline{F}(k)$ we turn to the Yule-Walker equations given in corollary 2.12.1 and 2.12.2. We have:-

$$\underline{\Gamma}(k+1) = [\underline{\Gamma}(k) \quad \underline{\Gamma}(k-1) \quad \dots \quad \underline{\Gamma}(k-p+1)] \begin{bmatrix} \underline{\phi}'_1 \\ \underline{\phi}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\phi}'_p \end{bmatrix}$$

where $k > q$ - for an $ARMA_m(p,q)$ process,
and $k > 0$ - for an $AR_m(p,q)$ process.

Thus we have by stacking p of these equations that:-

$$\begin{bmatrix} \underline{\Gamma}(k+1) \\ \underline{\Gamma}(k+2) \\ \cdot \\ \cdot \\ \underline{\Gamma}(k+p) \end{bmatrix} = \begin{bmatrix} \underline{\Gamma}(k) & \underline{\Gamma}(k-1) & \dots & \underline{\Gamma}(k-p+1) \\ \underline{\Gamma}(k+1) & \underline{\Gamma}(k) & \dots & \underline{\Gamma}(k-p+2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \underline{\Gamma}(k+p-1) & \dots & \dots & \underline{\Gamma}(k) \end{bmatrix} \begin{bmatrix} \underline{\phi}'_1 \\ \underline{\phi}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\phi}'_p \end{bmatrix} \quad (2.2.2)$$

We may obtain estimates of $\underline{\phi}_1, \dots, \underline{\phi}_2$ by solving this system of equations. However we do not know p , so we form this system for $p = 1, 2, \dots$ we have:-

$$\underline{\Gamma}(k+1) = \underline{\Gamma}(k)\underline{\Phi}'_{11} \quad ,$$

$$\begin{bmatrix} \underline{\Gamma}(k+1) \\ \underline{\Gamma}(k+2) \end{bmatrix} = \begin{bmatrix} \underline{\Gamma}(k) & \underline{\Gamma}(k-1) \\ \underline{\Gamma}(k+1) & \underline{\Gamma}(k) \end{bmatrix} \begin{bmatrix} \underline{\Phi}'_{21} \\ \underline{\Phi}'_{22} \end{bmatrix} \quad , \quad (2.2.3)$$

\vdots
 \vdots
 etc. for $k=q, q+1, \dots$.

Then we may estimate $\underline{\Phi}'_{11}, \underline{\Phi}'_{21}, \underline{\Phi}'_{22}, \underline{\Phi}'_{31}, \dots, \underline{\Phi}'_{pp}, \dots$. Then :-

$$\begin{aligned} \underline{\Phi}'_{ii} & \neq 0 & i \leq p \\ & = 0 & i > p ; \end{aligned}$$

and we may write:- $\underline{F}(j) = \hat{\underline{\Phi}}_{jj} \quad j = 1, 2, \dots$. (2.2.4)

Suppose in particular we let $k=q$, and write as in Tiao and Box(1981):-

$$\underline{A}(s, k) = \begin{bmatrix} \underline{\Gamma}(k) & \underline{\Gamma}(k-1) & \dots & \underline{\Gamma}(k-s+2) \\ \underline{\Gamma}(k+1) & \underline{\Gamma}(k) & \dots & \underline{\Gamma}(k-s+1) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \underline{\Gamma}(k+s-2) & \underline{\Gamma}(k+s-3) & \dots & \underline{\Gamma}(k) \end{bmatrix} \quad , \quad \underline{b}(s, k) = \begin{bmatrix} \underline{\Gamma}(k-s+1) \\ \underline{\Gamma}(k-s+2) \\ \vdots \\ \vdots \\ \underline{\Gamma}(k-1) \end{bmatrix}$$

$$\begin{aligned} \underline{g}'(s, k) & = [\underline{\Gamma}(k+s-1) \quad \underline{\Gamma}(k+s-2) \quad \dots \quad \underline{\Gamma}(k+1)] \quad , \quad \underline{c}(s, k) = \begin{bmatrix} \underline{\Gamma}(k+1) \\ \underline{\Gamma}(k+2) \\ \vdots \\ \vdots \\ \underline{\Gamma}(k+s-1) \end{bmatrix} \\ \underline{\Xi}'_{s-1} & = [\underline{\Phi}_{s1} \quad \underline{\Phi}_{s2} \quad \dots \quad \dots \quad \underline{\Phi}_{s, s-1}] \quad , \end{aligned} \quad (2.2.5)$$

Then we may rewrite our system of equations as:- ($k=q$)

$$\underline{\Gamma}(q+1) = \underline{\Gamma}(q)\underline{\Phi}'_{11} \quad ,$$

$$\begin{bmatrix} \underline{c}(2, q) \\ \underline{\Gamma}(q+2) \end{bmatrix} = \begin{bmatrix} \underline{A}(2, q) & \underline{b}(2, q) \\ \underline{g}'(2, q) & \underline{\Gamma}(q) \end{bmatrix} \begin{bmatrix} \underline{\Xi}'_1 \\ \underline{\Phi}'_{22} \end{bmatrix} \quad , \quad (2.2.6)$$

$$\begin{bmatrix} \underline{c}(3, q) \\ \underline{\Gamma}(q+3) \end{bmatrix} = \begin{bmatrix} \underline{A}(3, q) & \underline{b}(3, q) \\ \underline{g}'(3, q) & \underline{\Gamma}(q) \end{bmatrix} \begin{bmatrix} \underline{\Xi}'_2 \\ \underline{\Phi}'_{33} \end{bmatrix} \quad ,$$

\vdots
 \vdots
 etc.

Then using the results of e.g. Morrison(1976 p68), we have that:-

$$\begin{aligned} \underline{F}'(r) = \underline{\phi}'_{rr} &= \Gamma^{-1}(q)\Gamma(q+1), & r = 1 \\ & [\Gamma(q) - \underline{b}'(r,q)\underline{A}^{-1}(r,q)\underline{b}(r,q)]^{-1} & (2.2.7) \\ & [\Gamma(r) - \underline{b}'(r,q)\underline{A}^{-1}(r,q)\underline{c}(r,q)] & r > 1 \end{aligned}$$

In practice we do not know q , and in fact we calculate $\underline{F}'(r)$ using $q=0$. Having made this assumption we will be able to identify an $AR_m(p)$ process. If we have a mixed $ARMA_m(p,q)$ process then $\underline{R}(k)$ and $\underline{F}(k)$ will die away slowly. We must then decide primarily on the basis of the pattern of the two functions, or alternatively choose some other criteria, which we will discuss below. Before moving on however we need to consider how $\underline{F}(r)$ is estimated in practice. There are several ways this could be done, but we present here the approach used by the computer package available to us to calculate $\underline{F}(r)$. We have available an early version of the Wisconsin Multiple Time Series Package (WMTS-1), which follows the procedure outlined in Tiao and Box(1981).

Consider the transpose of the $AR_m(r)$ model;

$$\underline{V}'_t = \underline{V}'_{t-1}\underline{\phi}'_1 + \dots + \underline{V}'_{t-p}\underline{\phi}'_r + \underline{\varepsilon}'_t \quad (2.2.8)$$

Suppose that we have a sample $\underline{V}_1, \dots, \underline{V}_n$ then we may stack at least $n-p$ equations:-

$$\begin{bmatrix} \underline{V}'_{-r+1} \\ \underline{V}'_{-r+2} \\ \vdots \\ \underline{V}'_{-n} \end{bmatrix}_{(n-r) \times m} = \begin{bmatrix} \underline{V}'_{-r} \\ \underline{V}'_{-r+1} \\ \vdots \\ \underline{V}'_{-n-1} \end{bmatrix}_{(n-r) \times m} \underline{\phi}'_1 + \dots + \begin{bmatrix} \underline{V}'_{-1} \\ \underline{V}'_{-2} \\ \vdots \\ \underline{V}'_{-n-r} \end{bmatrix} \underline{\phi}'_r + \begin{bmatrix} \underline{\varepsilon}'_{-r+1} \\ \underline{\varepsilon}'_{-r+2} \\ \vdots \\ \underline{\varepsilon}'_{-n} \end{bmatrix} \quad (2.2.9)$$

If we let:-

$$\begin{aligned} \underline{y}_r &= \left[\underline{V}'_{-r+1} \dots \underline{V}'_{-n} \right]', & \underline{Y}_r &= \begin{bmatrix} \underline{V}'_{-r} & \underline{V}'_{-r-1} & \dots & \underline{V}'_{-1} \\ \vdots & \vdots & & \vdots \\ \underline{V}'_{-n-1} & \underline{V}'_{-n-2} & \dots & \underline{V}'_{-n-r} \end{bmatrix}, \\ \underline{\phi}' &= \left[\underline{\phi}'_1, \dots, \underline{\phi}'_r \right], \\ \underline{A}' &= \left[\underline{\varepsilon}'_{-r+1}, \underline{\varepsilon}'_{-r+2}, \dots, \underline{\varepsilon}'_{-n} \right], \end{aligned}$$

then we may rewrite (2.2.8) as:-

$$\underline{y}_r = \underline{y}'_r \underline{\phi}' + \underline{a} ,$$

and we may estimate $\underline{\phi}'$ by ordinary least squares (OLS),

$$\hat{\underline{\phi}}'_r = [\underline{y}'_r \underline{y}_r]^{-1} \underline{y}'_r \underline{y}_r . \quad (2.2.10)$$

Then if we do this for $r=1,2,\dots$ then we may define $\hat{\underline{f}}(r)$ to be the last $m \times m$ matrix of $\hat{\underline{\phi}}_r$.

As with the ACF we may easily examine the PACF by plotting it against its corresponding lag. However the WMTS-1 package also produces a schematic version for $\hat{\underline{f}}(r)$ in a similar fashion to that described for $\hat{R}(k)$. From standard linear model theory it is possible to calculate the standard errors of $\hat{\underline{f}}(r)$. If each element is divided by its standard error we obtain a standardized form of $\hat{\underline{f}}(r)$ ($\hat{\underline{f}}^S(r)$ say). We may represent these by writing a "+" if an element of $\hat{\underline{f}}^S(r)$ is greater than 2, or "-" if it is less than -2, or a "." if it lies between +2 and -2.

The ACF and PACF offer a means of identifying an $MA_m(q)$ or an $AR_m(p)$ process respectively. But what can we use if we have a mixed $ARMA_m(p,q)$ process? Tiao and Box(1981) suggest a criterion with the following properties:-

$$d(s,k) = \begin{cases} 0 & \text{if } s > p \text{ and } k \geq q \\ \neq 0 & \text{otherwise .} \end{cases}$$

Consider the equations (2.2.2) and (2.2.5),

$$\begin{bmatrix} \underline{a}(s,k) & \underline{b}(s,k) \\ \underline{g}'(s,k) & \underline{\Gamma}(k) \end{bmatrix} \begin{bmatrix} \underline{\Xi}'_{s-1} \\ \underline{\phi}'_{ss} \end{bmatrix} = \begin{bmatrix} \underline{c}(s,k) \\ \underline{\Gamma}(s+k) \end{bmatrix} \quad (2.2.11)$$

If $k \geq q$ and $s=p$ then $[\underline{\Xi}'_{s-1}, \underline{\phi}'_{ss}]$ are the exact parameters of the Yule-Walker equations. We define $d_{ij}(s,k)$ to be :-

$$d_{ij}(s,k) = \begin{vmatrix} \underline{a}(s,k) & \underline{c}_j(s,k) \\ \underline{g}'_j(s,k) & \gamma_{ij}(s+k) \end{vmatrix} \quad (2.2.12)$$

Then if $k \geq q$ and $s > p$, because of (2.2.11) the $\underline{c}_j(s,k)$ and $\gamma_{ij}(s+k)$ are linearly dependent on $\underline{a}(s,k)$ and $\underline{g}'_j(s,k)$ and consequently

$$d_{ij}(s,k) = 0 \quad \text{if } s > p \text{ and } k \geq q \text{ for } i,j=1,\dots,m. \quad (2.2.13)$$

Using (2.2.12) we may define the following:-

Definition 2.16

Let $\underline{D}(s,k) = \{d_{ij}(s,k)\}$; $s=1,2,\dots$; $k=0,1,\dots$; where $d_{ij}(s,k)$ is given by (2.2.12). Then the series of matrices $\underline{D}(1,0), \underline{D}(2,0), \dots, \underline{D}(1,1), \underline{D}(2,1), \dots$ of determinants forms a determinantal criteria for identifying $ARMA_m(p,q)$ models:-

$$\begin{aligned} \underline{D}(s,k) &= \underline{0} \text{ if } s > p \text{ and } k \geq q. \\ &\neq \underline{0} \text{ otherwise.} \end{aligned}$$

A further tool suggested by Tiao and Box(1981) for the identification of $AR_m(p)$ models takes the form of a likelihood ratio test. For an $AR_m(r)$ model consider testing the hypothesis $H_0: \underline{\Phi}_r = \underline{0}$. Let $\underline{SS}(r)$ be the matrix of residual sums of squares and products after fitting an $AR_m(r)$. Using the notation of (2.2.9) we have:-

$$\underline{SS}(0) = \underline{Y}'_0 \underline{Y}_0 \quad (2.2.14)$$

$$\underline{SS}(r) = [\underline{Y}_r - \underline{x}_{1r} \hat{\Phi}'_1 - \dots - \underline{x}_{rr} \hat{\Phi}'_r]' [\underline{Y}_r - \underline{x}_{1r} \hat{\Phi}'_1 - \dots - \underline{x}_{rr} \hat{\Phi}'_r],$$

where $[\underline{x}_{1r} : \underline{x}_{2r} : \dots : \underline{x}_{rr}] = \underline{Y}_r$, for $r = 1, 2, \dots$ (2.2.15)

The likelihood ratio statistic $\frac{|\underline{SS}(r)|}{|\underline{SS}(r-1)|}$ may be used to determine the order p of an $AR_m(p)$ process. Using the approximation of Bartlett(1938) we have that:-

$$M(r) = -(N - 1/2 - rm) \ln \left\{ \frac{|\underline{SS}(r)|}{|\underline{SS}(r-1)|} \right\} \quad (2.2.16)$$

is asymptotically $\chi^2_{(m)}^2$, where $N = m-r-1$.

Thus we may compute $M(r)$ for $r=1,2,\dots,s$ (say) and choose a suitable p when $M(r)$ is sufficiently small.

Various other statistical tools may be used for identification. For example two similar χ^2 statistics are presented by Tjøstheim and

Paulsen(1982). Also the inverse autocorrelation function has similar properties to the partial autocorrelation function. The fourier transform of the autocorrelation function results in a normalized cross-spectral matrix. If the inverse of this cross-spectral matrix is then formed and then the inverse fourier transform applied, the resulting coefficients will have analogous properties to the original autocorrelation function, but with the AR and MA operators interchanged (see for example Priestley(1981)). A further identification tool is the use of canonical variate analysis as suggested by Akaike(1974) and further discussed by Cooper and Wood(1982).

We now move on, and consider next the estimation of the parameter values.

2.3 Estimation

Once the order of the ARMA model has been determined we then require to estimate the parameter values themselves. Under the assumption of normality we may use the method of maximum likelihood, however the exact maximum likelihood estimator proves to be computationally expensive, and consequently various approximations have arisen, which, under certain conditions, prove to be close to the required result.

Two simple, but inefficient, methods of estimation have in fact already been mentioned for the $AR_m(p)$ process in the previous section. In the definition of the partial autocovariance function, the parameters ϕ_1, \dots, ϕ_p were estimated by solving the Yule-Walker equations, thus we may apply this technique here. We note, since the Yule-Walker equations apply for $ARMA_m(p,q)$ processes, that using the appropriate autocovariance function we may estimate ϕ_1, \dots, ϕ_p for the mixed model as well. A similar approach may be used to obtain estimates of $\theta_1, \dots, \theta_q$ via the inverse autocorrelation function since Yule-Walker type equations hold in an analogous way to that of the $AR_m(p)$ case. A recursive procedure for easily solving the Yule-Walker equations has been developed by Whittle(1963). This is the method employed by the SAS package to compute the partial autocorrelation function. However these Yule-Walker estimates are poor, especially for smaller sample sizes, and also if the roots of the characteristic equation $|\phi(z^{-1})| = 0$ lie close to the unit circle. Since the inverse autocorrelation function is itself

based on an estimate of the spectral density function there seems to be further room for error which is perhaps the reason this does not seem to have been done in practice.

The second type of estimation mentioned in the previous section was that of the OLS estimator (2.2.10). There is no similar estimator for the $MA_m(q)$ process, other than fitting a high order $AR_m(p)$ process and forming estimates of $\underline{\theta}_1, \dots, \underline{\theta}_q$ as:-

$$\underline{\theta}(B) = \underline{\phi}^{-1}(B).$$

The OLS estimate is more accurate than the Yule-Walker estimate but, it is still relatively poor compared to that of the exact likelihood estimate. This latter estimate and the conditional likelihood estimate are based on the likelihood function of an $ARMA_m(p,q)$ process which under the assumption of normality is:-

$$L(\underline{\phi}, \underline{\theta}, \underline{\Sigma} | \underline{V}_t) \propto |\underline{\Sigma}|^{-(n-p)/2} \exp\left\{-\frac{1}{2} \text{tr} \underline{\Sigma}^{-1} S(\underline{\phi}, \underline{\theta})\right\}, \quad (2.3.1)$$

where $\underline{\phi} = \phi_1, \dots, \phi_p$; $\underline{\theta} = \theta_1, \dots, \theta_q$; and $S(\underline{\phi}, \underline{\theta}) = \sum_{k=1-p-q}^n \underline{a}_k \underline{a}_k'$.

The \underline{a}_k may be estimated by:-

$$\begin{aligned} \hat{\underline{a}}_{p+1} &= \frac{V_{p+1}}{V_{p+1}} + \frac{\phi_1 V_p}{V_{p+1}} + \dots + \frac{\phi_p V_1}{V_{p+1}} - \frac{\theta_1 \hat{\underline{a}}_p}{V_{p+1}} - \dots - \frac{\theta_q \hat{\underline{a}}_{p-q+1}}{V_{p+1}} \\ &\vdots \\ \hat{\underline{a}}_n &= \frac{V_n}{V_n} + \frac{\phi_1 V_{n-1}}{V_n} + \dots + \frac{\phi_p V_{n-p}}{V_n} - \frac{\theta_1 \hat{\underline{a}}_{n-1}}{V_n} - \dots - \frac{\theta_q \hat{\underline{a}}_{n-q}}{V_n}, \end{aligned} \quad (2.3.2)$$

for $k = p+1, \dots, n$.

Since the estimates of $\underline{a}_{1-p-q}, \dots, \underline{a}_p$ are difficult to compute, an approximate likelihood estimate is given by setting these to their expected value, namely zero. This approximation forms the conditional likelihood estimate. i.e. the likelihood function conditional on $\underline{a}_{1-p-q} = \dots = \underline{a}_p = \underline{0}$. This is given by:-

$$L_c(\underline{\phi}, \underline{\theta}, \underline{\Sigma} | \underline{V}_t) \propto |\underline{\Sigma}|^{-(n-p)/2} \exp\left\{-\frac{1}{2} \text{tr} \underline{\Sigma}^{-1} S_c(\underline{\phi}, \underline{\theta})\right\}, \quad (2.3.3)$$

where now
$$S_c(\tilde{\Phi}, \tilde{\Theta}) = \sum_{k=p+1}^n \hat{a}_t \hat{a}'_t \quad (2.3.4)$$

The exact likelihood function may also be computed by various methods for example the Box and Jenkins(1976) method in which estimates of $\underline{a}_{1-p-q}, \dots, \underline{a}_p$ are obtained by "backforecasting". This method is applied to $AR_1(p)$ and $MA_1(q)$ processes and is extended to the mixed process ($ARMA_1(p,q)$) by Newbold(1974), to the $MA_m(q)$ process by Osborn(1977) and to the general $ARMA_m(p,q)$ process by Hillmer and Tiao(1979) who show that the exact likelihood may be written as:-

$$\mathcal{L}_e(\tilde{\Phi}, \tilde{\Theta}, \underline{\Sigma} | \underline{V}_t) \propto \mathcal{L}_c(\tilde{\Phi}, \tilde{\Theta}, \underline{\Sigma} | \underline{V}_t) \mathcal{L}_r(\tilde{\Phi}, \tilde{\Theta}, \underline{\Sigma} | \underline{V}_t), \quad (2.3.5)$$

where $\mathcal{L}_r(\tilde{\Phi}, \tilde{\Theta}, \underline{\Sigma} | \underline{V}_t)$ is a function of:-

- i) $\underline{V}_1, \dots, \underline{V}_p$ if $q = 0$,
- ii) $\underline{V}_1, \dots, \underline{V}_n$ if $q \neq 0$.

Having obtained an expression for the likelihood, it is still necessary to maximize it. This is done iteratively. An initial estimate of the parameters $\tilde{\Phi}^0, \tilde{\Theta}^0$ is given and these are used to form an estimate of $\underline{\Sigma}, \underline{\Sigma}^0$.

$$\underline{\Sigma}^0 = \frac{1}{n} \sum_{t=1-p-q}^n \hat{a}_t \hat{a}'_t. \quad (2.3.6)$$

A non-linear least squares routine such as that of Marquardt(1963) can then be employed to find $\tilde{\Phi}^1, \tilde{\Theta}^1$ which minimize $S(\tilde{\Phi}, \tilde{\Theta})$ and hence maximize $\mathcal{L}(\tilde{\Phi}, \tilde{\Theta}, \underline{\Sigma} | \underline{V}_t)$. From this we may repeat the procedure until we have convergence. This is the approach used by the WMTS-1 package, which employs a modification of the Marquardt algorithm written by Meeter(1965).

The complexity of the exact likelihood function makes it computationally inefficient. However if the initial guesses are chosen to be close to those of the final estimates very few of the complex iterations will be required. Thus we can improve the computational efficiency by letting the initial guesses be those as estimated by one of the approximate estimation procedures mentioned above.

Other methods of estimation are also available, based on the frequency domain approach of Whittle(1953). Akaike(1976) applies such an approach to fit State Space models, these being the techniques

employed by SAS.

Finally we note that confidence limits of the estimates are readily available and the asymptotic distribution of the estimates have been well studied by for example Hannan(1970).

2.4 Diagnostic Checks

Once a model has been fitted, it is necessary to ensure that it is fitting the data well. One approach is to examine the residuals; one would expect the residuals to behave like a white-noise process, if they are in fact estimates of the innovation series powering the process. A further way to examine the adequacy of the model is to show that the addition of more parameters to the model is superfluous. Thus we may approach diagnostic checks applied to the residuals and/or by "over-fitting tests".

The simplest way to examine the residuals is to treat them as a new set of data, and apply the identification techniques to them to see if they will be adequately represented by a "white-noise model". The residuals are readily available to do this, having been already calculated at the estimation stage. The WMTS-1 package automatically performs some useful processing of the estimated innovations. The cross-correlations are calculated, and using the indicator symbols described in section 2.2 we would expect these indicators to consist almost entirely of "."s. Thus for a two dimensional series the pattern of the autocorrelation function of the residuals may resemble something like this:-

```
          . . . . .
          . . . . + . . .
lags:     1   2   3   4   5
```

The odd "+" or "-" should not worry us, providing they are infrequent and not conforming to any pattern.

We could also examine the PACF but since it merely reciprocates the ACF for a white-noise process, this may not be worthwhile, unless there is some indication that we do not have a white-noise process from the examination of the ACF.

Overall χ^2 tests which examine all the residual crosscorrelations simultaneously have been developed. The well known portmanteau statistic (e.g. Box and Jenkins(1970)) for univariate time series has been generalized to multivariate time series by Hosking(1980a,1980b).

Hosking(1980a) also examines a whole battery of overfitting tests based on the lagrange-multiplier test procedure.

One simple overfitting test has in fact already been mentioned in 2.2, namely the likelihood ratio statistic (2.2.16) which we have labelled $M(r)$. We could also apply $M(r)$ to see if additional moving average parameters would improve the model. We note that in overfitting tests we try adding parameters either to the autoregressive side of the model or to the moving-average side, but not simultaneously, because of parameter redundancy. (See Box and Jenkins(1970)).

In this context we will use the diagnostic checks available to us on our computer software. Namely the ACF of the residual series, and where appropriate the likelihood ratio form of the overfitting test.

2.5 Forecasting

The ability to forecast is perhaps one of the most useful applications of time series analysis. It is important in business, finance, marketing, public administration and many other areas. There are many ways to approach the subject of forecasting. In some situations subjective forecasting is the best way to analyse future trends. Many managers, for example, need to make important decisions as to how to run their business based on their "know how" and past experience. Such intuitive prediction is invaluable, but it can be augmented by statistical techniques. Also, in situations where it is extremely difficult for anyone to subjectively predict future events, statistical techniques provide a necessary objective method of prediction.

Various statistical forecasting methods are available, not all of them based on time series analysis. For example in launching a new product a sales manager might well conduct a consumer survey to get a feel for the market. Time series analysis, however, lends itself very well to this area, and a whole battery of time series forecasting methods are available, not only the Box-Jenkins method, but many others as well; for example the Holt-Winters method. The Box-Jenkins approach considered here gives us a means of estimating the distribution of future observations conditional on present and past values. Once we know this distribution, we may use it to provide both point predictors, based on the mean, median or mode; and also interval predictors, usually in the form of a confidence region for whatever location parameter we

may have chosen. Point predictors give us a single value which estimates the future value of the variable of interest, whilst interval predictors indicate the degree of uncertainty attributable to these single values.

The most common location parameter used in practice is the mean, because it has the property of providing the minimum mean square error forecast. (MMSE forecast.) In other words if the forecast of Z_{t+l} at time t is $\underline{Z}_t(l)$, and the forecast error is given by $e_t(l) = Z_{t+l} - \underline{Z}_t(l)$, then if $\underline{Z}_t(l)$ is chosen such that it minimizes $E[e_{it}(l)^2], i=1, \dots, m$; then it is the MMSE forecast, which is simply $\underline{Z}_t(l) = E[Z_{t+l} | Z_t, Z_{t-1}, \dots]$, the mean of the distribution of Z_{t+l} conditional on Z_t, Z_{t-1}, \dots . Obviously if the distribution is the normal distribution, the mean, median and mode would all be identical. However if a transformation is applied prior to the analysis, or some other distribution assumed, this is no longer the case, and we need to differentiate between the various options of location parameters. Before considering these non-gaussian possibilities, let us consider the simple case of the stationary, normal distributed time series. In particular one that follows an ARMA_m(p,q) process:-

$$\begin{aligned} Z_t &= -\phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \\ &= e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots \quad ; \text{ where } \text{var}(e_t) = \underline{\Sigma}. \end{aligned}$$

Consider $Z_{t+l} = e_{t+l} + \psi_1 e_{t+l-1} + \dots + \psi_{l-1} e_{t+1} + \psi_l e_t + \psi_{l+1} e_{t-1} + \dots$ (2.5.1)

At time t , e_t, e_{t-1}, \dots are known, whereas e_{t+1}, \dots, e_{t+l} are unknown but have mean $\underline{0}$ and variance $\underline{\Sigma}$. Thus the distribution of Z_{t+l} at time t , conditional on Z_t, Z_{t-1}, \dots is normal with mean:-

$$\begin{aligned} \underline{Z}_t(l) &= E[Z_{t+l} | Z_t, Z_{t-1}, \dots] \\ &= \psi_l e_t + \psi_{l+1} e_{t-1} + \dots \end{aligned} \quad (2.5.2)$$

and variance:-

$$\begin{aligned} \text{Var}[Z_{t+l} | Z_t, Z_{t-1}, \dots] &= \text{Var}[e_t(l)] \\ &= \text{Var}[e_{t+l} + \psi_1 e_{t+l-1} + \dots + \psi_{l-1} e_{t+1}] \\ &= \underline{\Sigma} + \psi_1 \underline{\Sigma} \psi_1' + \dots + \psi_{l-1} \underline{\Sigma} \psi_{l-1}'. \end{aligned} \quad (2.5.3)$$

(2.5.2) is often more conveniently expressed as:-

$$\underline{Z}_t(\ell) = \hat{\theta}_1 \underline{e}_t + \dots + \hat{\theta}_q \underline{e}_{t+1-q} - \hat{\phi}_1 \underline{Z}_t(\ell-1) - \dots - \hat{\phi}_{\ell-1} \underline{Z}_t(1) - \hat{\phi}_\ell \underline{Z}_t - \dots - \hat{\phi}_p \underline{Z}_{t+1-p} \quad (2.5.4)$$

In practice the parameters are replaced by their estimated values $\hat{\theta}_1, \dots, \hat{\theta}_q, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\psi}_1, \dots$; as computed by the techniques described in the previous section. We also need estimates of $\underline{e}_t, \underline{e}_{t-1}, \dots$; these may also be replaced by the estimates resulting from the previous step of our Box-Jenkins method. Alternatively they may be regarded as the one step ahead forecast errors, so that if a forecaster is using a model which is well tried, for example, over a long time period, in which more observations become known but the model is not re-estimated, then estimates of $\underline{e}_t, \underline{e}_{t-1}, \dots$ are:-

$$\hat{\underline{e}}_t = \underline{e}_{t-1}(1) = \underline{Z}_t - \underline{Z}_{t-1}(1) \quad t = 0, \pm 1, 2, \dots \quad (2.5.5)$$

Thus as more \underline{Z}_t 's become known, the forecast of $\underline{Z}_{t+\ell}$ may be easily updated, giving successively more accurate estimates:- $\underline{Z}_t(\ell), \underline{Z}_{t+1}(\ell-1), \underline{Z}_{t+2}(\ell-2), \dots, \underline{Z}_{t+\ell-1}(1)$.

The expressions (2.5.2) or (2.5.4) provide us with a point predictor $\underline{Z}_t(\ell)$ of $\underline{Z}_{t+\ell}$. The confidence region for $\underline{Z}_{t+\ell}$ may also be produced using standard multivariate theory. Thus if $\text{Var}[\underline{e}_t(\ell)]$ is estimated by substituting estimates of $\underline{\Sigma}, \underline{\Psi}_1, \dots, \underline{\Psi}_{\ell-1}$ into (2.5.3) then, by the results of e.g. Morrison(1976) a $100(1-\alpha)\%$ confidence region is given by:-

$$(\underline{Z}_t(\ell) - \underline{Z}_{t+\ell})' \hat{\underline{\Sigma}}_{\underline{e}_t(\ell)}^{-1} (\underline{Z}_t(\ell) - \underline{Z}_{t+\ell}) \leq \chi_{\alpha, m}^2 \quad (2.5.6)$$

where $\hat{\underline{\Sigma}}_{\underline{e}_t(\ell)}$ is the estimate of $\text{Var}[\underline{e}_t(\ell)]$ based on (2.5.3),

and $\chi_{\alpha, m}^2$ is the $100\alpha\%$ point of the χ_m^2 distribution.

We note in passing, that the expression (2.5.6) above is not based on the F distribution, as is the usual case in the multivariate context, where the covariance matrix is the usual estimate of some multivariate variable \underline{x} :-

$$\underline{S} = \frac{1}{n-1} \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})' \quad , \text{ and } \quad \bar{\underline{x}} = \frac{1}{n} \sum_{i=1}^n \underline{x}_i \quad .$$

\underline{S} is known to follow a Wishart distribution, and consequently the equivalent expression to (2.5.6) for $\bar{\underline{x}}$ results in the Hotelling's T^2

distribution, which in turn may be rewritten in terms of F:-

$$n(\bar{\underline{x}} - \underline{x})\underline{S}^{-1}(\bar{\underline{x}} - \underline{x})' \leq \frac{(n-1)m}{(n-m)} F_{\alpha; m, n-m} \quad , \quad (2.5.7)$$

where $F_{\alpha; m, n-m}$ is the upper 100 α % point of the $F_{m, n-m}$ distribution. However in this time series context, the variance matrix within the quadratic expression ($\sum_{\underline{e}_t(1)}$ in (2.5.6)) does not follow the Wishart distribution, but since as shown by, for example, Hannan(1970) $\underline{\Sigma}$ and $\underline{\Psi}_i$ are asymptotically normally distributed with mean $\underline{\Sigma}$ and $\underline{\Psi}_i$ ($i=1,2,\dots$) respectively, then (2.5.6) gives us an approximate confidence region in this case.

Having developed the above forecasting procedure we now consider how these results are affected by a transformation. As mentioned previously a time series, $\{\underline{V}_t\}$ may be transformed to a series $\{\underline{W}_t\}$ by some instantaneous transformation, such as that of Box and Cox(1964), in order to have a stationary process $\{\underline{W}_t\}$. The ARMA model is then fitted to this new process. These transformations also hope to produce a gaussian time series so that the methods given above may be applied. Thus we have that

$$\underline{W}_t = f(\underline{V}_t) \quad , \quad t=0, \pm 1, \dots \quad ; \quad (2.5.8)$$

and using the methods described above we may readily form forecasts of $\underline{W}_{t+1}, \underline{W}_{t+2}, \dots$ as $\underline{W}_t(1), \underline{W}_t(2), \dots$. However interest is usually in the forecasts of $\underline{V}_{t+1}, \underline{V}_{t+2}, \dots$ so that what is usually required is an estimate of $E[\underline{V}_{t+1} | \underline{V}_t, \underline{V}_{t-1}, \dots]$. If we let the inverse transformation of (2.5.8) be:-

$$\underline{V}_t = g(\underline{W}_t) \quad , \quad t=0, \pm 1, \dots \quad (2.5.9)$$

Then the forecast of interest is

$$\begin{aligned} \underline{V}_t(k) &= E[\underline{V}_{t+k} | \underline{V}_t, \underline{V}_{t-1}, \dots] \\ &= E[\underline{V}_{t+k} | \underline{W}_t, \underline{W}_{t-1}, \dots] \\ &= E[g(\underline{W}_{t+k}) | \underline{W}_t, \underline{W}_{t-1}, \dots] \quad , \quad (2.5.10) \end{aligned}$$

$$\text{which, in general} \quad \neq g(E[\underline{W}_{t+k} | \underline{W}_t, \underline{W}_{t-1}, \dots]) \quad . \quad (2.5.11)$$

This problem may be tackled in many ways, one in particular is to consider the resulting distribution of \underline{V}_t , given \underline{W}_t is normal. This is perhaps most easily demonstrated by a simple example. Consider the well known transformation:-

$$\underline{W}_t = \ln(\underline{V}_t) \quad .$$

In general if $\underline{Y} = \ln(\underline{X})$ and $\underline{Y} \sim N_m(\underline{\mu}, \underline{\Sigma})$ then \underline{X} is said to follow the log-normal distribution with probability density function:-

$$\frac{1}{\prod_{i=1}^m X_i} \frac{1}{|2\pi \underline{\Sigma}|^{-\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(\ln \underline{X} - \underline{\mu})' \underline{\Sigma}^{-1} (\ln \underline{X} - \underline{\mu}) \right\} \quad . \quad (2.5.12)$$

We will denote this by $\Lambda_m(\underline{\mu}, \underline{\Sigma})$, m being the dimension and, $\underline{\mu}$ and $\underline{\Sigma}$ the parameters of the log-normal distribution. It is then easily shown that the mean of \underline{X} is

$$E(X_i) = \exp(\mu_i + \frac{1}{2}\sigma_{ii}) \quad ; \quad i=1, \dots, m. \quad (2.5.13)$$

and

$$\text{Cov}(X_i, X_j) = \{ \exp(\sigma_{ij}) - 1 \} \{ \exp(\mu_i + \mu_j + \frac{1}{2}(\sigma_{ii} + \sigma_{jj})) \} \quad ;$$

$$i, j=1, \dots, m; \quad (2.5.14)$$

hence

$$\text{Var}(X_i) = \{ \exp(\sigma_{ii}) - 1 \} \{ \exp(2\mu_i + \sigma_{ii}) \} \quad ; \quad i=1, \dots, m.$$

Thus we may now easily form estimates of \underline{V}_{t+l} , from those of \underline{W}_{t+l} .
Recalling that:-

$$(\underline{W}_{t+l} | \underline{W}_t, \underline{W}_{t-1}, \dots) \sim N_m(\underline{W}_t(l), \underline{\Sigma}_{e_t}(l)) \quad ,$$

and

$$\underline{V}_t = \exp(\underline{W}_t) \quad ,$$

we have

$$E[\underline{V}_{t+l}] = \underline{V}_t(l) = \exp\{ \underline{W}_t(l) + \frac{1}{2}\sigma_{tii}(l) \} \quad , \quad (2.5.15)$$

and

$$\text{Var}[\underline{V}_{t+l}] = \{ \exp\sigma_{tii}(l) - 1 \} \{ \exp(2\underline{W}_t(l) + \sigma_{tii}(l)) \} \quad ,$$

where

$$\{ \sigma_{tij}(l) \} = \underline{\Sigma}_{e_t}(l) \quad ; \quad (2.5.16)$$

$$i=1, \dots, m.$$

Finally we note that for this simple example, α 100(1- α)% confidence region for \underline{V}_{t+l} is simply:-

$$(\ln \underline{V}_{t+l} - \underline{W}_t(l))' \underline{\Sigma}_{e_t}(l)^{-1} (\ln \underline{V}_{t+l} - \underline{W}_t(l)) \leq \chi_{\alpha, m}^2 \quad . \quad (2.5.17)$$

We see, however, that although this provides an interval predictor for V_{t+l} , it is not centred on $E[V_{t+l}]$. Thus, from the above example, for a general transformation we see that we may form a point predictor of V_{t+l} providing we know the mean of the resulting "f-normal" distribution. Also we may form an interval predictor analogous to (2.5.17).

Another approach to the forecasting of V_{t+l} after performing a transformation is given by Granger and Newbold(1976), who consider an expansion of g in terms of hermite polynomials. For a univariate series such that

$$W_t = g(V_t),$$

Then we may write
$$W_t = \sum_{i=0}^{\infty} \alpha_i H_i(V_t), \quad (2.5.18)$$

where,
$$H_n(x) = \exp(x^2/2) \left[\frac{d}{dx} \right]^n \exp(-x^2/2)$$

$$= n! \sum_{m=0}^{[n/2]} (-1)^m \{2^m m! (n-2m)!\}^{-1} x^{n-2m}, \quad (2.5.19)$$

the Hermite polynomials, (e.g. $H_0(x)=1$, $H_1(x)=x$, $H_2(x)=x^2-1$).

And
$$\alpha_n = ((2\pi)^{1/2} n!)^{-1} \int_{-\infty}^{\infty} e^{-x^2/2} \left[\frac{d}{dx} \right]^n g(V_t) dV_t. \quad (2.5.20)$$

For example $E[W_{t+1} | W_t, \dots]$ is shown by Granger and Newbold to give the same results as above for $g=\log$. Such an approach requires that the function concerned be easily expanded this way, which means that (2.5.20) be easily computed.

Having defined the $ARMA_m(p,q)$ model and discussed its properties, we next consider the concept of Wiener-Granger causality.

2.6 Independence Between Time Series

In multivariate time series, questions often arise in connection with the inter-relationships between individual time series. For example, in economics knowledge of the influence that one economic variable has on another aids us in understanding the underlying economic system. This in turn enables the economist to forecast future economic

trends, and advise government bodies regarding policy. Many other examples exist, and various concepts such as causality and feedback are of particular interest in the modeling of both engineering and economic systems. Exact definitions of these concepts have been given e.g. Granger(1969,1982), Geweke(1982b). In particular the definition offered by Granger(1969), now commonly known as Wiener-Granger causality, offers a good working definition, with practical implications both in testing and in interpretation. Although there does seem to be much discussion and dispute over these definitions (see for example A. Zellner's comment to Geweke(1982a) and also Geweke(1982b).); these disputes are in general philosophical and will not be considered here. Wiener-Granger causality offers a widely accepted definition of dependence between time series and is presented below together with various methods to test and measure causality, feedback and independence between time series.

2.6.1 Some Definitions of Causality

We first require some notation. If Z_t is a stationary stochastic process, then let

Z_{t_p} represent the set of past values $\{Z_{t-j}; j=1,2,\dots\}$;

$Z_{t_{pp}}$ the set of past and present values $\{Z_{t-j}; j=0,1,2,\dots\}$;

$Z_{t_{fpp}}$ the set of future, present and past values $\{Z_{t-j}; j=0,\pm 1,\pm 2,\dots\}$;

$P_t(Z/W)$ be the minimum mean square error predictor of Z_t using the set W , with prediction error:-

$$\epsilon_t(Z/W) = Z_t - P_t(Z/W),$$

$$\sigma^2(Z/W) \text{ be the variance of } \epsilon_t(Z/W),$$

and Ω_{t_p} be all the information in the universe accumulated up to time $t-1$, and $(\Omega-Y)_{t_p}$ be all this information apart from the series Y_{t_p} . We then have :-

Defintion 2.17 (Wiener-Granger)

a) Causality: $Y_t \rightarrow X_t$

If $\sigma^2(X_t/\Omega_{tp}) < \sigma^2(X_t/(\Omega-Y)_{tp})$ then Y is said to cause X.

b) Feedback: $Y_t \leftrightarrow X_t$

If i) $\sigma^2(X_t/\Omega_{tp}) < \sigma^2(X_t/(\Omega-Y)_{tp})$, and

ii) $\sigma^2(Y_t/\Omega_{tp}) < \sigma^2(Y_t/(\Omega-X)_{tp})$, then both $Y_t \rightarrow X_t$ and $X_t \rightarrow Y_t$ occur

and we say we have feedback between X and Y.

c) Instantaneous Causality: $Y_t \cdot X_t$

If $\sigma^2(X_t/\Omega_{tp}, Y_{tpp}) < \sigma^2(X_t/\Omega_{tp})$ then Y is instantaneously causing X and X is instantaneously causing Y.

We note that by symmetry we may also have defined c) by:-

$$\sigma^2(Y_t/\Omega_{tp}, X_{tpp}) < \sigma^2(Y_t/\Omega_{tp}),$$

which consequently is equivalent to the inequality given.

The criteria involved in these definitions is that of predictability. Thus $Y_t \rightarrow X_t$ implies that we are better able to predict X_t given the past values of Y_t than if we did not employ Y_t 's past values. We note that in all the definitions the time series are stationary.

Granger(1969) also makes the following points.

- 1) The fact that the criterion used is the variance, implies that these definitions might be better named "causality in mean".
- 2) We will not in fact have the whole universe of information Ω_t but rather a subset U_t (say) of "useful" information.
- 3) In practice any predictors that are used are linear in nature.

Thus using these facts we could for example replace Ω_t by U_t , $\sigma^2(X/W)$ by $\sigma^{+2}(X/W)$; where, $\sigma^{+2}(X/W)$ represents the variance of the linear prediction error of X_t given a set W ; and replace causality, for example, by linear causality in mean with respect to a set U . However this leads to rather long-winded definitions, although these are more realistic. We will bear in mind what the true definitions should be

(i.e. the more long-winded ones) but instead use the more general definitions. It should be clear in the next section, and in later applications exactly what form of causality, feedback etc. we are implying. What is important is to understand that in whatever framework, be it linear or otherwise, $Y_t \rightarrow X_t$ for example, implies that within that framework we are better able to predict X_t if we use information about past values of Y_t over and above any other past information, such as past X_t . Obviously U_t may not in fact consist of "useful" information, but rather known and thought to be useful information. If some important variable is unknown, spurious relationships can occur between known variables (cf. correlation). Even if however, we only discover a spurious relationship which enables us, for example, to better forecast some variable then it is still of value. In any case our understanding of a system must start somewhere.

In later sections we will also require two further definitions:-

Definition 2.17 continued

d) (Linear) Independence: $Y_t \not\parallel X_t$

If X does not cause (linearly in mean) Y (with respect to U), Y does not cause (linearly in mean) X (with respect to U) and no instantaneous (linear) causality (in mean with respect to U) exists then Y is said to be (linearly) independent (in mean) of X (with respect to U).

c) Complete Dependence: $Y_t \Leftrightarrow X_t$

If $X_t \rightarrow Y_t$, $Y_t \rightarrow X_t$, and $X_t \cdot Y_t$ then Y_t and X_t are said to be completely dependent (linearly in mean with respect to U) on one another.

The term "independence" in definition 2.17d) should not be confused with the usual stochastic independence. Two variables z_1 and z_2 are independent if their joint distribution is the product of their individual marginal distributions. If our two time series X_t , Y_t are independent in this sense then they are also linearly independent in the sense of 2.17d), but the converse is not necessarily true. This is analogous to the fact that independent variables have zero correlation whilst variables with zero correlation are not in general necessarily independent. Thus 2.17d) may be thought of as analogous to X_t and Y_t

being uncorrelated, whilst 2.17a)-c) represent a directional form of correlation. This will be illustrated in definition 2.21, which is a directional squared correlation coefficient for time series.

2.6.2 Tests and Measures for Causality, Feedback and Independence

Having defined various types of causality between two time series we need to know how to model and test for them. This has been done recently by Geweke(1982a,1984) who systematically brings together and builds on earlier work done by, for example, Sims(1972), Granger(1969) and Pierce(1979). What follows is a brief description of Geweke's results together with a further test proposed by Haugh(1976).

Consider two multivariate processes $\{X_t\}$ and $\{Y_t\}$ of dimension k and l respectively. We will assume throughout that the joint process is stationary and invertible (maybe after transformation and differencing.) If we wish to predict future values of $\{X_t\}$, we may employ an appropriate time series model. If it is known how $\{Y_t\}$ influences $\{X_t\}$ then this may be included in the model. We will consider four possible information sets that could be used to predict $\{X_t\}$:-

- 1) $X_{t,p}$
- 2) $X_{t,p}$ and $Y_{t,p}$
- 3) $X_{t,p}$ and $Y_{t,pp}$
- 4) $X_{t,p}$ and $Y_{t,ppp}$

Geweke's(1982a) measures of dependence consist of comparing the above information sets with each other. Clearly we may symmetrically examine four information sets to predict $\{Y_t\}$, and may also compare these. We can express 1)-4) above in terms of four models for $\{X_t\}$ and similarly $\{Y_t\}$. We consider linear models, and collecting them pairwise for X_t and Y_t we have:-

$$1) \quad \phi_{11}^{(1)}(B)X_t = a_t^{(1)} \quad , \quad (2.6.1)$$

$$\phi_{22}^{(1)}(B)Y_t = b_t^{(1)} \quad , \quad (2.6.2)$$

where $a_t^{(1)}$, $b_t^{(1)}$ are white noise, but may be correlated with each other at various lags, and:-

$$\text{Var}(\underline{a}_t^{(1)}) = \underline{\Sigma}(\underline{X}|\underline{X}_p) \quad ; \quad \text{Var}(\underline{b}_t^{(1)}) = \underline{\Sigma}(\underline{Y}|\underline{Y}_p) \quad ;$$

and
$$\underline{\Phi}_{ii}^{(1)}(B) = \sum_{j=0}^{\infty} \underline{\Phi}_{ii,j}^{(1)} B^j \quad , \quad i=1,2 \quad ;$$

where
$$\underline{\Phi}_{ii,j}^{(1)} = \underline{I}_r \quad , \quad i=1,2; j=0; \quad \begin{matrix} r=k & \text{when } i=1 \\ r=l & \text{when } i=2 \end{matrix}$$

$$2) \quad \left. \begin{aligned} \underline{\Phi}_{11}^{(2)}(B)\underline{X}_t + \underline{\Phi}_{12}^{(2)}(B)\underline{Y}_t &= \underline{a}_t^{(2)} \quad , \\ \underline{\Phi}_{21}^{(2)}(B)\underline{X}_t + \underline{\Phi}_{22}^{(2)}(B)\underline{Y}_t &= \underline{b}_t^{(2)} \quad , \end{aligned} \right\} \quad (2.6.3)$$

where $\underline{a}_t^{(2)}$, $\underline{b}_t^{(2)}$ are white noise that may be correlated with one another but only contemporaneously, and where

$$\begin{aligned} \text{Var}(\underline{a}_t^{(2)'}, \underline{b}_t^{(2)'})' &= \underline{\Sigma}(\underline{X}, \underline{Y} | \underline{X}_p, \underline{Y}_p)_{(k+l) \times (k+l)} \\ &= \begin{bmatrix} \underline{\Sigma}(\underline{X} | \underline{X}_p, \underline{Y}_p)_{k \times k} & \underline{\Gamma}(\underline{X}, \underline{Y} | \underline{X}_p, \underline{Y}_p)_{k \times l} \\ \underline{\Gamma}(\underline{Y}, \underline{X} | \underline{X}_p, \underline{Y}_p)_{l \times k} & \underline{\Sigma}(\underline{Y} | \underline{X}_p, \underline{Y}_p)_{l \times l} \end{bmatrix} \end{aligned} \quad (2.6.4)$$

and
$$\underline{\Phi}_{ij}^{(2)}(B) = \sum_{r=0}^{\infty} \underline{\Phi}_{ij,r}^{(2)} B^r \quad , \quad i=1,2$$

but with $\underline{\Phi}_{11,0}^{(2)} = \underline{I}_k \quad ; \quad \underline{\Phi}_{12,0}^{(2)} = \underline{0} \quad ; \quad \underline{\Phi}_{21,0}^{(2)} = \underline{0} \quad ; \quad \underline{\Phi}_{22,0}^{(2)} = \underline{I}_l \quad . \quad (2.6.5)$

$$3) \quad \left. \begin{aligned} \underline{\Phi}_{11}^{(3)}(B)\underline{X}_t + \underline{\Phi}_{12}^{(3)}(B)\underline{Y}_t &= \underline{a}_t^{(3)} \quad , \\ \underline{\Phi}_{21}^{(3)}(B)\underline{X}_t + \underline{\Phi}_{22}^{(3)}(B)\underline{Y}_t &= \underline{b}_t^{(3)} \quad , \end{aligned} \right\} \quad (2.6.6)$$

where, $\underline{a}_t^{(3)}$ and $\underline{b}_t^{(3)}$ are uncorrelated white-noise with:-

$$\left. \begin{aligned} \text{Var}(\underline{a}_t^{(3)}) &= \underline{\Sigma}(\underline{X} | \underline{X}_p, \underline{Y}_{pp}) \quad ; \\ \text{Var}(\underline{b}_t^{(3)}) &= \underline{\Sigma}(\underline{Y} | \underline{Y}_p, \underline{X}_{pp}) \quad , \end{aligned} \right\} \quad (2.6.7)$$

where $\underline{\Phi}_{ij}^{(3)}(B)$ are as in 2) except that instead of (2.6.5) we have

$$\underline{\Phi}_{11,0}^{(3)} = \underline{I}_k \quad ; \quad \underline{\Phi}_{22,0}^{(3)} = \underline{I}_l \quad (2.6.8)$$

$$4) \quad \left. \begin{aligned} \underline{\phi}_{11}^{(4)} \underline{X}_t + \underline{\phi}_{12}^{(4)} \underline{Y}_t &= \underline{a}_t^{(4)} \quad , \\ \underline{\phi}_{21}^{(4)} \underline{X}_t + \underline{\phi}_{22}^{(4)} \underline{Y}_t &= \underline{b}_t^{(4)} \quad , \end{aligned} \right\} \quad (2.6.9)$$

$$\text{with} \quad \left. \begin{aligned} \text{Var}(\underline{a}_t^{(4)}) &= \underline{\Sigma}(\underline{X}|\underline{X}_p, \underline{Y}_{fpp}) \quad , \\ \text{Var}(\underline{b}_t^{(4)}) &= \underline{\Sigma}(\underline{Y}|\underline{Y}_p, \underline{X}_{fpp}) \quad , \end{aligned} \right\} \quad (2.6.10)$$

where $\underline{\phi}_{ii}^{(4)}(B)$ $i=1,2$ are of the same form as in 2) and 3), but

$$\underline{\phi}_{ij}^{(4)}(B) = \sum_{r=-\infty}^{\infty} \underline{\phi}_{ij,r}^{(4)} b^r \quad i \neq j=1,2 \quad (2.6.11)$$

Clearly all the above models may be re-written in matrix form:-

$$\begin{bmatrix} \underline{\phi}_{11}^{(i)}(B) & \underline{\phi}_{12}^{(i)}(B) \\ \underline{\phi}_{21}^{(i)}(B) & \underline{\phi}_{22}^{(i)}(B) \end{bmatrix} \begin{bmatrix} \underline{X}_t \\ \underline{Y}_t \end{bmatrix} = \begin{bmatrix} \underline{a}_t^{(i)} \\ \underline{b}_t^{(i)} \end{bmatrix} \quad (2.6.12)$$

, $i=1,2,3,4$;

with $\underline{\phi}_{12}^{(1)} = \underline{\phi}_{21}^{(1)} = \underline{0}$.

Clearly if

- i) $X \parallel Y$, all the models reduce to 1)
- ii) $X \rightarrow Y$ only, 3) reduces to 2) with, $\underline{\phi}_{12}^{(2)}(B)=\underline{0}$; $\underline{\phi}_{11}^{(2)}(B)=\underline{\phi}_{11}^{(1)}(B)$; $\underline{a}_t^{(2)}$ is uncorrelated with $\underline{b}_t^{(2)}$ i.e. $\underline{\Gamma}(\underline{X}, \underline{Y}|\underline{X}_p, \underline{Y}_p)=\underline{0}$.
- iii) $X \leftrightarrow Y$ only, 3) reduces to 2) i.e. $\underline{\Gamma}(\underline{X}, \underline{Y}|\underline{X}_p, \underline{Y}_p)=\underline{0}$.
- iv) X, Y only, 2) reduces to 1) with $\underline{a}_t^{(1)}$ and $\underline{b}_t^{(1)}$ correlated only at lag 0. 4) reduces to 3), but with $\underline{\phi}_{ij,r}^{(3)} = \underline{0} \quad i \neq j=1,2; r \neq 0$.
- v) $X \Leftrightarrow Y$ then $\underline{a}_t^{(1)}$ and $\underline{b}_t^{(1)}$ are correlated at various lags. Also model 2), 3), the first k equations of model 4), or the last k equations of model 4) are all alternative representations of the same form.

In the light of these properties Geweke(1982a) defines the following measures of causality. (We use a different terminology.)

Definition 2.18

The measure of linear causality from Y to X is:-

$$F_{Y \rightarrow X} = \ln \left[\frac{|\underline{\Sigma}(\underline{X}|\underline{X}_p)|}{|\underline{\Sigma}(\underline{X}|\underline{X}_p, \underline{Y}_p)|} \right] .$$

symmetrically

$$F_{X \rightarrow Y} = \ln \left[\frac{|\underline{\Sigma}(\underline{Y}|\underline{Y}_p)|}{|\underline{\Sigma}(\underline{Y}|\underline{Y}_p, \underline{X}_p)|} \right] .$$

Definition 2.19

The measure of instantaneous causality between X and Y is:-

$$F_{X.Y} = \ln \left[\frac{|\underline{\Sigma}(\underline{X}|\underline{X}_p)| \cdot |\underline{\Sigma}(\underline{Y}|\underline{Y}_p)|}{|\underline{\Sigma}(\underline{X}, \underline{Y}|\underline{X}_p, \underline{Y}_p)|} \right] .$$

Definition 2.20

The measure of total linear dependence

$$F_{X \Leftrightarrow Y} = \ln \left[\frac{|\underline{\Sigma}(\underline{X}|\underline{X}_p)| \cdot |\underline{\Sigma}(\underline{Y}|\underline{Y}_p)|}{|\underline{\Sigma}(\underline{X}, \underline{Y}|\underline{X}_p, \underline{Y}_p)|} \right] .$$

It follows directly from the above definitions that:-

$$F_{X \Leftrightarrow Y} = F_{X \rightarrow Y} + F_{Y \rightarrow X} + F_{X.Y} \quad (2.6.13)$$

Geweke(1982a) also gives the following results:-

$$F_{X \Leftrightarrow Y} = \ln \left[\frac{|\underline{\Sigma}(\underline{X}|\underline{X}_p)|}{|\underline{\Sigma}(\underline{X}|\underline{X}_p, \underline{Y}_{fpp})|} \right] = \ln \left[\frac{|\underline{\Sigma}(\underline{Y}|\underline{Y}_p)|}{|\underline{\Sigma}(\underline{Y}|\underline{Y}_p, \underline{X}_{fpp})|} \right] ,$$

$$F_{X \rightarrow Y} = \ln \left[\frac{|\underline{\Sigma}(\underline{X}|\underline{X}_p, \underline{Y}_{pp})|}{|\underline{\Sigma}(\underline{X}|\underline{X}_p, \underline{Y}_{fpp})|} \right] ,$$

$$F_{Y \rightarrow X} = \ln \left[\frac{|\underline{\Sigma}(\underline{Y}|\underline{Y}_p, \underline{X}_{pp})|}{|\underline{\Sigma}(\underline{Y}|\underline{Y}_p, \underline{X}_{fpp})|} \right] . \quad (2.6.14)$$

The motivation behind these measures should be clear from properties i)-v). For example, $F_{Y \rightarrow X}$ is the log-ratio of the prediction error matrices of \underline{X} based on past \underline{X} , vs. past \underline{X} and past \underline{Y} . Thus if $Y \nrightarrow X$ so that the denominator equals the numerator $F_{Y \rightarrow X} = 0$. Similarly for $F_{X \rightarrow Y}$, if there is no instantaneous causality, $\underline{a}_t^{(2)}$ will be uncorrelated to $\underline{b}_t^{(2)}$ so that $\Gamma(\underline{X}, \underline{Y} | \underline{X}_p, \underline{Y}_p) = \underline{0}$ and again the denominator will equal the numerator so that $F_{X \rightarrow Y} = 0$.

We note that:-

$$0 < F < \infty .$$

Pierce(1982) relates Geweke's measures to his R^2 measure for time series (Pierce(1979)). This R^2 measure is perhaps easier to interpret since:-

$$0 < R^2 < 1 ,$$

and it represents the proportion of variance explained by allowing the particular dependence under consideration to hold. It may be defined in terms of F .

Definition 2.21

$$R^2 = 1 - e^{-F} .$$

Thus for example, $R_{X \rightarrow Y}^2 = 1 - e^{-F_{X \rightarrow Y}}$, and represents the proportion of the prediction error for $\underline{Y} | \underline{Y}_p$ explained by \underline{X}_p .

Although R^2 is perhaps easier to interpret, it does not have an equivalent additive property like (2.6.13). It is possible to form a third type of measure, which is defined below.

Definition 2.22

$$W = e^F - 1 .$$

To understand the motivation behind W consider the null hypothesis that a given measure of dependence is zero, then, when estimates of the covariance matrices are made, F is the log-likelihood ratio test-statistic, R^2 the Lagrange-multiplier test, and W the Wald

test. A comparison of these three tests is discussed by Geweke et al(1983). The comparison is made in terms both of asymptotic theory, and by a simulation study.

For our purposes, we will utilize F, since it is well developed in this context Geweke(1982a,1984), it has the additive property (2.6.13), and the other two measures may be expressed in terms of it.

Having introduced the above measures, we need to know how to estimate them. The first step requires that we truncate all the lagged polynomial expressions in (2.6.1) through to (2.6.12) at a point p (say), so that:-

$$\phi_{ij,r}^{(k)} = 0 \quad |r| > p; k=1,2,3,4; i,j=1,2.$$

This leaves us with a finite number of parameters to estimate. Although p may be chosen in a similar manner to that for identification (section 2.2), we are not seeking here to identify an exact order for the model, but instead a truncation point for an infinite order model, which will finally result in a good estimate of F. Geweke(1982a) states that "p should be allowed to increase with sample size" so that consistency may be maintained. However the identification techniques of section 2.2 should aid us, if only to set a lower limit for p.

Having selected p, the parameters, residuals and hence the required variance-covariance matrices may be estimated using least squares or other standard techniques such as those in section 2.3. F so formed is then the maximum likelihood statistic, which asymptotically follows a chi-square distribution under the null hypothesis that the dependence relationship does not hold. The degrees of freedom are simply equal to the difference in the number of parameters of the two models estimated.

$$\begin{aligned} H_0: Y \nrightarrow X \quad (F_{Y \rightarrow X} = 0) & \quad \hat{nF}_{Y \rightarrow X} \sim \chi^2_{(k\ell p)}, \\ H_0: X \nrightarrow Y \quad (F_{X \rightarrow Y} = 0) & \quad \hat{nF}_{X \rightarrow Y} \sim \chi^2_{(k\ell p)}, \\ H_0: X \not\perp Y \quad (F_{X.Y} = 0) & \quad \hat{nF}_{X.Y} \sim \chi^2_{(k\ell)}, \\ H_0: X \parallel Y \quad (F_{X \Leftrightarrow Y} = 0) & \quad \hat{nF}_{X \Leftrightarrow Y} \sim \chi^2_{(k\ell(2p+\ell))}. \end{aligned} \quad (2.6.15)$$

If the null hypothesis does not hold then \hat{F} is an estimate of the causality measure F, and follows a non-central chi-square (asymptotically) with the same degrees of freedom, and non-centrality

parameter F. For example:-

$$n\hat{F}_{Y \rightarrow X}^a \sim \chi'^2(k \neq p, nF_{Y \rightarrow X}) \quad (2.6.16)$$

Thus confidence intervals may be estimated for the measures, and Geweke(1982a) suggests some approximations to the non-central chi-square distribution to aid this.

There are several other approaches to testing for independence between time series. Some of the more familiar are formulated in the frequency domain such as the phase and coherence. However we are confined to the time domain. We mention one final method for investigating the independence between two univariate series, as developed by Haugh(1976).

Definition 2.23

Let $\{X_t\}$ and $\{Y_t\}$ be two univariate series, and $\{a_t\}$ and $\{b_t\}$ two series formed by pre-whitening X_t and Y_t respectively. Then under $H_0 X \parallel Y$

$$S = n \sum_{k=-q}^q \hat{r}_{a,b}^2(k) \sim \chi_{2q+1}^2$$

$$S^* = \sum_{k=-q}^q \frac{\hat{r}_{a,b}^2(k)}{n - |k|} \sim \chi_{2q+1}^2$$

where $\hat{r}_{a,b}(k)$ is the sample cross-correlation of the estimated white noise series $\{a_t\}$ and $\{b_t\}$. (Definition 2.14). We will call S and S^* Haugh's test for independence.

The motivation for the Haugh test comes from the fact that if two series are independent then their cross-correlations should be zero. However we cannot simply look at the cross-correlations of the raw series since the autocorrelation in each series may inflate the values of the cross-correlations in a spurious manner. However, it is possible to investigate two white-noise series in this way since no autocorrelation exists. Thus by first pre-whitening, the resulting cross-correlations become easy to interpret. The resulting S or S^* test-statistics are then simply a sum of these cross-correlations, and clearly are very similar to the portmanteau goodness-of-fit statistic.

2.7 Summary

In this chapter we have defined the multivariate $ARMA_m(p,q)$ model, and examined some of its properties. In particular we have examined the four stages of the Box-Jenkins approach to fitting such models. Finally we have examined the concept of Wiener-Granger causality and presented some measures of dependence. In chapter 6 these measures will be extended to apply to compositional time series.

"And we know that all things work together for good
to them that love God..."

Romans 8:28 (KJV)

CHAPTER 3

Compositional Data

3.0 Introduction

In this chapter we consider the nature of compositional, or sum-constrained data. Such data occurs in many situations, and for many reasons, as was seen in chapter 1. As with all data, the type of analysis carried out depends on the exact context in which it occurs and the insight into some area of study that the analyst wishes to gain. Many of the examples given in chapter 1 have been investigated when the compositional data consists of independent stochastic observations. The geologist may take independent random soil samples; his interest is often that he wants to determine if the presence of one constituent influences the presence or absence of another. The public opinion poll of political preference may be carried out in each constituency on a single occasion, and interest may be in how the preference for one political party relates to the preferences for the remaining parties. Household expenditure may be studied by taking a sample from the population stratified by income size. A comparison of expenditure is then made within and between strata.

In this chapter we introduce a mathematical framework in which to study compositional data. We then discuss some of the problems in detail, and in particular those specific to ARMA models. We will briefly discuss some of the approaches formulated to overcome these difficulties, and in more detail the approach of Aitchison(1982). It is upon this that we shall develop a means of modelling compositional time series in subsequent chapters.

3.1 A formal description of compositional data

Our notation is similar to that of Aitchison(1982). Let \mathbb{R}^m be the m -dimensional real space, and let \mathbb{P}^m be the positive orthant of \mathbb{R}^m i.e.

$$\mathbb{P}^m = \{(w_1, w_2, \dots, w_m) : \underline{w} \in \mathbb{R}^m; w_i > 0, (i=1, \dots, m)\}. \quad (3.1.1)$$

Let the positive simplex of \mathbb{R}^m be \mathbb{S}^m i.e.

$$\mathbb{S}^m = \{(u_1, \dots, u_m) : \underline{u} \in \mathbb{R}^m; u_i > 0, (i=1, \dots, m); u_1 + \dots + u_m < 1\}. \quad (3.1.2)$$

Definition 3.1

Any vector $\underline{u} \in \mathbb{S}^m$ is said to be a composition.

A data set consisting of compositions is said to be compositional data.

Let $u_{m+1} = 1 - u_1 - \dots - u_m$ be the "fill-up-value" (FUV). (3.1.3)

and let $\underline{u}^{(m+1)} = (u_1, u_2, \dots, u_{m+1})'$ be the vector of the composition

inclusive of u_{m+1} . (3.1.4)

Definition 3.2

For $\underline{w} \in \mathbb{P}^{m+1}$ we will define the function T(w) by:-

$$T(\underline{w}) = \sum_{i=1}^{m+1} w_i,$$

and the function C(w) by:-

$$\{C(\underline{w})\}_i = \frac{w_i}{T(\underline{w})}, \quad i=1, 2, \dots, m.$$

Sometimes a compositional data set is formed from a data set on \mathbb{P}^{m+1} .

For example if $\underline{w} \in \mathbb{P}^{m+1}$ consists of the number of deaths in one particular country over a specified time period broken down by causes of death; then the total number of deaths is $T(\underline{w})$ and the composition by cause of death is $C(\underline{w})$. We state this formally.

Definition 3.3

For a composition $\underline{u} \in \mathbb{S}^m$ such that there exists $\underline{w} \in \mathbb{P}^{m+1}$ and where

$$\underline{u} = C(\underline{w}),$$

then \underline{w} is said to form the basis of \underline{u} .

Many studies may wish to focus on the relative proportions of some subset of a composition \underline{u} . This may easily be done by allowing this subset of c components (say) to form a basis for a new composition on \mathbb{S}^{c-1} .

Definition 3.4

Let $\underline{u}_{(c)}$ be a subset of c elements ($c \leq m$) of $\underline{u}^{(m+1)}$, where $\underline{u} \in \mathbb{S}^m$. Then the composition $C(\underline{u}_{(c)}) \in \mathbb{S}^{c-1}$ is said to be a subcomposition of \underline{u} .

Often we may wish to concentrate on broader categories, for example in the household expenditure study our categories may be :- "food", "heating", "clothing", "holiday", "hobbies" etc. We may wish to combine these into categories such as "essentials", "luxury" etc. Thus we have :-

Definition 3.5

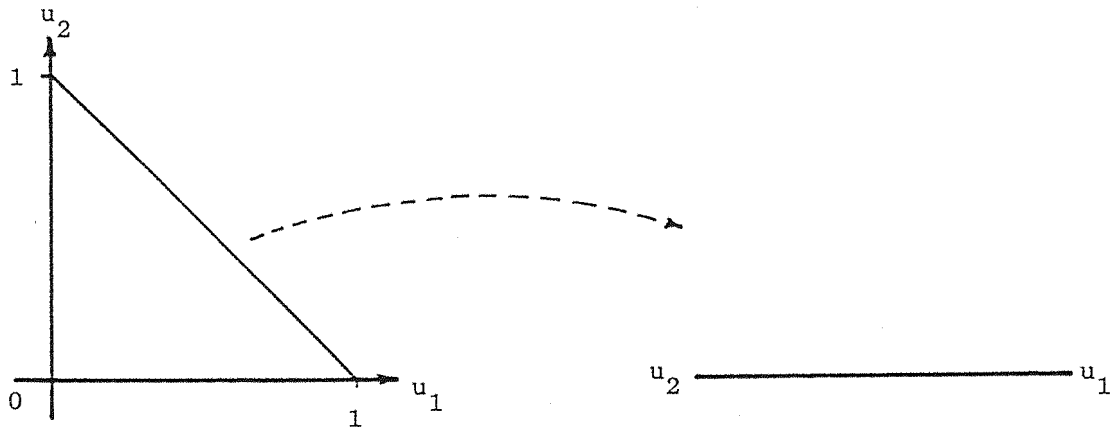
An amalgamation of a composition $\underline{u} \in \mathbb{S}^m$ is a composition $\underline{t} \in \mathbb{S}^k$, $k \leq m$; formed by combining some of the components of $\underline{u}^{(m+1)}$. If we assume that \underline{u} is arranged so that the combinations are between neighbouring u_i 's then an amalgamation $\underline{t} \in \mathbb{S}^m$ is formed by

$$t_j = u_{a_{j-1}+1} + \dots + u_{a_j}, \quad j=1, \dots, k+1,$$

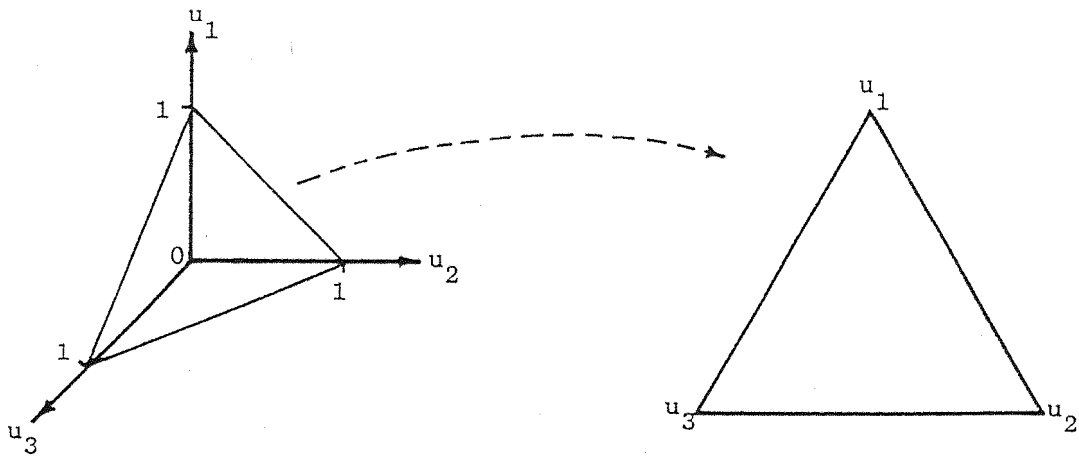
where a_0, \dots, a_{k+1} are integers such that,

$$0 = a_0 < a_1 < \dots < a_k < a_{k+1} = m+1.$$

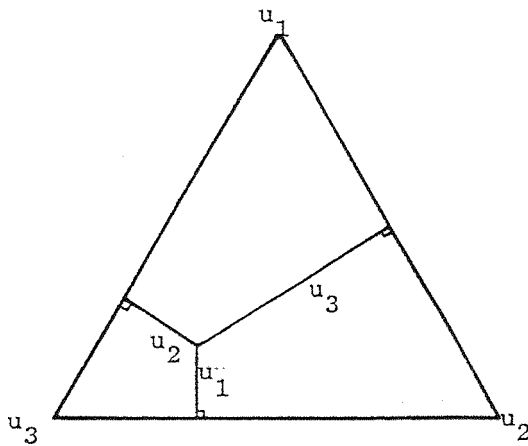
Compositional data can be represented graphically as follows. If $\underline{u} \in \mathbb{S}^1$, then the locus of $\underline{u}^{(2)}$ will yield a straight line between the points (1,0) and (0,1). Thus although $\underline{u}^{(2)}$ seems to be 2-dimensional in that it consists of u_1 and u_2 , it is uni-dimensional (hence \mathbb{S}^1), and this can be represented graphically thus:-



Similarly for $\underline{u} \in \mathbb{S}^2$ we may plot $\underline{u}^{(3)} = \{u_1, u_2, u_3\}$ in 2-dimensions:-

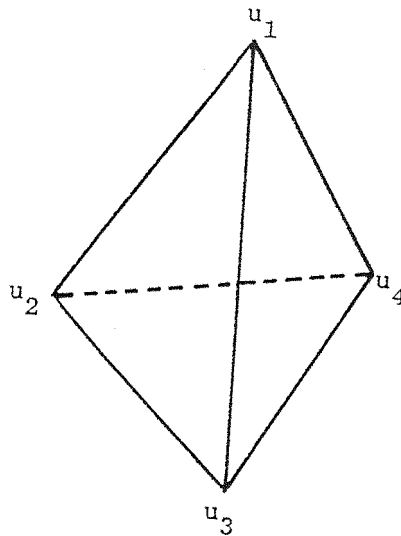


The coordinate system of the two dimensional ternary diagram is obtained from the perpendicular distances from each side with the perpendicular of the triangle scaled to be 1.



$$\underline{u} = \{u_1, u_2, u_3\}$$

Finally $\underline{u} \in \mathbb{S}^4$ may be represented as points lying in a tetrahedron:-



We now move on to consider some of the properties of compositional data.

3.2 The nature of compositional data

Because of the sum-constraint, data on the simplex have proved difficult to handle statistically. Problems arise in at least four interrelated areas: interpreting correlations, independence of the variables, the distribution of these variables, and modelling.

First we consider how the correlations of \underline{u} give spurious results. Let $\underline{u}_k^{(m+1)}$ $k=1,2,\dots,n$ be a compositional data set. Let the sample covariance and correlation matrices be

$$\underline{C} = \{ c_{ij} \}, \text{ where}$$

$$c_{ij} = \frac{1}{n} \sum_{k=1}^n (u_{ki} - \bar{u}_i)(u_{kj} - \bar{u}_j), \quad i,j=1,2,\dots,m+1;$$

$$\bar{u}_i = \frac{1}{n} \sum_{k=1}^n u_{ki}, \quad i=1,2,\dots,m+1. \quad (3.2.1)$$

Also let $\underline{P} = \{ \rho_{ij} \}$, where

$$\rho_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}, \quad i,j=1,2,\dots,m. \quad (3.2.2)$$

The inherent linear dependency in $\underline{u}^{(m+1)}$ is carried over into \underline{C} and \underline{P} in such a way as to make any interpretation of them difficult. A detailed examination is given by Chayes(1960) and is summarized below in the form of six lemmas.

Lemma 3.6

$$\sum_{i=1}^{m+1} c_{ij} = 0 \quad .$$

Every row (and hence every column) of the covariance matrix sums to zero.

Lemma 3.7

Of the $\binom{m+1}{2}$ correlations ρ_{ij} , $i=1,\dots,m+1, j=i+1,\dots,m+1$; at least m must be negative.

Lemma 3.9

If the standard deviation of one variable is greater than the sum of any j of the other standard deviations, at least $(j+1)$ of the covariances of that variable must be negative.

Lemma 3.10

If $c_{jj} \geq \sum_{\substack{i=1 \\ i \neq j}}^{m+1} c_{ii}$, then one or more of the covariances c_{kl} , $k \neq j, l \neq k$

must be positive, while all the covariances c_{jl}, c_{kj} , $l, k \neq j$, are all negative.

Lemma 3.11

i) For $m+1 = 3$, any assumed or observed set of variances completely fixes all three correlations.

ii) For $m+1=4$, if any two covariances having a common variable are known, the remaining four covariances may be expressed as an additive function of these two and the variances. Thus in a four variable closed table there will never be more than two potentially independent correlations.

The above relationships demonstrate that the correlations of compositional data are difficult to interpret. For example the use of $\rho = 0$ as a criterion for independence is no longer valid. (See later for definitions of independence.) Chayes(1960) suggests using $\rho = -1/m$, however it is clear that even using this as a criterion the data clearly cannot follow a normal distribution, and it is still difficult to infer independence.

In view of this several authors have tried to find new ways of defining independence. Before looking at these in more detail, we first examine a brief expansion of Chayes work to cross-correlations between two compositional data sets.

Consider x_k, y_k ($k=1, \dots, n$) such that $x_k, y_k \in \mathbb{S}^m$. Let c_{xy} be the sample cross-covariance matrix with elements:-

$$c_{x_i y_j} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(y_{kj} - \bar{y}_j), \text{ where} \tag{3.2.3}$$

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki}, \quad \bar{y}_j = \frac{1}{n} \sum_{k=1}^n y_{kj}; \quad i, j=1, \dots, m+1.$$

Then we have the following result:-

Lemma 3.12

$$\sum_{i=1}^{m+1} \left[c_{x_k y_i} + c_{x_i y_k} \right] = 0, \quad k=1, \dots, m+1. \quad (3.2.4)$$

That is the sum of the k^{th} row elements together with the k^{th} column elements (the $k^{\text{th}}, k^{\text{th}}$ element being added in twice) is zero.

Proof

$$\begin{aligned} \text{Let } X_{ij} &= x_{ij} - \bar{x}_i, \\ Y_{ij} &= y_{ij} - \bar{y}_j. \quad i=1, \dots, m+1; j=1, \dots, n. \end{aligned} \quad (3.2.5)$$

$$\text{Since } \sum_{i=1}^{m+1} x_{ij} = \sum_{i=1}^{m+1} y_{ij} = \sum_{i=1}^{m+1} \bar{x}_i = \sum_{i=1}^{m+1} \bar{y}_i = 1, \text{ then}$$

$$\sum_{i=1}^{m+1} X_{ij} = \sum_{i=1}^{m+1} Y_{ij} = 0; \quad j=1, \dots, n. \quad (3.2.6)$$

In particular rewriting (3.2.6) as :-

$$X_{1j} + X_{2j} + \dots + X_{m+1,j} = 0, \quad (3.2.7)$$

$$Y_{1j} + Y_{2j} + \dots + Y_{m+1,j} = 0. \quad (3.2.8)$$

If we multiply (3.2.7) and (3.2.8) together, sum over j and divide by n we obtain :-

$$\sum_{i=1}^{m+1} c_{x_i y_i} + \sum_{i=1}^{m+1} \sum_{\substack{j=1 \\ i \neq j}}^{m+1} c_{x_i y_j} = 0. \quad (3.2.9)$$

Similarly we may rewrite (3.2.6) as :-

$$-X_{1j} = X_{2j} + \dots + X_{m+1,j}, \quad (3.2.10)$$

$$-Y_{1j} = Y_{2j} + \dots + Y_{m+1,j}, \quad (3.2.11)$$

and obtain :-

$$C_{x_1 y_1} = \sum_{i=2}^{m+1} C_{x_i y_i} + \sum_{i=2}^{m+1} \sum_{\substack{j=2 \\ i \neq j}}^{m+1} C_{x_i y_j} . \quad (3.2.12)$$

On subtracting (3.2.12) from (3.2.9) we obtain :-

$$2C_{x_1 y_1} + \sum_{i=2}^{m+1} [C_{x_1 y_i} + C_{x_i y_1}] = 0 , \quad (3.2.13)$$

which generalizes to (3.2.4). □

This lemma clearly demonstrates that various linear dependencies exist between the elements of C_{xy} , purely because both \underline{x} and \underline{y} are compositional data. It is not easy to come up with hard and fast rules about spurious negative correlations, because now the diagonal elements may be negative; whereas previously the diagonal elements represented (positive) variances resulting in spurious negative correlations in the off-diagonal. If the diagonal element is positive then some of the off-diagonal elements must be negative and vice versa. Thus the matrix C_{xy} is also difficult to interpret. We could scrutinize this still further, but it is sufficient to realize the difficulties in examining the covariance and cross-covariance matrices of compositional variables.

The difficulty in interpretation of correlations led to an introduction of some new definitions of independence because compositional variables are necessarily linearly dependent on one another. We now briefly examine some of the earlier types of independence proposed. We will denote independence by the symbol "||". The first property was suggested by Mossiman(1962).

Definition 3.13

The random variables u_1, \dots, u_{m+1} ; $\underline{u} \in \mathbb{S}^m$ are said to be independent except for the constraint if \exists a basis $\underline{w} \in \mathbb{F}^{m+1}$ ($\underline{u} = C(\underline{w})$) such that

i) $\underline{u} \parallel \underline{w}$ (3.2.14)

ii) $\underline{u} \parallel T(\underline{w})$ (3.2.15)

The transformation $\underline{w} \rightarrow \underline{u}, T(\underline{w})$ is a 1-1 transformation. If (3.2.15) holds then no additional information about the vector of proportions \underline{u} is contained within $T(\underline{w})$; that is, the transformation $\underline{w} \rightarrow \underline{u}$ "loses" no information about \underline{u} . Hence if \underline{u} is "independent except for the constraint", it originates from a basis whose elements were themselves independent, and such that $T(\underline{w})$ is a function of the basis which, is independent of the composition under investigation.

The second concept of independence is that of "neutrality", as developed by Connor and Mossimann(196.). The idea has arisen from an often required wish by the analyst to eliminate one proportion, say u_1 ,

from consideration and instead concentrate on $\left[\frac{u_2}{1-u_1}, \frac{u_3}{1-u_1}, \dots, \frac{u_m}{1-u_1} \right]$ i.e. on $C(u_2, \dots, u_{m+1})$; usually because u_1 is of little or no interest. If u_1 is omitted it is required to know if the remaining composition is affected, i.e. is u_1 "neutral". We have :-

Definition 3.14

- (i) u_1 is said to be neutral in the vector $(u_1, u_2, \dots, u_{m+1})$ if

$$u_1 \parallel C(u_2, u_3, \dots, u_{m+1}). \quad (3.2.16)$$

- (ii) The vector (u_1, u_2, \dots, u_j) is said to be neutral in the vector $\underline{u}^{(m+1)}$ if

$$(u_1, u_2, \dots, u_j) \parallel C(u_{j+1}, \dots, u_{m+1}) \quad (3.2.17)$$

- (iii) If (ii) holds for all $j=1, \dots, m$ then \underline{u} is said to be completely neutral.

Note: The ordering in the vector for this definition is vital. If (u_1, \dots, u_m) is completely neutral it does not necessarily follow that $(u_2, u_1, u_3, \dots, u_m)$, for example, is completely neutral.

These two types of compositional independence have useful interpretations. Consider the geologists soil-sample. If the constituents which make up the sample evolved by a purely random process, i.e. if the presence of one constituent does not indicate the presence or

absence of another, then there exists a basis with independent elements. Further we would not expect the properties of the soil to vary according to how large a soil sample was taken. Hence we would have independence except for the constraint.

Neutrality offers a means of examining a subset of variables without loss. For example in the political opinion poll, if the minor parties are neutral we may examine C(CON,LAB) on their own, because the relationship between the two major parties is independent of the others.

These early definitions of compositional independence provide a means of solving the sum-constraint problem. However, although useful conceptually, it proved difficult to apply these definitions in a practical way, since great difficulty was found in forming statistical tests for them. This was primarily due to a lack of the useful parametric class of distributions on the simplex \mathbb{S}^m .

As Aitchison(1982) states "Undoubtedly the only familiar class of distributions on \mathbb{S}^m is the Dirichlet class..." The Dirichlet distribution may be regarded as a multivariate generalization of the beta distribution. It may be found by considering the density of C(\underline{w}) where w_i are independent gamma random variables. This fact implies that it contains a strong independence structure within it. Mossimann(1962) shows that if \underline{u} is independent except for the constraint then \underline{u} must follow a Dirichlet distribution; the converse is also true. Thus it is possible to calculate the correlations of the u 's under the assumption that they are independent except for the constraint and compare these to the sample correlations of a compositional data set under investigation. However Mossimann(1962) could not form any appropriate statistical tests because no distribution theory was available for such correlations under a Dirichlet model. Also, in order to model the u 's themselves such that they are not independent except for the constraint an alternative distribution must be found. One such distribution which moved towards this is the Generalized Dirichlet developed by Connor and Mossimann (1964). They form variables $\zeta_1, \zeta_2, \dots, \zeta_m$; such that :-

$$\left. \begin{aligned} \zeta_1 &= u_1 \\ \zeta_i &= \frac{u_i}{1 - \sum_{j=1}^{i-1} u_j} \quad , \quad i=2, \dots, m. \end{aligned} \right\} \quad (3.2.18)$$

They show that if the vector $(u_1, \dots, u_k) \parallel C(u_{k+1}, \dots, u_{m+1})$ for $k=1, \dots, r$ then ζ_k ; $k=1, \dots, r$ are mutually independent. Hence if \underline{u} is completely neutral ζ_1, \dots, ζ_m are mutually independent. Assuming that \underline{u} is completely neutral, and letting the density function of each of the ζ_i 's be a univariate beta distribution, the Generalized Dirichlet is formed by transforming the product of these beta distributions to a multivariate distribution in terms of \underline{u} . This reduces to the Dirichlet distribution under certain parametric restrictions. Obviously this result indicates that if \underline{u} follows a Dirichlet it must be neutral. Connor and Mossimann explain further that \underline{u} follows a Dirichlet if and only if \underline{u} is completely neutral for any permutation of the u_i 's.

Using the Generalized Dirichlet distribution it is again possible to estimate the correlations of the compositional data set under the hypothesis of complete neutrality. Again even knowing the expected values of these correlations we are not helped very much as we do not know their distribution. However, both Mossimann(1962) and Connor and Mossimann(1969) employ Fisher's z - transformation as a means of inference in some numerical examples; but as Mossimann(1962) himself points out "z - values may be given, although more with hope than confidence."

Summarizing, we have seen that new forms of independence must be used in exploring compositional data. The earlier forms of independence need further study, and so the search for definitions of compositional independence have continued. Also a more general distribution to employ on compositional data is required. The ones which we will be using are the result of the work of Aitchison(1981,1982), and are introduced later in this chapter.

3.3 Compositional time series

In this section we will demonstrate why the conventional approach to time series analysis fails to incorporate the nature of a compositional time series.

Definition 3.15

A time series $\{\underline{u}_t\}$, $t = 0, \pm 1, \pm 2, \dots$ such that $\underline{u}_t \in \mathbb{S}^m$ is said to be a compositional time series.

We first consider the problems arising in trying to use the usual multivariate ARMA model for a compositional time series. Consider the $AR_{m+1}(1)$ model for a compositional time series $\underline{u}_t^{(m+1)}$, with $\underline{u}_t \in \mathbb{S}^m$.

$$\underline{u}_t^{(m+1)} = -\underline{\phi} \underline{u}_{t-1}^{(m+1)} + \underline{\varepsilon}_t, \quad (3.3.1)$$

where
$$\sum_{i=1}^{m+1} u_{it} = \sum_{i=1}^{m+1} u_{i,t-1} = 1.$$

Summing down the columns of (3.3.1) we have :-

$$\sum_{i=1}^{m+1} \left[\sum_{j=1}^{m+1} -\phi_{ij} u_{j,t-1} + \varepsilon_{it} \right] = 1; \quad (3.3.2)$$

$$\Rightarrow \sum_{k=1}^{m+1} \left[u_{k,t-1} \phi_{+k} + \varepsilon_{kt} \right] = 1, \quad (3.3.3)$$

where
$$\phi_{+k} = -\sum_{i=1}^{m+1} \phi_{ik}, \quad k=1, \dots, m+1.$$

We may generalize (3.3.3) to higher order models, thus for the

$ARMA_{m+1}(p,q)$ process:-

$$\underline{u}_t^{(m+1)} + \underline{\phi}_1 \underline{u}_{t-1}^{(m+1)} + \dots + \underline{\phi}_p \underline{u}_{t-p}^{(m+1)} = \underline{\varepsilon}_t + \underline{\theta}_1 \underline{\varepsilon}_{t-1} + \dots + \underline{\theta}_q \underline{\varepsilon}_{t-q},$$

we obtain :-

$$\sum_{k=1}^{m+1} \left[\sum_{r=1}^p u_{k,t-r} \phi_{r,+k} + \sum_{s=0}^q \varepsilon_{k,t-s} \theta_{s,+k} \right] = 1, \quad (3.3.4)$$

where
$$\phi_{r,+k} = -\sum_{i=1}^{m+1} \left\{ \frac{\phi}{-r} \right\}_{ik}, \quad r=1, \dots, p;$$

$$\theta_{s,+k} = -\sum_{j=1}^{m+1} \left\{ \frac{\theta}{-s} \right\}_{jk}, \quad s=0, \dots, q;$$

and
$$\frac{\theta}{-0} = \frac{1}{-m+1}.$$

The equations (3.3.3) and (3.3.4) serve to demonstrate that there are many inherent linear restrictions within the model. This may be further seen by considering individual $ARMA_1(p,q)$ models for each variable u_{it} , $i=1, \dots, m+1$. For example, consider again the AR(1) model, and in

particular $m+1$ $AR_1(1)$ models for each of the $\{u_{it}\}$ series:-

$$\begin{aligned} u_{1t} &= e_{1t} - \phi_1 u_{1,t-1} , \\ u_{2t} &= e_{2t} - \phi_2 u_{2,t-1} , \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ u_{m+1,t} &= e_{m+1,t} - \phi_{m+1} u_{m+1,t-1} ; \end{aligned}$$

summing each equation gives:-

$$\sum_{i=1}^{m+1} \left[e_{it} - \phi_i u_{i,t-1} \right] = 1. \quad (3.3.5)$$

Clearly this is of identical form to (3.3.3). We could also construct a similar expression to (3.3.4) by the same route. In all of the above three expressions we see that not only are there linear restrictions on the u_i 's, but, also on the parameters of the model, and the "random" components.

A problem with the above approach is that we are either using $m+1$ univariate models (as in (3.3.5)) or an $(m+1)$ -dimensional model for data that is essentially m -dimensional. As an illustration let us partition $\underline{\phi}$ in (3.3.1) into :-

$$\underline{\phi} = \begin{bmatrix} \underline{\phi}_{11} & \underline{\phi}_{12} \\ \underline{\phi}_{21} & \underline{\phi}_{22} \end{bmatrix}, \text{ and } \underline{\varepsilon}_t = \begin{bmatrix} \underline{\varepsilon}_t \\ \varepsilon_{m+1,t} \end{bmatrix},$$

giving:-

$$\begin{bmatrix} \underline{u}_t \\ u_{t,m+1} \end{bmatrix} = \begin{bmatrix} \underline{\phi}_{11} & \underline{\phi}_{12} \\ \underline{\phi}_{21} & \underline{\phi}_{22} \end{bmatrix} \begin{bmatrix} \underline{u}_{t-1} \\ u_{m+1,t-1} \end{bmatrix} + \begin{bmatrix} \underline{\varepsilon}_t \\ \varepsilon_{m+1,t} \end{bmatrix} \quad (3.3.6)$$

$$\Rightarrow \underline{u}_t = \underline{\phi}_{11} \underline{u}_{t-1} + \underline{\phi}_{12} u_{m+1,t-1} + \underline{\varepsilon}_t$$

$$= \phi_{11} \underline{u}_{t-1} + \phi_{12} \left[1 - \sum_{i=1}^m u_{i,t-1} \right] + \underline{\varepsilon}_t^- .$$

So that re-parameterizing we have :-

$$\underline{u}_t = \underline{\mu} + \underline{\phi}^* \underline{u}_{t-1} + \underline{\varepsilon}_t^- , \quad (3.3.7)$$

where $\phi_{ij}^* = \phi_{ij} - \phi_{i,m+1}$; $i, j=1, \dots, m$; and $\underline{\mu} = \underline{\phi}_{12}$;

$u_{m+1,t}$ is then formed as $1 - \sum_{i=1}^m u_{it}$.

Hence the (m+1)-dimensional model (3.3.1) may be derived from the m-dimensional model in (3.3.7).

However even if we recognise the need to use an appropriate dimension for our models the linear approach still poses many problems. Firstly we recall that the assumption is usually made that the $\underline{\varepsilon}_t^-$ series in (3.3.7), for example, is normally distributed. That it is not may be seen by the fact that $1 > u_{it} > 0$ and consequently, however small the variance of the ε_{it} 's, extreme values must be ruled out in order to keep u_{it} permanently within its correct bounds. (This of course excludes the case of $\text{var}(\varepsilon_{it})=0$.) Because of the difficulty in defining any such bounds, it is obvious that the use of such a model in practice may produce incorrect results, especially in the case of forecasting. If we fit a standard model to m of the variables and use it to forecast into the future, what is our guarantee that $\underline{u}_{t+1} \in \underline{\mathbb{S}}^m$? As an illustration that this is not the case we turn to the following example :-

Example 3.16

Consider the Gallup Poll data given below;

t																
1-15	3.0	3.5	2.0	1.5	2.5	2.0	2.5	4.0	3.5	2.0	2.5	2.0	3.0	2.0	3.5	
16-30	2.5	3.5	2.5	2.0	2.0	2.0	2.5	2.0	2.5	2.5	2.0	2.0	2.5	2.5	3.0	
31-45	2.0	2.5	1.0	3.0	1.5	2.0	3.5	2.0	0.5	1.5	2.0	2.0	1.5	1.5	3.0	
46-60	3.0	2.0	3.5	1.5	3.5	2.0	3.5	3.0	2.5	2.0	4.0	5.5	3.5	2.5	4.5	
61-75	4.5	3.5	5.0	4.5	4.5	7.0	6.5	5.0	3.0	4.0	5.0	4.5	6.5	5.0	3.0	
76-90	3.0	2.0	2.5	1.5	1.5	2.0	2.5	1.5	1.5	1.5	1.5	1.5	1.5	1.0	1.0	2.0
91-105	1.0	1.0	1.0	0.5	0.5	1.5	1.0	1.0	0.5	1.0	1.0	0.0	1.0	0.5	0.5	
106-108	1.0	0.5	0.0													

This data set is from the GALLUP(C) opinion poll from January 1965 - December 1973, but in reverse order. It represents the percentage of people surveyed who, when asked which party they would vote for said they would not vote for one of the three major parties. This data may be considered as $u_t \in \mathbb{S}^1$ with u_{2t} being the percentage who would vote for a major party.

A plot of the correlations (fig 3.18) suggested either an AR(1) model or the model:-

$$\Delta u_t = \mu + \varepsilon_t, \quad \text{with } \mu \text{ a constant.}$$

On fitting both models using the ESP package, it was found that the second model produced the better fit. The portmanteau statistic, based on 12 autocorrelations was 4.0 for this model; which is not significant when compared to the $\chi^2_{(11)}$ distribution. The t-test statistic for the constant term was significant at the 10% level. It is not clear therefore if it should be retained, but for the purposes of example we have kept it in. (After applying the difference operator μ is usually assumed to be zero.)

Using this model the first three forecast are :-

$$-0.1858 \quad -0.3715 \quad -0.5573$$

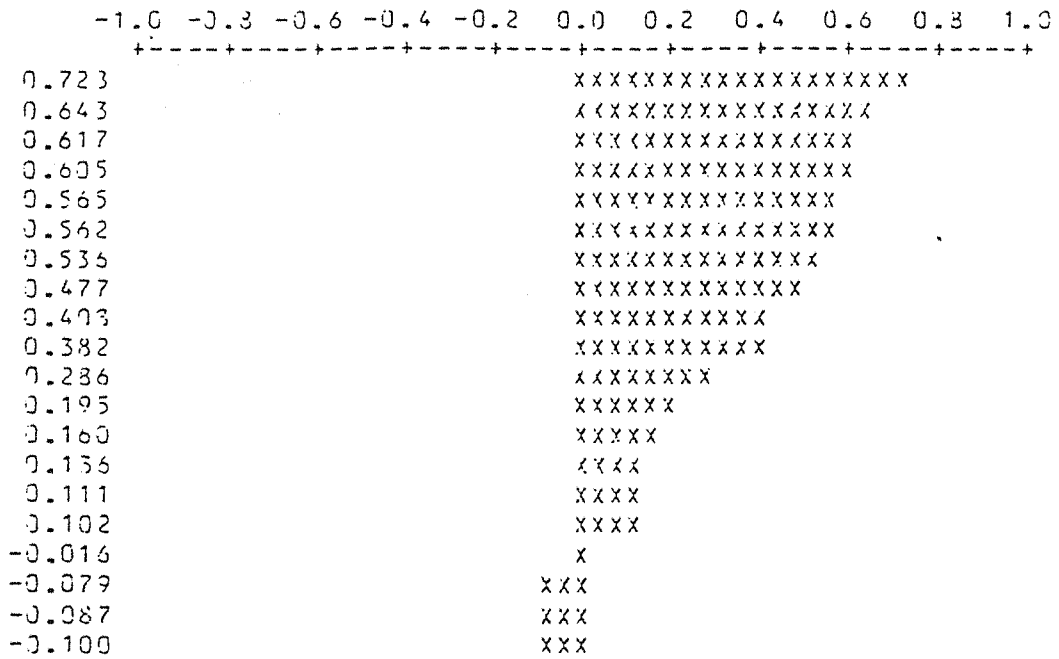
The forecasts are negative, and consequently do not lie on the simplex.

Whilst some may argue that this example is over-engineered, it still demonstrates the problem. Indeed even if another model were chosen, such as the AR(1) which does not produce negative forecasts, the confidence region around these forecasts still includes values outside the simplex.

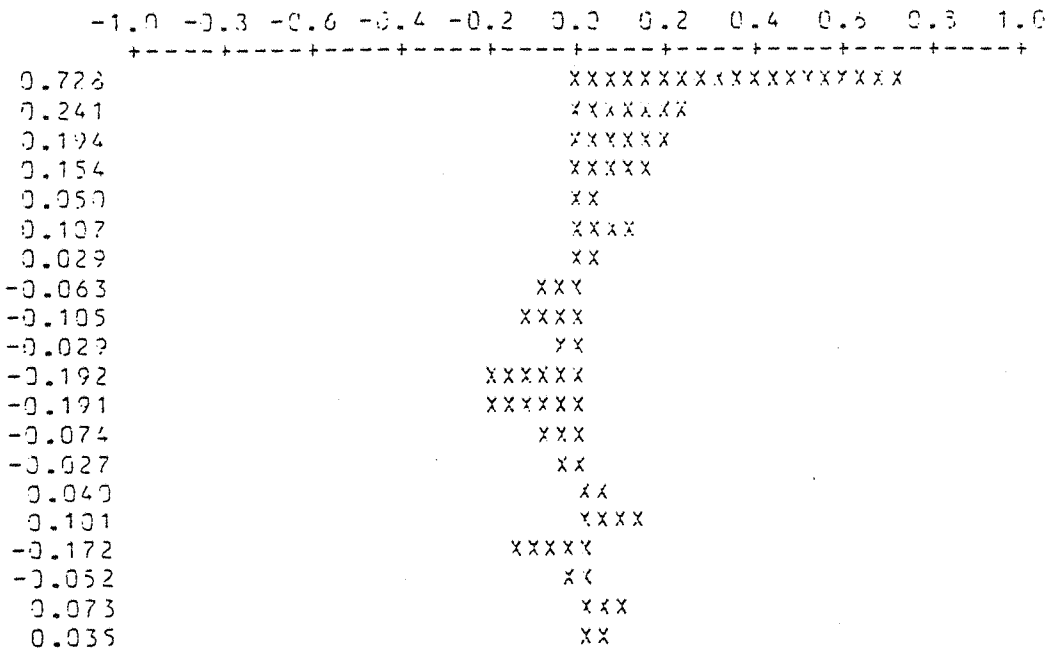
Another problem lies in the interpretation of the sample autocorrelation function. From definition 2.14 we see that it is derived from the autocovariance function, which has strong linear dependencies within it. To see this we need only consider lemma 3.12 and put :-

$$\underline{x} = \underline{u}_t^{(m+1)}, \quad \underline{y} = \underline{u}_{t-k}^{(m+1)};$$

from which we obtain :-



ACF of GALLUP(C) Series



PACF of GALLUP(C) Series

Figure 3.18

$$\sum_{i=1}^{m+1} \left[c_{li}(k) + c_{il}(k) \right] = 0 \quad ; \quad k=0, \pm 1, \dots; \quad l=1, \dots, m+1. \quad (3.3.8)$$

If we then hope to use the autocorrelation function for identification it is difficult to detect genuine departures from a white-noise process, particularly using the off-diagonal elements. Even if we consider looking at the autocorrelation function of $\{\underline{u}_t\}$ rather than $\{\underline{u}_t^{(m+1)}\}$ we are looking at exactly the same function, but with the last row and column of each of the matrices omitted. Hence interpretation is still difficult.

We could go on demonstrating how a linear model is inadequate, and there are many obvious examples, such as the sample partial autocorrelation function defined by (2.2.10), which must be based on \underline{u}_t and not $\underline{u}_t^{(m+1)}$ otherwise the $[\underline{Y}_r' \underline{Y}_r]$ matrix would be singular and consequently we would be unable to find its inverse. There are also some more sophisticated methods which will not be valid. For example Box and Tiao(1977) present a canonical covariate analysis of multiple time series which would break down in this context. (cf. Aitchison(1983) on principal component analysis on the simplex.)

In section 2.6 we discussed independence between time series, and in the previous section the problems of defining independence for data on the simplex. In the same way that it is necessary to create new forms of independence for stochastic data, it is also necessary to formulate new types of compositional causality, feedback, dependence etc. for compositional time series. The sum-constraint forces an automatic dependence between the series, making the definitions in section 2.6 difficult to interpret. This final problem will be addressed in chapter 6.

3.4 The proposed approach

In the last section we saw how the multivariate ARMA model cannot represent a compositional time series. The problem lies with the fact that the data lie on the positive simplex \mathbb{S}^m rather than the more general real space \mathbb{R}^m . This fact gives us an insight into a possible solution. The idea is a very simple one, namely to transform our data so as to map it onto the real space. What we seek is a function such that :-

$$f: \mathbb{S}^m \rightarrow \mathbb{R}^m.$$

The example cited earlier in (3.2.18) is such a transformation. We may add to the requirement that the data be transformed onto the real space, the further requirement being that it be transformed to follow a multivariate normal distribution. This is no new idea to either statistics in general or time series analysis, as can be seen by the well known paper by Box and Cox(1964).

Because of the simplicity and the availability of this approach, it is the one which we will take up here. Several possible transformations are available and many of them are introduced by Aitchison (1982). His transformations are familiar to the field of statistics, as they are multivariate versions of the logistic transformation. His choice of which transformation to use depends mainly on the particular application required. Three such transformations are :-

Definition 3.17: $f = a_m$

Let $\underline{u} \in \mathbb{S}^m$ and $\underline{v} = a_m(\underline{u})$, $\underline{v} \in \mathbb{R}^m$, where the function a_m is given by

$$a_m: \quad v_i = \ln \left(\frac{u_i}{u_{m+1}} \right) \quad ; \quad i=1, \dots, m;$$

with inverse

$$a_m^{-1}: \quad u_i = \frac{\exp(v_i)}{m + \sum_{j=1}^m \exp(v_j)} \quad ; \quad i=1, \dots, m;$$

$$u_i = \frac{1}{m + \sum_{j=1}^m \exp(v_j)} \quad ; \quad i=m+1.$$

We will call a_m the additive logistic transformation.

Defintion 3.18: $f = m_m$

Let $\underline{u} \in \mathbb{S}^m$ and $\underline{v} = m_m(\underline{u})$, with $\underline{v} \in \mathbb{R}^m$ then the function m_m is given by

$$m_m: \quad v_i = \ln \left(\frac{u_i}{1 - \sum_{j=1}^m u_j} \right) \quad ; \quad i=1, \dots, m;$$

with inverse h_m^{-1} :

$$u_i = \frac{\exp(v_i)}{\prod_{j=1}^i (1 - \exp(v_j))} ; i=1, \dots, m;$$

$$= \frac{1}{1 - \sum_{j=1}^m u_j} ; i=m+1.$$

We will call h_m the multiplicative logistic transformation.

Definition 3.19: $f=h_m$

Let $\underline{u} \in \mathbb{S}^m$ and $\underline{v} = h_m(\underline{u})$, with $\underline{v} \in \mathbb{R}^m$ then the function h_m is given by

$$h_m : v_i = \ln \left(\frac{u_i}{1-u_i} \right) \quad i=1;$$

$$= \ln \left(\frac{u_i}{\left[\begin{matrix} i-1 \\ 1 - \sum_{j=1} u_j \end{matrix} \right] \left[\begin{matrix} i \\ 1 - \sum_{j=1} u_j \end{matrix} \right]} \right) ;$$

$$i=2, \dots, m;$$

with inverse h_m^{-1} :

$$u_i = \frac{\exp(v_i)}{1 + \exp(v_i)} ; \quad i=1;$$

$$= \frac{\exp(v_i)}{\left[\begin{matrix} i-1 \\ 1 + \sum_{j=1} \exp(v_j) \end{matrix} \right] \left[\begin{matrix} i \\ 1 + \sum_{j=1} \exp(v_j) \end{matrix} \right]} ;$$

$$i=2, \dots, m;$$

$$= \frac{1}{\left[\begin{matrix} m \\ 1 + \sum_{j=1} \exp(v_j) \end{matrix} \right]} ; i=m+1.$$

We will call h_m the logistic hybrid transformation.

We note that all of the above three transformations have the same Jacobian, namely :-

$$\begin{pmatrix} D \underline{v} \\ D \underline{u} \end{pmatrix} = \begin{pmatrix} m+1 \\ \prod_{i=1} u_i \end{pmatrix}^{-1} \quad (3.4.1)$$

From these transformations we can derive other related transformations. For example we may replace the exponential function by any one-to-one transformation for

$$\mathbb{P}^1 \leftrightarrow \mathbb{R}^1,$$

although the one cited above is perhaps the most familiar one. Clearly any permutation of the \underline{u}_t 's before transforming represents a further simple variation. Aitchison gives two further ways of using the three transformations above as building blocks to obtain further transformations. Of these the so called "linear transformation method" is most useful in our context. This method involves replacing \underline{v}_t in any of the three definitions above, by \underline{Gv}_t , where \underline{G} is an $m \times m$ non-singular matrix. This is perhaps best illustrated by an example.

Example 3.20

Consider the matrix

$$\underline{G} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \text{ then } \underline{G}^{-1} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Define a new transformation $\underline{v}^* = \underline{Gv} = \underline{G}_m(\underline{u})$.

$$\begin{aligned} \text{Then } \underline{v}^* = \underline{Gv} \Rightarrow v_i^* &= \ln \frac{u_i}{u_{m+1}} - \ln \frac{u_{i-1}}{u_{m+1}} \\ &= \ln \frac{u_i}{u_{i-1}} ; \quad i=1, \dots, m-1. \\ v_m^* &= v_m = \ln \frac{u_m}{u_{m+1}} ; \quad i=m. \end{aligned}$$

So that we have:-

$$v_i^* = \ln \frac{u_i}{u_{i+1}} ; \quad i=1, \dots, m.$$

To find the inverse we note that $u_i = a_m^{-1}(v_i) = a_m^{-1}(G^{-1}v_i^*)$; $i=1, \dots, m$.

$$\begin{aligned}
 u_i &= \frac{\exp(v_i)}{1 + \sum_{j=1}^m \exp(v_j)}, \\
 &= \frac{\exp\left(\sum_{j=i}^m v_j^*\right)}{1 + \sum_{k=1}^m \exp\left(\sum_{j=k}^m v_j^*\right)}, \quad i=1, \dots, m; \\
 &= 1 - \sum_{j=1}^m u_j, \quad i=m+1.
 \end{aligned}$$

Of the three transformations mentioned above the one that Aitchison-(1982) has found most useful is the additive logistic transformation, and it is upon this transformation that Aitchison and Shen(1980) base their logistic-normal distribution. If we allow $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$, then we may form a transformed-normal class for $\underline{u} \in \mathbb{S}^m$ based on any suitable transformation. We now define two such distributions.

Definition 3.21

If $\underline{u} = a_m^{-1}(\underline{v})$, where $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$, then \underline{u} is said to follow the logistic-normal distribution: $L_m(\underline{\mu}, \underline{\Sigma})$. i.e. $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$ when the density function of \underline{u} is given by,

$$P_L(\underline{u}/\underline{\mu}, \underline{\Sigma}) = \frac{1}{|2\pi\underline{\Sigma}|^{1/2} \prod_{i=1}^{m+1} u_i} \exp\left\{-\frac{1}{2} \left(\ln \frac{\underline{u}}{u_{m+1}} - \underline{\mu}\right)' \underline{\Sigma}^{-1} \left(\ln \frac{\underline{u}}{u_{m+1}} - \underline{\mu}\right)\right\}.$$

Definition 3.22

If $\underline{u} = m_m^{-1}(\underline{v})$, where $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$, then \underline{u} is said to follow the multiplicative logistic-normal: $M_m(\underline{\mu}, \underline{\Sigma})$. i.e. $\underline{u} \sim M_m(\underline{\mu}, \underline{\Sigma})$ when the density function of \underline{u} is given by,

$$P_M(\underline{u}/\underline{\mu}, \underline{\Sigma}) = \frac{1}{|2\pi\underline{\Sigma}|^{1/2} \prod_{i=1}^{m+1} u_i} \exp\left\{-\frac{1}{2} (m_m(\underline{u}) - \underline{\mu})' \underline{\Sigma}^{-1} (m_m(\underline{u}) - \underline{\mu})\right\}.$$

We could go on in a similar manner to define further distributions on the simplex, but we shall not be using them. Of the two distributions above, the $L_m(\underline{\mu}, \underline{\Sigma})$ is the most developed, and is fully discussed by Aitchison and Shen(1980). Aitchison(1982) develops it further and discusses how it may be used to test for various types of independence defined on the simplex. These distributions do not restrict the data since the strong forms of independence implied by the Dirichlet assumption are no longer enforced. On the other hand these distributions do not contain these strong independence structures within them.

Not only may the $L_m(\underline{\mu}, \underline{\Sigma})$ and $M_m(\underline{\mu}, \underline{\Sigma})$ distributions be used to investigate independence, but various statistical models may be based on them. An example of this is that of a model for measurement error developed by Aitchison and Shen(1984). The a_m transformation is employed further by Aitchison(1983) and Aitchison(1984) in a new approach to principal component analysis for data on the simplex. In the following chapter we will introduce a new model for compositional time series, and develop its properties further in chapter 5 where we will also further discuss the $L_m(\underline{\mu}, \underline{\Sigma})$ and $M_m(\underline{\mu}, \underline{\Sigma})$ distributions.

Finally we end this section by noting that a transformation of the type given in example 3.20, when used to form an alternative logistic-transformation, will still preserve the same logistic-normal distribution as that of the underlying transformation. Thus, if \underline{v}^* in example 3.21 is allowed to follow a normal distribution, then \underline{u} will still follow a logistic-normal distribution. This is easily seen since if $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$, then $\underline{v}^* \sim N_m(G\underline{\mu}, G\underline{\Sigma}G')$, and $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$. Obviously this is true of all non-singular G , and in particular is true of a set of matrices which we will denote by $\underline{Z}(k)$. These are described below as we will find them useful in later chapters.

By $\underline{Z}(k)$ we will mean the $m \times m$ matrix with elements,

$$\begin{aligned} z_{ii}(k) &= 1 && i=1, \dots, m; i \neq k; \\ z_{ik}(k) &= -1 && i=1, \dots, m; \\ z_{ij}(k) &= 0 && \text{otherwise.} \end{aligned} \tag{3.4.2}$$

$$\text{i.e. } \underline{Z}(k) = \begin{bmatrix} 1 & 0 & \dots & -1 & \dots & 0 \\ 0 & 1 & \dots & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & \dots & 1 \end{bmatrix}$$

↑
kth column.

It is easily seen that :-

$$\underline{Z}(k)^{-1} = \underline{Z}(k). \quad (3.4.3)$$

If we repeat example 3.20 with $\underline{Z}(k)$ instead of \underline{G} and define :-

$$\underline{v}^\dagger = \underline{Z}(k)\underline{v} = \underline{Z}(k)a_m(\underline{u}),$$

then

$$\begin{aligned} v_i^\dagger &= \ln \frac{u_i}{u_{m+1}} - \ln \frac{u_k}{u_{m+1}} ; i=1, \dots, m; i \neq k; \\ &= -\ln \frac{u_k}{u_{m+1}} ; i=k. \end{aligned}$$

i.e.

$$\begin{aligned} v_i^\dagger &= \ln \frac{u_i}{u_k} ; i=1, \dots, m; i \neq k; \\ &= \frac{u_{m+1}}{u_k} ; i=k. \end{aligned}$$

Thus we see that the effect of $\underline{Z}(k)$ is to change u_k to be the reference variable in the transformation, as opposed to the FUV u_{m+1} . In other words:-

$$\underline{Z}(k)a_m(\underline{u}) = a_m(\underline{u}^\dagger),$$

where \underline{u}^\dagger is the permutation of \underline{u} such that u_k and u_{m+1} are interchanged. This set of matrices will be useful when we examine invariance properties under the choice of reference variable.

3.5 Independence of compositional data

In section 3.2 we examined the problem underlying compositional data sets, and in particular the difficulty in understanding independence in such data sets. Two forms of compositional independence (definitions 3.13 and 3.14) were given. Although these were useful it was difficult to produce any statistical tests for them. Allied to this

was the need for a new class of distributions on the simplex. In section 3.4 we introduced some new distributions based on transformations from the \mathbb{S}^m space to the \mathbb{R}^m space. In this section we explore some further concepts of independence developed by Aitchison(1982). It will be seen that the transformations a_m and m_m of section 3.4, together with their related distributions (L_m and M_m), provide an easy means of investigating these forms of independence.

Aitchison discusses two types of independence, namely extrinsic and intrinsic. Extrinsic analysis examines a composition under the assumption that it originated from a basis, and interest lies in the relationship of the composition to this basis. Intrinsic analysis concentrates on the proportions themselves, without any thought of a basis being involved. Neutrality (definition 3.14) is an example of intrinsic analysis, which is one of the types of independence considered by Aitchison.

The two sorts of extrinsic independence defined by Aitchison (1982) are as follows.

Definition 3.23

A basis $\underline{w} \in \mathbb{P}^{m+1}$ is compositionally invariant (CIB) if,

$$C(\underline{w}) \parallel T(\underline{w}).$$

Definition 3.24

A composition $\underline{u} \in \mathbb{S}^m$ is said to have basis independence (BI) if there exists a basis $\underline{w} \in \mathbb{P}^{m+1}$ such that :-

- i) $\underline{u} \parallel \underline{w}$
- (ii) $\underline{u} = C(\underline{w})$).

Clearly CIB is identical to (3.2.15) in definition 3.13, and is referred to as Lukac's condition by Mossimann(1962). Mossimann also considered BI under the name "partial independence except for the constraint". BI \cap CIB is equivalent to the concept of independence except for the constraint given by definition 3.13. Under this, as previously mentioned, the u 's must follow a Dirichlet distribution; so that if we wish to simultaneously model BI and CIB we should not appeal to the logistic-normal class.

Aitchison(1982) describes how to test for CIB and BI. Firstly for CIB, let $\tau = T(\underline{w})$, then using the transformed logistic-normal distribution we may model the \underline{u} 's as follows :-

$$\underline{u} \sim L_m(\underline{\alpha} + \underline{\beta}\tau, \underline{\Sigma}), \quad (3.5.1)$$

or
$$\underline{u} \sim L_m(\underline{\alpha} + \underline{\beta}\ln(\tau), \underline{\Sigma}). \quad (3.5.2)$$

Hence if $\underline{v} = a_m(\underline{u})$, then

$$\underline{v} \sim N_m(\underline{\alpha} + \underline{\beta}\tau, \underline{\Sigma}), \quad (3.5.3)$$

or
$$\underline{v} \sim N_m(\underline{\alpha} + \underline{\beta}\ln(\tau), \underline{\Sigma}). \quad (3.5.4)$$

The test for compositional invariance of a basis is then simply a test of the form $\underline{\beta} = \underline{0}$, for which we can appeal to the usual methods of multivariate analysis. (See for example Morrison(1976) chapter 5).

The presence of basis independence is indicated by $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$, where $\underline{\Sigma}$ has the following structure :-

$$\begin{aligned} \underline{\Sigma} &= \underline{\Sigma}_0 = \text{dg}(\omega_1, \dots, \omega_m) + \omega_{m+1} \underline{U}_m \\ &= \begin{bmatrix} \omega_1 + \omega_{m+1} & \omega_{m+1} & \omega_{m+1} & \dots & \omega_{m+1} \\ \omega_{m+1} & \omega_2 + \omega_{m+1} & \omega_{m+1} & \dots & \omega_{m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega_{m+1} & \omega_{m+1} & \omega_{m+1} & \dots & \omega_m + \omega_{m+1} \end{bmatrix}, \end{aligned} \quad (3.5.5)$$

$$\omega_i > 0 ; i=1, \dots, m+1.$$

It is worth considering how (3.5.5) arises. The result was originally derived by Aitchison(1981b).

For $\underline{w} \in \mathbb{P}^{m+1}$, and $\underline{u} = C(\underline{w}) \in \mathbb{S}^m$, let

$$\underline{v} = a_m(\underline{u}), \text{ and } \underline{x} = \ln(\underline{w}).$$

Also

$$\underline{\mu}_x = E[\underline{x}]_{(m+1) \times 1},$$

$$\underline{\mu}_v = E[\underline{v}]_{m \times 1},$$

$$\underline{\Omega} = \text{Var}(\underline{x})_{(m+1) \times (m+1)},$$

$$\underline{\Sigma} = \text{Var}(\underline{v})_{m \times m}.$$

$$(3.5.6)$$

Consider the matrix $\underline{Y}_{m \times (m+1)} = \begin{bmatrix} \underline{I}_m & \vdots & -\underline{e}_m \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}$, then it is easily seen

that $\underline{\mu}_V = \underline{Y}\underline{\mu}_X$, (3.5.7)

and $\underline{\Sigma} = \underline{Y}\underline{\Omega}\underline{Y}'$. (3.5.8)

Equations (3.5.7) and (3.5.8) represent a mapping of the mean and variance for \underline{x} onto those for \underline{v} . Now

$$\|\underline{w}\| \Leftrightarrow \|\ln \underline{w} (= \underline{x})\|,$$

$$\therefore \|\underline{w}\| \Leftrightarrow \|\underline{x}\|.$$

And consequently if we have basis independence, $\|\underline{w}\|$, then $\underline{\Omega}$ (the covariance matrix of \underline{x}) will be diagonal. Applying (3.5.8) to a diagonal matrix $\underline{\Omega}_0 = \text{dg}(\omega_1, \dots, \omega_{m+1})$, and because ω_i represents the variance of x_i $i=1, \dots, m+1$; $\omega_i > 0$, gives the required result (3.5.5).

Although $\underline{\Sigma}_0$ is necessary for BI, it is not sufficient. For a given $\underline{\mu}_V$ and $\underline{\Sigma}$, there is a class of $\underline{\mu}_X$ and $\underline{\Omega}$ from which they may be derived. From (3.5.7) and (3.5.8) it is clear that these are :-

$$\underline{\mu}_X = \begin{bmatrix} \underline{\mu}_V + \alpha \underline{e}_m \\ \alpha \end{bmatrix}, \tag{3.5.9}$$

$$\underline{\Omega} = \begin{bmatrix} \underline{\Sigma} + \underline{e}_m \underline{\beta}' + \underline{\beta} \underline{e}_m' + \gamma \underline{U}_m^T \underline{\beta} + \gamma \underline{e}_m & & \\ \vdots & \ddots & \vdots \\ \underline{\beta}' + \gamma \underline{e}_m' & & \vdots \\ & & \vdots & \gamma \end{bmatrix} \tag{3.5.10}$$

where $\alpha, \gamma, \underline{\beta} = (\beta_1, \dots, \beta_m)'$ are constants.

Hence substituting $\underline{\Sigma}_0$ into (3.5.10) we see that $\underline{\Sigma}_0$ may have originated from any $\underline{\Omega}$ of the form :-

$$\underline{\Omega}_1 = \begin{bmatrix} \omega_1 + 2\beta_1 + \omega_{m+1} + \gamma & \beta_1 + \beta_2 + \omega_{m+1} + \gamma & \dots & \beta_1 + \beta_m + \omega_{m+1} + \gamma & \beta_1 + \gamma \\ \beta_1 + \beta_2 + \omega_{m+1} + \gamma & \omega_2 + 2\beta_2 + \omega_{m+1} + \gamma & \dots & \beta_2 + \beta_m + \omega_{m+1} + \gamma & \beta_2 + \gamma \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_1 + \beta_m + \omega_{m+1} + \gamma & \beta_2 + \beta_m + \omega_{m+1} + \gamma & \dots & \omega_m + 2\beta_m + \omega_{m+1} + \gamma & \beta_m + \gamma \\ \beta_1 + \gamma & \beta_2 + \gamma & \dots & \beta_m + \gamma & \gamma \end{bmatrix} \tag{3.5.11}$$

Setting $\beta_1 = \beta_2 = \dots = \beta_m = -\gamma = -\omega_{m+1}$ yields $\underline{\Omega}_0$. We note that (3.5.11) is not the pattern presented by Aitchison(1981b), which is incorrect.

Aitchison(1981b) tests for BI by testing that $\underline{\Sigma} = \underline{\Sigma}_0$, but as we have seen this is strictly a test that $\underline{\Omega} = \underline{\Omega}_1$ for any values of $\underline{\beta}$ and γ . However, without knowledge of the basis, this is the best we can do.

Having derived $\underline{\Sigma}_0$, it is now possible to test for basis independence by testing that $\underline{\Sigma} = \underline{\Sigma}_0$. That is :-

$$H_0: \underline{\Sigma} = \underline{\Sigma}_0 \quad \text{vs.} \quad H_a: \underline{\Sigma} \text{ unrestricted.}$$

It is possible to estimate $\underline{\Sigma}_0$ using Newton-Raphson iterative methods (see appendix of Aitchison(1981b)). The unrestricted $\underline{\Sigma}$ is estimated by the usual $\hat{\underline{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\underline{v}_i - \hat{\underline{\mu}})'(\underline{v}_i - \hat{\underline{\mu}})$. The familiar procedure is to form the likelihood ratio test-statistic $\Lambda_0(D)$ (say), where D represents a given data set $D = \{\underline{v}_1, \dots, \underline{v}_n\}$. The likelihood ratio test-statistic is then usually compared to a χ^2 , with degrees of freedom equal to the number of restrictions. However because of the inequalities present in $\underline{\Sigma}_0$ the standard results for the distribution do not apply. However H_0 is embedded between two other hypothesis.

$$H_1: \underline{\Sigma} = \underline{\Sigma}_1 = \text{dg}(\omega_1, \dots, \omega_m) + \omega_{m+1} \underline{U}_m, \quad \omega_{m+1} \text{ unrestricted; (3.5.12)}$$

$$H_2: \underline{\Sigma} = \underline{\Sigma}_2 = \text{dg}(\omega_1, \dots, \omega_m). \quad (3.5.13)$$

Let $\Lambda_1(D)$ and $\Lambda_2(D)$ be the likelihood ratio test-statistics corresponding to H_1 and H_2 respectively, then

$$\Lambda_1(D) \underset{\sim}{\sim} \chi_{\frac{1}{2}m(m-1)-1}^2, \quad (3.5.14)$$

$$\Lambda_2(D) \underset{\sim}{\sim} \chi_{\frac{1}{2}m(m-1)}^2, \quad (3.5.15)$$

$$\text{and} \quad \Lambda_1(D) \leq \Lambda_0(D) \leq \Lambda_2(D) \quad (3.5.16)$$

Consequently a simple and "safe" approach is to compare $\Lambda_0(D)$ to a $\chi_{\frac{1}{2}m(m-1)}^2$ distribution. In summary we reject the null hypothesis that a composition possesses basis independence at a significance level of at most α , when

$$\Lambda_0(D) > \chi_{2m(m-1); \alpha}^2 \quad (3.5.17)$$

Having explored extrinsic analysis, we next consider intrinsic analysis. The first type of intrinsic compositional independence is the intrinsic counterpart to the extrinsic concept of basis independence, and is given below.

Definition 3.25

Let $\underline{u}^{(m+1)} \in \mathbb{S}^m$ be partitioned :- $\{\underline{u}^{(1)}, \underline{u}^{(2)}, \dots, \underline{u}^{(k)}\}$. If for every such partition the subcompositions are independent i.e. if

$$C(\underline{u}^{(1)}) \parallel C(\underline{u}^{(2)}) \parallel \dots \parallel C(\underline{u}^{(k)}), \quad (3.5.18)$$

for any $\underline{u}^{(1)}, \dots, \underline{u}^{(k)}$ such that

$$\bigcup_{i=1}^k \underline{u}^{(i)} = \underline{u}^{(m+1)}, \text{ and } \underline{u}^{(i)} \cap \underline{u}^{(j)} = 0 \quad ; i, j = 1, \dots, k; i \neq j; \quad (3.5.19)$$

$$k = 1, 2, \dots, \left\lceil \frac{m+1}{1} \right\rceil,$$

then \underline{u} is said to have complete subcompositional independence; (CSI).

To understand CSI and its relationship to BI we may derive a parametric hypothesis for its occurrence. If $\underline{u} \in \mathbb{S}^m$ possesses CSI, then it follows that :-

$$\frac{u_i}{u_j} \parallel \frac{u_k}{u_l};$$

$$\Rightarrow \ln \frac{u_i}{u_j} \parallel \ln \frac{u_k}{u_l} \quad ; i \neq j \neq k \neq l.$$

$$\text{Cov} \left[\ln \frac{u_i}{u_j}, \ln \frac{u_k}{u_l} \right] = 0, \quad i \neq j \neq k \neq l,$$

$$\text{Cov} \left[\ln \frac{u_i}{u_k}, \ln \frac{u_j}{u_k} \right] = \text{Cov} \left[\ln \frac{u_i}{u_l} + \ln \frac{u_l}{u_k}, \ln \frac{u_j}{u_n} + \ln \frac{u_n}{u_k} \right]$$

$$= \text{Cov} \left[\ln \frac{u_\ell}{u_k}, \ln \frac{u_n}{u_k} \right], \quad i \neq j \neq k \neq \ell \neq n;$$

= λ_k (say), since it is independent of i, j, ℓ and n .

If we similarly let

$$\lambda_{ij} = \text{Var} \left[\ln \frac{u_i}{u_j} \right], \quad i \neq j;$$

then
$$\lambda_{ij} = \text{Var} \left[\ln \frac{u_i}{u_k} + \ln \frac{u_k}{u_j} \right], \quad i \neq j \neq k;$$

$$= \lambda_{ik} + \lambda_{jk} - 2\lambda_k$$

$$\Rightarrow 2\lambda_k = \lambda_{ik} + \lambda_{jk} - \lambda_{ij}.$$

Symmetrically it also follows that

$$2\lambda_i = \lambda_{ik} + \lambda_{ij} - \lambda_{jk}, \quad i \neq j \neq k = 1, \dots, m+1.$$

Adding these last two equations yields

$\lambda_{ik} = \lambda_i + \lambda_k$, so we have shown that if \underline{u} has CSI then

$$\left. \begin{aligned} \text{Var} \left[\ln \frac{u_i}{u_k} \right] &= \lambda_i + \lambda_k, \\ \text{Cov} \left[\ln \frac{u_i}{u_k}, \ln \frac{u_j}{u_k} \right] &= \lambda_k, \\ \text{Cov} \left[\ln \frac{u_i}{u_k}, \ln \frac{u_j}{u_\ell} \right] &= 0 \quad \text{for } \forall i \neq j \neq k \neq \ell = 1, \dots, m; \end{aligned} \right\} \quad (3.5.20)$$

and for some constants $\lambda_i \quad i=1, \dots, m+1$.

Consequently for $\underline{v} = \underline{a}(\underline{u})$, where \underline{u} has CSI, $\underline{\Sigma} = \text{Var}(\underline{v})$ has the form $\underline{\Sigma}_a$,

where
$$\underline{\Sigma}_a = \text{dg}(\lambda_1, \dots, \lambda_m) + \lambda_{m+1} \underline{U}_m, \quad (3.5.21)$$

and the $\lambda_i \quad (i=1, \dots, m+1)$ have the interpretation given by equations

(3.5.20). There are no restrictions on the λ_i 's except that they ensure the non-negative definiteness of $\underline{\Sigma}_a$.

The similarity between $\underline{\Sigma}_a$ given by (3.5.21) and $\underline{\Sigma}_1$ (3.5.5) is most striking. In fact they would be identical except for the fact that the λ_i 's in $\underline{\Sigma}_a$ need not be strictly positive. This fact allows us to utilize the previous discussion on BI, and comparing (3.5.21) with (3.5.12) we may test the hypothesis of CSI by forming the test-statistic $\Lambda_1(D)$ and comparing it with values from a $\chi^2_{\frac{1}{2}m(m-1)-1}$ distribution. We must also assume that \underline{y} follows the Normal distribution in order for the covariance relationships to imply independence. This in turn results in \underline{u} following the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution, as in BI.

Before considering other forms of compositional independence we note that for $\underline{u} \in \mathbb{S}^1$ and $\underline{u} \in \mathbb{S}^2$, complete subcompositional invariance automatically holds.

The next group of definitions take their motivation from the common need to examine only a few components of the composition. One such concept resulting from such problems was introduced in section 3.2, and that was the concept of neutrality (Definition 3.15). Here it is redefined in a slightly modified form, together with some related concepts.

For the remainder of this section we will consider the partition:-

$$(\underline{u}^{(c)}, \underline{u}_{(c)}) = \underline{u}^{(m+1)} \in \mathbb{S}^m,$$

where $\underline{u}^{(c)} = (u_1, u_2, \dots, u_c),$

and $\underline{u}_{(c)} = (u_{c+1}, \dots, u_{m+1}).$

Definition 3.26

A subcomposition $C(\underline{u}^{(c)})$ has subcompositional invariance if

$$C(\underline{u}^{(c)}) \parallel T(\underline{u}^{(c)}) \quad - (SI_1)$$

similarly $C(\underline{u}_{(c)})$ may also be subcompositionally invariant :-

$$C(\underline{u}_{(c)}) \parallel T(\underline{u}_{(c)}) \quad - (SI_2).$$

Definition 3.27

A subcomposition $C(\underline{u}^{(c)})$ possesses conditional subcompositional invariance if

$$C(\underline{u}^{(c)}) \parallel C(\underline{u}_{(c)}) | T(\underline{u}^{(c)}) \quad - \quad (\text{CoSI})$$

Definition 3.28

Left neutrality occurs if

$$C(\underline{u}^{(c)}) \parallel \underline{u}_{(c)} \quad - \quad (N_1).$$

Right neutrality occurs if

$$C(\underline{u}_{(c)}) \parallel \underline{u}^{(c)} \quad - \quad (N_2).$$

Definition 3.29

A composition has partition independence if

$$\parallel C(\underline{u}^{(c)}), C(\underline{u}_{(c)}), T(\underline{u}_{(c)}) \quad - \quad (P).$$

We note from comparing definitions 3.26 through to 3.29 that

$$\begin{aligned} \text{i)} \quad N_1 &\equiv \text{CoSI} \cap \text{SI}_1, \\ N_2 &\equiv \text{CoSI} \cap \text{SI}_2. \end{aligned} \quad (3.5.22)$$

$$\begin{aligned} \text{ii)} \quad P &\equiv N_1 \cap N_2 \\ &\equiv N_1 \cap \text{SI}_2 \equiv N_2 \cap \text{SI}_1. \end{aligned} \quad (3.5.23)$$

It should be apparent that SI is a subcompositional counterpart of CIB (definition 3.23), since $\underline{u}^{(c)}$ is acting as a basis for $C(\underline{u}^{(c)}) \in \mathbb{S}^{c-1}$. Since $T(\underline{u}^{(c)}) + T(\underline{u}_{(c)}) = 1$, SI implies that not only is the subcomposition independent of $T(\underline{u}^{(c)})$, but also of $T(\underline{u}_{(c)})$. Thus if $C(\underline{u}^{(c)})$ is in any way dependent on $\underline{u}_{(c)}$ it must be a relationship via the proportional breakdown of $\underline{u}_{(c)}$ (i.e. $C(\underline{u}_{(c)})$), and not through the share of the original composition attributed to $\underline{u}^{(c)}$ (i.e. $T(\underline{u}^{(c)})$). The concept of CoSI on the other hand, suggests the opposite, namely that if $\underline{u}_{(c)}$ is not independent of $\underline{u}^{(c)}$, then it is only through its share of $\underline{u}_{(c)}$, $T(\underline{u}_{(c)})$. When $C(\underline{u}^{(c)})$ is completely independent of $\underline{u}_{(c)}$ then both of the above situations hold, and we have neutrality. This yields

(3.5.22). Neutrality as defined here is identical to the earlier definition 3.14 except that now we allow it from both ends of the ordered vector \underline{u} . This allows us to relate it easily to partition independence.

The above forms of independence may be tested as suggested by Aitchison(1982), by modelling the data so as to mimic the independence property as an easily testable parametric restriction. In this instance let

$$\left. \begin{aligned} \underline{v}_1 &= a_{c-1} \left\{ C(\underline{u}^{(c)}) \right\}, \\ \underline{v}_2 &= a_{m-c} \left\{ C(\underline{u}_{(c)}) \right\}, \\ \text{and } \underline{v}_3 &= a_1 \left\{ \left[T(\underline{u}^{(c)}), T(\underline{u}_{(c)}) \right]' \right\}. \end{aligned} \right\} \quad (3.5.24)$$

Next consider

$$\begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix} \sim N_{(m-1)} \left[\begin{pmatrix} \underline{\alpha}_1 + \underline{E}_1 \underline{v}_3 \\ \underline{\alpha}_2 + \underline{E}_2 \underline{v}_3 \end{pmatrix}, \begin{pmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{pmatrix} \right], \quad (3.5.25)$$

for which the independence concepts are equivalent to various parametric hypotheses as follows :-

$$\begin{aligned} SI_1 &\Leftrightarrow \underline{E}_1 = \underline{0}; \\ SI_2 &\Leftrightarrow \underline{E}_2 = \underline{0}; \\ CoSI &\Leftrightarrow \underline{\Sigma}_{12} = \underline{0}; \\ N_1 &\Leftrightarrow \underline{E}_1 = \underline{0}, \underline{\Sigma}_{12} = \underline{0}; \\ N_2 &\Leftrightarrow \underline{E}_2 = \underline{0}, \underline{\Sigma}_{12} = \underline{0}; \\ P &\Leftrightarrow \underline{E}_1 = \underline{0}, \underline{E}_2 = \underline{0}, \underline{\Sigma}_{12} = \underline{0}. \end{aligned} \quad (3.5.26)$$

We may then appeal to the usual multivariate techniques, as for example in Morrison(1976).

For all the forms of independence 3.26 through to 3.29 we may be interested in considering different values of c ; working through the vector \underline{u} for some specific order of the variables. For example, consider again the opinion poll data with variables :-

"CON", "LAB", "LIB", "OTHER", "DONT KNOW".

First, we may wish to see if the four variables may be considered together, without the DONT KNOWs, because really we are interested in political parties, and the votes they are likely to gain. We might then wish to drop OTHER, since it may be thought that such candidates are not likely to be elected. Continuing our "speculative licence" we may finally wish to drop LIB, since our interest is finally concerned with the two major parties, and which one might form the next government.

To this end we have the following further definition :-

Definition 3.30

For some specific ordered $\underline{u} \in \mathbb{S}^m$, \underline{u} has independence property (SI_1 , SI_2 , CoSI, etc.) of order k if the stated independence property holds for $c=1, \dots, k$. Further if $k = m$, then \underline{u} is said to have complete independence of the form stated.

For example, SI_1 of order k implies that

$$C(\underline{u}^{(c)}) \parallel T(\underline{u}^{(c)}), \text{ for } c=1, \dots, k;$$

and complete right neutrality is equivalent to complete neutrality as defined by iii) of definition 3.14.

SI_1 up to order k is conjectured by Aitchison(1982) to be equivalent to neutrality up to order k within the framework of transformed normal modelling. The distinction between CoSI of order k with that of neutrality does, however, seem to exist, but the concept CoSI of order k does not appear to have any useful applications. Consequently the only independence property examined in this context is neutrality. This does have a more readily defined application, as for example in the opinion poll data cited above.

To test for independence up to some order k becomes cumbersome if we use the method above via the equations (3.5.24) and (3.5.25). It would be necessary to compute $\underline{v}_1, \underline{v}_2$ and \underline{v}_3 for each of $c=1, \dots, k$; and test for independence at each stage. This is not only laborious, but also leads to a multi-stage test, with the usual problems of defining the critical region. However Aitchison(1981a) shows that right neutrality is equivalent to the independence of the components :-

$$\frac{u_1}{1-u_1}, \frac{u_2}{1-u_1-u_2}, \dots, \frac{u_m}{1-u_1-\dots-u_m}$$

(cf. (3.2.18)) Taking logs yields the m_m transformation. This leads to the result that if we let $\underline{u} \sim M_m(\underline{\mu}, \underline{\Sigma})$, then neutrality for a partition with $c=k$, neutrality up to order k , and complete neutrality, are equivalent to the following parametric forms of $\underline{\Sigma}$ respectively :-

$$\begin{bmatrix} \underline{\Sigma}_{11} & \underline{0} \\ \underline{0} & \underline{\Sigma}_{22} \end{bmatrix}, \begin{bmatrix} \text{dg}(\sigma_{11}, \dots, \sigma_{kk}) & \underline{0} \\ \underline{0} & \underline{\Sigma}_{22} \end{bmatrix} \text{ and } \text{dg}(\sigma_{11}, \dots, \sigma_{mm})$$

Where $\underline{\Sigma}_{11}$ is $k \times k$ and $\underline{\Sigma}_{22}$ is $(m-k) \times (m-k)$. Left neutrality may be similarly examined by reversing the order of \underline{u} .

We now consider one final form of independence.

Definition 3.31

$\underline{u} = (\underline{u}^{(c)}, \underline{u}_{(c)}) \in \mathbb{S}^m$ has partial independence restricted by $\underline{u}^{(c)}$ (PSI^(c)) if :-

i) $C(\underline{u}^{(c)}) \perp\!\!\!\perp C(\underline{u}_{(c)})$,

and ii) $C(\underline{u}_{(c)}) \in \mathbb{S}^{m-c}$ has complete subcompositional independence.

It is again possible to form a parametric equivalent to PSI^(c) within the transformed normal-class. If we form \underline{v}_1 and \underline{v}_2 as in (3.5.24) PSI^(c) is equivalent to :-

$$\underline{\Sigma}_{12} = \underline{0}, \text{ and } \underline{\Sigma}_{22} = \text{dg}(\lambda_{c+1}, \dots, \lambda_m) + \lambda_{m+1} \underline{U}_{m-c}, \quad (3.5.27)$$

in equation (3.5.25).

We have described a powerful set of compositional independence concepts, together with a means of expressing them parametrically via the $L_m(\underline{\mu}, \underline{\Sigma})$ and $M_m(\underline{\mu}, \underline{\Sigma})$ distributions. However the logistic-normal class of distributions is unable to model some of the stronger independence properties, such as complete neutrality of all possible permutations. As mentioned in section 3.4, this latter concept requires the Dirichlet class. A solution to this problem has been suggested by Aitchison(1985) via a more general distribution for data on the simplex, which contains within it both the $L_m(\underline{\mu}, \underline{\Sigma})$ and the Dirichlet distributions. The resulting distribution has only one additional parameter over and above that of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution, and certain parametric hypotheses reduce it to the above two distributions. This allows for inference to be made to distinguish between them. It should be noted, however, that this hybrid distribution is not algebraically exact, and numerical techniques are necessary to evaluate it. It is not intended to develop the applications of this additional distribution into the area of time series at this stage, although it does suggest a possible area for further work.

3.6 Summary

In this chapter we have discussed the nature of compositional data, and various problems encountered when trying to analyse it. These problems include: understanding independence, a need for a suitable distribution, and further problems related to compositional time series. We have discussed some possible solutions following the approach of Aitchison(1982). These transformations and concepts of independence will be extended to time series in subsequent chapters.

"... in a flash, in the twinkling of an eye ... we will be changed."

I Corinthians 15:52

CHAPTER 4

Transformed Multivariate Time Series Models

4.0 Introduction

In the previous chapter we introduced compositional time series and discussed some of their properties. We have seen the need to define some new types of models which will handle these time series. In this chapter we present some possible models that will overcome the difficulties discussed in chapter 3. These new models employ the transformations of section 3.4 which map the m -dimensional simplex onto the real space. Having defined these models we will investigate some of their properties and illustrate their use by means of an example.

4.1 Compositional Time Series Models

We first consider the additive logistic transformation. Let $\{\underline{u}_t\}$ be a compositional time series i.e.

$$\underline{u}_t \in \mathbb{S}^m, \quad t=0, \pm 1, \dots$$

and then let $\underline{v}_t = a_m(\underline{u}_t)$, $t=0, \pm 1, \dots$ (4.1.1)

where a_m is given by definition 3.17. We consider applying conventional time series modelling techniques to the $\{\underline{v}_t\}$ series which takes values in \mathbb{R}^m .

Definition 4.1

A compositional time series $\{\underline{u}_t\}$, $\underline{u}_t \in \mathbb{S}^m$ ($t=0, \pm 1, \dots$) is said to be a multivariate additive logistic autoregressive moving-average process of order p, q and dimension m : $ln_+ARMA_m(p, q)$, if $\{\underline{v}_t\}$ given by (4.1.1) is $ARMA_m(p, q)$.

We employ the a_m transformation because it has been well developed by Aitchison(1982,1983,1984,1985,1986) and by Aitchison and Shen- (1980,1984). Many of its useful properties and expedience in testing for various types of independence were discussed in chapter 3. The other transformation which was found to be useful in chapter 3 was the m_m transformation (definition 3.18). We examined how it proved to be useful when testing for neutrality, and consequently we will introduce a model based on this transformation, but we will not discuss its properties in as much detail.

$$\text{Let } \underline{v}_t = m_m(\underline{u}_t), t=0, \pm 1, \dots \quad (4.1.2)$$

Definition 4.2

A compositional time series $\{\underline{u}_t\}$, $\underline{u}_t \in \mathbb{S}^m$ ($t=0, \pm 1, \dots$) is a multivariate multiplicative logistic autoregressive moving-average process of order p, q and dimension m : $\ln_x \text{ARMA}_m(p, q)$, if \underline{v}_t given by (4.1.2) is $\text{ARMA}_m(p, q)$.

We will similarly define $\ln_+ \text{AR}_m(p)$, $\ln_+ \text{MA}_m(q)$, $\ln_x \text{AR}_m(p)$, and $\ln_x \text{MA}_m(q)$ to be $\ln_+ \text{ARMA}_m(p, 0)$, $\ln_+ \text{ARMA}_m(0, q)$, $\ln_x \text{ARMA}_m(p, 0)$, and $\ln_x \text{ARMA}_m(0, q)$ respectively.

We note from the above definitions that if $\{\underline{v}_t\}$ follows an $\text{ARMA}_m(p, q)$ process then the input series is also m -dimensional. Hence $\{\underline{u}_t\}$ will have an m -dimensional input series as we require.

It will be useful to obtain an expression for \underline{u}_t in terms of its own past values under the various models.

Let $\underline{\phi}_1, \dots, \underline{\phi}_p, \underline{\theta}_0, \dots, \underline{\theta}_q$ be the parameters of the $\text{ARMA}_m(p, q)$ process for \underline{v}_t given by (4.1.1), where $\underline{\theta}_0 = \underline{I}_m$;

$$\underline{a}_t = \exp(\underline{\varepsilon}_t), \quad (4.1.3)$$

$$E[\underline{v}_t] = \underline{v}, \quad (4.1.4)$$

so that $\underline{V}_t = \underline{v}_t - \underline{v}, \quad (4.1.5)$

and hence $\underline{V}_t = \underline{\theta}_0 \underline{\varepsilon}_t + \underline{\theta}_1 \underline{\varepsilon}_{t-1} + \dots + \underline{\theta}_q \underline{\varepsilon}_{t-q} - \underline{\phi}_1 \underline{V}_{t-1} - \dots - \underline{\phi}_p \underline{V}_{t-p} \quad (4.1.6)$

Also let

$$\eta_i = \exp \left\{ u_i + \sum_{j=1}^p \sum_{k=1}^m \phi_{ik}^{(j)} u_k \right\}, \quad (4.1.7)$$

and
$$\phi_{i,m+1}^{(k)} = - \sum_{j=1}^m \phi_{ij}^{(k)} ; \quad i=1, \dots, m; \quad k=1, \dots, p. \quad (4.1.8)$$

Where
$$\underline{\phi}_k = \{ \phi_{ij}^{(k)} \} ; \quad k=1, \dots, p.$$

Then we have :-

Lemma 4.3

If $\underline{u}_t \in \mathbb{S}^m$ follows an $\ln_+ \text{ARMA}_m(p,q)$ process then :-

$$u_{it} = \frac{\left[\prod_{j=1}^p \prod_{k=1}^{m+1} -\phi_{ik}^{(j)} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{ik}^{(j)}} \right] \eta_i}{1 + \sum_{r=1}^m \left\{ \left[\prod_{j=1}^p \prod_{k=1}^{m+1} -\phi_{rk}^{(j)} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{rk}^{(j)}} \right] \eta_r \right\}}, \quad i=1, \dots, m; \quad t=0, \pm 1, \dots \quad (4.1.9)$$

Proof

Recall from definition 3.17 and (4.1.1) that :-

$$\underline{u}_t = a_m^{-1}(\underline{v}_t),$$

and
$$u_{it} = \frac{\exp(v_{it})}{1 + \sum_{r=1}^m \exp(v_{rt})}, \quad i=1, \dots, m; \quad t=0, \pm 1, \dots \quad (4.1.10)$$

Now from (4.1.6)

$$\exp(v_{it}) = \left[\exp \left\{ \underline{v}_t + \underline{\varepsilon}_t + \frac{\theta_1}{1} \underline{\varepsilon}_{t-1} + \dots + \frac{\theta_q}{q} \underline{\varepsilon}_{t-q} - \frac{\phi_1}{1} (\underline{v}_{t-1} - \underline{v}) - \dots - \frac{\phi_p}{p} (\underline{v}_{t-p} - \underline{v}) \right\} \right]_i,$$

where $\left[\underline{\quad} \right]_i$ denotes the i^{th} element of the vector for $i=1, \dots, m$.

Hence

$$\exp(v_{it}) =$$

$$\exp\left\{-\sum_{j=1}^p \sum_{k=1}^m v_{k,t-j} \phi_{ik}^{(j)}\right\} \cdot \exp\left\{\sum_{j=0}^q \sum_{k=1}^m \epsilon_{k,t-j} \theta_{ik}^{(j)}\right\} \cdot \exp\left\{v_i + \sum_{j=1}^p \sum_{k=1}^m v_k \phi_{ik}^{(j)}\right\}$$

$$= \left[\prod_{j=1}^p \prod_{k=1}^m \exp(-v_{k,t-j} \phi_{ik}^{(j)}) \right] \left[\prod_{j=0}^q \prod_{k=1}^m \exp(\epsilon_{k,t-j} \theta_{ik}^{(j)}) \right] \eta_i,$$

from equation(4.1.7).

(4.1.11)

Putting $a_{k,t-j} = \exp(\epsilon_{k,t-j})$, and recalling from definition 3.17

that $v_{k,t-j} = \ln \frac{u_{k,t-j}}{u_{m+1,t-j}}$, we obtain :-

$$\exp(v_{it}) = \left[\prod_{j=1}^p \prod_{k=1}^m \left(\frac{u_{k,t-j}}{u_{m+1,t-j}} \right)^{-\phi_{ik}^{(j)}} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{ik}^{(j)}} \right] \eta_i$$

$$= \left[\prod_{j=1}^p \prod_{k=1}^{m+1} u_{k,t-j}^{-\phi_{ik}^{(j)}} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{ik}^{(j)}} \right] \eta_i, \quad \begin{matrix} i=1, \dots, m \\ t=0, \pm 1, \dots \end{matrix};$$

(4.1.12)

where $\phi_{i,m+1}^{(k)}$ is given by (4.1.8).

Substituting (4.1.12) into (4.1.10) gives the required result. \square

Corollary 4.4

If $\underline{u}_t \in \mathbb{S}^m$ follows an $\ln_+ AR_m(p)$ process then, using the same notation,

$$u_{it} = \frac{\left[\prod_{j=1}^p \prod_{k=1}^{m+1} u_{k,t-j}^{-\phi_{ik}^{(j)}} \right] a_{it} \eta_i}{1 + \sum_{r=1}^m \left\{ \left[\prod_{j=1}^p \prod_{k=1}^{m+1} u_{k,t-j}^{-\phi_{rk}^{(j)}} \right] a_{rt} \eta_r \right\}}, \quad \begin{matrix} i=1, \dots, m \\ t=0, \pm 1, \dots \end{matrix}$$

$$= 1 - \sum_{j=1}^m u_{jt}, \quad i=m+1$$

Corollary 4.5

If $\underline{u}_t \in \mathbb{S}^m$ follows an $\ln_+ MA_m(q)$ process then,

$$u_{it} = \left. \begin{aligned} & \frac{\left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{ik}^{(j)}} \right] \eta_i}{1 + \sum_{r=1}^m \left[\left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{rk}^{(j)}} \right] \eta_r \right]} \\ & = 1 - \sum_{j=1}^m u_{jt} \end{aligned} \right\} \begin{array}{l} , i=1, \dots, m \\ t=0, \pm 1, \dots \\ , i=m+1 \end{array}$$

where the notation is as in lemma 4.3, except that $\eta_i = \exp(v_i)$.

Corollary 4.6

If $\underline{u}_t \in \mathbb{S}^m$ follows an $\ln_x ARMA_m(p,q)$ process, then :-

$$u_{it} = \frac{\left[\prod_{j=1}^p \prod_{k=1}^m \left[\frac{u_{k,t-j}}{\sum_{r=k+1}^{m+1} u_{r,t-j}} \right]^{-\phi_{ik}^{(j)}} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{ik}^{(j)}} \right] \eta_i}{\prod_{s=1}^i \left[1 + \left[\prod_{j=1}^p \prod_{k=1}^m \left[\frac{u_{k,t-j}}{\sum_{r=k+1}^{m+1} u_{r,t-j}} \right]^{-\phi_{sk}^{(j)}} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{sk}^{(j)}} \right] \eta_s \right]},$$

$i=1, \dots, m ; t=0, \pm 1, \dots$

where η_i is as before, but noting that \underline{u}_k is now the mean of the \underline{v}_t series given by (4.1.2).

Proof

This follows the exact form of the proof for the lemma, except that after equation (4.1.11) we substitute,

$$v_{k,t-j} = \ln \frac{u_{k,t-j}}{1 - \sum_{r=1}^m u_{r,t-j}} = \ln \frac{u_{k,t-j}}{\sum_{r=k+1}^{m+1} u_{r,t-j}},$$

since $\sum_{i=1}^{m+1} u_{i,t-j} = 1$,

and then the expression for $\exp(v_{it})$ is

$$\exp(v_{it}) = \left[\prod_{j=1}^p \prod_{k=1}^m \left(\frac{u_{k,t-j}}{\sum_{r=k+1}^{m+1} u_{r,t-j}} \right)^{-\phi_{ik}^{(j)}} \right] \left[\prod_{j=0}^q \prod_{k=1}^m a_{k,t-j}^{\theta_{ik}^{(j)}} \right] \eta_i \quad (4.1.13)$$

From definition 3.18 u_{it} is then given by :-

$$u_{it} = \frac{\exp(v_{it})}{\prod_{j=1}^m \{1 + \exp(v_{jt})\}} \quad , \quad t=0, \pm 1, \dots \quad (4.1.14)$$

Substituting (4.1.13) into (4.1.14) then gives the required result. \square

We may obtain similar expressions for the $\ln_x \text{AR}_m(p)$ and $\ln_x \text{MA}_m(q)$ processes. However for the remainder of this chapter we will concentrate on the $\ln_+ \text{ARMA}_m(p,q)$ model. Many of the results for the $\ln_+ \text{ARMA}_m(p,q)$ model will have similar and obvious counter-parts for the $\ln_x \text{ARMA}_m(p,q)$ model, whilst other results may have less obvious or possibly no counter-part. Where possible we will try to indicate which is the case. What follows in the rest of this section can be easily formulated for the $\ln_x \text{ARMA}_m(p,q)$ model.

We now proceed with some further definitions.

Definition 4.7

An $\ln_+ \text{ARMA}_m(p,q)$ process \underline{u}_t is MA-invertible if it can be expressed as an $\ln_+ \text{MA}_m(\infty)$ process.

Clearly if a process \underline{u}_t is MA-invertible then the process \underline{v}_t formed by taking the a_m transformation is stationary, and hence a necessary and sufficient condition for \underline{u}_t to be MA-invertible is given by lemma 2.6.

Definition 4.8

An $\ln_+ARMA_m(p,q)$ process \underline{u}_t is AR-invertible if it can be expressed as an $\ln_+AR_m(\infty)$ process.

In a similar way the above definition is equivalent to the model for the $\{\underline{v}_t\}$ series being invertible, and hence the required conditions for this are given by lemma 2.7. We note that we could easily have interchanged the names in the above definitions, so that they refer to the polynomial equation of the parent ARMA model that is invertible; instead of which we have chosen to define the names so that they refer to the resulting form of the infinite polynomial. Also, we have not used the terms stationary and invertible as it is the $\{\underline{v}_t\}$ series and not the $\{\underline{u}_t\}$ series to which these terms refer.

Clearly if a process is MA-invertible we may express it as :-

$$u_{it} = \frac{\left[\prod_{j=0}^{\infty} \prod_{k=1}^m a_{k,t-j}^{(j)} \right] \eta_i}{1 + \sum_{r=1}^{m+1} \left\{ \left[\prod_{j=0}^{\infty} \prod_{k=1}^m a_{k,t-j}^{(j)} \right] \eta_r \right\}}, \quad \begin{array}{l} t=0, \pm 1, \dots \\ i=1, \dots, m \end{array} \quad (4.1.15)$$

where $\eta_i = \exp(v_i)$.

Similarly an AR-invertible process may be expressed as :-

$$u_{it} = \frac{\left[\prod_{j=1}^{\infty} \prod_{k=1}^{m+1} u_{k,t-j}^{-\pi_{ik}^{(j)}} \right] a_{it} \eta_i}{1 + \sum_{r=1}^{m+1} \left\{ \left[\prod_{j=1}^{\infty} \prod_{k=1}^{m+1} u_{k,t-j}^{-\pi_{rk}^{(j)}} \right] a_{rt} \eta_r \right\}}, \quad \begin{array}{l} t=0, \pm 1, \dots \\ i=1, \dots, m \end{array} \quad (4.1.16)$$

where $\pi_{i,m+1}^{(k)} = -\sum_{j=1}^m \pi_{ij}^{(k)}$,

and $\eta_i = \exp\left\{v_i + \sum_{j=1}^{\infty} \sum_{k=1}^m \pi_{ik}^{(j)} v_k\right\}$.

In (4.1.15) and (4.1.16) the ψ 's and π 's refer to the infinite polynomial parameters for $\{v_t\}$ expressed as an $MA_m(\infty)$ and $AR_m(\infty)$ process respectively.

One class of models we have not considered are those based on the $ARIMA_m(p,d,q)$ process. This is partly due to our comments in chapter 2 regarding such models, and also because the $ARIMA_m(p,d,q)$ may be thought of as a non-stationary $ARMA_m(p+d,q)$ process, for which the above models will suffice.

4.2 A Linear Approximation to the Model

So far we have suggested an approach to modelling a $\{u_t\}$ series, where $u_t \in \mathbb{S}^m$, via a transformed model such as the $\ln_+ARMA_m(p,q)$. However, in the survey context e.g. Scott et al (1977), such series have been analyzed by using the traditional linear $ARMA(p,q)$ model, although in general this has been restricted to univariate models. One may then wish to ask if, and under what conditions, such a model is appropriate.

To compare the two approaches consider the a_1 transformation of $u \in \mathbb{S}^1$ to $v \in \mathbb{R}^1$:-

$$v = \ln \frac{u}{1-u} .$$

This function is plotted in figure 4.9.

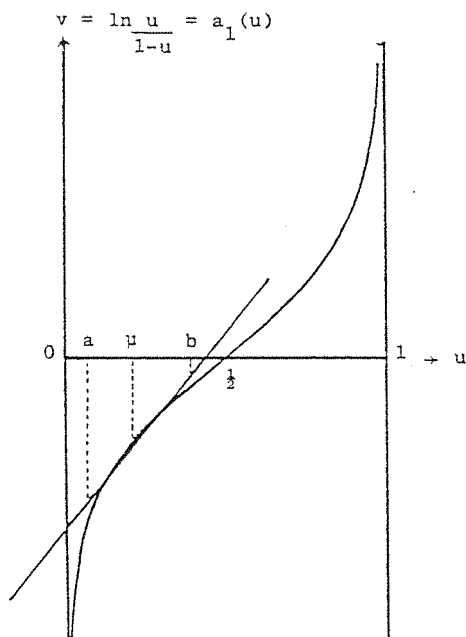


Figure 4.9

When u is "close" to some point μ (say), then $\ln \frac{u}{1-u}$ is approximately linear. By "close" we mean that $a \leq u \leq b$, where $a < \mu < b$, and where the choice of a and b depends on the degree of approximation required and the value of μ . The closer μ is to 0.5 the wider the range $[a,b]$ can be because the function is more "flat" around this point. Clearly we may extend such an approach to higher dimensions.

It is usual to examine a linear approximation about the point $\underline{\mu} = E[\underline{u}]$. We now form an approximation to the multiplicative $\ln_+ \text{ARMA}_m(p,q)$ model using a Taylor series expansion. It is hoped that this will provide a useful means of understanding the properties of the $\ln_+ \text{ARMA}_m(p,q)$ system. Consider (4.1.9) and let

$$i) \quad E[\underline{u}_t^{(m+1)}] = \underline{\mu}^{(m+1)}, \quad (4.2.1)$$

$$E[\underline{a}_t] = \underline{\alpha}, \quad (4.2.2)$$

$$ii) \quad \lambda_k = \frac{\ell_k}{T}, \quad (4.2.3)$$

$$\ell_k = \left[\begin{array}{ccc} p & m+1 & -\phi_{ki}^{(j)} \\ \prod_{j=1} & \prod_{i=1} & \mu_i \end{array} \right] \left[\begin{array}{cc} q & m \\ \prod_{j=0} & \prod_{i=1} \alpha_i \end{array} \right] \theta_{ki}^{(j)} \eta_k; \quad k=1, \dots, m. \quad (4.2.4)$$

$$T = 1 + \sum_{i=1}^m \ell_i, \quad (4.2.5)$$

$$\underline{\Lambda} = (\lambda_1, \dots, \lambda_m)'$$

$$iii) \quad \beta_{kr}^{(s)} = \frac{\lambda_k}{\mu_r} \left[-\phi_{kr}^{(s)} + \sum_{j=1}^m \phi_{jr}^{(s)} \lambda_j \right], \quad (4.2.6)$$

$$h_{kr}^{(s)} = \beta_{kr}^{(s)} - \beta_{k,m+1}^{(s)}, \quad (4.2.7)$$

$$\underline{H}_s = \{h_{kr}^{(s)}\}; \quad k,r=1, \dots, m; \quad s=1, \dots, p.$$

$$iv) \quad g_{kr}^{(s)} = \frac{\lambda_k}{\alpha_r} \left[\theta_{kr}^{(s)} - \sum_{j=1}^m \theta_{jr}^{(s)} \lambda_j \right], \quad (4.2.8)$$

$$\underline{G}_s = \{g_{kr}^{(s)}\}; \quad k,r=1, \dots, m; \quad s=0, \dots, q.$$

Theorem 4.10

If $\underline{u}_t \in \mathbb{S}^m$ follows a $\ln_r \text{ARMA}_m(p,q)$ process then a linear approximation to the model (4.1.9) is

$$\underline{u}_t = \underline{\mu} + \underline{H}_1(\underline{u}_{t-1} - \underline{\mu}) + \dots + \underline{H}_p(\underline{u}_{t-p} - \underline{\mu}) + \underline{G}_0(\underline{a}_t - \underline{\alpha}) + \dots + \underline{G}_q(\underline{a}_{t-q} - \underline{\alpha}) + \underline{L}_t,$$

where \underline{L}_t represents the second order and higher order terms of the Taylor expansion.

Proof

$$\text{Let } \underline{J}'_t = (\underline{u}'_{t-1}, \underline{u}'_{t-2}, \dots, \underline{u}'_{t-p}, \underline{a}'_t, \dots, \underline{a}'_{t-q}) \quad , \quad (4.2.9)$$

$$F_i(\underline{J}_t) = \begin{bmatrix} p & m+1 & -\phi_{ik}^{(j)} \\ \prod_{j=1} & \prod_{k=1} & u_{k,t-j} \end{bmatrix} \begin{bmatrix} q & m & \theta_{ik}^{(j)} \\ \prod_{j=0} & \prod_{k=1} & a_{k,t-j} \end{bmatrix} \eta_i \quad , \quad \begin{matrix} i=1, \dots, m \\ t=0, \pm 1, \dots \end{matrix} \quad (4.2.10)$$

$$f_k(\underline{J}_t) = \frac{F_k(\underline{J}_t)}{1 + \sum_{r=1}^m F_r(\underline{J}_t)} \quad ; \quad k=1, \dots, m \quad ; \quad t=0, \pm 1, \dots \quad (4.2.11)$$

Then from equation (4.1.9), the function $f_k(\underline{J}_t)$ is the right-hand side of the model formula. Hence we may rewrite (4.1.9) as

$$u_{it} = f_i(\underline{J}_t) \quad ; \quad i=1, \dots, m \quad ; \quad t=0, \pm 1, \dots \quad (4.2.12)$$

From (4.2.1), (4.2.2) and (4.2.9) we note that

$$E[\underline{J}'_t] = (\underline{\mu}', \underline{\mu}', \dots, \underline{\mu}', \underline{\alpha}', \dots, \underline{\alpha}') = \underline{N}' \quad (\text{say}) \quad (4.2.13)$$

The Taylor series expansion of $f_i(\underline{J}_t)$ about \underline{N} is given by

$$f_i(\underline{J}_t) = f_i(\underline{J}_t) \Big|_{\underline{J}_t = \underline{N}} + \sum_{k=1}^p \sum_{r=1}^{m+1} \left[\frac{\partial f_i(\underline{J}_t)}{\partial u_{r,t-k}} \Big|_{\underline{J}_t = \underline{N}} \right] (u_{r,t-k} - \mu_r) + \sum_{k=0}^q \sum_{r=1}^m \left[\frac{\partial f_i(\underline{J}_t)}{\partial a_{r,t-k}} \Big|_{\underline{J}_t = \underline{N}} \right] (a_{r,t-k} - \alpha_r) + R_{2ti} \quad , \quad (4.2.14)$$

where \underline{R}_t are the second and higher order terms. We may rewrite this in vector form as

$$\begin{aligned} \underline{f}(\underline{J}_t) &= \underline{f}(\underline{J}_t) \Big|_{\underline{J}_t = \underline{N}} + \sum_{i=1}^p \left[\frac{\partial \underline{f}(\underline{J}_t)}{\partial \underline{u}_{t-i}^{(m+1)}} \Big|_{\underline{J}_t = \underline{N}} \right] (\underline{u}_{t-i}^{(m+1)} - \underline{\mu}) + \\ &\quad \sum_{i=0}^q \left[\frac{\partial \underline{f}(\underline{J}_t)}{\partial \underline{a}_{t-i}} \Big|_{\underline{J}_t = \underline{N}} \right] (\underline{a}_{t-i} - \underline{\alpha}) + \underline{R}_{2t}, \end{aligned} \quad (4.2.15)$$

where $\underline{f}(\underline{J}_t) = (f_1(\underline{J}_t), \dots, f_m(\underline{J}_t))'$,

and $\underline{R}_{2t} = (R_{2t1}, R_{2t2}, \dots, R_{2tm})'$,

and terms of the form $\frac{\partial \underline{y}_{1 \times r}}{\partial \underline{x}_{1 \times s}}$ are the $r \times s$ matrix with elements z_{ij}

$$\text{and } z_{ij} = \frac{\partial y_i}{\partial x_j} \quad \begin{array}{l} i=1, \dots, r; \\ j=1, \dots, s. \end{array} \quad (4.2.16)$$

We will take each term of (4.2.14) in turn. First we note that

$$F_k(\underline{J}_t) \Big|_{\underline{J}_t = \underline{N}} = \lambda_k \text{ as in (4.2.4); } k=1, \dots, m. \quad (4.2.17)$$

and hence from (4.2.11)

$$f_k(\underline{J}_t) \Big|_{\underline{J}_t = \underline{N}} = \frac{\lambda_k}{1 + \sum_{i=1}^m \lambda_i} = \frac{\lambda_k}{T} = \lambda_k \text{ from (4.2.3)-(4.2.5).} \quad (4.2.18)$$

Secondly differentiating with respect to \underline{u}_t we find from (4.2.10)

$$\frac{\partial F_k(\underline{J}_t)}{\partial u_{r,t-s}} = -\phi_{kr}^{(s)} u_{r,t-s}^{-1} F_k(\underline{J}_t), \quad \begin{array}{l} k=1, \dots, m; \\ r=1, \dots, m+1; \end{array} \quad (4.2.19)$$

and thus,

$$\frac{\partial F_k(\underline{J}_t)}{\partial u_{r,t-s}} \Big|_{\underline{J}_t = \underline{N}} = \frac{-\phi_{kr}^{(s)}}{\mu_r} \lambda_k \text{ from (4.2.13) and (4.2.17).} \quad (4.2.20)$$

From (4.2.11)

$$\frac{\partial f_k(\underline{J}_t)}{\partial u_{r,t-s}} = \frac{\partial F_k(\underline{J}_t)}{\partial u_{r,t-s}} \Big/ \left[1 + \sum_{\zeta=1}^m F_{\zeta}(\underline{J}_t) \right] - \left[\frac{F_k(\underline{J}_t)}{\sum_{\zeta=1}^m \frac{\partial F_{\zeta}(\underline{J}_t)}{\partial u_{r,t-s}}} \right] \Big/ \left[1 + \sum_{\zeta=1}^m F_{\zeta}(\underline{J}_t) \right]^2. \quad (4.2.21)$$

Hence from (4.2.20) and (4.2.21) we obtain

$$\begin{aligned} \frac{\partial f_k(\underline{J}_t)}{\partial u_{r,t-s}} \Big|_{\underline{J}_t = \underline{N}} &= \frac{-\phi_{kr}^{(s)} \ell_k}{\mu_r \left[1 + \sum_{i=1}^m \ell_i \right]} + \frac{\ell_k \sum_{\zeta=1}^m \phi_{\zeta r}^{(s)} \ell_{\zeta}}{\mu_r \left[1 + \sum_{i=1}^m \ell_i \right]^2} \\ &= \frac{\lambda_k}{\mu_r} \left[-\phi_{kr}^{(s)} + \sum_{j=1}^m \phi_{jr}^{(s)} \lambda_j \right] \quad \text{from (4.2.3)-(4.2.6) ,} \\ &= \beta_{kr}^{(s)} \quad ; \quad k=1, \dots, m ; \quad r=1, \dots, m+1 ; \quad s=1, \dots, p. \end{aligned} \quad (4.2.22)$$

Similarly, since

$$\frac{\partial F_k(\underline{J}_t)}{\partial a_{r,t-s}} = \theta_{kr}^{(s)} a_{r,t-s}^{-1} F_k(\underline{J}_t) ,$$

whence

$$\frac{\partial F_k(\underline{J}_t)}{\partial a_{r,t-s}} \Big|_{\underline{J}_t = \underline{N}} = \frac{\theta_{kr}^{(s)}}{\alpha_r} \ell_k \quad ; \quad k, r=1, \dots, m ; \quad s=0, \dots, q ; \quad t=0, \pm 1, \dots \quad (4.2.23)$$

We have that

$$\begin{aligned} \frac{\partial f_k(\underline{J}_t)}{\partial a_{r,t-s}} \Big|_{\underline{J}_t = \underline{N}} &= \frac{\theta_{kr}^{(s)} \ell_k}{\alpha_r \left[1 + \sum_{i=1}^m \ell_i \right]} - \frac{\ell_k \sum_{\zeta=1}^m \theta_{\zeta r}^{(s)} \ell_{\zeta}}{\alpha_r \left[1 + \sum_{i=1}^m \ell_i \right]^2} \\ &= \frac{\lambda_k}{\alpha_r} \left[\theta_{kr}^{(s)} - \sum_{j=1}^m \theta_{jr}^{(s)} \lambda_j \right] \\ &= g_{kr}^{(s)} \quad ; \quad k, r=1, \dots, m ; \quad s=0, \dots, q, \text{ from (4.2.8).} \end{aligned} \quad (4.2.24)$$

Substituting (4.2.18), (4.2.22) and (4.2.24) into our expression for the Taylor series expansion (4.2.14) gives

$$f_k^{(j)}(t) = \lambda_k + \sum_{i=1}^p \sum_{r=1}^{m+1} \beta_{kr}^{(i)} (u_{r,t-i} - \mu_r) + \sum_{i=0}^q \sum_{r=1}^m g_{kr}^{(i)} (a_{r,t-i} - \alpha_r) + R_{2tk} \quad (4.2.25)$$

Finally we note that the summation in the second term of the above expression is over $m+1$ linearly dependent terms. We may re-parameterize this to be the sum of m linear independent terms thus :-

$$\begin{aligned} \sum_{r=1}^{m+1} \beta_{kr}^{(i)} (u_{r,t-i} - \mu_r) &= \\ & \sum_{r=1}^m \beta_{kr}^{(i)} (u_{r,t-i} - \mu_r) + \beta_{k,m+1}^{(i)} \left[1 - \sum_{j=1}^m u_{j,t-i} - 1 + \sum_{j=1}^m \mu_{j,t-i} \right] \\ &= \sum_{r=1}^m (\beta_{kr}^{(i)} - \beta_{k,m+1}^{(i)}) (u_{r,t-i} - \mu_r) \\ &= \sum_{r=1}^m h_{kr}^{(i)} (u_{r,t-i} - \mu_r) . \end{aligned} \quad (4.2.26)$$

Substituting (4.2.26) into (4.2.25) gives the required result, where

$$\underline{L}_t = \underline{R}_{2t} . \quad (4.2.27)$$

□

Corollary 4.11

$$\underline{H}_i = - \underline{E} \underline{\Phi}_i \underline{F} ; \quad i=1, \dots, p ; \quad (4.2.28)$$

$$\underline{G}_j = \underline{E} \underline{\Theta}_j \underline{A}^{-1} ; \quad j=0, \dots, q ; \quad (4.2.29)$$

where \underline{H}_i ($i=1, \dots, p$) and \underline{G}_j ($j=0, \dots, q$) are the parameters of the model in equation (4.1.9) and $\underline{\Phi}_i$ and $\underline{\Theta}_j$ are the parameters of the $\ln_+ \text{ARMA}_m(p, q)$.

$$\text{And} \quad \underline{E} = \underline{L} (\underline{I}_m - \underline{U}_m \underline{L}) , \quad (4.2.30)$$

$$\underline{F} = \underline{M}^{-1} + \frac{1}{\mu_{m+1}} \underline{U}_m , \quad (4.2.31)$$

$$\underline{A} = \text{dg}(\alpha_1, \dots, \alpha_m) , \quad (4.2.32)$$

$$\underline{M} = \text{dg}(\mu_1, \dots, \mu_m) , \quad (4.2.33)$$

$$\underline{L} = \text{dg}(\lambda_1, \dots, \lambda_m) . \quad (4.2.34)$$

Proof

From (4.2.6) and (4.2.7)

$$\begin{aligned}
 h_{kr}^{(s)} &= \beta_{kr}^{(s)} - \beta_{k,m+1}^{(s)} \\
 &= \frac{\lambda_k}{\mu_r} \left[-\phi_{kr}^{(s)} + \sum_{j=1}^m \phi_{jr}^{(s)} \lambda_j \right] - \frac{\lambda_k}{\mu_{m+1}} \left[-\phi_{k,m+1}^{(s)} + \sum_{j=1}^m \phi_{j,m+1}^{(s)} \lambda_j \right] \\
 &= \lambda_k \left[- \left[\frac{\phi_{kr}}{\mu_r} + \frac{\sum_{i=1}^m \phi_{ki}}{\mu_{m+1}} \right] + \sum_{j=1}^m \lambda_j \left[\frac{\phi_{jr}}{\mu_r} + \frac{\sum_{i=1}^m \phi_{ji}}{\mu_{m+1}} \right] \right] \quad ; k,r=1,\dots,m; \\
 & \quad s=1,\dots,p. \tag{4.2.35}
 \end{aligned}$$

+ - (i) - + + - (ii) - + + - (iii) - + + - (iv) - +

We may rewrite this in matrix form (term by term) :-

$$\begin{aligned}
 \underline{H}_s &= \underline{L} \left[- \underline{\Phi}_s \underline{M}^{-1} - \frac{1}{\mu_{m+1}} \underline{\Phi}_s \underline{U} + \frac{\underline{U} \underline{L}}{\mu_{m+1}} \left(\underline{\Phi}_s \underline{M}^{-1} + \frac{1}{\mu_{m+1}} \underline{\Phi}_s \underline{U} \right) \right] \\
 & \quad + - (i) - + + - (ii) - + + - (iii) - + + - (iv) - + \\
 &= \underline{L} (\underline{I}_m - \frac{\underline{U} \underline{L}}{\mu_{m+1}}) (-\underline{\Phi}_s) (\underline{M}^{-1} + \frac{1}{\mu_{m+1}} \underline{U}) \quad ; s=1,\dots,p; \tag{4.2.36} \\
 &= -\underline{E} \underline{\Phi}_s \underline{F} \text{ as required.}
 \end{aligned}$$

Similarly since

$$\begin{aligned}
 g_{kr}^{(s)} &= \frac{\lambda_k}{\alpha_r} \left[\theta_{kr}^{(s)} - \sum_{j=1}^m \theta_{jr}^{(s)} \lambda_j \right], \\
 \Rightarrow \underline{G}_s &= \underline{L} (\underline{\Theta}_s - \frac{\underline{U} \underline{L} \underline{\Theta}_s}{\mu_{m+1}}) \underline{A}^{-1}, \tag{4.2.37} \\
 &= \underline{L} (\underline{I}_m - \frac{\underline{U} \underline{L}}{\mu_{m+1}}) \underline{\Theta}_s \underline{A}^{-1} = \underline{E} \underline{\Theta}_s \underline{A}^{-1} \text{ as required, } s=0,\dots,q.
 \end{aligned}$$

□

Clearly the linear approximation given in the above theorem is in the form of an ARMA_m(p,q) model. However we note that $\underline{G}_0 \neq \underline{I}_m$ as in the usual specification, and we have additional terms \underline{L}_t and $\underline{\Lambda}$.

We may investigate conditions under which $\underline{L}_t \approx \underline{0}$. Recall that the innovation series for $\{\underline{v}_t\}$ is $\{\underline{\varepsilon}_t\}$, where $\text{Var}[\underline{\varepsilon}_t] = \underline{\Sigma}$. Then following the approach of Bickel and Doksum(1981) we would expect $\underline{L}_t \approx \underline{0}$ when $\|\underline{\Sigma}\|$

≈ 0 , where $\|\underline{\Sigma}\|$ is some norm of the covariance matrix, such as the determinant. This, for example might yield the criterion for a linear approximation as :-

$$\underline{L}_t \approx \underline{0} \quad , \quad (4.2.38)$$

when $|\underline{\Sigma}| < k$.

The choice of k will depend on how accurate we wish the linear approximation to be. Exact details may be obtained by examining higher order terms of the Taylor series expansion of the $\ln_+ \text{ARMA}_m(p,q)$ process. We leave this for further study, since we wish to develop the $\ln_+ \text{ARMA}_m(p,q)$ as such, and not to concentrate on its linear approximation. In general when k is small the values of the compositional time series will not vary that much. i.e. the values will not fluctuate from one extreme value to another.

The additional constant term $\underline{\Lambda}$ may also be sensibly dealt with. For the $\text{ARMA}_m(p,q)$ model in theorem 4.10 to be of standard form, we require that,

$$\underline{\Lambda} = \underline{\mu} \quad , \quad \text{i.e.} \quad \lambda_i = \mu_i \quad ; \quad i=1, \dots, m.$$

This may be seen in context if we note that λ_i is simply the value of $f_i(\underline{J}_t)$ (equations (4.2.13) and (4.2.18)) evaluated at the mean, and since from (4.2.12),

$$u_{it} = f_i(\underline{J}_t) \quad ; \quad i=1, \dots, m \quad ; \quad t=0, \pm 1, \dots \quad ;$$

then $\lambda_i = \mu_i$ is equivalent to,

$$\mu_i = f_i(\underline{N}) \quad ; \quad i=1, \dots, m$$

$$\text{i.e.} \quad E[u_{it}] = f_i(E[\underline{J}_t]) \quad . \quad (4.2.39)$$

Since (4.1.9) has been approximated by an $\text{ARMA}_m(p,q)$ process, then (4.2.39) is equivalent to that approximation applied to the expected values. i.e. when $\underline{L}_t \approx \underline{0}$ (4.2.39) may be obtained by taking expectations of the ARMA model:-

$$(\underline{u}_t - \underline{\mu}) = \underline{H}_1(\underline{u}_{t-1} - \underline{\mu}) + \dots + \underline{H}_p(\underline{u}_{t-p} - \underline{\mu}) + \underline{A}_t^* + \underline{G}_1^* \underline{A}_{t-1}^* + \dots + \underline{G}_q^* \underline{A}_{t-q}^* + \underline{L}_t^* ; \quad (4.2.40)$$

$$\text{where } \left. \begin{aligned} \underline{L}_t^* &= \underline{L}_t - (\underline{\mu} - \underline{\Lambda}) , \\ \underline{A}_t &= \underline{a}_t - \underline{\alpha} , \\ \underline{A}_t^* &= \underline{G}_0 \underline{A}_t , \\ \underline{G}_t^* &= \underline{G}_1 \underline{G}_0^{-1} ; \quad i=0, \dots, q . \end{aligned} \right\} \quad (4.2.41)$$

Hence (4.2.39) holds when the linear approximation is good.

The ARMA model (4.2.40) is a re-parameterized form of that given in theorem 4.10. Clearly if (4.2.38) and consequently also (4.2.39) hold, then $\underline{L}_t \approx \underline{0}$. If $\underline{L}_t \approx \underline{0}$ for whatever reason, then we have an ARMA_m(p,q) process in its standard form. Alternatively if we could assume that \underline{L}_t^* is an independent white-noise process then we also have an ARMA_m(p,q) process, since the sum of two independent moving-average processes MA_m(q₁) + MA_m(q₂) say, is itself MA_m(q₃), where q₃=max(q₁,q₂). (See for example Box and Jenkins(1976)).

From (4.1.3) we note that

$$a_{kt} = \exp\{\varepsilon_{kt}\} ; \quad k=1, \dots, m ; \quad t=0, \pm 1, \dots ;$$

$$\text{where } \underline{\varepsilon}_t \sim N_m(\underline{0}, \underline{\Sigma}).$$

Hence \underline{a}_t follows a multivariate log-normal distribution with mean

$$\underline{\alpha} : \alpha_k = E[a_{kt}] = \exp(\frac{1}{2}\sigma_{kk}) , \quad \left. \begin{aligned} &k=1, \dots, m ; \\ &t=0, \pm 1, \dots \end{aligned} \right\}$$

and variance $\underline{\Sigma}_A : \sigma_{ij}^{(A)} = \left[e^{\sigma_{ij} - 1} \right] \left[e^{\frac{1}{2}(\sigma_{ii} + \sigma_{jj})} \right] ; \quad i, j=1, \dots, m.$

$$(4.2.42)$$

If $\underline{L}_t^* \approx \underline{0}$ we have some interesting results as follows.

Lemma 4.12

The determinantal equations for the moving-average components of

- i) The parent ARMA_m(p,q) model for $\underline{v}_t = a_m(\underline{u}_t)$, $\underline{u}_t \in \mathbb{S}^m$.
 and ii) the linear approximation of the resulting $\ln_+ \text{ARMA}_m(p,q)$ model
 for \underline{u}_t , namely equation (4.2.40), with $\underline{L}_t^* = \underline{0}$

both have the same roots, and hence both models are invertible or otherwise.

Proof

The determinantal equation for the MA component of (4.2.40), $\underline{L}_t^* = \underline{0}$ is,

$$|\underline{G}_0^* z^q + \underline{G}_1^* z^{q-1} + \dots + \underline{G}_q^*| = 0 \quad (4.2.43)$$

We have from (4.2.41) that $\underline{G}_i^* = \underline{G}_i \underline{G}_0^{-1} \quad ; \quad i=0, \dots, q$;
 $= \underline{E} \underline{O}_i \underline{A}^{-1} (\underline{E} \underline{I}_m \underline{A}^{-1})^{-1}$, from (4.2.37),

and since $\underline{O}_0 = \underline{I}_m$ $= \underline{E} \underline{O}_i \underline{E}^{-1} \quad ; \quad i=0, \dots, q.$ (4.2.44)

\underline{E}^{-1} exists since from (4.2.30) $\underline{L}^{-1} = \text{dg}(1/\lambda_1, \dots, 1/\lambda_m)$,

and $|\underline{I}_m - \underline{U}_m \underline{L}| = 1 - \sum_{i=1}^m \lambda_i \neq 0$ (in general).

Substituting (4.2.44) into (4.2.43) gives,

$$|\underline{E}| |\underline{O}_0 z^q + \dots + \underline{O}_q| |\underline{E}^{-1}| = 0 \quad (4.2.45)$$

for which the roots are those of the determinantal equation for the parent ARMA_m(p,q) model. □

A consequence of this lemma is that if $\{\underline{u}_t\}$, $\underline{u}_t \in \mathbb{S}^m$ follows a $\ln_+ \text{ARMA}_m(p,q)$ model and is AR-invertible (definition 4.8), then the linear approximation (4.2.40) with $\underline{L}_t^* = \underline{0}$ will be invertible. A similar relationship may be found regarding the stationarity of its linear approximation. This is easily seen by noting that,

$$\underline{H}(B)^{-1} = (\underline{I} - \underline{E}(\underline{\Phi}_1 B + \dots + \underline{\Phi}_p B^p) \underline{E})^{-1} \quad (4.2.46)$$

and hence if $\underline{\Phi}^{-1}(B)$ exists (i.e. if it converges) so does $\underline{H}^{-1}(B)$, so that (4.2.40) may be written as an infinite moving-average. However in this case the roots of the determinantal equations are not so easily found to be the same. We have

$$|\underline{I}_m z^p + \underline{\Phi}_1 z^{p-1} + \dots + \underline{\Phi}_p| = 0 \quad , \quad (4.2.47)$$

and $|\underline{I}_m z^p + \underline{H}_1 z^{p-1} + \dots + \underline{H}_p| = 0 \quad . \quad (4.2.48)$

(4.2.48) may be written as

$$|\underline{E}| |\underline{E}^{-1} \underline{F}^{-1} z^p + \underline{\Phi}_1 z^{p-1} + \dots + \underline{\Phi}_p| |\underline{F}| = 0. \quad (4.2.49)$$

This equation is not the same as that of (4.2.47) (cf. (4.2.45)), unless of course $\underline{E} = \underline{F}^{-1}$, which is not necessarily the case. However consider \underline{F}^{-1} :-

$$\underline{F}^{-1} = \begin{bmatrix} \mu_1(1-\mu_1) & -\mu_2\mu_1 & \dots & -\mu_m\mu_1 \\ -\mu_1\mu_2 & \mu_2(1-\mu_2) & \dots & -\mu_m\mu_2 \\ \vdots & & \ddots & \vdots \\ -\mu_1\mu_m & -\mu_2\mu_m & \dots & \mu_m(1-\mu_m) \end{bmatrix} \quad (4.2.50)$$

$$= \underline{M}(\underline{I}_m - \underline{U} \underline{M}) \quad . \quad (4.2.51)$$

Consequently $\underline{F}^{-1} = \underline{E}$ if and only if $\underline{L} = \underline{M}$, see (4.2.30). i.e. if and only if $\mu_i = \lambda_i$ (from (4.2.33) and (4.2.34)). We recall that this is equivalent to (4.2.39). Hence we must assume that $\underline{L}_t = \underline{0}$, as in lemma 4.12. We have shown :-

Lemma 4.13

- i) If the parent ARMA_m(p,q) model for $\underline{v}_t = a_m(\underline{u}_t)$; $\underline{u}_t \in \mathbb{S}^m$ is stationary then so is the linear approximation (4.2.40) of the resulting ln₊ARMA_m(p,q) model.
- ii) If (4.2.39) holds, the determinantal equations for both AR components are the same.

In summary, in this section we have established a linear approximation for the multiplicative model. We have not considered higher order terms in order to avoid additional complications. These would demonstrate the conditions under which the linear approximation is a good one. We may crudely determine this via the choice of a "flat" region about $E[\underline{u}_t]$ as in the choice of a and b in Fig. 4.9, or k in (4.2.38).

We have also shown some interesting relationships at the linear approximation to the $\ln_+ARMA_m(p,q)$ model and its parent $ARMA_m(p,q)$. In particular we note that the stationarity and invertibility of the two ARMA models are equivalent.

4.3 Invariance Properties

In our definition of the $\ln_+ARMA_m(p,q)$ model we allowed the fill-up-value (FUV) $u_{m+1,t}$ to be the reference variable in the transformation of $\{\underline{u}_t\}$ to $\{\underline{v}_t\}$. However we could have selected any one of the remaining m u_{it} 's to be the reference variable. The resulting model would still be an $\ln_+ARMA_m(p,q)$ model, but what is its relationship to the original model? Will it make any difference to our analysis if we chose a different u_{it} ? What we must discover is whether the model is invariant to such a permutation of the u_{it} 's. In this section we aim to investigate these questions.

We begin by recalling the results at the end of section 3.4, and in particular the definition of the $\underline{Z}(k)$ matrix (3.4.2), which transforms $\underline{v}_t \rightarrow \underline{v}_t^\dagger$ thus :-

$$\underline{v}_t^\dagger = \underline{Z}(k)\underline{v}_t,$$

where \underline{v}_t and \underline{v}_t^\dagger are the logistic transformations of $\underline{u}_t \in \mathbb{S}^m$, based on the reference variables u_{m+1} and u_k respectively ($k=1, \dots, m$). We demonstrated in chapter 3 that under the usual multivariate theory if $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$, then \underline{u}^\dagger (with obvious notation) $\sim L_m(\underline{Z}(k)\underline{\mu}, \underline{Z}(k)\underline{\Sigma}\underline{Z}(k)')$. Aitchison(1982) further studied the invariance properties of the a_m transformation in the context of compositional independence.

Our first invariance property for time series is given below.

Theorem 4.14

$$\begin{aligned} \text{Let } \underline{v}_t^{(k)} &= \underline{Z}(k)\underline{v}_t \quad ; \quad t=0, \pm 1, \dots \quad ; \quad k=1, \dots, m. \\ & \\ &= \underline{Z}(k)(\underline{v}_t - \underline{u}) = \underline{v}_t^{(k)} - \underline{u}^{(k)} \quad . \end{aligned} \quad (4.3.1)$$

If $\{\underline{v}_t\}$ follows an ARMA_m(p,q) process, then $\{\underline{v}_t^{(k)}\}$ is also ARMA_m(p,q). Further the roots of the determinantal equations of both the AR and the MA components from the two models are identical so that the stationarity and invertibility conditions remain consistent.

Proof

Since $\{\underline{v}_t\}$ follows an ARMA_m(p,q) we may write :-

$$\underline{v}_t = \underline{\varepsilon}_t + \underline{\theta}_1 \underline{\varepsilon}_{t-1} + \dots + \underline{\theta}_q \underline{\varepsilon}_{t-q} - \underline{\phi}_1 \underline{v}_{t-1} - \dots - \underline{\phi}_p \underline{v}_{t-p} \quad ; \quad (4.3.2)$$

where $E[\underline{\varepsilon}_t] = \underline{0}$, and $\text{Var}[\underline{\varepsilon}_t] = \underline{\Sigma}$.

If we multiply by $\underline{Z}(k)$ the L.H.S. will be $\underline{v}_t^{(k)}$. Also, since $\underline{Z}^{-1}(k) = \underline{Z}(k)$ it follows that $\underline{v}_t = \underline{Z}(k)\underline{v}_t^{(k)}$, whence

$$\begin{aligned} \underline{v}_t^{(k)} &= \underline{\varepsilon}_t^{(k)} + \underline{Z}(k)\underline{\theta}_1 \underline{Z}(k)\underline{\varepsilon}_{t-1}^{(k)} + \dots + \underline{Z}(k)\underline{\theta}_q \underline{Z}(k)\underline{\varepsilon}_{t-q}^{(k)} - \\ &\quad \underline{Z}(k)\underline{\phi}_1 \underline{Z}(k)\underline{v}_{t-1}^{(k)} - \dots - \underline{Z}(k)\underline{\phi}_p \underline{Z}(k)\underline{v}_{t-p}^{(k)} \quad , \end{aligned} \quad (4.3.3)$$

$$\text{where} \quad \underline{\varepsilon}_t^{(k)} = \underline{Z}(k)\underline{\varepsilon}_t \quad ; \quad k=1, \dots, m. \quad (4.3.4)$$

$$\text{If we then put} \quad \underline{\theta}_i^{(k)} = \underline{Z}(k)\underline{\theta}_i \underline{Z}(k) \quad ; \quad i=0, \dots, q \quad ; \quad (4.3.5)$$

$$\text{and} \quad \underline{\phi}_j^{(k)} = \underline{Z}(k)\underline{\phi}_j \underline{Z}(k) \quad ; \quad j=1, \dots, p \quad ; \quad k=1, \dots, m; \quad (4.3.6)$$

we obtain

$$\underline{v}_t^{(k)} = \underline{\varepsilon}_t^{(k)} + \underline{\theta}_1^{(k)} \underline{\varepsilon}_{t-1}^{(k)} + \dots + \underline{\theta}_q^{(k)} \underline{\varepsilon}_{t-q}^{(k)} - \underline{\phi}_1^{(k)} \underline{v}_{t-1}^{(k)} - \dots - \underline{\phi}_p^{(k)} \underline{v}_{t-p}^{(k)} \quad . \quad (4.3.7)$$

From (4.3.5) and (4.3.6) , and since $|\underline{Z}(k)| \neq 0$,

$$\left. \begin{aligned} \underline{\theta}_i = \underline{0} &\Leftrightarrow \underline{\theta}_i^{(k)} = \underline{0} \quad ; \quad i=0, \dots, q \quad ; \\ \underline{\phi}_j = \underline{0} &\Leftrightarrow \underline{\phi}_j^{(k)} = \underline{0} \quad ; \quad j=1, \dots, p. \end{aligned} \right\} k=1, \dots, m; \quad (4.3.8)$$

Hence (4.3.3) is not only an ARMA model but it contains the same number of parameters as the original ARMA model (4.3.2), and in particular remains ARMA_m(p,q), with white-noise process $\underline{\varepsilon}_t^{(k)}$,

$$\begin{aligned} \text{where } E[\underline{\varepsilon}_t^{(k)}] &= \underline{0} \quad , \\ \text{and } \text{Var}[\underline{\varepsilon}_t^{(k)}] &= \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k). \end{aligned} \quad (4.3.9)$$

Consider now the roots of :-

$$\left| \underline{I}_m z^p + \underline{\phi}_1^{(k)} z^{p-1} + \dots + \underline{\phi}_p^{(k)} \right| = 0 \quad , \quad (4.3.10)$$

$$\equiv \left| \underline{Z}(k) \right| \left| \underline{I}_m z^p + \underline{\phi}_1 z^{p-1} + \dots + \underline{\phi}_p \right| \left| \underline{Z}(k) \right| = 0 \quad ,$$

$$\equiv \left| \underline{I}_m z^p + \underline{\phi}_1 z^{p-1} + \dots + \underline{\phi}_p \right| = 0 \quad , \quad (4.3.11)$$

which is the characteristic equation for the original AR component, and hence the roots are identical. The MA component follows in a similar manner. □

Corollary 4.15

Let $\underline{\phi}_j = \{\phi_{j,(r,s)}\} \quad r,s=1, \dots, m$; (note slight change in notation)

and $\underline{\phi}_j^{(k)} = \{\phi_{j,(r,s)}^{(k)}\} \quad , \text{ where}$

$$\left. \begin{aligned} \phi_{j,(r,s)}^{(k)} &= \phi_{j,(r,s)} - \phi_{j,(k,s)} \quad ; \quad s=1, \dots, m; \quad s \neq k \\ &= \phi_{j,(r,m+1)} - \phi_{j,(k,m+1)} \quad ; \quad s=k \\ &= \phi_{j,(r,k)} - \phi_{j,(k,k)} \quad ; \quad s=m+1 \\ &= -\phi_{j,(k,s)} \quad ; \quad s=1, \dots, m; \quad s \neq k \\ &= -\phi_{j,(k,m+1)} \quad ; \quad s=k \\ &= -\phi_{j,(k,k)} \quad ; \quad s=m+1 \end{aligned} \right\} \left. \begin{array}{l} r=1, \dots, m \\ r \neq k \\ \\ \\ r=k \end{array} \right\} \left. \begin{array}{l} j=1, \dots, p \\ k=1, \dots, m \end{array} \right\} \quad (4.3.12)$$

(We recall that $\phi_{j,(\ell,m+1)} = -\sum_{i=1}^m \phi_{j,(\ell,i)}$.)

$$\left. \begin{aligned} \text{Similarly if we define } \theta_{j,(\ell,m+1)} &= -\sum_{i=1}^m \theta_{j,(\ell,i)}, \\ \text{and } \theta_{j,(\ell,m+1)}^{(k)} &= -\sum_{i=1}^m \theta_{j,(\ell,i)}^{(k)} ; k=1, \dots, m. \end{aligned} \right\} \begin{array}{l} j=0, \dots, q \\ (4.3.13) \end{array}$$

$$\left. \begin{aligned} \theta_{j,(r,s)}^{(k)} &= \theta_{j,(r,s)} - \theta_{j,(k,s)} ; s=1, \dots, m ; s \neq k \\ &= \theta_{j,(r,m+1)} - \theta_{j,(k,m+1)} ; s=k \\ &= \theta_{j,(r,k)} - \theta_{j,(k,k)} ; s=m+1 \\ &= -\theta_{j,(k,s)} ; s=1, \dots, m ; s \neq k \\ &= -\theta_{j,(k,m+1)} ; s=k \\ &= -\theta_{j,(k,k)} ; s=m+1 \end{aligned} \right\} \begin{array}{l} \left. \begin{array}{l} r=1, \dots, m \\ r \neq k \end{array} \right\} j=0, \dots, q \\ r=k \left\{ \begin{array}{l} k=1, \dots, m \\ (4.3.14) \end{array} \right. \end{array}$$

Proof

a) Via the ARMA_m(p,q) model. From (4.3.5) and (4.3.6) we have

$$\underline{\theta}_i^{(k)} = \underline{Z}(k) \underline{\theta}_i \underline{Z}(k) \quad \text{and}$$

$$\underline{\phi}_j^{(k)} = \underline{Z}(k) \underline{\phi}_j \underline{Z}(k) \quad ,$$

where
$$\underline{Z}(k) = \begin{bmatrix} 1 & 0 & \dots & -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 0 & \dots & 1 \end{bmatrix} ;$$

multiplying out these matrix expressions gives the required result.

b) Via the ln₊ARMA_m(p,q) model (we give a sketch only).

$$\begin{aligned} \text{Let } F_r &= \begin{bmatrix} p & m+1 & -\phi_{j,(r,s)} \\ \Pi & \Pi & u_{s,t-j} \end{bmatrix} \begin{bmatrix} q & m & \theta_{j,(r,s)} \\ j=0 & s=1 & a_{s,t-j} \end{bmatrix} \eta_r ; r=1, \dots, m. \\ &= 1 ; r=m+1. \end{aligned} \quad (4.3.15)$$

Then the $ln_+ ARMA_m(p,q)$ model may be expressed as

$$u_{rt} = \frac{F_r}{\sum_{i=1}^{m+1} F_i}, \quad r=1, \dots, m+1; \quad (4.3.16)$$

(cf. (4.2.11) and (4.2.12))

$$= \frac{F_r / F_k}{\sum_{i=1}^{m+1} F_i / F_k} = \frac{F_r^{(k)}}{\sum_{i=1}^{m+1} F_i^{(k)}} \quad (\text{say}) \quad r=1, \dots, m+1. \quad (4.3.17)$$

where $F_r^{(k)} = F_r / F_k$; $r=1, \dots, m$; $r \neq k$

$$= 1 \quad ; \quad r=k$$

$$= 1 / F_k \quad ; \quad r=m+1.$$

If we consider each term of $F_r^{(k)}$ we obtain,

$$F_r^{(k)} = \frac{F_r}{F_k} = \left[\prod_{j=1}^p \prod_{s=1}^{m+1} u_{s,t-j}^{-\phi_{j,(r,s)} + \phi_{j,(k,s)}} \right] \left[\prod_{j=0}^q \prod_{s=1}^m a_{s,t-j}^{\theta_{j,(r,s)} - \theta_{j,(k,s)}} \right] \eta_r / \eta_k ;$$

$r=1, \dots, m$; $r \neq k$. (4.3.18)

Clearly the coefficients of the u 's correspond to $\phi_{j,(r,s)}^{(k)}$ of (4.3.12) for $r=1, \dots, m$; $r \neq k$, but with $s=k$, and $s=m+1$ interchanged. Since the whole point of the above theorem is to interchange the u_{kt} and $u_{m+1,t}$ values so that $\phi_{j,(r,k)}^{(k)}$ refers to the coefficient of $u_{m+1,t-j}$ (and vice versa) the expressions are as required. Similarly examining the result for $r=m+1$ above, gives the result for $r=k$ in (4.3.12). (Again k and $m+1$ interchanged.)

Consider now the coefficients of the $a_{s,t-j}$'s :-

$$\left[\prod_{j=0}^q \prod_{s=1}^m a_{s,t-j}^{\theta_{j,(r,s)} - \theta_{j,(k,s)}} \right] =$$

$$\left[\prod_{j=0}^q \prod_{s=1}^m \left(a_{s,t-j} / a_{k,t-j} \right)^{\theta_{j,(r,s)} - \theta_{j,(k,s)}} \right] \left[\prod_{j=0}^q \prod_{s=1}^m a_{k,t-j}^{\theta_{j,(r,s)} - \theta_{j,(k,s)}} \right] =$$

$$\left[\prod_{j=0}^q \prod_{\substack{s=1 \\ s \neq k}}^m a_{s,t-j}^{(k)} \theta_{j,(r,s)} - \theta_{j,(k,s)} \right] \left[\prod_{j=0}^q \prod_{s=1}^m a_{m+1,t-j}^{\left(-\sum_{s=1}^m \theta_{j,(r,s)} + \sum_{s=1}^m \theta_{j,(k,s)} \right)} \right] =$$

$$\prod_{j=0}^q \prod_{\substack{s=1 \\ s \neq k}}^m a_{s,t-j}^{(k)} \theta_{j,(r,s)} - \theta_{j,(k,s)} \quad (4.3.19)$$

(from (4.3.13)) .

Since $\underline{\varepsilon}_t^{(k)} = \underline{Z}(k)\underline{\varepsilon}_t$, and $\underline{a}_t = \exp \underline{\varepsilon}_t$, $\underline{a}_t^{(k)}$ corresponds to

$$\underline{a}_t^{(k)} = \exp \underline{\varepsilon}_t^{(k)} \quad ; \quad t=0, \pm 1, \dots \quad ; \quad k=1, \dots, m \quad ; \quad (4.3.20)$$

with $a_{m+1,t}^{(k)}$ corresponding to the k^{th} position of the vector $\underline{a}_t^{(k)}$.

Thus once again interchanging the k^{th} and $m+1^{\text{st}}$ indices between the $\underline{\Theta}$ and $\underline{\Theta}^{(k)}$ parameters the power of $a_{s,t-j}^{(k)}$ in $F_r^{(k)}$ is $\theta_{j,(r,s)}^{(k)}$ given by (4.3.13), for $r=1, \dots, m$; $r \neq k$: $j=0, \dots, q$: $k=1, \dots, m$: $t=0, \pm 1, \dots$: and $s=1, \dots, m+1$; $s \neq k$. Again we may repeat for $F_r^{(k)} = 1/F_k$; $r=m+1$ to give the remaining results. □

Theorem 4.14 demonstrates that the basic structural form of the $\ln_+ \text{ARMA}_m(p,q)$ model and its parent $\text{ARMA}_m(p,q)$ model is invariant to the choice of reference variable. Whatever $\text{ARMA}_m(p,q)$ model (e.g. (4.3.2) or (4.3.3)) is selected to represent the \underline{u}_t series, it is of the same order and has the same stationarity and invertibility properties. We may take this one step further.

Theorem 4.16

The $\ln_+ \text{ARMA}_m(p,q)$ model for a series $\underline{u}_t \in \mathbb{S}^m$ is totally invariant to the choice of reference variable. That is, any of the $\text{ARMA}_m(p,q)$ models (4.3.2) and (4.3.7) represent the same model for \underline{u}_t on the simplex, except that the $\underline{u}_t^{(m+1)}$'s have been permuted.

Proof

From (4.3.16) and (4.3.17) we have that

$$u_{rt} = \frac{F_r^{(k)}}{\sum_{i=1}^{m+1} F_i^{(k)}} ; r=1, \dots, m+1 ; k \in (1, \dots, m+1) ; \quad (4.3.21)$$

where we set $F_r^{(m+1)} = F_r$ as in (4.3.16). From (4.3.18), (4.3.12) and (4.3.13) we have that,

$$F_r^{(k)} = \begin{pmatrix} p & m+1 & -\phi_{j,(r,s)}^{(k)} \\ \prod_{j=1} & \prod_{s=1} & u_{s,t-j} \end{pmatrix} \begin{pmatrix} q & m & \theta_{j,(r,s)}^{(k)} \\ \prod_{j=0} & \prod_{s=1} & a_{s,t-j}^{(k)} \end{pmatrix} \eta_r^{(k)} \quad ; r=1, \dots, m+1 \\ r \neq k ;$$

$$= 1 \quad r=k ; \quad (4.3.22)$$

$$\text{where } \left. \begin{aligned} \eta_r^{(k)} &= \frac{\eta_r}{\eta_k} \quad r=1, \dots, m \\ &= \frac{1}{\eta_k} \quad r=k ; \end{aligned} \right\} k=1, \dots, m ;$$

$$= \eta_r \quad r=1, \dots, m ; \quad k=m+1.$$

Comparing (4.3.21) and (4.3.22) with (4.1.9) and the proof of lemma 4.3, we see that (4.3.21) is just the $\ln_+ARMA_m(p,q)$ process in terms of \underline{u}_t , corresponding to the $ARMA_m(p,q)$ model for $\underline{v}_t^{(k)}$, $k \in \{1, \dots, m+1\}$ given by (4.3.7). (Again we let $\underline{v}_t^{(m+1)} = \underline{v}_t$, $\phi_i^{(m+1)} = \phi_i$, $i=1, \dots, p$; $\theta_j^{(m+1)} = \theta_j$, $j=0, \dots, q$.) Hence for these $ARMA_m(p,q)$ models to be equivalent forms of the same $\ln_+ARMA_m(p,q)$ model we require that the models represented by (4.3.21) be equivalent. But we have from (4.3.21) that for any constant C,

$$u_{rt} = \frac{F_r^{(k)} / C}{\sum_{i=1}^{m+1} F_i^{(k)} / C} ; r=1, \dots, m+1 ; k \in (1, \dots, m+1) . \quad (4.3.23)$$

We may choose C to be $F_s^{(k)}$, for which we obtain (4.3.21) for $k=s$, as

can be seen by (4.3.16)-(4.3.18). Hence all the models are equivalent, except for a permutation of the u 's.

□

From (4.3.23) we see that (4.3.21) must have a constraint on the elements to give a unique representation. This constraint is that $F_k^{(k)} = 1$, and is equivalent to choosing u_{kt} as the reference variable. (c.f. the marginal constraints of a log-linear model for a contingency table e.g. Bishop et al (1975).)

Aitchison and Shen(1980) demonstrated that $L_m(\underline{\mu}, \underline{\Sigma})$ and $L_m(\underline{Z}(k)\underline{\mu}, \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k))$ represent the same distribution, but with a permutation of the elements in \mathbb{S}^m . Our result is the analogue of this for the $\ln_+ARMA_m(p,q)$ model. We may investigate the validity of the theorem above from many directions. For example, in (4.3.7) we may let $\underline{\varepsilon}_t^{(k)} \sim N_m(\underline{0}, \underline{\Sigma}^{(k)})$, so that the innovation series $\underline{e}_t = a_m^{-1}(\underline{\varepsilon}_t^{(k)})$ is $L_m(\underline{0}, \underline{\Sigma}^{(k)})$. Thus from Aitchison and Shen's (1980) result $\underline{\varepsilon}_t^{(1)}, \underline{\varepsilon}_t^{(2)}, \dots, \underline{\varepsilon}_t^{(m+1)}$ all represent the same innovation series on \mathbb{S}^m i.e. \underline{e}_t , but under a particular permutation. This is true for $\underline{v}_t^{(1)}, \dots, \underline{v}_t^{(m+1)}$ also, since they all represent \underline{u}_t . Hence all the $ARMA_m(p,q)$ models given by (4.3.7) represent $\underline{v}_t^{(k)}$ (and consequently \underline{u}_t) as a function of its own past and the past and present of $\underline{\varepsilon}_t^{(k)}$ (and therefore \underline{e}_t). i.e. all are expressions for \underline{u}_t in terms of its own past and the innovation series \underline{e}_t .

Another example occurs when we consider forecasts. Clearly if $\hat{\underline{v}}_{t+l}^{(k)}$ is a forecast for $\underline{v}_{t+l}^{(k)}$, then $\hat{\underline{v}}_{t+l}^{(k)} = \underline{Z}(k)\hat{\underline{v}}_{t+l}^{(k)}$ as can be seen by applying $\underline{Z}(k)$ to (2.5.2) or by noting that $\hat{\underline{v}}_{t+l}^{(k)}$ is obtained from (4.3.7) by recursively computing $\underline{v}_{t+i}^{(k)}$ $i=1, \dots, l$ with $\underline{\varepsilon}_{t+l}^{(k)} = \underline{0}$. As in (4.3.5) and (4.3.6) we note that in (2.5.1) $\underline{\Psi}_i^{(k)} = \underline{Z}(k)\underline{\Psi}_i^{(m+1)}\underline{Z}(k)$; $i=1, 2, \dots$. Then from (2.5.3)

$$\begin{aligned} \underline{\Sigma}_{\underline{v}_{t+l}^{(k)}}^{(k)} &= \text{Var}(\hat{\underline{v}}_{t+l}^{(k)} / \underline{v}_t^{(k)}, \dots) = \underline{\Sigma}^{(k)} + \underline{\Psi}_1^{(k)}\underline{\Sigma}^{(k)}\underline{\Psi}_1'^{(k)} + \dots + \underline{\Psi}_{l-1}^{(k)}\underline{\Sigma}^{(k)}\underline{\Psi}_{l-1}'^{(k)} \\ &= \underline{Z}(k)\underline{\Sigma}^{(m+1)}\underline{Z}'(k) + \underline{Z}(k)\underline{\Psi}_1^{(m+1)}\underline{Z}(k)\underline{\Sigma}^{(m+1)}\underline{Z}'(k)\underline{\Psi}_1'^{(m+1)}\underline{Z}'(k) + \dots \\ &\quad + \underline{Z}(k)\underline{\Psi}_{l-1}^{(m+1)}\underline{Z}(k)\underline{\Sigma}^{(m+1)}\underline{Z}'(k)\underline{\Psi}_{l-1}'^{(m+1)}\underline{Z}'(k) \end{aligned}$$

$$\begin{aligned}
&= \underline{Z}(k) \left[\underline{\Sigma}^{(m+1)} + \underline{\Psi}_1^{(m+1)} \underline{\Sigma}^{(m+1)} \underline{\Psi}'_1^{(m+1)} + \dots + \underline{\Psi}_{\ell-1}^{(m+1)} \underline{\Sigma}^{(m+1)} \underline{\Psi}'_{\ell-1}^{(m+1)} \right] \underline{Z}'(k) \\
&= \underline{Z}(k) \left[\text{Var}(\hat{\underline{v}}_{t+\ell}^{(m+1)} / \underline{v}_t^{(m+1)}, \dots) \right] \underline{Z}'(k) \\
&= \underline{Z}(k) \underline{\Sigma}_{\underline{v}_{t+\ell}}^{(m+1)} \underline{Z}'(k) \quad . \quad (4.3.24)
\end{aligned}$$

Hence the forecast for $\underline{u}_{t+\ell} = E[\underline{u}_{t+\ell} / \underline{u}_t, \underline{u}_{t-1}, \dots]$, where

$$\underline{u}_{t+\ell} \sim L_m \left[\underline{v}_t^{(m+1)}, \underline{\Sigma}_{\underline{v}_{t+\ell}}^{(m+1)} \right], \text{ or equivalently } L_m \left[\underline{v}_t^{(k)}, \underline{\Sigma}_{\underline{v}_{t+\ell}}^{(k)} \right] \text{ from}$$

(4.3.24), which is the same distribution except that the \underline{u}_t 's have been permuted.

In summary we have shown that whatever model we use on the \mathbb{R}^m space to express the $\ln_+ARMA_m(p,q)$ model; we achieve the same result on "returning" to the \mathbb{S}^m space. Consequently any results such as forecasts obtained from the model will be identical. We may map from one model to another using (4.3.1), (4.3.4)-(4.3.6) and (4.3.9). Maximum likelihood estimates of the parameters also remain consistent. We obtain :-

$$\hat{\underline{\Sigma}}(k) = \underline{Z}(k) \hat{\underline{\Sigma}}^{(m+1)} \underline{Z}'(k) \quad (4.3.25)$$

$$\hat{\underline{\Theta}}(k) = \underline{Z}(k) \hat{\underline{\Theta}}^{(m+1)} \underline{Z}(k) \quad (4.3.26)$$

$$\hat{\underline{\Phi}}(k) = \underline{Z}(k) \hat{\underline{\Phi}}^{(m+1)} \underline{Z}(k) \quad (4.3.27)$$

$$\hat{\underline{v}}(k) = \underline{Z}(k) \hat{\underline{v}}^{(m+1)} \quad (4.3.28)$$

$$\hat{\underline{\epsilon}}_t(k) = \underline{Z}(k) \hat{\underline{\epsilon}}_t^{(m+1)} \quad (4.3.29)$$

(See for example Mood et al (1974) p. 284.)

Because of (4.3.29) the diagnostic checks should be invariant, and because the models are of identical orders (i.e. because of (4.3.26) and (4.2.27)) the same model should be identified. It will be useful to examine briefly how to map some of the identification statistics between different models on \mathbb{R}^m . We establish relationships between population

parameters and also between their estimates.

Lemma 4.17

Let $\underline{v}_t = a_m(\underline{u}_t)$; $\underline{u}_t \in \mathbb{S}^m$ and $\underline{v}_t^{(k)} = \underline{Z}(k)\underline{v}_t$. Then

- i) $\underline{\Gamma}^{(k)}(s) = \underline{Z}(k)\underline{\Gamma}(s)\underline{Z}'(k)$
- ii) $\underline{C}^{(k)}(s) = \underline{Z}(k)\underline{C}(s)\underline{Z}'(k)$
- iii) $\underline{\Omega}^{(k)}(s) = \underline{Z}(k)\underline{\Omega}(s)\underline{Z}'(k)$
- iv) $\underline{S}^{(k)}(s) = \underline{Z}(k)\underline{S}(s)\underline{Z}'(k)$
- v) $\underline{P}^{(k)}(s) = \underline{Y}(k)\underline{P}(s)\underline{Y}'(k)$
- vi) $\underline{R}^{(k)}(s) = \hat{\underline{Y}}(k)\underline{R}(s)\hat{\underline{Y}}'(k)$; $s=0,\pm 1,\dots$; $k=1,\dots,m$;

where $\underline{\Gamma}(s), \dots, \underline{R}(s), \underline{\Gamma}^{(k)}(s), \dots, \underline{R}^{(k)}(s)$, are the various cross-covariance functions and cross-correlations functions of \underline{v}_t and $\underline{v}_t^{(k)}$ respectively. (Definitions 2.10, 2.11, 2.13 and 2.14.)

Finally
$$\underline{Y}(k) = \underline{\Delta}^{(k)-\frac{1}{2}} \underline{Z}(k)\underline{\Delta}^{\frac{1}{2}} , \tag{4.3.30}$$

where
$$\underline{\Delta} = dg(\gamma_{11}(0), \dots, \gamma_{mm}(m)) , \tag{4.3.31}$$

and
$$\underline{\Delta}^{(k)} = dg(\gamma_{11}^{(k)}(0), \dots, \gamma_{mm}^{(k)}(m)) . \tag{4.3.32}$$

And
$$\hat{\underline{Y}}(k) = \hat{\underline{\Delta}}^{(k)-\frac{1}{2}} \underline{Z}(k)\hat{\underline{\Delta}}^{\frac{1}{2}} , \tag{4.3.33}$$

with
$$\hat{\underline{\Delta}} = dg(c_{11}(0), \dots, c_{mm}(m)) , \tag{4.3.34}$$

and
$$\hat{\underline{\Delta}}^{(k)} = dg(c_{11}^{(k)}(0), \dots, c_{mm}^{(k)}(m)) . \tag{4.3.35}$$

Proof

Since $\underline{v}_t^{(k)} = \underline{Z}(k)\underline{v}_t$; $t=0,\pm 1,\dots$; $k=1,\dots,m$. all the results follow from standard multivariate theory e.g.

$$\begin{aligned} \text{i) } \underline{\Gamma}^{(k)}(s) &= E\left[\underline{v}_t^{(k)}\underline{v}_{t-s}^{\prime(k)}\right] = E\left[\underline{Z}(k)\underline{v}_t\underline{v}_{t-s}'\underline{Z}'(k)\right] = \underline{Z}(k)\left[E\underline{v}_t\underline{v}_{t-s}'\underline{Z}'(k)\right] \\ &= \underline{Z}(k)\underline{\Gamma}(s)\underline{Z}'(k) , \text{ as required. } (s=0,\pm 1,\dots ; k=1,\dots,m.) \end{aligned}$$

$$\text{ii) } \underline{S}^{(k)}(s) = \underline{C}^{(k)}(s)\underline{C}^{(k)}(0)^{-1} = \underline{Z}(k)\underline{C}(s)\underline{Z}'(k)\{\underline{Z}(k)\underline{C}(0)\underline{Z}'(k)\}^{-1}$$

$$= \underline{Z}(k)\underline{C}(s)\underline{C}(0)^{-1}\underline{Z}'(k) = \underline{Z}(k)\underline{S}(s)\underline{Z}(k) \text{ as required.}$$

$$\begin{aligned} \text{v) } \underline{P}^{(k)}(s) &= \underline{\Delta}^{(k)-\frac{1}{2}} \underline{\Gamma}(s) \underline{\Delta}^{(k)-\frac{1}{2}} = \underline{\Delta}^{(k)-\frac{1}{2}} \underline{Z}(k) \underline{\Gamma}(s) \underline{Z}'(k) \underline{\Delta}^{(k)-\frac{1}{2}} \\ &= \underline{\Delta}^{(k)-\frac{1}{2}} \underline{Z}(k) \underline{\Delta}^{\frac{1}{2}} \underline{\Delta}^{-\frac{1}{2}} \underline{\Gamma}(s) \underline{\Delta}^{-\frac{1}{2}} \underline{\Delta}^{\frac{1}{2}} \underline{Z}'(k) \underline{\Delta}^{(k)-\frac{1}{2}} = \underline{Y}(k) \underline{P}(s) \underline{Y}'(k) , \end{aligned}$$

as required. ($s=0, \pm 1, \dots$; $k=1, \dots, m.$)

□

In theory examination of the cross-correlation functions at the identification stage should yield the same model. If $\underline{v}_t^{(k)}$ $k \in (1, \dots, m+1)$ is $MA_m(q)$ then $\underline{P}(s)^{(k)} = \underline{0}$, $|s| > q$. Now from the lemma it follows that $\underline{P}^{(k)}(s) = \underline{0}$ if and only if $\underline{P}(s) = \underline{0}$. However in practice the elements of the sample cross-correlation matrices are examined to see if they are "small" in some sense. Could it be possible that the transformation of e.g. $\underline{R}(s) \rightarrow \underline{R}^{(k)}(s)$ distorts the elements so that in one the elements are "small", whilst in another they are "large"? This would result in different order models being selected, depending on which variable was used as the reference variable. It does seem unlikely that such a fluke would occur in practice, especially since all the sample estimates are asymptotically normal with mean equal to the corresponding population parameter. (See e.g. Hannan(1970).) We might try and look at individual elements of $\underline{R}^{(k)}(s)$ in terms of $\underline{R}(s)$, but when tried we found this added little to the picture. However, as we shall see, and by theorem 4.16 identification statistics, such as the likelihood ratio statistic $M(\ell)$ (2.2.16) are invariant. By examining these and also possibly computing $\underline{R}^{(k)}(s)$ for $k=1, \dots, m+1$ we may easily circumvent this. However, bearing in mind the theoretical results it does seem to be an unlikely event in any case. To illustrate this point we will consider a simple example.

Example 4.18

We will use the data set relating to the GALLUP(c) poll on political preference. The data runs from January 1965 to December 1973 and gives the results of a monthly poll for allegiance to the following political parties :-

Conservative, Labour, Liberal, and Other.

(CON_t) (LAB_t) (LIB_t) (OTH_t)

The data set lies on the \mathbb{S}^3 simplex. First we transform the data via the a_m function so that it lies on the \mathbb{R}^m space. We may choose any of the four variables to be the reference variable. With obvious notation we will form the series,

$$\underline{v}_t^{(CON)}, \underline{v}_t^{(LAB)}, \underline{v}_t^{(LIB)} \text{ and } \underline{v}_t^{(OTH)}.$$

e.g.
$$\underline{v}_t^{(LAB)} = \left[\ln \left[\frac{CON_t}{LAB_t} \right], \ln \left[\frac{LIB_t}{LAB_t} \right], \ln \left[\frac{OTH_t}{LAB_t} \right] \right]'$$

The sample autocovariance function $\underline{C}(j)$ and the sample autocorrelation function $\underline{R}(j)$ $j=0, \dots, 30$; were computed for each of the $\underline{v}_t^{(\sim)}$'s series, (" \sim " \in CON, LAB, LIB, OTH). In order to check the computation the $\underline{C}(j)$ and $\underline{R}(j)$ were also transformed via the $\underline{Z}(k)$ matrix ($k=1, \dots, 3$). This demonstrated that the computation was correct. Below are the results in symbolic form (see chapter 2) for the first 5 lags.

Reference

<u>Variable</u> ↓	<u>Lag</u> →	0	1	2	3	4	5
CON		+ . -	+ . -	+ . -	+ . -	+ . -	+ . -
		. + +	. + .	. + .	. + .	. + .	. + .
		- + +	- + +	- . +	- . +	- + +	- . +
LAB		+ + +	+ + +	+ + +	+ + +	+ + +	+ + +
		+ + +	+ + +	+ + +	+ + +	+ + +	+ + +
		+ + +	+ + +	+ + +	+ + +	+ + +	+ + +
LIB		+ + +	+ + .	+ + .	+ + .	+ + .	+ + .
		+ + .	+ + .	+ + .	+ + .	+ + .	+ + .
		+ . +	. . +	. . +	. . +	. . +	. - +
OTH		+ + +	+ + +	+ + +	+ + +	+ + +	+ + +
		+ + +	+ + +	+ + +	+ + +	+ + +	+ + +
		+ + +	+ + +	+ + +	+ + +	+ + +	+ + +

For all four series, the pattern of "+" and "-" does not disappear with higher order lags. All contain off-diagonal elements that are significantly different from zero. Hence the patterns presented by the $\underline{R}(j)$ all indicate a similar type of model, namely an $AR_3(p)$ or $ARMA_3(p,q)$ ($p \neq 0$) model.

$$\underline{F}^{(k)}(s) = \underline{Z}(k)\underline{F}(s)\underline{Z}(k) \quad ; \quad (4.3.37)$$

$$s=0, \pm 1, \dots; k=1, \dots, m ;$$

$$\widehat{\underline{F}}^{(k)}(s) = \underline{Z}(k)\widehat{\underline{F}}(s)\underline{Z}(k) . \quad (4.3.38)$$

Proof

The result may be seen immediately from the fact that $\underline{\Phi}_i^{(k)} = \underline{Z}(k)\underline{\Phi}_i\underline{Z}(k)$ $i=1, \dots, p$; and from the understanding that $\underline{F}(s)$ and $\widehat{\underline{F}}(s)$ are simply estimates of $\underline{\Phi}_\ell$ under the assumption that $\ell=p$. However we may also demonstrate the theorem by examination of (2.2.5)-(2.2.7) and (2.2.9) and (2.2.10) respectively. Firstly from (2.2.7)

$$\underline{F}'(1) = \underline{\Gamma}^{-1}(0)\underline{\Gamma}(1) , \text{ and}$$

$$\underline{F}'^{(k)}(1) = \underline{\Gamma}^{(k)-1}(0)\underline{\Gamma}^{(k)}(1)$$

$$= \{\underline{Z}(k)\underline{\Gamma}(0)\underline{Z}'(k)\}^{-1}\{\underline{Z}(k)\underline{\Gamma}(1)\underline{Z}'(k)\} = \underline{Z}'(k)\underline{F}'(1)\underline{Z}'(k) .$$

$$\therefore \underline{F}^{(k)}(1) = \underline{Z}(k)\underline{F}(1)\underline{Z}(k) . \quad (4.3.39)$$

For higher order terms, again from (2.2.7) we have

$$\underline{F}'(\ell) = \{\underline{\Gamma}(0) - \underline{b}'(\ell,0)\underline{A}^{-1}(\ell,0)\underline{b}(\ell,0)\}^{-1}\{\underline{\Gamma}(\ell) - \underline{b}'(\ell,0)\underline{A}^{-1}(\ell,0)\underline{c}(\ell,0)\} ; \quad \ell > 1 .$$

Dropping the second term of the bracketed terms for simplicity, and writing $\underline{b}(\ell,0) = \underline{b}(\ell)$ etc. gives,

$$\underline{F}'(\ell) = \{\underline{\Gamma}(0) - \underline{b}'(\ell)\underline{A}^{-1}(\ell)\underline{b}(\ell)\}^{-1}\{\underline{\Gamma}(\ell) - \underline{b}'(\ell)\underline{A}^{-1}(\ell)\underline{c}(\ell)\} ; \quad (4.3.40)$$

and similarly

$$\underline{F}'^{(k)}(\ell) = \{\underline{\Gamma}^{(k)}(0) - \underline{b}'^{(k)}(\ell)\underline{A}^{(k)}(\ell)\underline{b}^{(k)}(\ell)\}^{-1} \times \{\underline{\Gamma}^{(k)}(\ell) - \underline{b}'^{(k)}(\ell)\underline{A}^{(k)}(\ell)\underline{c}^{(k)}(\ell)\} . \quad (4.3.41)$$

$$\text{where } \underline{A}^{(k)}(\ell) = \begin{bmatrix} \underline{\Gamma}^{(k)}(0) & \underline{\Gamma}^{(k)}(-1) & \dots & \underline{\Gamma}^{(k)}(2-\ell) \\ \underline{\Gamma}^{(k)}(1) & \underline{\Gamma}^{(k)}(0) & \dots & \underline{\Gamma}^{(k)}(1-\ell) \\ \vdots & & \ddots & \vdots \\ \underline{\Gamma}^{(k)}(\ell-2) & \underline{\Gamma}^{(k)}(\ell-3) & \dots & \underline{\Gamma}^{(k)}(0) \end{bmatrix} = \underline{Z}^{\alpha k} \underline{A}(\ell) \underline{Z}'^{\alpha k} . \quad (4.3.42)$$

$$\underline{Z}^{\alpha k} = \underline{I}_{\ell-1} \otimes \underline{Z}(k) , \quad (4.3.43)$$

i.e. an $(\ell-1)m \times (\ell-1)m$ matrix with $\underline{Z}(k)$ in the diagonal.

Similarly it may easily be seen from (2.2.5) that,

$$\underline{b}^{(k)}(\ell) = \underline{Z}^{\alpha k} \underline{b}(\ell) \underline{Z}'(k), \quad \ell > 1 \quad (4.3.44)$$

$$\underline{c}^{(k)}(\ell) = \underline{Z}^{\alpha k} \underline{c}(\ell) \underline{Z}'(k). \quad (4.3.45)$$

Noting that $\underline{Z}^{\alpha k - 1} = \underline{Z}^{\alpha k}$, we may substitute (4.3.42) through to (4.3.45) into (4.3.41) to give,

$$\begin{aligned} \underline{F}'^{(k)}(\ell) &= \left[\underline{Z}(k) \underline{\Gamma}(0) \underline{Z}'(k) - \underline{Z}(k) \underline{b}'(\ell) \underline{Z}'^{\alpha k} \left(\underline{Z}^{\alpha k} \underline{A}(\ell) \underline{Z}'^{\alpha k} \right)^{-1} \underline{Z}^{\alpha k} \underline{b}(\ell) \underline{Z}'(k) \right]^{-1} \times \\ &\quad \left[\underline{Z}(k) \underline{\Gamma}(\ell) \underline{Z}'(k) - \underline{Z}(k) \underline{b}'(\ell) \underline{Z}'^{\alpha k} \left(\underline{Z}^{\alpha k} \underline{A}(\ell) \underline{Z}'^{\alpha k} \right)^{-1} \underline{Z}^{\alpha k} \underline{c}(\ell) \underline{Z}'(k) \right] \\ &= \{ \underline{Z}(k) (\underline{\Gamma}(0) - \underline{b}'(\ell) \underline{A}^{-1}(\ell) \underline{b}(\ell)) \underline{Z}'(k) \}^{-1} \times \\ &\quad \{ \underline{Z}(k) (\underline{\Gamma}(\ell) - \underline{b}'(\ell) \underline{A}^{-1}(\ell) \underline{c}(\ell)) \underline{Z}'(k) \} \\ &= \underline{Z}'(k) \underline{F}'(\ell) \underline{Z}'(k). \end{aligned}$$

$$\text{Hence } \underline{F}^{(k)}(\ell) = \underline{Z}(k) \underline{F}(\ell) \underline{Z}(k), \quad \ell=1,2,\dots \quad (4.3.46)$$

(4.3.39) and (4.3.46) are (4.3.37).

To prove (4.3.38) we recall that $\hat{\underline{F}}(s)$ is defined to be the last $m \times m$ matrix of $\hat{\underline{\Phi}}'_\ell$, where from (2.2.10),

$$\hat{\underline{\Phi}}'_\ell = (\underline{Y}'_\ell \underline{Y}_\ell)^{-1} \underline{Y}'_\ell \underline{Y}_\ell, \quad (4.3.47)$$

$$\text{whence } \hat{\underline{\Phi}}'_\ell(k) = (\underline{Y}'_\ell(k) \underline{Y}_\ell(k))^{-1} \underline{Y}'_\ell(k) \underline{Y}_\ell(k), \quad (4.3.48)$$

$$\text{where } \underline{Y}_\ell(k) = \begin{bmatrix} \underline{v}'_\ell(k) & \dots & \underline{v}'_1(k) \\ \underline{v}'_{\ell+1}(k) & \dots & \underline{v}'_2(k) \\ \vdots & & \vdots \\ \underline{v}'_{n-1}(k) & \dots & \underline{v}'_{n-\ell}(k) \end{bmatrix}, \quad \text{and } \underline{Y}_\ell(k) = \begin{bmatrix} \underline{v}'_{\ell+1}(k) \\ \underline{v}'_{\ell+2}(k) \\ \vdots \\ \underline{v}'_n(k) \end{bmatrix} \quad (4.3.49)$$

Putting $\underline{Z}^{\beta k} = \underline{I}_\ell \otimes \underline{Z}(k)$, this time an $m\ell \times m\ell$ block diagonal matrix, gives

$$\underline{Y}_\ell^{(k)} = \underline{Y}_\ell \underline{Z}_\ell^{\beta k}, \text{ and } \underline{Y}_\ell^{(k)} = \underline{Y}_\ell \underline{Z}'(k). \quad (4.3.50)$$

Substituting into (4.3.48) gives,

$$\hat{\underline{\Phi}}_\ell^{(k)} = (\underline{Z}_\ell^{\beta k} \underline{Y}_\ell' \underline{Y}_\ell \underline{Z}_\ell^{\beta k})^{-1} \underline{Z}_\ell^{\beta k} \underline{Y}_\ell' \underline{Y}_\ell \underline{Z}'(k) = \underline{Z}'^{\beta k} \hat{\underline{\Phi}}_\ell \underline{Z}'(k). \quad (4.3.51)$$

Thus the last $m \times m$ matrix of $\hat{\underline{\Phi}}_\ell^{(k)}$, i.e. $\hat{\underline{F}}^{(k)}(\ell)$ is given by

$$\hat{\underline{F}}^{(k)}(\ell) = \underline{Z}'(k) \hat{\underline{F}}'(\ell) \underline{Z}'(k), \quad \ell=1,2,\dots; k=1,\dots,m,$$

which gives (4.3.38). □

Corollary 4.20

$$\left| \underline{F}^{(k)}(s) \right| = \left| \underline{F}(s) \right|, \text{ and } \left| \hat{\underline{F}}^{(k)}(s) \right| = \left| \hat{\underline{F}}(s) \right|, \quad s=1,2,\dots; k=1,\dots,m.$$

From the results of the above theorem we may construct an element by element description of $\hat{\underline{F}}^{(k)}(s)$ in terms of the elements of $\underline{F}(s)$. This will be analogous to (4.3.12) and (4.3.14). (c.f. (4.3.37) and (4.3.38) with (4.3.5) and (4.3.6).) At the end of this chapter we present an example which incorporates $\hat{\underline{F}}(s)$ and $\hat{\underline{F}}^{(k)}(s)$. We next consider the determinantal criteria given by definition 2.16.

Theorem 4.21

If $\underline{D}(s, \ell)$ is the Box-Tiao determinantal criterion for $\{\underline{v}_t\}$ (definition 2.16), then the corresponding determinantal criterion $\underline{D}^{(k)}(s, \ell)$ for $\{\underline{v}_t^{(k)}\}$, where $\underline{v}_t^{(k)} = \underline{Z}(k) \underline{v}_t$, is

$$\underline{D}^{(k)}(s, \ell) = \underline{Z}(k) \underline{D}(s, \ell) \underline{Z}'(k).$$

Proof

We recall from definition 2.16 that $\underline{D}(s, \ell)$ consists of elements that are determinants of a matrix of the autocovariance function, with some of the rows and columns deleted. That is, from (2.2.12)

$$\underline{D}(s, \ell) = \{d_{ij}(s, \ell)\},$$

where

$$d_{ij}(s, \ell) = \begin{vmatrix} \underline{A}(s, \ell) & \underline{c}_j(s, \ell) \\ \underline{g}'_i(s, \ell) & \gamma_{ij}(s+\ell) \end{vmatrix}$$

$$= |\underline{\delta}_{ij}(s, \ell)| \quad (\text{say}) . \quad (4.3.52)$$

From (4.3.42) and (4.3.43) we have that,

$$\underline{A}^{(k)}(s, \ell) = \underline{Z}^{\alpha k} \underline{A}(s, \ell) \underline{Z}'^{\alpha k} . \quad (4.3.53)$$

Now $c_j^{(k)}(s, \ell)$ is the j^{th} row of $\underline{c}^{(k)}(s, \ell)$; thus from (4.3.45)

$$c_j^{(k)}(s, \ell) = \{ \underline{Z}^{\alpha k} \underline{c}(s, \ell) \underline{Z}'(k) \}_j$$

$$= \underline{Z}^{\alpha k} \{ \underline{c}(s, \ell) \underline{Z}'(k) \}_j \quad ; j, k=1, \dots, m$$

$$; s, \ell=0, 1, 2, \dots \quad (4.3.54)$$

Similarly from the definition of $\underline{g}(s, \ell)$, equation (2.2.5) we have,

$$\underline{g}^{(k)}(s, \ell) = \underline{Z}^{\alpha k} \underline{g}(s, \ell) \underline{Z}'(k) ,$$

whence

$$g'_i{}^{(k)}(s, \ell) = \{ \underline{Z}(k) \underline{g}'(s, \ell) \}_i \underline{Z}'^{\alpha k} \quad ; i, k=1, \dots, m$$

$$; s, \ell=0, 1, 2, \dots \quad (4.3.55)$$

Thus, substituting (4.3.53) through to (4.3.55) into $\delta_{ij}^{(k)}(s, \ell)$ we have,

$$\delta_{ij}^{(k)}(s, \ell) = \begin{bmatrix} \underline{Z}^{\alpha k} \underline{A}(s, \ell) \underline{Z}'^{\alpha k} & \underline{Z}^{\alpha k} \{ \underline{c}(s, \ell) \underline{Z}'(k) \}_j \\ \{ \underline{Z}(k) \underline{g}'(s, \ell) \}_i \underline{Z}'^{\alpha k} & \gamma_{ij}^{(k)}(s+\ell) \end{bmatrix}$$

$$= \underline{Z}^{\alpha k} \begin{bmatrix} \underline{A}(s, \ell) & \{ \underline{c}(s, \ell) \underline{Z}'(k) \}_j \\ \{ \underline{Z}(k) \underline{g}'(s, \ell) \}_i & \gamma_{ij}^{(k)}(s+\ell) \end{bmatrix} \underline{Z}'^{\alpha k} \quad (4.3.56)$$

where $\underline{Z}^{\alpha k} = \begin{bmatrix} \underline{Z}^{\alpha k} & \underline{0} \\ \underline{0}' & 1 \end{bmatrix} . \quad (4.3.57)$

We note that $|\underline{Z}^{\alpha 1k}| = 1$, so that

$$d_{ij}^{(k)}(s, \ell) = \left| \delta_{ij}^{(k)}(s, \ell) \right| = \left| \begin{array}{cc} \underline{A}(s, \ell) & \{\underline{c}(s, \ell)\underline{Z}'(k)\}_j \\ \{Z(k)\underline{g}'(s, \ell)\}_i & y_{ij}^{(k)}(s+\ell) \end{array} \right| . \quad (4.3.58)$$

$$\text{Now } \underline{Z}(k)\underline{g}'(s, \ell) = \begin{bmatrix} 1 & 0 & \dots & -1 & \dots & 0 \\ 0 & 1 & \dots & -1 & \dots & 0 \\ \vdots & & \ddots & & & \\ 0 & 0 & \dots & -1 & \dots & 0 \\ \vdots & & & & \ddots & \\ 0 & 0 & \dots & -1 & \dots & 1 \end{bmatrix} \begin{bmatrix} \underline{g}'_1(s, \ell) \\ \underline{g}'_2(s, \ell) \\ \vdots \\ \underline{g}'_m(s, \ell) \end{bmatrix} .$$

$$\begin{aligned} \text{Hence } \underline{g}'_i^{(k)}(s, \ell) &= \{Z(k)\underline{g}'(s, \ell)\}_i \\ &= \underline{g}'_i(s, \ell) - \underline{g}'_k(s, \ell) \quad , \quad i=1, \dots, m ; i \neq k ; \\ &= -\underline{g}'_k(s, \ell) \quad , \quad i=k. \end{aligned} \quad (4.3.59)$$

Similarly

$$\begin{aligned} \underline{c}_j^{(k)}(s, \ell) &= \underline{c}_j(s, \ell) - \underline{c}_k(s, \ell) \quad , \quad j=1, \dots, m ; j \neq k ; \\ &= -\underline{c}_k(s, \ell) \quad , \quad j=k. \end{aligned} \quad (4.3.60)$$

$y_{ij}^{(k)}(s, \ell)$ is given by (4.3.36), substituting (4.3.36), (4.3.59) and (4.3.60) into (4.3.58) gives,

$$\begin{aligned} d_{ij}^{(k)}(s, \ell) &= \left| \begin{array}{cc} \underline{A}(s, \ell) & \underline{c}_j(s, \ell) - \underline{c}_k(s, \ell) \\ \underline{g}'_i(s, \ell) - \underline{g}'_k(s, \ell) & y_{ij}(s+\ell) - y_{ik}(s+\ell) - y_{kj}(s, \ell) + y_{kk}(s+\ell) \end{array} \right| \\ &= \left| \begin{array}{cc} \underline{A}(s, \ell) & \underline{c}_j(s, \ell) \\ \underline{g}'_i(s, \ell) & y_{ij}(s+\ell) \end{array} \right| - \left| \begin{array}{cc} \underline{A}(s, \ell) & \underline{c}_j(s, \ell) \\ \underline{g}'_k(s, \ell) & y_{kj}(s+\ell) \end{array} \right| \end{aligned}$$

$$\begin{vmatrix} \underline{A}(s, l) & \underline{c}_k(s, l) \\ \underline{g}'_i(s, l) & \gamma_{ik}(s+l) \end{vmatrix} + \begin{vmatrix} \underline{A}(s, l) & \underline{c}_k(s, l) \\ \underline{g}'_k(s, l) & \gamma_{kk}(s+l) \end{vmatrix}$$

$$= d_{ij}(s, l) - d_{kj}(s, l) - d_{ik}(s, l) + d_{kk}(s, l) \quad , \quad \begin{matrix} i, j=1, \dots, m ; \\ i, j \neq k. \end{matrix}$$

$$d_{kj}^{(k)}(s, l) = \begin{vmatrix} \underline{A}(s, l) & \underline{c}_j(s, l) - \underline{c}_k(s, l) \\ -\underline{g}'_k(s, l) & \gamma_{kk}(s+l) - \gamma_{kj}(s+l) \end{vmatrix}$$

$$= \begin{vmatrix} \underline{A}(s, l) & \underline{c}_j(s, l) \\ -\underline{g}'_k(s, l) & -\gamma_{kj}(s+l) \end{vmatrix} + \begin{vmatrix} \underline{A}(s, l) & -\underline{c}_k(s, l) \\ -\underline{g}'_k(s, l) & \gamma_{kk}(s+l) \end{vmatrix}$$

$$= -d_{kj}(s, l) + d_{kk}(s, l) \quad , \quad j=1, \dots, m ; j \neq k .$$

Similarly,

$$d_{ik}^{(k)}(s, l) = d_{kk}(s, l) - d_{ik}(s, l) \quad , \quad i=1, \dots, m ; i \neq k ;$$

and also

$$d_{kk}^{(k)}(s, l) = \begin{vmatrix} \underline{A}(s, l) & -\underline{c}_k(s, l) \\ -\underline{g}'_k(s, l) & \gamma_{kk}(s+l) \end{vmatrix} = d_{kk}(s, l) .$$

Summarizing, we have shown that,

$$\left. \begin{aligned} d_{ij}^{(k)}(s, l) &= d_{ij}(s, l) - d_{ik}(s, l) - d_{kj}(s, l) + d_{kk}(s, l) \quad ; \quad i, j \neq k \\ &= d_{kk}(s, l) - d_{ik}(s, l) \quad ; \quad i \neq k, j = k \\ &= d_{kk}(s, l) - d_{kj}(s, l) \quad ; \quad i = k, j \neq k \\ &= d_{kk}(s, l) \quad ; \quad i = j = k \end{aligned} \right\} \begin{matrix} i=1, \dots, m \\ j=1, \dots, m \end{matrix}$$

$$(4.3.61)$$



Comparing (4.3.61) to (4.3.36), we see that the mapping of $\underline{D}(s, \ell)$ to $\underline{D}^{(k)}(s, \ell)$ is identical to that of $\underline{\Gamma}(r)$ to $\underline{\Gamma}^{(k)}(r)$, given by (i) of lemma 4.17, which is the required result.

□

Finally we examine the likelihood ratio statistic $M(\ell)$ as given by equation (2.2.16).

Lemma 4.22

With obvious notation, $M^{(k)}(\ell) = M(\ell)$, $\ell=0,1,\dots$ $k=1,\dots,m$.

Proof

From (2.2.15) we have,

$$\underline{SS}^{(k)}(\ell) =$$

$$\left[\begin{matrix} y_{\ell}^{(k)} - \underline{x}_{1\ell}^{(k)} \hat{\phi}_{1\ell}^{(k)} - \dots - \underline{x}_{\ell\ell}^{(k)} \hat{\phi}_{\ell\ell}^{(k)} \\ \vdots \\ y_{\ell}^{(k)} - \underline{x}_{1\ell}^{(k)} \hat{\phi}_{1\ell}^{(k)} - \dots - \underline{x}_{\ell\ell}^{(k)} \hat{\phi}_{\ell\ell}^{(k)} \end{matrix} \right] \left[\begin{matrix} y_{\ell}^{(k)} - \underline{x}_{1\ell}^{(k)} \hat{\phi}_{1\ell}^{(k)} - \dots - \underline{x}_{\ell\ell}^{(k)} \hat{\phi}_{\ell\ell}^{(k)} \\ \vdots \\ y_{\ell}^{(k)} - \underline{x}_{1\ell}^{(k)} \hat{\phi}_{1\ell}^{(k)} - \dots - \underline{x}_{\ell\ell}^{(k)} \hat{\phi}_{\ell\ell}^{(k)} \end{matrix} \right] \quad (4.3.62)$$

But from (4.3.50) we have that,

$$\left. \begin{aligned} y_{\ell}^{(k)} &= y_{\ell} Z^{(k)} ; \\ \underline{x}_{i\ell}^{(k)} &= \underline{x}_{i\ell} Z^{(k)} ; \quad i=1,\dots,\ell \end{aligned} \right\} \begin{aligned} &\ell=0,1,\dots \\ &k=1,\dots,m. \end{aligned} \quad (4.3.63)$$

Substituting into (4.3.63) gives,

$$\underline{SS}^{(k)}(\ell) = \underline{Z}^{(k)} \underline{SS}(\ell) \underline{Z}'^{(k)} \quad , \quad k=1,\dots,m ; \ell=0,1,\dots ; \quad (4.3.64)$$

$$\Rightarrow \left| \underline{SS}^{(k)}(\ell) \right| = \left| \underline{SS}(\ell) \right| \quad , \quad k=1,\dots,m ; \ell=0,1,\dots ; \quad (4.3.65)$$

and substituting this into (2.2.16) gives the required result.

□

We summarize the relationships established in this section for the parameters of the model and various identification statistics as follows. (We follow the same notation as above.)

$$i) \quad \underline{v}_t^{(k)} = \underline{Z}^{(k)} \underline{v}_t \quad , \quad \underline{\varepsilon}_t^{(k)} = \underline{Z}^{(k)} \underline{\varepsilon}_t \quad , \quad \underline{u}^{(k)} = \underline{Z}^{(k)} \underline{u} \quad , \quad (t=0,\pm 1,\dots) .$$

- ii) $\underline{Z}^{(k)} = \underline{Z}(k)\underline{Z}'(k)$, $\underline{D}^{(k)}(s, \ell) = \underline{Z}(k)\underline{D}(s, \ell)\underline{Z}'(k)$, $(s, \ell=0, 1, \dots)$.
 $\underline{C}^{(k)}(s) = \underline{Z}(k)\underline{C}(s)\underline{Z}'(k)$, $\underline{\Gamma}^{(k)}(s) = \underline{Z}(k)\underline{\Gamma}(s)\underline{Z}'(k)$, $(s=0, \pm 1, \dots)$,
 $\underline{SS}^{(k)}(\ell) = \underline{Z}(k)\underline{SS}(\ell)\underline{Z}'(k)$, $(\ell=0, 1, \dots)$.
- iii) $\underline{P}^{(k)}(s) = \underline{Y}(k)\underline{P}(s)\underline{Y}'(k)$, $\underline{R}^{(k)}(s) = \underline{Y}(k)\underline{R}(s)\underline{Y}'(k)$, $(s=0, \pm 1, \dots)$.
- iv) $\underline{\Omega}^{(k)}(s) = \underline{Z}(k)\underline{\Omega}(s)\underline{Z}(k)$, $\underline{S}^{(k)}(s) = \underline{Z}(k)\underline{S}(s)\underline{Z}(k)$, $(s=0, \pm 1, \dots)$,
 $\underline{F}^{(k)}(s) = \underline{Z}(k)\underline{F}(s)\underline{Z}(k)$, $\hat{\underline{F}}^{(k)}(s) = \underline{Z}(k)\hat{\underline{F}}(s)\underline{Z}(k)$, $(s=0, 1, \dots)$,
 $\underline{\Phi}_i^{(k)} = \underline{Z}(k)\underline{\Phi}_i\underline{Z}(k)$, $(i=1, \dots, p)$, $\underline{\Theta}_j^{(k)} = \underline{Z}(k)\underline{\Theta}_j\underline{Z}(k)$, $(j=0, \dots, q)$,
- v) $\underline{M}^{(k)}(\ell) = \underline{M}(\ell)$, $\ell=0, 1, \dots$.

Where $k=1, \dots, m$.

In summary, we have seen that the $\ln_+ARMA_m(p, q)$ model is invariant to choice of reference variable. We have considered the mapping of various parameters from one model to another. Further the various identification criteria are invariant, for example

$$\underline{D}(s, \ell) = \underline{0} \Leftrightarrow \underline{D}^{(k)}(s, \ell) = \underline{0} \text{ , } s > p, \text{ and } \ell \geq q \text{ .}$$

However the transformation of the sample autocorrelation and partial autocorrelation functions, may result in slight variations of interpretation since the criteria for elements being zero is that they are "small". Usually "small" refers to being less than 2 standard deviations in modulus. In general, identification of an $ARMA_m(p, q)$ model is a subjective process, and since the exact identification criteria are invariant we would expect little or no variation in the choice of model.

4.4 An Example of Analysis on Two Compositional Time Series

Let us take two data sets and apply the four stages of Box-Jenkins modelling. The two data sets are both political opinion polls, one the GALLUP(c) series already described in example 4.18, the second is a similar one run by the National Opinion Poll (NOP). This consists of

monthly survey results between September 1961 and August 1970 inclusive, and is again on voting intentions with four possible options :-

Conservative, Labour, Liberal, Other,
 CON LAB LIB OTH .

In both data sets we have $CON_t + LAB_t + LIB_t + OTH_t = 100$ (%), $t=1, \dots, 108$. The variable OTH_t often took the value zero, consequently sending the corresponding value of \underline{v}_t to $\pm\infty$. This "zero value problem" was encountered by Aitchison(1982) and is a possible area for further research. Aitchison suggests setting zeros to an extremely low value, which may be determined by the number of decimal places to which the data is recorded. For example if the data is accurate to one decimal place any data less than 0.05 would have been rounded down to 0.00. In this case we might reset 0.0 to, for example 0.025, which lies halfway between the possible range of values for data recorded as zero. In example 4.18 the zeros were reset to 0.005. This method will be reasonable provided there are not too many zeros. For our data sets this was not the case. At this stage we wish to consider reasonable data, so that we examine instead the subcomposition with the variable OTH_t omitted. (We do examine the full data set in chapter 6.)

The variables we use are now CON_t , LAB_t and LIB_t , where now, $CON_t + LAB_t + LIB_t = 100$. Using both data sets in this \mathbb{S}^2 space we compare three possible approaches to modelling them.

i) As three separate univariate time series $ARMA_1(p,q)$,

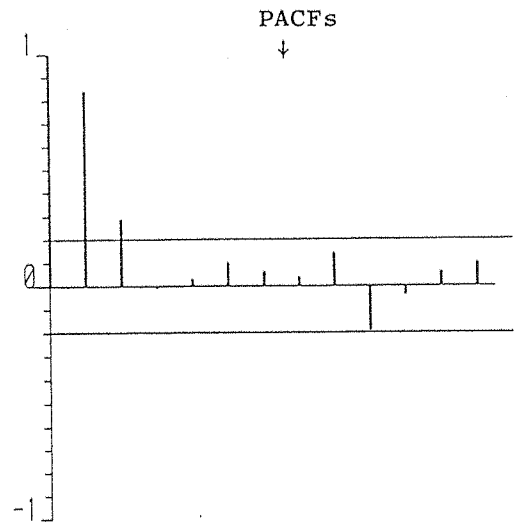
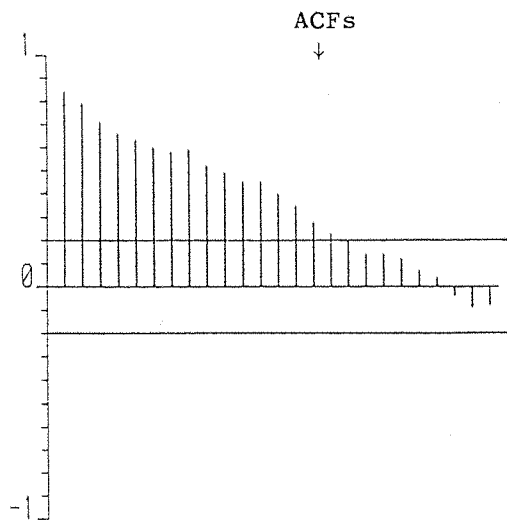
1. CON_t 2. LAB_t 3. LIB_t .

ii) As an $ARMA_2(p,q)$ process by omitting one variable,

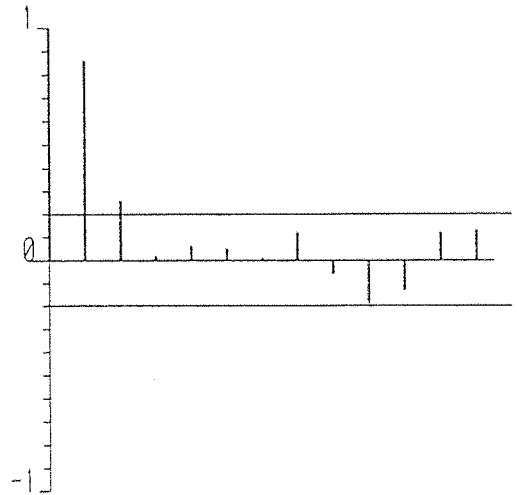
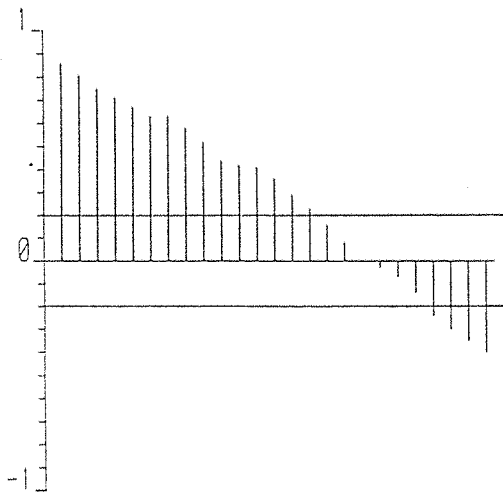
4. $\begin{bmatrix} CON_t \\ LAB_t \end{bmatrix}$ 5. $\begin{bmatrix} CON_t \\ LIB_t \end{bmatrix}$ 6. $\begin{bmatrix} LAB_t \\ LIB_t \end{bmatrix}$.

iii) As an $\ln_+ ARMA_2(p,q)$ process, i.e. as the following $ARMA_2(p,q)$ processes,

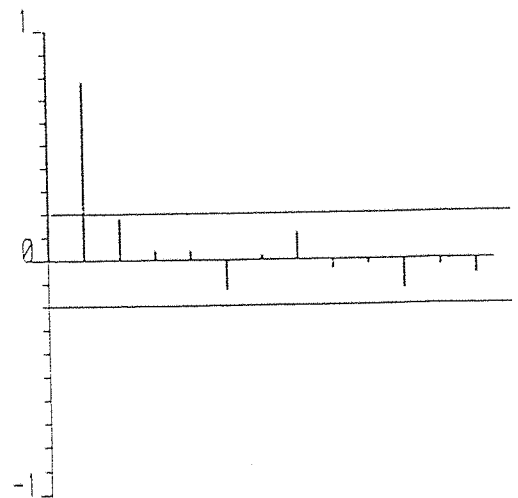
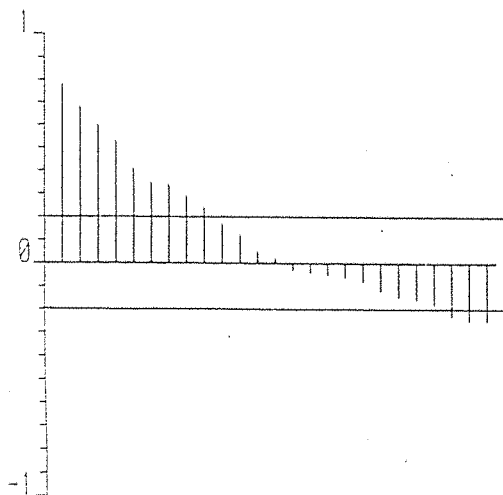
Figure 4.23 Plot of ACF and PACF for GALLUP(C) Series.



a) Conservative



b) Labour



c) Liberal

For $n=108$ (the number of observations in the series),

+ denotes a value greater than $2/\sqrt{n}$,

- denotes a value less than $2/\sqrt{n}$, and

. denotes a non-significant value based on the above criterion.

	+ - .	+ - .	+ - .	+ - .	+ - .	+ - .	+ - .	+ - -
	- + -	- + -	- + -	- + -	- + .	- + .	- + .	- + .
	. - +	. - +	. - +	. - +	. - +	. - +	+ - +	+ - +
Lag	1	2	3	4	5	6	7	8
	+ - -	+ - -	+ - -	+ - -	+ - -	+ . -	+ . -	+ . -
	- + .	- + .	- + .	- + +	- + +	- + +	- + +	- . +
	+ - +	+ - .	+ - .	+ - .	+ - .	+ - .	+ - .	+ - .
Lag	9	10	11	12	13	14	15	16
	+ . -	. . -	. . -	. . -	. + -	. + -	. + -	. + -
	- . +	- . +	- . +	- . +	. . +	. - +	. - +	. - +
	+ - .	+ . .	+ . .	+ -	. . -
Lag	17	18	19	20	21	22	23	24
	. + -	. + -	. + -	. + -				
	. - +	. - +	. - +	. - +				
	. . -	. . -	. . -	. . -				
Lag	25	26	27	28				

Figure 4.24 : Summary of Cross-correlations for the GALLUP(c) Poll Data for the Variables CON, LAB and LIB.

Key:-

+ denotes a value greater than 2 standard errors,

- denotes a value less than 2 standard errors, and

. denotes a non-significant value base on the above criterion.

Series

↓

(CON,LAB)'	+ .	+ -
	. +	+ +
(CON,LIB)'	+ .	+ +	. .	- .	. .
	. +	. +	- .	. .	- .
(LAB,LIB)'	+ .	+ -	. .	- .	. .
	. +	. +	. +	+ .	. .	+ .
Lag	1	2	3	4	5	6	7	8	9	10	

Figure 4.25 : Schematic Representation of the Partial Autocorrelation Function for Series 4, 5, and 6 of the GALLUP(c) Poll Data.

For key to the meaning of +, - and . see Figure 4.24.

a) Cross-correlations for series 7: $(\ln(\text{CON}/\text{LIB})_t, \ln(\text{LAB}/\text{LIB})_t)'$.

	++	++	++	++	++	++	++	++	++	++	+.	+.
	++	++	++	++	++	+.	+.	+.	+.	+.	+.	..
Lag 1	2	3	4	5	6	7	8	9	10	11	12	
	+.
	-.	-.	-.	-.	-.	-.	-.	--	--	--	--	--
	13	14	15	16	17	18	19	20	21	22	23	24

b) Cross-correlations for series 8: $(\ln(\text{CON}/\text{LAB})_t, \ln(\text{LIB}/\text{LAB})_t)'$.

	++	++	++	++	++	++	++	+.	+.	+.	+.	+.
	++	++	++	++	++	++	++	++	++	++	++	+.
Lag 1	2	3	4	5	6	7	8	9	10	11	12	
	+.	+.	+.	+ -	. -	. -	. -	. -	. -	. -	. -	--
	+.	+.	+.	+.	+.	+.	+.	+ -	. -	. -	. -	. -
	13	14	15	16	17	18	19	20	21	22	23	24

c) Cross-correlations for series 9: $(\ln(\text{LAB}/\text{CON})_t, \ln(\text{LIB}/\text{CON})_t)'$.

	+.	+.	+.	+.	+.	++	++	++	++	++	++	++
	+.	+.	+.	+.	+.	+.	+.	+.	+.	-.	-.	-.
Lag 1	2	3	4	5	6	7	8	9	10	11	12	
	++	++	++	++	+.	+.	+.	+.	+.	+.	+.	-.
	-.	-.	-.	-.	-.	-.	-.	-.	-.
	13	14	15	16	17	18	19	20	21	22	23	24

Figure 4.26 : Summary of the Autocorrelation Function for Series 7, 8, and 9 of the GALLUP(c) Poll Data.

For key to the meanings of +,- and . see Figure 4.25, and for details of series see Figure 4.26.

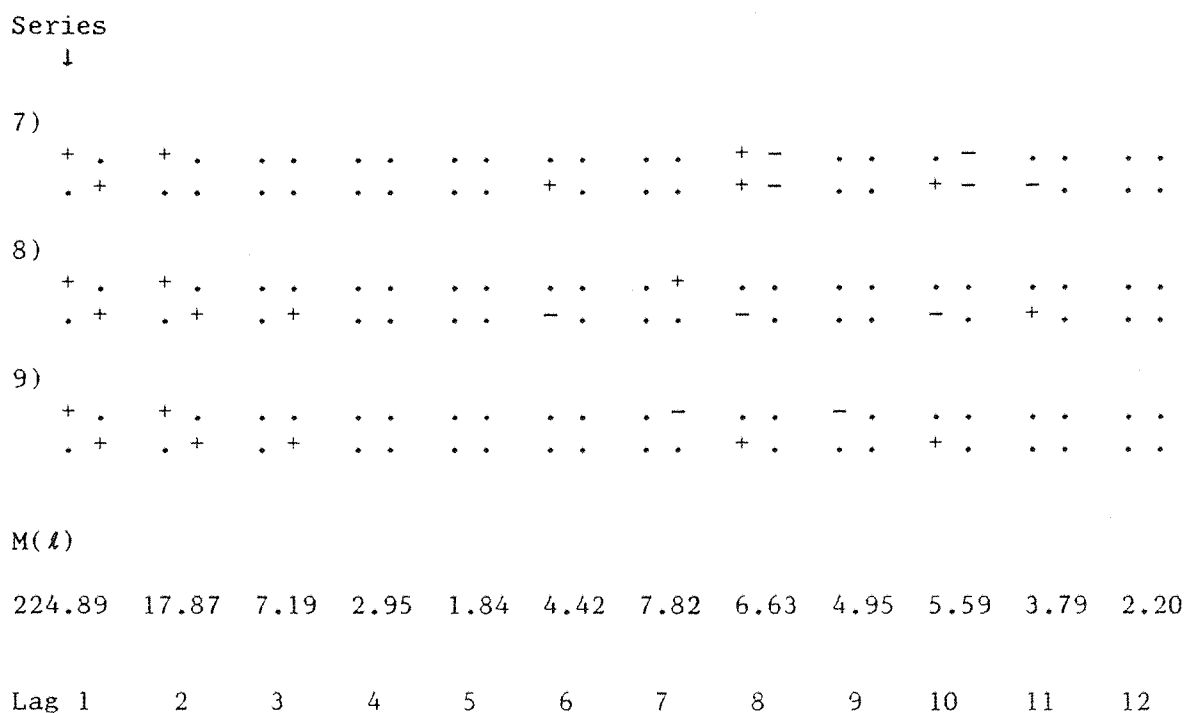


Figure 4.27 : Schematic representation of the Partial Autocorrelation Function, and values of the $M(l)$ statistic (Equation (2.2.16)), for Series 7, 8, and 9 of the GALLUP(c) Poll Data.

$$7. \begin{pmatrix} \ln(\text{CON}/\text{LIB})_t \\ \ln(\text{LAB}/\text{LIB})_t \end{pmatrix} \quad 8. \begin{pmatrix} \ln(\text{CON}/\text{LAB})_t \\ \ln(\text{LIB}/\text{LAB})_t \end{pmatrix} \quad 9. \begin{pmatrix} \ln(\text{LAB}/\text{CON})_t \\ \ln(\text{LIB}/\text{CON})_t \end{pmatrix} .$$

We would expect models 4,5, and 6 to produce similar results because of the sum-constraint. For example model 5 could be rewritten as,

$$\begin{pmatrix} \text{CON}_t \\ 100-\text{CON}_t-\text{LAB}_t \end{pmatrix} , \text{ that is as a linear transform of 4.}$$

From the previous section we expect models 7-9 to give identical results.

The nine models enable us to compare the $\ln_+ \text{ARMA}_m(p,q)$ approach to that of a linear approach which ignores the linear constraint on the variables. It also serves to numerically enhance the theory of the previous section.

The models above were fitted using two statistical packages, the Economic and Statistical Package (ESP) for the univariate analysis, and the Wisconsin Multiple Time Series package (WMTS-1) for the multivariate analysis. Since the data represent the same phenomenon, we present only the results for the GALLUP poll at the identification stage, since the analogous results for the NOP data were virtually identical. Figures 4.23 through to 4.27 give the ACF and PACF for all nine series, the univariate series are presented in graphical form, whilst the multivariate series are in symbolic form using the "+", "-", "." notation described in chapter 2. Figure 4.23 contains all the univariate results. Figure 4.24 is a summary of the cross-correlation matrices for the series $(\text{CON}, \text{LAB}, \text{LIB})'_t$, from which we can obtain the appropriate cross-correlations of our series in (ii) above by taking the required subset of these matrices. The diagonal elements correspond to the ACF of the univariate series given in Figure 4.23. It is necessary to compute the PACFs for the series in (ii) separately. This is for two reasons. Firstly it is not possible to compute the PACF for the tri-variate series because the sum-constraint will cause the $\underline{Y}'_t \underline{Y}_t$ matrix of equation (2.2.10) to be singular. Consequently it can not be inverted. Secondly the PACF is by definition a parameter estimate under the assumption that the data follow an AR model. Extra variables alter the parameters, so that the sub-matrices would not correspond to the PACFs

Table 4.28 : Parameter Estimates for the Univariate Models.

Data Set	Var.	Model	Parameter Estimates	t-test for Parameter Ests.		Portmanteau Test	
				Value	dof.	Value	dof.
G A L L U P	CON 1	AR(1)	$\hat{\phi} = 0.8529$	16.67 ^{***}	106	22.5 ^{**}	10
			Const. = 6.4912	2.82 ^{***}	106	34.8 ^{**}	22
	LAB 2		$\hat{\phi} = 0.8709$	18.40 ^{***}	106	17.3 [*]	10
			Const. = 5.7220	2.69 ^{***}	106	26.2	22
	LIB 3		$\hat{\phi} = 0.8149$	13.76 ^{***}	106	11.1	10
			Const. = 2.1300	3.05 ^{***}	106	13.1	22
N O P	CON 1	$\hat{\phi} = 0.9147$	24.17 ^{***}	106	19.6 ^{**}	10	
		Const. = 3.7517	2.23 ^{**}	106	27.0	22	
	LAB 2	$\hat{\phi} = 0.8940$	21.3 ^{***}	106	26.8 ^{***}	10	
		Const. = 4.6799	2.49 ^{**}	106	36.3 ^{**}	22	
	LIB 3	$\hat{\phi} = 0.9182$	23.54 ^{***}	106	26.5 ^{***}	10	
		Const. = 0.9720	1.85	106	28.1	22	
G A L L U P	CON 1	IMA(1,1)	$\hat{\theta} = 0.4351$	4.99 ^{***}	106	12.0	11
						26.3	23
	LAB 2		$\hat{\theta} = 0.3519$	3.88 ^{***}	106	13.2	11
						26.8	23
	LIB 3		$\hat{\theta} = 0.5016$	5.93 ^{***}	106	3.9	11
						6.2	23
N O P	CON 1	$\hat{\theta} = 0.3127$	3.40 ^{***}	106	11.3	11	
					16.7	23	
	LAB 2	$\hat{\theta} = 0.3096$	3.36 ^{***}	106	14.4	11	
					21.6	23	
	LIB 3	$\hat{\theta} = 0.3344$	3.69 ^{***}	106	14.1	11	
					16.0	23	

Table 4.28 ...continued.

Data Set	Var.	Model	Parameter Estimates	t-test for Parameter Ests.		Portmanteau Test	
				Value	dof.	Value	dof.
G A L L U P	CON 1	ARMA(1,1)	$\hat{\phi} = 0.9500$	31.63 ^{***}	105	11.5 24.8	9 21
			$\hat{\theta} = 0.3792$	3.81 ^{***}	105		
			Const = 2.1671	1.58	105		
	LAB 1		$\hat{\phi} = 0.9470$	30.53 ^{***}	105	12.8 24.4	9 21
			$\hat{\theta} = 0.3160$	3.14 ^{***}	105		
			Const = 2.3500	1.64	105		
LIB 1	$\hat{\phi} = 0.9739$	42.66 ^{***}	105	4.1 6.2	9 21		
	$\hat{\theta} = 0.4661$	4.99 ^{***}	105				
	Const = 0.3395	1.56	105				
N O P	CON 1	$\hat{\phi} = 0.9594$	45.94 ^{***}	105	11.2 16.0	9 21	
		$\hat{\theta} = 0.2826$	2.93 ^{***}	105			
		Const = 1.7944	1.87	105			
	LAB 1	$\hat{\phi} = 0.9489$	34.02 ^{***}	105	15.3 [*] 22.4	9 21	
		$\hat{\theta} = 0.2754$	2.77 ^{***}	105			
		Const = 2.3365	1.85	105			
LIB 1	$\hat{\phi} = 0.9780$	44.94 ^{***}	105	14.7 [*] 16.7	9 21		
	$\hat{\theta} = 0.3208$	33.14 ^{***}	105				
	Const = 0.2582	0.77	105				

Key:- * - 10%
 ** - 5%
 *** - 1% } Significance Level.

Table 4.29 : Parameter Estimates for the Multivariate Models.

Data Set	Variables in the model	Model	Parameter Estimates
GALLUP	4. (CON LAB)	ARMA 2 (1 , 1)	Const = $\begin{pmatrix} 1.71 \\ -6.48 \end{pmatrix}$ $\hat{\Phi} = \begin{pmatrix} 0.96^{***} & 0.00 \\ 0.12 & 1.03^{***} \end{pmatrix}$ $\hat{\Theta} = \begin{pmatrix} 0.56^{***} & 0.18 \\ 0.21 & 0.50^{***} \end{pmatrix}$
			Const = $\begin{pmatrix} 1.76 \\ 1.88^{**} \end{pmatrix}$ $\hat{\Phi} = \begin{pmatrix} 0.96^{***} & 0.00 \\ -0.05^{***} & 1.03^{***} \end{pmatrix}$ $\hat{\Theta} = \begin{pmatrix} 0.38^{***} & -0.17 \\ -0.08 & 0.69^{***} \end{pmatrix}$
	6. (LAB LIB)		Const = $\begin{pmatrix} 5.15^* \\ -2.79^{***} \end{pmatrix}$ $\hat{\Phi} = \begin{pmatrix} 0.91^{***} & -0.12 \\ 0.05^{***} & 1.08^{***} \end{pmatrix}$ $\hat{\Theta} = \begin{pmatrix} 0.30^{***} & -0.22 \\ 0.08 & 0.77^{***} \end{pmatrix}$
			Const = $\begin{pmatrix} -0.12 \\ 3.45 \end{pmatrix}$ $\hat{\Phi} = \begin{pmatrix} 0.98^{***} & 0.03 \\ -0.01 & 0.94^{***} \end{pmatrix}$ $\hat{\Theta} = \begin{pmatrix} 0.23^* & -0.09 \\ 0.14 & 0.37^{***} \end{pmatrix}$
	5. (CON LIB)		Const = $\begin{pmatrix} 2.56 \\ 0.31 \end{pmatrix}$ $\hat{\Phi} = \begin{pmatrix} 0.95^{***} & -0.03 \\ 0.00 & 0.96^{***} \end{pmatrix}$ $\hat{\Theta} = \begin{pmatrix} 0.32^{***} & 0.09 \\ -0.09 & 0.28^{***} \end{pmatrix}$
			6. (LAB LIB)
GALLUP	4. (CON LAB)	IMA 2 (1 , 1)	$\hat{\Theta} = \begin{pmatrix} 0.57^{***} & 0.17 \\ -0.03 & 0.32^{**} \end{pmatrix}$
			$\hat{\Theta} = \begin{pmatrix} 0.40^{***} & -0.17 \\ -0.04 & 0.49^{**} \end{pmatrix}$
	6. (LAB LIB)		$\hat{\Theta} = \begin{pmatrix} 0.36^{***} & 0.03 \\ 0.04 & 0.53^{***} \end{pmatrix}$
			$\hat{\Theta} = \begin{pmatrix} 0.23^* & -0.12 \\ 0.16 & 0.41^{***} \end{pmatrix}$
	5. (CON LIB)		$\hat{\Theta} = \begin{pmatrix} 0.35^{***} & 0.12 \\ -0.09 & 0.29^{***} \end{pmatrix}$
			6. (LAB LIB)

Table 4.29 ...continued.

Data Set	Variables in the model	Model ↓	Parameter Estimates
G A L L U P	7 { ln(LAB/CON) ln(LIB/CON)	A R M A 2 (1 , 1)	Const = $\begin{bmatrix} -0.04 \\ 0.05 \end{bmatrix}$ $\hat{\Phi} = \begin{bmatrix} 0.95^{***} & -0.03 \\ 0.06 & 1.02^{***} \end{bmatrix}$ $\hat{\Theta} = \begin{bmatrix} 0.33^{***} & 0.02 \\ 0.06 & 0.70^{***} \end{bmatrix}$
	8 { ln(CON/LAB) ln(LIB/LAB)		Const = $\begin{bmatrix} 0.04 \\ 0.08 \end{bmatrix}$ $\hat{\Phi} = \begin{bmatrix} 0.92^{***} & 0.03 \\ -0.16^{***} & 1.05^{***} \end{bmatrix}$ $\hat{\Theta} = \begin{bmatrix} 0.34^{***} & -0.02 \\ -0.41^{***} & 0.68^{***} \end{bmatrix}$
	9 { ln(CON/LIB) ln(LAB/LIB)		Const = $\begin{bmatrix} -0.04 \\ -0.08 \end{bmatrix}$ $\hat{\Phi} = \begin{bmatrix} 1.08^{***} & -0.06^* \\ 0.16^{***} & 0.89^{***} \end{bmatrix}$ $\hat{\Theta} = \begin{bmatrix} 0.75^{***} & -0.05 \\ 0.41^{***} & 0.27^* \end{bmatrix}$
N O P	7 { ln(LAB/CON) ln(LIB/CON)	I M A 2 (1 , 1)	Const = $\begin{bmatrix} 0.01 \\ -0.06 \end{bmatrix}$ $\hat{\Phi} = \begin{bmatrix} 0.95^{***} & 0.01 \\ 0.00 & 0.96^{***} \end{bmatrix}$ $\hat{\Theta} = \begin{bmatrix} 0.28^{***} & -0.02 \\ 0.10 & 0.35^{***} \end{bmatrix}$
	8 { ln(CON/LAB) ln(LIB/LAB)		Const = $\begin{bmatrix} -0.01 \\ -0.07 \end{bmatrix}$ $\hat{\Phi} = \begin{bmatrix} 0.95^{***} & -0.01 \\ -0.01 & 0.96^{***} \end{bmatrix}$ $\hat{\Theta} = \begin{bmatrix} 0.26^{**} & 0.02 \\ -0.20 & 0.37^{***} \end{bmatrix}$
	9 { ln(CON/LIB) ln(LAB/LIB)		Const = $\begin{bmatrix} 0.06 \\ 0.07 \end{bmatrix}$ $\hat{\Phi} = \begin{bmatrix} 0.97^{***} & -0.00 \\ 0.01 & 0.94^{***} \end{bmatrix}$ $\hat{\Theta} = \begin{bmatrix} 0.45^{**} & -0.10 \\ 0.20 & 0.20 \end{bmatrix}$
G A L L U P	7 { ln(LAB/CON) ln(LIB/CON)	I M A 2 (1 , 1)	$\hat{\Theta} = \begin{bmatrix} 0.34^{***} & 0.07 \\ 0.05 & 0.56^{***} \end{bmatrix}$
	8 { ln(CON/LAB) ln(LIB/LAB)		$\hat{\Theta} = \begin{bmatrix} 0.41^{***} & -0.07 \\ -0.20 & 0.49^{**} \end{bmatrix}$
	9 { ln(CON/LIB) ln(LAB/LIB)		$\hat{\Theta} = \begin{bmatrix} 0.61^{***} & -0.05 \\ 0.20 & 0.29^* \end{bmatrix}$
N O P	7 { ln(LAB/CON) ln(LIB/CON)	I M A 2 (1 , 1)	$\hat{\Theta} = \begin{bmatrix} 0.31^{***} & -0.02 \\ 0.11 & 0.37^{***} \end{bmatrix}$
	8 { ln(CON/LAB) ln(LIB/LAB)		$\hat{\Theta} = \begin{bmatrix} 0.29^{***} & 0.03 \\ -0.20 & 0.39^{***} \end{bmatrix}$
	9 { ln(CON/LIB) ln(LAB/LIB)		$\hat{\Theta} = \begin{bmatrix} 0.47^{***} & -0.11 \\ 0.19 & 0.20^{***} \end{bmatrix}$

Figure 4.30 : Summary of Cross-correlation Matrices for Residual Series.

1. Residuals from $AR_2(2)$ fitted to a) series 4.(CON/LAB)' and b) series 9. ($\ln(\text{CON/LIB}) \ln(\text{LAB/LIB})$)'.

a)	+.
	1	2	3	4	5	6	7	-

	13	14	15	16	17	18	19

b)
	1	2	3	4	5	6	7

	13	14	15	16	17	18	19

1. Residuals of $ARMA_2(1,1)$ a) series 4. b) series 9.

a)	+.
	1	2	3	4	5	6	7	-

	13	14	15	16	17	18	19

b)
	1	2	3	4	5	6	7

	13	14	15	16	17	18	19

1. Residuals of $IMA_2(1,1)$ a) series 4. b) series 9.

a)	+.
	1	2	3	4	5	6	7	-

	13	14	15	16	17	18	19

b)
	1	2	3	4	5	6	7

	13	14	15	16	17	18	19

As expected the $AR_i(1)$ ($i=1,2$) models were insufficient to describe the data. For example, in table 4.28, the portmanteau statistic was often significant, although for the LIB series of the GALLUP(c) poll, the fit seems reasonably good. For the IMA(1,1) and ARMA(1,1) models the portmanteau statistic is vastly reduced across all the series including the afore mentioned LIB series. This is so even accounting for degrees of freedom, as can be seen by the reduced number of significant results. For these univariate series and for the bivariate series there seems to be little difference between the IMA_i(1,1) and the ARMA_i(1,1) models, all the estimates of $\underline{\phi}$ in the ARMA_i(1,1) are close to \underline{I}_i ($i=1,2$). The $AR_i(2)$ process was also found to give similarly good results. Of all these three models, the ARMA_i(1,1) model seems to do the best, although its improvement over the other two models is often minimal. For example for series 4 in figure 4.30, the $AR_2(2)$ and IMA₂(1,1) model have respectively 10 and 5 significant residual cross-correlations, whilst the ARMA₂(1,1) model has only 3.

Our data sets have also been examined by Scott et al (1977), who based on the structure of the survey from which the data comes, in particular the overlapping nature of the samples, produce a strong intuitive argument for using the mixed ARMA(1,1) model. Consequently we select this model.

Table 4.31 gives the forecasts one to twenty steps ahead by each method. The univariate forecasts have been normalized so that their sum is 100. The forecast of the omitted variable in series 4. to 6. is obtained as the difference from 100 of the two included variables. Finally for the transformed series, we have taken the inverse (i.e. a_m^{-1}) of the forecasts, however as described in section 2.5 this will not necessarily give us the minimum mean square error forecast. This problem is addressed in chapter 5. We note also that we could have normalized the univariate forecasts at each step, and used the normalized forecasts to produce the next-step ahead forecast. However ESP or indeed any similar computer package does not easily facilitate this, and although a program in e.g. FORTRAN would easily accomplish this, it is evident in our current example that this would add very little to the results.

If we examine the forecasts themselves we see that the one-step ahead forecasts via each method are very similar, but the similarity

Table 4.31 (a) : GALLUP(c) Forecasts

Lag	Variable	Reference Variable of $\ln ARMA_2(1,1)$			Omitted Variable of $ARMA_2(1,1)$			Combined Univariate Forecast	Actual (where known)
		Con.	Lab.	Lib.	Con.	Lab.	Lib.		
1	CON	36.141	36.198	36.211	36.225	36.226	36.116	36.922	
	LAB	40.043	40.119	40.167	39.827	39.784	39.985	42.368	
	LIB	23.816	23.683	23.621	23.949	23.990	23.899	26.710	
2	CON	36.268	36.348	36.366	36.546	36.549	36.401	37.155	
	LAB	38.838	38.944	39.015	38.599	38.543	38.806	42.374	
	LIB	24.894	24.708	24.618	24.854	24.909	24.793	20.471	
3	CON	36.364	36.470	36.494	36.857	36.861	36.674	37.378	
	LAB	37.664	37.803	37.898	37.372	37.299	37.626	42.382	
	LIB	25.972	25.727	25.608	25.771	25.841	25.700	20.240	
4	CON	36.431	36.565	36.595	37.157	37.162	36.936	37.591	
	LAB	36.521	36.696	36.815	36.142	36.052	36.445	42.391	
	LIB	27.048	26.739	26.589	26.701	26.787	26.619	20.018	
5	CON	36.471	36.635	36.673	37.447	37.453	37.186	37.795	
	LAB	35.410	35.625	35.768	34.908	34.800	35.261	42.401	
	LIB	28.119	27.741	27.559	27.645	27.748	27.553	19.804	
6	CON	36.485	36.682	36.728	37.726	37.734	37.426	37.990	
	LAB	34.332	34.588	34.757	33.670	33.541	34.072	42.413	
	LIB	29.183	28.731	28.515	28.604	28.725	28.502	19.598	
7	CON	36.474	36.707	36.763	37.997	38.006	37.656	38.176	
	LAB	33.287	33.586	33.780	32.426	32.275	32.878	42.425	
	LIB	30.239	29.706	29.457	29.578	29.719	29.467	19.399	
8	CON	36.441	36.713	36.779	38.258	38.269	37.876	38.354	
	LAB	32.275	32.620	32.839	31.174	31.001	31.676	42.439	
	LIB	31.284	30.667	30.383	30.569	30.731	30.448	19.207	
9	CON	36.388	36.701	36.777	38.511	38.524	38.086	38.524	
	LAB	31.296	31.688	31.932	29.913	29.715	30.466	42.454	
	LIB	32.316	31.610	31.291	31.577	31.761	31.448	19.022	
10	CON	36.315	36.673	36.760	38.755	38.770	38.288	38.688	
	LAB	30.350	30.791	31.060	28.642	28.419	29.246	42.469	
	LIB	33.335	32.536	32.180	32.603	32.811	32.466	18.843	
11	CON	36.224	36.630	36.729	38.992	39.009	38.481	38.844	36.5
	LAB	29.437	29.929	30.221	27.360	27.109	28.015	42.485	38.5
	LIB	34.339	33.441	33.050	33.649	33.883	33.504	18.671	15.0
12	CON	36.118	36.574	36.685	39.221	39.240	38.665	38.994	34.0
	LAB	28.556	29.099	29.415	26.064	25.784	26.772	42.502	39.0
	LIB	35.325	34.327	33.900	34.714	34.975	34.563	18.504	17.0
13	CON	35.998	36.507	36.630	39.444	39.465	38.842	39.138	35.5
	LAB	27.708	28.303	28.642	24.755	24.444	25.515	42.519	51.0
	LIB	36.295	35.191	34.728	35.801	36.091	35.644	18.343	13.5
14	CON	35.864	36.429	36.566	39.659	39.683	39.011	39.275	46.5
	LAB	26.891	27.539	27.900	23.431	23.087	24.242	42.538	42.0
	LIB	37.245	36.032	35.534	36.910	37.231	36.747	18.187	11.5
15	CON	35.720	36.342	36.493	39.868	39.895	39.172	39.408	43.0
	LAB	26.104	26.806	27.190	22.089	21.711	22.954	42.556	45.5
	LIB	38.176	36.851	36.318	38.043	38.395	37.874	18.036	11.5
16	CON	35.565	36.248	36.412	40.071	40.100	39.326	39.534	44.0
	LAB	25.348	26.105	26.509	20.730	20.315	21.648	42.575	46.0
	LIB	39.087	37.647	37.079	39.199	39.585	39.026	17.891	10.0
17	CON	35.401	36.148	36.326	40.268	40.300	39.474	39.656	47.5
	LAB	24.622	25.433	25.857	19.352	18.898	20.322	42.594	41.0
	LIB	39.977	38.420	37.817	40.380	40.802	40.204	17.750	11.5
18	CON	35.231	36.042	36.235	41.460	41.495	39.615	39.773	
	LAB	23.924	24.789	25.233	17.953	17.458	18.977	42.614	
	LIB	40.845	39.168	38.532	41.588	42.048	41.408	17.613	
19	CON	35.054	35.933	36.139	40.646	40.684	39.750	39.885	
	LAB	23.255	24.175	24.637	16.532	15.994	17.610	42.634	
	LIB	41.692	39.893	39.224	42.822	43.322	42.641	17.481	
20	CON	34.871	35.820	36.041	40.827	40.869	39.878	39.993	
	LAB	22.613	23.587	24.067	15.088	14.504	16.219	42.654	
	LIB	42.516	40.593	39.892	44.085	44.627	43.903	17.353	

Table 4.31 (b) : NOP Forecasts

Variable	Lag	Reference Variable of $\epsilon_{n+1} \text{ARMA}_2(1,1)$			Omitted Variable of $\text{ARMA}_2(1,1)$			Combined Univariate Forecast	Actual (where known)
		Con.	Lab.	Lib.	Con.	Lab.	Lib.		
1	CON	48.639	48.642	48.646	48.626	48.625	48.633	47.383	
	LAB	44.362	44.359	44.363	44.337	44.339	44.332	45.408	
	LIB	6.998	6.999	6.991	7.037	7.036	7.035	7.209	
2	CON	48.607	48.612	48.618	48.571	48.570	48.583	46.469	46.2
	LAB	44.324	44.318	44.325	44.297	44.299	44.286	46.102	47.0
	LIB	7.069	7.070	7.057	7.133	7.131	7.131	7.429	6.8
3	CON	48.574	48.580	48.588	48.516	48.515	48.533	45.564	
	LAB	44.288	44.281	44.290	44.259	44.263	44.245	46.787	
	LIB	7.138	7.139	7.122	7.225	7.222	7.223	7.650	
4	CON	48.539	48.548	48.557	48.462	48.460	48.482	44.667	
	LAB	44.256	44.247	44.259	44.224	44.230	44.206	47.464	
	LIB	7.205	7.205	7.184	7.314	7.310	7.312	7.869	
5	CON	48.503	48.514	48.524	48.408	48.406	48.432	43.781	46.8
	LAB	44.228	44.217	44.231	44.192	44.199	44.171	48.131	44.8
	LIB	7.269	7.270	7.245	7.420	7.395	7.397	8.088	8.4
6	CON	48.467	48.479	48.491	48.354	48.352	48.382	42.905	
	LAB	44.202	44.189	44.205	44.163	44.171	44.138	48.789	
	LIB	7.331	7.332	7.304	7.483	7.477	7.480	8.306	
7	CON	48.430	48.443	48.458	48.301	48.299	48.333	42.041	40.2
	LAB	44.179	44.165	44.182	44.137	44.146	44.109	49.437	50.8
	LIB	7.392	7.392	7.360	7.562	7.555	7.559	8.522	9.0
8	CON	48.392	48.406	48.423	48.249	48.247	48.283	41.188	42.8
	LAB	44.158	44.143	44.161	44.112	44.122	44.081	50.075	49.8
	LIB	7.450	7.451	7.415	7.639	7.631	7.635	8.737	7.4
9	CON	48.354	48.370	48.388	48.198	48.195	48.235	40.347	
	LAB	44.140	44.123	44.143	44.090	44.101	44.057	50.702	
	LIB	7.506	7.507	7.468	7.712	7.704	7.709	8.950	
10	CON	48.316	48.332	48.353	48.147	48.144	48.186	39.519	39.0
	LAB	44.124	44.106	44.127	44.070	44.082	44.034	51.319	53.0
	LIB	7.561	7.562	7.520	7.783	7.774	7.780	9.162	8.0
11	CON	48.277	48.295	48.317	48.097	48.094	48.138	38.705	38.3
	LAB	44.109	44.091	44.113	44.052	44.064	44.014	51.923	54.0
	LIB	7.613	7.614	7.570	7.852	7.842	7.848	9.372	7.7
12	CON	48.239	48.257	48.282	48.047	48.045	48.091	37.904	
	LAB	44.097	44.078	44.100	44.035	44.048	43.995	52.517	
	LIB	7.664	7.665	7.618	7.918	7.907	7.914	9.580	
13	CON	48.200	48.219	48.246	47.999	47.996	48.044	37.116	41.8
	LAB	44.087	44.066	44.090	44.020	44.034	43.979	53.098	50.2
	LIB	7.713	7.714	7.665	7.981	7.970	7.977	9.785	8.0
14	CON	48.162	48.182	48.210	47.951	47.948	47.998	36.343	
	LAB	44.077	44.057	44.080	44.007	44.021	43.964	53.668	
	LIB	7.761	7.762	7.710	8.042	8.030	8.038	9.989	
15	CON	48.124	48.144	48.174	47.904	47.901	47.953	35.585	39.2
	LAB	44.070	44.049	44.072	43.995	44.010	43.951	54.225	52.3
	LIB	7.807	7.807	7.753	8.101	8.089	8.096	10.190	8.5
16	CON	48.085	48.107	48.139	47.858	47.855	47.908	34.842	
	LAB	44.064	44.042	44.066	43.985	44.000	43.939	54.770	
	LIB	7.851	7.852	7.795	8.158	8.145	8.153	10.388	
17	CON	48.048	48.069	48.103	47.813	47.810	47.865	34.113	43.2
	LAB	44.059	44.036	44.060	43.975	43.991	43.928	55.303	49.9
	LIB	7.893	7.894	7.836	8.212	8.199	8.207	10.584	6.9
18	CON	48.010	48.032	48.068	47.768	47.766	47.821	33.400	
	LAB	44.055	44.032	44.056	43.967	43.983	43.919	55.823	
	LIB	7.935	7.936	7.876	8.265	8.251	8.260	10.777	
19	CON	47.973	47.996	48.033	47.725	47.723	47.779	32.702	42.4
	LAB	44.052	44.029	44.053	43.960	43.977	43.911	56.331	49.5
	LIB	7.975	7.975	7.914	8.315	8.301	8.310	10.967	8.1
20	CON	47.937	47.960	47.999	47.682	47.680	47.737	32.019	45.2
	LAB	44.050	44.027	44.051	43.954	43.971	43.904	56.826	47.2
	LIB	8.013	8.014	7.951	8.364	8.349	8.359	11.155	7.6

decreases the further ahead we predict. The forecasts obtained from series 4. to 6. are alike, as are those for the series 7. to 9., although this is more so for the former series than the latter. The difference within each type of the multivariate methods is due to rounding errors in the estimation, forecasting etc. of the models. Some of the extraneous difference between the series based on the $\ln_+ARMA_m(p,q)$ model, is due to the additional rounding error in the a_m transformation and its inverse.

Where known the actual values have been written in. Comparing these to the forecasts it seems that especially for the GALLUP(c) series the univariate forecasts do the best. For the NOP data the multivariate methods give the better results at lag 5 and at lag 20. At lag 20 the univariate forecast preserves the correct status within the data (i.e. LAB > CON > LIB), but the multivariate forecasts are numerically closer. The fact that the univariate method does so well should not disappoint us too much as it has often been found that multivariate ARMA models perform badly. Further the WMTS-1 package is an early prototype of a multiple time series package which proved to have many bugs. The multivariate estimates themselves have an almost diagonal structure to them indicating that the univariate model may be more appropriate to these data. In order to come to a real understanding of the above techniques many more data sets need to be similarly analyzed. It may be possible to improve on the $\ln_+ARMA_m(p,q)$ forecasts by developing a less crude way of taking the inverse of the \underline{v}_t series.

The two multivariate methods give similar results. This is indicative of the fact that the logistic transformation is nearly linear for the range of values taken by our data sets. This fact has kept the linear forecasts within the correct range (c.f. example 3.16). It would perhaps be useful to work with some series that have greater variation, and/or have values closer to 0.0 or 1.0.

Although the linear methods used here do not produce forecasts outside the range they may still give a confidence interval that does fall outside. A meaningful confidence interval for the forecasts may be easily developed within the $\ln_+ARMA_m(p,q)$ class of model. This is also developed in the next chapter.

What this section has perhaps confirmed is that the $\ln_+ARMA_m(p,q)$ is invariant to the choice of reference variable, apart from rounding error. It also serves to point out the need of the discussion in the

next chapter, and provides an example of the possible use of an $\ln_+ARMA_m(p,q)$ process.

To see the invariance property exhibited on the parameters as well as the forecasts, consider the $ARMA_2(1,1)$ model fitted to series 7. for the GALLUP(c) data. If we consider $\underline{Z}(2)$ i.e.

$$\underline{Z}(2) = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix}$$

and use this to map the estimates for model 7. to those using LIB as the reference variable. i.e. we map,

$$\begin{bmatrix} \ln(LAB/CON) \\ \ln(LIB/CON) \end{bmatrix} \quad \text{to} \quad \begin{bmatrix} \ln(LAB/LIB) \\ \ln(CON/LIB) \end{bmatrix} ,$$

then we have series 9. but with the first and second variable interchanged. Thus for example,

$$\underline{Z}(2) \times CONST = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -0.040 \\ 0.045 \end{bmatrix} = \begin{bmatrix} -0.085 \\ -0.045 \end{bmatrix} , \text{ and rewriting this as}$$

$$\begin{bmatrix} -0.045 \\ -0.085 \end{bmatrix} \text{ so that the variables are as in series 9. we have a}$$

second estimate of CONST for series 9. We may compare this to the estimates in table 4.29 and note that they are reasonably close.

Similarly,

$$\underline{Z}(2) \hat{\underline{\Phi}}^{(CON)} \underline{Z}(2) \Rightarrow \hat{\underline{\Phi}}^{(LIB)} = \begin{bmatrix} 1.083 & -0.059 \\ 0.163 & 0.887 \end{bmatrix} ,$$

$$\underline{Z}(2) \hat{\underline{\Theta}}^{(CON)} \underline{Z}(2) \Rightarrow \hat{\underline{\Theta}}^{(LIB)} = \begin{bmatrix} 0.761 & -0.058 \\ 0.424 & 0.269 \end{bmatrix} ,$$

$$\underline{Z}(1) CONST^{(CON)} \Rightarrow CONST^{(LIB)} = \begin{bmatrix} 0.040 \\ 0.085 \end{bmatrix} ,$$

$$\underline{Z}(1) \hat{\underline{\Phi}}^{(LIB)} \underline{Z}(1) \Rightarrow \hat{\underline{\Phi}}^{(CON)} = \begin{bmatrix} 0.946 & -0.027 \\ 0.058 & 1.022 \end{bmatrix} ,$$

Again the estimates of $\hat{\phi}^{(LIB)}$, $\hat{\theta}^{(LIB)}$, $\hat{CONST}^{(LAB)}$ and $\hat{\phi}^{(CON)}$ compare favourably to the estimates in table 4.29. These represent only a subsection of the possible $\underline{Z}(k)$ transformations that we could perform on the estimates. However they are more than sufficient to demonstrate the theory. We will briefly examine these results at the end of chapter 5, when we have examined the forecasting problem.

4.5 Summary

In this chapter we have introduced two new classes of model, the $\ln_+ARMA_m(p,q)$ and the $\ln_xARMA_m(p,q)$ model and have indicated how similar models might be developed. We have investigated the properties of these models. A Taylor series expansion of the $\ln_+ARMA_m(p,q)$ model has been developed, and this may be easily repeated for the $\ln_xARMA_m(p,q)$ model. An investigation of the $\ln_+ARMA_m(p,q)$ with regard to the choice of reference variable suggests that the basic structure of the model is invariant to this choice. It also appears that the identification statistics are robust, as illustrated by the example of the last section. Section 4.4 also gives evidence that estimation and diagnostic checks are invariant, which confirms the theory outlined in this chapter. What remains is a more detailed look at the forecasting problem, and this will be the subject of the next chapter.

Finally we note that when using the $\ln_xARMA_m(p,q)$ model, it is not so straight forward to examine what happens under a permutation of the u_{it} 's. No equivalent of the $\underline{Z}(k)$ matrix exists. This model in fact requires not only the choice of reference variable, but of the exact order of all the variables. However the main purpose of the $\ln_xARMA_m(p,q)$ model will be seen later in chapter 6, and it is to investigate concepts such as neutrality (definition 3.28) for which the ordering of the u_{it} 's is vital in any case.

"Who told of this from the beginning so that we could know, or
beforehand so that we could say, 'He was right' ? "

Isaiah 41:26

CHAPTER 5

Distributional Properties of Logistic Time Series Models in Relation to Forecasting

5.0 Introduction

In chapter 2 and section 2.5 the $ARMA_m(p,q)$ model was discussed with special reference to forecasting. We recall in particular that the MMSE forecast is the mean of Z_{t+l} (say) conditional on Z_t, Z_{t-1}, \dots . When it is not $\{Z_t\}$ itself that follows an ARMA process, but rather a function of it, $y_t = f(Z_t)$ (say), then the MMSE forecast is,

$$Z_t(l) = E\{f^{-1}(y_{t+l})/Z_t, Z_{t-1}, \dots\}, \quad (5.0.1)$$

with variance $Var[Z_t(l)] = Var\{f^{-1}(y_{t+l})/Z_t, Z_{t-1}, \dots\}$. (5.0.2)

These expressions may be difficult to compute for non-linear functions even if the ARMA process for y is known. An example was given in 2.5 where f was the log function. There we wished to produce forecasts of a series $\{X_t\}$ for which $W_t = \ln X_t$. It was shown that the MMSE forecast of X_{t+l} is then the mean of the log-normal distribution $\Lambda_m(W_t(l), \Sigma_{e_t}(l))$ as in (2.5.15) and (2.5.16).

In chapter 4 we introduced a model based on the a_m function. For $u_t \in \mathbb{S}^m$ and $v_t = a_m(u_t)$ we have that,

$$(v_{t+l}/v_t, v_{t-1}, \dots) \sim N_m(v_t(l), \Sigma_{e_t}(l)) \quad , \quad (5.0.3)$$

and recalling from definition 3.21 we have,

$$(u_{t+l}/u_t, u_{t-1}, \dots) \sim L_m(v_t(l), \Sigma_{e_t}(l)) \quad , \quad (5.0.4)$$

i.e. a logistic-normal with parameters $\underline{v}_t(\ell)$ and $\underline{\Sigma}_{e_t}(\ell)$, where $\{e_t\}$ is now the innovation series of $\{\underline{v}_t\}$. Thus the MMSE forecast of \underline{u}_t is given by the mean of the logistic-normal distribution with parameters $\underline{v}_t(\ell)$ and $\underline{\Sigma}_{e_t}(\ell)$. Similarly if we had modeled \underline{u}_t via the \ln_x ARMA $_m(p,q)$ the MMSE forecast of \underline{u}_t would be the mean of the multiplicative logistic-normal (definition 3.22) $M_m(\underline{v}_t(\ell), \underline{\Sigma}_{e_t}(\ell))$. In both cases it is clear that we have a problem. Aitchison and Shen(1980) note that although the moments of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution do exist, they cannot be readily evaluated. Although no such reference exists for the $M_m(\underline{\mu}, \underline{\Sigma})$ distribution, at least not to our knowledge, a similar result about its moments is expected. (Recall that for $m=1$ both distributions are the same in any case.) Thus we see that we require a detailed investigation of the logistic-normal distributions to investigate the estimation of means and variances. We may also consider other location parameters such as the mode, or in the univariate case the median. Finally we need to examine how to use interval predictors.

5.1 Some Basic Properties of the Logistic-Normal Distributions

Although the logistic-normal does appear in earlier literature, Aitchison and Shen(1980) seem to have made the first concentrated effort to give it a clear identity. Aitchison(1982) also discusses further properties, the first of these is related to the idea of a marginal distribution for the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution.

5.5.1 "Marginal" Distribution Property

In the discussion following definitions 3.21 and 3.22 we showed that if $\underline{u} \in \mathbb{S}^m$, $\underline{v} = a_m(\underline{u})$, $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$ and $\underline{v}^* = \underline{G}\underline{v}$, where \underline{G} is $c \times m$ so that $\underline{v}^* \sim N_m(\underline{G}\underline{\mu}, \underline{G}\underline{\Sigma}\underline{G}')$, then \underline{u} is $L_m(\underline{\mu}, \underline{\Sigma})$ and $\underline{u}^* = a_m^{-1}(\underline{v}^*) \sim L_m(\underline{G}\underline{\mu}, \underline{G}\underline{\Sigma}\underline{G}')$. Expanding $\underline{v}^* = \underline{G}\underline{v}$ we have,

$$v_i^* = \sum_{j=1}^m g_{ij} v_j \quad ; \quad i=1, \dots, c \quad ; \quad (5.1.1)$$

and from definition 3.17,

$$\begin{aligned}
u_i^* &= \frac{\exp v_i^*}{1 + \sum_{j=1}^c \exp v_j^*} = \frac{\exp \left[\sum_{j=1}^m g_{ij} v_j \right]}{1 + \sum_{k=1}^c \exp \left[\sum_{j=1}^m g_{kj} v_j \right]} \\
&= \frac{\exp \left[\sum_{j=1}^m g_{ij} \left(\ln \frac{u_j}{u_{m+1}} \right) \right]}{1 + \sum_{k=1}^c \exp \left[\sum_{j=1}^m g_{kj} \left(\ln \frac{u_j}{u_{m+1}} \right) \right]} = \frac{\prod_{j=1}^m \left(\frac{u_j}{u_{m+1}} \right)^{g_{ij}}}{1 + \sum_{k=1}^c \prod_{j=1}^m \left(\frac{u_j}{u_{m+1}} \right)^{g_{kj}}}, \\
& \qquad \qquad \qquad i=1, \dots, c. \qquad (5.1.2)
\end{aligned}$$

In particular let

$$\left. \begin{aligned}
g_{ii} &= 1 \\
g_{i,c+1} &= -1 \\
g_{ij} &= 0 \quad \text{otherwise,}
\end{aligned} \right\} i=1, \dots, c, \qquad c < m, \qquad (5.1.3)$$

so that $\underline{G} = \begin{bmatrix} 1 & 0 & \dots & -1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & -1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \cdot & \vdots & \vdots & & & \vdots \\ 0 & 1 & & 1 & -1 & 0 & \dots & 0 \end{bmatrix}$. (5.1.4)

Substituting these values of \underline{G} into (5.1.2) we obtain,

$$\underline{u}_i^* = C(u_1, \dots, u_{c+1}) \qquad (5.1.5)$$

Thus we have the following lemma.

Lemma 5.1

If $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$ then $C(u_1, \dots, u_{c+1}) \sim L_c(\underline{G}\underline{\mu}, \underline{G}\underline{\Sigma}\underline{G}')$ where \underline{G} is given by (5.1.3).

We note that this lemma enables us to find the distribution of a subcomposition of \underline{u} , but not a subset of \underline{u} . Thus it is not strictly speaking a marginal distribution. In order to find such a marginal distribution we need to integrate out some of the u 's. This produces untractable expressions and so we have not developed it here.

Clearly we may permute the rows of \underline{G} to form a subcomposition with any combination of the u 's. This also follows from the fact that the L_m distribution is invariant to a permutation of the u 's. The structure of the M_m distribution, however, depends on the order of the u 's. Its equivalent to lemma 5.1 is now given below.

Lemma 5.2

If $\underline{u} \in \mathbb{S}^m$ is such that $\underline{u} \sim M_c(\underline{\mu}, \underline{\Sigma})$, then

$$(u_1, u_2, \dots, u_c, T(u_{c+1}, \dots, u_{m+1})) \sim M_m(\underline{\mu}_1, \underline{\Sigma}_{11}) \quad (5.1.6)$$

and $C(u_{c+1}, u_{c+2}, \dots, u_{m+1}) \sim M_{m-c}(\underline{\mu}_2, \underline{\Sigma}_{22})$, (5.1.7)

where $\underline{\mu}_1$, $\underline{\Sigma}_{11}$, $\underline{\mu}_2$, $\underline{\Sigma}_{22}$ come from the partition of $\underline{\mu}$ and $\underline{\Sigma}$:

$$\underline{\mu}_{m \times 1} = \begin{bmatrix} \underline{\mu}_1 \\ \vdots \\ \underline{\mu}_2 \end{bmatrix}, \quad \underline{\Sigma}_{m \times m} = \begin{bmatrix} \underline{\Sigma}_{11} & \vdots & \underline{\Sigma}_{12} \\ \vdots & \vdots & \vdots \\ \underline{\Sigma}_{21} & \vdots & \underline{\Sigma}_{22} \end{bmatrix},$$

for $c=1, \dots, m-1$ (5.1.8)

Proof

Let $\underline{v} = m(\underline{u})$ so that $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$ (definition 3.22).

Then $\underline{Gv} = \underline{v}^\dagger \sim N_m(\underline{G}\underline{\mu}, \underline{G}\underline{\Sigma}\underline{G}')$ as in the L_m case, where \underline{G} is $d \times m$

and $v_i^\dagger = \sum_{j=1}^m g_{ij} v_j$, $i=1, \dots, d$. (5.1.9)

Now $\underline{u}_i^\dagger = m^{-1}(\underline{v}^\dagger)$.

$$\Rightarrow u_i^\dagger = \frac{\exp v_i^\dagger}{\prod_{k=1}^i (1 + \exp v_k^\dagger)}, \quad i=1, \dots, d;$$

$$= \frac{\prod_{j=1}^m \exp g_{ij} v_j}{\prod_{k=1}^i \left(1 + \prod_{j=1}^m \exp g_{kj} v_j \right)}, \quad \text{from (5.1.9);}$$

$$\begin{aligned}
 & \prod_{j=1}^m \left[\frac{u_j}{1 - \sum_{\ell=1}^j u_\ell} \right]^{\xi_{ij}} \\
 = & \prod_{k=1}^i \left[1 + \prod_{j=1}^m \left[\frac{u_j}{1 - \sum_{\ell=1}^j u_\ell} \right]^{\xi_{kj}} \right]
 \end{aligned} \tag{5.1.10}$$

Now consider $\underline{G} = [\underline{I}_c | \underline{0}]_{c \times m}$,

$$\begin{aligned}
 \text{i.e.} \quad \xi_{ii} &= 1 & i=1, \dots, c; \\
 \xi_{ij} &= 0 & \text{otherwise, } i=1, \dots, c; j=1, \dots, m.
 \end{aligned} \tag{5.1.11}$$

Substituting (5.1.11) into (5.1.10) we have,

$$\begin{aligned}
 u_i^\dagger &= \frac{\left[\frac{u_i}{1 - \sum_{\ell=1}^i u_\ell} \right]}{\prod_{k=1}^i \left[1 - \frac{u_k}{1 - \sum_{\ell=1}^k u_\ell} \right]} \\
 &= \frac{\left[\frac{u_i}{1 - \sum_{\ell=1}^i u_\ell} \right]}{\prod_{k=1}^i \left[\frac{1 - \sum_{\ell=1}^{k-1} u_\ell}{1 - \sum_{\ell=1}^k u_\ell} \right]} \\
 &= \frac{u_i}{1 - \sum_{\ell=1}^i u_\ell} \frac{(1-u_1)(1-u_1-u_2) \dots (1-u_1-\dots-u_{i-1})}{1(1-u_1)(1-u_1-u_2) \dots (1-u_1-\dots-u_{i-1})} \\
 &= u_i; \quad i=1, \dots, c;
 \end{aligned} \tag{5.1.12}$$

$$\begin{aligned} \text{and } u_{c+1}^\dagger &= 1 - \sum_{j=i}^c u_j^\dagger = 1 - \sum_{j=1}^c u_j \\ &= T(u_{c+1}, \dots, u_{m+1}) . \end{aligned}$$

Also $G\mu = \mu_1$ and $G\Sigma G' = \Sigma_{11}$ from which we obtain (5.1.6).

For (5.1.7) we repeat the procedure, but now put $G = [0 | I_{m-c}]_{(m-c) \times m}$,

$$\begin{aligned} \text{that is } \varepsilon_{ij} &= 1 \quad i=1, \dots, c ; j=i+c ; \\ &= 0 \quad \text{otherwise, } i=1, \dots, c ; j=1, \dots, m. \end{aligned} \quad (5.1.13)$$

Again substituting into (5.1.10) gives,

$$\begin{aligned} u_i^\dagger &= \frac{\begin{pmatrix} u_{i+c} \\ \hline 1 - \sum_{\ell=1}^{i+c} u_\ell \end{pmatrix}}{\prod_{k=c+1}^{c+i} \begin{pmatrix} k-1 \\ \hline 1 - \sum_{\ell=1}^{k-1} u_\ell \\ \hline k \\ \hline 1 - \sum_{\ell=1}^k u_\ell \end{pmatrix}} \\ &= \frac{u_{c+i}}{1 - \sum_{\ell=1}^{c+i} u_\ell} \frac{(1-u_1 - \dots - u_{c+1}) \dots (1-u_1 - \dots - u_{c+i})}{1(1-u_1 - \dots - u_c) \dots (1-u_1 - \dots - u_{c+i-1})} \\ &= \frac{u_{i+c}}{1-u_1 - \dots - u_c} \quad (\text{after cancelling}) , \\ &= \frac{u_{i+c}}{u_{c+1} + u_{c+2} + \dots + u_{m+1}} , \quad i=1, \dots, c ; \\ &= 1 - \sum_{i=1}^{m-c} u_i^\dagger = \frac{u_{m+1}}{u_{c+1} + u_{c+2} + \dots + u_{m+1}} \quad i= m+1-c . \end{aligned} \quad (5.1.14)$$

i.e. $\underline{u}^\dagger = C(u_{c+1}, \dots, u_{m+1})$ and $G\mu = \mu_2$; $G\Sigma G' = \Sigma_{22}$ from which we may derive (5.1.7). □

It is upon the above result that Aitchison(1981a) bases his test of neutrality, although it is not specifically expressed in the form of lemma 5.2. It can be seen that whilst (5.1.7) is an identical expression to that for the L_m distribution, (5.1.6) offers us an alternative form of a marginal distribution. Here we have in fact extracted u_1, \dots, u_c as is required of a marginal distribution, but we have also included a fill-up-value which is the sum of the remaining u 's: $u_{c+1} + \dots + u_{m+1}$. Clearly if we limit the marginal distribution to be one which describes data on the simplex, then any form of subset of the u 's must be formed in such a way as to result in a composition. Hence we must obtain the distribution of either a subcomposition as in lemma 5.1 and equation (5.1.7), or as in (5.1.6).

5.1.2 "Conditional" Distribution Property

We first quote the result of Aitchison and Shen(1980) :-

Lemma 5.3

If $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$ then,

$$C(u_1, \dots, u_{c+1}) \mid C(u_{c+1}, \dots, u_{m+1}) \sim$$

$$L_c \left(\underline{\mu}_1 + \underline{e}_c \ln r_1 + \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} (\ln \underline{r} - \underline{\mu}_2), (\underline{\Sigma}_{11} - \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21}) \right)$$

Where, as before, $\underline{\mu}_1, \underline{\mu}_2, \underline{\Sigma}_{11}, \underline{\Sigma}_{22}, \underline{\Sigma}_{12}$ follow from (5.1.8) the $c, m-c$ partition of $\underline{\mu}$ and $\underline{\Sigma}$ (but for the L_m distribution), and where

$$\underline{r} = (r_1, \dots, r_{m-c})' \quad \text{and} \quad r_i = \frac{u_{c+i}}{u_{m+1}}, \quad i=1, \dots, m-c.$$

Proof

The result follows from Aitchison and Shen(1980) and from standard multivariate-normal theory. □

Lemma 5.4

If $\underline{u} \sim M_m(\underline{\mu}, \underline{\Sigma})$ and $u_{c+1}^r = 1 - \sum_{i=1}^c u_i$ then,

$$(u_1, \dots, u_c, u_{c+1}^r) \mid C(u_{c+1}, u_{c+2}, \dots, u_{m+1}) \sim$$

$$M_c(\underline{\mu}_1 + \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} (\underline{t}_2 - \underline{\mu}_2), (\underline{\Sigma}_{11} - \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21})), \quad (5.1.15)$$

and $C(u_{c+1}, \dots, u_{m+1}) \mid (u_1, \dots, u_c, u_{c+1}^\dagger) \sim$

$$M_{m-c}(\underline{\mu}_2 + \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} (\underline{t}_1 - \underline{\mu}_1), (\underline{\Sigma}_{22} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{12})), \quad (5.1.16)$$

where $\underline{t}_1 = m_c(\underline{u}_1)$,

$$\underline{t}_2 = m_{m-c}(\underline{u}_2) = m_{m-c}(C(\underline{u}_2)),$$

and $(\underline{u}_1 \mid \underline{u}_2)'$ is the $c, m-c$ partition of \underline{u} ; with $\underline{\mu}$ and $\underline{\Sigma}$ partitioned as in (5.1.8).

Proof

Again this result follows directly from standard multivariate normal distribution theory and lemma 5.2.

Some of these distributional properties will prove useful in chapter 6, where various parameter restrictions on the above distributions may be related to differing forms of independence. The results of section 5.1.1 and 5.1.2 also serve to demonstrate how we may simply understand the logistic-normal distribution in terms of lower dimensions. This may prove useful in the next section, when some of the results we require are easily formulated for the univariate case.

5.2 Location Parameters

There appears to be little or no work on the location parameters of the L_m and M_m distributions, mainly because of the intractability of the results. However the univariate distribution is relatively well investigated, probably due to its early use and its relative simplicity. For example Johnson(1949) discusses its properties, where in his case, it is called the S_B distribution. (See also Johnson and Kotz(1970) vol. 2). We note that the density function $L_1(\mu, \sigma^2)$ is given by:-

$$P_L(u \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{u(1-u)} \exp\left\{ -\frac{1}{2\sigma^2} \left(\ln \frac{u}{1-u} - \mu \right)^2 \right\}, \quad (5.2.1)$$

$0 < u < 1.$

Also $v = \ln \frac{u}{1-u} \sim N(\mu, \sigma^2),$ (5.2.2)

$$\Rightarrow z = \frac{v-\mu}{\sigma} \sim N(0,1) \quad . \quad (5.2.3)$$

Now Johnson's S_B distribution (standing for "bounded system") is written

$$\text{as} \quad z = \gamma + \delta \ln \frac{u}{1-u} \quad \text{where} \quad z \sim N(0,1). \quad (5.2.4)$$

Comparing this to (5.2.3) it is simply the $L_1(-\gamma/\delta, 1/\delta^2)$ distribution. We note that the L_1 distribution is equivalent to the M_1 distribution, so that the study of the univariate L_1 (M_1) provides a stepping stone to both multivariate distributions. We shall now proceed with a study of the location parameters, and begin with the median.

5.2.1 The Median

For the univariate distribution (5.2.1) the median is :-

$$\frac{e^\mu}{1 + e^\mu} \quad . \quad (5.2.5)$$

The concept of a median is, however, difficult to extend to multivariate distributions. For example, Haldane(1948) describes two possible definitions which he terms the arithmetic median and the geometric median. An idea of ordering multivariate data is discussed by Barnett-(1976) and further definitions of a multivariate median are based on this. (See also Green and Silverman(1979), Scheult et al's (1976) discussion of Barnett's (1976) paper and Sibson (1984)). Some of the definitions above give different results under rotation and/or change of scale, whilst others refer more readily to multivariate samples. Thus we avoid the median, since it has such an ambiguous generalization to the multivariate case.

However, we note that (5.2.5) is simply the inverse of the logistic transformation. Thus it does perhaps serve to justify the possible use of the inverse transformation a_m^{-1} and m_m^{-1} as point predictors even in the multivariate case. We hope to compare these with the mode and mean.

5.2.2 The Mode

Again we will start with the univariate distribution; taking logs and differentiating (5.2.1) with respect to u gives us,

$$\frac{d}{du} \ln P_L(u) = -\frac{1}{u} + \frac{1}{1-u} - \frac{1}{\sigma^2} \left[\ln \frac{u}{1-u} - \mu \right] \frac{1}{u(1-u)} \quad ,$$

so that the modal value must satisfy :-

$$2\sigma^2 u - \sigma^2 + \mu = \ln \frac{u}{1-u} . \quad (5.2.6)$$

Johnson(1949) then considers (using our notation) the equations,

$$y = 2\sigma^2 u - \sigma^2 + \mu , \quad (5.2.7)$$

and
$$y = \ln \frac{u}{1-u} . \quad (5.2.8)$$

A plot of the line (5.2.7) and the curve (5.2.8) is given in figure 5.5. The intersection between the curve and the line represent solutions to (5.2.6). It is easily seen that there are some positions of the line for which there are 3 solutions, these represent two modal values and one minimum.

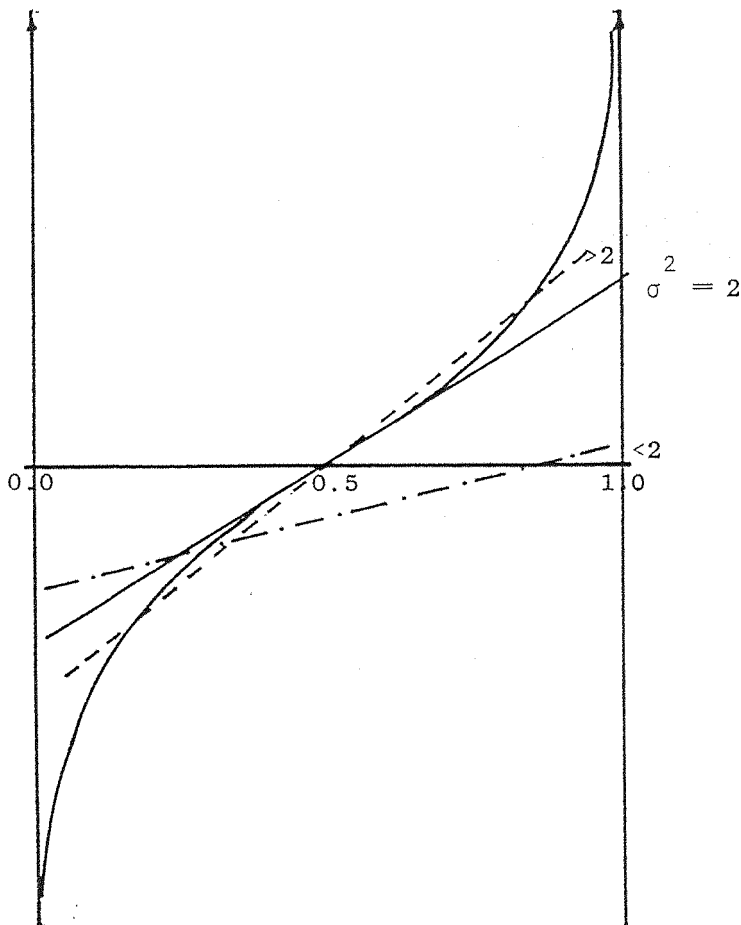


Figure 5.5 Plot of $y = \ln \frac{u}{1-u}$ and of $y = 2\sigma^2 u - \sigma^2 + \mu$ for various values of σ^2 and μ .

Consider the tangents to the curve which are parallel to the line. The gradient of the line is simply $2\sigma^2$ and the gradient of the tangent is given by,

$$\frac{d}{du} \left[\ln \frac{u}{1-u} \right] = \frac{1}{u(1-u)} .$$

Thus the tangent is parallel to the line when,

$$2\sigma^2 = \frac{1}{u(1-u)} , \quad (5.2.9)$$

$$\Rightarrow u = \frac{1}{2} \pm \frac{1}{2} \sqrt{1 - 2/\sigma^2} . \quad (5.2.10)$$

There are then three possible cases :- $\sigma^2 < 2$, $\sigma^2 = 2$, and $\sigma^2 > 2$. For a real solution to (5.2.9) $\sigma^2 \geq 2$, and for three distinct solutions (i.e two modes) $\sigma^2 > 2$. The other requirement for three intersections is that the line must be "close" to the u-axis. Johnson(1949) demonstrated that L_1 is bi-modal when :-

$$\sigma^2 > 2 , \quad (5.2.11)$$

and
$$|\mu| < \sigma^2 \sqrt{1-2/\sigma^2} - 2 \tanh^{-1} \sqrt{1-2/\sigma^2} . \quad (5.2.12)$$

The equation (5.2.12) represents an exact definition of how close the line must be.

If $\sigma^2 > 2$ we may compute the limiting value of μ for various choices of σ^2 . Johnson(1949) does this in terms of his parameters. For our parameters we have the following table.

max μ	0.00	0.16	0.42	0.73	1.07	1.43	1.81	2.20	2.61	3.02	3.44
σ^2	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0

Table 5.6 : Limiting values for which $L_1(M_1)$ will be bi-modal.

We also note that if $\mu = 0$ L is symmetrical otherwise it is skewed.

To find the actual values we must solve (5.2.6) numerically. However from the above we may determine whether the distribution is uni-modal. It would be useful to obtain a similar result for our multivariate distributions. For the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution we need to maximize,

$$f(\underline{u}) = |2\pi\underline{\Sigma}|^{-\frac{1}{2}} \left[\prod_{j=1}^{m+1} u_j \right]^{-1} \exp \left\{ -\frac{1}{2} \left[\ln \frac{u_m}{u_{m+1}} - \mu \right]' \underline{\Sigma}^{-1} \left[\ln \frac{u_m}{u_{m+1}} - \mu \right] \right\}, \quad \underline{u}_m \in \mathbb{S}^m. \quad (5.2.13)$$

Taking logs and differentiating w.r.t. \underline{u} we have:-

$$\begin{aligned} \frac{\partial}{\partial \underline{u}} \ln(f(\underline{u})) &= - \left[\frac{1}{u_1}, \frac{1}{u_2}, \dots, \frac{1}{u_m} \right]' + e_m \frac{1}{u_{m+1}} - \underline{D} \underline{\Sigma}^{-1} \left[\ln \frac{u_m}{u_{m+1}} - \mu \right] \\ &= \underline{0}, \quad \underline{u}_m \in \mathbb{S}^m, \end{aligned} \quad (5.2.14)$$

where
$$\underline{D} = \text{dg} \left[\frac{1}{u_1}, \frac{1}{u_2}, \dots, \frac{1}{u_m} \right]' + \frac{1}{u_{m+1}} e_m.$$

Let
$$\underline{f}_m = \left[\frac{1}{u_{m+1}} - \frac{1}{u_1}, \frac{1}{u_{m+1}} - \frac{1}{u_2}, \dots, \frac{1}{u_{m+1}} - \frac{1}{u_m} \right]'$$
 then we need

to find the solution of,

$$\underline{f}_m - \underline{D} \underline{\Sigma}^{-1} \left[\ln \frac{u_m}{u_{m+1}} - \mu \right] = \underline{0}. \quad (5.2.15)$$

We recall from chapter 4 and corollary 4.11 that the matrix \underline{F} was defined to be (equation (4.2.31)),

$$\underline{F} = \text{dg} \left[\frac{1}{\mu_1}, \dots, \frac{1}{\mu_m} \right] + \frac{1}{\mu_{m+1}} \underline{U}_m.$$

For which \underline{F}^{-1} is given by (4.2.50). Thus noting that \underline{D} is of the same form we have,

$$\underline{D}^{-1} = \begin{bmatrix} u_1(1-u_1) & -u_2 u_1 & \dots & -u_m u_1 \\ -u_1 u_2 & u_2(1-u_2) & \dots & -u_m u_2 \\ \vdots & \vdots & \ddots & \vdots \\ -u_1 u_m & -u_2 u_m & \dots & u_m(1-u_m) \end{bmatrix}, \quad (5.2.16)$$

$$\text{and } |D| = \frac{1}{\prod_{i=1}^{m+1} u_i} \quad (5.2.17)$$

Thus we may rewrite (5.2.15) as ,

$$\ln \frac{u_m}{u_{m+1}} = \underline{\Sigma D}^{-1} \underline{f}_m + \underline{\mu} \quad (5.2.18)$$

From (5.2.16),

$$\begin{aligned} \left[\underline{D}^{-1} \underline{f}_m \right]_i &= \frac{u_i}{u_{m+1}} \left[1 - \sum_{j=1}^m u_j \right] - \left[1 - m u_i \right] , \quad i=1, \dots, m ; \\ &= (m+1)u_i - 1 , \quad \text{where } [-]_i \text{ denotes the } i^{\text{th}} \text{ row.} \end{aligned} \quad (5.2.19)$$

Thus substituting into (5.2.18) gives,

$$\ln \frac{u_m}{u_{m+1}} = \underline{\Sigma} \{ (m+1) \underline{u} - \underline{e}_m \} + \underline{\mu} \quad (5.2.19)$$

which for $m=1$ gives exactly (5.2.6) as before. We may then examine,

$$\underline{y} = (m+1) \underline{\Sigma u} - \underline{\Sigma e}_m + \underline{\mu} \quad (5.2.20)$$

and

$$\underline{y} = \ln \frac{u_m}{u_{m+1}} \quad (5.2.21)$$

to see under what circumstances the multivariate distribution is uni-modal. It is much more complicated than in the univariate case. The equivalent concept to the tangent of the curve being parallel to the line in figure 5.5 is now that,

$$\frac{\partial}{\partial \underline{u}} \left[\ln \frac{u_m}{u_{m+1}} \right] = \frac{\partial}{\partial \underline{u}} \left[(m+1) \underline{\Sigma u} - \underline{\Sigma e}_m - \underline{\mu} \right] \quad (5.2.22)$$

where when \underline{x} is $n \times 1$, $\underline{y} = \underline{f}(\underline{x})$ is $n \times 1$, and by $\frac{\partial}{\partial \underline{x}} [\underline{f}(\underline{x})]$ we mean the $n \times n$ matrix of partial derivatives.

Evaluating (5.2.22) yields,

$$\underline{D} = (m+1) \underline{\Sigma} \quad ,$$

where \underline{D} is given by (5.2.14). i.e. :-

$$\text{dg} \left(\frac{1}{u_1}, \dots, \frac{1}{u_m} \right) + \frac{1}{u_{m+1}} u_m = (m+1)\underline{\Sigma} \quad (5.2.23)$$

This represents m -equations in m unknowns. ($u_{m+1} = 1 - \sum_{i=1}^m u_i$). It is easier to solve if we take the inverse of both sides to give,

$$\underline{D}^{-1} = \frac{1}{(m+1)} \underline{\Sigma}^{-1} \quad (5.2.24)$$

Let $\{\underline{\Sigma}^{-1}\}_{ij} = \sigma^{ij}$, then from (5.2.16) we have that,

$$u_i(1 - u_i) = \frac{1}{m+1} \sigma^{ii}, \quad i=1, \dots, m \quad (5.2.25)$$

Solving for u_i gives,

$$u_i = \frac{1 \pm \sqrt{1 - \frac{4\sigma^{ii}}{m+1}}}{2}, \quad i=1, \dots, m. \quad (5.2.26)$$

If we compare this with (5.2.10) it is identical when $m=1$. Again there are three possibilities,

$$\sigma^{ii} > \frac{m+1}{4},$$

$$\sigma^{ii} = \frac{m+1}{4},$$

$$\sigma^{ii} < \frac{m+1}{4}.$$

The third of these corresponds to the earlier requirement that $\sigma^2 > 2$, and yields two solutions to (5.2.26). Hence when,

$$\sigma^{ii} < \frac{m+1}{4}, \quad (5.2.27)$$

we may conclude that the plane (5.2.20) may intersect the curve (5.2.21) in three places in the u_i direction. The number of modes will therefore be dependent on how many times the equation (5.2.27) holds; and on the "closeness" of the line to the curve.

Consider the limiting case, i.e. when,

$$\sigma^{ii} = \frac{m+1}{4}, \quad i=1, \dots, m. \quad (5.2.28)$$

i.e. when the diagonal elements of $\underline{\Sigma}^{-1}$ are identical. For $L_m(\underline{\mu}, \underline{\Sigma})$ to be symmetrical in all elements u_i , $i=1, \dots, m+1$, we require that $\underline{\mu} = \underline{0}$ and $\underline{\Sigma} = \text{dg}(\sigma, \dots, \sigma) + \underline{U}_m \sigma$, as can be seen if we note that $\underline{\mu}_{(k)} = \underline{Z}(k)\underline{\mu} = \underline{\mu}$, and $\underline{\Sigma}_{(k)} = \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k) = \underline{\Sigma}$. For the symmetrical case we therefore have that,

$$\underline{\Sigma} = \begin{bmatrix} 2\sigma & \sigma & \dots & \sigma \\ \sigma & 2\sigma & \dots & \sigma \\ \vdots & \vdots & \ddots & \vdots \\ \sigma & \sigma & \dots & 2\sigma \end{bmatrix}, \quad (5.2.29)$$

$$\underline{\Sigma}^{-1} = \begin{bmatrix} m/\sigma & -1/\sigma & \dots & -1/\sigma \\ -1/\sigma & m/\sigma & \dots & -1/\sigma \\ \vdots & \vdots & \ddots & \vdots \\ -1/\sigma & -1/\sigma & \dots & m/\sigma \end{bmatrix}. \quad (5.2.30)$$

In this instance the limiting case is,

$$\frac{m}{(m+1)\sigma} = \frac{m+1}{4}, \quad (5.2.31)$$

i.e. that
$$\sigma = \frac{4m}{(m+1)^2}. \quad (5.2.32)$$

Consequently if $\sigma > 4m/(m+1)^2$ the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution with $\underline{\Sigma}$ given by (5.2.29) may be multi-modal, depending on $\underline{\mu}$. When $\sigma < 4m/(m+1)^2$ it will be uni-modal. Once we get away from the non-symmetrical case it becomes difficult to assess when the distribution will be multi-modal. Clearly the elements of $\underline{\Sigma}_{(k)}$ and hence $\underline{\Sigma}_{(k)}^{-1}$ may be of differing magnitudes for $k \in (1, \dots, m+1)$. We may however investigate the conditions,

$$\sigma_{ij}^{(k)} < \frac{4m}{(m+1)^2}, \quad \sigma_{ii}^{(k)} < \frac{8m}{(m+1)^2}, \quad (5.2.33)$$

$$\sigma_{ij}^{(k)} < \frac{-(m+1)}{4m}, \quad \sigma_{ii}^{(k)} < \frac{m+1}{4}; \quad (5.2.34)$$

$$k \in (1, \dots, m+1),$$

as a possible "rule-of-thumb" criteria for uni-modality. This will be investigated in section 5.2.4 where we will illustrate a few L_2 and L_3 distributions. The actual values of the mode(s) must be evaluated numerically.

5.2.3 The Mean

The mean of the univariate distributions has been shown by Johnson(1949) to be,

$$\mu_u = \frac{1}{2\sqrt{2\pi}} e^{-\mu^2/2\sigma^2} \left\{ \sigma + 2\sigma \sum_{n=1}^{\infty} e^{-n^2\sigma^2/2} \cosh\frac{n}{2}(\sigma^2 + 2\mu) \operatorname{sech}\frac{n\sigma^2}{2} + \frac{4\pi}{\sigma} \sum_{n=1}^{\infty} e^{-\frac{1}{2}(2n-1)^2\pi^2/\sigma^2} \frac{\sin\frac{(2n-1)\pi\mu}{\sigma^2} \operatorname{cosech}(2n-1)\frac{\pi^2}{\sigma^2}}{\sigma^2} \right\} /$$

$$\left\{ 1 + 2 \sum_{n=1}^{\infty} e^{-2n^2\pi^2/\sigma^2} \cos\frac{2n\pi\mu}{\sigma^2} \right\} . \quad (5.2.35)$$

This expression is somewhat complicated and again must be computed numerically. Johnson also gives various recurrence relationships, and other expressions relating to higher order moments in terms of the partial differential of μ_u with respect to μ . Thus it seems that exact relationships such as (5.2.35) are more complicated to evaluate than is a numerical integration, and consequently there is little point in trying to extend (5.2.35) to higher dimensions.

There is, however, a more useful result that may be extended to higher order distributions. It consists of an approximation which was developed by Aitchison and Begg(1976). They consider the integral,

$$\int_{-\infty}^{\infty} \frac{e^v}{1 + e^v} p(v) dv = E \left[\frac{e^v}{1 + e^v} \right] = E[u] = \mu_u . \quad (5.2.36)$$

In our case $p(v)$ is the density function of a normal $N(\mu, \sigma^2)$ distribution since for L_1 (or M_1) we assume $v \sim N(\mu, \sigma^2)$. Aitchison and Begg instead evaluate (5.2.36) for $p(v)$ equal to the density function of the student distribution. However in the course of approximating (5.2.36) $p(v)$ is taken to be nearly normal in any case. Thus in this instance we actually slightly improve on Aitchison and Begg's result.

Consider the function,

$$\frac{e^v}{1 + e^v} .$$

This has the range of values (0,1) and can be viewed as a distribution function corresponding to a density function,

$$\frac{e^v}{(1 + e^v)^2}, \text{ since } \frac{e^v}{1 + e^v} = \int_{-\infty}^v \frac{e^t}{(1 + e^t)^2} dt \quad (5.2.37)$$

(For example see Johnson and Kotz(1970).)

This distribution, known as the logistic distribution may be approximated by a normal distribution with mean zero and an appropriate variance σ_1^2 . (c.f. the student t distribution.) Making this substitution into (5.2.36) we have,

$$\mu_u \approx \tilde{\mu}_u = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^v \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-t^2/2\sigma_1^2} dt \right\} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(v-\mu)/2\sigma^2} dv \quad (5.2.38)$$

If we then substitute $k=t-v$ and $V=v$ we obtain after some manipulation,

$$\begin{aligned} \tilde{\mu}_u &= \int_{-\infty}^0 \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma^2)}} \exp\{-(k+\mu)^2/2(\sigma_1^2 + \sigma^2)\} dk \\ &= \Phi \left(\frac{1}{\sqrt{\sigma_1^2 + \sigma^2}} \right), \end{aligned} \quad (5.2.39)$$

where Φ is the cumulative density function of the standard normal.

Aitchison and Begg also suggest another approximation obtained by taking a Taylor series expansion of $\frac{e^v}{1 + e^v}$ about μ . Since this latter approximation was then shown to give worse results we do not consider it.

Table 5.7 below gives the values of μ_u and σ_u^2 as calculated by Johnson via expression (5.2.35) and also $\tilde{\mu}_u$ with $\sigma_1^2 = 2.942$. This value, suggested by Aitchison and Begg is chosen so as to make the 90% quantiles of the logistic and the $N(0, \sigma_1^2)$ distribution agree. We note that the values of $\tilde{\mu}_u$ are close to the corresponding values of μ_u .

σ	μ	μ_u	$\tilde{\mu}_u$	σ_u	σ	μ	μ_u	$\tilde{\mu}_u$	σ_u
2.0	0.00	0.500	0.500	0.314	1.0	1.50	0.221	0.225	0.155
	1.00	0.352	0.352	0.296		2.00	0.155	0.159	0.125
	2.00	0.225	0.224	0.249		2.50	0.105	0.104	0.095
	3.00	0.130	0.127	0.187	0.5	0.00	0.500	0.500	0.118
	4.00	0.068	0.044	0.126		0.25	0.441	0.444	0.117
	5.00	0.032	0.029	0.077		0.50	0.384	0.390	0.112
1.0	0.00	0.500	0.500	0.208	0.75	0.330	0.337	0.100	
	0.50	0.398	0.401	0.201	1.00	0.279	0.288	0.097	
	1.00	0.303	0.308	0.182	1.25	0.234	0.242	0.087	

Table 5.7 : moments of $L_1(\mu, \sigma^2)$ distribution

Let us now consider how we may extend the above result to multivariate distributions. Aitchison and Begg(1976) indicate that this may be done easily, but do not develop the result. We need to evaluate the integral I, where,

$$\begin{aligned}
 I &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{e^{v_k}}{1 + \sum_{i=1}^m e^{v_i}} \phi(\underline{v}/\underline{\mu}, \underline{\Sigma}) dv_1, dv_2, \dots, dv_m \\
 &= E[u_k] , \quad k=1, \dots, m .
 \end{aligned}
 \tag{5.2.40}$$

Where $\phi(\underline{v}/\underline{\mu}, \underline{\Sigma})$ is the normal density function with mean $\underline{\mu}$ and variance $\underline{\Sigma}$:-

$$\phi(\underline{v}/\underline{\mu}, \underline{\Sigma}) = \frac{1}{|2\pi\underline{\Sigma}|^{-\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underline{v}-\underline{\mu})' \underline{\Sigma}^{-1}(\underline{v}-\underline{\mu})\right\} ,
 \tag{5.2.41}$$

Now we may re-write,

$$\frac{e^{v_k}}{1 + \sum_{i=1}^m e^{v_i}} = \frac{1}{1 + \sum_{i=1}^m e^{-w_i}} ,
 \tag{5.2.42}$$

where $w_i = \begin{cases} -v_i + v_k & i=1, \dots, m ; i \neq k \\ v_k & i=k. \end{cases}$

That is $\underline{w} = -\underline{v}_{(k)} = -\underline{Z}(k)\underline{v}$.

Substituting this into (5.2.40) and noting that the jacobian equals unity we obtain,

$$I = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{1}{1 + \sum_{i=1}^m e^{-w_i}} \phi(\underline{w}; -\underline{Z}(k)\underline{\mu}, \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k)) d\underline{w} \quad . \quad (5.2.43)$$

Now $\frac{1}{1 + \sum_{i=1}^m e^{-w_i}}$ is the cumulative distribution function of the multi-

variate logistic distribution, (not to be confused with the Logistic-normal distribution $L_m(\underline{\mu}, \underline{\Sigma})$) see, for example, Johnson and Kotz(1972), and may be approximated by a multivariate normal density function,

$$\frac{1}{1 + \sum_{i=1}^m e^{-w_i}} \approx \int_{-\infty}^{w_1} \int_{-\infty}^{w_2} \dots \int_{-\infty}^{w_m} \phi(\underline{L}; \underline{0}, \underline{\Sigma}_1) d\underline{L} \quad . \quad (5.2.44)$$

As in the univariate case we may substitute (5.2.44) into (5.2.43) and obtain the integral of two normal densities, one over the complete range. Thus after a suitable transformation we would hope to eliminate many of the terms and reduce (5.2.44) substituted into (5.2.43) to a single cumulative normal density function as in (5.2.39). Hence substituting (5.2.44) into (5.2.43) gives.

$$I \approx \tilde{I} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{w_1} \int_{-\infty}^{w_2} \dots \int_{-\infty}^{w_m} \phi(\underline{L}; \underline{0}, \underline{\Sigma}_1) d\underline{L} \right\} \phi(\underline{w}; -\underline{Z}(k)\underline{\mu}, \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k)) d\underline{w} \quad ,$$

Let $\underline{\mu}(k) = \underline{Z}(k)\underline{\mu}$, and $\underline{\Sigma}(k) = \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k)$, then

$$\tilde{I} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{w_1} \int_{-\infty}^{w_2} \dots \int_{-\infty}^{w_m} \frac{1}{|2\pi\underline{\Sigma}_1|^{1/2}} \exp\{-\frac{1}{2}\underline{L}'\underline{\Sigma}_1^{-1}\underline{L}\} d\underline{L} \times \quad (5.2.45)$$

$$\frac{1}{|2\pi\underline{\Sigma}(k)|^{1/2}} \exp\{-\frac{1}{2}(\underline{w} + \underline{\mu}(k))' \underline{\Sigma}^{-1}(k) (\underline{w} + \underline{\mu}(k))\} d\underline{w}$$

We now perform the transformation,

$$\begin{aligned} \underline{S} &= \underline{L} - \underline{w} & \underline{w} &= \underline{T} \\ & \Leftrightarrow & & \\ \underline{T} &= \underline{w} & \underline{L} &= \underline{S} + \underline{T} \end{aligned} \quad (5.2.46)$$

The jacobian is again unity, substituting into (5.2.45) gives :-

$$\begin{aligned} \tilde{I} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^0 \cdots \int_{-\infty}^0 \frac{1}{(|2\pi\underline{\Sigma}_1| |2\pi\underline{\Sigma}(k)|)^{\frac{1}{2}}} \times \\ & \exp -\frac{1}{2} \left\{ (\underline{S} + \underline{T})' \underline{\Sigma}_1^{-1} (\underline{S} + \underline{T}) + (\underline{T} + \underline{\mu}(k))' \underline{\Sigma}^{-1}(k) (\underline{T} + \underline{\mu}(k)) \right\} d\underline{S} d\underline{T} \end{aligned}$$

Completing the square in the exponential term, and reversing the order of integration gives :-

$$\begin{aligned} \tilde{I} &= \int_{-\infty}^0 \cdots \int_{-\infty}^0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(|2\pi\underline{\Sigma}_1| |2\pi\underline{\Sigma}(k)|)^{\frac{1}{2}}} \exp -\frac{1}{2} \left\{ (\underline{S} + \underline{\mu}(k))' (\underline{\Sigma}_1 + \underline{\Sigma}(k))^{-1} (\underline{S} + \underline{\mu}(k)) + \right. \\ & \left. \left[\underline{T} + (\underline{\Sigma}_1^{-1} + \underline{\Sigma}^{-1}(k))^{-1} (\underline{\Sigma}_1^{-1} \underline{S} + \underline{\Sigma}^{-1}(k) \underline{\mu}(k)) \right]' (\underline{\Sigma}_1^{-1} + \underline{\Sigma}^{-1}(k)) \times \right. \\ & \left. \left[\underline{T} + (\underline{\Sigma}_1^{-1} + \underline{\Sigma}^{-1}(k))^{-1} (\underline{\Sigma}_1^{-1} \underline{S} + \underline{\Sigma}^{-1}(k) \underline{\mu}(k)) \right] \right\} d\underline{T} d\underline{S} \\ &= \int_{-\infty}^0 \cdots \int_{-\infty}^0 \frac{1}{(|2\pi(\underline{\Sigma}_1 + \underline{\Sigma}(k))|)^{\frac{1}{2}}} \exp -\frac{1}{2} \left\{ (\underline{S} + \underline{\mu}(k))' (\underline{\Sigma}_1 + \underline{\Sigma}(k))^{-1} (\underline{S} + \underline{\mu}(k)) \right\} \times \\ & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{(|2\pi(\underline{\Sigma}_1 + \underline{\Sigma}(k))|)^{\frac{1}{2}}}{(|2\pi\underline{\Sigma}_1| |2\pi\underline{\Sigma}(k)|)^{\frac{1}{2}}} \exp -\frac{1}{2} \left\{ \left[\underline{T} + (\underline{\Sigma}_1^{-1} + \underline{\Sigma}^{-1}(k))^{-1} (\underline{\Sigma}_1^{-1} \underline{S} + \underline{\Sigma}^{-1}(k) \underline{\mu}(k)) \right]' \times \right. \\ & \left. (\underline{\Sigma}_1 (\underline{\Sigma}_1 + \underline{\Sigma}(k))^{-1} \underline{\Sigma}(k))^{-1} \left[\underline{T} + (\underline{\Sigma}_1^{-1} + \underline{\Sigma}^{-1}(k))^{-1} (\underline{\Sigma}_1^{-1} \underline{S} + \underline{\Sigma}^{-1}(k) \underline{\mu}(k)) \right] \right\} d\underline{T} d\underline{S} \end{aligned}$$

We note that $(\underline{\Sigma}_1^{-1} + \underline{\Sigma}^{-1}(k)) = (\underline{\Sigma}_1 (\underline{\Sigma}_1 + \underline{\Sigma}(k))^{-1} \underline{\Sigma}(k))^{-1}$ which has been substituted into the second exponent. The integration over \underline{T} is that of

a normal density function over its entire range and is therefore equal to unity. Hence,

$$\begin{aligned} \tilde{I} &= \int_{-\infty}^0 \dots \int_{-\infty}^0 \frac{1}{(2\pi(\underline{\Sigma}_1 + \underline{\Sigma}(k)))^{1/2}} \exp \left\{ -\frac{1}{2} \left\{ (\underline{S} + \underline{\mu}(k))' (\underline{\Sigma}_1 + \underline{\Sigma}(k))^{-1} (\underline{S} + \underline{\mu}(k)) \right\} \right\} d\underline{S} \\ &= \Phi\{\underline{0}; -\underline{\mu}(k), [\underline{\Sigma}_1 + \underline{\Sigma}(k)]\} \quad . \end{aligned} \quad (5.2.47)$$

The result is summarized in the following lemma.

Lemma 5.8

If $\underline{u} \sim L_m(\underline{\mu}, \underline{\Sigma})$ then we may approximate $E(u_k)$ by the value of the m -dimensional normal cumulative density function with mean $-\underline{Z}(k)\underline{\mu}$ and variance $\underline{\Sigma}_1 + \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k)$ evaluated at zero, i.e.

$$E(u_k) \approx \Phi_m(0 / -\underline{\mu}(k), \underline{\Sigma}(k) + \underline{\Sigma}_1) \quad ; \quad k=1, \dots, m+1$$

where

$$\underline{\mu}(k) = \underline{Z}(k)\underline{\mu} \quad ,$$

and

$$\underline{\Sigma}(k) = \underline{Z}(k)\underline{\Sigma}\underline{Z}'(k) \quad , \quad (\underline{Z}_{(m+1)} = \underline{I}_m).$$

Tables exist for computing bi-variate and tri-variate normal densities e.g. Owen(1956) and Steck(1958). However in order to use the lemma we require a suitable choice of the matrix $\underline{\Sigma}_1$ which we recall must be chosen so as to facilitate a good approximation of the logistic distribution by a $N_m(0, \underline{\Sigma}_1)$ density function. A sensible initial choice is to take $\underline{\Sigma}_1$ to be the value of the covariance function for the logistic distribution.

For a logistic random variable L Johnson and Kotz(1972) give the following results,

$$\begin{aligned} \text{Var}(L_i) &= \pi^2/3, & i=1, \dots, m; \\ \text{Cov}(L_i, L_j) &= \pi^2/6, & i, j=1, \dots, m, \quad i \neq j. \end{aligned} \quad (5.2.48)$$

Hence $\underline{\Sigma}_1$ must at least be of the form,

$$\begin{bmatrix} \sigma & \sigma/2 & \dots & \sigma/2 \\ \sigma/2 & \sigma & \dots & \sigma/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma/2 & \sigma/2 & \dots & \sigma/2 \end{bmatrix} = \sigma/2 \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2 \end{bmatrix} \quad (5.2.49)$$

Two obvious possibilities for the choice of σ are then,

- i) $\sigma = \pi^2/3 = 3.290$ since it is the value for the true distribution,
- ii) $\sigma = 2.942$ as in the univariate case.

Having noted these two possibilities another is,

- iii) $\sigma = 3$ since it lies between i) and ii) and for simplicity.

We may easily compare these possibilities in a numerical study. A small study of this kind is presented in the next section, and it will be seen that all three values give reasonable results.

To obtain a similar result to lemma 5.8 for the $M_m(\underline{\mu}, \underline{\Sigma})$ distribution is not so straight forward. As mentioned earlier there is no equivalent matrix to the $Z(k)$ matrix, partly because the variables v in the M_m distribution by their very nature depend on the ordering of the u 's. In fact we would have to make a separate distributional approximation for each u_i in turn, i.e.

$$\frac{e^{v_1}}{1 + e^{v_1}}, \quad \frac{e^{v_2}}{(1 + e^{v_1})(1 + e^{v_2})}, \quad \frac{e^{v_3}}{(1 + e^{v_1})(1 + e^{v_2})(1 + e^{v_3})} \dots \text{etc.}$$

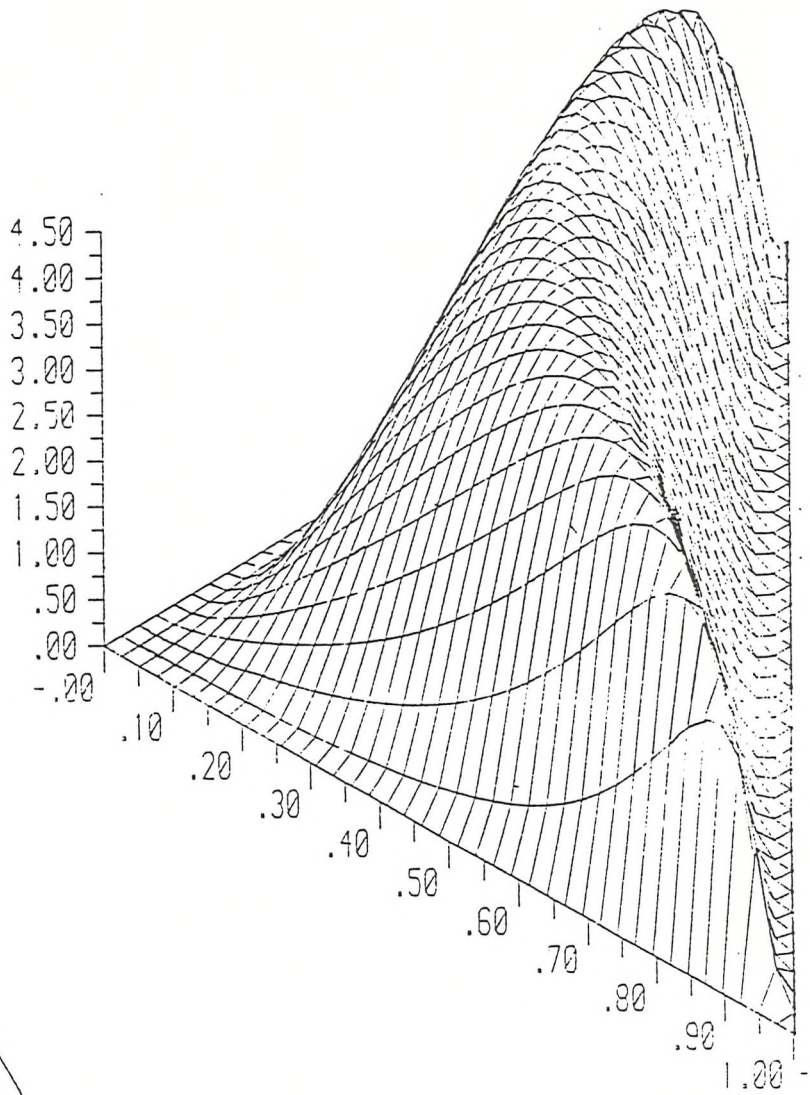
Although it seems likely that these may represent functions that can be approximated by a normal density function we do not pursue this here.

5.2.4 A Numerical Study of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution

Various analytical methods for evaluating the mean, and the mode of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution have been explored in 5.2.1-5.2.3. We have concluded that the only way to evaluate these and other such values e.g. the variance, is numerically. However we have derived an approximation to the mean (5.2.47) and indicated that the distribution is not necessarily uni-modal. The purpose of this numerical study is to compare the approximation (5.2.47) to the true mean, and to gain a better "feel" for the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution. In particular we hope to understand when the distribution is multi-modal and compare various location parameters such as $a_m^{-1}(\underline{\mu})$, the mode and the mean. A greater insight into this distribution will enable us to tackle the forecasting problem.

A selection of $L_2(\underline{\mu}, \underline{\Sigma})$ and $L_3(\underline{\mu}, \underline{\Sigma})$ distributions were studied. They were chosen so as to encompass several properties; some are symmetrical, some are multi-modal, some have $\underline{\mu}=\underline{0}$, whilst some have $\underline{\mu}\neq\underline{0}$;

a) 3-d Plot.



b) Contour Plot

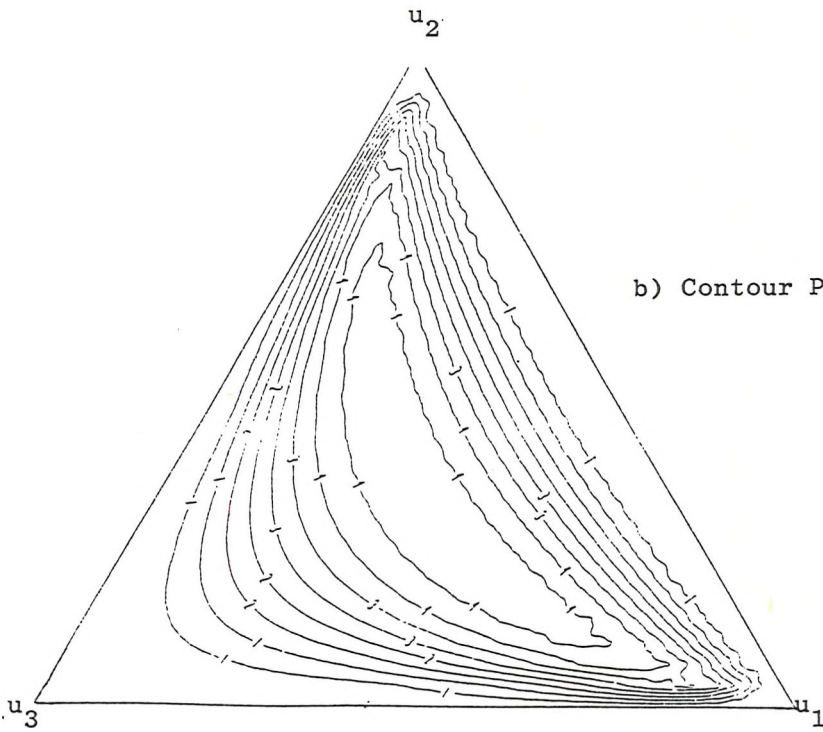


Figure 5.9 Plot of the $L_2(0, I_2)$.

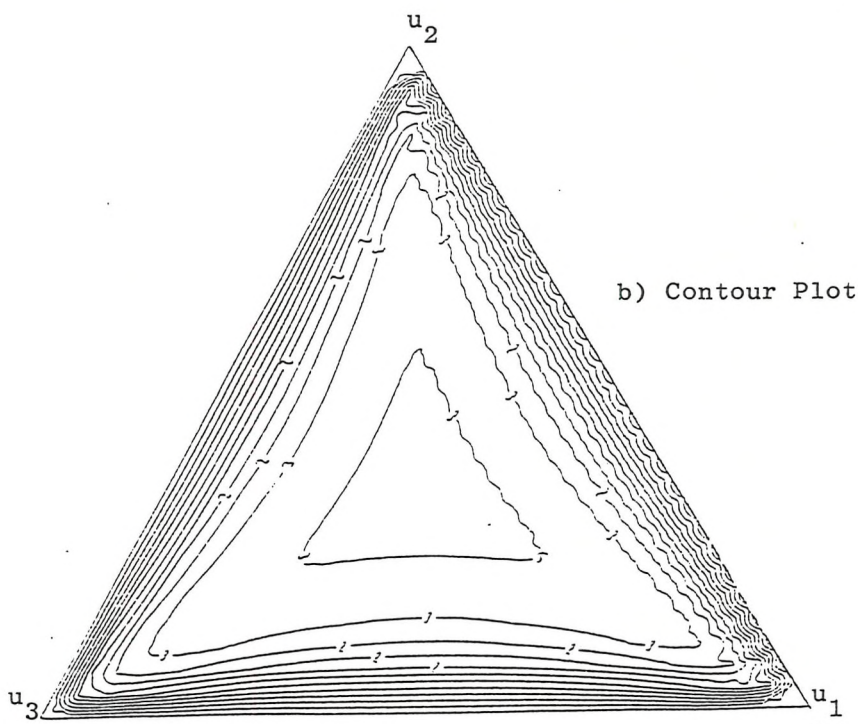
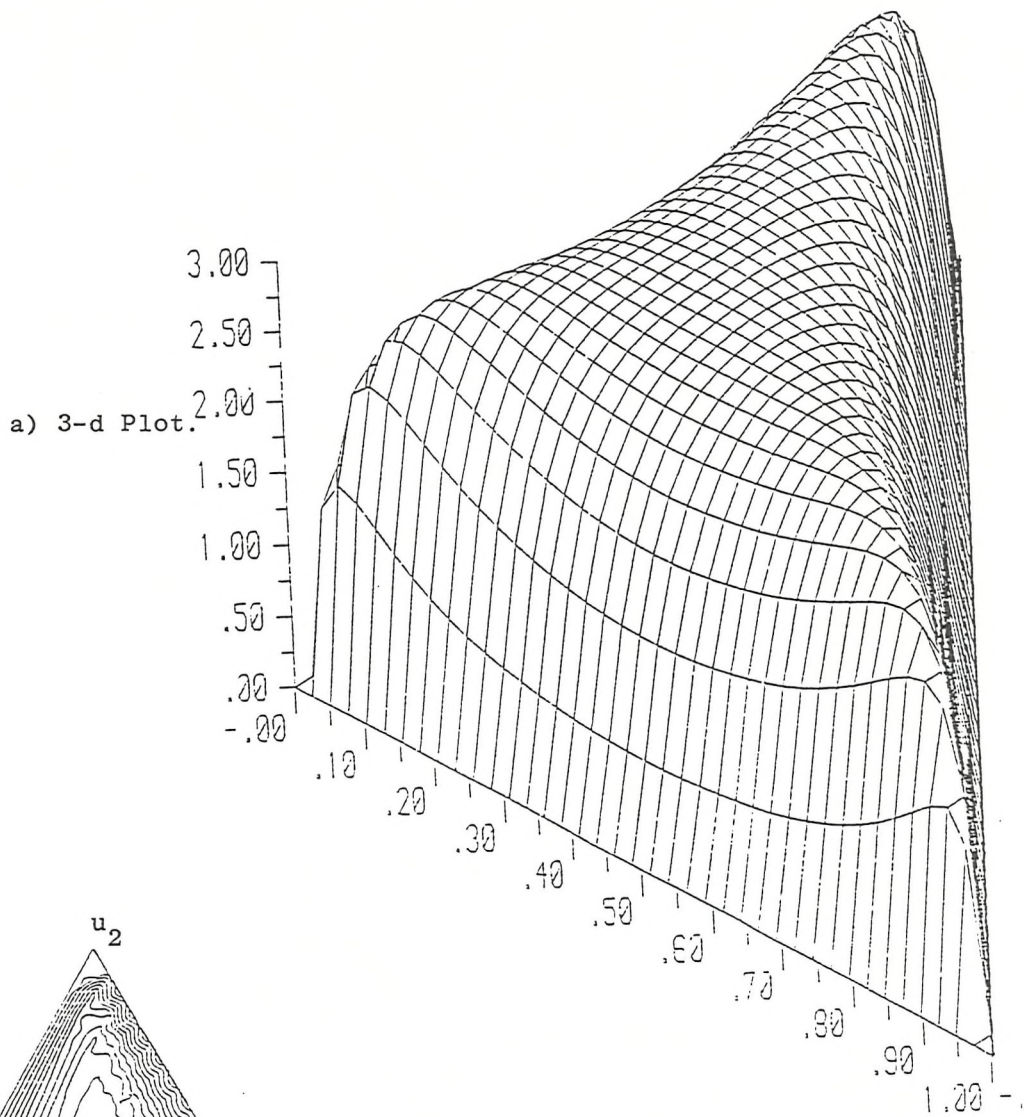
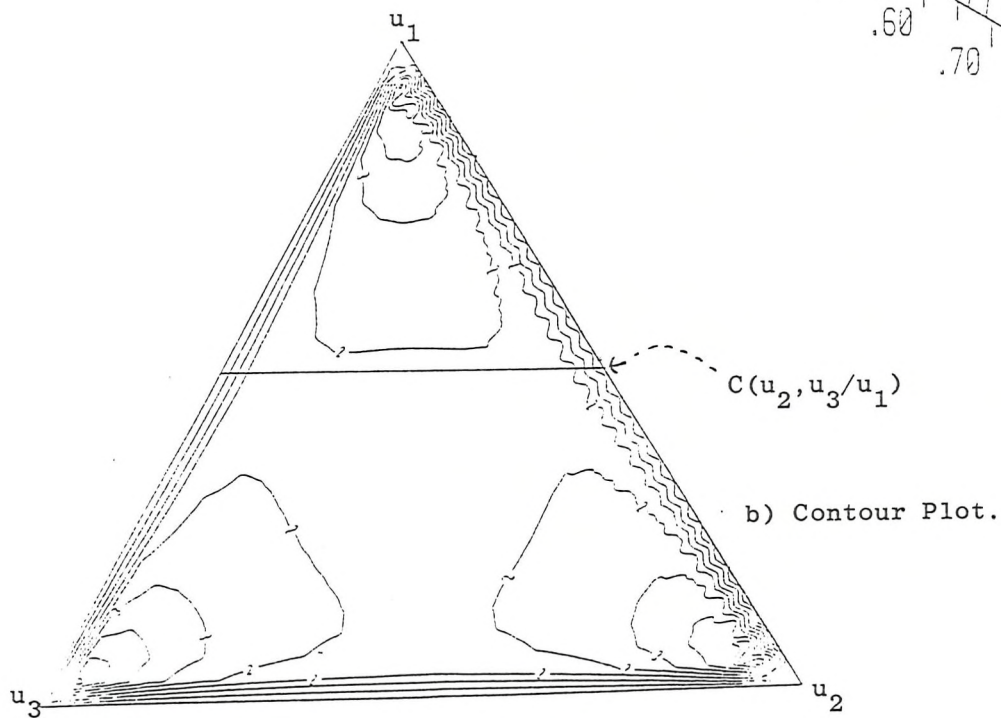
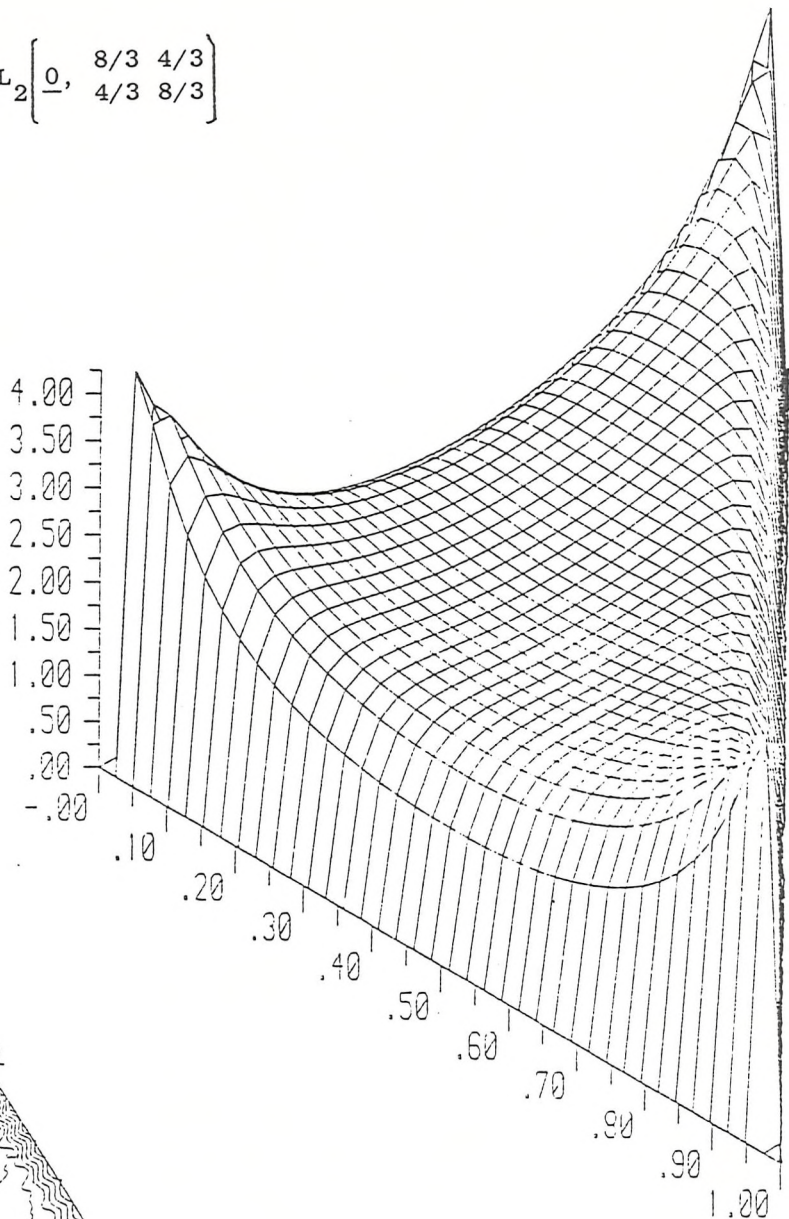


Figure 5.10 Plot of $L_2 \begin{bmatrix} 0, & 16/9 & 8/9 \\ 0, & 8/9 & 16/9 \end{bmatrix}$

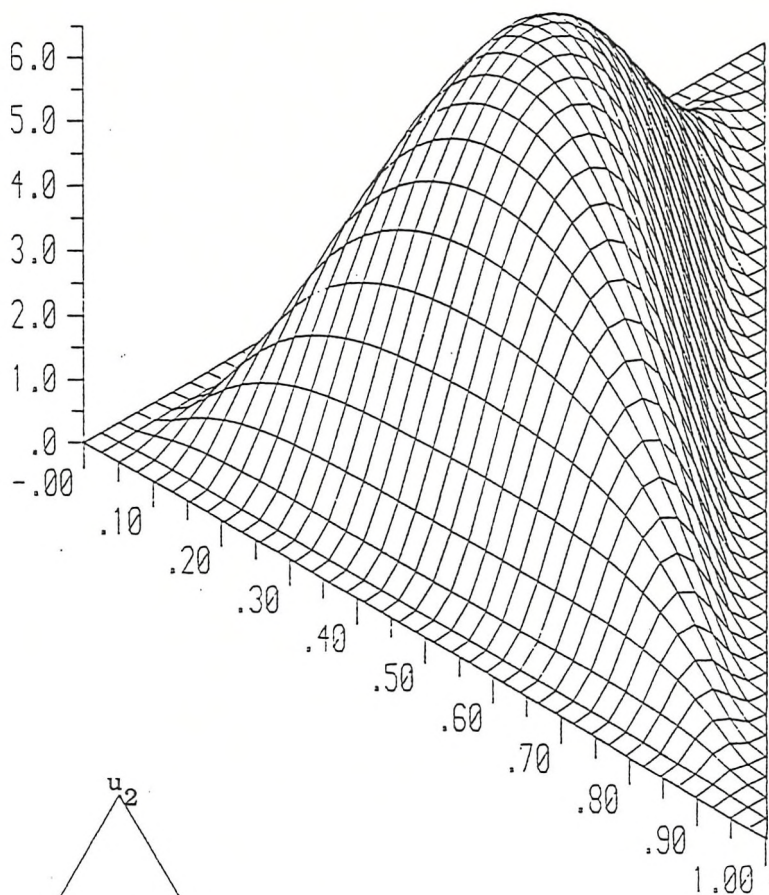
Figure 5.11 Plot of $L_2 \left[\begin{matrix} 0, & 8/3 & 4/3 \\ 4/3 & 8/3 \end{matrix} \right]$

a) 3-d Plot.



b) Contour Plot.

a) 3-d Plot.



b) Contour Plot.

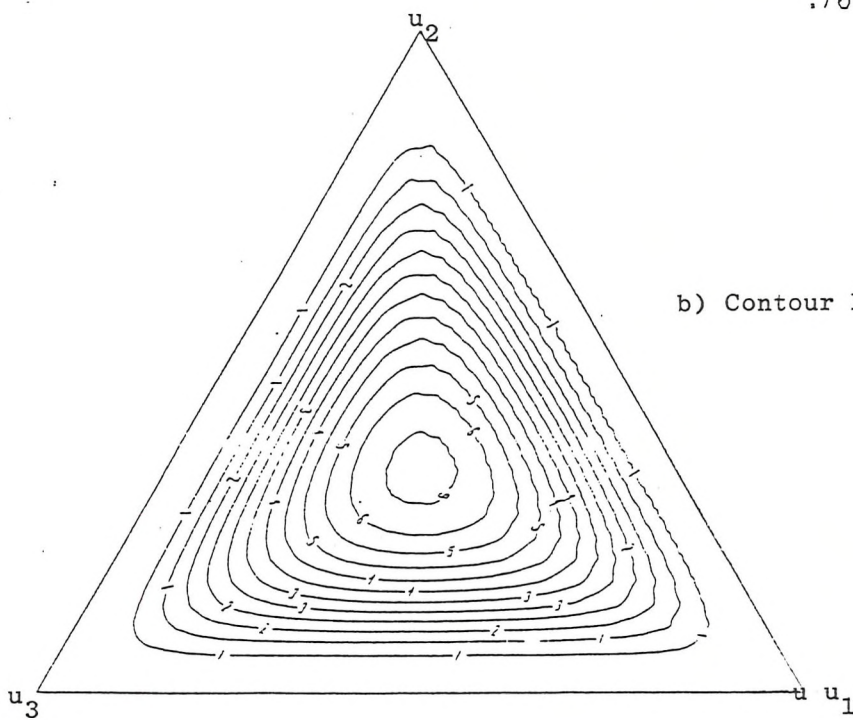
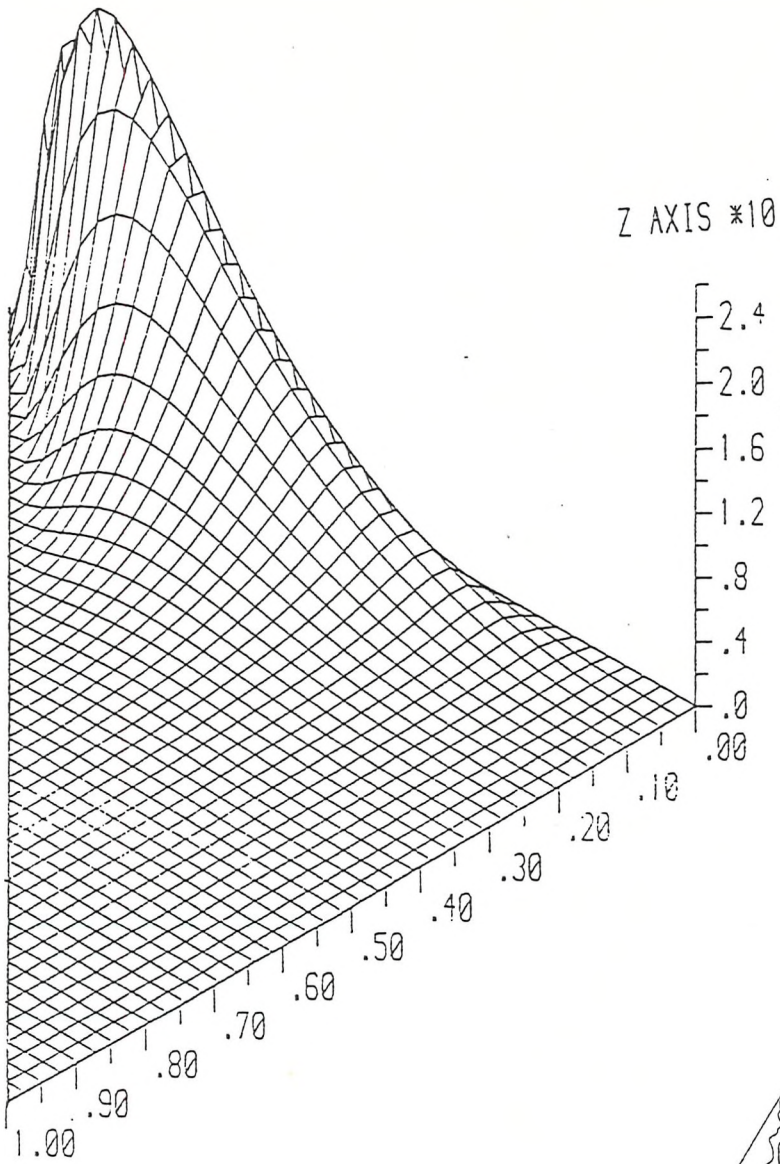


Figure 5.12 Plot of $L_2 \begin{bmatrix} 0 & 0.8 & 0.4 \\ 0.4 & 0.8 & 0 \end{bmatrix}$ distribution.



a) 3-d Plot.

b) Contour Plot.

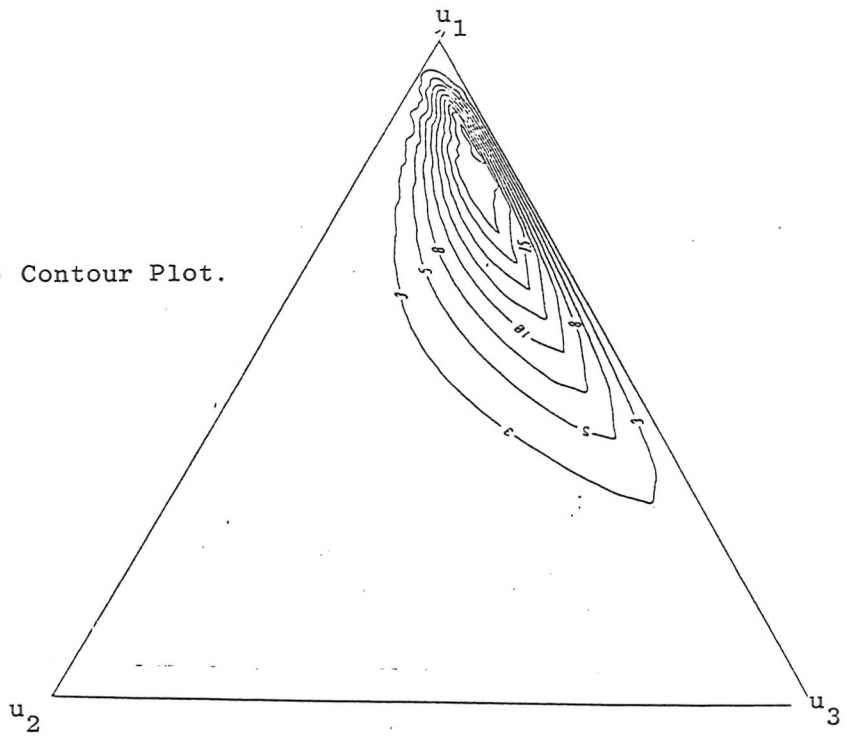
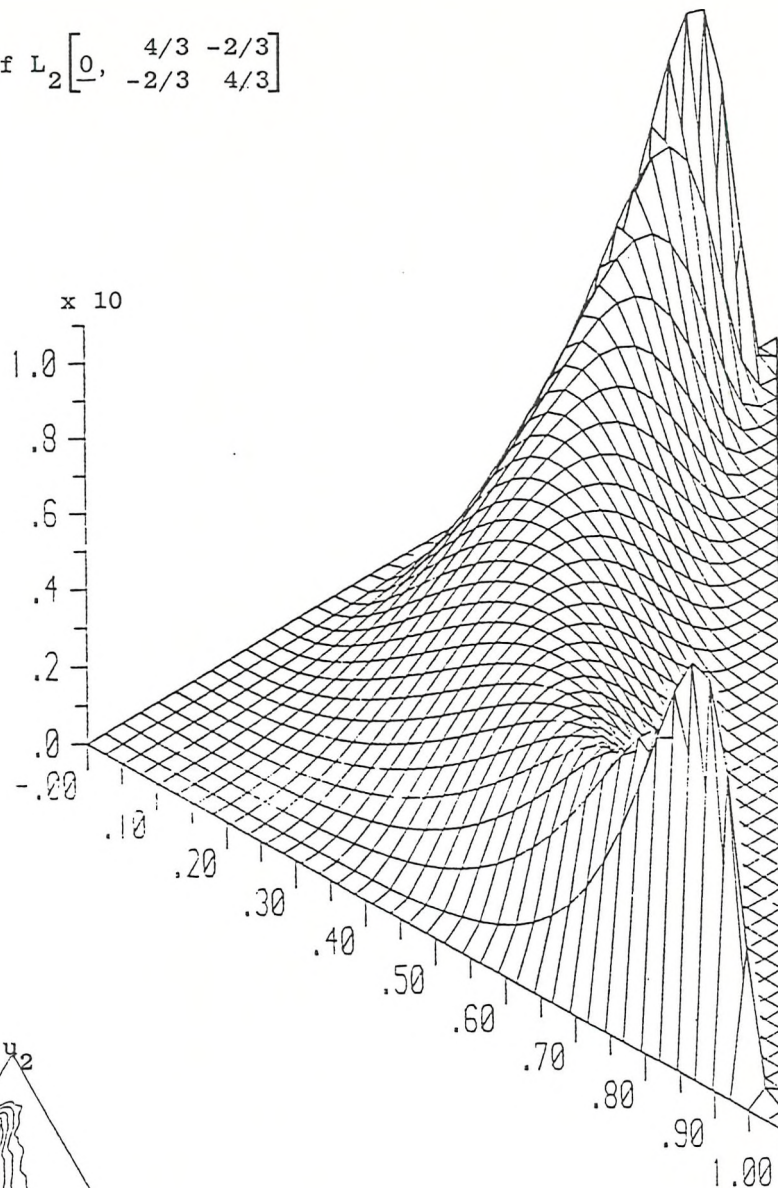


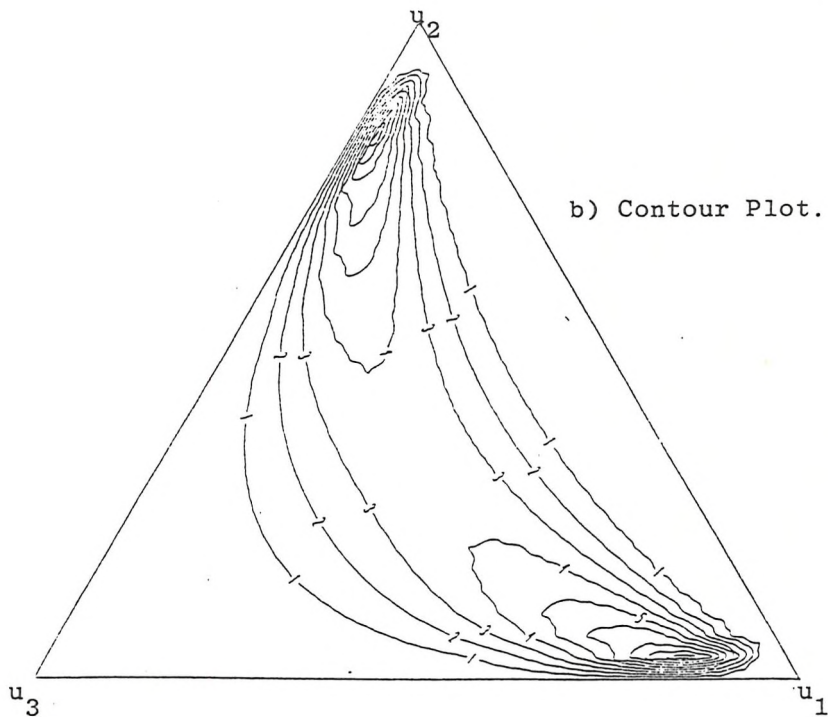
Figure 5.13 Plot of $L_2 \begin{bmatrix} 1.0 & 0.8 & 0.4 \\ -1.0 & 0.4 & 0.8 \end{bmatrix}$

Figure 5.14 Plot of $L_2 \begin{bmatrix} 0 & 4/3 & -2/3 \\ -2/3 & 4/3 & 0 \end{bmatrix}$

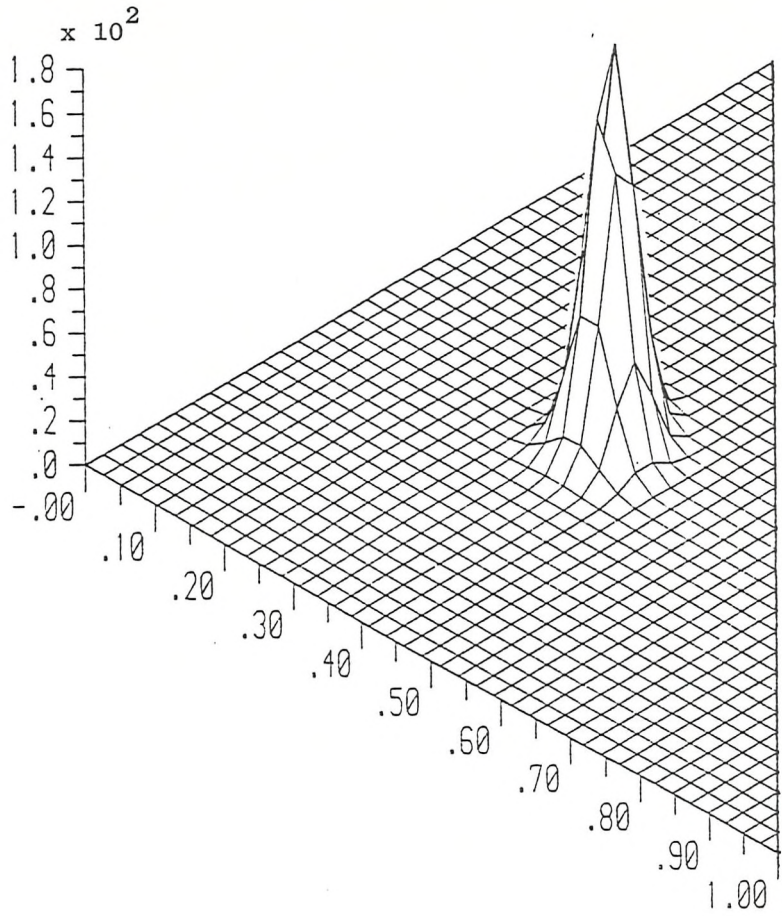
a) 3-d Plot.



b) Contour Plot.



a) 3-d Plot.



b) Contour Plot.

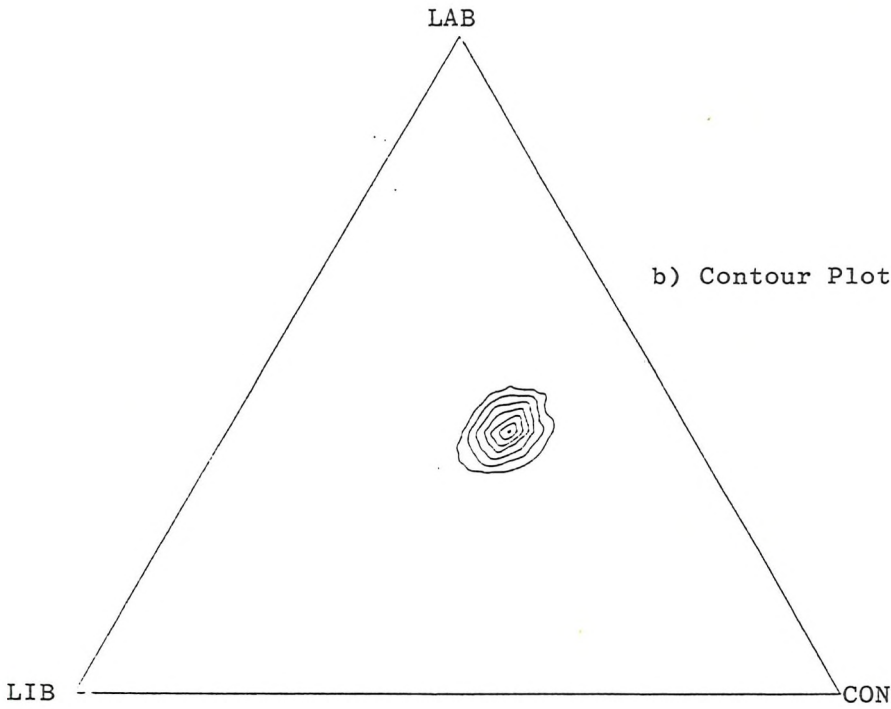


Figure 5.15 Plot of $L_2 \begin{bmatrix} 0.4272 & 0.0413 & 0.0351 \\ 0.5309 & 0.0351 & 0.0452 \end{bmatrix}$, i.e. the one-step ahead forecast distribution of the GALLUP(C) Series.

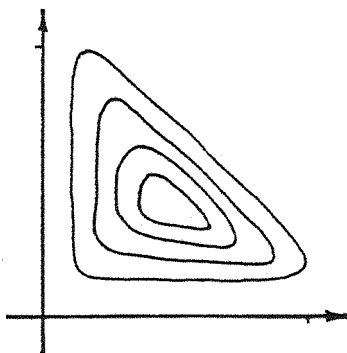
and so on. Initially only the L_2 distribution was studied and contour plots and 3-d plots such as those in figures 5.9-5.15 were produced. The first distribution considered was that based on the standard normal distribution, i.e. $L_2(\underline{0}, \underline{I}_m)$ as in figure 5.9. Some others were chosen to be symmetrical in all variables. For the $L_2(\underline{\mu}, \underline{\Sigma})$ distribution to be symmetrical we require that,

$$\underline{\mu} = \underline{0} \quad \text{and} \quad \underline{\Sigma} = \begin{bmatrix} 2a & a \\ a & 2a \end{bmatrix}, \text{ for some constant } a. \quad (5.2.50)$$

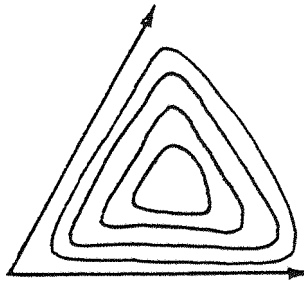
Figure 5.10 represents the case when $a = 8/9$, which from (5.2.32) is the limiting case. Figures 5.11-5.14 are a selection of tri-modal, bi-modal, uni-modal, skew and symmetrical distributions. Figure 5.15 is the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution where $\underline{\mu}$ is the one-step-ahead prediction of \underline{v}_t for the GALLUP poll series studied in section 4.4 and $\underline{\Sigma}$ is the corresponding covariance matrix.

In addition to the distributions plotted a few other L_2 distributions were studied, together with a few L_3 distributions. These were chosen to be either slight variations of the symmetrical situation (e.g. by varying μ_1 so that $\mu_1 \neq 0$), or in the L_3 case the symmetrical distribution itself. Only uni-modal L_3 distributions were studied since it is difficult to locate the modes without the aid of a graphical representation, and consequently difficult to evaluate these modes numerically.

The contour plots of the $L_2(\underline{\mu}, \underline{\Sigma})$ distribution were produced by the GINO package. The computer algorithm in fact produces a plot of the form,



which is then "skewed" and scaled to give :-



The final diagonal has been drawn in. It should be noted that as a result of this together with rounding errors, and interpolation between points etc, the plot near the diagonal is slightly distorted. However this distortion is not so great as to lose the basic shape of the distribution displayed.

In addition to the graphical plots various parameters were computed for each $L_m(\underline{\mu}, \underline{\Sigma})$ distribution. These were,

- i) Mean $E[u_i]$.
- ii) Second moment $E[u_i^2]$.
- iii) Variance $v[u_i]$
- iv) Estimated mean $E[\hat{u}_i]$ using $\sigma = 2.942$
- v) Estimated mean $E[\hat{u}_i]$ using $\sigma = 3.000$
- vi) Estimated mean $E[\hat{u}_i]$ using $\sigma = 3.290$
- vii) Inverse of $\underline{\mu}$ (\underline{i} say),

} using lemma 5.8.

$$i_k = \frac{e^{\mu_k}}{1 + \sum_{j=1}^m e^{\mu_j}}, \quad k=1, \dots, m;$$

$$= \frac{1}{1 + \sum_{j=1}^m e^{\mu_j}}, \quad k=m+1.$$

viii) Mode or modes \underline{m}^K (say), where $\kappa = 1, \dots, K$; and $K =$ number of modes.

ix) Maximum value of the $L_m(\underline{\mu}, \underline{\Sigma})$ density function, \max^K i.e. value at the mode(s) listed in (viii).

Table 5.16 : Parameter Values for Some $L_m(\mu, \Sigma)$ Distributions

Models Considered are :-

- a) $L_2(0, I_2)$ b) $L_2\left[0, \begin{matrix} 8/3 & 4/3 \\ 4/3 & 8/3 \end{matrix}\right]$ c) $L_2\left[0, \begin{matrix} 0.8 & 0.4 \\ 0.4 & 0.8 \end{matrix}\right]$
- d) $L_2\left[\begin{matrix} 1.0 & 0.8 & 0.4 \\ -1.0 & 0.4 & 0.8 \end{matrix}\right]$ e) $L_2\left[0, \begin{matrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{matrix}\right]$
- f) $L_2\left[\begin{matrix} 0.42720 & 0.04130 & 0.03510 \\ 0.53090 & 0.03510 & 0.04520 \end{matrix}\right]$ (one-step ahead forecast distribution for GALLUP(c)) g) $L_2\left[\begin{matrix} 1.0 \\ -1.0 \end{matrix}, I_2\right]$

Model	a_m^{-1}	Mode(s) (Height at Mode)	Estimated Mean (With $\sigma =$)			Mean	Variance	
			(2.94 3.00 3.29)				$E[u_i^2]$	\downarrow
a)	3333	3333	3446	3444	3437	3454	1626	0433
	3333	3333 (4.2972)	3446	3444	3437	3454	1626	0433
	3333	3333	3109	3112	3126	3092	1169	0213
b)	3333	0230 0230 9539	3324	3324	3325	3333	1481	0369
	3333	0230 9539 0230	3324	3324	3325	3333	1481	0369
	3333	9539 0230 0230 (All 4.2526)	3333	3333	3333	3333	1481	0369
c)	3333	3333	3339	3339	3339	3333	1357	0246
	3333	3333 (6.2050)	3339	3339	3339	3333	1357	0246
	3333	3333	3333	3333	3333	3333	1357	0246
d)	6652	8241	6402	6382	6652	6311	4260	0277
	0900	0437 (24.9993)	0919	0931	0900	1044	0166	0057
	2447	1322	2685	2694	2447	2645	0918	0219
e)	3333	0031 9559	3608	3604	3588	3614	1953	0647
	3333	9559 0031	3608	3604	3588	3614	1953	0647
	3333	0409 0409 (Both 16.5708)	2801	2808	2839	2772	0934	0166
f)	3621	3628	3658	3655	3643	3614	1314	0008
	4017	4031 (184.1880)	4066	4061	4028	4009	1618	0010
	2362	2341	2316	2325	2366	2377	0578	0013
g)	3672	6925	3728	3725	3709	3726	1840	0452
	3006	0942 (4.9486)	3169	3169	3170	3188	1427	0411
	3322	2133	3103	3106	3106	3086	1165	0213

$\times 10^{-4}$ (except Height at Mode).

Table is continued on the next page.

Table 5.16 ...continued

Models are :-

$$h) L_2 \begin{bmatrix} 0.0 & 1 \frac{1}{2} \\ -0.1 & \frac{1}{2} 1 \end{bmatrix} \quad i) L_2 \begin{bmatrix} 0.1 & 1 -\frac{1}{2} \\ -0.1 & -\frac{1}{2} 1 \end{bmatrix} \quad j) L_3(0, \underline{I}_3)$$

$$k) L_3 \begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \end{bmatrix}, \underline{I}_3 \quad l) L_3 \begin{bmatrix} 1 \frac{1}{2} \frac{1}{2} \\ \frac{1}{2} 1 \frac{1}{2} \\ \frac{1}{2} \frac{1}{2} 1 \end{bmatrix} \quad m) L_3 \begin{bmatrix} 0.1 & 1 \frac{1}{2} \frac{1}{2} \\ -0.1 & \frac{1}{2} 1 \frac{1}{2} \\ 0.1 & \frac{1}{2} \frac{1}{2} 1 \end{bmatrix}$$

Model	a_m^{-1}	Mode(s) (Height at Mode)	Estimated Mean (With $\sigma =$)			Mean	Variance	
			(2.94 3.00 3.29)				$E[u_1^2]$	\downarrow
h)	3443	3541	3433	3432	3429	3428	1478	0302
	3115	2918 (4.9937)	3135	3136	3143	3143	1269	0281
	3443	3541	3433	3432	3429	3428	1478	0302
i)	3672	8925	3818	3814	3793	3825	2021	0558
	3006	0140 (10.8856)	3290	3289	3282	3315	1617	0518
	3322	0935	2891	2897	2925	2860	0948	0130
j)	2500	2500	2612	2611	2604	2610	1023	0342
	2500	2500 (16.2544)	2612	2611	2604	2610	1023	0342.
	2500	2500	2612	2611	2604	2610	1023	0342
	2500	2500	2163	2168	2189	2171	0576	0342
k)	2756	8896	2844	2841	2827	2824	1166	0368
	2256	0227 (33.2623)	2388	2388	2388	2402	0892	0315
	2494	0254	2608	2607	2600	2606	1021	0342
	2494	0623	2159	2164	2186	2168	0574	0105
l)	2500	2500	2500	2500	2500	2500	0828	0203
	2500	2500 (22.9872)	2476	2477	2477	2500	0828	0203
	2500	2500	2525	2524	2523	2500	0828	0203
	2500	2500	2500	2500	2500	2500	0828	0203
m)	2756	3057	2749	2747	2738	2727	0967	0224
	2256	2044 (23.4546)	2239	2241	2250	2282	0704	0183
	2494	2450	2519	2519	2518	2495	0826	0203
	2494	2450	2495	2495	2496	2495	0826	0203

$\times 10^{-4}$ (except Height at Mode).

The results are presented in table 5.16.

In the calculation of the mean and variance Gauss-Hermite quadrature was used. This type of quadrature is used to approximate an integral of the form $\int_{-\infty}^{\infty} f(x) dx$ and is known to give exact solutions if the function is of the form,

$$f(x) = e^{-b(x-a)^2} \sum_{i=0}^{2n-1} c_i x^i, \quad b > 0.$$

For the multiple integrals the quadrature formula is applied to each dimension in turn. We employed a NAG library routine to do this.

In our case we need to find the integral over \mathbb{R}^m of the function $a_m^{-1}(v_i) = e^{-\frac{v_i}{1 + \sum_{j=1}^m v_j}}$ ($i=1, \dots, m$) multiplied by the density function of a $N_m(\underline{\mu}, \underline{\Sigma})$ distribution. Thus if we regard the normal density function as resembling $e^{-b(x-a)^2}$, then in very loose terms a good approximation will be given when,

$$a_m^{-1}(v_i) \approx \sum_{k=0}^{2n} c_k v^k \quad ; \quad i=1, \dots, m+1 ; (v_{m+1} = 1).$$

Or in the case of evaluating $E[u_i^2]$ when,

$$\left[a_m^{-1}(v_i) \right]^2 \approx \sum_{k=0}^{2n} c_k v^k \quad ; \quad i=1, \dots, m ; (v_{m+1} = 1).$$

The NAG procedure in fact is iterative and increases n at each step through the following values: 4, 5, 6, 8, 10, 12, 14, 16, 20, 24, 32, 48, 64. It was found that at most six iterations ($n=12$) were required to obtain an accuracy of order 10^{-6} . $\text{Var}[u_i]$ was computed from $E[u_i]$ and $E[u_i^2]$. We note that we could have easily computed $\text{Cov}[u_i, u_j]$ $i \neq j$ in a similar manner although we have chosen not to do so.

When it came to evaluating the estimated mean $(iv-v_i)$ above, it was found that tables were unnecessarily complicated and that it was less laborious to evaluate the cumulative normal density numerically. This required the use of a NAG routine to evaluate the integral over the negative orthant of \mathbb{R}^m of the $N_m(-\underline{\mu}(k), \underline{\Sigma}(k) + \underline{\Sigma}_1)$ function, for which we require the use of Gauss-Laguerre quadrature. This latter quadrature

is used to estimate integrals of the form, $\int_{-\infty}^a f(x) dx$, and is known to be exact if $f(x)$ is of the form :

$$f(x) = e^{-bx} \sum_{i=0}^{2n-1} c_i x^i .$$

We note that in this type of quadrature the exponent no longer has a quadratic form which may be related to the normal density function. This implies that we would expect less accuracy from the numerical algorithm in the estimation of iv)-vi) compared to i)-iii). Although a more detailed scrutiny could enable us to see just how close our functions are to those preferred for the quadrature procedure, the hint given by the exponent term perhaps explains why more iterations were required to evaluate the estimated means. In fact up to 12 iterations ($n=48$) were required of the Gauss-Laguerre quadrature to compute iv)-vi) compared to the previous 6. (This is to the same accuracy of 10^{-6}).

The computation of the estimated mean values was further lengthened by the need to evaluate $\mu_{(k)}$ and $\Sigma_{(k)}$ and the determinant and inverse for each variable. (In the evaluation of the exact mean and variance, μ and Σ remain fixed for each variable.) Thus the approximation to the mean required more computation than that for the exact value. This was reflected by the length of the respective SUBROUTINE codes developed to evaluate them.

This lengthy computation of the approximate mean value is due, however, to the particular algorithm used to evaluate the cumulative density of the multivariate normal. It may be that alternatives are available. Indeed if a user has access only to a statistical package which evaluates this cumulative density, but not to a general numerical analysis package, then the approximate mean will still prove useful. Otherwise $E[u_i]$ is only of academic interest, since the exact value is more easily computerized, and more accurate. The only exception to this is perhaps the univariate case.

If we examine the results themselves, the problem in computing the estimated mean values is again highlighted. For example, in the second and third models the approximation to the mean should have given us the same answer as the exact mean, but the numerical algorithm has failed to converge to the known approximate value. The value of the exact means are, however, correct. (See also the 12th model). However, the results appear to be accurate to at least two decimal places and still provide

us with a rough yard stick for comparison.

Throughout it is clear that the approximation to the mean is reasonably good. It is difficult to determine which value of σ should be used to estimate it. This is partly due to the limited accuracy described above, but also because there are occasions when each of the choices of σ gives a better result. A more extensive and accurate study may aid the choice of σ , but meanwhile it would seem sensible to let $\sigma = 3.0$ for simplicity.

In many of the examples in table 5.16 there is a marked difference between the various location parameters especially when the distribution is not uni-modal and/or when it is skewed. Although this is perhaps obvious is is none the less a reminder that care needs to be exercised in the choice of which location parameters to use as point-predictors. It should be noted, however, that in the forecasting context the occurrence of a multi-modal distribution is likely to be rare. For this to occur we would require the u_{it} 's to be close to the extreme values (i.e. $u_{it} \approx 1$, $u_{jt} \approx 0$, $j=1, \dots, m+1$, $j \neq i$), and for them to "jump" from one extreme to another (i.e. if when $t=1$ $i=i_1$, then at time $t=2$ $i=i_2$ (say)). This will be further examined in section 5.5.

As an example of "well behaved data" consider figure 5.17, which is a plot of the opinion poll data analysed in section 4.4. The points are all mid-way between Conservative and Labour and only have small values for Liberal. This data is relatively tightly packed, and no points are close to the apexes of the triangle. It is therefore expected that the varying point predictors for this data will give similar results. This is readily seen to be the case since the 6th L_m distribution of table 5.16 which corresponds to figure 5.15 is the distribution of the one-step-ahead forecast of the GALLUP(c) series.

If we examine the plots we see several interesting points. The $L_2(\underline{0}, \underline{I}_2)$ distribution in Figure 5.9 is symmetrical about the line $u_2 = u_3$. Figures 5.10-5.12 are all symmetrical since $\underline{\mu} = \underline{0}$ and the covariance matrix is of the form (5.2.50). Figure 5.10 is (from (5.2.33) and (5.2.34)) the limiting case between uni-modalality and multi-modality,

and consequently is very flat. $L_2 \left[\underline{0}, \begin{bmatrix} 8/3 & 4/3 \\ 4/3 & 8/3 \end{bmatrix} \right]$ in Figure 5.11 is

tri-modal. Comparing this with our "rule-of-thumb" neither (5.2.33) nor

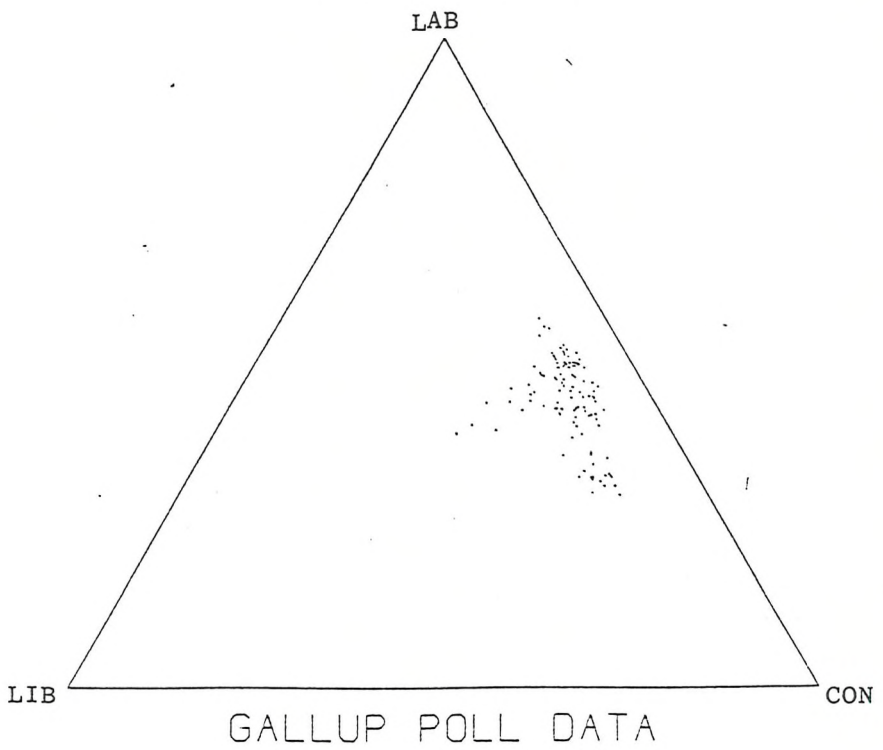
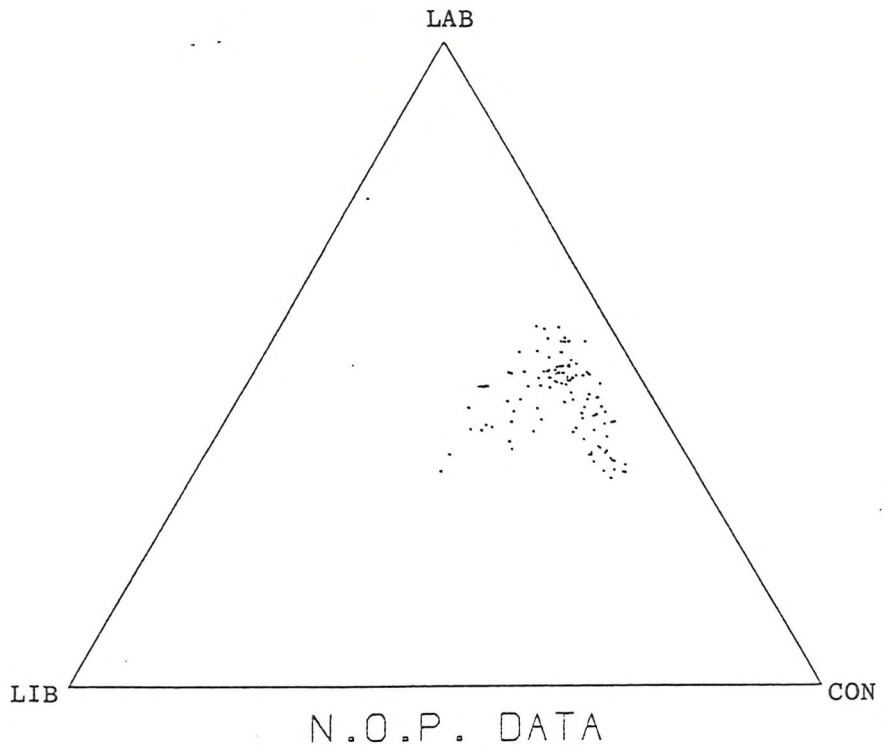


Figure 5.17 Plot of Opinion Poll Data.

(5.2.34) hold, $\left[\begin{bmatrix} 8/3 & 4/3 \\ 4/3 & 8/3 \end{bmatrix}^{-1} = \begin{bmatrix} 0.5 & -0.25 \\ -0.25 & 0.5 \end{bmatrix} \right]$ and since $\underline{\mu} = \underline{0}$ this is

what we would expect. Figure 5.12 is uni-modal, and its symmetrical structure gives the characteristic bell shape. The distribution in Figure 5.13 has the value of $\underline{\Sigma}$ as does that in figure 5.12, but now $\underline{\mu} \neq \underline{0}$ resulting in the mass of the density functions being shifted to one apex. Figure 5.14 is an example of a bi-modal L_2 density function. We note that for this density function,

$$\underline{\Sigma} = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}, \quad \underline{\Sigma}_{(1)} = \begin{bmatrix} 4/3 & 2 \\ 2 & 4 \end{bmatrix}, \quad \underline{\Sigma}_{(2)} = \begin{bmatrix} 4 & 2 \\ 2 & 4/3 \end{bmatrix}, \text{ and}$$

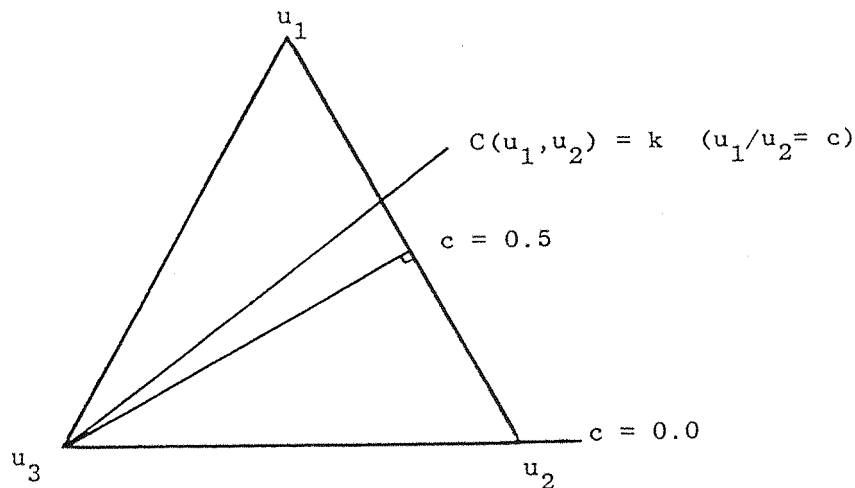
$$\underline{\Sigma}^{-1} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}, \quad \underline{\Sigma}_{(1)}^{-1} = \begin{bmatrix} 3 & -1.5 \\ -1.5 & 1 \end{bmatrix}, \quad \underline{\Sigma}_{(2)}^{-1} = \begin{bmatrix} 1 & -1.5 \\ -1.5 & 1 \end{bmatrix}.$$

Comparing these with (5.2.33) and (5.2.34) we note that,

$$\sigma_{22}^{(1)} = 4 > \frac{16}{9}, \quad \sigma_{11}^{(2)} > \frac{16}{9}; \quad \sigma_{12}^{(1)} = \sigma_{12}^{(2)} = 2 > \frac{8}{9}; \quad \text{and } \sigma^{12} = 0.5 > -0.375.$$

Hence under both "rules-of-thumb" there is some evidence for a multi-modal distribution. However the evidence is not conclusive, and illustrates the difficulty in finding a rule for multivariate distributions.

One might hope to get some idea of the conditions for which these L_2 distributions are not uni-modal by considering their conditional distributions. For example, in figure 5.11a a line is marked which represents $C(u_2, u_3)/u_1 = 0.5$. At this point the conditional distribution is uni-modal, however for a value of u_1 close to 1.0 the marginal distribution is bi-modal. If this examination were repeated for $C(u_1, u_2)/u_3$ and $C(u_1, u_3)/u_2$ we would be able to deduce an over all tri-modal distribution. However, from section 5.1.2 the only conditional distribution that is tractable and results in an L_1 distribution is of the form $C(u_2, u_3)/C(u_1, u_2)$. For this conditional distribution, however, the locus of the line passes through the apex of the triangle at u_3 .



Consequently it does not ever seem to become bi-modal. It may be possible to determine something about the multi-modal distribution by examining the location of the mode as the conditional distribution is allowed to vary across k . This approach does seem rather unsystematic and so we do not pursue it here.

In conclusion we see that three possible location parameters are available which may be used as point predictors: the mean, the mode and the inverse of the logistic transformation. The choice of which one to use will depend on other factors such as whether the distribution is uni-modal, and the variances of the prediction errors for \underline{v}_t . In the situation when the distribution is multi-modal, it may be more sensible to use the modes. This problem will be further addressed in section 5.5. When the distribution is uni-modal, the mean is the point predictor best suited to most purposes since it provides the MMSE forecast. However when the distribution of the prediction is tightly packed, as in the one-step-ahead forecast for the GALLUP poll series (figure 5.15) it is possible to use the inverse of the a_m transformation as an approximation to the mean. Table 5.16 gives an indication of when the inverse is close to the mean. In the fully symmetrical case it equals the mean. It does well when the $\underline{\Sigma}$ parameter is of the form (5.2.29), and less well otherwise. This can be seen by comparing the L_2 and L_3 distributions labelled d, h and m with those labelled e.g. i and m. The first group have $\underline{\Sigma}$ of the form (5.2.29) whilst the second do not. All have non-zero $\underline{\mu}$. The mean of the first group is more closely approximated by a_m^{-1} in the first group than in the second.

In general it is suggested that the rule-of-thumb described earlier by equations (5.2.33) and (5.2.34) be applied to indicate uni-

modality. If more than one mode is present, which predictor to use will be a subjective matter, depending on the application. In the uni-modal case the mean may be easily calculated numerically, or approximated using $\sigma=3$. If it is clear that the prediction distribution is dense around one area then a_m^{-1} may be used. This will be the case when the elements of $\underline{\Sigma}$ are small, where $\underline{\Sigma}$ is the prediction error covariance matrix for the appropriate forecast \underline{v}_t . The decision to use a_m^{-1} will depend both on the size of these elements (e.g. in figure 5.15 these are of order 10^{-2}), and on the accuracy of the forecast required.

5.3 Some Examples of the $M_m(\underline{\mu}, \underline{\Sigma})$ Distribution

It is possible to carry out a similar investigation to that of the previous section for the $M_m(\underline{\mu}, \underline{\Sigma})$ distribution. However, as we require limited use of this distribution, for brevity we do not do so here. Nevertheless we produce a few plots of the $M_m(\underline{\mu}, \underline{\Sigma})$ class. Three examples of the $M_m(\underline{\mu}, \underline{\Sigma})$ distribution are represented in figures 5.18-5.20. These should help to illustrate a few of the properties of the multiplicative logistic density function.

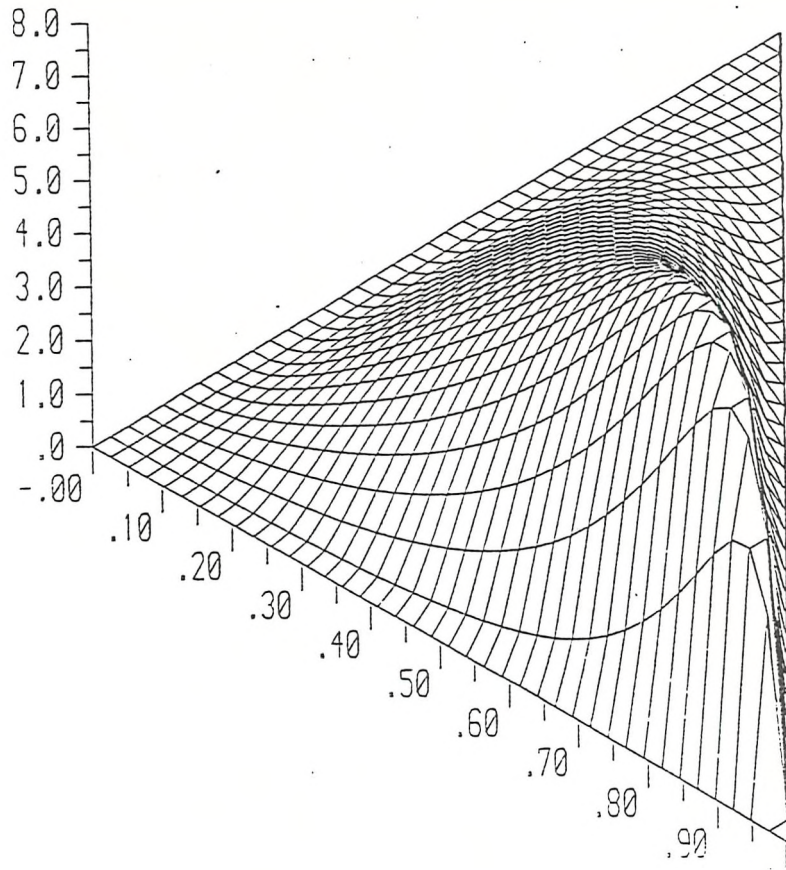
The first figure illustrates the $M_2(\underline{0}, \underline{I}_2)$ distribution. As with the $L_2(\underline{0}, \underline{I}_2)$ distribution it is symmetrical between only two variables, but whereas before the symmetry was between u_1 and u_2 , it is now between u_2 and u_3 . To examine this recall that for $\underline{u} \in \mathbb{S}^m$, $\underline{u} \sim M_m(\underline{\mu}, \underline{\Sigma})$, then $\underline{v} = m_m(\underline{u})$, where $\underline{v} \sim N_m(\underline{\mu}, \underline{\Sigma})$. In this instance ($m=2$), $\underline{v} = m_2(\underline{u})$, i.e.

$$v_1 = \ln \frac{u_1}{1 - u_1} \quad ,$$

$$v_2 = \ln \frac{u_2}{1 - u_1 - u_2} = \ln \frac{u_2}{u_3} = -\ln \frac{u_3}{u_2} \quad .$$

Hence v_2 may be thought of as the log-ratio of u_2 with u_3 . If there is no correlation with v_1 , and with zero $\underline{\mu}$ the distribution will be symmetrical in u_2 and u_3 . This illustrates that despite the difficulty in permuting the elements of \underline{u} when comparing M_m distributions on the same data set, it is none the less possible to interchange the m^{th} and $m+1^{\text{th}}$ element. For the L_2 distribution we recall that $v_1 = \ln(u_1/u_3)$ and $v_2 = \ln(u_2/u_3)$, so that both u_1 and u_2 are referenced by u_3 . Hence if v_1 and v_2 have identical means and variances, u_1 and u_2 will be symmetrical in the resulting L_m distribution.

a) 3-d Plot.



b) Contour Plot.

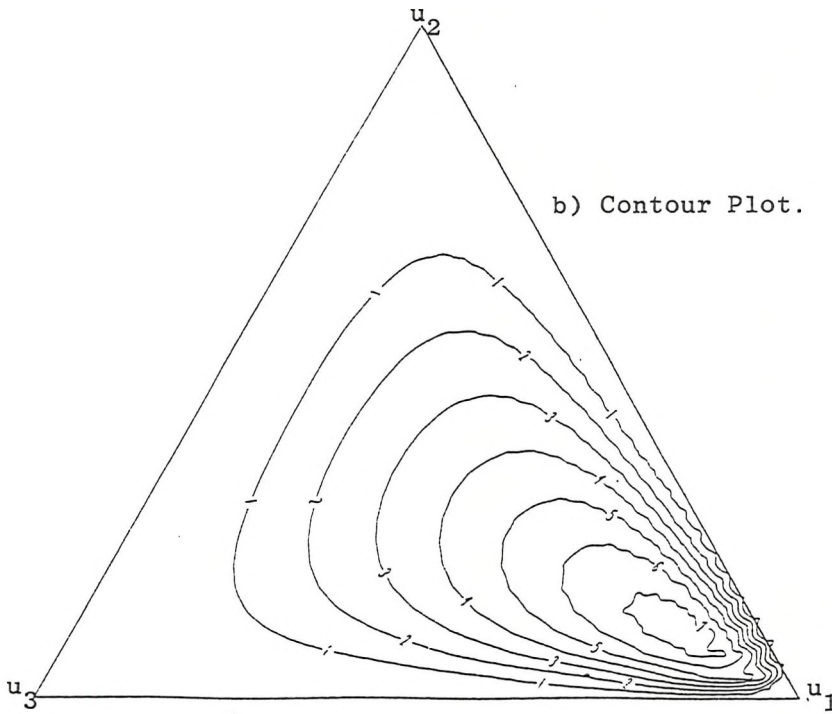
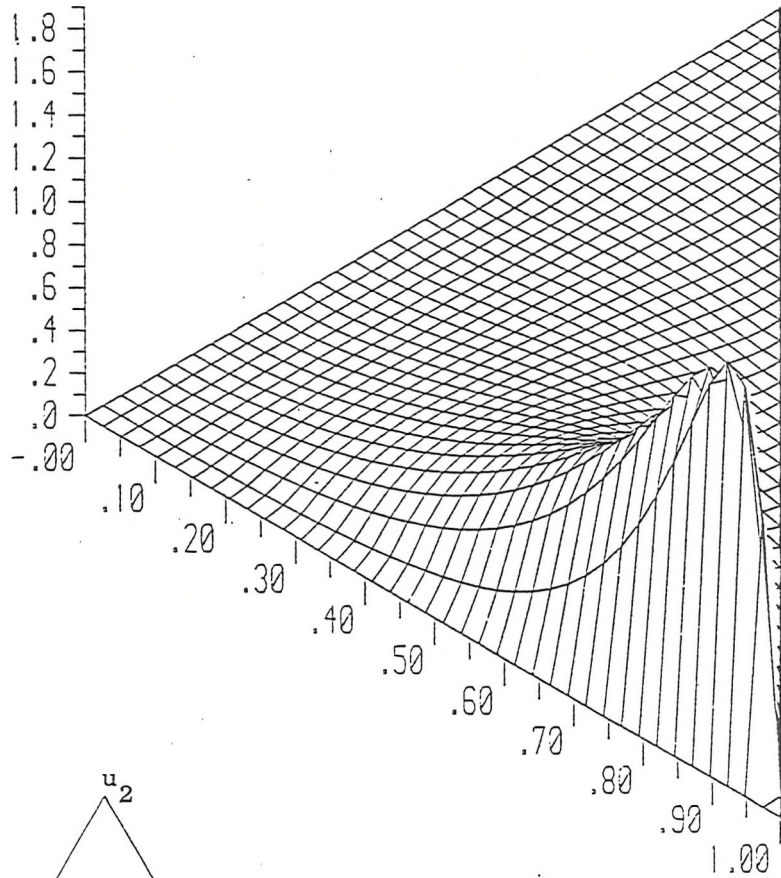


Figure 5.18 Plot of the $M_2(0, I_2)$ Distribution.

a) 3-d Plot.



b) Contour Plot.

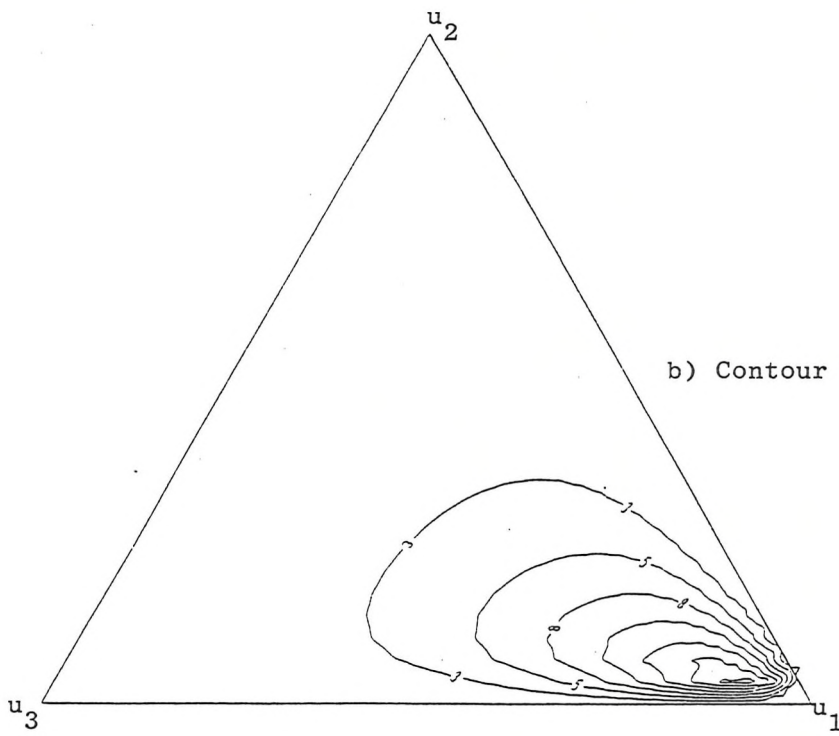
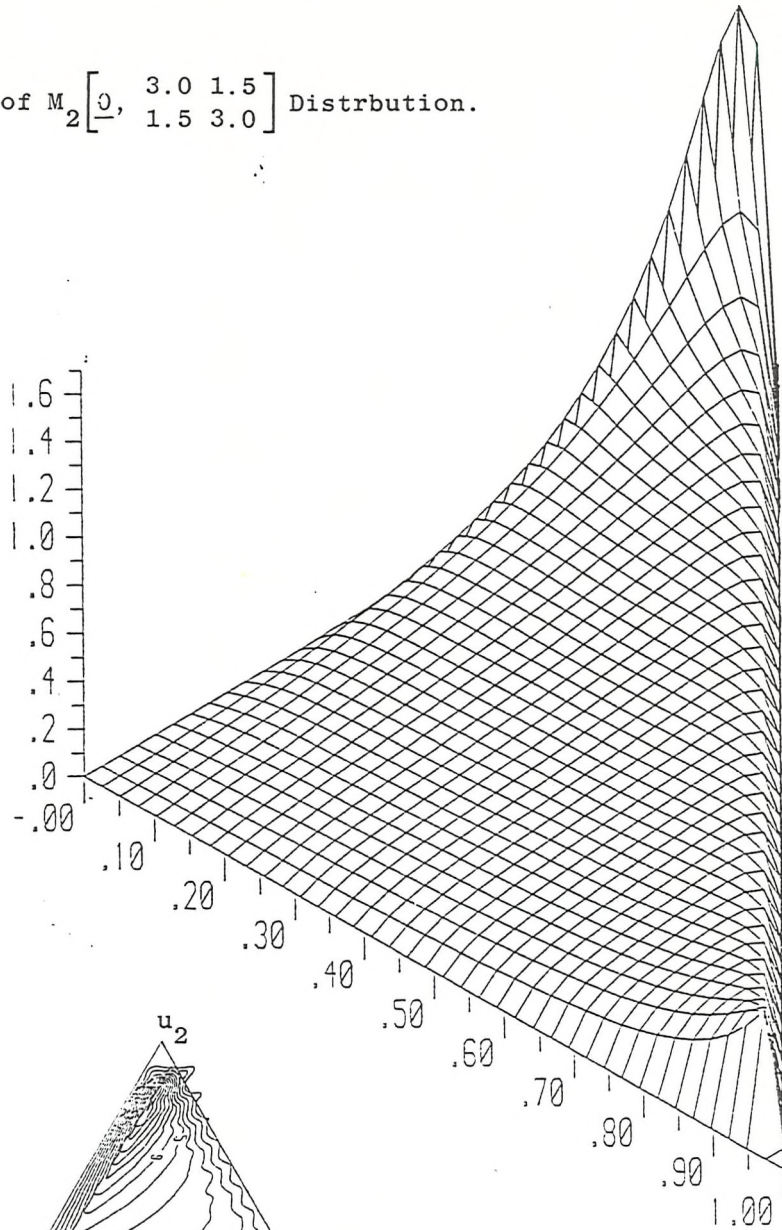


Figure 5.19 Plot of $M_2 \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}, I_2$ Distribution.

Figure 5.20 Plot of $M_2 \begin{bmatrix} 0 & 3.0 & 1.5 \\ 1.5 & 3.0 & 0 \end{bmatrix}$ Distribution.

a) 3-d Plot.



b) Contour Plot.

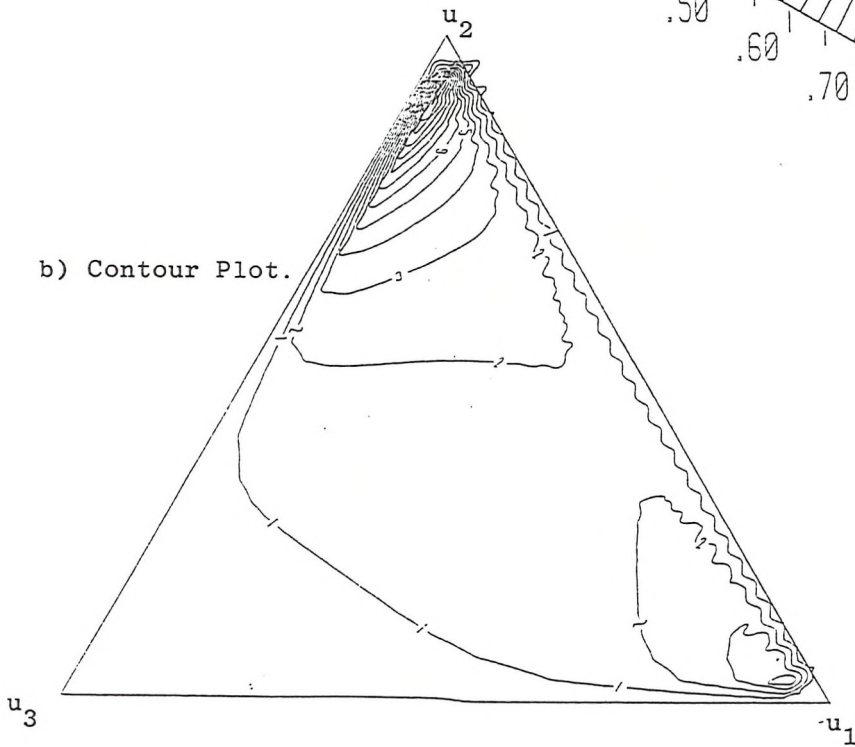


Figure 5.19 illustrates the M_m distribution for $\mu \neq 0$. As in figure 5.13 most of the mass is close to $u_1 = 1$, but in this case it is more pronounced. The final figure, figure 5.20 illustrates a bi-modal density function. This example has one mode higher than the other. It illustrates that care must again be exercised in choosing an appropriate point predictor, should the $\ln_x \text{ARMA}_m(p,q)$ model be used for forecasting.

A final characteristic of the M_m distribution is that it is not easy to construct a parameterization that will produce a fully symmetric distribution. This is indicative of the order-dependence of the $M_m(\underline{u})$ transformation.

5.4 Confidence Region Based on the Logistic-Normal Distribution

Often a time series analyst is not only interested in producing a single valued forecast, but in obtaining a region in which the future value lies. This region defines an interval predictor. Aitchison and Shen(1984) point out that a confidence region is easily obtained for a sample of \underline{u} 's $\in \mathbb{S}^m$ using standard multivariate normal distribution theory. Consider n independent estimates of a composition \underline{u} ; u_i , $i=1, \dots, n$. A $100(1-\alpha)\%$ confidence region for \underline{u} , where $u_i \sim L_m(\mu, \Sigma)$ is :-

$$n \left[\bar{\underline{v}} - \ln \frac{\underline{u}}{u_{m+1}} \right] S_{\underline{v}}^{-1} \left[\bar{\underline{v}} - \ln \frac{\underline{u}}{u_{m+1}} \right] \leq \frac{(n-1)m}{n-m} F_{\alpha; m, n-m} \quad (5.4.1)$$

where $\underline{v} = a_m(\underline{u})$, $\bar{\underline{v}} = \frac{1}{n} \sum_{i=1}^n \underline{v}_i$, $S_{\underline{v}} = \frac{1}{n-1} \sum_{i=1}^n (\underline{v}_i - \bar{\underline{v}})(\underline{v}_i - \bar{\underline{v}})'$, and $F_{\alpha; m, n-m}$ is the 100α upper percentage point of the $F_{m, n-m}$ distribution.

This is akin to the fact that a $100(1-\alpha)\%$ confidence region for $\mu = E(\underline{v})$, $\underline{v} \sim N_m(\mu, \Sigma)$ is,

$$n(\bar{\underline{v}} - \mu) S_{\underline{v}}^{-1} (\bar{\underline{v}} - \mu) \leq \frac{(n-1)m}{n-m} F_{\alpha; m, n-m} \quad (5.4.2)$$

as in Morrison(1976).

In section 2.5 we discussed forecasting a vector time series. For a time series \underline{Z}_t we may estimate the l -step ahead forecast of \underline{Z}_{t+l} by $\underline{Z}_t(l)$ and its variance by $\text{Var}[\underline{e}_t(l)]$, resulting in a confidence region for \underline{Z}_{t+l} given by (2.5.6). Using the same notation as in (2.5.6), but instead, considering the series \underline{v}_t , a $100(1-\alpha)\%$ confidence region for $\underline{v}_{t+l} \sim \text{ARMA}_m(p,q)$ is:-

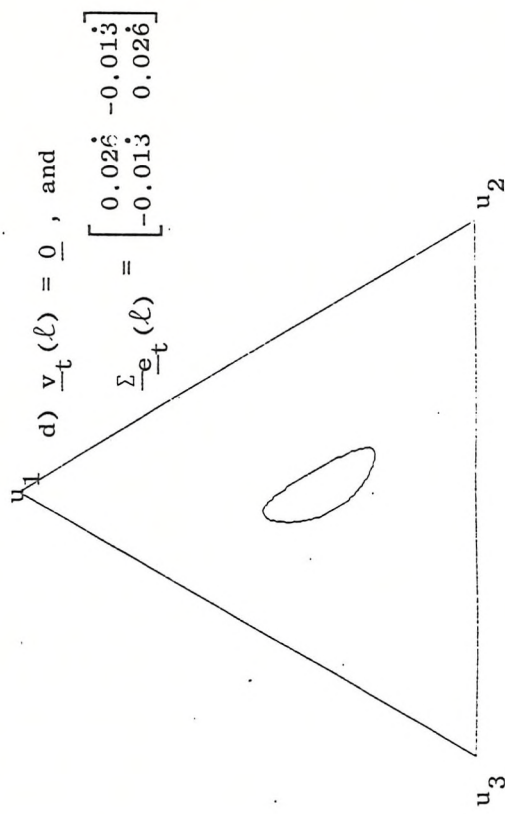
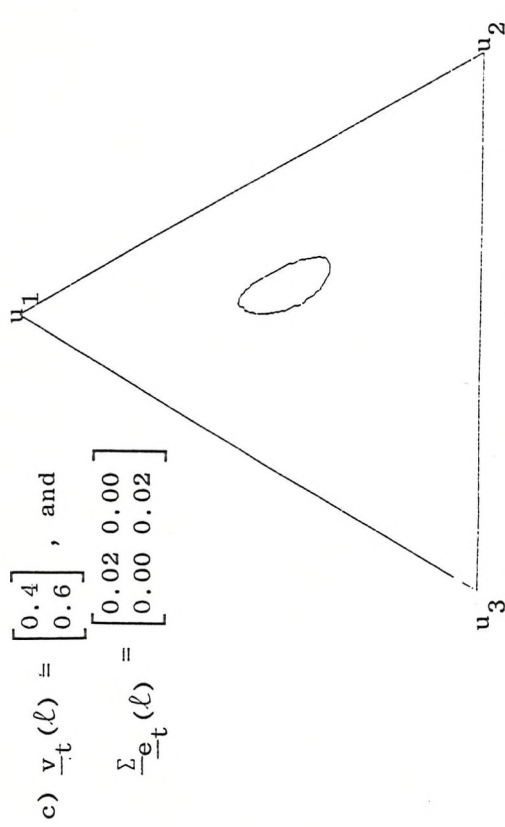
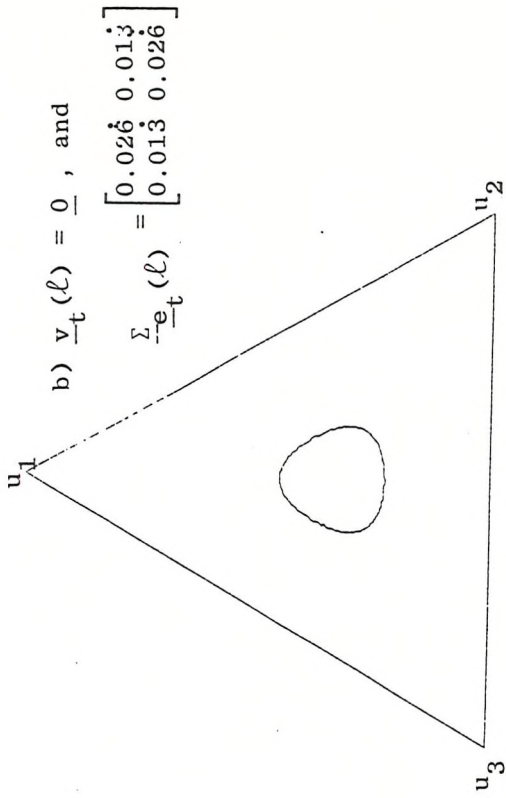
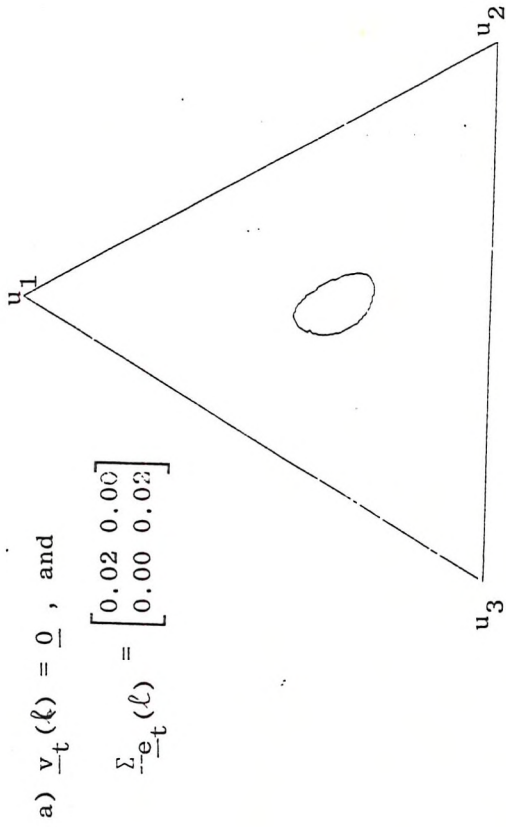


Figure 5.21 : Four examples of a 95% Confidence Region for $\underline{u}_{t+\ell}$.

$$(\underline{v}_t(l) - \underline{v}_{t+l})' \underline{\Sigma}_{\underline{e}_t}^{-1}(l) (\underline{v}_t(l) - \underline{v}_{t+l}) \leq \chi_{\alpha; m}^2, \quad (5.4.3)$$

Comparing (5.4.3) with (5.4.2), and in a similar manner to (5.4.1) we thus obtain a $100(1-\alpha)\%$ confidence region for $\underline{u}_t \sim \ln_+ \text{ARMA}_m(p, q)$ as :-

$$\left[\underline{v}_t(l) - \ln \frac{\underline{u}_{t+l}}{\underline{u}_{t+l, m+1}} \right]' \underline{\Sigma}_{\underline{e}_t}^{-1}(l) \left[\underline{v}_t(l) - \ln \frac{\underline{u}_{t+l}}{\underline{u}_{t+l, m+1}} \right] \leq \chi_{\alpha; m}^2. \quad (5.4.4)$$

Four examples of a 95% confidence region for \underline{u}_{t+l} are given in figure 5.21.

5.5 A Simulated Example

In section 5.2 various location parameters of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution function were considered. All of these could be used as point predictors for forecasting time series $\underline{u}_t \in \mathbb{S}^m$. When applied to our GALLUP poll data these various predictors were almost indistinguishable (see table 5.16f), since all of the data points for the series were well grouped away from the extremities possible. (Figure 5.15). In this instance we would have little concern about which point predictor to use. Also the resulting distribution is uni-modal. Consequently series similar in nature will also be easily handled. But what of other series? In particular compositional time series that originate from bi- or tri- modal white-noise series. What do such series look like? How do their forecasts perform? It is these questions that this example is designed to address by producing a simulated example with the properties required.

The first step was to simulate a white-noise series from a $N_m(\underline{\mu}, \underline{\Sigma})$ distribution, where $\underline{\mu}$ and $\underline{\Sigma}$ are chosen so as to make the daughter $L_m(\underline{\mu}, \underline{\Sigma})$ distribution tri- or bi- modal. This may be readily done using the NAG-library.

500 points from a $N_2 \left[\underline{0}, \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \right]$ were generated. If these points are \underline{v}_t^e , where \underline{v}_t^e is white-noise then the resulting \underline{u}_t^e series is formed by taking the inverse of the logistic transformation. Thus $\underline{u}_t^e = a_m^{-1}(\underline{v}_t^e)$. For the series of 500 points described above the resulting white-noise $\underline{u}_t^e \in \mathbb{S}^2$ series was computed and a plot of this is given in figure 5.22.

As expected the points are well scattered over all the range of possible values (i.e. over all the triangle), but are more dense towards the extremities. That is, nearer the coordinates (1,0,0), (0,1,0) and (0,0,1) which correspond to apexes of the triangular axes. This corresponds to the underlying tri-modal distribution. (c.f. section 5.2).

Using the \underline{v}_t^e series plus a further 100 points, 500 points from an $AR_2(1)$ process were generated by:-

$$\underline{v}_t = \begin{bmatrix} 0.8 & 0.3 \\ -0.4 & -0.5 \end{bmatrix} \underline{v}_{t-1} + \underline{v}_t^e$$

The first 100 points of the now augmented \underline{v}_t^e series were used to "start up" the $AR_2(1)$ process. Finally the resulting \underline{u}_t series was formed as $\underline{u}_t = a_m^{-1}(\underline{v}_t)$. Thus \underline{u}_t is an $ln_+AR_2(1)$ model with parameters,

$$\underline{\phi} = \begin{bmatrix} 0.8 & 0.3 \\ -0.4 & -0.5 \end{bmatrix}, \text{ and } \underline{\Sigma} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}.$$

This series is plotted in figure 5.23. What is apparent from figure 5.23 is that many values occur at or near (1,0,0) and (0,1,0) and very few centrally. Thus the autoregressive components have exaggerated the tendency for the series to take values close to the extremities.

In order to see what sort of forecasts would be produced, we analyzed this data using WMTS-1. The pattern of the cross-correlation matrices for $\{\underline{v}_t\}$ were:-

	+	-	+	-	+	-	+	.	+	.	+	.	+
	+	-
lag 1			2		3		4		5		6		7		8		9		10	
	+	.	+	-	.	-

	11		12		13		14		15		16		17		18		19		20	

A graph of the autocorrelation function for v_{1t} and v_{2t} , and the cross-correlation function are given in figure 5.24.

A schematic representation of the PACF is,

	+	+	.	.	.	-
	-	-
lag 1			2		3		4		5		6		7		8		9		10	

These both suggest an $AR_2(1)$ model, as we would expect

Finally the parameters of an $AR_2(1)$ model were estimated. The

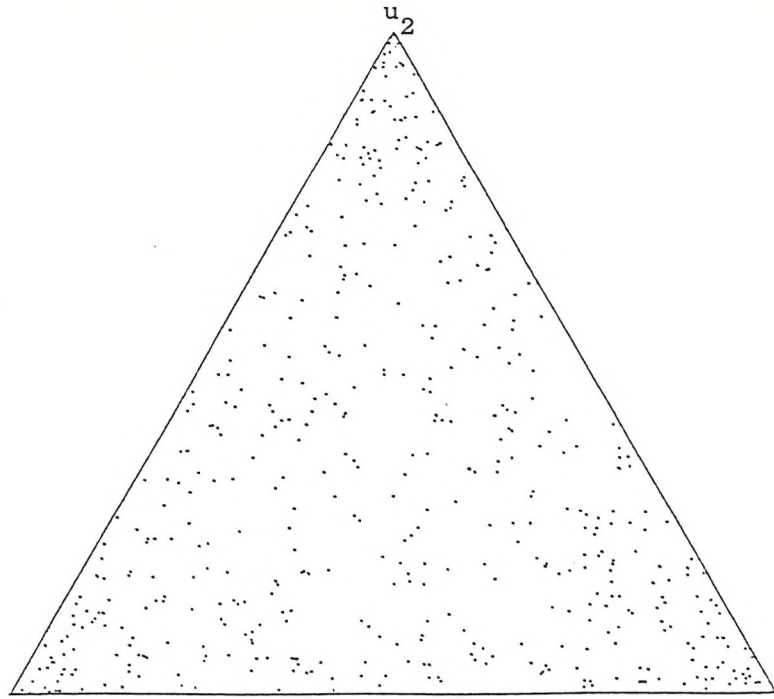


Figure 5.22 Simulated White Noise Series, based on $L_2\left[0, \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}\right]$.

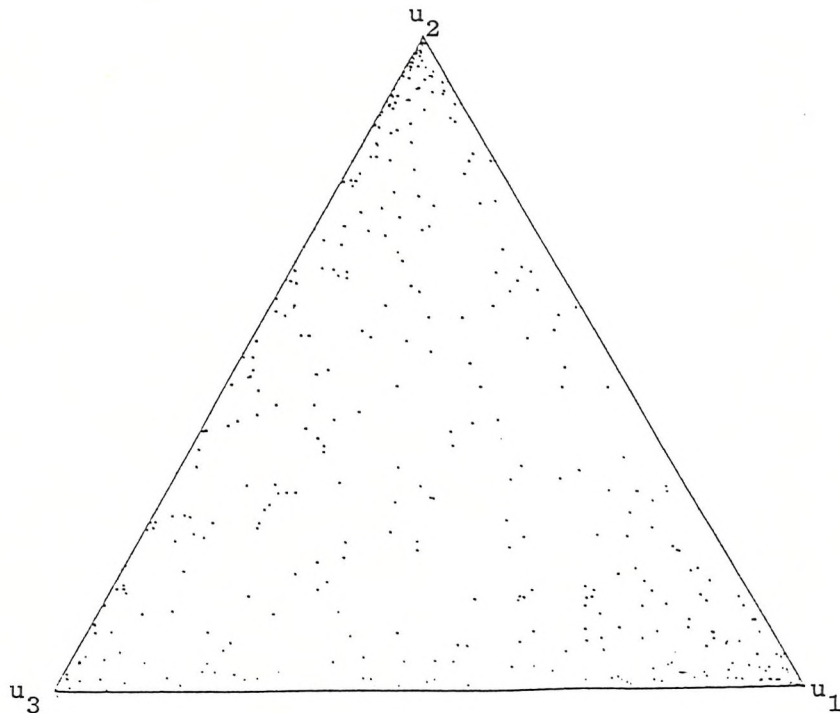
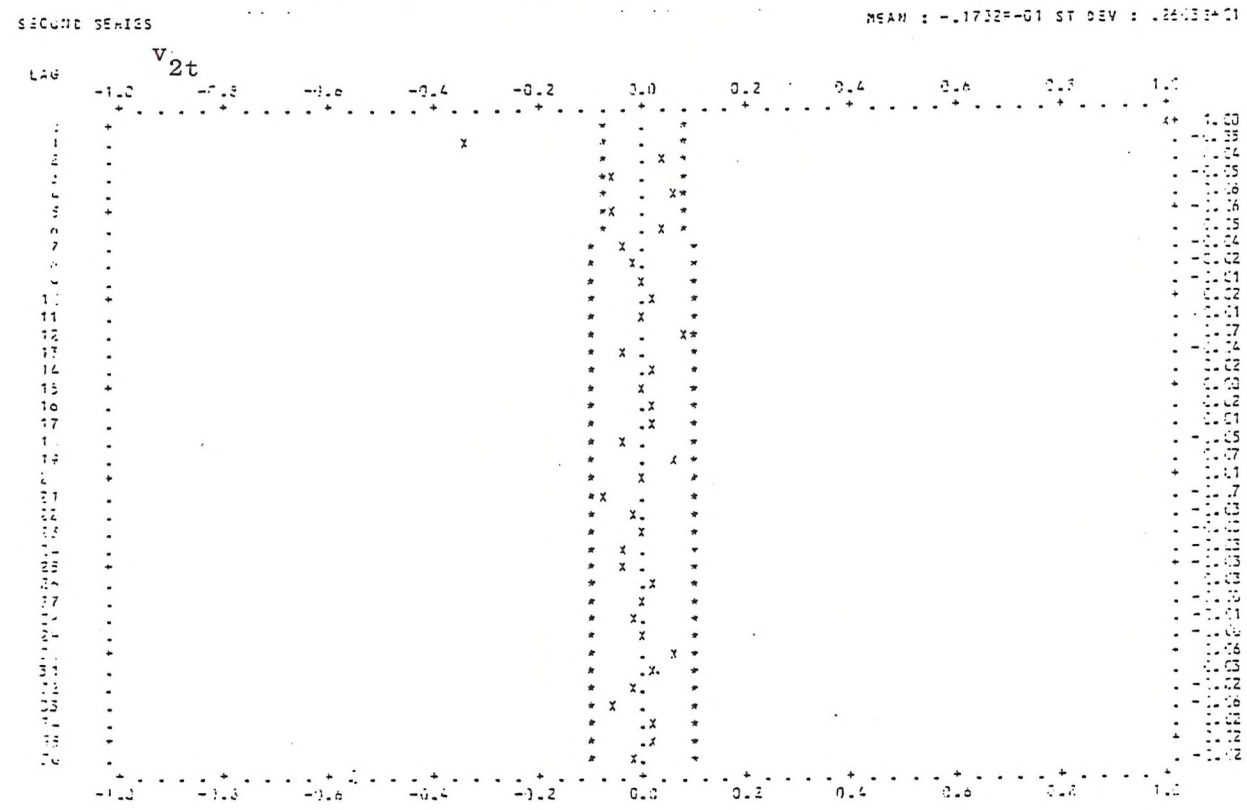
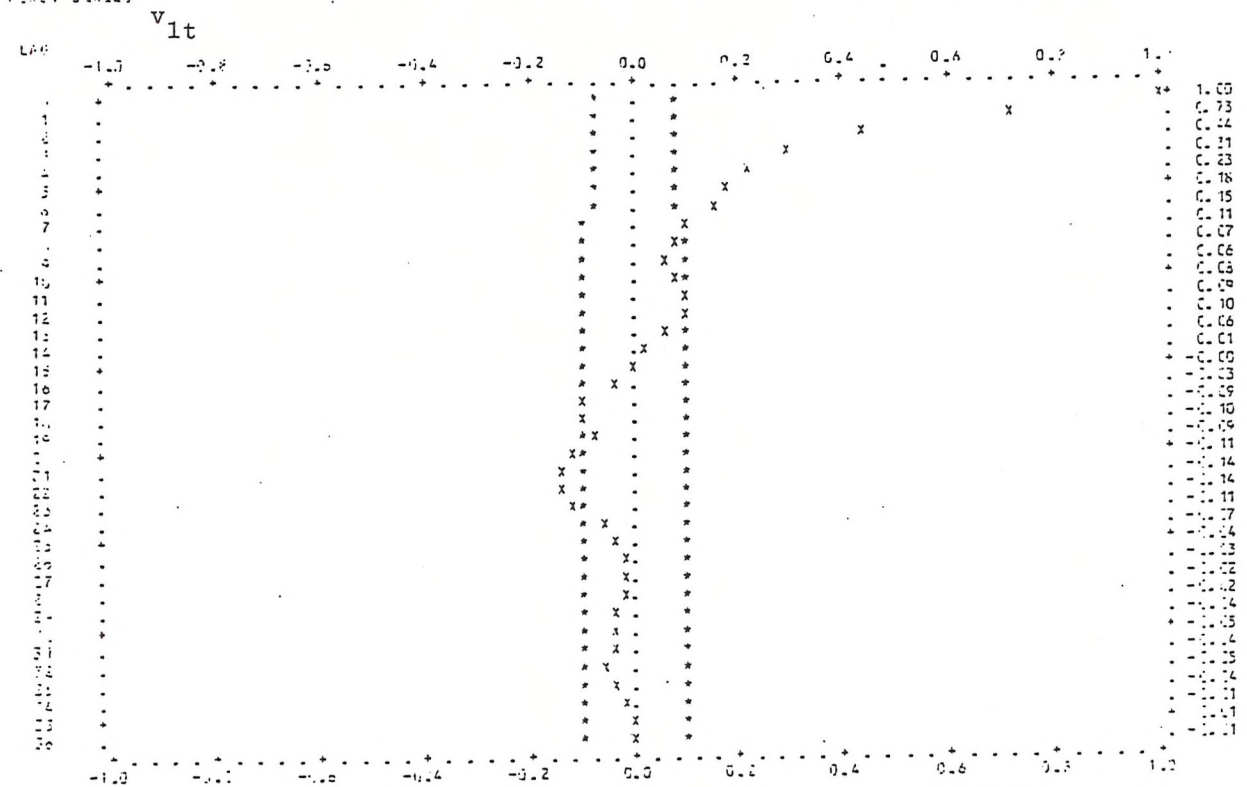


Figure 5.23 Simulated $\ln_+AR_2(1)$ series using white-noise series in figure 5.22 above; where $\underline{\phi} = \begin{bmatrix} 0.8 & 0.3 \\ -0.4 & -0.5 \end{bmatrix}$.

Figure 5.24a) ACFs of v_t Series.

ESTIMATED SAMPLE AUTO CORRELATIONS BASED ON 500 OBSERVATIONS OF THE SERIES
 MEAN : $-1.4027E-01$ ST DEV : $.21773E+01$



constant term was found to be nearly zero (as in fact it is), and the model was re-estimated assuming no constant term. The estimates produced were,

$$\hat{\underline{\Phi}} = \begin{bmatrix} 0.787 & 0.361 \\ -0.416 & -0.448 \end{bmatrix}, \text{ and } \underline{\Sigma} = \begin{bmatrix} 3.896 & 1.917 \\ 1.917 & 4.215 \end{bmatrix},$$

these correspond reasonably well to the true model.

Using this model, starting from the 500th observation, various forecasts were produced, together with their corresponding variances:-

$$\underline{v}_t(1) = \begin{bmatrix} 0.69 \\ 0.15 \end{bmatrix}, \quad \underline{\Sigma}_{\underline{e}_t}(1) = \begin{bmatrix} 3.90 & 1.92 \\ 1.92 & 4.21 \end{bmatrix}$$

$$\underline{v}_t(2) = \begin{bmatrix} 0.60 \\ -0.36 \end{bmatrix}, \quad \underline{\Sigma}_{\underline{e}_t}(2) = \begin{bmatrix} 7.97 & -1.03 \\ -1.03 & 6.48 \end{bmatrix}$$

From the forecasts $\underline{v}_t(l)$ we may then compute various forecasts $\underline{u}_t(l)$. Figure 5.25 shows a plot of the $L_2(\underline{v}_t(1), \underline{\Sigma}_{\underline{e}_t}(1))$ distribution. Most of the mass is around (1,0,0) with a ridge along $u_3=0$ and $u_2=0$. There are three modal values near (1,0,0), (0,1,0) and (0,0,1). The largest is clearly the one near (1,0,0). Evaluating these modes numerically gives,

$$\begin{aligned} \text{Max} &= 66.5909 \text{ at } (0.9974, 0.0011, 0.0140), \\ \text{Max} &= 7.7140 \text{ at } (0.0065, 0.0027, 0.9908), \text{ and} \\ \text{Max} &= 20.9875 \text{ at } (0.0025, 0.9961, 0.0013). \end{aligned}$$

Thus a sensible choice of predictor may be to say $\underline{u}_t(1)$ will be close to (1,0,0) with a high probability, close to (0,1,0) with less probability, and close to (0,0,1) with a small probability. If one were to use the mean then numerically this is at,

$$\text{Mean} = (0.4261, 0.3058, 0.2681).$$

whereas the inverse is at,

$$a_m^{-1}(\underline{v}_t(1)) = (0.4805, 0.2797, 0.2398).$$

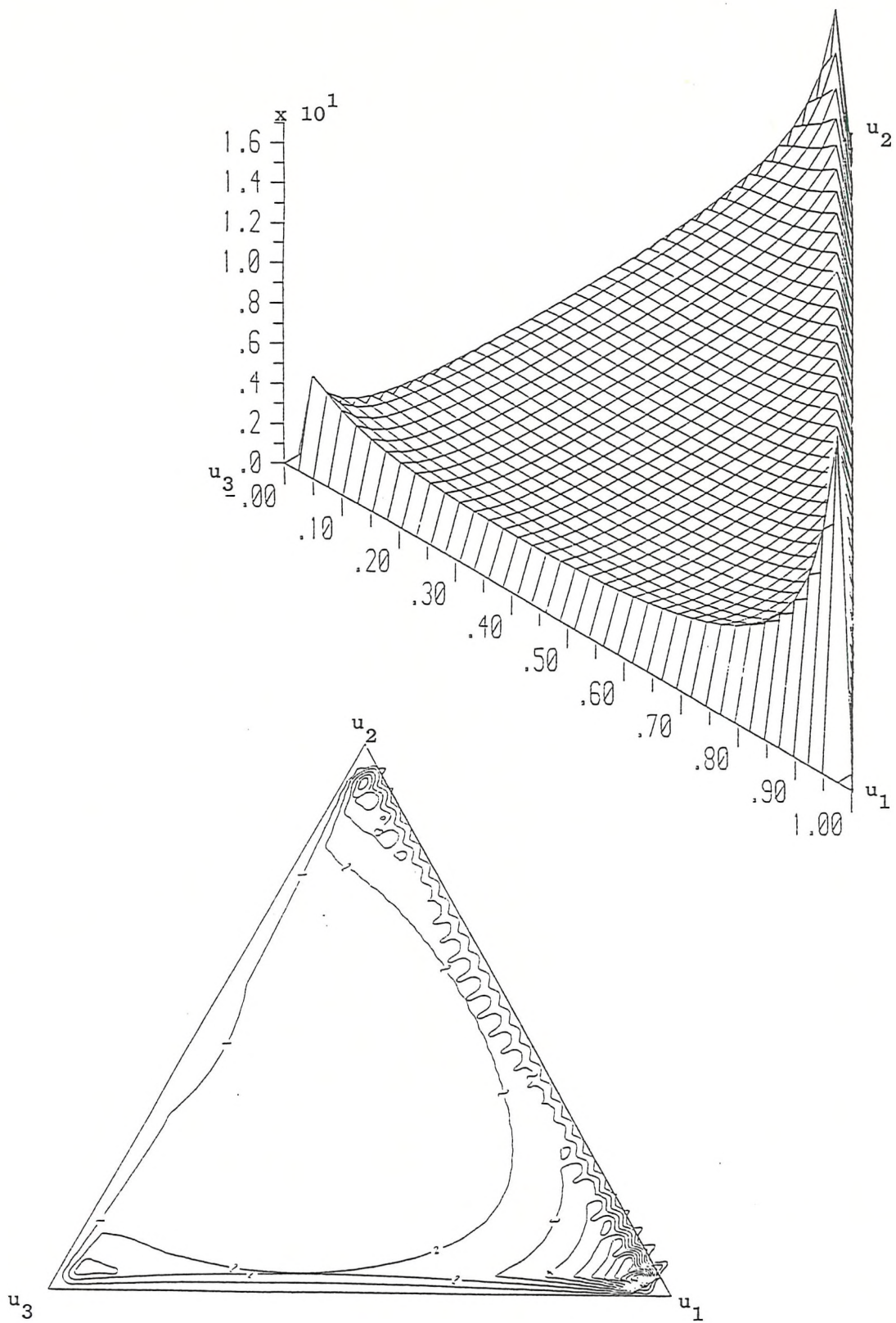


Figure 5.25 Plot of $L_2(v_{-t}(1), \Sigma_{-t}(1))$ Distribution for simulated series.

Both these latter predictors of \underline{u}_{t+1} are fairly close to the central point $(1/3, 1/3, 1/3)$. Also they show a higher value for u_1 and the lowest for u_3 , reflecting the heights of the modes. What is clear, is that the use of the mean or inverse gives a very different picture of the value of $\underline{u}_t(1)$, compared to that given by the modal values. It seems that these two alternatives are equal and opposite. The best choice of predictor will be that which meets the need of the analyst. For example, if u_1, u_2, u_3 are the proportion of sales made by a car manufacturer, and it is required to know how many of each model to produce for next month, the mean should minimize costs for any mistakes in the forecast and the actual values (all other things being equal). If instead the manufacturer wanted to know which car would be the most popular, the modes may be used since it demonstrates that this is u_1 more clearly than does the mean. Or in a car show room, where only one of the three models may be selected for display, the modes might again prove useful.

A further forecast is illustrated in figure 5.26 which is the 95% confidence interval for \underline{u}_{t+1} . The confidence region is extremely large and virtually covers the whole domain of the \mathbb{S}^2 space. Its large size reflects the large value of $\underline{\Sigma}$ and the tri-modal nature of the underlying L_m distribution featured in figure 5.25. It is slightly shifted towards the $(1, 0, 0)$ coordinate, as one might expect. Because of the high mass along the $u_3=0$ and $u_2=0$ axis, a small region around these axes will account for the 5% of the density outside the 95% confidence band. The centre of the region is at $a_m^{-1}(\underline{v}_t(1))$, and consequently these high mass ridges are partially excluded from the confidence region. An alternative interval predictor which would incorporate the high density areas can be obtained by taking the contour of the L_m distribution, inside which the mass is representative of the proportion of the density required to obtain a given size confidence region. The exact nature of this would require further study, but might result in something similar to figure 5.27. Figure 5.27 is the contour lying approximately between those labelled "1" and "2" in figure 5.25. It should be noted that this contour should be regarded as being closer to the $u_3=0$ and $u_2=0$ axis and to the apexes than the contour in figure 5.26. Once again it covers a large region as one would expect in view of the density of the L_m distribution (c.f. figure 5.15).

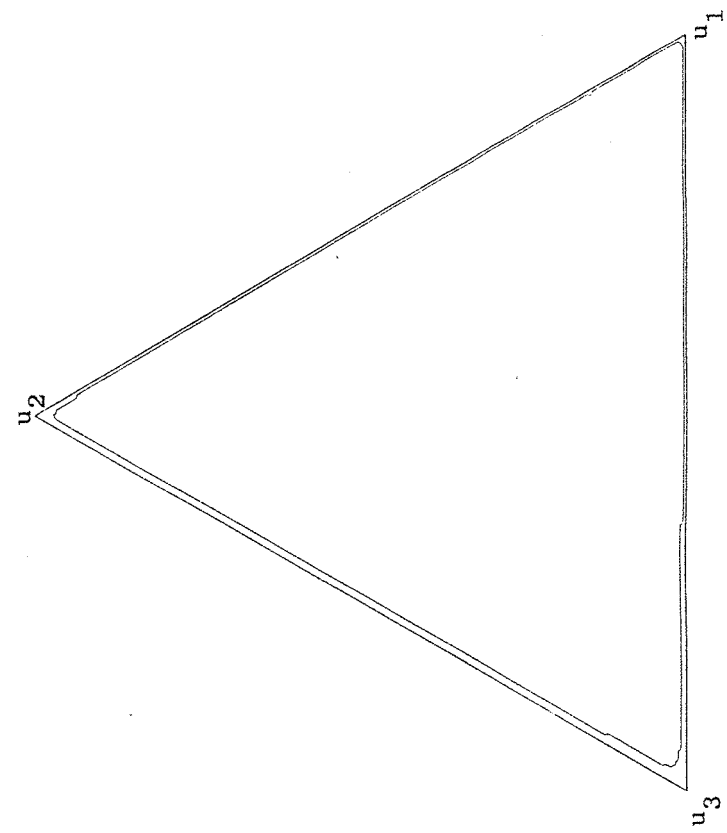


Figure 5.26 95% Confidence Region for \underline{u}_{t+1} .

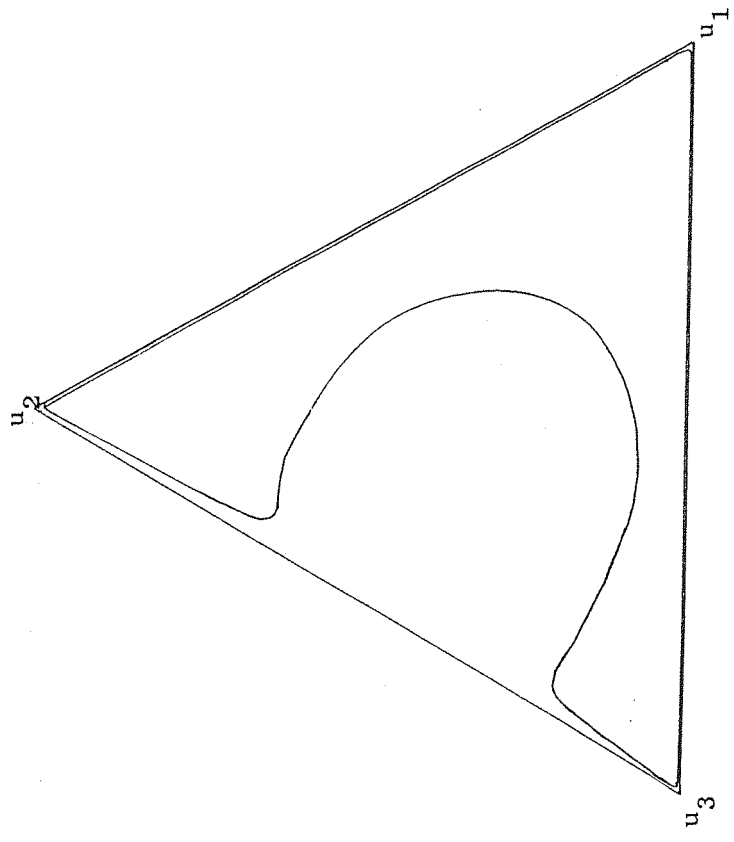


Figure 5.27 Speculative Confidence Region for \underline{u}_{t+1} .

In summary this example sheds light on the difficulty of forecasting compositional data sets, when the underlying distribution is multi-modal. It will depend on the context as to which forecast the analyst should give. Further this example illustrates the fact that if a distribution is multi-modal, then its modes are all close to the extremities. Two points may be drawn from this. Firstly for such data an alternative model may be formulated and prove preferable, if as it seems, the data may be approximated by a discrete state series. (e.g. transition probabilities could then be computed). Secondly it is hard to envisage such difficult data occurring in practice. We would not really expect the pattern of car sales in our example above to look anything like the simulated series. The opinion poll data certainly did not resemble such a pattern. It is hard to imagine that the preference for the Conservative party (say) was 99%, but that in the next month it was now the Labour party that had a preference of 99%. Similarly in the geological context, soil structure will change slowly along a spatial direction, or over time. Thus the need to model compositional time series with the properties described above seems to be rare almost non-existent.

5.6 Summary

In this chapter we have investigated the properties of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution and to a lesser extent the $M_m(\underline{\mu}, \underline{\Sigma})$ distribution. Various examples of these distributions have been examined. It has been shown that the mean and the mode must be estimated numerically, although an approximation to the mean is easily produced. The distributions are not uni-modal for some values of $\underline{\mu}$ and $\underline{\Sigma}$.

The results have been related to the $\ln_+ARMA_m(p,q)$ and $\ln_xARMA_m(p,q)$ models, especially in the context of forecasting. For the opinion poll data it was shown that the various point predictors were virtually identical. It seems that for the majority of data sets this will be the case. However, should this not be the case, a simulated example illustrating the more obscure types of data and the possible pit falls, has been examined.

It seems that forecasting in this context must be done with great care, and the choice of which predictor to use will thus depend on the nature of the data in question, and the reason for forecasting.

"Do you know the laws of the heavens?
Can you set up their dominion over the earth?"

Job 38:33

CHAPTER 6

Dependence and Independence in Compositional Time Series

6.0 Introduction

In section 2.6 we discussed various concepts of causality, feedback etc. between multivariate time series. As discussed in chapter 3 the sum-constraint on compositional data induces an automatic dependence. If we wish to understand the interrelationships between compositional time series it is therefore necessary to develop new forms of dependence. This is the aim of this chapter. The resulting types of dependence and independence have various applications which will be discussed as each new concept is introduced and developed. All of these new concepts arise by integrating the results of section 3.5 with those of 2.6. In section 3.5 we discussed two types of compositional dependence:- extrinsic and intrinsic. These two types are developed below in section 6.1 and 6.2 respectively.

6.1 Extrinsic Analysis

We recall from section 3.5 that extrinsic analysis of compositional data is concerned with investigating the relationship of a composition to the basis from which the composition is conceived to have originated. Similar questions to those of section 3.5 arise in this time series context. For example, consider the household expenditure survey; the relationship between income and the proportional breakdown of expenditure on various commodities may be

examined via the concept of compositional invariance (definition 3.23). If the data consist of a time series, such as would occur if we were to examine a repeated survey on household expenditure we may examine more detailed questions. Here income is τ_t for a given month (say), and assuming all income is spent on n commodities with respective expenditures W_{1t}, \dots, W_{nt} in month t we have that $\tau_t = T(\underline{W}_t)$, and that the proportional breakdown of expenditure is $C(\underline{W}_t) = \underline{U}_t$. We may then investigate the relationship between τ_t and \underline{U}_t . For example how does a pay rise alter spending? Is there a time delay before spending habits are altered to keep in line with others in the same income bracket? In other words does past income 'cause' spending habits. Using the definitions of Wiener-Granger causality (definition 2.17) we may examine if $\tau_t \rightarrow \underline{U}_t$. If one or more of the categories on spending consists of investment in a monetary scheme that will later produce a change in income then clearly the past pattern of expenditure will cause income i.e. $\underline{U}_t \rightarrow \tau_t$ (definition 2.17). Finally the relationship between expenditure and income may be instantaneous which would occur if $\tau_t \cdot \underline{U}_t$. Hence using definition 2.17 we now consider a time series counterpart to compositional invariance (definition 3.23).

Definition 6.1 : (Wiener-Granger) Dependence of a basis.

Let $\underline{U}_t \in \mathcal{S}^m$ be a compositional time series such that $\underline{W}_t \in \mathcal{P}^{m+1}$ is the basis of \underline{U}_t ; i.e. $\underline{U}_t = C(\underline{W}_t)$ and also let

$$\tau_t = T(\underline{W}_t)$$

Then using the notation of definition 2.17 (Wiener-Granger causality) if:-

- i) $\underline{U}_t \parallel \tau_t$ then \underline{W}_t is compositionally invariant ($C \parallel B$)
- ii) $\underline{U}_t \rightarrow \tau_t$ the basis has compositional dependence ($C \rightarrow B$)
- iii) $\tau_t \rightarrow \underline{U}_t$ the basis has compositional causality ($B \rightarrow C$)

- iv) $\underline{U}_t \leftrightarrow \tau_t$ the basis has compositional feedback ($B \leftrightarrow C$)
- v) $\underline{U}_t \cdot \tau_t$ the basis has instantaneous compositional dependence (C.B)
- vi) $\underline{U}_t \Leftrightarrow \tau_t$ the basis has complete compositional dependence ($C \Leftrightarrow B$)

The above definitions have all been included for completeness, and some may prove to have more obvious applications. These definitions could be more vigorously named along the lines of section 2.6 i.e. dependence, causality, feedback are linear in mean with respect to the basis/composition. To include this would produce rather lengthy names. A further reason is that if we were to develop the concepts of causality in for example a non-linear context, then the analogous derived time series definitions may easily be imported into definition 6.1. More importantly, the linear relationship will be via $\underline{v}_t = a_m(\underline{U}_t)$ and not \underline{U}_t as will be seen below. Thus, definition 6.1 should perhaps be in terms of \underline{v}_t 's relationship to τ_t . However since there is a one to one correspondence between \underline{v}_t and \underline{U}_t this amounts to roughly the same thing; and in keeping to the form above we remain more in line with the concept we are trying to achieve. In other words, we want to investigate the relationship between the composition (i.e. \underline{U}_t) and τ_t , \underline{v}_t being a useful intermediate aid to us.

Of the definitions it can be seen that $C \perp\!\!\!\perp B$ and $C \Leftrightarrow B$ are at opposite ends of the scale; the first implies that we lose no information about the compositional time series in ignoring the basis from which it came, unless we have an additional interest in τ_t . The second implies just the opposite, and depending on the particular interpretation required means we may do better to analyse \underline{W}_t instead of \underline{U}_t and τ_t .

The next step is to devise a means of testing the various types of compositional dependence. The \underline{U}_t 's may

be readily modelled by a ℓ_{n+} ARMA_m (p,q) process, and inference can then be made using the F-measure of section 2.6. The procedure is as follows: first form \underline{v}_t from \underline{U}_t i.e.

$$\underline{v}_t = \mathbf{a}_m (\underline{U}_t) , \text{ and also let}$$

$$\tau_t = T (\underline{W}_t) .$$

Then as in definitions 2.18 through to 2.20 we have:-

Definition 6.2 : measures of time series dependence on a basis.

i) The measure of compositional dependence is

$$F_{\underline{U} \rightarrow \tau}^* = \ln \left(\frac{|\underline{\Sigma}(\tau/\tau_p)|}{|\underline{\Sigma}(\tau/\tau_p, \underline{v}_p)|} \right)$$

ii) The measure of compositional causality is

$$F_{\tau \rightarrow \underline{U}}^* = \ln \left(\frac{|\underline{\Sigma}(\underline{v}/\underline{v}_p)|}{|\underline{\Sigma}(\underline{v}/\underline{v}_p, \tau_p)|} \right)$$

iii) The measure of instantaneous compositional dependence is

$$F_{\underline{U}, \tau}^* = \ln \left(\frac{|\underline{\Sigma}(\underline{v}/\underline{v}_p, \tau_p)| \cdot |\underline{\Sigma}(\tau/\tau_p, \underline{v}_p)|}{|\underline{\Sigma}(\underline{v}, \tau/\underline{v}_p, \tau_p)|} \right)$$

iv) The measure of total linear dependence

$$F_{\underline{U} \leftrightarrow \tau}^* = \ln \left(\frac{|\underline{\Sigma}(\underline{v}/\underline{v}_p)| \cdot |\underline{\Sigma}(\tau/\tau_p)|}{|\underline{\Sigma}(\underline{v}, \tau/\underline{v}_p, \tau_p)|} \right)$$

As before

$$F_{\underline{U} \leftrightarrow \tau}^* = F_{\underline{U} \rightarrow \tau}^* + F_{\tau \rightarrow \underline{U}}^* + F_{\underline{U} \cdot \tau}^* \quad (6.1.1)$$

These definitions are related to the time series measures thus:-

$$\left. \begin{aligned} F_{\underline{U} \leftrightarrow \tau}^* &= F_{\underline{v} \leftrightarrow \tau}^* , \\ F_{\underline{U} \rightarrow \tau}^* &= F_{\underline{v} \rightarrow \tau}^* , \\ F_{\tau \rightarrow \underline{U}}^* &= F_{\tau \rightarrow \underline{v}}^* , \quad \text{and } F_{\underline{U} \cdot \tau}^* = F_{\underline{v} \cdot \tau}^* \end{aligned} \right\} (6.1.2)$$

Clearly if \underline{v}_t is univariate we could also use Haugh's test to investigate compositional invariance. As before under the hypothesis of no dependence.

$nF^* \sim \chi^2(r)$, where r is the number of parameters no longer needed if the hypothesis is true.

When some form of dependence does exist between the basis and composition it may be only through a subset of the composition. In the context of a usual set of multivariate time series we can easily examine the subset independently. However because of the transformation of \underline{U}_t to \underline{v}_t it may prove difficult to disentangle exactly which subset of the \underline{U}_t 's is involved. However many of the partition independence properties will assist us. These were introduced in section 3.5, and are extended to the time series context in the next section. The basic idea is to form two subcompositions, allowing only one to interrelate with the τ_t series whilst these are still themselves modelled jointly. This may be done by invoking conditional F measures (Geweke (1984)). In fact any number of possibilities can result from this approach, but since these are simple extensions we do not develop them here. They do require a knowledge of which subset to scrutinise. However, with a little common sense there are

some cases where we may disentangle \underline{v}_t and its relationship to τ_t .

Consider $\underline{W}_t \in \mathbb{P}^3$, and assume that $B \rightarrow C$ exists but only through the first element. Suppose further that

$$\begin{bmatrix} \log W_1 \\ \log W_2 \\ \log W_3 \end{bmatrix}_t = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \lambda \\ 0 \\ 0 \end{bmatrix} \tau_{t-1} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_t \quad (6.1.3)$$

We have chosen to use $\log \underline{W}$ so as easily to map this onto \underline{v}_t , but we note that $\tau_{t-1} = W_{1,t-1} + W_{2,t-1} + W_{3,t-1}$ so that we do not strictly have a linear model in \underline{W}_t .

If we pre-multiply (6.1.3) by $\begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$

and $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$ we obtain respectively:-

$$\begin{bmatrix} v_{1,1} \\ v_{1,2} \end{bmatrix}_t = \begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \end{bmatrix} + \begin{bmatrix} -\lambda \\ -\lambda \end{bmatrix} \tau_{t-1} + \begin{bmatrix} e_2 - e_1 \\ e_3 - e_1 \end{bmatrix}_t \quad \text{(i)}$$

$$\begin{bmatrix} v_{2,1} \\ v_{2,2} \end{bmatrix}_t = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_3 - \mu_2 \end{bmatrix} + \begin{bmatrix} -\lambda \\ 0 \end{bmatrix} \tau_{t-1} + \begin{bmatrix} e_1 - e_2 \\ e_3 - e_2 \end{bmatrix}_t \quad \text{(ii) (6.1.4)}$$

$$\begin{bmatrix} v_{3,1} \\ v_{3,2} \end{bmatrix}_t = \begin{bmatrix} \mu_1 - \mu_3 \\ \mu_2 - \mu_3 \end{bmatrix} + \begin{bmatrix} -\lambda \\ 0 \end{bmatrix} \tau_{t-1} + \begin{bmatrix} e_1 - e_2 \\ e_3 - e_2 \end{bmatrix}_t \quad \text{(iii)}$$

where $v_{i,t} = a_m(\underline{U}_t)$ with $U_{i,t}$ being the reference variable $i = 1, 2, 3$. The coefficients of τ_{t-1} are indicative of the original model. In (ii) and (iii) $(\lambda, 0)$ suggests the relationship $\tau_{t-1} \rightarrow U_{1,t}$, and $(-\lambda, -\lambda)'$ in (i) implies:-

$\tau_{t-1} \rightarrow W_{1t}$ (since U_1 is the reference variable) or
 $\tau_{t-1} \rightarrow U_{2t}$ and U_{3t} , or both $\tau_{t-1} \rightarrow \underline{W}$. By examining
various permutations of the $\ln_+ARMA_m(p,q)$ model it is
therefore possible to gain some insight into the model.
This may be easily done via the $\underline{Z}(k)$ matrix used in
chapter 3, (3.4.2).

The above example illustrates how inspection of the
parameters may aid us in understanding the nature of the
causal relationships in our data. There remains a cautionary
note, however, which stems from a result that is similar
in nature to that of section (3.5). Recall the structure
of $\underline{\Sigma}_y$, the covariance matrix for y that originated from
 U 's basis, \underline{W} consisting of independent components i.e.
 \underline{W} . The $\underline{\Sigma}_y$ given by (3.5.5) not only could be derived
from $\underline{\Sigma}_W$ being diagonal but from any matrix of the form
(3.5.10). Similarly recall (3.5.9). As in (3.5.9) the
models given by (6.1.4) may also be derived from any model
such as (6.1.3) but where the coefficient of τ_{t-1} is now:-
 $(\lambda + a, a, a)$. Thus although models such as (6.1.4) are
necessary for (6.1.3) they are not sufficient. This is
likely to carry over when examining subcompositions and
their relationship to τ_t and as before we may have to make
do with necessity.

Despite its flaws, compositional dependence of a basis
provides a useful tool for many statisticians. It may also
prove to be useful to economists where the coefficient of
 $\log \tau$ (rather than τ) is related to income elasticity.

Another question an investigator may have about the
composition and its basis concerns the interrelationships
between the individual components $W_{1t}, \dots, W_{m+1,t}$. If
 $W_{1t} \rightarrow W_{2t}$ then what will be the relationship between U_{1t}
and U_{2t} ? If \underline{W}_t is known then it is not necessary to
examine \underline{U}_t , however if \underline{W}_t is not known then what would be
useful would be to be able to make inference about \underline{W}_t
based on \underline{U}_t . The loss of information in having only \underline{U}_t

means that it will not always be possible to understand what went on in the basis. If we fit an $\ln_+ARMA_D(p,q)$ model to \underline{U}_t with various different choices of reference variable, and then inspect the values of the resulting parameter estimates, it may be possible to gain some insight into the relationships within \underline{W}_t . That is, we may apply intuition and common sense in a similar way to (6.1.3) and (6.1.4) above. Some forms of dependence have easier solutions than others. For example if $\underline{W}_t = (\underline{W}'_{1t}, \underline{W}'_{2t})'$ and we suspect that $\underline{W}_{1t} \rightarrow \underline{W}_{2t}$ then we may test our hypothesis using subcompositional dependence etc. as will be discussed in section 6.2, and which is a fairly straight forward procedure. It is not possible to examine all the possible relationships on \underline{W}_t , especially those that are extremely complicated, or give intractable results. So what we have chosen to do is to examine an important special case. In particular we will examine the case where the basis consists of $m+1$ independent series. This gives rise to the time series equivalent of basis independence (definition 3.24). It is virtually identical except that by ' $\perp\!\!\!\perp$ ' we mean that the individual auto-correlated W_{it} series are independent of one another, whereas in definition 3.24 meant independence in the statistical sense of W_{ij} being independent observations (i.e. not auto-correlated observations) of independent random variables.

Definition 6.3

$\underline{U}_t \in \mathcal{S}^m (t = 0, \pm 1, \dots)$ is said to have basis independence if there exists a basis $\underline{W}_t \in \mathcal{P}^m$ for \underline{U}_t such that \underline{W}_{it} ($i=1, \dots, m+1$) are independent. That is

- i) $\perp\!\!\!\perp \underline{W}_t$
- ii) $\underline{U}_t = C(\underline{W}_t)$

We will describe this by $\perp\!\!\!\perp B$.

In section 3.5 it was shown that a composition possessing basis independence led to a particular structure of $\underline{\Sigma}$ the

covariance matrix of $\underline{v}_t = a_m(\underline{U}_t)$. The structure for $\underline{\Sigma}$ given by (3.5.5) may be applied to time series. However, it is now necessary to extend this result to the autocovariance function, since the data are now autocorrelated. As might be expected the pattern represented by (3.5.5) becomes the pattern for each cross-covariance matrix $\underline{\Gamma}(k)$.

Lemma 6.4

If $\underline{U}_t \in \mathcal{S}^m$ $t = 0, \pm k, \dots$ has basis independence the autocovariance function $\underline{\Gamma}(k)$ (definition 2.10) of $\underline{v}_t = a_m(\underline{U}_t)$ has the same pattern as (3.5.5), that is:-

$$\underline{\Gamma}(k) = \text{dg}(\gamma_1^{(k)}, \gamma_2^{(k)}, \dots, \gamma_m^{(k)}) + \gamma_{m+1}^{(k)} \underline{U}_m \quad (6.1.5)$$

$$\gamma_i^{(0)} > 0, \quad i = 1, \dots, m+1$$

Proof

Let:- \underline{W}_t be the basis of \underline{U}_t

$$\underline{x}_t = \log(\underline{W}_t),$$

$$\underline{\Omega}(k) = \text{a.c.f. of } \underline{x}_t \quad k = 0, \pm 1, \dots,$$

$$\underline{Y} = [\underline{I}_m \quad ; \quad -\underline{e}_m],$$

$$\underline{v}_t = a_m(\underline{U}_t),$$

and $\underline{\Gamma}(k) = \text{a.c.f. of } \underline{v}_t$.

Then from the definition $\perp\!\!\!\perp \underline{W}_t$

$$\Leftrightarrow \perp\!\!\!\perp \underline{x}_t$$

$$\Rightarrow \underline{\Omega}(k) \text{ is diagonal } k = 0, \pm 1, \dots$$

Also we have that, $\underline{v}_t = \underline{Y}\underline{x}_t$,

$$\Rightarrow \underline{\Gamma}(k) = \underline{Y}\underline{\Omega}(k)\underline{Y}' ; k = 0, \pm 1, \dots$$

For $\underline{\Gamma}(0)$ elements are $\text{var}(\underline{x}_t)$ so that $\gamma_i^{(0)} > 0, i = 1, \dots, m + 1.$

□

Although (6.1.5) is necessary for $\underline{\Gamma} \parallel B$, it is not sufficient. This is not only because of normality assumptions, but as in section 3.5 any $\underline{\Omega}(k)$ of the form,

$$\underline{\Omega}(k) = \begin{bmatrix} \gamma_1^{(k)} + 2\beta_1^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \beta_1^{(k)} + \beta_2^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \dots & \beta_1^{(k)} + \beta_m^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \beta_1^{(k)} + \alpha \\ \beta_1^{(k)} + \beta_2^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \gamma_2^{(k)} + 2\beta_2^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \dots & \beta_2^{(k)} + \beta_m^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \beta_2^{(k)} + \alpha \\ \vdots & \vdots & & \vdots & \vdots \\ \beta_1^{(k)} + \beta_m^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \beta_2^{(k)} + \beta_m^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \dots & \gamma_m^{(k)} + 2\beta_m^{(k)} + \gamma_{m+1}^{(k)} + \alpha & \beta_m^{(k)} + \alpha \\ \beta_1^{(k)} + \alpha & \beta_2^{(k)} + \alpha & \dots & \beta_m^{(k)} + \alpha & \alpha \end{bmatrix} \quad (6.1.6)$$

will yield $\underline{\Gamma}(k) = \underline{Y} \underline{\Omega}(k) \underline{Y}'$ of the same structure as (6.1.5).

Having seen what the structure of $\underline{\Gamma}(k)$ will be under $\underline{\Gamma} \parallel B$, we need to examine next a means of testing for $\underline{\Gamma} \parallel B$. When Aitchison (1981) introduced basis independence he suggested testing for a covariance matrix of the required structure as discussed in section 3.5 (equation (3.5.17)). However, here we have not just one matrix $\underline{\Sigma}$, but a series of such matrices. This is not the only difficulty. The auto-covariance function is not easy to interpret as may be seen by Haugh (1976), who in searching for a test of independence between ordinary time series suggests the S statistic given by definition 2.23. As we saw there, it was not possible to test for diagonal $\underline{\Gamma}(k)$, (i.e. independence assuming normality) without first pre-whitening each series. To

examine \underline{U}_t in compositional time series we seek an equivalent way to pre-whiten our series. An alternative is to develop Bartlett's (1946) result for the expected value of the sampled autocovariance function. This latter approach has so far been avoided because the expressions lead to intractable algebra. We therefore follow the first approach. For this we seek an ARMA process with ACF given by (6.1.5) and then consider pre-whitening the series using this particular model. We begin by examining what the ARMA process would be for \underline{v}_t if \underline{W}_t are independent series such that each $x_{it} = \log W_{it}$ ($i = 1, \dots, m+1$) follow independent ARMA₁(p,q) processes. For $\underline{U}_t = C(\underline{W}_t)$, $\underline{x}_t = \log \underline{W}_t$ assume

$$x_{it} \sim \text{ARMA}_1(p_i, q_i) \quad i = 1, \dots, m+1$$

i.e.
$$\phi_i(B) x_{it} = \theta_i(B) e_{it} \quad (6.1.7)$$

where the e_{it} are independent white-noise series, with variance σ_i^2 .

Since the e_{it} are independent the x_{it} are necessarily independent.

Consider the series $x_{it} - x_{m+1t} = v_{it}$ $i = 1, \dots, m$, hence $\underline{v}_t = a_m(\underline{U}_t)$, then using the result of e.g. Box and Jenkins (1976)

$$v_{it} \sim \text{ARMA}(P_i, Q_i), \text{ where}$$

$$P_i \leq p_i + p_{m+1} \quad (6.1.8)$$

$$Q_i \leq \max(p_i + q_{m+1}, p_{m+1} + q_i) \quad (6.1.9)$$

The resulting ARMA models take the form:-

$$\begin{aligned} \phi_i(B) \phi_{m+1}(B) v_{it} &= -\phi_i(B) \theta_{m+1}(B) e_{m+1,t} + \phi_{m+1}(B) \theta_i(B) e_{i,t} \\ &= \theta_i^{(m+1)}(B) a_{it} \quad (\text{say}) \quad i=1, \dots, m. \end{aligned} \quad (6.1.10)$$

Equality of (6.1.8) and (6.1.9) holds if there are no identical roots in the AR and MA components of (6.1.10).

We will not cancel out any common factors at this stage, so that we may easily compare the m equations represented by (6.1.10).

If we were to use (6.1.10) to pre-whiten our series we first need to know how the a_{it} series were related to one another, since unlike the $m+1$ e_{it} series they are not independent. From (6.1.10) we have

$$\theta_i^{(m+1)}(B) a_{it} = -\theta_{m+1}(B)\phi_i(B)e_{m+1,t} + \phi_{m+1}(B)\theta_i(B) e_{i,t} \quad (6.1.11)$$

Assuming the L.H.S. is invertible let

$$\alpha_i(B) = \left[\theta_i^{(m+1)}(B) \right]^{-1} \theta_{m+1}(B) \phi_i(B) \quad (6.1.12)$$

$$\beta_i(B) = \left[\theta_i^{(m+1)}(B) \right]^{-1} \phi_{m+1}(B) \theta_i(B) \quad \text{and hence we}$$

may rewrite (6.1.11) as:-

$$a_{i,t} = \alpha_i(B) e_{m+1,t} + \beta_i(B) e_{i,t} \quad (6.1.13)$$

Since the $e_{i,t}$'s are independent we have

$$\text{Cov}(a_{i,t}, a_{j,t}) = \sigma_{m+1}^2 \sum_{k=0}^{\infty} \alpha_{i,k} \alpha_{j,k} \quad \begin{array}{l} i, j = 1, \dots, m \\ i \neq j \end{array} \quad (6.1.14)$$

$$\text{var}(a_{i,t}) = \sigma_{m+1}^2 \sum_{k=0}^{\infty} \alpha_{i,k}^2 + \sigma_i^2 \sum_{k=0}^{\infty} \beta_{i,k}^2 \quad i = 1, \dots, m.$$

Consequently, after pre-whitening each of the v_{it} series we do not have an obvious structure to look for amongst the resulting residuals. In particular the a_{it} 's are no longer mutually independent. However consider the following situation:-

Let $\alpha_i(B) = \alpha_j(B) = \beta_{m+1}(B)$ (say),

$$\text{and } \sigma_i^2 \sum_{k=0}^{\infty} \beta_{i,k}^2 = \lambda_i, \quad i = 1, \dots, m+1 \quad (6.1.15)$$

then (6.1.14) yields:-

$$\text{Cov}(a_{i,t}, a_{j,t}) = \begin{cases} \lambda_{m+1} & i, j = 1, \dots, m \\ & i \neq j \end{cases} \quad (6.1.16)$$

$$\text{var}(a_{i,t}) = \lambda_i + \lambda_{m+1} \quad i = 1, \dots, m$$

Equation (6.1.16) represents the now familiar structure of a covariance matrix for $a_m(\underline{U})$ when \underline{U} processes basis independence. Thus if we were to pre-whiten using identical ARMA processes for the \underline{v}_t we could then examine the cross-variance matrix of the residuals to see if it is of the form given by (6.1.16). We may test this via the likelihood ratio statistic (3.5.17); if (6.1.16) does hold then we would have basis independence. However at first sight there seems no reason to suppose we should have identical ARMA₁(p,q) processes for each of the \underline{v}_{it} series. However, further evidence exists for identical structures. The following results that indicate this are somewhat complex. We develop them in stages. Intermediate results and proofs can be found in Appendix A. The first result describes how the individual elements of the autocovariance function are related across lags for an ARMA_m(1,q), where the autocovariance is that given by (6.1.5). It will be seen that the generation of the individual series, $(\gamma_i^{(r)}, \gamma_i^{(r+1)}, \dots)$ $r \geq q$, are identical for all $i = 1, \dots, m+1$ except for one value of i, k (say). This indicates an identical AR component for the \underline{v}_{it} series, apart from an additional term responsible for the generation of the k series $(\gamma_k^{(r)}, \gamma_k^{(r+1)}, \dots)$.

Lemma 6.5

Consider an ARMA_m(1,q) process with autocovariance function

$$\begin{aligned} \underline{\Gamma}(s) &= \text{dg}\{\gamma_1^{(s)}, \gamma_2^{(s)}, \dots, \gamma_m^{(s)}\} + \gamma_{m+1}^{(s)} \underline{U}_m \\ &= \underline{\Lambda}(s) + \gamma_{m+1}^{(s)} \underline{U}_m \quad (\text{say}) \end{aligned} \quad (6.1.17)$$

Then for any $r > q$ and $s > 0$ the ratio

$$\lambda_i^{(s,r)} = \frac{\gamma_i^{(s+r)}}{\gamma_i^{(r)}} \text{ is constant with respect to } i; \quad (6.1.18)$$

$i = 1, \dots, m+1$ except for one value of i , k (say) i.e.

$$\begin{aligned} \lambda_i^{(s,r)} &= \lambda^{(s,r)} \text{ for } i = 1, \dots, m+1 \\ &\quad i \neq k, \\ &\quad \text{where } k \in \{1, 2, \dots, m+1\} \end{aligned}$$

Proof

See Appendix A.

Continuing on from this lemma we now find the exact solution for the AR parameter ϕ . It will be shown to be the sum of two matrices. One represents the identical relation for the m terms that are equivalently generated, whilst the other deals with the k^{th} term, which we will refer to as the rogue parameter. The first matrix is the identity matrix multiplied by a scalar. The second is either a zero matrix except for row k , $k \in (1, \dots, m)$ which apart from a constant $(1/v^{(r)}) \gamma_k^{(r)}$ in notation below) is the k^{th} row of $\underline{\Gamma}(r)^{-1}$. For $k = m+1$ it is the product of the inverse of the diagonal component of $\underline{\Gamma}(r)$ (i.e. $\underline{\Lambda}(r)$) multiplied by \underline{U}_m . The effect of these choices

for the second matrix is to "knock-out" the product of it with $\Gamma(r)$'s $\gamma_i^{(r)}$ terms except for $i = k$, leaving only $\gamma_k^{(r)}$ affected. The result is given in Theorem 6.6.

Theorem 6.6

For an ARMA_m(1,q) process with autoregressive parameter ϕ and autocovariance function as in (6.1.17), i.e.

$$\underline{\Gamma}(r) = \underline{\Lambda}(r) + \gamma_{m+1}^{(r)} \underline{U}_m$$

the only solution of $\underline{\Phi}$ possible is:-

$$\underline{\Phi}' = \phi \underline{I}_m + \beta \underline{A}(k) \tag{6.1.18}$$

for some constants ϕ and β , where

$$k \in \{1, 2, \dots, m+1\},$$

and

$$\{\underline{A}(k)\}_{ij} = \begin{cases} d_{(k)}^{(q)} & i = j = k \quad i, j = 1, \dots, m \\ -\frac{1}{\gamma_i^{(q)}} & j = k \quad ; \quad k = 1, \dots, m; \\ 0 & \text{otherwise} \end{cases} \tag{6.1.19}$$

$$\underline{A}_{(m+1)} = \underline{\Lambda}(q)^{-1} \underline{U}, \quad k = m + 1.$$

$$d_{(k)}^{(s)} = \sum_{\substack{i=1 \\ i \neq k}}^{m+1} \frac{1}{\gamma_i^{(s)}}, \quad s = 0, 1, \dots, .$$

Also for this solution the $\underline{\Gamma}(r+s)$ matrices may be generated from $\underline{\Gamma}(r)$ thus:-

$$\underline{\Gamma}(r+s) = \phi^s \underline{\Gamma}(r) + \delta_{s,r,k} \underline{B}(k) \quad \begin{matrix} r = q, q + 1, \dots, \\ s = 0, 1, \dots ; \end{matrix}$$

where

$$\delta_{s,r,k} = \phi^{r-s} \sum_{i=1}^s \binom{s}{i} \phi^{s-i} \beta^i d \binom{q}{k}^{i-1} (1+\gamma \binom{r}{k}) d \binom{r}{k} ,$$

and

$$\{\underline{B}(k)\}_{ij} = \begin{cases} 1 & i = j = k , \quad k = 1, \dots, m ; \\ 1 & k = m + 1 ; i, j = 1, \dots, m ; \\ 0 & \text{otherwise} , \quad j = 1, \dots, m. \end{cases} \quad (6.1.20)$$

Proof

See Appendix A.

We now consider how this relates to basis independence.

Corollary 6.7

If $\underline{U}_t \in \mathcal{S}^m$, follows an ARMA_m(1,q) process with autoregressive parameter $\underline{\phi}$, and \underline{U}_t possesses basis independence as given in definition 6.3, then $\underline{\phi}$ is of the form given in theorem 6.6.

Proof

Follows directly from lemma 6.4 and Theorem 6.6.

The coefficient matrix $\underline{\phi}$ generates $\underline{\Gamma}(k)$, $k > q$ from $\underline{\Gamma}(q)$ by multiplication. It multiplies each $\gamma_i^{(q)}$ by the same constant, but over and above this, just one of the γ_i 's is generated separately in a non-uniform way. Consequently apart from the rogue series, each $\underline{v}_t = a_m(\underline{U}_t)$ series that follows this model under basis independence follows the same auto-regressive process. Comparing this with our earlier comments on pre-whitening, we again see some evidence for fitting the same ARMA process. However we still have to consider higher-order ARMA models, and the $\underline{\theta}_1, \dots, \underline{\theta}_q$ parameters of the moving-average component.

The generalization to higher-order processes is not that straight forward. We have a much larger matrix

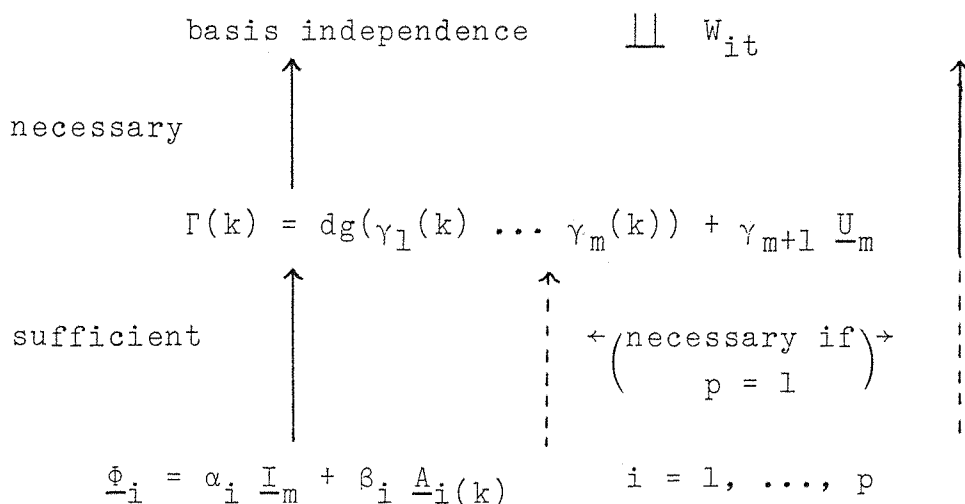
to invert when $p > 1$ since the Yule-Walker equations are now a system of p matrix equations in p unknown matrices. It has not proved possible to derive exact results for higher-order models. However generalizations of theorem 6.6 for the ARMA(2,q) and ARMA(p,q) models are given in Appendix A. (Lemma A.2 - Corollary A.6). These generalizations provide possible solutions for $\underline{\phi}_1, \dots, \underline{\phi}_p$ under basis independence. The solutions have not been shown to be unique. The solutions given are none the less compatible with theorem 6.6. The resulting structure of each of the $\underline{\phi}_i$'s $i = 1, \dots, p$ is as in (6.1.18)

$$\phi_i \underline{I}_m + \beta_i \underline{A}_i \quad i = 1, \dots, p.$$

The second matrix (apart from the constant) is responsible for the generation of the rogue series, and can be seen to have the same pattern as in (6.1.18) except that the elements are now functions of $\gamma_i^{(q)}, \gamma_i^{(q+1)}, \dots, \gamma_i^{(q+p-1)}$ and not just $\gamma_i^{(q)}$. It may be easily shown that \underline{A}_i $i = 1, \dots, p$ all generate the same rogue series. (i.e. \underline{A}_1 cannot generate a different rogue series from \underline{A}_2 etc).

This generalization to higher order models does, however, provide a sufficient structure of the $\underline{\phi}$ parameters for $\underline{\Gamma}(k) = dg(\gamma_1(k), \dots, \gamma_m(k)) + \gamma_{m+1}(k) \underline{U}_m$ ($k=0, \pm 1, \dots$). For the ARMA_m(1,q) process this structure is both necessary and sufficient. Therefore one might use this structure to test for basis independence. It would be necessary to pre-whiten each series using an AR_m(p) process of sufficient order so as to ignore the MA_m(q) component and then examine the resulting residual series. There are several problems with this. Firstly, which series should be the rogue series? Secondly, what will be the structure of the resulting residual series? And finally the structure of the $\underline{\Gamma}(k)$'s is necessary for Basis Independence but by (6.1.6) it is not sufficient. The first problem may be overcome by using an identical AR_m(p)

process, i.e. assuming that there is no rogue series. The second problem would then need to be investigated under the assumption of identical $AR_m(p)$ processes. This would still leave us with the third problem, which implies that the structure for the $\underline{\phi}$'s is neither necessary nor sufficient for basis independence. This can be seen diagrammatically:-



Further algebraic manipulation indicated that the cross-correlation matrix of the residual series may be of the same structure as $\underline{\Gamma}(k)$, and also that the $\underline{\theta}$'s could be of the same form as the $\underline{\phi}$'s. However with no concrete solution to any of these parameters further investigations have been omitted.

The question remains as to why there should be an identical structure to all the v_{ti} series apart from one. One may presume that the identical structure may occur because of the normalizing process i.e. $\underline{U}_t = C(\underline{W}_t) = \underline{W}_t / T(\underline{W}_t)$. However this does not explain the rogue series unless of course it is a mathematical truth which is not a statistical reality. Perhaps the next step might be to carry out an extensive simulation study whereby independent W_{it} series were generated, \underline{U}_t series then formed and an $\ln_+ ARMA_M(p,q)$ model fitted.

We now note one final property of basis independence.

In section 3.5 for $m = 1$ or $m = 2$ the series must necessarily possess basis independence. In time series this is also necessarily true for $m = 1$, but not $m = 2$. Although $\underline{\Gamma}(0)$ must be as in (6.1.5) for $m = 2$, $\underline{\Gamma}(r)$, $r \neq 0$ is not necessarily of the same structure, unless $\underline{\Gamma}(r) = \underline{\Gamma}'(r)$; $r=0, \pm 1, \dots$; i.e. we require that $\underline{\gamma}_{12}(r) = \underline{\gamma}'_{21}(r)$. A similar restriction holds on the parameters. For example,

$$\begin{aligned} \text{If } \underline{\Phi} &= \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} = a\underline{I}_2 + \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \underline{U}_m \quad (\text{from theorem 6.28}) \\ &= \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} + \begin{bmatrix} d_1 & d_1 \\ d_2 & d_2 \end{bmatrix} . \end{aligned}$$

So that

$$\left. \begin{aligned} \phi_{11} &= d_1 + a \\ \phi_{12} &= d_1 \Rightarrow a = \phi_{11} - \phi_{12} \\ \phi_{21} &= d_2 \\ \phi_{22} &= a + d_2 \quad a = \phi_{22} - \phi_{21} \end{aligned} \right\} \Rightarrow \phi_{11} - \phi_{12} = \phi_{22} - \phi_{21} \quad (6.1.21)$$

The equation (6.1.21) represents a linear restriction on the parameters within $\underline{\Phi}$.

The difference between basis independence and time series basis lies in the fact that for the former we require:- (using the same notation as before)

$$\underline{W}_i \perp \underline{W}_j \quad \text{i.e.} \quad \underline{W}_i \perp \underline{W}_j \quad \forall i \neq j = 1, \dots, m+1 \quad (6.1.22)$$

whereas for the latter we require:-

$$\begin{aligned}
\perp\!\!\!\perp W_t \quad \text{i.e.} \quad W_{it} \perp\!\!\!\perp W_{jt} & \quad t = 0, \pm 1, \dots \\
& \quad \forall i \neq j = 1, \dots, m+1 \\
\text{and} \quad W_{it} \perp\!\!\!\perp W_{j,t+k} & \quad k = 0, \pm 1, \dots
\end{aligned}
\tag{6.1.23}$$

Consequently the independence in time series is a lagged independence as well as a contemporaneous one. When a sample of size n (say) of \underline{W} is taken \underline{W}_ℓ ($\ell = 1, \dots, n$) then $W_{i\ell} \perp\!\!\!\perp W_{j,\ell+k}$ for all i, j and k if this holds for $k = 0$ since each of the \underline{W}_ℓ are independent of one another. For a time series, however, \underline{W}_t ($t = 1, \dots, n$) the observations are autocorrelated which leads to the necessity of the further independence condition in (6.1.23).

6.2 Intrinsic Analysis

Intrinsic analysis is concerned with examination of the internal relationships within a compositional data set. Such an analysis is carried out either because no basis exists (conceptually perhaps), or because interest lies within the composition itself. In a public opinion poll of political preference we may be concerned with the relationships between, for example, the two major parties. Little interest as far as the purpose of the survey is concerned lies in the basis. This basis is a function of the sample size obtained by the survey investigator. The way that this achieved sample size varies across time e.g. due to non-response may be of interest to the designer of the survey (e.g. to improve the design) but at the end of the day the results are collected so as to be representative of the total population. Consequently the analysis required is intrinsic, but extrinsic analysis might be used to check the validity of the results (e.g. $\underline{U}_t \perp\!\!\!\perp \tau_t$ is required (definition 6.1)). In other situations intrinsic analysis is the only option, as with the geologist collecting soil or rock samples.

By its nature intrinsic analysis is concerned with the relationships between different elements, or subsets of elements. That is, with non-overlapping subcompositions and amalgamations (definition 3.4 and 3.5). Does, for example, past preference for one of the minor parties influence the 'swing' to one of the major parties. These sorts of questions are examined using the partition independence properties in section 6.2.2. We first, however, consider the intrinsic counter-part to basis independence given by definition 6.3.

6.2.1 Complete Subcompositional Independence

The property that we wish to examine here is whether the components of the composition are all independent of each other, that is no dependence exists between any possible non-overlapping subcompositions. This is the time series version of definition 3.25, the difference being that the independence is required at all lags (c.f. (3.5.18) with (6.2.2) below).

Definition 6.8

A compositional time series $\underline{U}_t \in \mathcal{S}^m$; $t = 0, \pm 1, \dots$ is said to possess complete subcompositional independence

$\coprod \underline{U}_t$ if for every partition of \underline{U}_t :-

$$\underline{U}_t = \{\underline{U}_t^{(1)}, \dots, \underline{U}_t^{(k)}\} \quad k \leq m + 1, \text{ such that}$$

$$\sum_{i=1}^k \underline{U}_t^{(i)} = \underline{U}_t,$$

(6.2.1)

$$\underline{U}_t^{(i)} \cap \underline{U}_t^{(j)} = 0$$

$$i, j = 1, \dots, k$$

$$i \neq j$$

the subcompositional time series are independent i.e.

$$C(\underline{U}_{t_1}^{(1)}) \perp\!\!\!\perp C(\underline{U}_{t_2}^{(2)}) \perp\!\!\!\perp \dots \perp\!\!\!\perp C(\underline{U}_{t_k}^{(k)}) \quad (6.2.2)$$

$$\forall t_1, \dots, t_k, \quad t_i = 0, \pm 1, \dots$$

If $\perp\!\!\!\perp \underline{U}_t$, then $\underline{v}_t = a_m(\underline{U}_t)$ will have a structure to its autocovariance function as can be seen in the following lemma.

Lemma 6.9

If $\underline{U}_t \in \mathcal{S}^m$; $t = 0, \pm 1, \dots, m > 2$, has complete subcompositional independence then the autocovariance function of $\underline{v}_t = a_m(\underline{U}_t)$ is

$$\underline{\Gamma}(s) = \text{diag}(\gamma_1^{(s)}, \dots, \gamma_m^{(s)}) + \gamma_{m+1}^{(s)} \quad (6.2.3)$$

$$s = 0, \pm 1, \dots$$

Proof

If \underline{U}_t has complete subcompositional independence then

$$\log \left\{ \frac{U_{ti}}{U_{tj}} \right\} \perp\!\!\!\perp \log \left\{ \frac{U_{t+s,k}}{U_{t+s,l}} \right\} \quad i, j, k, l \text{ all different}$$

$$s = 0, \pm 1 \dots$$

$$\therefore \text{Cov} \left[\log \left\{ \frac{U_{ti}}{U_{tj}} \right\}, \log \left\{ \frac{U_{t+s,k}}{U_{t+s,l}} \right\} \right] = 0 \quad (6.2.4)$$

Consider first

$$\gamma_{ijk}^{(s)} = \text{Cov} \left[\log \frac{U_{ti}}{U_{tk}}, \log \frac{U_{t+s,i}}{U_{t+s,k}} \right] = \text{Cov} \left[\text{Log} \frac{U_{ti}}{U_{tk}} + \log \frac{U_{t\ell}}{U_{tk}}, \log \frac{U_{t+s,i}}{U_{t+s,k}} \right]$$

$$= \text{Cov} \left[\text{Log} \frac{U_{t\ell}}{U_{tk}}, \log \frac{U_{t+s,i}}{U_{t+s,k}} \right] = \gamma_{\ell jk}^{(s)}$$

for $i \neq j \neq k \neq \ell = 1, \dots, m$ (6.2.5)

Similarly

$$\gamma_{ijk}^{(s)} = \gamma_{ilk}^{(s)} \quad i \neq j \neq k \neq l = 1, \dots, m$$

so that $\gamma_{ijk}^{(s)}$ is independent of i or j so that

$$\text{Cov} \left[\log \left\{ \frac{U_{ti}}{U_{tk}} \right\}, \log \left\{ \frac{U_{t+s,i}}{U_{t+s,k}} \right\} \right] = \gamma_k^{(s)} \quad (6.2.6)$$

Finally consider

$$\begin{aligned} & \text{Cov} \left[\log \left\{ \frac{U_{ti}}{U_{tj}} \right\}, \log \left\{ \frac{U_{t+s,i}}{U_{t+s,j}} \right\} \right] \\ &= \text{Cov} \left[\log \left\{ \frac{U_{ti}}{U_{tk}} \right\} + \log \left\{ \frac{U_{tk}}{U_{tj}} \right\}, \log \left\{ \frac{U_{t+s,i}}{U_{t+s,l}} \right\} + \log \left\{ \frac{U_{t+s,l}}{U_{t+s,j}} \right\} \right] \\ &= \text{Cov} \left[\log \left\{ \frac{U_{tk}}{U_{tj}} \right\}, \log \left\{ \frac{U_{t+s,l}}{U_{t+s,i}} \right\} \right] + \text{Cov} \left[\log \left\{ \frac{U_{tk}}{U_{tj}} \right\}, \log \left\{ \frac{U_{t+s,l}}{U_{t+s,j}} \right\} \right] \end{aligned}$$

since the other terms equal zero
from (6.2.4)

$$= \gamma_i^{(s)} + \gamma_j^{(s)} \quad (6.2.7)$$

The result follows directly from (6.2.6) and (6.2.7) since

$$\left\{ \underline{\Gamma}(s) \right\}_{ij} = \text{Cov} \left[\ln \frac{U_{ti}}{U_{t,m+1}}, \ln \frac{U_{t+s,j}}{U_{t+s,m+1}} \right] = \gamma_{m+1}^{(s)} \left. \begin{array}{l} i \neq j = 1, \dots, m \\ i = j = 1, \dots, m. \end{array} \right\}$$

□

The difference between $\underline{\Gamma}(s)$ in lemma 6.4 and its equivalent form in lemma 6.9 lies firstly in the fact that in the former $\gamma_i^{(2)} > 0$ (6.1.5), whereas no restrictions hold on the $\gamma_i^{(0)}$'s in lemma 6.9 except that they must obviously ensure positive definiteness of $\underline{\Gamma}(0)$ (for which $\gamma_i^{(0)} > 0$ need not necessarily hold (e.g. $\gamma_{m+1}^{(0)} = 0$). In lemma 6.4, the $\gamma_i^{(0)}$ represent variances and so they must be positive values. This is the identical difference between the two structures of Σ in the static formulation of these concepts of independence (section 3.5). The second distinction occurs when $m = 2$. At the end of section 6.1 we saw that basis independence was no longer automatic when considering time series. However as in 3.5 for intrinsic analysis complete subcompositional independence trivially holds in the time series context. This may be seen from the definition since what we require is:-

$$C(U_{ti}, U_{tj}) \perp\!\!\!\perp C(U_{t+s,k})$$

i.e. $\log \left\{ \frac{\gamma_{ti}}{\gamma_{tj}} \right\} \perp\!\!\!\perp 1$ which always holds. (6.2.8)

$$i \neq j \neq k = 1, 2, 3.$$

If we compare this to section 3.5, this trivial result was also apparent because all 2 x 2 covariance matrices can necessarily be expressed in the required form:-

$$\Sigma = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} + \begin{bmatrix} \lambda_3 & \lambda_3 \\ \lambda_3 & \lambda_3 \end{bmatrix}$$

On crossing to time series it is now the ACF that must have this structure, so that under basis independence of

a time series

$$\underline{\Gamma}(s) = \begin{bmatrix} \gamma_1^{(s)} & 0 \\ 0 & \gamma_2^{(s)} \end{bmatrix} + \begin{bmatrix} \gamma_3^{(s)} & \gamma_3^{(s)} \\ \gamma_3^{(s)} & \gamma_3^{(s)} \end{bmatrix} \quad s = \pm 1, \dots \quad (6.2.9)$$

which reduces to $\underline{\Gamma}(s) = \underline{\Gamma}'(s)$.

In the intrinsic case however whilst complete sub-compositional independence holds automatically as described above, the structure of $\underline{\Gamma}(s)$ need not necessarily be as in (6.2.9). That is the form of Σ in the static version is wholly consistent with the definition of complete subcompositional independence at $m = 2$ so that both Σ and the definition lead to the same conclusion. In the time series version this is no longer the case. Part of the problem can be seen by examining the proof of lemma 6.9, (6.2.5) depends on the indices i, j, k and l all being different but that cannot hold when they may each be one of 1, 2 or 3. The fact that the proof relies on $m > 2$ can be seen also by noting (6.2.4); again this requires 4 indices, without which we are reduced to the trivial equation (6.2.8).

Although this is something of an anomaly, we will copy section 3.5 and for compositional time series on \mathcal{S}^2 assume that complete compositional independence holds trivially.

Corollary 6.10

If $\underline{U}_t \in \mathcal{S}^m$ follows a $L_m(\underline{u}, \underline{\Gamma}(0))$ distribution, the $\underline{\Gamma}(s)$ given by (6.2.3) is a sufficient condition for complete subcompositional independence.

Proof

$$\begin{aligned}
 & \text{Consider cov} \left[\log \left\{ \frac{U_{ti}}{U_{tj}} \right\}, \log \left\{ \frac{U_{t+s,k}}{U_{t+s,l}} \right\} \right] \\
 &= \text{cov} \left[\log \left\{ \frac{U_{ti}}{U_{t,m+1}} \right\} - \log \left\{ \frac{U_{tj}}{U_{t,m+1}} \right\}, \log \left\{ \frac{U_{t+s,k}}{U_{t+s,m+1}} \right\} - \log \left\{ \frac{U_{t+s,l}}{U_{t+s,m+1}} \right\} \right] \\
 &= \gamma_{m+1}(s) - \gamma_{m+1}(s) - \gamma_{m+1}(s) + \gamma_{m+1}(s) \\
 &= 0,
 \end{aligned}$$

$$\left[\text{Recall from (6.2.6) } \text{cov} \left[\log \left\{ \frac{U_{ti}}{U_{t,m+1}} \right\}, \log \left\{ \frac{U_{tj}}{U_{t,m+1}} \right\} \right] = \gamma_{m+1}(s) \right. \\
 \left. \qquad \qquad \qquad i \neq j \neq m+1 \right]$$

which under normality

$$\begin{aligned}
 \Rightarrow \log \left\{ \frac{U_{ti}}{U_{tj}} \right\} \perp\!\!\!\perp \log \left\{ \frac{U_{t+s,k}}{U_{t+s,l}} \right\} & \equiv C(U_{ti}, U_{tj}) \perp\!\!\!\perp C(U_{t+s,k}, U_{t+s,l}) \\
 & i \neq j \neq k \neq l = 1, \dots, m+1 \\
 & t, s = 0, \pm 1, \dots
 \end{aligned}$$

\Rightarrow complete subcompositional independence. □

From lemma 6.9, a test for $\perp\!\!\!\perp U_t$ is a test of (6.2.3). Hence we may use the results derived in section 6.1 that are based on a similar premise. In fact the structure for $\Gamma(k)$ (6.2.3) is now both necessary and sufficient. Consequently the structure suggested by theorem 6.6 and lemma A.2 - theorem A.5 for ϕ_i is now sufficient for $i = 1, \dots, p$ and all p . Despite this we are again left with what seems an intractable problem. However, the definition 6.8 suggests a further approach. We require that for each

possible partition the resulting subcompositions $C(\underline{U}_t^{(i)})$ $i = 1, \dots, k$ are independent. Let

$C(\underline{U}_t^{(i)}) \in \mathcal{S}^{m_i}$ $i = 1, \dots, k$, then this is equivalent to

$$\prod a_{m_i} \left(C(\underline{U}_t^{(i)}) \right) \quad i = 1, \dots, k \text{ for all partitions } k.$$

It should be noted that this reduces to

$$\prod a_{m_i} \left(C(\underline{U}_t^{(i)}) \right) \quad i = 1, \dots, k \text{ for all partitions}$$

such that $m_i \in \{0, 1\}$

If we form each of these partitions, complete subcompositional independence may be tested by requiring independence in each partition.

For example consider $\underline{U}_t \in \mathcal{S}^3$ then let

$$V_t^{(ij)} = \log \left\{ \frac{U_{ti}}{U_{tj}} \right\} \quad i \neq j \neq 1, 2, 3, 4$$

then complete subcompositional independence \Leftrightarrow

$$V_t^{(12)} \prod V_{t+s}^{(34)} \cap V_t^{(13)} \prod V_{t+s}^{(24)} \cap V_t^{(14)} \prod V_{t+s}^{(23)}$$

$$t, s = 0, \pm 1, \dots$$

Independence between the resulting series may be easily checked using any method in section 2.6 (e.g. the S test). This may be easily extended to higher dimensions, but we must form $m(m+1)/2$ ($= \sum_{i=1}^m i$) v_t series and make $3 \times \binom{m+1}{4}$ comparisons. This is not only computationally inefficient but also requires that care must be taken to ensure the correct critical region is obtained as is necessary with multiple comparisons.

In the absence of any concrete method this may perhaps be a useful approach.

Despite the difficulties it is perhaps worth noting that this form of independence is probably the least useful of the intrinsic concepts. The partition independence properties have very ready applications and these are discussed below. If the analyst wishes to study only a subset of his series he may use these methods to discover the implications of discarding some series. Complete subcompositions independence implies that he may quite happily study any subset of the data without undue loss. More usually such a subset is pre-chosen on the basis of interest for which partition independence properties will suffice. In this regard complete subcomposition independence may be seen as a generalization of some of the concepts described below. It also offers something in the way of an explanation of the data. For example if it holds for the public opinion poll data it would give us insight into the way public opinions arise. In a poll on voting intentions in a four party system it would imply that a swing to say the first party away from the second was not related to the swing between the third and fourth parties, nor the swing between the first and third to that of the second and fourth etc. It is difficult to envisage to what extent such knowledge of a compositional time series would be useful.

6.2.2 Partition Independence Properties

The concept discussed in the previous section refers to every possible partition of the compositional time series $\underline{U}_t \in \mathcal{S}^m$. Now instead we consider the independence properties of a particular partition or set of partitions. To start consider the partition:-

$$\underline{U}_t^{(C)} = (U_{t1}, \dots, U_{tC}) \in \mathcal{S}^C$$

$$\underline{U}_{t(C)} = (U_{t,C+1}, \dots, U_{t,m+1}) \in \mathcal{S}^{m-C+1} \quad t = 0, \pm 1, \dots$$
(6.2.10)

$$\underline{U}_t^{(m+1)} = \{\underline{U}_t^{(C)}, \underline{U}_{t(C)}\}$$
(6.2.11)

Let the fill-up-value for the series $\underline{U}_t^{(C)}$ be $\tau_t(C)$ where

$$\tau_t(C) = T(\underline{U}_{t(C)})$$
(6.2.12)

and similarly the FUV for $\underline{U}_{t(C)}$ is $\tau_t^{(C)}$

$$\tau_t^{(C)} = T(\underline{U}_t^{(C)}) = 1 - \tau_t(C)$$
(6.2.13)

Whence $(\underline{U}_t^{(C)}, \tau_t(C))$ forms a compositional time series by amalgamating the last $m-C$ components of the \underline{U}_t series. $\tau_t(C)$ may be regarded as the share of \underline{U}_t accounted for by $\underline{U}_{t(C)}$. Similarly $(\tau_t^{(C)}, \underline{U}_{t(C)})$ is a compositional time series. We have a one to one mapping from the original series to these two newly formed ones:-

$$\underline{U}_t \leftrightarrow (\underline{U}_{t(C)}, \tau_t^{(C)}, \underline{U}_t^{(C)}) \leftrightarrow (\underline{U}_{t(C)}, \tau_t^{(C)}); (\underline{U}_t^{(C)}, \tau_t(C))$$
(6.2.14)

Having formed this particular partition, various hypotheses about the independence of its components may be considered. The first of these is time series subcompositional dependence.

Definition 6.11 : Time series subcompositional dependence

For $\underline{U}_t \in \mathcal{S}^m$ and the partition given by (6.2.10) through to (6.2.13) and using the notation of definition 2.17 (Wiener-Granger causality)

- i) If $C(\underline{U}_t^{(C)}) \perp\!\!\!\perp \tau_t^{(C)}$ then $\underline{U}_t^{(C)}$ is subcompositionally invariant. (S $\perp\!\!\!\perp$ T)₁

- ii) If $C(\underline{U}^{(C)}) \rightarrow \tau^{(C)}$; then $\underline{U}_t^{(C)}$ has subcompositional causality. (S \rightarrow T)₁
- iii) If $\tau^{(C)} \rightarrow C(\underline{U}^{(C)})$; $\underline{U}_t^{(C)}$ has subcompositional dependence. (T \rightarrow S)₁
- iv) If $C(\underline{U}^{(C)}) \leftrightarrow \tau^{(C)}$; $\underline{U}_t^{(C)}$ has subcompositional feedback. (S \leftrightarrow T)₁
- v) If $C(\underline{U}^{(C)}) \cdot \tau^{(C)}$; $\underline{U}_t^{(C)}$ has instantaneous subcompositional dependence. (S \cdot T)₁
- vi) If $C(\underline{U}^{(C)}) \Leftrightarrow \tau^{(C)}$; $\underline{U}_t^{(C)}$ has complete subcompositional dependence. (S \Leftrightarrow T)₁

If we compare definition 6.11 with definition 6.1 it is easily seen that subcompositional dependence is the exact intrinsic counterpart of time series dependence of a basis. $\underline{W}_t, \underline{U}_t, \tau_t$ are replaced by $\underline{U}_t^{(C)}, C(\underline{U}_t^{(C)})$, and $T(\tau_t^{(C)})$ i.e. $\underline{U}_t^{(C)}$ is used as a basis, and the term composition is replaced by subcomposition. Consequently much that was said previously is valid here, so that testing and modelling is as before, but interpretation is different.

It is also possible to define a symmetric set of relationships on $\underline{U}_t^{(C)}$, we will denote these by (S $\perp\!\!\!\perp$ T)₂, (S \rightarrow T)₂, ..., (S \Leftrightarrow T)₂, as the distinction between the two will prove useful below; although in fact with reordering they amount to the same thing as (S $\perp\!\!\!\perp$ T)₁, etc.

The interpretation of the above subcompositional concept varies at least notionally from its extrinsic parent. Here we are studying the relationship between a subset of the variables and their "share" of the full composition. As an illustration consider again the example of a household expenditure survey. Suppose we are interested in studying the proportional breakdown spent on individual types of food given the proportion

of income spent on food. Does someone who spends 50% of their income on food buy a different basket full of goods than someone who spends say 10%? In the extrinsic case we were considering the amount spent on food £500 or £100 say, whereas now it is the proportion of income that is under investigation. In this example one would prefer the extrinsic analysis where possible. However if the basis were not known, intrinsic analysis may still give useful results. If we assume that those on a high income spend proportionally less on food although their total expenditure is higher so that they buy more luxury foods (e.g. expensive cuts of meat vs. bread). Then a higher value of U_{tj} for luxury foods will be related to a higher income in the extrinsic case, but a lower proportion in the intrinsic. Thus if data is lost the intrinsic approach may still offer an extrinsic interpretation.

A purely intrinsic example may be found by considering the opinion poll. Does the swing to one of the major parties vary according to how much of the vote the remaining parties take, and to how many people don't vote. If there is no difference then $S \perp\!\!\!\perp T$. If the parties are evenly matched e.g. $U_t(\text{CON}) \approx U_t(\text{LAB})$, voters in the next few elections may be motivated to vote for one of the major parties. This would result in a higher turn out at the polls and a smaller vote for the remaining parties. If instead there is a strong seat the opposite may occur. In such a case we clearly have $S \rightarrow T$. Similarly if in fact a high vote for the major parties favours one party we have $S \leftarrow T$, and so on.

We now consider the next form of partition independence

Definition 6.12 : Time series conditional subcompositional dependence.

For the partition given by (6.2.10) - (6.2.13) and

definition 2.17

- i) If $C(\underline{U}_t^{(C)}) \perp\!\!\!\perp C(\underline{U}_{t(C)})/\tau_t^{(C)}$ then the partition is said to have conditional subcompositional independence. $(S^C \perp\!\!\!\perp S_C/\tau)$
- ii) If $C(\underline{U}_t^{(C)}) \rightarrow C(\underline{U}_{t(C)})/\tau_t^{(C)}$ the subcomposition $C(\underline{U}_t^{(C)})$ is said to conditionally cause $C(\underline{U}_{t(C)})$. $(S^C \rightarrow S_C/\tau)$
- iii) If $C(\underline{U}_t^{(C)}) \cdot C(\underline{U}_{t(C)})/\tau_t^{(C)}$ the partition has conditional instantaneous dependence. $(S^C \cdot S_C/\tau)$
- iv) If $C(\underline{U}_t^{(C)}) \leftrightarrow C(\underline{U}_{t(C)})/\tau_t^{(C)}$ the partition has conditional feedback. $(S^C \leftrightarrow S_C/\tau)$
- v) If $C(\underline{U}_t^{(C)}) \Leftrightarrow C(\underline{U}_{t(C)})/\tau_t^{(C)}$ the partition has complete conditional dependence. $(S^C \Leftrightarrow S_C/\tau)$

Where $A \sim B/\alpha$ implies that in definition 2.17 $\Omega = \{A, B, \alpha\}$ for ' \sim ' $\in \{\perp\!\!\!\perp, \rightarrow, \cdot, \leftrightarrow, \Leftrightarrow\}$.

Definition 6.13 : (left and right) neutrality of time series

For the partition (6.2.10) - (6.2.13) of $\underline{U}_t \in \mathcal{S}^m$ we have

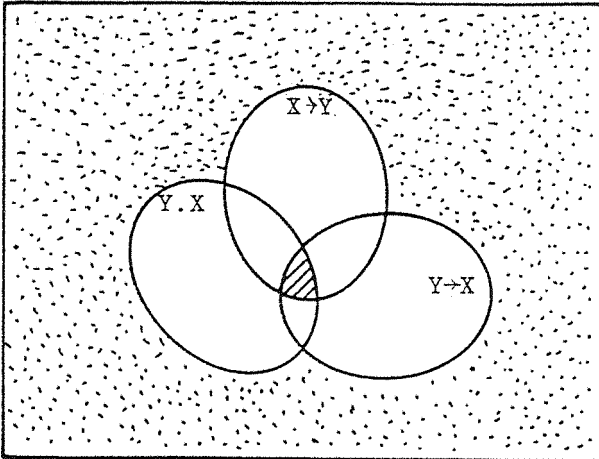
- i) Left neutrality if $C(\underline{U}_t^{(C)}) \perp\!\!\!\perp \underline{U}_{t(C)}$ $(N_1 \perp\!\!\!\perp)$
- ii) If $C(\underline{U}_t^{(C)}) \rightarrow \underline{U}_{t(C)}$ we have left non-neutral causation $(N_1 \rightarrow)$
 but if $C(\underline{U}_t^{(C)}) \neq \underline{U}_{t(C)}$ we similarly have left neutral non-causality. $(N_1 \neq)$
- iii) If $\underline{U}_{t(C)} \rightarrow C(\underline{U}_t^{(C)})$ we say we have left non-neutral dependence $(N_1 \leftarrow)$
 and $\underline{U}_{t(C)} \neq C(\underline{U}_t^{(C)})$ is left neutral non-dependence $(N_1 \neq)$

- iv) If $C(\underline{U}_t^{(C)}) \leftrightarrow \underline{U}_{t(C)}$ the partition is said to have left non-neutral feedback $(N_1 \leftrightarrow)$
- v) If $C(\underline{U}_t^{(C)}) \cdot \underline{U}_{t(C)}$ the partition is said to have left instantaneous non-neutrality $(N_1 \cdot)$
- vi) If $C(\underline{U}_t^{(C)}) \leftrightarrow \underline{U}_{t(C)}$ the partition is said to have complete left non-neutrality $(N_1 \leftrightarrow)$
- vii) Interchanging $\underline{U}_t^{(C)}$ with $\underline{U}_{t(C)}$ in i) - vi) above gives the corresponding definitions of right neutrality $(N_2 \sim)$
- where $\sim \in \{\lll, \rightarrow, \neq, \leftarrow, \neq, \cdot, \cdot\}$.

Definition 6.14 : Partition independent time series

- i) For the above partition (6.2.10)-(6.2.13) the time series $\underline{U}_t \in \mathbb{S}^m$ is said to have partition independence if $\lll (C(\underline{U}_t^{(C)}), C(\underline{U}_{t(C)}), \tau_t(C))$ $(P \lll)$
- ii) Similarly the partition has full feedback if
- $$C(\underline{U}_t^{(C)}) \leftrightarrow C(\underline{U}_{t(C)})$$
- $$C(\underline{U}_t^{(C)}) \leftrightarrow \tau_t$$
- $$C(\underline{U}_{t(C)}) \leftrightarrow \tau_t \quad (P \leftrightarrow)$$
- iii) The partition has full instantaneous dependence if
- $$C(\underline{U}_t^{(C)}) \cdot C(\underline{U}_{t(C)})$$
- $$C(\underline{U}_t^{(C)}) \cdot \tau_t$$
- $$C(\underline{U}_{t(C)}) \cdot \tau_t \quad (P \cdot)$$
- iv) The partition has complete dependence if both ii) and iii) above hold. $(P \leftrightarrow)$

From the above set of definitions it is clear that there is a difference between the static forms of independence and those considered here. In the static case two quantities are either independent or not independent. Here, although the two extremes, independence or full dependence also exist, between them lie other forms of independence/dependence. Schematically we have:-



$$//// = X \rightarrow Y \cap Y \rightarrow X \cap X.Y = X \Leftrightarrow Y.$$

$$::: = X \not\rightarrow Y \cap Y \not\rightarrow X \cap X/Y = X \perp\!\!\!\perp Y$$

Thinking of e.g. $X \perp\!\!\!\perp Y$ as a subset of $X \not\rightarrow Y$ allows us in turn to think of $X \not\rightarrow Y$ as a form of independence. This gives rise to the wording of the definitions 6.11-6.14. For example in definition 6.13 the forms of dependence are described as non-neutral, and the forms of independence as neutral but with e.g. non-causality. Similarly in definition 6.12 it is necessary throughout to be reminded of the conditioning. If $S_c \neq S^C/\tau$ we have a form of conditional independence, but if $S_c = S^C/\tau$ we have a form of conditional dependence.

All of the partition independence properties above have useful applications depending upon the context of the compositional data set under investigation. The use of subcompositional dependence has already been discussed. In order to understand some of the other properties consider again the example of a political opinion

poll. In particular let us assume the following variables.

% preference for	i) Conservatives	(CON)
	ii) Labour	(LAB)
	iii) Liberal	(LIB)
	iv) Don't know	(DK)
	v) Other	(OTH)

If interest is in the relationship between the two major parties the inclusion of the other variables may prove superfluous. Thus we need to check to see if this is the case. If variables iii) - v) in no way influence i) and ii) i.e. if they are 'neutral' we may analyze CON and LAB on their own, via $C(\text{CON}, \text{LAB})$. In the static case we required

$$C(\text{CON}, \text{LAB}) \perp\!\!\!\perp \text{LIB}, \text{DK}, \text{OTH} \quad \text{i.e. } N_1$$

However, here we are primarily concerned that $\text{LIB}, \text{DK}, \text{OTH} \not\perp C(\text{CON}, \text{LAB})$ and possibly $\text{LIB}, \text{DK}, \text{OTH} \not\perp C(\text{CON}, \text{LAB})$, i.e. $N_1 \not\perp$ and $N_1 \not\perp$. If we have $N_1 \rightarrow$, then omitting iii) - v) from our analysis does not lose any information about i) and ii) except that we cannot compare the influence of CON on the variables iii) - v) with that of LAB. If we wish to predict the future value of $C(\text{CON}, \text{LAB})$ e.g. to see who would win the next general election, then even supposing we had $N_1 \cdot$, as well as $N_1 \rightarrow$, we may still only seek to analyze $C(\text{CON}, \text{LAB})$ and ignore LIB, DK and OTH. Thus $N_1 \not\perp$ is only necessary when for example we know the values of LIB, DK and OTH in advance of CON and LAB. Consequently the major neutrality concept in this instance might be $N_1 \not\perp$. If it was the case that $N_1 \leftarrow$ occurred we may still reduce the need to include variables ii) - v) as explanatory variables in the analysis of i) and ii). It may be that the effect of variables iii) - v) is only due to the fact that the preference is not for one of the major parties, and not specifically due to which alternative to the major parties is chosen i.e. we have $T(\text{LIB}, \text{DK}, \text{OTH}) \rightarrow C(\text{CON}, \text{LAB})$ but

$C(\text{LIB, DK, OTH}) \neq C(\text{CON, LAB})/T(\text{LIB, DK, OTH})$. This would allow the analyst to reduce their data set to

- i) CON
- ii) LAB
- iii) NMP = $T(\text{LIB, DK, OTH})$ (Not Major Party)

If interest lies in the smaller parties we would need to investigate N_2 properties. Partition independence would then be an amalgamation of the N_1 and N_2 concepts.

In order to test for the various types of independence we may again utilize the methods outlined in section 2.6. and Geweke (1982, 1984). Also we may define measures of dependence in a similar manner to definition 6.2. Since such measures are a generalization of definitions such as 6.2 we now give only a brief summary of possible measures. As already mentioned subcompositional dependence is the direct analogue of time series dependence of a basis and so its measures may be defined in terms of definition 6.2.

Definition 6.15 : Measures of time series subcompositional dependence.

These measures $F_{(S \rightarrow T)_1}^*$, $F_{(T \rightarrow S)_1}^*$, $F_{(S.T)_1}^*$ and $F_{(S \leftrightarrow T)}^*$, are defined as in definition 6.2. but with the word compositional replaced by subcompositional, u_t by $u_t^{(c)}$, V_t by $C(U_t^{(c)})$ and τ_t by $\tau_t(c)$ with τ in the r.h.s. replaced by $a_1(\tau_t(c))$. Similarly $F_{(S \rightarrow T)_2}^*$ is defined symmetrically in terms of $U_t(c)$.

For example the measure of instantaneous subcompositional dependence on $U_t(c)$ is

$$F_{(S.T)_2}^* = \ln \frac{|\Sigma(V(c)/V(c), p, \tau^*(c), p) / |\Sigma(\tau^*(c)/\tau^*(c), p, V(c), p)|}{|\Sigma(V(c), \tau^*(c)/V(c), p, \tau^*(c), p|}$$

where $\underline{V}(C) = a_{m-C+1}(C(\underline{U}(C)))$, $\tau(C)^* = a_1(\tau(C))$.

Definition 6.16 : Measures of time series conditional subcompositional dependence.

For the partition given by (6.2.10)-(6.2.13), with

$$\tau^* = a_1(\tau_t(C)) \quad , \quad \underline{V}_t^{(C)} = a_{C-\underline{t}}(C(\underline{U}_t^{(C)})) \quad , \quad \underline{V}_t(C) = a_{m-C+1}(C(\underline{U}_t(C)))$$

and as in section 2.6

Z_p the past values $\{Z_{t-j} ; j = 1, 2, \dots\}$

i) The measure of conditional subcompositional causality from $C(\underline{U}_t^{(C)})$ to $C(\underline{U}_t(C))$ is

$$F_{S^C \rightarrow S_C / \tau}^* = \ln \frac{|\Sigma(\underline{V}(C) / \underline{V}(C), p; \tau_p^*)|}{|\Sigma(\underline{V}(C) / \underline{V}(C), p; \tau_p^*; \underline{V}_p^{(C)})|}$$

ii) The measure of conditional instantaneous dependence is

$$F_{S^C . S_C / \tau}^* = \ln \frac{|\Sigma(\underline{V}(C) / \underline{V}(C), p; \underline{V}_p^{(C)}; \tau_p^*)| \cdot |\Sigma(\underline{V}^{(C)} / \underline{V}(C), p; \underline{V}_p^{(C)}; \tau_p^*)|}{|\Sigma(\underline{V}(C), \underline{V}^{(C)} / \underline{V}(C), p; \underline{V}_p^{(C)}, \tau_p^*)|}$$

iii) The measure of complete conditional dependence is

$$F_{S^C \rightleftharpoons S_C / \tau}^* = \ln \frac{|\Sigma(\underline{V}(C) / \underline{V}(C), p, \tau_p^*)| \cdot |\Sigma(\underline{V}^{(C)} / \underline{V}_p^{(C)}; \tau_p^*)|}{|\Sigma(\underline{V}(C), \underline{V}^{(C)} / \underline{V}(C), p; \underline{V}_p^{(C)}; \tau_p^*)|}$$

$$= F_{S^C \rightarrow S_C / \tau}^* + F_{S_C \rightarrow S^C / \tau}^* + F_{S^C . S_C / \tau}^*$$

Definition 6.17 : Measures of time series neutrality.

With the same notation as above.

$$i) F_{N_{1\rightarrow}}^* = \ln \frac{|\underline{\Sigma}(\underline{V}(C), \tau^*/\underline{V}(C), p, \tau_p^*)|}{|\underline{\Sigma}(\underline{V}(C), \tau^*/\underline{V}(C), p, \tau_p^*, \underline{V}_p^{(C)})|}$$

$$ii) F_{N_{1\leftarrow}}^* = \ln \frac{|\underline{\Sigma}(V^{(C)}/\underline{V}_p^{(C)})|}{|\underline{\Sigma}(\underline{V}^{(C)}/\underline{V}_p^{(C)}, \underline{V}(C), p, \tau_p^*)|}$$

$$iii) F_{N_{1\cdot}}^* = \ln \frac{|\underline{\Sigma}(\underline{V}(C), \tau^*/\underline{V}(C), p, \tau_p^*, \underline{V}_p^{(C)})| \cdot |\underline{\Sigma}(V^{(C)}/\underline{V}(C), p, \tau_p^*, \underline{V}_p^{(C)})|}{|\underline{\Sigma}(\underline{V}(C), \tau^*, \underline{V}^{(C)}/\underline{V}(C), p, \tau_p^*, \underline{V}_p^{(C)})|}$$

$$iv) F_{N_{1\Leftarrow}}^* = F_{N_{1\rightarrow}}^* + F_{N_{1\leftarrow}}^* + F_{N_{1\cdot}}^*$$

v) $F_{N_{2\rightarrow}}^*$, $F_{N_{2\leftarrow}}^*$, $F_{N_{2\cdot}}^*$, $F_{N_{2\Leftarrow}}^*$ are as in i)-iv) with $\underline{V}(C)$ and $\underline{V}^{(C)}$ interchanged.

Definition 6.18 : Measures of partition independence

Let $\underline{A} = \{\underline{V}(C), p, \underline{V}_p^{(C)}, \tau_p^*\}$ then

$$i) F_{\underline{P}}^* = \ln \frac{|\underline{\Sigma}(\underline{V}(C)/\underline{A})| \cdot |\underline{\Sigma}(\underline{V}^{(C)}/\underline{A})| \cdot |\underline{\Sigma}(\tau^*/\underline{A})|}{|\underline{\Sigma}(\underline{V}(C), \underline{V}^{(C)}, \tau^*/\underline{A})|}$$

$$ii) F_{\underline{P}\Leftrightarrow}^* = \ln \frac{|\underline{\Sigma}(\underline{V}(C)/\underline{V}(C), p)| \cdot |\underline{\Sigma}(\underline{V}^{(C)}/\underline{V}_p^{(C)})| \cdot |\underline{\Sigma}(\tau^*/\tau_p^*)|}{|\underline{\Sigma}(\underline{V}(C), \underline{V}^{(C)}, \tau^*/\underline{A})|}$$

$$iii) F_{\underline{P}\leftrightarrow}^* = F_{\underline{P}\Leftrightarrow}^* - F_{\underline{P}}^*$$

As before tests of the hypothesis $F^* = 0$ are obtained by comparing $n\hat{F}^*$ with the chi-square distribution with the appropriate degrees of freedom.

All the definitions in this section have referred to the basic partition:-

$$\underline{U}_t \leftrightarrow (C(\underline{U}_t^{(C)}), C(\underline{U}_t(C)), \tau_t)$$

This partition may be mapped onto the real plane via the a_m transformation in such a way as to retain the properties of the partition i.e.

$$\begin{aligned} (C(\underline{U}_t^{(C)}), C(\underline{U}_t(C)), \tau_t) &\leftrightarrow (a_C(C(\underline{U}_t^{(C)})), a_{m-C+1}(C(\underline{U}_t(C))), a_1(\tau_t)) \\ &= (V_t^{(C)}, \underline{V}_t(C), \tau_t^*) \quad t = 0, \pm 1, \dots \end{aligned}$$

The various covariance matrices required in definitions 6.15 through to 6.18 are then the covariance matrices of appropriate multivariate AR(p) models. Consider for example

$$\begin{bmatrix} \phi_{11}(B) & \phi_{12}(B) & \phi_{13}(B) \\ \phi_{21}(B) & \phi_{22}(B) & \phi_{23}(B) \\ \phi_{31}(B) & \phi_{32}(B) & \phi_{33}(B) \end{bmatrix} \begin{bmatrix} \underline{V}_t^{(C)} \\ \underline{V}_t(C) \\ \tau_t^* \end{bmatrix} = \begin{bmatrix} \underline{a}_t^{(C)} \\ \underline{a}_t(C) \\ a_t^* \end{bmatrix}, \quad \begin{array}{l} (6.2.14a) \\ (6.2.14b) \\ (6.2.14c) \end{array}$$

where:-

$\underline{a}_t = (\underline{a}_t^{(C)}, \underline{a}_t(C), a_t^*)'$ is a white noise process with

$$\text{Var}(\underline{a}_t) = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \quad (6.2.14d)$$

Then the measure in Definition 6.17 for example is defined in terms of quantities such as

$$\Sigma(V^{(C)}/\underline{A}) = \Sigma_{11} \quad (6.2.15)$$

$$\text{and } \Sigma(\underline{V}(C), \tau^*/\underline{A}) = \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix} = \underline{\Sigma}^* \quad (6.2.16)$$

Assuming also that

$$\underline{\phi}_{11}^+(B) \underline{V}_t^{(C)} = \underline{a}_t^{(C)+}, \text{ where } \text{var}(\underline{a}_t^{(C)+}) = \underline{\Sigma}_{11}^+ \quad (6.2.17)$$

$$\text{and } \begin{bmatrix} \underline{\phi}_{22}^+(B) & \underline{\phi}_{23}^+(B) \\ \underline{\phi}_{32}^+(B) & \underline{\phi}_{33}^+(B) \end{bmatrix} \begin{bmatrix} \underline{V}_t^{(C)} \\ \tau_t^* \end{bmatrix} = \begin{bmatrix} \underline{a}_t^{(C)+} \\ \underline{a}_t^+ \end{bmatrix},$$

$$\text{where } \text{var} \begin{bmatrix} \underline{a}_t^{(C)+} \\ \underline{a}_t^+ \end{bmatrix} = \begin{bmatrix} \Sigma_{22}^+ & \Sigma_{23}^+ \\ \Sigma_{32}^+ & \Sigma_{33}^+ \end{bmatrix} \quad (6.2.18)$$

Then

$$\Sigma(\underline{V}^{(C)}/\underline{V}_p^{(C)}) = \Sigma_{11}^+ \quad (6.2.19)$$

$$\text{and } \Sigma(\underline{V}(C), \tau^*/\underline{V}(C)p; \tau_p^*) = \begin{bmatrix} \Sigma_{22}^+ & \Sigma_{23}^+ \\ \Sigma_{32}^+ & \Sigma_{33}^+ \end{bmatrix} = \underline{\Sigma}^+ \quad (6.2.20)$$

Comparing (6.2.15) - (6.2.20) with definition 6.17 it follows that

$$F_{N_1}^* \rightarrow = \ln(|\underline{\Sigma}^+|/|\underline{\Sigma}^*|), \quad F_{N,+}^* = \ln(|\underline{\Sigma}_{11}^+|/|\underline{\Sigma}_{11}|), \quad F_{N_1}^* = \ln(|\underline{\Sigma}_{11}| \cdot |\underline{\Sigma}^*|/|\underline{\Sigma}|)$$

Thus these measures involve the comparison of (6.2.14), which jointly models $\underline{V}_t^{(C)}$ and $(\underline{V}_t^{(C)}, \tau_t^*)$, with that of (6.2.17) and (6.2.18), which represent independent models for $\underline{V}_t^{(C)}$ and $(\underline{V}_t^{(C)}, \tau_t^*)$. Clearly if $\underline{\phi}_{12}(B) = \underline{0}$, $\underline{\phi}_{13}(B) = \underline{0}$, then (6.2.14a) is identical to (6.2.17) with $\underline{\phi}_{11}(B) = \underline{\phi}_{11}^+(B)$, $\underline{\Sigma}_{11} = \Sigma_{11}^+$ so that $F_{N_1}^* \rightarrow = 0$. Hence we

may regard N_{1+} as a parametric hypothesis about model (6.2.14). Similarly N_{1+} is equivalent to $\underline{\phi}_{21}(B) = \underline{0}$ and $\underline{\phi}_{31}(B) = \underline{0}$, N_1 to $\underline{\Sigma}_{12} = \underline{0}$ and $\underline{\Sigma}_{13} = \underline{0}$.

A full list of the parametric hypothesis using (6.2.14) is given in table 6.19.

Table 6.19 : Summary of partition independence properties.

	Independence Property	Parametric hypothesis								
		$\underline{\phi}_{12}(B)$	$\underline{\phi}_{21}(B)$	$\underline{\phi}_{13}(B)$	$\underline{\phi}_{31}(B)$	$\underline{\phi}_{23}(B)$	$\underline{\phi}_{32}(B)$	$\underline{\Sigma}_{12}$	$\underline{\Sigma}_{13}$	$\underline{\Sigma}_{23}$
Conditional subcompositional dependence	$S^C \perp\!\!\!\perp S_C/\tau$	$\underline{=0}$	$\underline{=0}$					$\underline{=0}$		
	$S^C \rightarrow S_C/\tau$		$\underline{\neq 0}$							
	$S^C \cdot S_C/\tau$							$\underline{\neq 0}$		
	$S^C \leftrightarrow S_C/\tau$	$\underline{\neq 0}$	$\underline{\neq 0}$							
	$S^C \Leftrightarrow S_C/\tau$	$\underline{\neq 0}$	$\underline{\neq 0}$					$\underline{\neq 0}$		
Neutrality	$N_1 \perp\!\!\!\perp$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$			$\underline{=0}$	$\underline{=0}$	
	$N_1 \rightarrow$		$\underline{\neq 0}$		$\underline{\neq 0}$					
	$N_1 \leftarrow$	$\underline{\neq 0}$		$\underline{\neq 0}$						
	$N_1 \cdot$							$\underline{\neq 0}$	$\underline{\neq 0}$	
	$N_1 \leftrightarrow$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$					
	$N_1 \Leftrightarrow$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$			$\underline{\neq 0}$	$\underline{\neq 0}$	
Partition	$P \perp\!\!\!\perp$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$	$\underline{=0}$
	$P \leftrightarrow$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$			
	$P \cdot$							$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$
	$P \Leftrightarrow$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$	$\underline{\neq 0}$

Other hypotheses may be extracted from the table, N_2 is the symmetric version of N_1 with $\underline{\phi}_{13}(B)$, $\underline{\phi}_{31}(B)$ and $\underline{\Sigma}_{13}$ replaced by $\underline{\phi}_{23}(B)$, $\underline{\phi}_{32}(B)$ and $\underline{\Sigma}_{23}$. Also the opposite hypothesis is obtained by replacing " $= \underline{0}$ " by " $\neq \underline{0}$ " and vice-versa, e.g.

$$N_1 \neq : \underline{\phi}_{21}(B) = \underline{0}, \underline{\phi}_{31}(B) = \underline{0}.$$

For subcompositional dependence we need to consider the following models.

$$\begin{bmatrix} \underline{\phi}_{11}^*(B) & \underline{\phi}_{13}^*(B) \\ \underline{\phi}_{31}^*(B) & \underline{\phi}_{33}^*(B) \end{bmatrix} \begin{bmatrix} \underline{v}_t^{(C)} \\ \tau_t^* \end{bmatrix} = \begin{bmatrix} \underline{b}_t^{(C)} \\ b_t^* \end{bmatrix}, \quad (6.2.21a)$$

$$\text{where Var} \begin{bmatrix} \underline{b}_t^{(C)} \\ b_t^* \end{bmatrix} = \begin{bmatrix} \underline{\Sigma}_{11}^* & \underline{\Sigma}_{13}^* \\ \underline{\Sigma}_{31}^* & \underline{\Sigma}_{33}^* \end{bmatrix} \quad (6.2.21c)$$

and

$$\begin{bmatrix} \underline{\phi}_{22}^*(B) & \underline{\phi}_{23}^*(B) \\ \underline{\phi}_{32}^*(B) & \underline{\phi}_{33}^*(B) \end{bmatrix} \begin{bmatrix} \underline{v}_t^{(C)} \\ \tau_t^* \end{bmatrix} = \begin{bmatrix} \underline{b}_t^{(C)} \\ b_t^+ \end{bmatrix}, \quad (6.2.22a)$$

$$\text{where Var} \begin{bmatrix} \underline{b}_t^{(C)} \\ b_t^+ \end{bmatrix} = \begin{bmatrix} \underline{\Sigma}_{22}^* & \underline{\Sigma}_{23}^* \\ \underline{\Sigma}_{32}^* & \underline{\Sigma}_{33}^+ \end{bmatrix} \quad (6.2.22c)$$

The subcompositional dependence properties may now be defined in terms of these models. The off-diagonal elements in the $\underline{\phi}$ and $\underline{\Sigma}$ matrices are responsible for the various relationships between $\underline{v}_t^{(C)}$ and τ_t^* in (6.2.21) and (6.2.22).

Table 6.20 : Summary of subcompositional dependence properties.

Independence Property	Parametric hypothesis		
	$\underline{\phi}_{13}^*(B)$ ($\underline{\phi}_{23}^*(B)$),	$\underline{\phi}_{31}^*(B)$ ($\underline{\phi}_{32}^*(B)$),	$\underline{\Sigma}_{13}$ ($\underline{\Sigma}_{23}^*$)
$S \perp\!\!\!\perp T_1$ ($S \perp\!\!\!\perp T_2$)	$= \underline{0}$	$= \underline{0}$	$= \underline{0}$
$S \rightarrow T_1$ ($S \rightarrow T_2$)		$\neq \theta$	
$T \rightarrow S_1$ ($S \rightarrow S_2$)	$\neq \underline{0}$		
$S \leftrightarrow T_1$ ($S \leftrightarrow T_2$)	$\neq \underline{0}$	$\neq \underline{0}$	
$S.T_1$ ($S.T_2$)			$\neq \underline{0}$
$S \Leftrightarrow T_1$ ($S \Leftrightarrow T_2$)	$\neq \underline{0}$	$\neq \underline{0}$	$\neq \underline{0}$

We note that model (6.2.21) is equivalent to (6.2.14) for the $\underline{V}_t^{(C)}$ and τ_t^* series if $\underline{\phi}_{12}(B) = \underline{0}$, $\underline{\phi}_{32}(B) = \underline{0}$, i.e. $N_2 \neq$, in which case $\underline{\phi}_{13}(B) = \underline{\phi}_{13}^*(B)$, $\underline{\phi}_{31}(B) = \underline{\phi}_{31}^*(B)$ and $\underline{\Sigma}_{13} = \underline{\Sigma}_{13}^*$. Similarly the equations (6.2.22) are equivalent to (6.2.14b) and (6.2.14c) if $\underline{\phi}_{21}(B) = \underline{0}$, $\underline{\phi}_{31}(B) = \underline{0}$ i.e. if $N_1 \neq$. Now recall that

$$N_2 \perp\!\!\!\perp \subset N_2 \not\leftrightarrow \subset N_2 \neq$$

Hence under all these neutrality types (6.2.14a) and (6.2.14c) remain identical to (6.2.21a) and (6.2.22b). In going from $N_2 \neq$ to $N_2 \not\leftrightarrow$ we required $\underline{\phi}_{21}(B) = \underline{0}$, and $\underline{\phi}_{23}(B) = \underline{0}$, and in going on to $N_2 \perp\!\!\!\perp$ $\underline{\Sigma}_{12} = \underline{0}$, $\underline{\Sigma}_{23} = \underline{0}$, which affect (6.2.14b) and (6.2.14d) only. The only off-diagonal ϕ 's in (6.2.14a) and (6.2.14c) or now equivalently (6.2.21) are $\underline{\phi}_{13}(B)$ and $\underline{\phi}_{31}(B)$, and from table 6.10 we see that

$$\begin{aligned}
P \not\leftrightarrow &\equiv N_2 \not\leftrightarrow \cap (S \not\leftrightarrow T)_1, \text{ similarly} \\
&\equiv N_1 \not\leftrightarrow \cap (S \not\leftrightarrow T)_2, \\
&\equiv N_1 \not\leftrightarrow \cap N_2 \not\leftrightarrow.
\end{aligned}
\tag{6.2.23}$$

Also we have,

$$\begin{aligned}
P \perp\!\!\!\perp &\equiv N_2 \perp\!\!\!\perp \cap (S \perp\!\!\!\perp T)_1, \\
&\equiv N_1 \perp\!\!\!\perp \cap (S \perp\!\!\!\perp T)_2 \\
&\equiv N_1 \perp\!\!\!\perp \cap N_2 \perp\!\!\!\perp.
\end{aligned}
\tag{6.2.24}$$

We may compare (6.2.23) and (6.2.24) with the equivalent result of Aitchison (1982) that of (3.5.23). A similar result to (3.5.22) is obtained if we note that

$$S_C \not\leftrightarrow S^C/\tau \equiv N_2 \not\leftrightarrow, \tag{6.2.25a}$$

$$\text{and } S_C \not\leftrightarrow S^C/\tau \equiv N_1 \not\leftrightarrow \tag{6.2.25b}$$

(6.2.25a) refers to the fact above that $N_2 \not\leftrightarrow$ makes (6.2.14) for $\underline{V}_t^{(C)}$ and τ_t^* equivalent to (6.2.21). If only $S^C \rightarrow S^C/\tau$ occurs i.e. $\Phi_{12}(B) = \underline{0}$, then only (6.2.14a) is equivalent to (6.2.21a). Again comparing with table 6.19 we have that

$$N_1 \leftarrow \equiv S_C \not\leftrightarrow S^C/\tau \cap (T \not\leftrightarrow S)_1 \text{ similarly} \tag{6.2.26a}$$

$$N_2 \not\leftrightarrow \equiv S^C \not\leftrightarrow S_C/\tau \cap (T \not\leftrightarrow S)_2 \tag{6.2.26b}$$

(6.2.26a) also gives rise to (6.2.25b).

However we can go no further e.g. it is not the case that

$$N_1 \not\leftrightarrow \equiv S_C \not\leftrightarrow S^C/\tau \cap (T \not\leftrightarrow S)_1$$

Although $S_C \not\leftrightarrow S^C/\tau$ reduces (6.2.14a) to (6.2.21a) it does not reduce (6.2.14c) to (6.2.21b). Thus the requirement of $N_1 \leftrightarrow$ that $\Phi_{31}(B) = 0$ cannot be met via $(T \leftrightarrow S)$, since in general $\Phi_{31}^*(B) \neq \Phi_{31}(B)$. Thus we must conclude that whilst (3.5.22) states that

$$N_1 \perp\!\!\!\perp \equiv S_C \perp\!\!\!\perp S^C/\tau \wedge (T \perp\!\!\!\perp S)_1, \quad ,$$

in the time series context this is not true.

It should be noted that (6.2.26) may be derived using the result of Geweke (1984), that for three series X_t , Y_t and Z_t , with obvious notation that

$$F_{X \rightarrow Y/Z} = F_{XZ \rightarrow Y} - F_{Z \rightarrow Y} \text{ i.e.} \tag{6.2.27}$$

$$F_{XZ \rightarrow Y} = F_{X \rightarrow Y/Z} + F_{Z \rightarrow Y},$$

whence

$$F_{S_C, \tau \rightarrow S^C} = F_{S_C \rightarrow S^C/\tau} + F_{\tau \rightarrow S^C}$$

These are the measures relating to (6.2.26a).

The main reason for being interested in expressions such as (6.2.23), (6.2.24) and (6.2.26) is that it enables us to produce a "lattice" of hypotheses similar to that produced by Aitchison (1982). This allows us to use the simpler types of independence as building blocks for higher forms of independence, and in the other direction when a stronger form of independence does not hold we may find a weaker one that does. Thus, for example the NMP (Not Major Party) variable above would be a way of reducing the opinion poll data if neutrality did not hold but the weaker conditional non-causality did. In the light of this it is a pity that (3.5.22) does not hold in the time series context. However again referring back to the

opinion poll, the need to reduce the data set in many contexts required only $N_1 \neq$ or $N_2 \neq$ so that (6.2.26) perhaps offers the most useful combination of hypotheses.

The difference between Aitchison's (1982) results and those considered here primarily lies in the auto-correlated nature of the data and the added difficulty in testing for independence of such series. If we compare (3.5.25) with (6.2.14), (6.2.21) or (6.2.22) and (3.5.26) with table 6.19 or 6.20 an insight into this is gained. We may postulate that $\underline{\beta}_1$ and $\underline{\beta}_2$ in (3.5.25) are equivalent to $\underline{\phi}_{13}(B)$ and $\underline{\phi}_{23}(B)$ in (6.2.14) or $\underline{\phi}_{13}^*(B)$ and $\underline{\phi}_{23}^*(B)$ in (6.2.21) and (6.2.22) respectively. Similarly the $\underline{\Sigma}_{12}$'s in the models may be thought to correspond. In (3.5.25) these parameters perfectly explain all the required relationships, the $\underline{\beta}$'s the subcompositional independence properties, and $\underline{\Sigma}_{12}$ the conditional subcompositional independence property. However, now, in addition to the analogue of the $\underline{\beta}$'s the $\underline{\phi}_{13}$'s and $\underline{\phi}_{23}$'s we also have $\underline{\phi}_{31}$, $\underline{\Sigma}_{13}$, $\underline{\phi}_{32}$ and $\underline{\Sigma}_{23}$. Each one representing a different direction of the subcompositional independence property. Similarly in addition to $\underline{\Sigma}_{12}$ we have $\underline{\phi}_{12}$ and $\underline{\phi}_{21}$ representing the conditional subcompositional property. Thus in Aitchison's model one property is modelled via the mean and the other by the covariance matrix. In the time series case both forms of independence are modelled by both types of parameters. One may try to form an alternative model where this is not the case, but this would become unnecessarily complicated in time series modelling terms.

To illustrate this point we may combine (6.2.21a) and (6.2.22a) to give us

$$\underline{\phi}_{11}^*(B) \underline{v}_t^{(C)} + \underline{\phi}_{13}^*(B) \tau_t^* = \underline{b}_t^{(C)} \quad (6.2.29a)$$

$$\underline{\phi}_{22}^*(B) \underline{v}_t^{(C)} + \underline{\phi}_{23}^*(B) \tau_t^* = \underline{b}_t^{(C)} \quad (6.2.29b)$$

In such a model we would require both $\underline{\phi}_{13}^*(B)$ and $\underline{\phi}_{23}^*(B)$ to be a lagged polynomial in positive and zero lags as well as negative lags i.e.

$$\underline{\phi}_{i3}^*(B) = \sum_{j=-\infty}^{\infty} \underline{\phi}_{i3,j} B^j \quad \text{instead of the previous} \quad (6.2.30)$$

$$\underline{\phi}_{i3}^*(B) = \sum_{j=1}^{\infty} \underline{\phi}_{i3,j} B^j \quad \text{for } i = 1, 2.$$

This is necessary in order to incorporate $V \rightarrow \tau$ and $v \rightarrow \tau$ as well as $\tau \rightarrow v$. The conditional subcompositional property would have to be modelled via $\underline{b}_t^{(C)}$ and $\underline{b}_t(C)$ and their cross-correlation. Again, to allow for the varying directions, both series will be auto-correlated making it necessary to examine a further model of the form

$$\begin{bmatrix} \underline{\psi}_{11}(B) & \underline{\psi}_{12}(B) \\ \underline{\psi}_{21}(B) & \underline{\psi}_{22}(B) \end{bmatrix} \begin{bmatrix} \underline{b}_t^{(C)} \\ \underline{b}_t(C) \end{bmatrix} = \begin{bmatrix} \underline{e}_t^{(C)} \\ \underline{e}_t(C) \end{bmatrix},$$

where now $\begin{bmatrix} \underline{e}_t^{(C)} \\ \underline{e}_t(C) \end{bmatrix}$ is

$$\text{white noise with variance-covariance} \begin{bmatrix} \underline{\Sigma}_{e11} & \underline{\Sigma}_{e12} \\ \underline{\Sigma}_{e21} & \underline{\Sigma}_{e22} \end{bmatrix} \quad (6.2.31)$$

Clearly (6.2.29) through to (6.2.31) are much more complicated to estimate. It is possible to explore such a model further and relate it back to our earlier models. To a certain extent Geweke (1984) does this when he forms a measure

of F decomposed by frequency. Thus the feasibility of (6.2.29) - (6.2.31) can be explored further. For the time being we prefer to stick to the simpler models as it seems unlikely that we will gain much by not doing so. The above model does serve as a useful comparison, however to (3.5.25).

Our final form of partition independence is the natural extension of definition 3.31.

Definition 6.21 :

a) $\underline{U}_t = (\underline{U}_t^{(C)}, \underline{U}_{t(C)}) \in \mathcal{S}^m$ has partial subcompositional independence restricted by $\underline{U}_t^{(C)}$ if

i) $C(\underline{U}_t^{(C)}) \perp\!\!\!\perp C(\underline{U}_{t(C)})$

ii) If $C(\underline{U}_{t(C)}) \in \mathcal{S}^{m-C}$ has complete subcompositional independence. $S^C \perp\!\!\!\perp (S_{(C)} \textcircled{\perp\!\!\!\perp})$.

b) $\underline{U}_t \in \mathcal{S}^m$ has partial subcompositional independence with causality to $\underline{U}_t^{(C)}$ if

i) $C(\underline{U}_t^{(C)}) \neq C(\underline{U}_{t(C)})$

ii) If $C(\underline{U}_{t(C)}) \in \mathcal{S}^{m-C}$ has complete subcompositional independence. $S^C \not\perp\!\!\!\perp (S_{(C)} \textcircled{\perp\!\!\!\perp})$.

Equation a) is identical to definition 3.31 except that we have \underline{U}_t $t = 0, \pm 1, \dots$ here. The second form of independence; b) is one in which $C(\underline{U}_{t(C)})$ is allowed to stand on its own as consisting of a subcomposition with

completed subcompositional independence. However, it is allowed to cause $C(\underline{U}_t^{(C)})$, be instantaneously related to $C(\underline{U}_t^{(C)})$, but $C(\underline{U}_t^{(C)})$ is not allowed to cause it (i.e. $C(\underline{U}_t^{(C)})$). One might define further forms of partial independence in which $C(\underline{U}_t^{(C)}) \rightarrow C(\underline{U}_t^{(C)})$, but conceptually this seems to make little sense. The basic idea behind making $C(\underline{U}_t^{(C)})$ have complete subcompositional independence is to try to capture the non-compositional idea that the elements of $\underline{U}_t^{(C)}$ are all independent of one another. To make these elements dependent on the remaining subcomposition $\underline{U}_t^{(C)}$ whilst this is so seems to detract from the basic idea.

The corresponding parametric hypothesis follows as before. For the model

$$\begin{bmatrix} \underline{\phi}_{11}(B) & \underline{\phi}_{12}(B) \\ \underline{\phi}_{21}(B) & \underline{\phi}_{22}(B) \end{bmatrix} \begin{bmatrix} \underline{U}_t^{(C)} \\ \underline{U}_t^{(C)} \end{bmatrix} = \begin{bmatrix} \underline{e}_t^{(C)} \\ \underline{e}_t^{(C)} \end{bmatrix} = \underline{e}_t$$

$$= \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$$

We require under $S^C \perp\!\!\!\perp (S_{(C)} \textcircled{\perp\!\!\!\perp})$ that $\underline{\phi}_{12}(B) = \underline{0}$, $\underline{\phi}_{21}(B) = \underline{0}$, $\underline{\Sigma}_{12} = \underline{0}$ and $\underline{\phi}_{22}(B), \underline{\Sigma}_{22} \in X$.

and under $S^C \not\perp\!\!\!\perp (S_{(C)} \textcircled{\perp\!\!\!\perp})$ that $\underline{\phi}_{21}(B) = \underline{0}$ and $\underline{\phi}_{22}(B), \underline{\Sigma}_{22} \in X$.

Where X is the parametric set required for complete subcompositional independence.

$$\text{e.g. } \underline{\Sigma}_{22} = \text{diag}(\sigma_1, \dots, \sigma_m) + \sigma_{m+1} \underline{U}$$

However, we again have the difficulty of testing for

complete subcompositional independence, so that both $S_{\perp(S_{(C)})}^C$ and $S_{\perp(S_{(C)})}^C$ will prove difficult to detect in practice.

6.2.3 Progressing Partitions

In the previous section we examined various partition independence properties. Often we may wish to examine several partitions of U_t , and in particular "progress" through the composition. For example consider our opinion poll data again and order it thus

"Don't know", "Other", "Lib", "Lab", "Con"

Suppose we are interested only in examining the relationship between the two major parties, but wish to be as comprehensive in our approach as possible so that all variables are examined. If, however, the additional variables are superfluous to the Lab, Con relationship we may omit them from our analysis. We might believe that the variable most likely not to be superfluous is the next major party, after that 'other' parties, and finally the "Don't know" variable to play little part in determining the Lab Con conflict. If this were so we may progressively work through the composition to see which variables are not needed. We would examine the following partitions

- a) ("Don't know" | C(Other, Lib, Lab, Con))
- b) (C(Don't know, Other) | T(Don't know, Other) | C(Lib, Lab, Con))
- c) (C(Don't know, Other, Lib) | T(Don't know, Other, Lib) | C(Lab, Con))

If for a), b) and c) we have $N_2 \neq$ then we need only examine C(Lab, Con). However we might find, for example that N_2 only held for a) and b) so that examination of C(Lib, Lab, Con) was necessary; and so on.

What would therefore be useful is a way of describing

this progression through partitions. This comes via the idea of order given below.

Definition 6.22

For some specific ordered $\underline{U}_t \in \mathcal{S}^m : (U_1, U_2, \dots, U_{m+1})'_t$, \underline{U}_t has independence property of order k if the stated independence property, as given by definitions 6.11-6.14 holds for $C = 1, \dots, k$ in the partition $(\underline{U}_t^{(C)}, \underline{U}_{t(C)}, \tau_{t(C)})$.

If $k = m$ then \underline{U}_t is said to have complete independence of the form stated.

As in section 3.5 it is rather cumbersome to try to test for an independence property of order k using the methods above. The solution previously was to employ the m_m transformation rather than a_m . We therefore consider the $\&n_x$ ARMA $_m(P,q)$ model.

Let $\underline{x}_t = m_m(\underline{U}_t)$, and

$$\begin{aligned} \underline{\Phi}(B) \underline{x}_t &= \begin{bmatrix} \underline{\Phi}_{11}(B)_{k \times k} & \underline{\Phi}_{12}(B)_{k \times (m-k)} \\ \underline{\Phi}_{21}(B)_{(m-k) \times k} & \underline{\Phi}_{22}(B)_{(m-k) \times (m-k)} \end{bmatrix} \begin{bmatrix} \underline{x}_t^{(k)}_{k \times 1} \\ \underline{x}_t^{(k)}_{(m-k) \times 1} \end{bmatrix} \\ &= \begin{bmatrix} \underline{\varepsilon}_t^{(k)} \\ \underline{\varepsilon}_t^{(k)} \end{bmatrix} = \underline{\varepsilon}_t, \end{aligned} \tag{6.2.31}$$

where $\text{Var}[\underline{\varepsilon}_t] = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$

Various forms of independence are represented in table 6.23. The equivalent for left neutrality is obtained by reversing the order of \underline{U}_t before taking the m_m transformations.

no a priori knowledge of C is required, it does require knowledge of the most likely variables that can be dropped so as to order \underline{U}_t .

6.2.4. : Some numerical examples of partition independence

In example 4.18 and section 4.4 we analyzed two data sets of political opinion polls. The GALLUP(C) poll and the N.O.P. We recall that both data sets consisted of 108 observations on 4 variables:-

1. CON_t
2. LAB_t
3. LIB_t
4. OTH_t

Because of the frequency of zeros occurring in the 4th series OTH_t we choose to model

$$C(CON, LAB, LIB)$$

The accuracy of such an approach rests on the neutrality of OTH_t to CON_t , LAB_t and LIB_t . What we have is the partition

$$(CON, LAB, LIB | OTH) = (\underline{U}_t^{(C)} | \underline{U}_t(C))$$

from which we may form:-

$$C(CON LAB LIB) = C(\underline{U}_t^{(C)}) = \underline{U}_t$$

$$OTH = \tau_t(C) \quad (= \underline{U}_t(C))$$

Since the second subcomposition consists of only one variable and hence $\tau_t(C) = \underline{U}_t(C)$, the two concepts $(S \sim T)_1$, and $N_1 \sim$ merge into one. Further since $C(OTH) = 1$, $S_C \perp\!\!\!\perp S_C/\tau$ is trivially satisfied.

Thus we wish to investigate the relationship between \underline{U}_t and $\tau_t(C)$. The first stage is to map our two series onto the real space via the a_m transformation. Let

$$\underline{V}_t = a_2(\underline{U}_t) = \begin{pmatrix} \ln(\text{CON}_t/\text{LIB}_t) \\ \ln(\text{LAB}_t/\text{LIB}_t) \end{pmatrix} \quad \text{say}$$

$$\tau_t = a_1(\tau_t(C)) = \ln(\tau_t(C)/1-\tau_t(C))$$

As mentioned in section 4.4, one way to avoid zero values of OTH_t is to reset it to a small position value. This we have done, replacing 0.0 by 0.05. With the data thus transformed the following AR models were fitted:-

AR₂(2) to the series \underline{V}_t

AR₁(2) to the series τ_t

and

AR₃(2) to the series $(\underline{V}_t', \tau_t)'$.

We will let the error covariance matrices for these models be defined as

$$\underline{\Sigma}_1(2 \times 2) \quad \underline{\Gamma}_1(1 \times 1) \quad \text{and} \quad \underline{\Gamma}_{3 \times 3} = \begin{bmatrix} \underline{\Sigma}_2 & \underline{C} \\ \underline{C}' & \underline{\Gamma}_2 \end{bmatrix} \quad \text{respectively.}$$

Using these we may construct the various estimates of measures of neutrality.

$$\hat{F}_{V \leftrightarrow \tau} = \hat{F}_{N_1}^* = \log(|\hat{T}_1| \cdot |\hat{\Sigma}_1| / |\hat{\Gamma}|)$$

$$\hat{F}_{V \rightarrow \tau} = \hat{F}_{N_1}^* = \log(|\hat{T}_1| / |\hat{T}_2|)$$

$$\hat{F}_{\tau \rightarrow V} = \hat{F}_{N_1}^* = \log(|\hat{\Sigma}_1| / |\hat{\Sigma}_2|)$$

$$\hat{F}_{V.\tau} = \hat{F}_{N_1}^* = \log(|\hat{\Sigma}_2| \cdot |\hat{T}_2| / |\hat{I}|)$$

Also

$$n\hat{F}_{N_1}^* \Leftrightarrow \sim \chi_{10}^2$$

$$n\hat{F}_{N_1}^* \rightarrow \sim \chi_4^2$$

$$n\hat{F}_{N_1}^* \leftarrow \sim \chi_4^2$$

$$n\hat{F}_{N_1}^* \sim \chi_2^2$$

under the hypothesis N_1^c , $F_{N_1}^* \sim = 0$ for $\sim \in \{\Leftrightarrow \rightarrow \leftarrow \}$.

The AR(2) model was chosen since the various F measures have been developed on the assumption of autoregressive models. Thus although the ARMA(1,1) model was used previously in section 4.4. here we use the AR(2) since it produces a reasonable fit and it provides a good approximation to the ARMA(1,1) model.

The models were fitted using the WMTS-1 package and MINITAB for the multivariate and univariate models respectively. The results for the two data sets are given in table 6.24.

Table 6.24 : Estimates of T_1 , $\underline{\Sigma}_1$, $\underline{\Gamma}$ and the F measures for the GALLUP(C) series and N.O.P. series.

	GALLUP(C) Series	N.O.P. Series
$T_1(T_1)$	0.6735	0.4319
$\underline{\Sigma}_1$	$\begin{bmatrix} 0.0451 & 0.0396 \\ 0.0396 & 0.0499 \end{bmatrix}$	$\begin{bmatrix} 0.0391 & 0.0342 \\ 0.0342 & 0.0407 \end{bmatrix}$
$ \underline{\Sigma}_1 $	6.86939×10^{-4}	4.20541×10^{-4}
$\underline{\Gamma}$	$\begin{bmatrix} 0.0448 & 0.0391 & -0.0108 \\ 0.0391 & 0.0491 & -0.0181 \\ -0.0108 & -0.0181 & 0.4756 \end{bmatrix}$	$\begin{bmatrix} 0.0360 & 0.0317 & -0.0293 \\ 0.0317 & 0.0382 & -0.0334 \\ -0.0293 & -0.0334 & 0.3815 \end{bmatrix}$
$ \underline{\Gamma} $	3.11550×10^{-4}	1.29206×10^{-4}
$ \underline{\Sigma}_2 $	6.65904×10^{-4}	3.66963×10^{-4}
$ T_2 $	0.47555	0.381549
n	108	108
$\hat{nF}_{N_1}^* \Leftrightarrow \sim \chi_{10}^2$	42.70***	36.78***
$\hat{nF}_{N_{1+}} \sim \chi_4^2$	37.58***	13.38***
$\hat{nF}_{N_{1+}} \sim \chi_4^2$	3.36	14.72***
$\hat{nF}_{N_1} \sim \chi_2^2$	1.76	8.68**

Looking at the results we see that the GALLUP(C) poll has only N_{1+} , whilst the N.O.P. series has N_{\leftrightarrow} . Thus our models in 4.4 are adequate for the GALLUP(C) series, whilst the N.O.P. series might well be predicted more accurately if OTH_t were included.

The difference between the two series may occur for a variety of reasons. Firstly, it may be due to the large number of zeros present in the OTH_t series. When $OTH_t \neq 0$ the values in the GALLUP(C) series tended to be somewhat smaller than in the N.O.P. data. This makes OTH_t very "flat" in the first case, causing it to add little to the remaining variables. The fact that OTH_t is often zero distorts the true situation in any case. Secondly, the WMTS-1 package proved to be highly unreliable making its results very suspect. Finally, there may be differences in the two surveys, e.g. wording of the questions asked, sampling design etc. If the results here are accurate then an investigation into these latter possibilities may shed light on why the difference is occurring.

For our second example we use GALLUP(A) series. This series is based on a slightly different set of questions than the (C) series, which allows for the response "Don't know". It thus contains the five variables:-

i) - (iv) as before

v) Don't know DTK_t

In this series OTH_t again had to be doctored to remove the zeros. The DTH_t variable remained non-zero throughout. Using this series we are able to distinguish between $S \sim T$ and $N \sim$ and examine $S^C \sim S_C/\tau$ on the same partition as above, but with an extra variable in the $\underline{U}_t(C)$ set. Whence we have:-

$$CON \quad LAB \quad LIB \quad | \quad OTH \quad DTK \quad = \quad \underline{U}_t^{(C)} \quad | \quad \underline{U}_t(C)$$

from which we may form

$$C(\text{CON LAB LIB}) = C(\underline{U}_t^{(C)}) = \underline{U}_t^{*(C)}$$

$$C(\text{OTH DTK}) = C(\underline{U}_{t(C)}) = \underline{U}_{t(C)}^*$$

$$T(\text{CON LAB LIB}) = \tau_t^{(C)} = 1 - \tau_{t(C)}$$

Transforming these from the \mathcal{S}^m space to the \mathcal{R}^m space gives

$$\underline{X}_t = a_2(\underline{U}_t^{*(C)}) = \begin{matrix} \log(\text{CON}_t/\text{LIB}_t) \\ \log(\text{LAB}_t/\text{LIB}_t) \end{matrix}$$

$$\underline{Y}_t = a_1(\underline{U}_{t(C)}^*) = \log\left(\frac{\text{OTH}_t}{\text{DTK}_t}\right)$$

$$\underline{Z}_t = a_1(\tau_t^{(C)}) = \log \tau_t^{(C)} / \tau_{t(C)}$$

To compute the various F measures we fitted AR(2) models to the following combinations of variables, with covariance matrices $\underline{\Sigma}$

$$1. \begin{pmatrix} \underline{X}_t \\ \underline{Y}_t \\ \underline{Z}_t \end{pmatrix} \underline{\Sigma} = \underline{A} = \begin{bmatrix} \underline{AD} & \underline{A}_{12} & \underline{A}_{13} \\ \underline{A}_{21} & \underline{AE} & \underline{A}_{23} \\ \underline{A}_{31} & \underline{A}_{32} & \underline{AF} \end{bmatrix} = \begin{bmatrix} \underline{AD} & \underline{A}_{12} & \underline{A}_{13} \\ \underline{A}_{21} & & \\ \underline{A}_{31} & \underline{AC} & \end{bmatrix}; \underline{AB} = \begin{bmatrix} \underline{AD} & \underline{A}_{13} \\ \underline{A}_{31} & \underline{AF} \end{bmatrix}$$

$$2. \begin{pmatrix} \underline{X}_t \\ \underline{Z}_t \end{pmatrix} \underline{\Sigma} = \underline{B} = \begin{bmatrix} \underline{BD} & \underline{B}_{12} \\ \underline{B}_{21} & \underline{BF} \end{bmatrix}$$

$$3. \begin{pmatrix} \underline{Y}_t \\ \underline{Z}_t \end{pmatrix} \underline{\Sigma} = \underline{C} = \begin{bmatrix} \underline{CE} & \underline{C}_{12} \\ \underline{C}_{21} & \underline{CF} \end{bmatrix}$$

$$4. \underline{X}_t \underline{\Sigma} = \underline{D},$$

$$5. \underline{Y}_t \underline{\Sigma} = \underline{E},$$

$$6. \underline{Z}_t \underline{\Sigma} = \underline{F}.$$

The measures computed were then

$$\hat{nF}_{X \rightarrow Z} = \hat{nF}_{(S \rightarrow T)_1}^* = 108 \times \log \left(\frac{|\hat{F}|}{|\hat{BF}|} \right) = 15.163^{***} \sim \chi_4^2$$

$$\hat{nF}_{Z \rightarrow X} = \hat{nF}_{(T \rightarrow S)_1}^* = 108 \times \log \left(\frac{|\hat{D}|}{|\hat{BD}|} \right) = 12.037^* \sim \chi_4^2$$

$$\hat{nF}_{X.Z} = \hat{nF}_{(T.S)_1}^* = 108 \times \log \left(\frac{|\hat{BD}| \cdot |\hat{BF}|}{|\hat{B}|} \right) = 9.420^{**} \sim \chi_2^2$$

$$\hat{nF}_{Y \rightarrow Z} = \hat{nF}_{(S \rightarrow T)_2}^* = 108 \times \log \left(\frac{|\hat{F}|}{|\hat{CF}|} \right) = 17.442^{***} \sim \chi_2^2$$

$$\hat{nF}_{Z \rightarrow Y} = \hat{nF}_{(T \rightarrow S)_2}^* = 108 \times \log \left(\frac{|\hat{E}|}{|\hat{CE}|} \right) = -4.544 \equiv 0.000 \sim \chi_2^2$$

$$\hat{nF}_{Y.Z} = \hat{nF}_{(T.S)_2}^* = 108 \times \log \left(\frac{|\hat{CF}| \cdot |\hat{CE}|}{|\hat{C}|} \right) = 0.009 \sim \chi_1^2$$

$$\hat{nF}_{X \rightarrow Y/Z} = \hat{nF}_{(S^C \rightarrow S_C/\tau)}^* = 108 \times \log \left(\frac{|\hat{CE}|}{|\hat{AE}|} \right) = -8.263 \equiv 0.000 \sim \chi_4^2$$

$$\hat{nF}_{Y \rightarrow X/Z} = \hat{nF}_{(S_c \rightarrow S^C/\tau)}^* = 108 \times \log \left(\frac{|\hat{BD}|}{|\hat{AD}|} \right) = -12.204 \equiv 0.000 \sim \chi_4^2$$

$$\hat{nF}_{X.Y/Z} = \hat{nF}_{(S^C.S_C/\tau)}^* = 108 \times \log \left(\frac{|\hat{AD}| \cdot |\hat{AE}|}{\left| \frac{\hat{AD}}{\hat{A}_{-21}} \frac{\hat{A}_{12}}{\hat{AF}} \right|} \right) = 1.0937 \sim \chi_2^2$$

$$\hat{nF}_{YZ \rightarrow X} = \hat{nF}_{N_1 \leftarrow}^* = 108 \times \log \left(\frac{|\hat{D}|}{|\hat{AD}|} \right) = -0.166 \equiv 0.000 \sim \chi_8^2$$

$$\hat{nF}_{X \rightarrow YZ} = \hat{nF}_{N_1 \rightarrow}^* = 108 \times \log \left(\frac{|\hat{C}|}{|\hat{AC}|} \right) = -3.590 \equiv 0.000 \sim \chi_8^2$$

$$\hat{nF}_{X.YZ} = \hat{nF}_{N_1 \cdot}^* = 108 \times \log \left(\frac{|\hat{AD}| \cdot |\hat{AC}|}{|\hat{A}|} \right) = 6.331 \sim \chi_4^2$$

$$\hat{nF}_{XZ \rightarrow Y} = \hat{nF}_{N_2 \leftarrow}^* = 108 \times \log \left(\frac{|\hat{E}|}{|\hat{AE}|} \right) = -12.806 \equiv 0.000 \sim \chi_6^2$$

$$\hat{nF}_{Y \rightarrow XZ} = \hat{nF}_{N_2 \rightarrow}^* = 108 \times \log \left(\frac{|\hat{B}|}{|\hat{AB}|} \right) = -1.156 \equiv 0.000 \sim \chi_6^2$$

$$\hat{nF}_{Y.XZ} = \hat{nF}_{N_2 \cdot}^* = 108 \times \log \left(\frac{|\hat{AB}| \cdot |\hat{AE}|}{|\hat{A}|} \right) = 1.0937 \sim \chi_3^2$$

$$\hat{nF}_{X,Y,Z} = \hat{nF}_P^* = 108 \times \log(|\hat{A}\hat{D}| \cdot |\hat{A}\hat{E}| \cdot |\hat{A}\hat{F}| / |\hat{A}|) = 6.438 \quad \sim \chi_5^2$$

$$\hat{nF}_{X,Y,Z} = \hat{nF}_{P \Leftrightarrow}^* = 108 \times \log(|\hat{D}| \cdot |\hat{E}| \cdot |\hat{F}| / |\hat{A}|) = 22.666 \quad \sim \chi_2^2$$

As required $\hat{nF}_{N_1}^* = \hat{nF}_{(T \rightarrow S)_1}^* + \hat{nF}_{(S_C \rightarrow S_C/\tau)}^*$,

and $\hat{nF}_{N_2}^* = \hat{nF}_{(T \rightarrow S)_2}^* + \hat{nF}_{(S_C \rightarrow S_C/\tau)}^*$.

Unlike our previous examples we did not use WMTS-1 to estimate the multivariate models. The program failed with this data set and so we used the SAS package instead. The PROC STATESPACE Statement allows the user to fit statespace models. Multivariate ARMA models may be expressed as statespace models and then fitted within this framework. For some of our estimates \hat{F} was negative, which in theory should not occur. This appears to be due to rounding error, and assuming $F = 0.00$ would indicate less accuracy in fitting the higher parametric models. Geweke (1982) suggests using OLS or even the Yule-Walker estimates on the grounds that both are equivalent to the M.L. estimates. The Yule-Walker estimates certainly would produce positive F 's despite being otherwise less accurate. Again some of this inconsistency may be due to the occurrence of zero's in the OTH_t series:- The measures relating to only X_t and Z_t remained positive (as can be seen above). Y_t is the only variable likely to be influenced substantially by the zero's in the OTH_t series. This may of course be only coincidence.

Examining the results we see that we have neutrality from both the right and left, that is:-

$$N_1 \quad \underline{\underline{\quad}} \quad \text{and} \quad N_2 \quad \underline{\underline{\quad}}$$

However we also have $(S \Leftrightarrow T)_1$ since $\hat{nF}_{X \Leftrightarrow Z}^* = 36.620$, the

sum of the three required components above. This is χ^2_8 , and highly significant. Also its components are all significant, the least one being $(T \rightarrow S)_1$, which is a component of $N_1 \Leftrightarrow$. We further have that $(S \rightarrow T)_2$. Both $(S \Leftrightarrow T)_1$ and $(S \rightarrow T)_2$ seem to contradict $N_1 \perp\!\!\!\perp$ and $N_2 \perp\!\!\!\perp$. As already mentioned if our intention is to be free to predict either of the subcompositions by omitting some of the variables then $(S \rightarrow T)_2$ does not infringe on this. Neither do $(S \rightarrow T)_1$, or $(S.T)_1$ so that the only contraction occurs with the presence of $(T \rightarrow S)_1$, but as already mentioned this is only just significant. Thus we conclude that both $\underline{U}_t^{(C)} = (\text{CON LAB LIB})$ and $\underline{U}_t(C) = (\text{OTH DTK})$ may be modelled independently of each other without undue loss of information. If we wish to investigate a series of the form

- i) One of three major parties (TMP)
 - ii) Not one of three major parties (NMP)
- i.e. i) = T(CON LAB LIB) ii) = T(OTH DTK)

we would do better to include both $\underline{U}_t^{(C)}$ and $\underline{U}_t(C)$ to improve accuracy. This means that the preference for one of the major political parties over the other is not influenced by the choice between minor parties or undecided voters. A similar converse property holds. The proportion preferring a main line party over another party or those undecided i.e. TMP vs NMP does seem to be influenced by which parties are involved.

Finally we note that we also have $P \perp\!\!\!\perp$ which suggests that all three components do not influence each other. Again this is contradictory, in the fact that both $(S \Leftrightarrow T)_1$ and $P \perp\!\!\!\perp$ hold. This may in part be due to the strange negative values of some of the F's. Thus these are subtracting from the significant parts of $F_{P \Leftrightarrow}$ reducing it so that it is no longer significant. Hence we must assume that with other estimation procedures better

results will occur. The final interpretation depends on the nature of the analysis required.

6.5 : Summary and Conclusion

In this chapter we have examined independence in the time series context and in the compositional context. In bridging the two, various useful time series compositional independence properties have been developed. However, throughout it is clear that many additional problems occur in moving from the independence properties of Aitchison (1982) to those of this chapter. Many of these problems occur because of the directional nature of the relationships that exist between time series. These lagged dependencies make it difficult to express parametrically basis independence and complete subcompositional independence.

Various initial attempts to do so have been made here but clearly further work is needed. Also the partition independence properties do not fit together so neatly as they do in the stationary case. This too needs further development.

Perhaps the most useful exercise is the numerical examples. These give us an insight into the problems in a clear way. They serve to demonstrate which concepts are likely to be more useful and which may possibly need to be re-defined.

The next step is to investigate some of these forms of independence through some simulation runs. This would, for example, aid us in understanding the parametric form of a series which has basis independence.

"I give you learning so do not forsake my teaching."

Proverbs 4:2

CHAPTER 7

Conclusion

7.0 Conclusion

This thesis has sought to develop methods for the analysis of compositional time series. The approach has been to combine the existing theory of time series analysis with that of compositional data analysis. The resulting models have been scrutinized as to their properties and applicability. Although only two models have been considered in detail, the $\ln_+ARMA_m(p,q)$ and the $\ln_xARMA_m(p,q)$, the approach to handling compositional time series is quite general. Thus for $\underline{v}_t \sim ARMA_m(p,q)$; $\underline{v}_t = f(\underline{u}_t) \in \mathbb{S}^m$ we may form other f - $ARMA_m(p,q)$ models. Consequently we have developed not just a specific method, but steps towards an even broader approach. Clearly this is an area for further research.

The two models we have examined provide a practical and straight forward representation of compositional time series. The $\ln_+ARMA_m(p,q)$ model utilizes an already popular transformation, and as was seen in theorem 4.16 gives rise to a symmetrical model on the simplex. This invariance property makes it appealing for many applications. On the other hand the $\ln_xARMA_m(p,q)$ model supplies us with both an alternative to the $\ln_+ARMA_m(p,q)$ model, and with an "ordered" description of the possible structure of \underline{u}_t . Both models may be used for forecasting, and each has particular applications with regards to understanding the interrelationships of the components of \underline{u}_t and the sum of its basis τ_t (where this exists).

There are still many problems to be overcome. For example, although we examined the forecasting problem in some depth, we still have not come up with a hard and fast rule of which point predictor to

use. The approximation to the mean, although interesting, does not seem to provide an easy solution. Its main application might be for approximating the mean of the univariate model. Clearly the whole problem of forecasting compositional time series is not as neat as it is in other areas of time series analysis.

The results of chapter 6 provide an interesting framework in which to understand the relationships between the constituents of the composition. These proved to be more complicated than their contemporaneous counterparts, as can be seen by the discussion of basis independence, complete subcompositional invariance and the directional nature of the other forms of independence. Of these, those given by definitions 6.1 and 6.11-6.14 are probably the most useful, and fortunately are the most straight forward to investigate. However, even these contained contradictions in e.g. the inability to derive a lattice of hypotheses for the partition independence properties.

A further result of this study of compositional time series is to highlight areas of interest in both of the separate areas of time series analysis and analysis of compositional data. In the time series context, is it the case, as possibly suggested by the example in section 4.4, that the univariate model will produce better forecasts? Similarly we can discern a need for further tests to see if several time series are independent. In compositional data analysis can we find an exact rule to determine the number of modes of the $L_m(\underline{\mu}, \underline{\Sigma})$ distribution? How do other distributions on the simplex behave? Is there a way out of the zero value problem? Clearly there is a need for further work in these areas and in the combined area of compositional time series.

In summary we see that we have so far investigated two specific models for compositional time series, pointed the way to a much broader approach, considered some areas of application, and gained an insight into the separate areas of time series analysis and the analysis of compositional data. There are still some loopholes to be filled, but as the old proverb states, "to walk a thousand miles one must take the first step".

7.1 Further Work

We have already seen in the preceding chapters and in the last section some possible areas for further work. We summarize some of the main possibilities below.

1. Any statistical method is best tried, and further developed by application to real data. This is perhaps the greatest need of our work so far.
2. The models may be similarly tested via simulated data. For example, a Monte Carlo study to determine the small sample properties etc. of the tests of independence.
3. Other transformations which map \mathbb{S}^m onto \mathbb{R}^m , and hence other f -ARMA_m(p,q) models.
4. Can the forecasting problem be further developed? For example, by using the Taylor series expansion method to approximate the mean. (See Aitchison and Begg(1976)).
5. The application of the models to repeated surveys.
6. Spectral analysis of compositional time series.
7. Comparison with other statistical models. e.g. the invariance property of the \ln_+ ARMA_m(p,q) model seems to have a connection with the models used for contingency tables. The choice of reference variable is similar to the choice of which margin to constrain.
8. Applications in other areas, e.g. income elasticity.
9. The zero value problem.
10. The contradictions in the formulation of the various independence properties.

11. An exact rule for the uni-modality of the $L_m(\underline{\mu}, \underline{\Sigma})$ and $M_m(\underline{\mu}, \underline{\Sigma})$ distributions.
12. A further model that incorporates both the stronger forms of compositional independence, as represented by the Dirichlet distribution, and the weaker forms as represented by the $L_m(\underline{\mu}, \underline{\Sigma})$ or $M_m(\underline{\mu}, \underline{\Sigma})$. This might be along the lines of the $A(\underline{\mu}, \underline{\Sigma})$ distribution of Aitchison-(1985).
13. The development of other forms of multivariate analysis of compositional time series data, e.g by combining the time series results of Brillinger(1981) with the compositional results of Aitchison(1983) for principal component analysis.
14. A development of alternative approaches to compositional time series. e.g. fitting a model with linear restrictions to the raw data. This is in fact what was done in section 4.4 when the series were modelled by omitting one variable.

APPENDIX A

Lemma A.1

Let a, b_1, \dots, b_n be $n + 1$ constants such that

$$a(b_i + b_j) - b_i b_j = c \quad i \neq j = 1, \dots, n$$

where c is also a constant, then at least n of the $n + 1$ constants $\{a, b_1, \dots, b_n\}$ must be equal.

Proof

We are given

$$a(b_i + b_j) - b_i b_j = c \quad (a)$$

$$a(b_i + b_k) - b_i b_k = c \quad (b)$$

$$a(b_i + b_k) - b_i b_k = c \quad (c)$$

$$(a) - (b) \quad (a - b_i)(b_j - b_k) = 0 \quad (d)$$

$$(a) - (c) \quad (a - b_j)(b_i - b_k) = 0 \quad (e)$$

$$(d) \Rightarrow a - b_i = 0 \quad (f)$$

and/or

$$b_j - b_k = 0 \quad (g)$$

$$(e) \Rightarrow a - b_j = 0 \quad (h)$$

and/or

$$b_i - b_k = 0 \quad (i)$$

$i, j, k = 1, \dots, n$
 $i \neq j \neq k$

There are four possibilities:-

$$(f) \text{ and } (h) \quad a = b_s \quad s = 1, \dots, n$$

$$s \neq k$$

or

$$(f) \text{ and } (i) \quad a = b_s \quad s = 1, \dots, n$$

$$s \neq j$$

or

$$(g) \text{ and } (h) \quad a = b_s \quad s = 1, \dots, n \\ s \neq i$$

or

$$(g) \text{ and } (i) \quad b_r = b_s, \quad r, s = 1, \dots, n$$

thus each possibility has all constants equal bar one. \square

Proof of Lemma 6.5

From corollary 2.12.2, the Yule-Walker equations for an ARMA_m(1,q) process are

$$\underline{\Gamma}(r)\underline{\Phi}' = \underline{\Gamma}(r+1) \quad r > q \quad (A.1)$$

where $\underline{\Phi}$ is the matrix of coefficients for the model, i.e.:-

$$\underline{v}_t + \underline{\Phi}\underline{v}_{t+1} = \underline{\Theta}(B) \underline{\epsilon}_t.$$

$$(A.1) \Rightarrow \underline{\Gamma}(r) \underline{\Phi}'^s = \underline{\Gamma}(r+s)$$

$$\begin{aligned} \underline{\Gamma}(r+2s) &= \underline{\Gamma}(r) \underline{\Phi}'^s \underline{\Phi}'^s \\ &= \underline{\Gamma}(r) [\underline{\Gamma}(r)^{-1} \underline{\Gamma}(r+s)] [\underline{\Gamma}(r)^{-1} \underline{\Gamma}(r+s)] \\ &= \underline{\Gamma}(s+r) \underline{\Gamma}(r)^{-1} \underline{\Gamma}(s+r) \end{aligned} \quad (A.2)$$

$$s > q \quad r > 0$$

Substituting (6.1.17) into (A.2) gives

$$\begin{aligned} \underline{\Gamma}(r+2s) &= (\underline{\Lambda}(s+r) + \gamma_{m+1}^{(s+r)} \underline{U}_m) \underline{\Gamma}(r)^{-1} (\underline{\Lambda}(s+r) + \gamma_{m+1}^{(s+r)} \underline{U}_m) \\ &= \underbrace{\underline{\Lambda}(s+r) \underline{\Gamma}(r)^{-1} \underline{\Lambda}(s+r)}_I + \underbrace{\gamma_{m+1}^{(s+r)} (\underline{U}_m \underline{\Gamma}(r)^{-1} \underline{\Lambda}(s+r))}_{II} + \underbrace{\gamma_{m+1}^{(s+r)} \underline{\Lambda}(s+r) \underline{\Gamma}(r)^{-1} \underline{U}_m}_{III} \\ &\quad + \underbrace{\gamma_{m+1}^{(s+r)2} \underline{U}_m \underline{\Gamma}(r)^{-1} \underline{U}_m}_{IV} \end{aligned} \quad (A.3)$$

Now

$$\Gamma(r)^{-1} = \frac{1}{v(r)} \begin{bmatrix} \left(\frac{1}{\gamma_1(r)} (v^{(r)} - \frac{1}{\gamma_1(r)}) \right) & \frac{-1}{\gamma_1(r)\gamma_2(r)} \cdots & \frac{-1}{\gamma_1(r)\gamma_m(r)} \\ \frac{-1}{\gamma_2(r)\gamma_1(r)} \left(\frac{1}{\gamma_2(r)} (v^{(r)} - \frac{1}{\gamma_2(r)}) \right) \cdots & \frac{-1}{\gamma_2(r)\gamma_m(r)} \\ \vdots & \vdots & \vdots \\ \frac{-1}{\gamma_m(r)\gamma_1(r)} \cdots \frac{-1}{\gamma_m(r)\gamma_2(r)} \cdots & \cdots & \left(\frac{1}{\gamma_m(r)} (v^{(r)} - \frac{1}{\gamma_m(r)}) \right) \end{bmatrix}$$

where $v^{(r)} = \sum_{i=1}^{m+1} \frac{1}{\gamma_i(r)}$ (A.4)

Substituting (A.4) into (A.3) gives

$$(I)_{ij} = \frac{\lambda_i^{(s,r)^2}}{v(r)} (\gamma_i^{(r)} v^{(r)} - 1) \quad i = j$$

$$i, j = 1, \dots, m$$

$$= -\lambda_i^{(s,r)} \lambda_j^{(s,r)} \quad i \neq j$$

and where $\lambda_i^{(s,r)}$ is given by (6.1.18) $i = 1, \dots, m$

$$\left. \begin{array}{l} (II)_{ij} \\ (III)_{ji} \end{array} \right\} = \lambda_{m+1}^{(s,r)} \lambda_j^{(s,r)} \quad i, j = 1, \dots, m$$

and

$$(IV)_{ij} = \frac{\gamma_{m+1}^{(s+r)^2}}{v(r)} \cdot \frac{1}{\gamma_{m+1}(r)} (v^{(r)} - \frac{1}{\gamma_{m+1}(r)}) \quad i, j = 1, \dots, m$$

$$= \frac{C_{m+1}^{(s+r)}}{v(r)} \quad (\text{say}).$$

Combining the four parts of (A.3) together gives

$$\begin{aligned} \gamma_{m+1}^{(r+2s)} &= \left\{ \Gamma(r+2s) \right\}_{ij} \quad i \neq j \\ &= \frac{1}{\nu(s)} \left(C_{m+1}^{(s+r)} + \lambda_{m+1}^{(s,r)} (\lambda_i^{(s,r)} + \lambda_j^{(s,r)}) - \lambda_i^{(s,r)} \lambda_j^{(s,r)} \right) \end{aligned}$$

Since L.H.S. is independent of i and j ,

$$\lambda_{m+1}^{(s,r)} (\lambda_i^{(s,r)} + \lambda_j^{(s,r)}) - \lambda_i^{(s,r)} \lambda_j^{(s,r)} = p_{m+1}^{(s,r)} \quad (\text{A.5})$$

where $p_{m+1}^{(s,r)}$ is a constant independent of i and j

Comparing (A.5) with Lemma A.1 implies

$$\lambda_i^{(s,r)} = \lambda_j^{(s,r)}, \quad i, j = 1, \dots, m+1$$

except for $i, j \neq k$ say

Finally we must show that k is the same for every $s > 0, r \geq q$

$$\text{Let } \frac{\gamma_i^{(a+q)}}{\gamma_i^{(q)}} = \lambda^{(a,q)} \quad (\text{A.6})$$

$$i \neq k_1 \text{ (say)}$$

$$\frac{\gamma_i^{(b+q)}}{\gamma_i^{(q)}} = \lambda^{(b,q)} \quad (\text{A.7})$$

$$i \neq k_2$$

for \forall integer $a > b \geq 0$

Dividing (A.6) by (A.7) gives

$$\frac{\gamma_i^{(a+q)}}{\gamma_i^{(b+q)}} = \lambda^{*(a,b,q)}$$

and $i \neq k_1$
 $i \neq k_2$

$$= \frac{\gamma_i^{((a-b)+(b+q))}}{\gamma_i^{(b+q)}} = \lambda^{(a-b, b+q)} \quad \text{since } a > b$$

$i \neq k_3$

By lemma A.1 there must be only one value of i for which the above ratio may vary and hence

$$k_1 = k_2 = k_3$$

That is k is the same for any values of a, b in $\lambda^{(a-b, b+q)}$ $a > b > 0$ i.e. in any $\lambda^{(r,s)}$ $s > 0, r \geq q$. □

Proof of Theorem 6.6

From lemma 6.4 there are two cases to consider, either the rogue parameter corresponds to the reference variable ($k = m+1$) or to one of the other m variables ($k \in (1, \dots, m)$). We have that (e.g. from (A.1))

$$\begin{aligned} \underline{\Phi}'^s &= \underline{\Gamma}(r)^{-1} \underline{\Gamma}(s+r) \\ &= \underline{\Gamma}(r)^{-1} (\underline{\Lambda}(s+r) + \gamma_{m+1}^{(s+r)} \underline{U}_m) \end{aligned} \quad (\text{A.8})$$

Substituting for $\underline{\Gamma}(r)^{-1}$ from (A.4) we obtain:-

$$\begin{aligned} \underline{\Phi}'^s &= \underline{H} + \underline{G} \quad , \\ \underline{H} &= \underline{\Gamma}(r)^{-1} \underline{\Lambda}(r+s) \quad , \\ \underline{G} &= \underline{\Gamma}(r)^{-1} \gamma_{m+1}^{(r+s)} \underline{U} \quad \text{so that} \end{aligned}$$

$$\{\underline{H}\}_{ij} = \begin{cases} \frac{1}{v^{(r)}} \frac{\gamma_i^{(s+r)}}{\gamma_i^{(s)}} (v^{(r)} - \frac{1}{\gamma_i^{(r)}}) & i = j \\ -\frac{1}{v^{(r)}} \frac{\gamma_j^{(s+r)}}{\gamma_i^{(r)} \gamma_j^{(r)}} & i \neq j \end{cases}, \quad i, j = 1, \dots, m$$

$$\{\underline{G}\}_{ij} = \frac{\gamma_{m+1}^{(s+r)}}{v^{(r)} \gamma_{m+1}^{(r)}} \cdot \frac{1}{\gamma_i^{(r)}} \quad i, j = 1, \dots, m \quad (\text{A.9})$$

Case 1 k = m+1

By lemma 6.4 we may substitute

$$\frac{\gamma_i^{(s+r)}}{\gamma_i^{(r)}} = \lambda^{(s,r)} \quad k = 1, \dots, m \text{ into (A.9)}$$

$$\{\underline{H}\}_{ij} = \begin{cases} \lambda^{(s,r)} (1 - \frac{1}{\gamma_i^{(r)} v^{(r)}}) & i = j \\ \lambda^{(s,r)} \times \frac{-1}{\gamma_i^{(r)} v^{(r)}} & i \neq j \end{cases}, \quad i, j = 1, \dots, m$$

$$\text{i.e. } \underline{H} = \lambda^{(s,r)} \underline{I}_m - \left(\frac{\lambda^{(s,r)}}{v^{(r)}} \right) \text{dg} \left(\frac{1}{\gamma_1^{(r)}}, \dots, \frac{1}{\gamma_m^{(r)}} \right) \underline{V}_m$$

similarly

$$\underline{G} = \frac{\gamma_{m+1}^{(s+r)}}{v^{(r)} \gamma_{m+1}^{(r)}} \text{dg} \left(\frac{1}{\gamma_1^{(s)}}, \dots, \frac{1}{\gamma_m^{(r)}} \right) \underline{V}_m$$

So that from (A.9) and noting that $\text{dg}\left(\frac{1}{\gamma_1(r)}, \dots, \frac{1}{\gamma_m(r)}\right) = \underline{\Lambda}^{-1}(r)$, we obtain

$$\underline{\phi}'^s = \lambda^{(s,r)} \underline{I}_m + \alpha^{(s,r)} \underline{\Lambda}^{-1}(r) \underline{U}_m$$

$$\text{where } \alpha^{(s,r)} = \frac{1}{\nu(r)} \left[\frac{\gamma_{m+1}^{(s+r)}}{\gamma_{m+1}(r)} - \lambda^{(s,r)} \right]$$

Thus for $s = 1$ and $r = q$

$$\underline{\phi}' = \phi \underline{I}_m + \beta \underline{\Lambda}(q)^{-1} \underline{U}_m,$$

$$\text{where } \lambda^{(1,q)} = \phi \text{ and } \alpha^{(1,q)} = \beta. \quad (\text{A.10})$$

For $k = m+1$ this is as required by the theorem in (6.1.18) and (6.1.19).

Next we need to demonstrate (6.1.20) for $k = m+1$. From (A.10) we have

$$\begin{aligned} \underline{\phi}'^s &= (\phi \underline{I}_m + \beta \underline{\Lambda}(q)^{-1} \underline{U}_m)^s \\ &= \sum_{i=0}^s \binom{s}{i} \phi^{s-i} (\beta \underline{\Lambda}(q)^{-1} \underline{U}_m)^i \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} \text{Now } (\underline{\Lambda}(q)^{-1} \underline{U}_m)^i &= (\underline{\Lambda}(q)^{-1} \underline{U}_m)^{i-2} \underline{\Lambda}(q)^{-1} \cdot \underline{U}_m \underline{\Lambda}(q)^{-1} \underline{U}_m \\ &= (\underline{\Lambda}(q)^{-1} \underline{U}_m)^{i-2} \underline{\Lambda}(q)^{-1} d_{m+1}^{(q)} \underline{U}_m \\ &= d_{m+1}^{(q)} (\underline{\Lambda}(q)^{-1} \underline{U}_m)^{i-1}, \quad i = 1, 2, \dots \end{aligned} \quad (\text{A.12})$$

$$\text{since } \left\{ \underline{U}_m \underline{L} \underline{U}_m \right\}_{ij} = \sum_{u,v=1}^m \left\{ \underline{L} \right\}_{u,v}, \quad i, j=1, \dots, m \quad (\text{A.13})$$

$$\text{and } \sum_{u,v=1}^m \left\{ \underline{\Lambda}(q)^{-1} \right\}_{u,v} = \sum_{i=1}^m \frac{1}{\gamma_i(q)} = d_{m+1}^{(q)} \quad (\text{A.14})$$

So from (A.12) we have

$$(\underline{\Lambda}(q)^{-1} \underline{U}_m)^i = (d_{m+1}^{(q)})^{i-1} \underline{\Lambda}(q)^{-1} \underline{U}_m \quad (\text{A.15})$$

and substituting this into (A.11), and recalling that:-

$$\begin{aligned} \underline{\Gamma}(q+s) &= \underline{\Gamma}(q) \underline{\phi}'^s \\ \underline{\Gamma}(q+s) &= \phi^s \underline{\Gamma}(q) + \left(\sum_{i=1}^s \binom{s}{i} \phi^{s-i} \beta^i d_{m+1}^{(q)i-1} \right) \underline{\Gamma}(q) \underline{\Lambda}(q)^{-1} \underline{U}_m \end{aligned} \quad (\text{A.16})$$

$$\begin{aligned} \text{Now } \underline{\Gamma}(q) \underline{\Lambda}(q)^{-1} \underline{U}_m &= (\underline{\Lambda}(q) + \gamma_{m+1}^{(q)} \underline{U}_m) \underline{\Lambda}(q)^{-1} \underline{U}_m \\ &= \underline{U}_m + \gamma_{m+1}^{(q)} d_{m+1}^{(q)} \underline{U}_m \quad \text{by (A.13) and (A.14)} \end{aligned}$$

∴ (A.16) becomes

$$\underline{\Gamma}(q+s) = \phi^s \underline{\Gamma}(q) + \left(\sum_{i=1}^s \binom{s}{i} \phi^{s-i} \beta^i d_{m+1}^{(q)i-1} \right) (1 + \gamma_{m+1}^{(q)} d_{m+1}^{(q)}) \underline{U}_m \quad (\text{A.17})$$

which is (6.1.20) for $r = q$, clearly since the diagonal $\underline{\Lambda}(q)$ is only modified by the first term of (A.17)

$$\underline{\Lambda}(r) = \phi^{r-q} \underline{\Lambda}(q) \quad r = q,$$

$$\text{so that we may write } \underline{\Lambda}(q)^{-1} = \phi^{r-q} \underline{\Lambda}(r)^{-1} \quad (\text{A.18})$$

substituting (A.18) into (A.16) and recalling that $\underline{\Gamma}(r+s) = \underline{\Gamma}(r) \underline{\phi}'^s$ (e.g. from (A.8)) we obtain (6.1.20) for all $r = q, q+1, \dots$

Case 2 $k \neq m+1$

Without loss of generality let $k = 1$.

$$\text{Then } \frac{\gamma_i^{(s+r)}}{\gamma_i(r)} = \lambda^{(s,r)}, \quad k=2,3, \dots, m+1.$$

$$\text{Let } \lambda_1^{(s,r)} = \frac{\gamma_1^{(s+r)}}{\gamma_1(r)}$$

$$\{\underline{H}\}_{ij} = \begin{cases} \frac{1}{v(r)} \lambda_1^{(s,r)} (v(r) - \frac{1}{\gamma_1(r)}) & i = j = 1 \\ \frac{1}{v(r)} \lambda^{(s,r)} (v(r) - \frac{1}{\gamma_i(r)}) & i = j \neq 1 \\ - \frac{1}{v(r)} \frac{\lambda_1^{(s,r)}}{\gamma_i} & j = 1, \\ & i = 2, \dots, m \\ - \frac{1}{v(r)} \frac{\lambda^{(s,r)}}{\gamma_i} & j = 2, \dots, m \\ & i = 1, \dots, m \\ & i \neq j \end{cases}$$

and

$$\{\underline{G}\}_{ij} = \frac{1}{v(r)} \frac{\lambda^{(s,r)}}{\gamma_i} \quad i, j = 1, 2, \dots, m.$$

$$\text{Hence } \underline{\Phi}'^s = \underline{H} + \underline{G}$$

$$= \lambda^{(s,r)} \underline{I}_m + \frac{1}{v(r)} \begin{bmatrix} \left(\lambda_1^{(s,r)} (v(r) - \frac{1}{\gamma_1(r)}) + \frac{\lambda^{(s,r)}}{\gamma_1} - \lambda^{(s,r)} v(r) \right) & 0 \dots 0 \\ - \frac{\lambda_1^{(s,r)}}{\gamma_2} + \frac{\lambda^{(s,r)}}{\gamma_2} & 0 \dots 0 \\ \vdots & \\ - \frac{\lambda_1^{(s,r)}}{\gamma_2} + \frac{\lambda^{(s,r)}}{\gamma_2} & 0 \dots 0 \end{bmatrix}$$

$$= \lambda^{(s,r)} \underline{I}_m + \frac{\lambda_1^{(s,r)} - \lambda^{(s,r)}}{\nu(s)} \begin{bmatrix} \nu^{(r)} - \frac{1}{\gamma_1^{(r)}} & 0 & \dots & 0 \\ -\frac{1}{\gamma_2^{(r)}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{\gamma_m^{(r)}} & 0 & \dots & 0 \end{bmatrix} \quad (\text{A.19})$$

Putting $\phi = \lambda^{(s,r)}$, $\beta = \frac{\lambda_1^{(s,r)} - \lambda^{(s,r)}}{\nu(s)}$, and noting that

$$\nu^{(r)} - \frac{1}{\gamma_1^{(r)}} = \sum_{i=1}^{m+1} \frac{1}{\gamma_i^{(r)}} - \frac{1}{\gamma_1^{(r)}} = \sum_{\substack{i=1 \\ i \neq 1}}^{m+1} \frac{1}{\gamma_i^{(r)}} = d_1^{(r)},$$

and for $r = q$; $s = 1$ we have

$$\underline{\phi}' = \phi \underline{I}_m + \beta \underline{A}(1) \quad \text{as in (6.1.18)}$$

Finally we need to show that (6.1.20) is correct. Clearly we have that

$$\underline{\phi}'^s = \phi^s \underline{I}_m + \sum_{i=1}^s \binom{s}{i} \phi^{s-i} \beta^i \begin{bmatrix} \underline{r}^{(q)} & \underline{0} \end{bmatrix}^i,$$

$$\text{where } \underline{r}^{(s)} = (d_1^{(q)}, -\frac{1}{\gamma_2^{(s)}}, \dots, -\frac{1}{\gamma_m^{(s)}})' \quad s=q, q+1 \dots \quad (\text{A.20})$$

$$\text{Now } \begin{bmatrix} \underline{r}^{(q)} & \underline{0} \end{bmatrix}^i = d_1^{(q)i-1} \begin{bmatrix} r_1^{(q)} & \underline{0} \end{bmatrix} \quad (\text{A.21})$$

Substituting this into (A.20) and then into (A.8), for $r=q$

we have

$$\Gamma(s+q) = \Gamma(q) \left[\phi^s \underline{I}_m + \sum_{i=1}^s \binom{s}{i} \phi^{s-i} \beta^i d_1^{(q)i-1} \begin{bmatrix} \underline{r}^{(q)} \\ \underline{0} \end{bmatrix} \right] \quad (\text{A.22})$$

$$\text{Now } \Gamma(q) \begin{bmatrix} \underline{r}^{(q)} \\ \underline{0} \end{bmatrix} = \begin{bmatrix} \gamma_1^{(q)} + \gamma_{m+1}^{(q)} & \gamma_{m+1}^{(q)} & \dots & \gamma_{m+1}^{(q)} \\ \gamma_{m+1}^{(q)} & \gamma_2^{(q)} + \gamma_{m+1}^{(q)} & \dots & \gamma_{m+1}^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m+1}^{(q)} & \gamma_{m+1}^{(q)} & \dots & \gamma_m^{(q)} + \gamma_{m+1}^{(q)} \end{bmatrix} \begin{bmatrix} d_1^{(q)} & 0 & \dots & 0 \\ -\frac{1}{\gamma_2^{(q)}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{\gamma_m^{(q)}} & 0 & \dots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} d_1^{(q)} (\gamma_1^{(q)} + \gamma_{m+1}^{(q)}) - \gamma_{m+1}^{(q)} \sum_{i=2}^m \frac{1}{\gamma_i^{(q)}} & 0 & \dots & 0 \\ d_1^{(q)} (\gamma_{m+1}^{(q)}) - 1 - \gamma_{m+1}^{(q)} \sum_{i=2}^m \frac{1}{\gamma_i^{(q)}} & & & \\ \vdots & & & \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_1^{(q)} d_1^{(q)} + \gamma_{m+1}^{(q)} \left(\sum_{i=2}^{m+1} \frac{1}{\gamma_i^{(q)}} - \sum_{i=2}^m \frac{1}{\gamma_i^{(q)}} \right) & 0 & \dots & 0 \\ \gamma_{m+1}^{(q)} \left(\sum_{i=2}^{m+1} \frac{1}{\gamma_i^{(q)}} - \sum_{i=2}^m \frac{1}{\gamma_i^{(q)}} \right) - 1 & 0 & \dots & 0 \\ \vdots & & & \vdots \end{bmatrix}$$

$$= \begin{bmatrix} 1 + \gamma_1^{(q)} d_1^{(q)} & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} \quad (\text{A.23})$$

substituting into (A.22)

$$\Gamma(s+q) = \phi^s \Gamma(q) + \left(\sum_{i=1}^s \binom{s}{i} \phi^{s-i} \beta^i d_1^{(q)s-i} \right) (1+\gamma_1^{(q)} d_1) \underline{B}(1) \quad (\text{A.24})$$

as in (6.1.20) for $r = q$.

$$\text{Finally since } \underline{r}^{(q)} = \left\{ \sum_{i=2}^{m+1} \frac{1}{\gamma_i^{(q)}}, -\frac{1}{\gamma_2^{(q)}}, \dots, -\frac{1}{\gamma_m^{(q)}} \right\},$$

i.e. \underline{r} consists of terms in $\frac{1}{\gamma_i}$, $i=2, \dots, m+1$.

However all $\gamma_i^{(r)}$ $r \geq q$ are generated by only the first term in (A.24) so that $\gamma_i^{(r)} = \phi^{r-q} \gamma_i^{(q)}$ $i = 2, \dots, m+1$.

Hence in general

$$\begin{aligned} \Gamma(r) \left[\underline{r}^{(q)} \quad \underline{0} \right] &= \Gamma(r) \frac{\phi^{r-q}}{\phi^{r-q}} \left[\underline{r}^{(q)} \quad \underline{0} \right] \\ &= \Gamma(r) \phi^{r-q} \left[\underline{r}^{(r)} \quad \underline{0} \right] \\ &= \phi^{r-q} \begin{bmatrix} 1+\gamma_1^{(r)} d_1^{(r)} & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} \end{aligned}$$

So that (A.23) and hence (A.24) generalizes to (6.1.20) for all $r = q, q+1, \dots$.

□

Alternative proof $k \neq m+1$

The second case may also be derived by using the \underline{Z} matrix. For example if $k = m+1$, then

$$\underline{\Phi}' = \phi \underline{I}_m + \beta \underline{\Lambda}^{(q)-1} \underline{U}$$

$$\text{and } \underline{z}(1) = \begin{bmatrix} -1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ -1 & 0 & \dots & 1 \end{bmatrix}$$

Permuting the system so that variable 1 is now the reference variable gives

$$\begin{aligned} \underline{\phi}^{*'} &= \underline{z}'(1) \underline{\phi}' \underline{z}(1) = \phi \underline{I}_m + \beta \underline{z}'(1) \underline{\Lambda}(q)^{-1} \underline{U} \underline{z}(1) \\ &= \phi \underline{I}_m + \beta \begin{bmatrix} \sum_{i=1}^m \frac{1}{\gamma_i(q)} & 0 & \dots & 0 \\ -\frac{1}{\gamma_2(q)} & 0 & \dots & 0 \\ \vdots & & & \\ -\frac{1}{\gamma_m(q)} & 0 & \dots & 0 \end{bmatrix} \end{aligned}$$

which has re-ordered the $\underline{u} \in \mathcal{S}^m$ vector on which this theory is based to (u_{m+1}, u_2, u_m, u_1) , thus $\underline{\phi}^{*'}$ is the form of $\underline{\phi}'$ when the first variable in the vector corresponds to the rogue parameter.

□

Lemma A.2

For an ARMA_m(2,q) process, such that the a.c.f. $\underline{\Gamma}(r)$ is of the form:-

$$\begin{aligned} \underline{\Gamma}(r) &= \text{dg}\{\gamma_1^{(r)}, \gamma_2^{(r)}, \dots, \gamma_m^{(r)}\} + \gamma_{m+1}^{(r)} \underline{U}_m \\ &= \underline{\Lambda}(r) + \gamma_{m+1}^{(r)} \underline{U}_m \quad \text{as in (6.1.5),} \end{aligned}$$

the $\underline{\Gamma}(r)$ matrices for $r = q+2, q+3, \dots$ may be recursively generated from $\underline{\Gamma}(q), \underline{\Gamma}(q+1), \underline{\phi}_1$ and $\underline{\phi}_2$, if $\underline{\phi}_1$ and $\underline{\phi}_2$ are of the form:-

$$\underline{\Phi}'_1 = \phi_1 \underline{I}_m + \underline{D}_1 \underline{U}_m \quad (\text{A.25})$$

$$\underline{\Phi}'_2 = \phi_2 \underline{I}_m + \underline{D}_2 \underline{U}_m$$

and \underline{D}_1 and \underline{D}_2 are the solutions of

$$\begin{bmatrix} \underline{D}_1 \\ \underline{D}_2 \end{bmatrix} = \begin{bmatrix} \underline{\Lambda}(q+1) & \underline{\Lambda}(q) \\ \underline{\Lambda}(q+2) & \underline{\Lambda}(q+1) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \underline{I}_m \\ \alpha_2 \underline{I}_m \end{bmatrix}, \quad \text{for some constants} \quad (\text{A.26})$$

α_1 and α_2 ,

so that \underline{D}_1 and \underline{D}_2 are diagonal dg $(d_1^{(i)}, \dots, d_m^{(i)})$ $i=1, 2$ (say).

Let $\delta_i = \sum_{j=1}^m d_j^{(i)}$ $i=1, 2$, then the $\underline{\Gamma}(r)$ matrices are

generated through the following equations:-

$$\underline{\Lambda}(r+2) = \phi_1 \underline{\Lambda}(r+1) + \phi_2 \underline{\Lambda}(r)$$

$$\text{i.e. } \gamma_i^{(r+2)} = \phi_1 \gamma_i^{(r+1)} + \phi_2 \gamma_i^{(r)} \quad i=1, \dots, m \quad r > q$$

$$\text{and } \gamma_{m+1}^{(r+2)} = (\phi_1 + \delta_1) \gamma_{m+1}^{(r+1)} + (\phi_2 + \delta_2) \gamma_{m+1}^{(r)} + \alpha_{r-q+1}$$

(A.27)

with α_{r-q+1} $r > q + 2$ generated by:-

$$\underline{\Lambda}(r+1) \underline{D}_1 + \underline{\Lambda}(r) \underline{D}_2 = \alpha_{r-q+1} \underline{I}_m \quad (\text{A.28})$$

and α_1, α_2 are as chosen in (A.26).

Proof

$$\begin{aligned} \text{Let } \underline{\phi}'_1 &= \phi_1 \underline{I}_m + \underline{A}_1 \\ \underline{\phi}'_2 &= \phi_2 \underline{I}_m + \underline{A}_2 \end{aligned} \tag{A.29}$$

where \underline{A}_1 and \underline{A}_2 are chosen so that they have no effect on $\gamma_i^{(r)}$ $i=1, \dots, m$ $r > q + 1$ in the recursive generation of $\underline{\Gamma}(r)$, but only influence the $\gamma_{m+1}^{(r)}$, $r = q+2, q+3, \dots$ series. If \underline{A}_1 and \underline{A}_2 are like this, then (A.27) automatically follows for $i \neq m+1$.

From the Yule-Walker equations (corollary 2.12.2)

$$\begin{aligned} \underline{\Gamma}(r+2) &= \underline{\Gamma}(r+1) \underline{\phi}'_1 + \underline{\Gamma}(r) \underline{\phi}'_2 \quad r = q, q+1, \dots \\ &= [\underline{\Lambda}(r+1) + \gamma_{m+1}^{(r+1)} \underline{U}_m] [\phi_1 \underline{I}_m + \underline{A}_1] + [\underline{\Lambda}(r) + \gamma_{m+1}^{(r)} \underline{U}_m] [\phi_2 \underline{I}_m + \underline{A}_2] \\ &= \phi_1 \underline{\Lambda}(r+1) + \phi_2 \underline{\Lambda}(r) + (\gamma_{m+1}^{(r+1)} \phi_1 + \gamma_{m+1}^{(r)} \phi_2) \underline{U}_m \\ &\quad + \underline{\Lambda}(r+1) \underline{A}_1 + \underline{\Lambda}(r) \underline{A}_2 + \gamma_{m+1}^{(r+1)} \underline{U} \underline{A}_1 + \gamma_{m+1}^{(r)} \underline{U} \underline{A}_2 \end{aligned} \tag{A.30}$$

We require that the last four terms of (A.30) be of the form $c \underline{U}_m$, where c is a constant

$$\text{If we let } \underline{A}_1 = \underline{D}_1 \underline{U}_m$$

$$\underline{A}_2 = \underline{D}_2 \underline{U}_m \text{ we obtain one such solution.}$$

The last four terms become

$$(\underline{\Lambda}(r+1) \underline{D}_1 + \underline{\Lambda}(r) \underline{D}_2) \underline{U}_m + (\gamma_{m+1}^{(r+1)} \delta_1^* + \gamma_{m+1}^{(r)} \delta_2^*) \underline{U}_m \tag{A.31}$$

$$\text{where } \delta_s^* = \sum_{ij} \{ \underline{D}_s \}_{ij} \quad s = 1, 2$$

The second term is now of the required form, we require that

$$(\underline{\Lambda}(r+1) \underline{D}_1 + \underline{\Lambda}(r) \underline{D}_2) \underline{U}_m = c^+ \underline{U}_m$$

so that $\underline{\Lambda}(r+1) \underline{D}_1 + \underline{\Lambda}(r) \underline{D}_2 = c_{r-q+1} \underline{I}_m \quad r = q, q+1, \dots$

which is (A.28)

To solve for $\underline{D}_1, \underline{D}_2$ we need at least the first two equations i.e. for $r = q$ and $r = q+1$

$$\begin{bmatrix} \underline{\Lambda}(q+1) & \underline{\Lambda}(q) \\ \underline{\Lambda}(q+2) & \underline{\Lambda}(q+1) \end{bmatrix} \begin{bmatrix} \underline{D}_1 \\ \underline{D}_2 \end{bmatrix} = \begin{bmatrix} c_1 & \underline{I}_m \\ c_2 & \underline{I}_m \end{bmatrix} \text{ which gives (A.26)}$$

Hence δ_s^* of (A.31) is the same as δ_s of (A.27) $s = 1, 2$

Substituting for \underline{A}_1 and \underline{A}_2 into (A.29) gives (A.25) and finally substituting for $\underline{A}_1, \underline{A}_2$ and (A.28) and (A.31) into (A.30) we obtain the recursive equation for $\gamma_{m+1}^{(r)}$ $r = q+2, \dots$ in (A.27). □

Corrollary A.3

The exact solution of \underline{D}_1 and \underline{D}_2 is as follows

$$d_i^{(1)} = \frac{1}{\eta_i} (\alpha_1 \gamma_i^{(q+1)} - \alpha_2 \gamma_i^{(q)})$$

$$d_i^{(2)} = \frac{1}{\eta_i} (-\alpha_1 \gamma_i^{(q+2)} + \alpha_2 \gamma_i^{(q+1)})$$

$$= \frac{1}{\eta_i} ((\alpha_2 - \alpha_1 \phi_1) \gamma_i^{(q+1)} - \alpha_1 \phi_2 \gamma_i^{(q)}) \quad , \quad i=1, \dots, m, (m+1) \\ i \neq k$$

where $\underline{D}_j = \text{dg}(d_1^{(j)}, \dots, d_m^{(j)}) \quad j = 1, 2$

$$\begin{aligned} \text{and } \eta_i &= \gamma_i^{(q+1)2} - \gamma_i^{(q)} \gamma_i^{(q+2)} \\ &= \gamma_i^{(q+1)2} - \phi_1 \gamma_i^{(q+1)} \gamma_i^{(q)} - \phi_2 \gamma_i^{(q)2} \end{aligned}$$

Proof

$$\text{From (A.26)} \quad \begin{bmatrix} \underline{D}_1 \\ \underline{D}_2 \end{bmatrix} = \begin{bmatrix} \underline{\Lambda}(q+1) & \underline{\Lambda}(q) \\ \underline{\Lambda}(q+2) & \underline{\Lambda}(q+1) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \underline{I}_m \\ \alpha_2 \underline{I}_m \end{bmatrix}$$

which since $\underline{\Lambda}(r)$ $r = 0, \pm 1, \dots$ is diagonal yields.

$$\begin{bmatrix} \underline{D}_1 \\ \underline{D}_2 \end{bmatrix} = (\underline{\Lambda}(q+1)^2 - \underline{\Lambda}(q)\underline{\Lambda}(q+1))^{-1} \begin{bmatrix} \underline{\Lambda}(q+1) & -\underline{\Lambda}(q) \\ -\underline{\Lambda}(q+2) & \underline{\Lambda}(q+1) \end{bmatrix} \begin{bmatrix} \alpha_1 \underline{I}_m \\ \alpha_2 \underline{I}_m \end{bmatrix}$$

examining this term by term gives the required results, in terms of $\gamma_i^{(q)}$, $\gamma_i^{(q+1)}$ and $\gamma_i^{(q+2)}$. Recalling that

$$\gamma_i^{(q+2)} = \phi_1 \gamma_i^{(q+1)} + \phi_2 \gamma_i^{(q)} \quad \text{gives the result in terms of } \gamma_i^{(q)} \text{ and } \gamma_i^{(q+1)}.$$

□

Corollary A.4

For $k \neq m+1$, that is if we generate $\underline{\Gamma}(r)$ $r > q+1$, such that another variable is the rogue variable, then $\underline{\phi}_1$ and $\underline{\phi}_2$ are given by:-

$$\left. \begin{aligned} \underline{\phi}_1' &= \phi_1 \underline{I}_m + \underline{A}^{(1)}(k) \\ \underline{\phi}_2' &= \phi_2 \underline{I}_m + \underline{A}^{(2)}(k) \end{aligned} \right\} \text{for } k \in \{1, \dots, m\} \quad (\text{A.32})$$

where

$$\left\{ \begin{array}{l} \underline{A}_i^{(\ell)} \\ \underline{A}_k^{(\ell)} \end{array} \right\}_{ij} = \left. \begin{array}{ll} -d_j^{(\ell)} & i = k \quad j \neq k \\ \delta_k^{(\ell)} & i = j = k \\ 0 & \text{otherwise} \end{array} \right\} i, j = 1, \dots, m; \ell = 1, 2 \quad (\text{A.33})$$

and $d_i^{(\ell)}$ is given by corollary (A.3) and $\delta_k^{(\ell)} = \sum_{\substack{i=1 \\ j \neq k}}^{m+1} d_j^{(i)}$

For these parameters the $\underline{\Gamma}(r)$ matrices are now generated through:-

$$\begin{aligned} \gamma_i^{(r+2)} &= \phi_1 \gamma_i^{(r+1)} + \phi_2 \gamma_i^{(r)} \\ \gamma_k^{(r+2)} &= (\phi_1 + \delta_1) \gamma_k^{(r+1)} + (\phi_2 + \delta_2) \gamma_k^{(r)} \quad r = q, q+1, \dots \\ & \quad i = 1, \dots, m+1 \\ & \quad i \neq k \quad (\text{A.34}) \end{aligned}$$

Proof

For $k \neq m+1$ $\underline{\phi}'_i = \underline{Z}'_i(k) \underline{\phi}^{*'}_i \underline{Z}_i(k)$ $i = 1, 2$ and

$\underline{\phi}^{*'}_i$ is $\underline{\phi}$ as given in Lemma (A.2), (A.25).

These give the result directly, as can be seen by noting that $\underline{\phi}'$ in theorem 6.6 for $k = m+1$ was also of the same form. Thus we may apply the alternative proof $k \neq m+1$ given above, and by similar argument obtain (A.34). □

Theorem A.5

For an ARMA_m(p,q) process, such that

$$\underline{\Gamma}(r) = \underline{\Lambda}(r) + \gamma_{m+1}^{(r)} \underline{U}_m \quad r = 0, \pm 1, \dots$$

the $\underline{\Gamma}(r)$ matrices for $r = q+p+1, q+p+2, \dots$ may be recursively generated from $\underline{\Gamma}(q), \underline{\Gamma}(q+1), \dots, \underline{\Gamma}(q+p)$ and $\underline{\Phi}_1, \dots, \underline{\Phi}_p$ if $\underline{\Phi}_i$ ($i=1, \dots, p$) are of the form

$$\underline{\Phi}_i = \phi_i \underline{I}_m + \underline{D}_i \underline{U} \quad i = 1, \dots, p$$

where ϕ_i ($i=1, \dots, p$) are constants, and \underline{D}_i ($i=1, \dots, p$) are solutions of

$$\begin{bmatrix} \underline{D}_1 \\ \underline{D}_2 \\ \vdots \\ \underline{D}_p \end{bmatrix} = \begin{bmatrix} \underline{\Lambda}(q+p-1) & \underline{\Lambda}(q+p-2) & \dots & \underline{\Lambda}(q) \\ \underline{\Lambda}(q+p) & \underline{\Lambda}(q+p-1) & \dots & \underline{\Lambda}(q+1) \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\Lambda}(q+2p-2) & \underline{\Lambda}(q+2p-3) & \dots & \underline{\Lambda}(q+p-1) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \underline{I}_m \\ \alpha_2 \underline{I}_m \\ \vdots \\ \alpha_p \underline{I}_m \end{bmatrix} \quad (\text{A.35})$$

after which α_{r-q+1} $r \geq q+p$ is generated by

$$\underline{\Lambda}(r+p-1) \underline{D}_1 + \underline{\Lambda}(r+p-2) \underline{D}_2 + \dots + \underline{\Lambda}(r) \underline{D}_p = \alpha_{r-q+1} \underline{I}_m \quad (\text{A.36})$$

The matrices \underline{D}_i will be shown to be diagonal. Let

$$\underline{D}_i = \text{dg}(d_1^{(i)}, d_2^{(i)}, d_2^{(i)}, \dots, d_m^{(i)}) \quad i=1, \dots, p,$$

$$\delta_{m+1}^{(i)} = \sum_{j=1}^m d_j^{(i)} \quad i = 1, \dots, p$$

For this solution to the autoregressive parameters the $\underline{\Gamma}(r)$ will be generated via

$$\gamma_i^{(r+p)} = \phi_1 \gamma_i^{(r+p-1)} + \dots + \phi_p \gamma_i^{(r)} \quad (\text{A.37})$$

$$\gamma_{m+1}^{(r+p)} = (\phi_1 + \delta_{m+1}^{(1)}) \gamma_{m+1}^{(r+p-1)} + \dots + (\phi_p + \delta_{m+1}^{(p)}) \gamma_{m+1}^{(r)} + \alpha_{r-q+1}$$

$$r = q, q+1, \dots$$

so that the rogue variable for this solution is simply the $m+1^{\text{th}}$ i.e. $k = m+1$.

Proof

The result follows from the direct extension of Lemma A.2.

Corollary A.6

For $k \neq m+1$ the parameter values for ϕ_1, \dots, ϕ_p are given by:-

$$\phi_i = \phi_i I_m + \frac{A(i)}{(k)} \quad \begin{matrix} i = 1, \dots, p \\ k \in \{1, \dots, m\} \end{matrix}$$

where

$$\left\{ \frac{A(i)}{(k)} \right\}_{j\ell} = \left. \begin{matrix} -d_\ell^{(i)} & j = k & \ell \neq k \\ \delta_k^{(i)} & j = \ell = k \\ 0 & \text{otherwise} \end{matrix} \right\} \begin{matrix} j, \ell = 1, \dots, m; \\ i = 1, \dots, p \end{matrix}$$

where $\delta_k^{(i)} = \sum_{\substack{j=1 \\ j \neq k}}^{m+1} d_j^{(i)}$, and $d_1^{(i)}, d_2^{(i)}, \dots, d_{m+1}^{(i)}$

$i = 1, \dots, p$ may be found by solving

$$\begin{bmatrix} d_\ell^{(1)} \\ d_\ell^{(2)} \\ \vdots \\ d_\ell^{(p)} \end{bmatrix} = \begin{bmatrix} \gamma_\ell^{(q+p-1)} & \gamma_\ell^{(q+p-2)} & \dots & \gamma_\ell^{(q)} \\ \gamma_\ell^{(q+p)} & \gamma_\ell^{(q+p-1)} & \dots & \gamma_\ell^{(q+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_\ell^{(q+2p-2)} & \gamma_\ell^{(q+2p-3)} & \dots & \gamma_\ell^{(q+p-1)} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \quad \begin{matrix} \ell = 1, \dots, m+1. \\ \ell \neq k \end{matrix} \quad (A.38)$$

and the γ 's are generated by

$$\gamma_i^{(r+p)} = \phi_i \gamma_i^{(r+p-1)} + \dots + \phi_p \gamma_i^{(r)} \quad \begin{matrix} i = 1, \dots, m+1 \\ i \neq k \end{matrix}$$

$$\gamma_k^{(r+p)} = (\phi_{1+\delta_k^{(p)}}) \gamma_k^{(r+p-1)} + \dots + (\phi_{p+\delta_k^{(p)}}) \gamma_k^{(r)} + \alpha_{r-q+1} \quad (\text{A.39})$$

$$r = q, q+1, \dots$$

$\alpha_1, \dots, \alpha_p$ are constants and for $r > q + p, \alpha_{r-q+1}$ are generated by:-

$$\Lambda_k^{(r+p-1)} \underline{D}_{k,1} + \dots + \Lambda_k^{(r)} \underline{D}_{k,p} = \alpha_{r-q+1} \underline{I}_m \quad (\text{A.40})$$

$$\Lambda_k^{(s)} = \text{dg}(\gamma_i^{(s)}; i=1, \dots, m+1 \quad i \neq k) \quad s = 0, \pm 1, \dots;$$

$$\underline{D}_{k,s} = \text{dg}(d_i^{(s)}; i = 1, \dots, m+1 \quad i \neq k)$$

Proof

The result follows exactly before by use of $\underline{Z}(k)$ matrix, e.g. $k = 1$

$$\underline{Z}'_{(1)} = \begin{bmatrix} -1 & -1 & \dots & -1 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

We have that $\underline{\Gamma}^*(r) = \text{dg}(\gamma_1^{(r)}, \dots, \gamma_m^{(r)}) + \gamma_{m+1}^{(r)} \underline{U}_m \quad r = 0, \pm 1, \dots$

$$\text{Let } \underline{\Gamma}^*(r) = \underline{Z}_{(1)} \underline{\Gamma}(r) \underline{Z}'_{(1)} \quad (\text{A.41})$$

$$= \text{dg}(\gamma_{m+1}^{(r)}, \gamma_2^{(r)}, \dots, \gamma_m^{(r)}) + \gamma_1^{(r)} \underline{U}_m \quad r = 0, \pm 1, \dots$$

If the rogue k is 1, then it corresponds to the term multiplying the \underline{U}_m matrix. This implies that the parameters ϕ_i^* $i = 1, \dots, p$ of the corresponding ARMA $_m(p,q)$ are identical to that given in theorem A.5, but with the indices on the $\underline{\Gamma}$'s and derived parameter interchanged between $m+1$ and 1. Thus we replace $\underline{\Lambda}(r)$ in (A.35) by $\underline{\Lambda}_1^*(r)$:-

$$\underline{\Lambda}(r) = \text{dg}(\gamma_1^{(r)}, \dots, \gamma_m^{(r)})$$

$$\underline{\Lambda}_1^*(r) = \text{dg}(\gamma_{m+1}^{(r)}, \gamma_2^{(r)}, \dots, \gamma_m^{(r)})$$

or re-ordering $\underline{\Lambda}_1(r) = \text{dg}(\gamma_2^{(1)}, \dots, \gamma_{m+1}^{(r)})$

Hence (A.35) becomes

$$\begin{bmatrix} \underline{D}_{1,1} \\ \underline{D}_{1,2} \\ \underline{D}_{1,p} \end{bmatrix} = \begin{bmatrix} \underline{\Lambda}_1(q+p-1) & \dots & \underline{\Lambda}_1(q) \\ \vdots & \ddots & \vdots \\ \underline{\Lambda}(q+2p-2) & \dots & \underline{\Lambda}_1(q+p-1) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 \underline{I}_m \\ \vdots \\ \alpha_p \underline{I}_m \end{bmatrix} \quad (\text{A.42})$$

which yields (A.38) for $l = 2, \dots, m+1$ because the \underline{D} and $\underline{\Lambda}$'s (and in fact \underline{I}_m) matrices are diagonal. (We note that (A.35) similarly yields (A.38) for $l = 1, \dots, m$ so that it holds for all $l = 1, \dots, m+1$.) We may also use $\underline{\Lambda}_1^*(r)$ in (A.42) providing we also re-order the $\underline{D}_{1,i}$ to $\underline{D}_{1,i}^*$:-

$$\underline{D}_{1,i} = \text{dg}(d_2^{(i)}, \dots, d_{m+1}^{(i)})$$

$$\underline{D}_{1,i}^* = \text{dg}(d_{m+1}^{(i)}, d_2^{(i)}, \dots, d_m^{(i)})$$

$i = 1, \dots, p$

But whichever form of (A.42) we use we will still obtain (A.38). Similarly (A.40) will now hold since it corresponds to (A.36) with $1^{\text{st}} + m+1^{\text{st}}$ indicies interchanged. From the theorem the parameters will be

$$\begin{aligned} \underline{\phi}_i^* &= \phi_i \underline{I}_m + \underline{D}_{1,i}^* \underline{U}_m \quad i = 1, \dots, p \\ &= \phi_i \underline{I}_m + \begin{bmatrix} d_{m+1}^{(i)} & & 0 \\ & d_2^{(i)} & \\ & & \ddots \\ 0 & & & d_m^{(i)} \end{bmatrix} \underline{U}_m \end{aligned}$$

Thus to transform $\underline{\phi}_i^*$ to the parameters corresponding to the ARMA model with a.c.f. $\underline{\Gamma}(r)$, and not $\underline{\Gamma}(r)^*$ we again take the $\underline{Z}(1)$ transformation. We recall from chapter 4 that if

$$\underline{\Gamma}^*(r), \quad \underline{\phi}_i^* \quad \text{refers to } a_m (u_{t,m+1}, u_{t,2} \dots u_{tm}, u_{t1})$$

then $\underline{\Gamma}(r)$, $\underline{\phi}_i$ will correspond to $a_m (u_{t,1}, u_{t,2}, \dots u_{t,m+1})$ where,

$$\underline{\Gamma}(r) = \underline{Z}(1) \underline{\Gamma}^*(r) \underline{Z}'(1) \quad \text{the inverse of (A.41)}$$

$$\underline{\phi}_i = \underline{Z}(1) \underline{\phi}_i^* \underline{Z}(1)$$

so that $\underline{\phi}_i' = \underline{Z}'(1) (\phi_i \underline{I}_m + \underline{D}_{1,i}^* \underline{U}_m) \underline{Z}'(1)$

$$= \phi_i \underline{I}_m + \begin{bmatrix} -1 & -1 & \dots & -1 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} d_{m+1}^{(i)} \\ d_2^{(i)} \\ \vdots \\ d_m^{(i)} \end{bmatrix} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & \dots & -1 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$= \phi_i \underline{I}_m + \begin{bmatrix} -d_{m+1}^{(i)} & -d_2^{(i)} & \dots & -d_m^{(i)} \\ 0 & d_2^{(i)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_m^{(i)} \end{bmatrix} \begin{bmatrix} -1 & 0 & \dots & 0 \\ -1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ -1 & 0 & \dots & 0 \end{bmatrix}$$

$$= \phi_i \underline{I}_m + \begin{bmatrix} + \sum_{j=2}^{m+1} d_j^{(i)} & 0 & \dots & 0 \\ -d_2^{(i)} & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ -d_m^{(i)} & 0 & \dots & 0 \end{bmatrix}$$

$$= \phi_i \underline{I}_m + \underline{A}_{(1)}^{(i)}$$

$$i = 1, \dots, p$$

as required.



REFERENCES

- Akaike, H. (1974) A new look at the statistical model identification. IEEE Trans. Auto. Control 19 719-723.
- Akaike, H. (1976) Canonical correlation analysis of time series and the use of an information criterion. System Identification: Advances and Case Studies, Ed. R.K. Mehra and D.G. Lainiotis. Academic Press New York.
- Aitchison J. (1981a) Distributions on the simplex for the analysis of neutrality. Statistical Distributions in Scientific Work. Vol. 4, Ed. C. Taillie, G.P. Patil and B.A. Baldessari; D. Reidel Pub. Company.
- Aitchison J. (1981b) A new approach to null correlations of proportions. J. Math. Geol. 13 175-189
- Aitchison J. (1982) The statistical analysis of compositional data. J. Roy. Statist. Soc. B 44 139-177.
- Aitchison J. (1983) Principal component analysis of compositional data. Biometrika 70 57-65.
- Aitchison J. (1984) Reducing the dimensionality of compositional data Sets. J. Math. Geol. 16 617-636.
- Aitchison J. (1985) A general class of distributions on the simplex. J. Roy. Statist. Soc. B 47 136-146.
- Aitchison J. (1986) The Statistical Analysis of Compositional Data. Chapman and Hall, New York.
- Aitchison J. and Begg C.B. (1976) Statistical diagnosis when basic cases are not classified with certainty. Biometrika 63 1-12.
- Aitchison J. and Shen S.M. (1980) Logistic-normal distributions: Some properties and uses. Biometrika 67 261-72.
- Aitchison J. and Shen S.M. (1984) Measurement error in compositional data. J. Math. Geol. 16 637-650.
- Barnett V. (1976) The Ordering of Multivariate Data. J. Roy. Statist. Soc. A 139 318-355.
- Bartlett M.S. (1946) On the theoretical properties of the autocorrelated time series. J. Roy. Statist. Soc. B 8 27-41.
- Bickel P.J. and Doksum K.A. (1981) An analysis of transforms revisited. J. Amer. Statist. Assoc. 88 296-311
- Blight B.J.N. and Scott A.J. (1973) A stochastic model for repeated surveys. J. Roy. Statist. Soc. B 35 61-66.
- Box G.E.P. and Cox D.R. (1964) Analysis of transformations. J. Roy. Statist. Soc. B 26 211-252.

- Box G.E.P and Jenkins G.M.** (1976) Time Series Analysis: Forecasting and Control. Holden-Day , San Francisco.
- Box G.E.P. and Tiao G.C.** (1977) A canonical analysis of multiple time series. Biometrika 64 355-365.
- Brillinger D.R.** (1981) Time Series Analysis Theory. Holden-Day, San Fransisco.
- Chayes. F** (1960) On the correlation between variables of constant sum. J. Geophys. Res. 65 4185-4193.
- Connor R.J. and Mosimann J.E.** (1969) Concepts for proportions with a generalization of the dirichlet distribution. J. Amer. Statist. Assoc. 64 194-206.
- Cooper D.M. and Wood E.F.** (1982) Identifying Multivariate Time Series Models. J. Time Series Analy. 3 153-164
- Fuller W.A.** (1976) Introduction to Statistical Time Series. Wiley, New York.
- Geweke J.** (1982a) Measurement of linear dependence and feed back between multiple time series. J. Amer. Statist. Assoc. 77 304-324.
- Geweke J.** (1982b) Causality, exogeneity, and inference. Advances in Quantitative Economics. Ed. D. Hildenbrand, Cambridge Univ. Press.
- Geweke J.** (1984) Measures of conditional linear dependence and feedback between time series. J. Amer. Statist. Assoc. 79 907-915
- Geweke J. , Meese R. and Dent W.** (1983) Comparing alternative tests of causality in temporal systems: Analytical results and experimental evidence. J. Econometrics 21 161-194.
- Granger C.W.J.** (1969) Investigating causal relationships by econometric models and cross-spectral models. Econometrica 37 424-438.
- Granger C.W.J.** (1982) Generating mechanisms, models, and causality. Advances in Quantitative Economics. Ed. D. Hildenbrand, Cambridge Univ. Press.
- Granger C.W.J. and Newbold P.** (1976) Forecasting transformed series. J. Roy. Statist. Soc. B. 38 189-203.
- Green P.J. and Silverman B.W.** (1979) Constructing the convex hull of a set of points in the plane. Comp. Jor. 22 262-266.
- Haldane J.B.S.** (1948) Note on the median of a multivariate distribution. Biometrika 35 414-415.
- Hannan E.J.** (1969) The identification of vector mixed autoregressive moving average systems. Biometrika 56 223-225.
- Hannan E.J.** (1970) Multiple Time Series. Wiley, New York.
- Harvey A.C.** (1981) Time Series Models. Philip Allen Pub. Ltd. Oxford.

Haugh L.D. (1976) Checking the independence of two covariance stationary time series: A univariate residual cross-correlation approach. J Amer. Statist. Assoc. 71 378-385.

Hillmer S.C. and Tiao G.C. (1979) Likelihood function of stationary multiple autoregressive moving average models. J. Amer. Statist. Assoc. 74 652-660.

Hosking J.R.M (1980a) Tests of Time Series Model Specifications. Unpublished Ph.D. Thesis, University of Southampton.

Hosking J.R.M (1980b) Lagrange-multiplier tests of time series models. J. Roy. Statist. Soc. B 42 170-181.

Johnson N.L. (1949) Systems of frequency curves generated by methods of translation. Biometrika 36 149-176.

Johnson N.L. and Kotz S. (1970) Distributions in Statistics: Volumes 2 and 3: Continuous Univariate Distributions. Houghton Mifflin Comp. Boston.

Johnson N.L. and Kotz S. (1972) Distributions in Statistics: Volume 4: Continuous Multivariate Distributions. Houghton Mifflin Comp. Boston.

Lütkepohl H. (1982) Differencing multiple time series: Another look at Canadian money and income data. J. Time Series Analy. 3 235-243.

Marquardt D.W (1963) An algorithm for least squares estimation of non-linear parameters. J Soc. Indust. Applic. Math. 2 431-441.

Meeter D.A. (1965) Non-linear least squares (GAUSHAUS) . Supplementary Program Series No. 603, The University of Wisconsin Computing Center, Madison, Wisconsin.

Mood A.M. , Graybill F.A. and Boes D.C. (1974) Introduction to the Theory of Statistics. McGraw-Hill, Tokyo.

Morrison D.F. (1976) Multivariate Statistical Methods. McGraw-Hill, Tokyo.

Mosimann J.E. (1962) On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. Biometrika 49 65-82.

Newbold P. (1974) The exact likelihood function for a mixed autoregressive-moving average process. Biometrika 61 423-426.

Osborn D.R. (1977) Exact and approximate maximum likelihood estimates for vector moving average processes. J. Roy. Statist. Soc. B. 39 114-118.

Owen D.B. (1956) Tables for computing bivariate normal probabilities. Ann. Math. Statist. 27 1075-1090.

Pierce D.R. (1979) R^2 Measures for time series. J Amer. Statist. Assoc. 74 901-910.

Pierce D.R. (1982) Comment. J. Amer. Statist. Assoc. 77 315-316.

- Priestly M.B. (1981) Spectral Analysis and Time Series. Academic Press, New York.
- Seheult A. , Diggle P. and Evans D. (1976) Discussion of Professor Barnett's Paper. J. Roy. Statist. Soc. A. 139 351-352.
- Scott A.J. and Smith T.M.F. (1974) Analysis of repeated surveys using time series methods. J. Amer. Statist. Assoc. 69 674-678.
- Scott A.J. , Smith T.M.F. and Jones R.G. The application of time series methods to the analysis of repeated surveys. Inter. Statist. Review. 43 13-28
- Sibson R. (1984) Present position and potential developments: Some personal views. Multivariate analysis. J. Roy. Statist. Soc. 147 198-207.
- Sims A. (1972) Money, income and causality. American Economic Review 62 540-552.
- Smith T.M.F. (1978) Principles and problems in the analysis of repeated surveys. Survey Sampling and Measurement. Ed. N. Krishnam. Academic Press, New York.
- Steck G.P. (1958) A table for computing trivariate normal probabilities. Annal. Math. Statist. 29 780-800.
- Tiao G.C. and Box G.E.P. (1981) Modeling multiple time series with applications. J. Amer. Statist. Assoc. 76 802-816.
- Tjøstheim D and Paulsen J. (1982) Empirical identification of multiple time series. J. Time Series Analy. 3 265-282.
- Whittle P. (1953) The analysis of multiple stationary time series. J. Roy. Statist. Soc. B 15 125-139.
- Whittle P. (1963) On the fitting of multivariate autoregressions, and the general canonical factorization of a spectral density matrix. Biometrika 50 129-134.
- Zellner A. (1982) Comment. J. Amer. Statist. Assoc. 77 313-314.