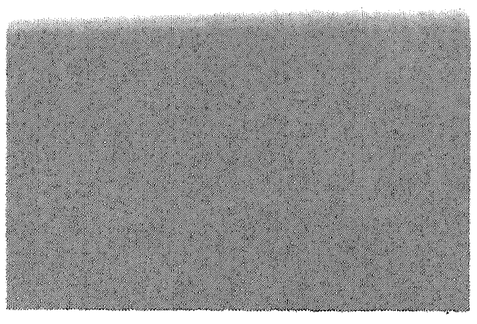
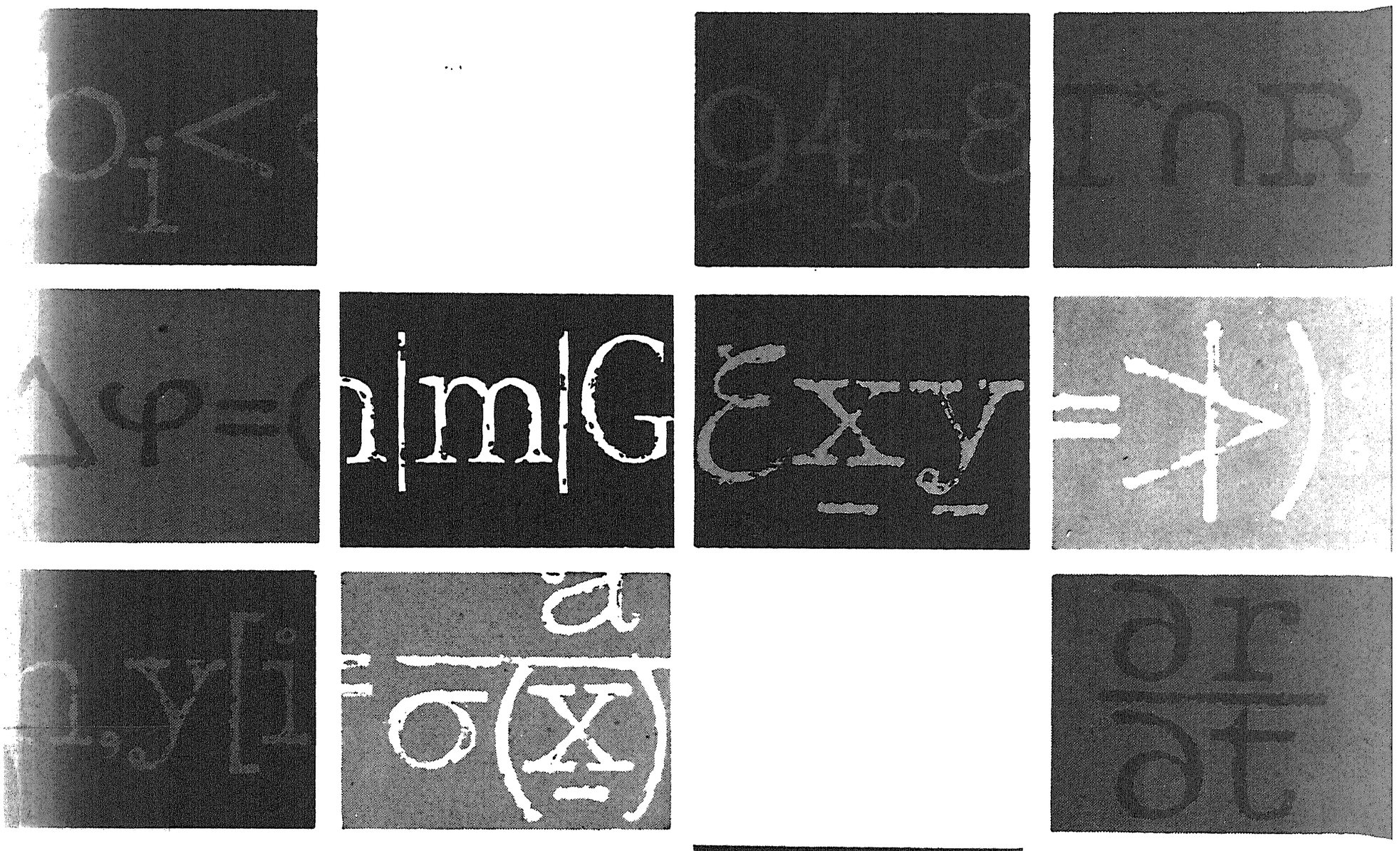


IDENTIFICATION AND INFORMATIVE SAMPLE SIZE

H.H. TIGELAAR



MATHEMATICAL CENTRE TRACTS

MC

147

MATHEMATICAL CENTRE TRACTS 147

IDENTIFICATION AND INFORMATIVE SAMPLE SIZE

H.H. TIGELAAR

MATHEMATISCH CENTRUM AMSTERDAM 1982

1980 Mathematics subject classification: 60G10, 62-00, 62M10, 62M15, 93E12

ISBN 90 6196 235 8

ACKNOWLEDGEMENTS

I am grateful to Prof. dr. B.B. van der Genugten for the many inspiring discussions, and I thank the Mathematical Centre for the opportunity to publish this monograph in their series Mathematical Centre Tracts and all those at the Mathematical Centre who have contributed to its technical realization.

CONTENTS

INTRODUCTION 1

List of notations and abbreviations 5

I A GENERAL APPROACH TO IDENTIFICATION

1.1 Classes of identifiable statistical statements 6

1.2 Conditional identifiability and informational independence 10

1.3 Identification and statistical procedures 18

1.4 Identification in statistical decision theory 19

1.5 Identification in Bayesian inference 22

1.6 Finite informative samples from stochastic processes and the problem of minimum informative sample size 25

1.7 Identification and prediction 28

1.8 Weak concepts of observational equivalence and strongly informative samples 31

1.9 Local identifiability 34

II INFORMATIVE SAMPLE SIZES IN UNIVARIATE STATIONARY MODELS

2.1 Introduction and preliminary results 37

2.2 The minimum informative sample size for MA(q) processes 43

2.3 Informative samples from AR(p) processes 47

2.4 Informative samples from ARMA(p,q) processes 49

2.5 Informative samples for the spectral measure; predictability 54

2.6 Application to linear regression with MA-errors and stationary lagged explanatory variables 58

III INFORMATIVE SAMPLE SIZES IN MULTIVARIATE STATIONARY MODELS

3.1 Introduction 61

3.2 The fundamental lemma and some special matrix theory 63

3.3 The minimum informative sample size for multivariate MA(q) processes 76
3.4 Informative samples from multivariate AR(p) processes 81
3.5 Informative samples from multivariate ARMA(p,q) processes; predictability 83

IV DYNAMIC SIMULTANEOUS EQUATIONS WITH MA-ERRORS
4.1 Introduction 95
4.2 Informative samples for the reduced form 99
4.3 Informative samples for the structural form 109
4.4 The non homogeneous case 116
4.5 The non mixed case 123

APPENDIX
A.1 Univariate spectral theory and stochastic difference equations 124
A.2 Some multivariate spectral theory 134

REFERENCES 139

SUBJECT INDEX 143

INTRODUCTION.

One of the fundamental problems in models of mathematical statistics is that of *identifiability*, that is the occurrence of *observationally equivalent* parametervalues. Two parametervalues are called observationally equivalent if they correspond to the same probability distribution. Clearly, one cannot distinguish between two such values on the basis of observations, and any attempt to do so is a priori meaningless. For example in a cointossing experiment it does not make sense to say something about the value of the coin (the unknown parameter), on the basis of the outcome head or tail (the observation). We shall refer to the definition of identifiability given above as the *classical definition* in contrast with more recent concepts of identifiability (for a survey see SCHÖNFELD [31]). The problem is often to see whether there exist observationally equivalent parameter values. In spite of its fundamentality only little attention had been paid to this kind of problem until 1950, when KOOPMANS and RIERSOL ([23] and [27]) tackled the problem for relatively simple linear relationships. Further BOSE introduced the concept of *estimability* ([3]), a concept closely related to identifiability, but less fundamental. However, when the models under consideration became more complicated, the identifiability problems became - from a mathematical point of view - more interesting, and often more difficult. Therefore it is not surprising that the most difficult identifiability problems arise in multivariate analysis, as for example in *factor analysis* and in *econometric models* (simultaneous equations). It is a remarkable fact, that FISHER, who treated the latter identifiability problem in 1966 [13], defines observational equivalence (and thus identifiability) in a way that is only valuable for the specific model under consideration. This is a dangerous approach as it may suggest that this definition can be generalized in a trivial way to models with lagged (dependent) variables (stochastic difference equations). This, however, is definitely not the case, and one of our goals is to make this point clear.

Another class of statistical problems where difficult identifiability problems arise is the statistical analysis of time series. Most important and widely used are stationary time series. The special identifiability problems are first recognized by HANNAN who in a fundamental paper ([18]) in 1968 treats the mixed autoregressive moving average model (ARMA):

$$\sum_{k=0}^{p} A_k \, \underline{x}_{t-k} = \sum_{j=0}^{q} B_j \, \underline{\varepsilon}_{t-j} \quad , \qquad t = 0, \pm 1, \ldots \quad ,$$

where the m-variate random process $\{\underline{x}_t\}$ is observable. (Throughout this thesis random variables will be denoted by underlined (lower case) letters).

In 1971 HANNAN, one of the leading authors in the field of identifiability in time series, also treated the multiple equation model with moving average errors ([17]). We also refer to DEISTLER, who treated models with stationary explanatory variables. ([4] and [5].)

Although most authors refer to the fundamental paper of HANNAN, and consider the ARMA case as completely solved, there is one important but unrecognized problem unsolved. To see what this problem is, it should be noted that instead of "probability distribution" in the classical definition it is more realistic to read: "probability distribution of the observed sample". Now the basic tool in the papers of HANNAN and DEISTLER is unique factorization of spectral densities and in this approach one has to study the probability law of the whole (observable) process rather than that of some finite sample. Since in practice one always has a finite sample, the identifiability problems have, in fact, only partially been solved. As far as we know, problems of this kind are not treated in the literature. Only recently MARAVALL [25] proved to be aware of it in the summary of his thesis. MARAVALL studied *local* identifiability in dynamic shock error models in contrast to the classical definition which is sometimes called *global* identifiability.

We shall not pay much attention to local identifiability in this thesis. Furthermore we shall restrict our attention to stochastic processes in discrete time. Identifiability problems for processes in continuous time are hardly found in literature; we refer to WESTCOTT [36].

Fairly general approaches to the theory of identifiability have been made by SCHÖNFELD [31] and more recently by van der GENUGTEN [14]. Following SCHÖNFELD one can easily get the impression that there is a close connection between identification and estimation and therefore that identifiability problems are part of estimation theory. This, however, is rather misleading. As van der GENUGTEN points out, identifiability problems may arise in other statistical problems such as hypothesis testing.

In Chapter I we shall present a general approach to identifiability, that enables us to recognize identifiability problems in all kinds of statistical problems, in particular in *statistical prediction problems*.

Although we shall not be concerned with Bayesian inference and statistical decision theory, we shall make one excursion into those fields. KADANE [22] says:

> "One general question unresolved in this literature is, whether Bayesian theory requires a different definition of identification from the classical one".

Or ROTHENBERG ([29] p. 14):

> "We leave unanswered the question of an appropriate Bayesian definition of identification".

MORALES ([26] p. 20) reports:

> "The concept of identification in a Bayesian context is not alltogether clear. We shall adopt the view of considering a structure 'identified' if the posterior density of the parameters of the model is not 'flat' on a subspace of the parameter space. This point of view may not be entirely satisfactory".

Although the problem is not completely ignored, (as is in fact

done by LINDLEY ([24] p 46, footnote 34)) the only author dealing with this problem is KADANE who presented a Bayesian approach in 1975 using the classical definition. As this in our opinion is not quite satisfactory we present an alternative approach in § 1.5. KADANE also pays some attention to the role of identification in statistical decision theory, a topic hardly treated in literature. However, the question what identifiability really means in a decision-theoretic setting remains unanswered. A few ideas are presented in § 1.2.

In Chapter II univariate stationary models are treated, and in Chapter III the corresponding multivariate models. We treat them separately, not only for sake of clarity but also because most multivariate problems are essentially more difficult than the corresponding univariate ones, and the 'obvious' generalization may be false. The results of these two chapters may have some interest outside the probabilistic setting as they can be seen as results in the theory of matrices with rational functions of a complex variable as elements.

In Chapter IV we shall deal with dynamic simultaneous equations with moving average errors, using results of Chapter III.

LIST OF NOTATIONS AND ABBREVIATIONS

$\emptyset$	empty set
V^c	complement of the set V
1_V	indicator function of the set V
$\mathbb{R}$	set of real numbers
$\mathbb{R}^+$	set of positive real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{C}$ (m)	set of complex m × m matrices
$\mathbb{C}^+$ (m)	set of hermitian positive definite m × m matrices
$\mathbb{C}^k$ (m)	k-fold cartesian product of $\mathbb{C}$(m) with itself.
$\bar{z}$	complex conjugate of z
I_m	m × m unit matrix
A'	transpose of the matrix A
A*	complex conjugate transpose of A
tr A	trace of the square matrix A
r [A]	rank of the matrix A
$A \geq 0$	if $A \in \mathbb{C}(m)$ and A is semi definite positive
ker A	nullspace of the matrix A
< A >	linear space spanned by the columns of A
$L^\perp$	orthogonal complement of the linear subspace L
$\underline{x}, \underline{y}, \underline{\varepsilon}, \ldots$	random variables or - vectors
E {.}	expectation
V {.}	covariance matrix
i.i.d.	independently identically distributed
l.i.m.	limit in the mean i.e. $\underset{n \to \infty}{\text{l.i.m.}}\ \underline{x}_n = \underline{x}$ if $\lim_{n\to\infty} E\{\lvert \underline{x}_n - \underline{x} \rvert^2\} = 0$
$\lVert . \rVert$	euclidean norm of vector or matrix
δ_t	Kronecker δ i.e. $\delta_0 = 1$ and $\delta_t = 0$, $t \neq 0$
A : = B	A is defined by B
iff	if and only if
□	end of proof

CHAPTER I

A GENERAL APPROACH TO IDENTIFICATION

1.1 CLASSES OF IDENTIFIABLE STATISTICAL STATEMENTS

In (non-sequential) statistical inference the observational material (the sample) is considered to be a realization of some random vector or process $\underline{x}$ that takes its values in a measurable space $(X, \mathcal{B})$ (the sample space). The only thing the statistician knows about the true distribution of $\underline{x}$ is that it belongs to a given class $\mathcal{P}$ of probability distributions on $(X, \mathcal{B})$.

In most statistical problems the class $\mathcal{P}$ admits a natural and simple parametric representation. More precisely, a mapping P is given from a known parameterspace θ into a given class of probability distributions on $(X, \mathcal{B})$. The range of this mapping is $\mathcal{P}$ and if P_θ denotes the image of $\theta \in \theta$ under P then we can shortly write $\mathcal{P} = \{P_\theta \mid \theta \in \theta\}$. The corresponding statistical problem will be denoted by the triple $(\underline{x}, \mathcal{P}, \theta)$. The goal of a statistician is to know something more about the true parametervalue than that it belongs to θ. Thus it is natural to consider subsets of θ and to identify them with statistical statements. This leads to the following definition.

DEFINITION 1.1.1 A *statistical statement* is a subset $\theta_o \subset \theta$.

REMARK. A possible interpretation is that a statement is true iff the unknown parametervalue belongs to it.

The parametric formulation is very attractive because of the direct interpretation of the parameter. However, it can introduce the problem of *identification*. Suppose there exist $\theta_o \in \theta_o$ and $\theta_1 \in \theta_o^c$ with $P_{\theta_o} = P_{\theta_1}$. We say that θ_o and θ_1 are *observationally equivalent* if $P_{\theta_o} = P_{\theta_1}$. The statistician

should then refuse to make the statement θ_o (or θ_o^c) because it discriminates between the observationally equivalent values θ_o and θ_1 while θ_o indicates that θ_o is true and θ_1 that it is false. Therefore a natural concept in statistical inference is the identifiability of statements.

DEFINITION 1.1.2 The statistical statement θ_o is called *identifiable w.r.t.* $\mathcal{P}$ or equivalently $\underline{x}$ is said to be *informative* for θ_o, if for all θ_1, $\theta_1 \in \theta$ we have the implication

$$\theta_o \in \theta_o \ , \ \theta_1 \in \theta_o^c \Rightarrow P_{\theta_o} \neq P_{\theta_1} \ .$$

It should be noted that observational equivalence is an equivalence relation on θ and therefore induces a dissection of θ into equivalence classes, called observational equivalence-classes. Thus, if we accept the axiom of choice, it is formally always possible to avoid identifiability problems by defining a new parameterspace consisting of one element out of each equivalence class. Of course such a reduction may be difficult to perform in practice, but that would not be a fundamental objection. This reduction is in general not reasonable because it may destroy the simple and natural form of the parameterspace in which case the parameter looses its natural interpretation.

From a mathematical point of view it is interesting to consider classes of statements.

THEOREM 1.1.3 Let J be the class of all identifiable statements. Then we have

a) $\theta_o \in J \iff \theta_o^c \in J$,

b) $\theta_\nu \in J$, $\nu \in N \Rightarrow \bigcap_{\nu \in N} \theta_\nu \in J$ for arbitrary index set N.

The simple proof is omitted.

REMARK. a) and b) imply

c) $\theta_\nu \in \mathcal{J},\ \nu \in N \Rightarrow \bigcup_{\nu \in N} \theta_\nu \in \mathcal{J}$,

d) $\emptyset \in \mathcal{J},\ \theta \in \mathcal{J}$.

In most statistical problems the statistician is not interested in all statements but merely in a certain class of statements. If $\{\theta_\nu\}_{\nu \in N}$ is a class of statements the statistician is interested in, which means that he is willing to say whether θ_ν is true or false for all $\nu \in N$, then he should also be interested in all statements that can be formed from them by taking complements and / or intersections. Thus the statistician is in fact interested in a class of statements with the properties a) and b). Therefore we define

DEFINITION 1.1.4 A class of statements with the properties a) and b) is called an *informational class*.

REMARK 1. Statements of an informational class are not necessarily identifiable.

REMARK 2. If $\mathcal{J}_o$ is the smallest informational class that contains a given set of statements $\{\theta_\nu\}_{\nu \in N}$ then $\mathcal{J}_o$ is said to be *generated* by $\{\theta_\nu\}_{\nu \in N}$.

Two simple examples will illustrate the ideas.

EXAMPLE 1.1.5 If the statistician is interested in *point estimation* he will consider all one-point subsets (singletons). The smallest informational class that contains all singletons is the class of all statements.

EXAMPLE 1.1.6 If the statistician is dealing with a hypothesis testing problem, he will consider only two complementary subsets θ_o and θ_o^c. The smallest informational class that contains θ_o (and θ_o^c) is $\{\theta_o, \theta_o^c, \emptyset, \theta\}$ and will be denoted by $\mathcal{J}_{\theta_o}$.

Both examples are special cases of the more general situation where the statistician is primarily interested in the value taken by a given mapping $\varphi : \Theta \to \Lambda$ from Θ into some space Λ. In such cases attention is restricted to statements that can be formulated in terms of φ. Formally

DEFINITION 1.1.7 A statistical statement θ_o is said to be in terms of $\varphi : \Theta \to \Lambda$ if there exists a subset $\Lambda_o \subset \Lambda$ such that $\theta_o = \varphi^{-1}(\Lambda_o)$.

LEMMA 1.1.8 The class of all statements in terms of φ is an informational class.

The proof is very simple and will be omitted. The informational class is said to be generated by φ and will be denoted by $\mathcal{J}_\varphi$.

EXAMPLE 1.1.9 (see also examples 1.1.5 and 1.1.6)

a) If $\varphi : \Theta \to \Theta$ is the identity map, then $\mathcal{J}_\varphi$ is the class of all subsets.

b) If $\varphi := 1_{\theta_o}$ is the indicatorfunction of a subset θ_o of Θ, then $\mathcal{J}_\varphi = \{\theta_o, \theta_o^c, \emptyset, \Theta\} = \mathcal{J}_{\theta_o}$.

DEFINITION 1.1.10 The mapping φ is called identifiable w.r.t. $\mathcal{P}$, or equivalently $\underline{x}$ is said to be informative for φ, if every statement in terms of φ is identifiable.

REMARK. It follows from the remark on theorem 1.1.3 that for φ to be identifiable it is sufficient that $\varphi^{-1}(\{\lambda\})$ is identifiable for all $\lambda \in \Lambda$.

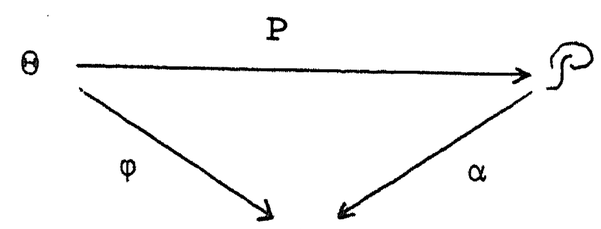
EXAMPLE 1.1.11 (see also example 1.1.9). The statement θ_o is identifiable iff its indicator function 1_{θ_o} is identifiable.

The following lemma shows that a mapping φ is identifiable iff it is constant on observational equivalence classes and thus definition 1.1.10 is indeed what we intuitively want it to be.

LEMMA 1.1.12 $\underline{x}$ is informative for $\varphi : \Theta \to \Lambda$ iff for all Θ_1, $\Theta_2 \in \Theta$ the implication $\varphi(\Theta_1) \neq \varphi(\Theta_2) \Rightarrow P_{\Theta_1} \neq P_{\Theta_2}$ holds.

PROOF. Let $\underline{x}$ be informative for φ and $\varphi(\Theta_1) \neq \varphi(\Theta_2)$. Then the statements $\varphi^{-1}(\{\varphi(\Theta_1)\})$ and $\varphi^{-1}(\{\varphi(\Theta_2)\})$ are identifiable since φ is and $\varphi^{-1}(\{\varphi(\Theta_1)\}) \cap \varphi^{-1}(\{\varphi(\Theta_2)\}) = \emptyset$ since $\varphi(\Theta_1) \neq \varphi(\Theta_2)$. Hence $P_{\Theta_1} \neq P_{\Theta_2}$.
Conversely let the implication hold and $\lambda \in \Lambda$ be such that $\varphi^{-1}(\{\lambda\}) \neq \emptyset$ and $\varphi^{-1}(\{\lambda\}) \neq \Theta$. It follows immediately that $\varphi^{-1}(\{\lambda\})$ is identifiable and since λ was arbitrary the result follows from the remark following def. 1.1.10 □.

REMARK. From the lemma it follows that $\underline{x}$ is informative for φ iff there exists a mapping $\alpha : \mathcal{P} \to \Lambda$ such that $\varphi = \alpha \circ P$.
In particular it follows that $\underline{x}$ is informative for its moments.



So far the consideration of classes of statements does not give new results. However it turns out that it is a fruitful basis for the presentation of new ideas and for extending the theory to Bayesian statistics and statistical decision theory. This will be done in the next sections.

1.2 CONDITIONAL IDENTIFIABILITY AND INFORMATIONAL INDEPENDENCE

Let θ_o be an arbitrary statement. Then we shall denote the dissection $\{\theta_o, \theta_o^c\}$ of θ by $\mathcal{D}_{\theta_o}$. More generally we consider arbitrary dissections $\{D_\nu\}_{\nu \in N}$ of θ. A mapping $\varphi : \theta \to \Lambda$ induces a dissection $\mathcal{D}_\varphi = \{D_\lambda\}_{\lambda \in \Lambda}$ of θ where $D_\lambda := \varphi^{-1}(\{\lambda\})$. (Every dissection can be generated in this way by a function).
Consider the dissection $\mathcal{D} = \{D_\nu\}_{\nu \in N}$. We shall call the

values θ_1 and θ_2 $\mathfrak{D}$-*equivalent* if they belong to the same D_ν and write $\theta_1 \overset{\mathfrak{D}}{\sim} \theta_2$. It is often easy to see that for some $\mathfrak{D}$, $\mathfrak{D}$ - equivalent values of θ are not observational equivalent. Therefore we define

DEFINITION 1.2.1 The statement θ_1 is said to be identifiable *conditional* on the dissection $\mathfrak{D}$, or equivalently $\underline{x}$ is called informative for θ conditional on $\mathfrak{D}$ if for all θ_1 , $\theta_2 \in \theta$ the following implication holds

$$\left.\begin{array}{l} \theta_1 \in \theta_1 \ , \ \theta_2 \in \theta_1^c \\ \theta_1 \overset{\mathfrak{D}}{\sim} \theta_2 \end{array}\right\} \Rightarrow P_{\theta_1} \neq P_{\theta_2}.$$

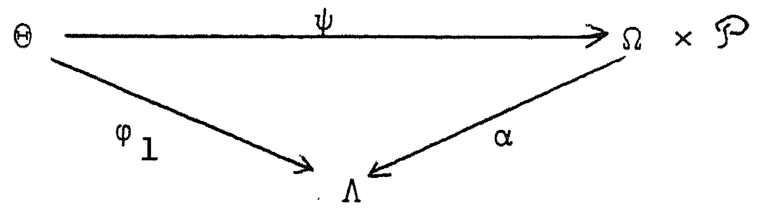
The mapping $\varphi_1 : \theta \to \Lambda$ is said to be identifiable conditional on $\mathfrak{D}$ if every statement in terms of φ_1 is. The statement θ_1 or the mapping φ_1 is called identifiable conditional on φ if $\mathfrak{D} = \mathfrak{D}_\varphi$.

In the same way as lemma 1.1.8 we have

LEMMA 1.2.2 The mapping $\varphi_1 : \theta \to \Lambda$ is identifiable conditional on $\varphi_2 : \theta \to \Omega$ iff for all θ_1 , $\theta_2 \in \theta$ the following implication holds

$$\left.\begin{array}{l} \varphi_1(\theta_1) \neq \varphi_1(\theta_2) \\ \varphi_2(\theta_1) = \varphi_2(\theta_2) \end{array}\right\} \Rightarrow P_{\theta_1} \neq P_{\theta_2} .$$

REMARK. It follows from this lemma that $\underline{x}$ is informative for $\varphi_1 : \theta \to \Lambda$ conditional on $\varphi_2 : \theta \to \Omega$ iff there exists a mapping $\alpha : \Omega \times \mathcal{P} \to \Lambda$ such that $\varphi_1 = \alpha \circ \psi$ where $\psi : \theta \to \Omega \times \mathcal{P}$ is defined by $\psi(\theta) = (\varphi_2(\theta), P_\theta)$, $\theta \in \theta$.



THEOREM 1.2.3 (*Conditional identification theorem*).
If $\underline{x}$ is informative for φ_1 conditional on φ_2 and $\underline{x}$ is informative for φ_2, then $\underline{x}$ is informative for φ_1.

PROOF. Suppose $\varphi_1(\theta_1) \neq \varphi_1(\theta_2)$. If $\varphi_2(\theta_1) = \varphi_2(\theta_2)$ we have $P_{\theta_1} \neq P_{\theta_2}$ by lemma 1.2.2. If $\varphi_2(\theta_1) \neq \varphi_2(\theta_2)$ we have $P_{\theta_1} \neq P_{\theta_2}$ since $\underline{x}$ is informative for φ_2.□

EXAMPLE 1.2.4 Consider the following simple model

$$a\,\underline{x} = \mu + \underline{\varepsilon}\ ,$$

where $E\{\underline{\varepsilon}\} = 0$, $V\{\underline{\varepsilon}\} = 1$, and where $a \in A \subset \mathbb{R}^+$ and $\mu \in M \subset \mathbb{R}$ are unknown constants. The random variable $\underline{\varepsilon}$ is unobservable and $\underline{x}$ is observable. If $\xi \in Z$ is a parameter that characterizes the distribution of $\underline{\varepsilon}$ then we may put $\theta := (a, \mu, \xi)$ and the natural parameterspace is $\theta = A \times M \times Z$.
Let the functions φ_1 and φ_2 be defined by $\varphi_1(\theta) = a$, $\theta \in \theta$ and $\varphi_2(\theta) = \mu$, $\theta \in \theta$. Since $A \subset \mathbb{R}^+$, different a-values correspond to different variances of $\underline{x}$ and therefore to different distributions of $\underline{x}$. Hence $\underline{x}$ is informative for φ_1. On the other hand, if a is held fixed, different μ-values correspond to different expectations of $\underline{x}$ and thus $\underline{x}$ is informative for φ_2 conditional on φ_1. By the conditional identification theorem it follows that $\underline{x}$ is informative for φ_1. Note that $\mu_1 \neq \mu_2$ does not imply different expectations for $\underline{x}$.

Intuitively one could expect that if the mappings φ_1 and φ_2 are in some sense 'independent', conditional identifiability should imply identifiability. In the next section we develop such a concept of independence.

Let $\theta_o \subset \theta$ be a statistical statement. Then there surely exist identifiable statements that have θ_o as a subset (e.g. θ), and it is easily seen that there exists a uniquely determined smallest identifiable statement with this property (take the union of all observational equivalence classes that have nonempty intersection with θ_o). Therefore we put

DEFINITION 1.2.5 The *identifiable hull* $\mathcal{H}(\theta_o)$ of θ_o is the smallest identifiable statement that has θ_o as a subset $\mathcal{H}(\theta_o) = \{ \theta \in \Theta \mid \exists\, \theta_o \in \theta_o,\ P_\theta = P_{\theta_o} \}$. Clearly, $\mathcal{H}(\theta_o) = \theta_o$ iff θ_o is identifiable.

REMARK. Since $\mathcal{H}(\theta_o) \cap \mathcal{H}(\theta_1)$ is an identifiable statement as intersection of identifiable statements we always have $\mathcal{H}(\theta_o \cap \theta_1) \subset \mathcal{H}(\theta_o) \cap \mathcal{H}(\theta_1)$.

DEFINITION 1.2.6 The informational classes J_o and J_1 are called *informationally independent* if for all $\theta_o \in J_o$ and $\theta_1 \in J_1$ we have $\mathcal{H}(\theta_o \cap \theta_1) = \mathcal{H}(\theta_o) \cap \mathcal{H}(\theta_1)$. Two mappings φ_1 and φ_2 are informationally independent if the informational classes J_{φ_1} and J_{φ_2} are. Two statements θ_o and θ_1 are informationally independent if J_{θ_o} and J_{θ_1} are.

Before we can establish the relation between conditional identifiability and informational independence we need the following lemma.

LEMMA 1.2.7 Let $\underline{x}$ be informative for φ_2 conditional on φ_1. If $\theta_o \in J_{\varphi_2}$, $\theta_1 \in \theta_o$ and $\theta_1 := \varphi_1^{-1}(\{\varphi_1(\theta_1)\})$ then $\mathcal{H}(\theta_o \cap \theta_1) \cap \theta_1 = \theta_o \cap \theta_1$.

PROOF. One way ($\supset$) being trivial we only have to prove $\mathcal{H}(\theta_o \cap \theta_1) \cap \theta_1 \subset \theta_o \cap \theta_1$. If $\theta_o \cap \theta_1 = \emptyset$ this is trivial, so suppose $\theta_o \cap \theta_1 \neq \emptyset$. Let $\theta_o \in \mathcal{H}(\theta_o \cap \theta_1) \cap \theta_1$ be arbitrary. Then $\theta_o \in \theta_1$ and thus we have to prove $\theta_o \in \theta_o$. Suppose $\theta_o \in \theta_o^c$. Then by the definition of $\mathcal{H}(\theta_o \cap \theta_1)$ there exists $\theta_2 \in \theta_o \cap \theta_1$ such that $P_{\theta_o} = P_{\theta_2}$. We also have $\theta_o \neq \theta_2$ and $\varphi_1(\theta_o) = \varphi_1(\theta_2) = \varphi_1(\theta_1)$ since $\theta_o, \theta_2 \in \theta_1$. But then we have $P_{\theta_o} \neq P_{\theta_2}$ since φ_2 (and thus θ_o) is identifiable condi-

tional on φ_1. Thus we have a contradiction and the lemma is proved. □

THEOREM 1.2.8 If $\underline{x}$ is informative for φ_2 conditional on φ_1 and φ_1 and φ_2 are informationally independent then $\underline{x}$ is informative for φ_2.

PROOF. Let $\theta_o \in J_{\varphi_2}$ be arbitrary and $\Theta_1 \in \theta_o$. If $\Theta_2 \in \theta_o^c$ then we have to prove that $P_{\Theta_1} \neq P_{\Theta_2}$. Put $\theta_1 := \varphi_1^{-1}(\{\varphi_1(\Theta_1)\})$. Then $\Theta_1 \in \theta_1$ and since $\underline{x}$ is informative for φ_2 conditional on φ_1 we have by lemma 1.2.7

$$(1.2.1) \qquad \mathcal{K}(\theta_o \cap \theta_1) \cap \theta_1 = \theta_o \cap \theta_1 .$$

By informational independence we also have

$$(1.2.2) \qquad \mathcal{K}(\theta_o \cap \theta_1) = \mathcal{K}(\theta_o) \cap \mathcal{K}(\theta_1) ,$$

and since $\theta_1 \subset \mathcal{K}(\theta_1)$, (1.2.1) and (1.2.2) imply

$$\mathcal{K}(\theta_o \cap \theta_1) = \theta_o \cap \theta_1 .$$

Thus $\theta_o \cap \theta_1$ is an identifiable statement with $\Theta_1 \in \theta_o \cap \theta_1$ and $\Theta_2 \in (\theta_o \cap \theta_1)^c$. Hence $P_{\Theta_1} \neq P_{\Theta_2}$ and the theorem is proved. □

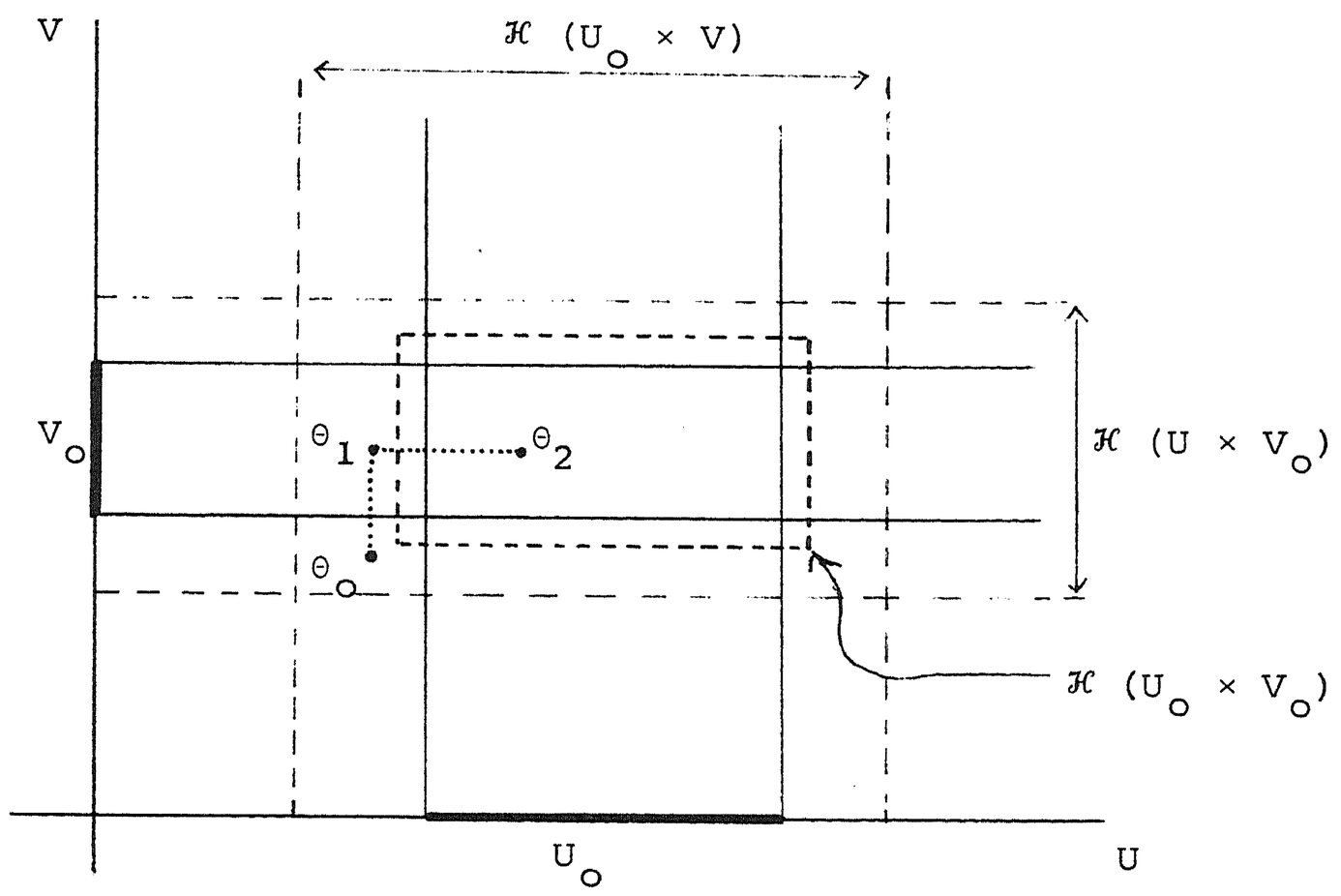
We shall consider one important case more closely. Suppose θ is of the form $\theta \subset U \times V$ and let $\Theta = (\varphi_1(\Theta), \varphi_2(\Theta))$, $\Theta \in \theta$ where $u := \varphi_1(\Theta) \in U$ and $v := \varphi_2(\Theta) \in V$. Thus $\varphi_1(\Theta)$ and $\varphi_2(\Theta)$ are *projections* of Θ on U and V, respectively, and the question arises when φ_1 and φ_2 are informationally independent. We have

THEOREM 1.2.9 The projections φ_1 and φ_2 are informationally independent iff all classes of observationally equivalent values

of Θ are of the form $U_o \times V_o$, $U_o \subset U$, $V_o \subset V$.

PROOF. (If) Let U_o and V_o be two arbitrary subsets of U and V respectively. Then we have to prove

$$\mathcal{H}(U_o \times V_o) = \mathcal{H}(U_o \times V) \cap \mathcal{H}(U \times V_o) .$$



One inclusion ($\subset$) being trivial we only have to prove the other ($\supset$). Let $\Theta_o \in \mathcal{H}(U_o \times V) \cap \mathcal{H}(U \times V_o)$. Because $\Theta_o \in \mathcal{H}(U \times V_o)$ there exists $\Theta_1 \in U \times V_o$ with $P_{\Theta_o} = P_{\Theta_1}$ and since the observational equivalence class to which Θ_o belongs is of the form $U_1 \times V_1$, Θ_1 can be chosen such that $\varphi_1(\Theta_1) = \varphi_1(\Theta_o)$ and $\Theta_1 \in U \times V_o$. In the same way there exists $\Theta_2 \in U_o \times V_o$ such that $P_{\Theta_2} = P_{\Theta_1}$ and $\varphi_2(\Theta_2) = \varphi_2(\Theta_1)$. But then we have $P_{\Theta_o} = P_{\Theta_2}$ and so $\Theta_o \in \mathcal{H}(U_o \times V_o)$.

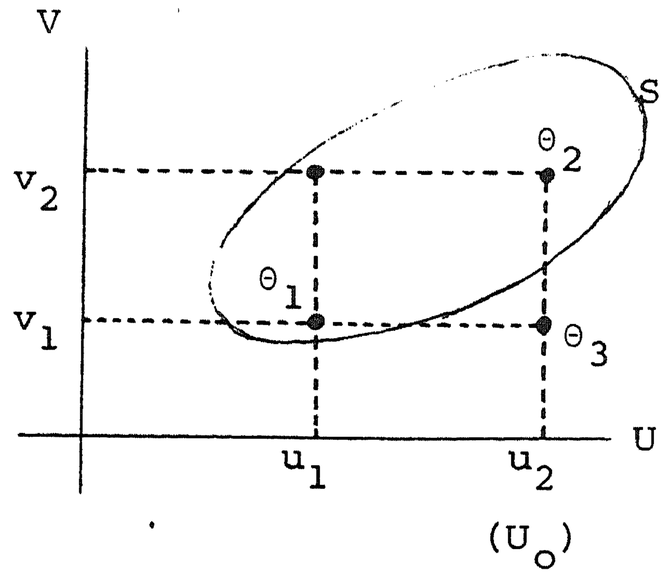
(Only if) Let φ_1 and φ_2 be informationally independent and suppose there exists an equivalence class $S \subset \Theta$ which is not the cartesian product of subsets of U and V. Then there exist

$\theta_1 = (u_1 , v_1) \in S$ and $\theta_2 = (u_2 , v_2) \in S$ such that either $\theta_3 := (u_2 , v_1) \notin S$ or $\theta_4 := (u_1 , v_2) \notin S$.

Suppose $\theta_3 \notin S$.

(see figure).

Choose $U_o := \{u_2\}$ and $V_o := \{v_1\}$. Then $U_o \times V_o = \{\theta_3\}$ and so $\mathcal{H}(U_o \times V_o) \cap S = \emptyset$ since S is an equivalence class and $\theta_3 \notin S$.



We also have $\mathcal{H}(U_o \times V) \supset S$ since $\theta_2 \in (U_o \times V) \cap S$.

Similarly $\mathcal{H}(U \times V_o) \supset S$. Hence

$$S \subset \mathcal{H}(U_o \times V) \cap \mathcal{H}(U \times V_o) .$$

Since φ_1 and φ_2 are informationally independent we have $\mathcal{H}(U_o \times V) \cap \mathcal{H}(U \times V_o) = \mathcal{H}(U_o \times V_o)$ and so $S \subset \mathcal{H}(U_o \times V_o)$. This contradicts $\mathcal{H}(U_o \times V_o) \cap S = \emptyset$ and proves the theorem.□

EXAMPLE 1.2.10 Consider the standard univariate linear regression model

$$\underline{y} = X \beta + \underline{\varepsilon} \quad , \quad E\{\underline{\varepsilon}\} = 0 \quad ,$$

where $\underline{y}$ is the n-vector of observations, X is a known $n \times k$ matrix of k explanatory variables, β is a k-vector of unknown regression coefficients and $\underline{\varepsilon}$ is an n-vector of (unobservable) errors. If $\xi \in V$ is a parameter that characterizes the distribution of $\underline{\varepsilon}$ and $\beta \in U \subset \mathbb{R}^k$, we may put $\theta := (\beta , \xi)$ and the natural choice for θ is $U \times V$. Note that the distribution P_θ of $\underline{y}$ depends on β through $E\{\underline{y}\} = X \beta$. Thus if $\theta_1 = (\beta_1 , \xi_1)$ and $\theta_2 = (\beta_2 , \xi_2)$ are observationally equivalent we must have $X \beta_1 = X \beta_2$. But then (β_1 , ξ_1) and (β_2 , ξ_1) are observationally equivalent and also (β_1 , ξ_2) and (β_2 , ξ_2).

Thus the observational equivalence classes are of the form $U_o \times V_o$, $U_o \subset U$, $V_o \subset V$ and so β and ξ are informationally independent by theorem 1.2.9. It follows from theorem 1.2.8 that in order to investigate identifiability of ξ we may consider $\underline{y} - X\beta$ as observable, and for identification of β we may consider $\underline{y} - \underline{\varepsilon} = X\beta$ as observable. The latter implies the well-known results that if $\theta = \mathbb{R}^k \times V$, $n \geq k$ then a necessary and sufficient condition for identifiability of β is $r[X] = k$, and that for identifiability of $d'\beta$, $d \in \mathbb{R}^k$ a necessary and sufficient condition is $d \in \langle X' \rangle$.

EXAMPLE 1.2.11 *(Error in variables model)* Let the variables μ_t and ν_t be related through

$$\mu_t = \alpha + \beta \nu_t \ , \qquad t = 1, 2, \ldots$$

Suppose we can only observe μ_t and ν_t with observational error, i.e. we observe at $t = 1, \ldots , n$

$$\underline{x}_t := \nu_t + \underline{\varepsilon}_t \ , \quad \underline{y}_t := \mu_t + \underline{\eta}_t \ ,$$

where $E\{\underline{\varepsilon}_t\} = E\{\underline{\eta}_t\} = 0$. Let $\xi \in V$ be a parameter characterizing the distribution of $(\underline{\varepsilon}_t , \underline{\eta}_t)$, $t = 1, 2, \ldots , n$. Put $\theta := (\alpha, \beta, \nu_1 \ldots, \nu_n , \xi)$. Obviously, if θ is such that ν_t is allowed to be constant over time, (α , β) is in general not identifiable. We shall show that if θ is a subset of $\mathbb{R}^{n+2} \times V$ such that ν_t is not constant over time, then (α , β) is identifiable. The model can be put in the form of a linear regression model with unknown regressor ν_t

$$\underline{y}_t = \alpha + \beta \nu_t + \underline{\eta}_t \ , \qquad t = 1, 2, \ldots, n \ .$$

Since ν_t is not constant over time, it follows from example 1.2.10 that (α , β) is identifiable conditional on $(\nu_1 \ldots \nu_n)$. However, $(\nu_1 \ldots \nu_n)$ is identifiable since it is the expectation of the observable vector $(\underline{x}_1 \ldots \underline{x}_n)$. Thus it follows by theorem 1.2.3 that (α , β) is identifiable.

1.3 IDENTIFICATION AND STATISTICAL PROCEDURES

In this section we present some new ideas on identification that, roughly speaking, tell the statistician what statistical procedures should be forbidden in the presence of observationally equivalent θ's.

Let J_o denote the informational class of statements the statistician is interested in. Note that since $\theta \in J_o$ there is no value of θ that is excluded a priori by the statistician from being the true value. Although not all statements in J_o need to be identifiable, there exist identifiable statements in J_o (e.g. θ) and there exists a largest informational class $\tilde{J}_o$ of identifiable statements in J_o. Suppose J_o is endowed with a σ-field $\mathcal{S}$ of subclasses such that $\tilde{J}_o \in \mathcal{S}$.

DEFINITION 1.3.1 A statistical procedure is a measurable mapping $d : (X, \mathcal{B}) \to (J_o, \mathcal{S})$ with the interpretation that if x is observed, then the statistician makes the statement $d(x)$.

Before a statistician answers the question what 'good' statistical procedures are, he should answer the question what procedures he will consider as a priori meaningless. Intuitively it seems reasonable to ignore procedures that can produce unidentifiable statements with positive probability for some $\theta \in \theta$. This motivates the following definition.

DEFINITION 1.3.2 A statistical procedure d is called *ignorable* if $P_\theta \{d(\underline{x}) \in \tilde{J}_o\} < 1$ for some $\theta \in \theta$.

Thus the statistician will only consider statistical procedures d with

$$P_\theta\{ d(\underline{x}) \in \tilde{J}_o \} = 1 \quad , \qquad \theta \in \theta \ .$$

EXAMPLE 1.3.3 Consider the case where J_o is the class of all statistical statements. This implies that the statistician is

interested in point estimates of θ , Then definition 1.3.2 implies that a non-ignorable estimator takes its values almost surely in one-point observational equivalence classes or, equivalently, with probability one ($\forall_\theta$) it does not discriminate between observational equivalent values of θ. In the case that all observational equivalence classes contain more than one element (or : there is no identifiable *singleton*) the statistician should refuse to produce point estimates. It is important to mention however, that this does not imply that every other statistical statement such as region estimates or acceptation of a hypothesis, is a priori meaningless.

1.4 IDENTIFICATION IN STATISTICAL DECISION THEORY

In this section we extend the statistical problem $(\underline{x}, \mathcal{P}, \theta)$ to the statistical decision problem $(\underline{x}, \mathcal{P}, \theta, A, L)$, where -A is the space of *actions* (or *strategies*) for the statistician.

-L is a mapping from $\theta \times A$ into a space C, called the space of *consequences*. In most cases L takes real values and is then called the *loss function* or *pay-off function*, and $L(\theta, \alpha)$ is interpreted as the penalty for the statistician for taking action α if θ is the true parameter.

Let $L_\alpha(\theta) := L(\theta, \alpha)$, $\theta \in \Theta$ denote the *section* of L at $\alpha \in A$. Thus L_α is a mapping from θ into the space of consequences C.

When the statistician has to make up his mind whether he shall take action α or not, he will base his decision on the consequence $L_\alpha(\theta)$. This is,however, impossible if $L_\alpha(\theta)$ is unknown but he may hope that if $\underline{x}$ is informative for L_α, he can make a choice which is not a pure gamble.

Let A_o denote the set of actions α such that $\underline{x}$ is informative for L_α. Suppose further that A is endowed with a σ-field $\mathcal{A}$ of subsets of A such that $A_o \in \mathcal{A}$. A measurable mapping

$d : (X, \mathcal{B}) \to (A, \mathcal{A})$, with the interpretation that action $d(x)$ is taken if x is observed, is called a *decision rule*.

DEFINITION 1.4.1 A decision rule d is called ignorable if $P_\theta \{d(\underline{x}) \in A_o\} < 1$ for some $\theta \in \theta$.

Thus the statistician should only consider decision rules d with $P_\theta \{d(\underline{x}) \in A_o\} = P_\theta \{L_{d(\underline{x})} \text{ identifiable } \} = 1$, $\theta \in \theta$.

In fact, any statistical problem can be considered as a special case of a statistical decision problem. To see this we take $A = \mathcal{J}_o$, the informational class of statements the statistician is interested in and for L a mapping into a set consisting of two consequences c_o and c_1 $(c_o \neq c_1)$ to be interpreted as 'true' and 'false' respectively ,

$$(1.4.1) \qquad L(\theta, \alpha) = \begin{cases} c_o & , \ \theta \in \alpha \\ \\ c_1 & , \ \theta \notin \alpha \ . \end{cases}$$

A decision rule is now a statistical procedure and the following theorem shows that then definitions 1.3.2 and 1.4.1 are equivalent.

THEOREM 1.4.2 If A is a class of statements and L is given by (1.4.1), then a decision rule d is ignorable iff it is ignorable as a statistical procedure.

PROOF. The theorem follows if we can prove: $L_{d(\underline{x})}$ identifiable a.s. iff $d(\underline{x})$ identifiable a.s. Let $\alpha \in A$ be arbitrary and suppose that L_α is identifiable. Then $L_\alpha^{-1}(\{c_o\}) = \alpha$ is identifiable. On the other hand suppose α is identifiable. Then also $L_\alpha^{-1}(\{c_o\})$ is identifiable as the complement of $\alpha = L_\alpha^{-1}(\{c_o\})$. Thus by the remark following definition 1.1.10 L_α is identifiable. □

If all actions are statistical statements, then a decision-rule is a statistical procedure and the question arises for which loss functions the definitions of ignorability are equivalent for *all* rules d. Thus we are looking for functions $L_\alpha : \Theta \to C$ such that L_α is identifiable iff α is identifiable. Necessary and sufficient for L_α to have this property is

$$L_\alpha(\theta_1) = L_\alpha(\theta_2) \iff (\theta_1 , \theta_2 \in \alpha \vee \theta_1 , \theta_2 \in \alpha^c) .$$

Thus L_α must be constant on α and on α^c.
Hence L is of the form

$$(1.4.2) \qquad L(\theta , \alpha) = \begin{cases} c_o(\alpha) & , \ \theta \in \alpha \\ c_1(\alpha) & , \ \theta \notin \alpha \end{cases} ,$$

with $c_o(\alpha) \neq c_1(\alpha)$ for all $\alpha \in A$.

EXAMPLE 1.4.3 *(Estimation)* Let $\Theta = A = \mathbb{R}$ and consider the usual quadratic loss function

$$L(\theta , \alpha) := (\theta - \alpha)^2 , \ \theta \in \Theta \quad \alpha \in A .$$

Since L_α is not constant for $\theta \neq \alpha$, this loss function is clearly not of the form (1.4.2).

EXAMPLE 1.4.4 *(Hypothesis testing)* Let $\Theta = \mathbb{R}^k$ and consider the problem of testing $H_o : \theta \in \Theta_o$ against $H_1 : \theta \in \Theta_o^c$ where Θ_o is a (measurable) subset of Θ.
Let $A := \{\Theta_o , \Theta_o^c\}$, and take

$$L(\theta , \alpha) := 1 - 1_\alpha(\theta) \qquad \alpha \in A , \ \theta \in \Theta .$$

Obviously it is of the form (1.4.2). Note, that a decision rule for this problem is not ignorable iff the statement Θ_o is identifiable.

1.5 IDENTIFICATION IN BAYESIAN INFERENCE

Let $\mathcal{A}$ be a σ-field of subsets of θ, and τ some measure on $(\theta, \mathcal{A})$, called the *prior measure*. Following KADANE [22] we shall interpret τ as an *opinion* on θ adopted by the statistician before $\underline{x}$ is observed. It is clear that subsets of θ of τ-measure zero cannot play a role anymore and that the classical concept of identification becomes rather meaningless. Furthermore, when the statistician has the opinion τ on $(\theta, \mathcal{A})$ this implies that he will consider only $\mathcal{A}$-measurable statistical statements. It should be noted that there exist identifiable statements in $\mathcal{A}$ (e.g. θ), and a good Bayesian definition of identifiability is of course such that these statements are also identifiable in the Bayesian sense. This leads to the following definition.

DEFINITION 1.5.1 A statistical statement $\theta_o \in \mathcal{A}$ is called identifiable w.r.t. τ or equivalently $\underline{x}$ is said to be informative for θ_o w.r.t. τ, if for some $N \in \mathcal{A}$ with $\tau\{N\} = 0$ and for all $\theta_1, \theta_2 \in \theta - N$ we have the implication: $\theta_1 \in \theta_o, \theta_2 \in \theta_o^c \Rightarrow P_{\theta_1} \neq P_{\theta_2}$.

REMARK. By choosing for τ a measure that assigns positive mass to all points of θ, identifiability w.r.t. τ is equivalent to identifiability. However in most practical situations this would imply that the measure τ is not σ-finite. Since such models are analytically not very attractive, the statistician shall prefer σ-finite prior measures.

Obviously all statements of τ-measure 0 and their complements are identifiable w.r.t. τ and the analogue of theorem 1.1.3 is that the class of all statements which are identifiable w.r.t. τ is a sub σ-field of $\mathcal{A}$.

Since functions of θ are now functions on the measurable space $(\theta, \mathcal{A})$, it will be clear that the only functions φ the statistician can be interested in, are measurable functions. Let $(\Lambda, \mathcal{L})$ be some measurable space, where $\mathcal{L}$ is a σ-field of subsets of Λ.

DEFINITION 1.5.2 The measurable function $\varphi : (\Theta, \mathcal{A}) \to (\Lambda, \mathcal{L})$ is called identifiable w.r.t. τ, or equivalently $\underline{x}$ is said to be informative for φ w.r.t. τ if $\underline{x}$ is informative for $\varphi^{-1}(\Lambda_o)$ w.r.t. τ for all $\Lambda_o \in \mathcal{L}$.

Although definition 1.5.2 is the obvious generalization of def. 1.1.10 it gives rise to a remarkable difference between classical and Bayesian theory since the obvious analogue of lemma 1.1.12 does not hold. More precisely we have

LEMMA 1.5.3 If there exists a set $N \in \mathcal{A}$ with $\tau\{N\} = 0$ such that for all $\theta_1, \theta_2 \in \Theta - N$ we have the implication $\varphi(\theta_1) \neq \varphi(\theta_2) \Rightarrow P_{\theta_1} \neq P_{\theta_2}$, then $\underline{x}$ is informative for φ w.r.t. τ.

PROOF. Let Λ_o be arbitrary with $\tau\{\varphi^{-1}(\Lambda_o)\} > 0$ and $\tau\{\varphi^{-1}(\Lambda_o^c)\} > 0$. (If $\varphi^{-1}(\Lambda_o)$ or $\varphi^{-1}(\Lambda_o^c)$ has τ-measure zero, nothing remains to prove).
Let $\theta_1 \in \varphi^{-1}(\Lambda_o) - N$ and $\theta_2 \in \varphi^{-1}(\Lambda_o^c) - N$. Then $\varphi(\theta_1) \neq \varphi(\theta_2)$ since $\varphi(\theta_1) \in \Lambda_o$ and $\varphi(\theta_2) \in \Lambda_o^c$. Hence $P_{\theta_1} \neq P_{\theta_2}$ and so $\underline{x}$ is informative for $\varphi^{-1}(\Lambda_o)$ w.r.t. τ. □

That the converse does not hold in general can be seen as follows. Let Λ be a non-denumerable set and $\mathcal{L}$ a σ-field of subsets of Λ that contains all one point subsets. Suppose the measurable function $\varphi : (\Theta, \mathcal{A}) \to (\Lambda, \mathcal{L})$ is identifiable w.r.t. some prior measure τ.
Let $\lambda \in \Lambda$ be arbitrary and let $N_\lambda \in \mathcal{A}$ be a null set such that for all $\theta_1, \theta_2 \in \Theta - N_\lambda$ the implication $\theta_1 \in \varphi^{-1}(\{\lambda\})$, $\theta_2 \in (\varphi^{-1}(\{\lambda\}))^c \Rightarrow P_{\theta_1} \neq P_{\theta_2}$ holds. Then a set N, such that for all $\theta_1, \theta_2 \in \Theta - N$ we have $\varphi(\theta_1) \neq \varphi(\theta_2) \Rightarrow P_{\theta_1} \neq P_{\theta_2}$, is equal to $\bigcup_{\lambda \in \Lambda} N_\lambda$.

Since Λ was non-denumerable, N need not be a null set, or may

even be nonmeasurable.

In order to obtain the equivalence in lemma 1.5.3 we need a stronger concept of identifiability of functions w.r.t. τ.

DEFINITION 1.5.4 The measurable function $\varphi : (\Theta, \mathcal{A}) \to (\Lambda, \mathcal{L})$ is called *uniformly identifiable w.r.t.* τ, or equivalently $\underline{x}$ is said to be *uniformly informative* for φ w.r.t. τ, if there exists a set $N \in \mathcal{A}$ with $\tau\{N\} = 0$ such that for all $\Lambda_o \in \mathcal{L}$ and $\theta_1, \theta_2 \in \Theta - N$ we have the implication $\theta_1 \in \varphi^{-1}(\Lambda_o)$, $\theta_2 \in \varphi^{-1}(\Lambda_o^c) \Rightarrow P_{\theta_1} \neq P_{\theta_2}$.

The difference between definitions 1.5.2 and 1.5.4. is that in the former definition the set N may depend on Λ_o. It is now easily seen that the function φ is uniformly identifiable w.r.t. τ iff there exists a null set $N \in \mathcal{A}$ such that for all $\theta_1, \theta_2 \in \Theta - N$ we have $\varphi(\theta_1) \neq \varphi(\theta_2) \Rightarrow P_{\theta_1} \neq P_{\theta_2}$.

Although the concept of identification introduced in definitions 1.5.1 and 1.5.2 is different from that in KADANE [22] (who in fact uses the classical concept) the next theorem will show that the concept of identification introduced in this section is just what we intuitively want it to be.

For terminological convenience we shall restrict ourselves to the case where τ is a probability measure. It is then called the *prior distribution* and the identity on $(\Theta, \mathcal{A})$ can be considered as a random object $\underline{\theta}$. To avoid problems with conditional probabilities we shall suppose that $\underline{\theta}$ is a random vector and that $\underline{x}$ takes its values in a complete separable metric space. Suppose further that for all $X_o \in \mathcal{B}$ the function $\Psi : (\Theta, \mathcal{A}) \to ([0,1], \mathcal{B}_o)$ where $\mathcal{B}_o$ is the Borelfield on $[0,1]$, is defined by $\Psi(\theta) := P_\theta\{X_o\}$ and is $\mathcal{A}$-measurable. (For details on conditional probabilities see ASH [2] p. 262 - 265) Then P_θ can be interpreted as conditional distribution of $\underline{x}$ given θ and we may write down the *posterior distribution* τ_x, the distribution of $\underline{\theta}$ conditional on x. It is interpreted as the opinion on θ after $\underline{x}$ has been observed, and roughly speaking one could expect the opinion on θ (non degenerate) to be

changed with positive probability if $\underline{x}$ is informative for $\underline{\theta}$ w.r.t. τ. Let P denote the joint distribution of $\underline{x}$ and $\underline{\theta}$. We have

THEOREM 1.5.5 If $0 < \tau\{\theta_o\} < 1$ for some $\theta_o \in \mathcal{A}$ and if $\underline{x}$ is informative for θ_o w.r.t. τ then $P\{\tau_{\underline{x}} \neq \tau\} > 0$.

PROOF. We have $\tau_x = \tau$ a.e. (x) iff the random variables $\underline{x}$ and $\underline{\theta}$ are independent or equivalently $P_\theta = F$ a. e. (θ) where F denotes the marginal distribution of $\underline{x}$. Since $\underline{x}$ is informative for θ_o w.r.t. τ, there exists a null-set $N \in \mathcal{A}$ such that $\theta_1 \in \theta_o - N$, $\theta_2 \in \theta_o^c - N$ implies $P_{\theta_1} \neq P_{\theta_2}$. Since both $\theta_o - N$ and $\theta_o^c - N$ have positive τ-measure we cannot have $P_\theta = F$ a.e. (θ). Hence $\{ x \mid \tau_x \neq \tau\}$ must have positive probability (unconditional to θ). □

REMARK 1. If all $\theta_o \in \mathcal{A}$ have τ-measure 0 or 1, $\underline{\theta}$ is identifiable w.r.t. τ regardless what P_θ is. Of course such an opinion τ is not a very realistic one.

REMARK 2. Conversely, if $P\{\tau_{\underline{x}} \neq \tau\} > 0$ then $\underline{x}$ and $\underline{\theta}$ are not independent and so there exist $\theta_o, \theta_1 \in \theta$ with $P_{\theta_o} \neq P_{\theta_1}$. Hence there exist nontrivial identifiable statements (i.e. statements $\theta_o \subset \theta$ with $\theta_o \neq \emptyset$ and $\theta_o^c \neq \emptyset$). It is however not clear whether they are $\mathcal{A}$-measurable.

1.6 FINITE INFORMATIVE SAMPLES FROM STOCHASTIC PROCESSES AND THE PROBLEM OF MINIMUM INFORMATIVE SAMPLE SIZE

Let $\{\underline{x}_t ; t = 1, 2, \ldots\}$ be a (possibly vector-valued) observable stochastic process in discrete time. If the distribution of the process is characterized by some unknown parameter $\theta \in \theta$ we put $\mathcal{P}^{(\infty)} = \{P_\theta^{(\infty)} \mid \theta \in \theta\}$ where $P_\theta^{(\infty)}$ denotes the (infinite dimensional) distribution of the process $\{\underline{x}_t\}$.

Suppose we observe the process $\{ \underline{x}_t \}$ at $t = 1, 2, \ldots, n$. Then our actual sample is $(\underline{x}_1 , \underline{x}_2 , \ldots \underline{x}_n)$. The joint distribution $P_\theta^{(n)}$ of this sample is a marginal distribution of $P^{(\infty)}$. Let $\mathcal{P}^{(n)} = \{ P_\theta^{(n)} \mid \theta \in \Theta \}$.

It will be clear now, that inferences about a function $\varphi : \Theta \to \Lambda$, based on the sample $(\underline{x}_1 , \underline{x}_2 , \ldots \underline{x}_n)$ only make sense if φ is identifiable w.r.t. $\mathcal{P}^{(n)}$ rather than w.r.t. $\mathcal{P}^{(\infty)}$. Therefore the *sample* must be informative for φ. Obviously identifiability with respect to $\mathcal{P}^{(n)}$ implies identifiability with respect to $\mathcal{P}^{(\infty)}$, but the converse does not hold in general. Although it follows from KOLMOGOROVS extension theorem that $P_\theta^{(\infty)}$ is determined by the sequence $P_\theta^{(n)}$, $n = 1, 2, \ldots$ there need not exist some finite sequence $P_\theta^{(1)}, \ldots, P_\theta^{(N)}$ that determines $P_\theta^{(\infty)}$ uniquely for all $\theta \in \Theta$. To be more precise, let $\theta_1 , \theta_2 \in \Theta$ be such that $P_{\theta_1}^{(\infty)} \neq P_{\theta_2}^{(\infty)}$ and let $N = N(\theta_1 , \theta_2)$ be the smallest value of n such that $P_{\theta_1}^{(n)} \neq P_{\theta_2}^{(n)}$. Since N need not be bounded in θ_1 and θ_2, it will be clear that in general identification w.r.t. $\mathcal{P}^{(\infty)}$ is an essentially weaker concept than identification w.r.t. $\mathcal{P}^{(n)}$ for some finite n. This fact is important because e.g. in estimation theory it implies that the existence of consistent estimators for $\varphi(\theta)$ does not guarantee a finite sample to be informative for $\varphi(\theta)$. The following example may clarify this.

EXAMPLE 1.6.1 Let $\underline{x}_1 , \underline{x}_2 , \ldots$ be a sequence of independent Bernoulli trials with $P\{ \underline{x}_t = 1 \} = \theta_t$, $P\{ \underline{x}_t = 0 \} = 1 - \theta_t$, $0 \leq \theta_t \leq 1$ such that the sequence $\theta = (\theta_1 , \theta_2 , \ldots)$ has a well defined Césaro limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \theta_t := \varphi(\theta) .$$

Suppose the θ_t are unknown, and the statistician wants to estimate $\varphi(\theta)$. The natural choice for θ is

$$\Theta = \{ \theta \mid \theta = (\theta_1 , \theta_2 \ldots), \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \theta_t = \varphi(\theta) \} .$$

The obvious estimator based on the sample $\underline{x}_1 , \underline{x}_2 \ldots \underline{x}_n$ is

$$\hat{\underline{\varphi}}_n := \frac{1}{n} \sum_{t=1}^{n} \underline{x}_t .$$

It is easily seen to be asymptotically unbiased. Furthermore, for its variance we have

$$V\{\hat{\underline{\varphi}}_n\} = \frac{1}{n^2} \sum_{t=1}^{n} \theta_t (1 - \theta_t) \leq \frac{1}{4n} \to 0, \quad n \to \infty.$$

Thus $\hat{\underline{\varphi}}_n$ is a (weakly) consistent estimator for $\varphi(\theta)$. Nevertheless it is easily seen that no finite sample is informative for $\varphi(\theta)$.

The previous discussion motivates the following definition.

DEFINITION 1.6.2 The sample size n is called informative for the function $\varphi : \Theta \to \Lambda$ if for all $\theta_1 , \theta_2 \in \Theta$ we have the implication $\varphi(\theta_1) \neq \varphi(\theta_2) \Rightarrow P_{\theta_1}^{(n)} \neq P_{\theta_2}^{(n)}$.

REMARK. If the proces $\{ \underline{x}_t \}$ is strictly stationary, and the sample size n is informative for φ, then any sample of the process taken at n consecutive time points is informative for φ.

If the sample size n is informative for φ, then any sample size $N \geq n$ is informative for φ because of the implication $P_{\theta_1}^{(n)} \neq P_{\theta_2}^{(n)} \Rightarrow P_{\theta_1}^{(n+1)} \neq P_{\theta_2}^{(n+1)}$, $\theta_1 , \theta_2 \in \Theta$. Therefore an interesting but often very difficult problem is to find the *minimum informative sample size*, or at least an upperbound for it. In the case of a sample from a moving-average process the minimum can be found and in mixed autoregressive moving average models an upperbound can be found (Chapter II univariate, and Chapter III multivariate).

1.7 IDENTIFICATION AND PREDICTION

In this section we consider the *statistical prediction problem*, that is the problem of predicting the random variable (or - vector) $\underline{y}$ on the basis of the observable vector $\underline{x}$ where the joint distribution of $\underline{x}$ and $\underline{y}$ depends on some unknown parameter $\theta \in \Theta$. The problem needs some special attention since it does not fit the framework of § 1.1 - § 1.3 in a trivial way. Let P_θ denote the *joint* distribution of $\underline{x}$ and $\underline{y}$ and put $\mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$. Then the triple $((\underline{x} , \underline{y}) , \mathcal{P} , \Theta)$ is not a relevant statistical problem since only $\underline{x}$ is observable and not $(\underline{x} , \underline{y})$. The relevant statistical problem is represented by $(\underline{x} , \mathcal{P}_1 , \Theta)$ where $\mathcal{P}_1 = \{ P_{1,\theta} \mid \theta \in \Theta \}$ and $P_{1,\theta}$ denotes the marginal distribution of $\underline{x}$.
If the statistician is interested in prediction of $\underline{y}$ on the basis of $\underline{x}$,he is, in fact, interested in statements in terms of the mapping $\Psi : \Theta \to \mathcal{P}$ defined by $\Psi(\theta) = P_\theta$, $\theta \in \Theta$. This leads to the following definition.

DEFINITION 1.7.1 $\underline{y}$ is said to be *predictable* w.r.t. $\underline{x}$ if $\underline{x}$ is informative for the mapping Ψ.

Thus $\underline{y}$ is predictable w.r.t. $\underline{x}$ iff we have the following implication for all $\theta_1 , \theta_2 \in \Theta$: $P_{\theta_1} \neq P_{\theta_2} \Rightarrow P_{1,\theta_1} \neq P_{1,\theta_2}$ (see lemma 1.1.12).
Two examples may clarify the ideas.

EXAMPLE 1.7.2 Let $\underline{x}$ and $\underline{y}$ be independent normal variables with unknown mean θ and unit variances. Then for all $\theta_1 , \theta_2 \in \Theta$ we have the implications $P_{\theta_1} \neq P_{\theta_2} \Rightarrow \theta_1 \neq \theta_2 \Rightarrow P_{1,\theta_1} \neq P_{1,\theta_2}$.
Hence $\underline{y}$ is predictable w.r.t. x.

EXAMPLE 1.7.3 Let $(\underline{x} , \underline{y})$ be binormal with zero means, unit variances and unknown correlation-coefficient θ. Then we have $P_{1,\theta_1} = P_{1,\theta_2}$ for all $\theta_1 , \theta_2 \in \Theta = [-1 , 1]$ and so $\underline{y}$

is not predictable w.r.t. $\underline{x}$. It should be noted, that if θ were known, then $\underline{y}$ would be predictable in a trivial way. However, in that case we don't have a real *statistical* prediction problem, but merely a *probabilistic* prediction problem which can be defined as a triple $((\underline{x} , \underline{y}) , \mathcal{P} , \theta)$ for which the conditional distribution of $\underline{y}$ given $\underline{x}$ does not depend on θ (a.s. $P_{1,\theta} \ \forall_\theta$). Clearly observational equivalence of values of θ is irrelevant for such prediction problems.

Let $\underline{x} = (\underline{x}_1 , \underline{x}_2 \ldots , \underline{x}_n)$ be a sample taken from a stochastic process $\{\underline{x}_t , t = 1, 2, \ldots\}$ with probability law $P_\theta^{(\infty)}$, $\theta \in \Theta$. Let $N \geq n+1$. Then $\underline{y} = \underline{x}_N$ is predictable if the sample size n is informative for $P_\theta^{(N)}$ (def. 1.6.2) or equivalently if for all $\theta_1 , \theta_2 \in \Theta$ we have the implication

$$(1.7.1) \qquad P_{\theta_1}^{(N)} \neq P_{\theta_2}^{(N)} \Rightarrow P_{\theta_1}^{(n)} \neq P_{\theta_2}^{(n)} .$$

Usually the index t represents time; therefore $\{\underline{x}_{n+1} , \underline{x}_{n+2}, \ldots\}$ is called the *future* of the process. Thus the future is predictable iff (1.7.1) holds for all $N \geq n+1$. In analogy with def. 1.6.2 we put

DEFINITION 1.7.4 The sample size n is called *predictive* if (1.7.1) holds for all $N \geq n+1$.

Of course, if the sample size n is predictive, any sample size $m \geq n$ is predictive and so the problem of the *minimum predictive sample size* makes sense. The problem is closely related to that of minimum informative sample sizes. It differs so far as in (1.7.1) the set $\{ P_\theta^{(N)} \mid N = n+1 , n+2 , \ldots \}$ depends on n. The relation to the minimum informative sample size (if it exists) is given in the next theorem.

THEOREM 1.7.5 If N_φ is the minimum informative sample size for $\varphi : \Theta \to \Lambda$ and n_o is the minimum predictive sample size

then we have $n_o \geq N_\varphi$. If φ is 1 - 1 then $n_o = N_\varphi$.

PROOF. Since N_φ is minimal informative for φ we have for $\theta_1 , \theta_2 \in \theta$

$$\varphi(\theta_1) \neq \varphi(\theta_2) \Rightarrow P_{\theta_1}^{(N_\varphi)} \neq P_{\theta_2}^{(N_\varphi)} \neq P_{\theta_1}^{(N_\varphi-1)} \neq P_{\theta_2}^{(N_\varphi-1)} ,$$

hence $n_o \geq N_\varphi$.

If φ is 1 - 1 we also have for all $\theta_1 , \theta_2 \in \theta$

$$P_{\theta_1}^{(N_\varphi+1)} \neq P_{\theta_2}^{(N_\varphi+1)} \Rightarrow \theta_1 \neq \theta_2 \iff \varphi(\theta_1) \neq \varphi(\theta_2) ,$$

which implies that N_φ is predictive and so $N_\varphi = n_o$. □

EXAMPLE 1.7.6 Consider the linear regression model of example 1.2.10. We rewrite it as

$$\underline{Y}_t = x_t' \beta + \underline{\varepsilon}_t, \quad t = 1, 2, \ldots , \qquad x_t \in \mathbb{R}^k .$$

Suppose the $\underline{\varepsilon}_t$ are i.i.d variables with common distribution-function F_ξ , $\xi \in V$. We shall prove that $\underline{Y} := (\underline{Y}_1 \ldots \underline{Y}_n)'$ is predictive for $\underline{Y}_{n+1}$ iff $\underline{Y}$ is informative for $x_{n+1}' \beta$.

PROOF. *(Only if)* Let $x_{n+1}' \beta_1 \neq x_{n+1}' \beta_2$. Then we have also $P_{\theta_1}^{(n+1)} \neq P_{\theta_2}^{(n+1)}$ and since $\underline{Y}$ is predictive for $\underline{Y}_{n+1}$ this implies $P_{\theta_1}^{(n)} \neq P_{\theta_2}^{(n)}$. Hence $\underline{Y}$ is informative for $x_{n+1}' \beta$ (lemma 1.1.12).

(If) Let $P_{\theta_1}^{(n+1)} \neq P_{\theta_2}^{(n+1)}$. Assume $P_{\theta_1}^{(n)} = P_{\theta_2}^{(n)}$. Then we have $F_{\xi_1} = F_{\xi_2}$ and $X_n \beta_1 = X_n \beta_2$ where X_n denotes the $n \times k$ matrix of regressors. Since $P_\theta^{(n+1)}$ is completely determined by F_ξ and $X_{n+1} \beta$, where

$$X_{n+1} := \begin{bmatrix} X_n \\ \\ x'_{n+1} \end{bmatrix},$$

we must have $X_{n+1}\,\beta_1 \neq X_{n+1}\,\beta_2$ and so $x'_{n+1}\,\beta_1 \neq x'_{n+1}\,\beta_2$. But this implies $P_{\theta_1}^{(n)} \neq P_{\theta_2}^{(n)}$ since $\underline{y}$ is informative for $x'_{n+1}\,\beta$. Thus we have a contradiction proving that $\underline{y}$ is predictive for $\underline{y}_{n+1}$. □

1.8 WEAK CONCEPTS OF OBSERVATIONAL EQUIVALENCE AND STRONGLY INFORMATIVE SAMPLES

The concept of identification introduced in the preceding sections is in fact based on the classical concept of observational equivalence, that is on *equality of distributions*. As SCHÖNFELD [31] allready pointed out, any other equivalence relation on $\mathcal{P}$ can serve as a basis of a (weaker) alternative concept of observational equivalence, and so a *stronger* notion of informativeness. Let $\sim$ denote an arbitrary equivalence relation on $\mathcal{P}$. Then the values $\theta_1 \in \theta$ and $\theta_2 \in \theta$ are called weakly $\sim$-observational equivalent if $P_{\theta_1} \sim P_{\theta_2}$. It is easily seen, that all results of the preceding sections remain valid if (in)-equality of distributions is replaced by (non-) $\sim$ equivalence.

We shall consider two possibilities for $\sim$, where the first one (M_r-equivalence) is the most important (particularly from a practical point of view) and the second (t-equivalence) is of some theoretic importance since it enables us to see a link between the classic concept of identification and *sufficient statistics*.

Let $M_r(\theta)$ denote the set of all moments up to order r of the distribution P_θ (r = 1, 2, ...).

DEFINITION 1.8.1 If $M_r(\theta_1) = M_r(\theta_2)$ then the distributions P_{θ_1} and P_{θ_2} are said to be *M_r-equivalent* : $P_{\theta_1} \overset{M_r}{\sim} P_{\theta_2}$. The values θ_1 and θ_2 are then said to be *weakly observational equivalent to the order r*. The corresponding concept of informativeness (identifiability) is called *strong r-th order informativeness (strong r-th order identifiability)*

Although the corresponding definitions of (minimum) strongly r-th order informative sample size and strong r-th order predictability are obvious, as an example we shall give the definition of r-th order predictive sample sizes. Let $M_r^{(n)}(\theta)$ denote the set of all moments up to order r of the sample $\underline{x} = (\underline{x}_1 , \underline{x}_2 \cdots \underline{x}_n)$ taken from the process $\{\underline{x}_t , t = 1, 2, \ldots\}$. Then the sample size n is called *strongly r-th order predictive* if for all $N \geq n+1$ and $\theta_1 , \theta_2 \in \theta$ we have the implication

$$M_r^{(N)}(\theta_1) \neq M_r^{(N)}(\theta_2) \Rightarrow M_r^{(n)}(\theta_1) \neq M_r^{(n)}(\theta_2) .$$

Clearly if we are studying weakly stationary processes and regression models, then r = 2 is the most important case. Of course, if the observable random vector is *normal*, informativeness is equivalent to strong second order informativeness. The main advantage however of using definition 1.8.1 rather than the classical concept is, that it is not necessary to give a complete specification of the class $\mathcal{P}$.

EXAMPLE 1.8.2 (Continuation of example 1.2.10)
Clearly the function $\mu(\theta) := X\beta = E_\theta\{ \underline{Y} \}$ is strongly first-order identifiable regardless the distribution of $\underline{\varepsilon}$. Furthermore if $\varphi(\theta) := (\beta , Q)$ where $Q = Q(\theta) := E_\theta\{ \underline{\varepsilon}\, \underline{\varepsilon}^* \}$, then a sufficient condition for φ to be strongly second-order identifiable is $r(X) = k$. If $\theta = \mathbb{R}^k \times V$ then the condition is also necessary.

In the next section we shall develop the concept of

t-informativeness. Let $(T, \mathcal{T})$ be a measurable space, and let $t : (X, \mathcal{B}) \to (T, \mathcal{T})$ be a measurable function. Then t induces the probability distribution $P_\theta\, t^{-1}$ on $(T, \mathcal{T})$. The random quantity $t(\underline{x})$ is called a *statistic* and is interpreted as a *reduction* of the observational material.

DEFINITION 1.8.3 If $P_{\theta_1} t^{-1} = P_{\theta_2} t^{-1}$ then the distributions P_{θ_1} and P_{θ_2} are said to be *t - equivalent*: $P_{\theta_1} \overset{t}{\sim} P_{\theta_2}$. The values θ_1 and θ_2 are then said to be *t - observationally equivalent*. The corresponding concept of informativeness is called *t - informativeness*.

The function t transforms the statistical problem $(\underline{x}, \mathcal{P}, \theta)$ into the new statistical problem $(t(\underline{x}), \mathring{\mathcal{P}}, \theta)$ where

$$\mathring{\mathcal{P}} = \{ \mathring{P}_\theta \mid \mathring{P}_\theta = P_\theta\, t^{-1}, \theta \in \theta \}.$$

THEOREM 1.8.4 $\underline{x}$ is t - informative for $\varphi : \theta \to A$ iff $t(\underline{x})$ is informative for φ.

PROOF. Follows immediately from def. 1.8.3 and lemma 1.1.12.□

Obviously we have for all $\theta_1, \theta_2 \in \theta$

$$(1.8.1) \qquad P_{\theta_1} t^{-1} \neq P_{\theta_2} t^{-1} \Rightarrow P_{\theta_1} \neq P_{\theta_2} .$$

Thus t - informativeness is stronger than informativeness. Of course, if t is 1 - 1 the concepts are identical but this is not a necessary condition, and the question arises for what functions t the equivalence holds. In general the answer is not very interesting as the following example may show.

EXAMPLE 1.8.5 Let $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_n)$ be a random sample from $N(0, \theta)$, $\theta \in (0, \infty)$ and $t(\underline{x}) := \underline{x}_1$. Then equivalence

holds in a trivial way.

The point is, that statistics like $t(\underline{x})$ in the above example ignore a lot of information. *Sufficient* statistics, however, are not 1 - 1 in general but don't ignore any relevant information. We have

THEOREM 1.8.6 Let $t(\underline{x})$ be sufficient for θ. Then $\underline{x}$ is informative for θ iff $t(\underline{x})$ is informative for θ.

PROOF. Let $\tilde{P}_\theta$ denote the joint distribution of $\underline{x}$ and $t(\underline{x})$. Then for all θ_1 , $\theta_2 \in \theta$ we have $P_{\theta_1} \neq P_{\theta_2} \Rightarrow \tilde{P}_{\theta_1} \neq \tilde{P}_{\theta_2}$. Since $t(\underline{x})$ is sufficient, the conditional distribution of $\underline{x}$ given $t(x)$ does not depend on θ and since $\tilde{P}_\theta$ is completely determined by this conditional distribution and $P_\theta\ t^{-1}$ (the marginal distribution of $t(\underline{x})$). We also have

$$P_{\theta_1} \neq P_{\theta_2} \Rightarrow P_{\theta_1}\ t^{-1} \neq P_{\theta_2}\ t^{-1} \quad ,$$

for all θ_1 , $\theta_2 \in \theta$. Combination with (1.8.1) completes the proof. □

1.9 LOCAL IDENTIFIABILITY

So far we considered the identification problem from the point of view that a fixed parameterspace θ is given. Therefore there was no need to refer to it explicitly. However sometimes it is convenient to consider the problem for subsets $\mathring{\theta}$ of θ. Since the identifiability then depends on the specific choice $\mathring{\theta} \subset \theta$, in this section we shall refer to it by saying that a statement (or a function) is identifiable w.r.t. $\mathring{\theta}$, or equivalently that $\underline{x}$ is informative w.r.t. $\mathring{\theta}$ for that statement (or that function). Thus e.g. a statement $\theta_o \subset \mathring{\theta}$ is identifiable w.r.t. $\mathring{\theta}$ if for all θ_o , $\theta_1 \in \mathring{\theta}$ we have

$\theta_o \in \theta_o$, $\theta_1 \in \mathring{\theta} - \theta_o \Rightarrow P_{\theta_o} \neq P_{\theta_1}$.

If a topology is given on Θ, we have the possibility of considering the identifiabilityproblem in a neighbourhood of a given point. Let the statisticican be interested in the function $\varphi : \Theta \to \Lambda$, where the space Λ is endowed with a topology τ. Without loss of generality we can assume that φ is onto.

DEFINITION 1.9.1 The function φ is said to be *locally identifiable* for φ at $\lambda_o \in \Lambda$, or equivalently $\underline{x}$ is called *locally informative* for φ at λ_o if there exists a neigbourhood $U_{\lambda_o} \in \tau$ of λ_o such that φ is identifiable w.r.t. $\varphi^{-1}(U_{\lambda_o})$.
The function φ is said to be locally identifiable on $\Lambda_o \subset \Lambda$ if it is locally identifiable at all $\lambda_o \in \Lambda_o$.

REMARK. Local identifiability in all points $\lambda \in \Lambda$ does not imply identifiability in the sense of def. 1.1.10.

Obviously local identifiability is a weaker concept of identifiability. Therefore we shall only use this concept in situations where non-identifiability can be sharpened to non-local identifiability or even *nowhere local identifiability* on some set $\Lambda_o \subset \Lambda$. We shall shortly say that φ is nowhere locally identifiable if it is nowhere locally identifiable on Λ. The following lemma is helpful .

LEMMA 1.9.2 Let $D \subset \Lambda$ be a set that is dense in Λ and let $\varphi : \Theta \to \Lambda$ be nowhere locally identifiable on D. Then φ is nowhere locally identifiable.

PROOF. Suppose φ is locally identifiable in $\lambda_o \in \Lambda$. Then there exists a neighbourhood $U_{\lambda_o} \in \tau$ of λ_o such that for all $\lambda_1 , \lambda_2 \in U_{\lambda_o}$ and $\theta_1 \in \varphi^{-1}\{\lambda_1\}$, $\theta_2 \in \varphi^{-1}\{\lambda_2\}$ we have $\lambda_1 \neq \lambda_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$. However, since D is dense in Λ there exists a $d_o \in D$ with $d_o \in U_{\lambda_o}$. But then φ is locally identifia-

ble at d_o which contradicts the fact that φ is nowhere locally identifiable on D. □

By replacing P_θ by $M_r(\theta)$ (the set of all moments up to order r of P_θ) we obtain the definition of *local r^{th}-order identifiability* and *local r^{th}-order informativeness*. We could also define concepts as locally informative sample size etc. but since we will not use all of these concepts, we shall not do so here.

The concepts introduced here, will be used to obtain strong results in problems of minimum informative sample size for moving average processes. The important concept then is local second-order informative sample size.

CHAPTER II

INFORMATIVE SAMPLE SIZES IN UNIVARIATE STATIONARY MODELS

2.1 INTRODUCTION AND PRELIMINARY RESULTS

In this chapter we shall derive informative and predictive sample sizes for a class of weakly stationary univariate processes $\{ \underline{x}_t , t = 1, 2, \ldots \}$. For reasons of mathematical elegance all processes are allowed to be complex valued, and are defined on $t = 0, \pm 1, \pm 2, \ldots$ (Every weakly stationary process $\{\underline{x}_t , t = 1, 2, \ldots \}$ can be extended to a weakly stationary process on $\ldots -1, 0, 1, 2, \ldots$).
We are interested in weakly stationary observable processes $\{ \underline{x}_t \}$ that satisfy a linear stochastic difference equation of the form

$$(2.1.1) \qquad \sum_{k=o}^{p} a_k \, \underline{x}_{t-k} = \sum_{m=o}^{q} b_m \, \underline{\varepsilon}_{t-m} \quad , \quad t = 0, \pm 1, \pm 2, \ldots$$

with $p, q \geq o$ and $a_o = b_o = 1$, and where $\{ \underline{\varepsilon}_t, t = 0, \pm 1, \ldots\}$ is a (non-observable) *white noise* process ,
(i.e $E \{ \underline{\varepsilon}_t \, \overline{\underline{\varepsilon}}_s \} = \delta_{t-s} \, \sigma^2, \quad t, s = 0, \pm 1, \ldots , \sigma^2 > 0$)
with $E \{ \underline{\varepsilon}_t \} = 0$ for all t.
The integers p and q are supposed to be known.
If $p = 0 , q \geq 1$, then $\{ \underline{x}_t\}$ is called a *moving average* process of order q, if $p \geq 1, q = 0$, an *autoregressive* process of order p and if $p \, q \geq 1$, a *mixed autoregressive moving average* process of order (p, q). The usual abbreviations are MA(q), AR(p) and ARMA(p, q) respectively. The (complex) coefficients $a_1, a_2 \ldots a_p , b_1 \ldots b_q$ and σ^2 are supposed to be unknown.
If $\xi \in Z$ is a parameter that characterizes the distribution of the process $\{ \underline{\varepsilon}_t \}$, then we write $\sigma^2 = \sigma^2(\xi)$ and put $\Theta := (a_1, a_2 \ldots a_p, b_1 \ldots b_q , \xi)$. In order to describe the parameter space Θ and the spectral measure F_Θ of $\{ \underline{x}_t \}$ we

introduce the generating functions

$$A(z) := \sum_{k=0}^{p} a_k z^k \quad , \quad B(z) := \sum_{m=0}^{q} b_m z^m \quad , \quad z \in \mathbb{C} \ .$$

To ensure the existence of a unique weakly stationary solution $\{ \underline{x}_t \}$ of (2.1.1) we must restrict the range of $(a_1, \ldots, a_p)$ to $S_p \subset \mathbb{C}^p$ with $S_p := \{ (a_1, \ldots, a_p) \mid A(z) \neq 0 \ , \ |z| = 1 \}$. (see appendix theorem A.1.4). Thus we must choose θ such that $\theta \subset S_p \times \mathbb{C}^q \times Z$. Models of this type are frequently used, and particularly in *econometrics*, where often only small samples are available, the question of informative sample sizes is interesting. The function of interest in practice is $\Psi(\theta) := (a_1, \ldots a_p, b_1 \ldots b_q \ , \sigma^2)$. As $\underline{x}_t$ is linearly expressible in the process $\{ \underline{\varepsilon}_t \}$ (appendix theorem A.1.4), it follows that $E_\theta \{ \underline{x}_t \} = 0 \quad , \theta \in \theta$. Furthermore we don't make any assumptions about the type of distributions P_θ and therefore we restrict our attention to the second order properties of the process, that is to the *covariance function* $\gamma_s = \gamma_s(\theta) := E_\theta \{ \underline{x}_t \bar{\underline{x}}_{t-s} \}$ $t,s = 0, \pm 1, \ldots$ Thus the "natural" concept is second order strong informativeness. From now on we shall omit the word "strong(ly)".
Note that $\gamma_{-s} = \bar{\gamma}_s$, $s = 0, 1, 2, \ldots$ Because of the 1 - 1 correspondence between covariance functions and spectral measures (concentrated on $(-\pi \ , \ \pi]$) of weakly stationary processes (see [11] p. 633) we could equivalently look at the spectral measure F_θ of $\{ \underline{x}_t \}$. It is known to be absolutely continuous with density

$$(2.1.2) \qquad f_\theta(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2 \quad , \quad \lambda \in (-\pi \ , \ \pi]$$

(see appendix, the discussion following theorem A.1.4).
Thus if $R_\theta^{(n)}$ denotes the covariance matrix of $\underline{x}_1 \ , \ \underline{x}_2 \ , \ \ldots \ \underline{x}_n$, for all $\theta_1, \theta_2 \in \theta$ we have the implication

(2.1.3) $$R_{\Theta_1}^{(n)} \neq R_{\Theta_2}^{(n)} \Rightarrow F_{\Theta_1} \neq F_{\Theta_2} \quad , \qquad n = 1, 2, \ldots$$

Thus a necessary condition for some sample size to be second-order informative for $\psi(\theta)$ is that different Ψ - values imply different spectral measures. Since the spectral measure depends on θ only through the absolute values of the polynomials $A(z)$ and $B(z)$ on the unit circle it is not surprising that the following lemma is fundamental.

LEMMA 2.1.1 Let the function $P(z)$ be analytic on $|z| < \rho$ for a $\rho > 1$, and $P(0) \neq 0$. Then there exists a function $Q(z)$, analytic on $|z| < \rho$ satisfying

a) $Q(z) \neq 0 \quad , \qquad |z| < 1$

b) $|Q(e^{-i\lambda})| = |P(e^{-i\lambda})| \quad , \quad \lambda \in (-\pi, \pi]$.

Furthermore, $Q(z)$ is uniquely determined except for a multiplicative constant of modulus one. In the special case that $P(z)$ is a polynomial of degree p, the function $Q(z)$ also is a polynomial of degree p.

PROOF. Existence (see also RIESZ [28]) If $P(z)$ has no zeros in $|z| < 1$, we can take $Q(z) = P(z)$. If $0 < |z_o| < 1$ is a zero of $P(z)$ then we write $P(z) = (z - z_o) P_1(z)$ with $P_1(z)$ analytic on $|z| < \rho$. Let $Q_1(z) := (z \bar{z}_o - 1) P_1(z)$. Then, counting a zero of multiplicity n_o as n_o zeros, $Q_1(z)$ has one zero less inside the unit circle than $P(z)$ and since $|e^{-i\lambda} \bar{z}_o - 1| = |\bar{z}_o - e^{i\lambda}| = |z_o - e^{-i\lambda}|$, we have

$$|Q_1(e^{-i\lambda})| = |e^{-i\lambda} \bar{z}_o - 1| \, |P_1(e^{-i\lambda})| =$$

$$= |e^{-i\lambda} - z_o| \, |P_1(e^{-i\lambda})| = |P(e^{-i\lambda})| \text{, so that}$$

$Q_1(z)$ satisfies condition b). The procedure can be repeated for every zero of $P(z)$ inside the unit circle and since there are only a finite number of them we can find in finitely many steps an analytic function $Q(z)$ satisfying a) and b). It

follows from the construction that if $P(z)$ is a polynomial of degree p, then $Q(z)$ is also a polynomial of degree p.

Uniqueness Let $\tilde{Q}(z)$ be an analytic function satisfying a) and b). Then we have to prove $\tilde{Q}(z) = Q(z)\, e^{i\alpha}$ for some real α. If $z_1 \ldots z_m$ are the zeros of $P(z)$ on the unit circle (if any) then also $Q(z)$ and $\tilde{Q}(z)$ have these zeros and we can write

$$Q(z) = q(z) \prod_{k=1}^{m} (z - z_k) \quad \text{with} \quad q(z) \neq 0 \,, \quad |z| \leq 1$$

$$\tilde{Q}(z) = \tilde{q}(z) \prod_{j=1}^{m} (z - z_j) \quad \text{with} \quad \tilde{q}(z) \neq 0 \,, \quad |z| \leq 1 \,.$$

Hence, it suffices to prove $\tilde{q}(z) = q(z)\, e^{i\alpha}$. Consider the function

$$h(z) := \frac{q(z)}{\tilde{q}(z)} \,, \quad |z| < \rho_o$$

with $\rho_o > 1$ such that $q(z)\,\tilde{q}(z) \neq 0 \,, \quad |z| < \rho_o$. Then both $h(z)$ and $1/h(z)$ can be expanded into a powerseries

$$(2.1.4) \qquad h(z) = \sum_{k=0}^{\infty} m_k z^k \,, \quad \frac{1}{h(z)} = \sum_{j=0}^{\infty} n_j z^j \,, \; |z| < \rho_o \,.$$

We also have using b) and the continuity of $h(z)$ on $|z| < \rho_o$

$$(2.1.5) \qquad |h(e^{-i\lambda})|^2 = \left| \frac{q(e^{-i\lambda})}{\tilde{q}(e^{-i\lambda})} \right|^2 = \lim_{\eta \to e^{-i\lambda}} \left| \frac{Q(\eta)}{\tilde{Q}(\eta)} \right|^2 =$$

$$= \lim_{\eta \to e^{-i\lambda}} \left| \frac{p(\eta)}{p(\eta)} \right|^2 = 1 \,, \quad \lambda \in (-\pi \,, \pi] \,.$$

Hence $1/\ h(e^{-i\lambda}) \ = \ \overline{h(e^{-i\lambda})}$ and with (2.1.4) for $z = e^{-i\lambda}$ we find

$$\frac{1}{h(e^{-i\lambda})} = \sum_{j=0}^{\infty} n_j\, e^{-ij\lambda} = \overline{h(e^{-i\lambda})} = \sum_{k=0}^{\infty} \bar{m}_k\, e^{ik\lambda}, \quad \lambda\in(-\pi,\pi].$$

Since both the Fourier series converge absolutely, it follows by equating coefficients that $n_o = \bar{m}_o$ and $n_k = m_k = 0$, $k \geq 1$. Thus $h(z) = n_o$ or equivalently $q(z) = n_o\, \tilde{q}(z)$ $\forall$ z and from (2.1.5) it follows that $|n_o| = 1$. This proves the lemma. □

It follows from the lemma and (2.1.2) that if $\theta = S_p \times \mathbb{C}^q \times Z$ then there exist $\theta_1 = (a_1 \ldots a_p\ ,\ b_1 \ldots b_q\ ,\ \xi)$ and $\theta_2 = (\tilde{a}_1 \ldots \tilde{a}_p\ ,\ \tilde{b}_1 \ldots \tilde{b}_q\ ,\ \xi)$ with $\theta_1 \neq \theta_2$ and $f_{\theta_1} = f_{\theta_2}$ Hence by (2.1.3) we have $R^{(n)}_{\theta_1} = R^{(n)}_{\theta_2}$, n= 1, 2, ... and so no sample size is second-order informative for the coefficients $(a_1 \ldots a_p\ ,\ b_1 \ldots b_q)$. We may hope, however, that it is sufficient to prescribe A(z) and B(z) to have no zeros inside (or outside) the unit circle. The "natural" choice is no zeros inside the unit circle. Therefore we define

$$S_p^{AR} := \{\ (a_1 \ldots a_p)\ \mid\ A(z) \neq 0,\ \forall\ |z| \leq 1\ \}\ ,$$

and

$$S_q^{MA} := \{\ (b_1 \ldots b_q)\ \mid\ B(z) \neq 0,\ \forall\ |z| < 1\ \}\ .$$

In the next sections we shall derive informative sample sizes for the case $\theta \subset S_p^{AR} \times S_q^{MA} \times Z$. The following lemmas will be needed.

LEMMA 2.1.2 If in (2.1.1) $A(z) \neq 0$ for $|z| \leq 1$, then the covariance function γ_s , s = 0, ± 1, ... of $\{x_t\}$ satisfies

$$(2.1.6) \qquad \sum_{k=0}^{p} a_k \gamma_{s-k} = \begin{cases} b_q \sigma^2 & , \quad s = q \\ 0 & , \quad s = q+1 , q+2 , \ldots \end{cases}$$

PROOF. Let the process $\{\underline{\varepsilon}_t\}$ have the spectral representation

$$\underline{\varepsilon}_t = \int_{-\pi}^{\pi} e^{it\lambda} \underline{z}_\varepsilon\{d\lambda\} \qquad t = 0, \pm 1, \ldots$$

Then the process $\{\underline{x}_t\}$ can be written

$$\underline{x}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} \underline{z}_\varepsilon\{d\lambda\} \qquad t = 0, \pm 1, \ldots$$

and so (see appendix, (A.1.7))

$$E\{\underline{x}_{t-s} \bar{\underline{\varepsilon}}_t\} = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} e^{i(t-s)\lambda} \frac{\bar{B}(e^{-i\lambda})}{\bar{A}(e^{-i\lambda})} e^{-it\lambda} \, d\lambda =$$

$$= \frac{\sigma^2}{2\pi i} \int_{|z|=1} z^{s-1} \frac{B(z)}{A(z)} \, dz = \sigma^2 \delta_s \qquad s = 0,1,2,\ldots$$

where the last equality follows from Cauchy's residu's theorem, since $A(z)$ has no zeros in $|z| \leq 1$, and $A(0) = B(0) = 1$. Thus multiplying (2.1.1) by $\bar{\underline{x}}_{t-s}$ for $s = q, q+1, \ldots$ and taking expectations gives (2.1.6). □

REMARK. In the case of an autoregressive process ($q = 0$) equations (2.1.6) are often called the YULE-WALKER equations.

LEMMA 2.1.3 (FEJÉR and RIESZ) If

$$f(\lambda) := \frac{1}{2\pi} \sum_{k=-q}^{q} \gamma_k e^{-ik\lambda} \geq 0 , \qquad \lambda \in (-\pi , \pi]$$

and $f(\lambda)$ is not identically zero, $f(\lambda)$ can be represented in the form

$$f(\lambda) = \frac{1}{2\pi} \left| \sum_{k=0}^{q} a_k e^{-ik\lambda} \right|^2 ,$$

where $a_o > 0$ and $\sum_{k=0}^{q} a_k z^k \neq 0 , \quad |z| < 1.$

The representation is unique.

PROOF. see HANNAN [20] p. 63 . □

We conclude this section with some conventions for notational convenience. In definitions 1.1.10 and 1.2.1 the space Λ is arbitrary. In particular Λ is allowed to be a space of *measures* or a space of *functions*. In the first case we shortly call the measure $\lambda_\Theta\{.\} = \varphi(\Theta)$ identifiable if the mapping φ is. In the second e.g. if Λ is a space of functions of the complex variable z, we call the function $\lambda_\Theta(z) := \varphi(\Theta)$ identifiable if φ is.
Thus we may speak of informative sample sizes for the spectral measure F_Θ, or for the functions $A(z)$, $B(z)$, $B(z)/A(z)$ etc. Clearly a sample is informative for $A(z)$ iff it is informative for $(a_1 , \ldots a_p)$.

2.2 THE MINIMUM INFORMATIVE SAMPLE SIZE FOR MA(q) PROCESSES

Before we deal with the general case, we consider a simple example .

EXAMPLE 2.2.1 Consider the MA(1) process
$\underline{x}_t = \underline{\varepsilon}_t + \alpha \underline{\varepsilon}_{t-1} , \quad t = 0, \pm 1, \ldots$ Let $\gamma_s , \quad s = 0, \pm 1, \ldots$ denote the covariance function of $\{\underline{x}_t\}$. A simple approach to finding second-order informative sample sizes is solving the *covariance equations*

$$\gamma_o = \sigma^2 (1 + |\alpha|^2) \text{ and } \quad \gamma_1 = \sigma^2\alpha \quad .$$

Even in this simple case one can feel the need of a general powerful approach; as is easily verified, the equations are in general not uniquely solvable for α and σ^2, and it is hard to see what are reasonable conditions.

Suppose we are sampling from the MA(q) process

$$\underline{x}_t = \sum_{k=0}^{q} b_k \underline{\varepsilon}_{t-k} \quad , \quad t = 0, \pm 1, \ldots \quad (b_o = 1) \quad .$$

We have

<u>THEOREM 2.2.2</u> If in the MA(q) case $\theta = S_q^{MA} \times Z$ then the sample size q+1 is second-order informative for $(B(z) , \sigma^2)$.

PROOF. Since the covariance function γ_s vanishes for $|s| \geq q+1$, for all $\theta_1 , \theta_2 \in \theta$ we have the implication

$$(2.2.1) \qquad F_{\theta_1} \neq F_{\theta_2} \Rightarrow R_{\theta_1}^{(q+1)} \neq R_{\theta_2}^{(q+1)} \quad .$$

This implies that it is sufficient to prove the whole process to be second-order informative for $(B(z) , \sigma^2)$, or equivalently to prove the spectral density $f_\theta(\lambda)$ to have a unique decomposition

$$f_\theta(\lambda) = \frac{\sigma}{2\pi} \left| B(e^{-i\lambda}) \right|^2 , \quad \lambda \in (-\pi , \pi] .$$

Suppose there exist $\tilde{B}(z)$ and $\tilde{\sigma}^2$ such that

$$f_\theta(\lambda) = \frac{\tilde{\sigma}^2}{2\pi} \left| \tilde{B}(e^{-i\lambda}) \right|^2 , \quad \lambda \in (-\pi , \pi] .$$

Since $B(z)$ and $\tilde{B}(z)$ have no zeros inside the unit circle, by lemma 2.1.1 we have

$$\sigma\ B(z) = \tilde{\sigma}\ \tilde{B}(z)\ e^{i\alpha}, \quad z \in \mathbb{C}$$

for a real α. As $B(0) = \tilde{B}(0) = 1$, it follows that $\alpha = 0$, $\sigma = \tilde{\sigma}$ and so $B(z) = \tilde{B}(z)$, $z \in \mathbb{C}$. This proves the sample size q+1 to be second-order informative for $(B(z), \sigma^2)$. □

If σ^2 is completely unknown, i.e. if $\sigma^2(Z) = (0,\infty)$ we can prove that the sample size q is not second-order informative for $(B(z), \sigma^2)$, and therefore that q+1 is the minimum second-order informative sample size. We shall not do this but we shall prove a much stronger result on nowhere local identifiability. Here "local" is with respect to the Euclidean topology on $\mathbb{C}^{q+1}$ restricted to $S_q^{MA} \times (0,\infty)$. We need the following lemma.

LEMMA 2.2.3 Let $D := \{(b_1 \ldots b_q) \mid B(z) \neq 0, |z| \leq 1\}$
Then D is a dense subset of S_q^{MA}.

PROOF. See DEISTLER DUNSMUIR and HANNAN [8], Lemma 1. □

THEOREM 2.2.4 If in the MA(q) case $\theta = S_q^{MA} \times Z$ and σ^2 is completely unknown, then the sample size q is nowhere locally second-order informative for $(B(z), \sigma^2)$.

PROOF. By lemma 2.2.3 and lemma 1.9.2 it follows that it suffices to prove the theorem for $\theta = D \times Z$. Let $(B(z), \sigma^2)$ be arbitrary such that $B(z) \neq 0$, $|z| \leq 1$. Then the theorem is proved if we can find a sequence $(B^{(n)}(z), \sigma_n^2)$ converging to $(B(z), \sigma^2)$ (convergence of coefficients) such that $B^{(n)}(z) \neq 0$ on $|z| \leq 1$ for all n, and such that $(B^{(n)}(z), \sigma_n^2)$ is not identical with $(B(z), \sigma^2)$ but generates the same covariances γ_s for $s = 0, \pm 1, \ldots \pm (q-1)$.
Introduce the covariance generating function

$G(z) := \sum_{s=-q}^{q} \gamma_s z^{-s}$, $z \neq 0$. Using the convention $b_k = 0$ for $k < 0$ and $k > q$ we can write

$\gamma_s = \sigma^2 \sum_{k=-\infty}^{\infty} b_k \bar{b}_{k+s}$, $s = 0, \pm 1, \ldots$. Hence

$$G(z) = \sigma^2 \sum_{s=-q}^{q} \sum_{k=-\infty}^{\infty} b_k \bar{b}_{k+s} z^{-s} = \sigma^2 \sum_{k=-\infty}^{\infty} b_k \sum_{s=-\infty}^{\infty} \bar{b}_{k+s} z^{-s-k} z^k =$$

$$= \sigma^2 B(z) \overline{B(\frac{1}{\bar{z}})} \quad , \quad z \neq 0 \quad .$$

Note, that $G(e^{-i\lambda}) = \sigma^2 | B(e^{-i\lambda}) |^2 > 0$, $\lambda \in (-\pi, \pi]$, since $B(z)$ has no zero on $|z| = 1$. Thus we may define $\delta > 0$ by

$$\delta := \min_{\lambda \in (-\pi, \pi]} G(e^{-i\lambda}) \quad .$$

Let $n > \frac{1}{\delta}$ and put

$$G^{(n)}(z) = G(z) + \frac{1}{4n} z^q + \frac{1}{4n} z^{-q} \quad , \quad z \neq 0 \quad .$$

Then $G^{(n)}(e^{-i\lambda})$ is also real valued and for all $\lambda \in (-\pi, \pi]$. We have

$$G^{(n)}(e^{-i\lambda}) = G(e^{-i\lambda}) + \frac{1}{2n} \cos q\lambda \geq \frac{1}{2} \delta \quad .$$

Furthermore,

$$(2.2.2) \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{is\lambda} G^{(n)}(e^{-i\lambda}) d\lambda = \begin{cases} \gamma_s & |s| = 0,1,\ldots,q-1 \\ \gamma_s + \frac{1}{4n} & |s| = q \\ 0 & |s| = q+1, q+2, \ldots \end{cases}$$

Hence by lemma 2.1.3 $G^{(n)}(e^{-i\lambda})$ can be decomposed into

$\sigma_n^2 \mid B^{(n)}(e^{-i\lambda})\mid^2$, $\lambda \in (-\pi, \pi]$ where $B^{(n)}(z)$ is a polynomial of degree $\leq q$ without zeros in $|z| \leq 1$. Since σ^2 is completely unknown, σ_n^2 may be chosen to obtain $B^{(n)}(0) = 1$. We may even express $\sigma_n B^{(n)}(z)$ explicitly in $G^{(n)}(z)$ as

$$(2.2.3) \quad \sigma_n B^{(n)}(z) = \exp\left[\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{e^{i\lambda}+z}{e^{i\lambda}-z} \log G^{(n)}(e^{-i\lambda})d\lambda\right], z \in \mathbb{C}$$

(see HANNAN [20] p. 142 theorem 5). From (2.2.2) and (2.2.3) it follows that we must have

$$\lim_{n\to\infty} \sigma_n B^{(n)}(z) = \sigma B(z) \quad , \quad z \in \mathbb{C} \ .$$

Taking $z = 0$ it follows that $\sigma_n \to \sigma$, $n \to \infty$ and so $B^{(n)}(z) \to B(z)$, $n \to \infty$, $z \in \mathbb{C}$. Thus we found a sequence $(B^{(n)}(z) , \sigma_n^2)$ converging to $(B(z) , \sigma^2)$ with the desired properties.
This proves the theorem . □

REMARK 1. Note that the minimum informative sample size equals the number of unknown parameters of the model.

REMARK 2. For $q \geq 1$, b_q is allowed to be zero.

2.3 INFORMATIVE SAMPLES FROM AR(p) PROCESSES

Suppose we are sampling from the weakly stationary process $\{\underline{x}_t\}$ satisfying

$$(2.3.1) \quad \sum_{k=0}^{p} a_k \underline{x}_{t-k} = \underline{\varepsilon}_t \ , \quad t = 0, \pm 1, \ldots , a_o = 1 \ .$$

We have

THEOREM 2.3.1 If in the AR(p) case $\theta = S_p^{AR} \times Z$ then the sample size p+1 is second order-informative for $(A(z), \sigma^2)$.

PROOF. Since $A(z) \neq 0$, $|z| \leq 1$ lemma 2.1.2 is applicable with q = 0 and yields

$$\sum_{k=0}^{p} a_k \gamma_{s-k} = \sigma^2 \delta_s , \qquad s = 0, 1, 2, \ldots$$

Using $\gamma_{-s} = \bar{\gamma}_s$, the first p + 1 of these equations can be written in matrix notation as follows

$$(2.3.2) \qquad (a_1 \ldots a_p, \sigma^2) W + (\gamma_o \ldots \gamma_p) = 0 ,$$

where W is the (p + 1) × (p + 1) matrix given by

$$(2.3.3) \quad W = \begin{bmatrix} \gamma_1 & \gamma_o & \cdots & \gamma_{p-1} \\ \bar{\gamma}_2 & \gamma_1 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \bar{\gamma}_p & \bar{\gamma}_{p-1} & \cdots & \gamma_o \\ -1 & 0 & \ldots & 0 \end{bmatrix} = \begin{bmatrix} \bar{\gamma}_1 & & \\ \bar{\gamma}_2 & & \\ \vdots & R^{(p)} & \\ \bar{\gamma}_p & & \\ -1 & 0 \ \ldots \ 0 & \end{bmatrix} .$$

Suppose the sample size p + 1 is not second-order informative for $(A(z), \sigma^2)$. Then there exist pairs $(A^{(1)}(z), \sigma_1^2) \neq (A^{(2)}(z), \sigma_2^2)$ that generate the same γ_s for $s = 0, \pm 1, \ldots \pm p$. But then it follows from (2.3.2) that W must be singular and by (2.3.3) this is equivalent to singularity of $R^{(p)}$. Hence there exist numbers $c_o \ldots c_{p-1}$, not all zero, such that

$$(2.3.4) \quad \sum_{k=0}^{p-1} c_k \underline{x}_{t-k} = 0 \qquad \text{(a.s.)} \quad t = 0, \pm 1, \ldots$$

Since the process $\{\underline{x}_t\}$ has the spectral density

$$\frac{\sigma^2}{2\pi} \frac{1}{|A(e^{-i\lambda})|^2} , \qquad \lambda \in (-\pi , \pi] ,$$

the relation (2.3.4) implies

$$\frac{\sigma^2}{2\pi} \frac{1}{|A(e^{-i\lambda})|^2} \left| \sum_{k=0}^{p-1} c_k e^{-ik\lambda} \right|^2 = 0$$

a.e. w.r.t. Lebesgue measure. As $A(z) \neq 0$ on $|z| = 1$, it follows that $\sum_{k=0}^{p-1} c_k e^{-ik\lambda} = 0$ and so $c_k = 0$, $k = 0, 1, \ldots, p-1$. Thus we have a contradiction and the theorem is proved. □

REMARK. It is not necessary that σ^2 is completely unknown (compare th. 2.2.1) since we dit not prove the sample size $p + 1$ to be minimal. It is not easily seen how we could prove the minimality due to the fact that lemma 2.1.3 has no analogue applicable to spectra of autoregressive processes.

2.4 INFORMATIVE SAMPLES FROM ARMA(p,q) PROCESSES

In spite of the remarkable difference in the proofs in the MA(q) und AR(p) case, combining the methods points the way to treating the ARMA(p,q) case. However, since the spectral density (2.1.2) (and so the covariance function) depends only on the ratio $B(z) / A(z)$, we cannot expect any sample size to be second-order informative for $(A(z) , B(z) , \sigma^2)$ if $A(z)$ and $B(z)$ are allowed to have common factors. Thus $A(z)$ and $B(z)$ must have no zeros in common, or, equivalently, $(A(z) , B(z)) \neq (0,0)$, $z \in \mathbb{C}$. Let

$$S_{p,q}^{ARMA} := \{ (a_1 \ldots a_p, b_1 \ldots b_q) \mid A(z) \neq 0, |z| \leq 1, \; B(z) \neq 0 \; |z| < 1 ,$$
$$(a_p, b_q) \neq (0,0) \text{ and } (A(z), B(z)) \neq (0,0) \quad \forall \; z \}$$

clearly

$$S_{p,q}^{ARMA} \subset S_p^{AR} \times S_q^{MA} \ .$$

Notice that the condition $(a_p \ , \ b_q) \neq (0 \ , \ 0)$ is automatically satisfied if $p = 0$ or $q = 0$ since $a_o = b_o = 1$.
Before we can establish a result on the ARMA(p,q) case we need the following lemma.

LEMMA 2.4.1 Let the function $\varphi(z)$ be analytic on $|z| < \rho$ for some $\rho > 1$, and $\varphi(z) \neq 0$, $|z| < 1$.

$$(2.4.1) \qquad \int_{-\pi}^{\pi} e^{is\lambda} \, |\varphi(e^{-i\lambda})|^2 \, d\lambda = 0 \quad , \qquad |s| = q \ , \ q+1, \ldots$$

then $\varphi(z)$ is a polynomial whose degree is at most q-1.

PROOF. From relation (2.4.1) it follows that $|\varphi(e^{-i\lambda})|^2$ can be written in the form

$$\sum_{k=-q+1}^{q-1} \alpha_k \, e^{ik\lambda} \quad , \quad \lambda \in (-\pi \ , \ \pi]$$

with $\alpha_k = \bar{\alpha}_{-k}$, $k=0,1,\ldots q-1$. Hence by lemma (2.1.3) there exists a polynomial $P(z)$ of degree at most q-1 such that

$$|\varphi(e^{-i\lambda})|^2 = |P(e^{-i\lambda})|^2 , \quad \lambda \in (-\pi \ , \ \pi]$$

and $P(z) \neq 0$, $|z| < 1$.
But then by lemma 2.1.1 we have $\varphi(z) = e^{i\alpha} P(z)$, $|z| < \rho$ for some real α and so $\varphi(z)$ is a polynomial of degree $\leq$ q-1. □

THEOREM 2.4.2 If in the ARMA(p,q) case $\theta = S_{p,q}^{ARMA} \times Z$ then the sample size p + q + 1 is second-order informative for $(A(z) \ , \ B(z) \ , \ \sigma^2)$.

PROOF. The proof consists of three steps; in the first step we prove the sample size p + q + 1 to be second-order informative for A(z). In the second step we prove it to be second-order informative for $(B(z)\ ,\ \sigma^2)$ *conditional* on A(z) (see def. 1.2.1 and lemma 1.2.2). In the third step we apply the conditional identification theorem (th. 1.2.3) to obtain the result. For q = 0 theorem 2.3.1 applies. Thus suppose $q \geq 1$.
1) Since $A(z) \neq 0$, $|z| \leq 1$ it follows from lemma 2.1.2 that

$$\sum_{k=0}^{p} a_k \gamma_{s-k} = 0 \ , \qquad s = q+1 \ , \ q+2 \ , \ \ldots \tag{2.4.3}$$

Since $a_o = 1$, the first p of these equations can be put into matrixform

$$(a_1 \ \ldots \ a_p)\ \tilde{R}_q + (\gamma_{q+1} \ \ldots \ \gamma_{q+p}) = 0 \ , \tag{2.4.4}$$

where the p x p matrix $\tilde{R}_q$ is given by

$$\tilde{R}_q = \begin{bmatrix} \gamma_q & \gamma_{q+1} & \cdots & \gamma_{q+p-1} \\ \gamma_{q-1} & \gamma_q & \ldots & \gamma_{q+p-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{q-p+1} & \gamma_{q-p+2} & \cdots & \gamma_q \end{bmatrix} .$$

Suppose the sample size p + q + 1 is not second-order informative for A(z). Then there exist $A^{(1)}(z)$ and $A^{(2)}(z) \neq A^{(1)}(z)$ that correspond to the same γ_s for $s = 0, \pm 1, \ldots \pm (p+q)$.
From (2.4.4) it then follows that $\tilde{R}_q$ must be singular. (Note that $\tilde{R}_q$ is not a principal minor of a covariance matrix if $q \geq 1$ so that the method of proof of theorem 2.3.1 fails.)
We shall derive a contradiction.
Let $c_o \ \ldots \ c_{p-1}$ be numbers, not all zero such that

$$(2.4.5) \qquad \sum_{k=0}^{p-1} c_k \gamma_{s-k} = 0 \quad , \qquad s=q \ , \ q+1 \ , \ldots \ q+p-1 \ .$$

From (2.4.3) we have

$$(2.4.6) \qquad \gamma_s = - \sum_{k=1}^{p} a_k \gamma_{s-k} \ , \qquad s = q+1 \ , \ q+2 \ , \ldots \ ,$$

and in particular

$$(2.4.7) \qquad \gamma_{q+p-s} = - \sum_{k=1}^{p} a_k \gamma_{q+p-s-k} \ , \ s = 0, \ 1, \ \ldots, \ p-1 .$$

Multiplying by c_s summing over s and using (2.4.5) yields

$$(2.4.8) \qquad \sum_{s=0}^{p-1} c_s \gamma_{q+p-s} = - \sum_{k=1}^{p} a_k \sum_{s=0}^{p-1} c_s \gamma_{q+p-s-k} = 0.$$

Thus the range of validity of (2.4.5) is extended to the value $s = q+p$. However the same procedure can be repeated an arbitrary number of times, which proves the validity of (2.4.5) for all $s \geq q$.

Let $$\underline{Y}_t := \sum_{k=0}^{p-1} c_k \underline{x}_{t-k} \quad , \ t = 0, \pm 1, \ \ldots$$

Without loss of generality we may suppose $c_o \neq 0$. (If $c_o = \ldots = c_{j-1} = 0$, $c_j \neq 0$ then we can take $\underline{Y}_t = \sum_{k=0}^{p-j-1} \tilde{c}_k \underline{x}_{t-k}$ where $\tilde{c}_k = c_{k+j}$).

Introducing the generating function $C(z) := \sum_{k=0}^{p-1} c_k z^k$ the spectral density of the process $\{\underline{Y}_t\}$ can be written

$$(2.4.9) \qquad g(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{C(e^{-i\lambda}) \ B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2 \ , \ \lambda \in (-\pi \ , \ \pi] \ .$$

Furthermore, since (2.4.5) holds for all $s \geq q$,

$$(2.4.10) \qquad E\{\underline{Y}_{t+s}\ \bar{\underline{x}}_{t-j}\} = \sum_{k=0}^{p-1} c_k\, \gamma_{s+j-k} = 0 \quad , \ s + j \geq q \ .$$

Multiplying with $\bar{c}_j$, summing over j, and using (2.4.10) gives

$$(2.4.11) \qquad E\{\underline{Y}_{t+s}\ \bar{\underline{Y}}_t\} = \sum_{j=0}^{p-1} \bar{c}_j\, E\{\underline{Y}_{t+s}\ \bar{\underline{x}}_{t-j}\} =$$

$$= \sum_{j=0}^{p-1} \bar{c}_j \sum_{k=0}^{p-1} c_k\, \gamma_{s+j-k} = 0 \ , \ s \geq q .$$

Using (2.4.9), we see that this is equivalent to

$$(2.4.12) \qquad \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} e^{is\lambda} \left| \frac{C(e^{-i\lambda})\ B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2 d\lambda = 0, \ |s| \geq q.$$

By lemma 2.1.1, C(z) can be chosen such that $C(z) \neq 0$, $|z| < 1$. Hence, lemma 2.4.1 is applicable if $q \geq 1$ because $C(z)\ B(z)/A(z)$ is analytic on $|z| < \rho$ for some $\rho > 1$ since $A(z) \neq 0$, $|z| \leq 1$. It follows that $C(z)\ B(z)\ /\ A(z)$ is a polynomial of degree at most q-1. But then it follows that we must have $a_p = b_q = 0$ since A(z) and B(z) have no common factors.

Thus we have a contradiction proving the sample size p + q + 1 to be second-order informative for A(z).

2) We now fix A(z) and put

$$(2.4.13) \qquad \underline{z}_t = \sum_{k=0}^{p} a_k\, \underline{x}_{t-k} = \sum_{j=0}^{q} b_j\, \underline{\varepsilon}_{t-j}, \quad t = 0, \pm 1, \ldots$$

Then $\{\underline{z}_t\}$ is a MA(q) process with unknown coefficients $(b_1 \ldots b_q)$ and σ^2, and according to theorem 2.2.2 a sample of size q+1 of this process is second-order informative for $(B(z)\ ,\ \sigma^2)$.

Using (2.4.13) it follows that a sample of size $p + q + 1$ of the process $\{\underline{x}_t\}$ is second-order informative for $(B(z)\ ,\ \sigma^2)$ conditional on $A(z)$ (see also lemma 1.2.2).

3) Application of theorem 1.2.3 to 1) and 2) gives the desired result.□

REMARK 1. It follows from the proof that if $A(z)$ and $B(z)$ have no factor in common and $(a_p\ ,\ b_q) \neq (0,\ 0)$, then the matrix $\tilde{R}_q$ is nonsingular. Hence we may express $(a_1\ \ldots\ a_p)$ by (2.4.4) explicitly in the first $p + q + 1$ covariances. This may be important to initiate estimation procedures (see HANNAN [19]). In the next section we shall see that as far as identification is concerned, the condition $(a_p\ ,\ b_q) \neq (0,\ 0)$ can be dropped.

REMARK 2. The proof given here is of special interest because HANNAN proved in [19] for the m-variate case a matrix like $\tilde{R}_q$ to be nonsingular. However in chapter III we shall show that HANNAN'S proof is not correct and even that the result is false in general, if $m > 1$.

2.5 INFORMATIVE SAMPLES FOR THE SPECTRAL MEASURE; PREDICTABILITY

Let $\underline{x}_1\ ,\ \underline{x}_2\ ,\ldots,\ \underline{x}_n$ be a sample from a weakly stationary process $\{\underline{x}_t\}$ with spectral measure $F_\theta\ ,\ \theta \in \Theta$. Two fundamental problems are

1) finding second-order informative sample sizes for F_θ
2) second-order predictability of the future, or equivalently finding second-order predictive sample sizes.

The following lemma shows that these problems are equivalent.

LEMMA 2.5.1 The sample size n is second-order predictive iff it is second-order informative for the spectral measure.

PROOF. Let the sample size n be second-order predictive, and let $\Theta_1, \Theta_2 \in \Theta$ such that $F_{\Theta_1} \neq F_{\Theta_2}$. Since the spectral measure is uniquely determined by the covariances of the process $\{\underline{x}_t\}$ there exist an integer N such that $R^{(N)}_{\Theta_1} \neq R^{(N)}_{\Theta_2}$.

If $N \leq n$ it follows that $R^{(n)}_{\Theta_1} \neq R^{(n)}_{\Theta_2}$ and if $N > n$ we have $R^{(n)}_{\Theta_1} \neq R^{(n)}_{\Theta_2}$ since n was predictive. Thus the sample size n is second-order informative for F_Θ.

Let the sample size n be second-order informative for F_Θ. Then for all $\Theta_1, \Theta_2 \in \Theta$ and all N we have the implications

$$R^{(N)}_{\Theta_1} \neq R^{(N)}_{\Theta_2} \Rightarrow F_{\Theta_1} \neq F_{\Theta_2} \Rightarrow R^{(n)}_{\Theta_1} \neq R^{(n)}_{\Theta_2},$$

which proves the sample size n to be predictive. □

Let $\{\underline{x}_t\}$ be an ARMA(p,q) process. Then it follows from (2.1.2) and lemma 2.1.1 that the sample size n is second-order informative for the spectral measure iff it is informative for $\sigma \; B(z)/_{A(z)}$, provided $\theta \subset S^{AR}_q \times S^{MA}_q \times Z$.
In particular, if the sample size n is second-order informative for $(A(z), B(z), \sigma^2)$, it is second-order informative for the spectral measure thus we may hope that relatively weak conditions on $(A(z), B(z), \sigma^2)$ can lead to predictability of the future. We have the following result.

THEOREM 2.5.2 If in the ARMA(p,q) case $\theta = S^{AR}_p \times \mathbb{C}^q \times Z$ then the sample size p + q + 1 is second-order informative for the spectral measure, or, equivalently, second-order predictive.

PROOF. Assume the sample size p + q + 1 is not second-order informative for the spectral measure. Then there exist $\Theta, \tilde{\Theta} \in \theta$ with $F_\Theta \neq F_{\tilde{\Theta}}$ such that

$$(2.5.1) \qquad (\gamma_o\ ,\ \gamma_1 \cdots \gamma_{p+q}) = (\tilde{\gamma}_o\ ,\ \tilde{\gamma}_1 \cdots \tilde{\gamma}_{p+q}) \quad ,$$

where γ_s and $\tilde{\gamma}_s$ denote the covariance functions corresponding to F_θ and $F_{\tilde{\theta}}$ respectively.

Let $A(z) = \sum_{k=0}^{p} a_k z^k$ correspond to θ and $\tilde{A}(z) = \sum_{k=0}^{p} \tilde{a}_k z^k$ to $\tilde{\theta}$. Since both $A(z)$ and $\tilde{A}(z)$ have no zeros on $|z| \leq 1$ it follows from lemma 2.1.2 that

$$(2.5.2) \quad \begin{cases} \sum_{k=0}^{p} a_k \gamma_{s-k} = 0 \quad , & s = q+1,\ q+2,\ \ldots \\ \sum_{k=0}^{p} \tilde{a}_k \tilde{\gamma}_{s-k} = 0 \quad , & s = q+1,\ q+2,\ \ldots \end{cases}$$

By (2.5.1) it follows

$$(2.5.3) \qquad \sum_{k=0}^{p} \tilde{a}_k \gamma_{s-k} = 0 \quad , \qquad s = q+1,\ q+2,\ \ldots\ ,\ q+p.$$

We shall prove (2.5.3) to be valid for $s = q + p + 1$ and so by induction for all $s \geq q + p + 1$.

Since $a_o = \tilde{a}_o = 1$ we have

$$\gamma_s = - \sum_{j=1}^{p} a_j \gamma_{s-j} \quad , \qquad s = q+1,\ q+2,\ \ldots$$

Hence

$$\sum_{k=0}^{p} \tilde{a}_k \gamma_{p+q+1-k} = - \sum_{k=0}^{p} \tilde{a}_k \sum_{j=1}^{p} a_j \gamma_{p+q+1-k-j} =$$

$$= - \sum_{j=1}^{p} a_j \sum_{k=0}^{p} \tilde{a}_k \gamma_{p+q+1-j-k} = 0 \quad ,$$

where the last equality follows from (2.5.3).

Thus

$$\gamma_{p+q+1} = -\sum_{k=1}^{p} \tilde{a}_k \gamma_{p+q+1-k} = -\sum_{k=1}^{p} \tilde{a}_k \tilde{\gamma}_{p+q+1-k} = \tilde{\gamma}_{p+q+1} ,$$

and by induction it follows that we must have $\gamma_s = \tilde{\gamma}_s$ for all s. But then it follows that the spectral measures F_θ and $F_{\tilde{\theta}}$ must be identical which contradicts our assumption, and proves the theorem. □

Theorem 2.5.2 enables us to prove a slight generalization of theorem 2.4.2. Let $\tilde{S}^{ARMA}_{p,q}$ be obtained from $S^{ARMA}_{p,q}$ by deleting the condition $(a_p , b_q) \neq (0,0)$. Then we have

THEOREM 2.5.3 If in the ARMA(p,q) case $\theta = \tilde{S}^{ARMA}_{p,q} \times Z$ then the sample size $p + q + 1$ is second-order informative for $(A(z) , B(z) , \sigma^2)$.

PROOF. By theorem 2.5.2 the sample size $p + q + 1$ is second-order informative for the spectral density which is known to be of the form

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \varphi(e^{-i\lambda}) \right|^2 , \qquad \lambda \in (-\pi , \pi] .$$

Here $\varphi(z)$ is rational, nonzero in $|z| < 1$, and analytic on $|z| < \rho$ for some $\rho > 1$. Furthermore $\varphi(0) = 1$. Thus by lemma 2.1.1 $f(\lambda)$ determines $(\varphi(z), \sigma^2)$ uniquely. But then $A(z)$ and $B(z)$ are uniquely determined by $f(\lambda)$ since they are known to have no common factor and $A(0) = B(0) = 1$. This proves the theorem. □

REMARK. The generalization is possible since we do not need the condition $(a_p , b_q) \neq (0,0)$ for the numerator and denominator of the rational function $A(z) / B(z)$ to have no common factor. We shall see that in the multivariate case such a generalization is not possible.

2.6 APPLICATION TO LINEAR REGRESSION WITH MA-ERRORS AND STATIONARY LAGGED EXPLANATORY VARIABLES

The methods used in the preceding sections admit applications in several kinds of problems.
We give two examples.

a) *Linear regression with MA-errors*

We consider the regression model

$$\underline{Y}_t = \beta' x_t + \sum_{k=0}^{q} b_k \underline{\varepsilon}_{t-k} , \qquad t = 0, \pm 1, \ldots \; b_o = 1$$

where $\{\underline{\varepsilon}_t\}$ is white noise with $E\{\underline{\varepsilon}_t\} = 0$ and $E\{|\underline{\varepsilon}_t|^2\} = \sigma^2 > 0$, $\beta \in \mathbb{R}^k$ is a vector of (unknown) regression coefficients and $x_t \in \mathbb{R}^k$ $t = 0, \pm 1, \ldots$ is a sequence of predetermined regressors. Since the model is essentially that of example 1.2.10, we may treat the identification of β in the same way we did there. Therefore suppose a sample of $\{\underline{Y}_t\}$ taken at $t = 1, 2, \ldots, n$ is first-order informative for β. Putting $B(z) = \sum_{k=0}^{q} b_k z^k$ it follows from theorem 2.2.2 that if $n \geq q+1$ and $B(z) \neq 0$, $|z| < 1$ then the sample is second-order informative for $(B(z), \sigma^2)$ conditional on β.
Hence by (the conditional indentification) theorem 1.2.3 it follows that we may treat the identification of β and $(B(z), \sigma^2)$ separately, the latter by considering $\underline{Y}_t - \beta' x_t$, $t = 1, \ldots, n$ as observable.

b) *Linear regression with stationary lagged explanatory variables and MA-errors*

Consider the model

$$\underline{Y}_t = \sum_{k=0}^{p} a_k \underline{x}_{t-k} + \sum_{j=0}^{q} b_j \underline{\varepsilon}_{t-j} , \; t = 0, \pm 1, \ldots \qquad b_o = 1 , \tag{2.6.1}$$

where $\{\underline{x}_t\}$ is a weakly stationary (observable) process with

$E\{\underline{x}_t\} = 0$ and known absolutely continuous spectral measure with density f_x; $\{\underline{\varepsilon}_t\}$ is white noise with $E\{\underline{\varepsilon}_t\} = 0$ and $E\{|\underline{\varepsilon}_t|^2\} = \sigma^2 > 0$. (It is not essential that f_x known, because we can treat its identifiability separately). The process $\{\underline{x}_t\}$ and $\{\underline{\varepsilon}_t\}$ are supposed to be mutually orthogonal.

Let $\xi \in Z$ characterize the distribution of $\{\underline{\varepsilon}_t\}$ and put $\theta := (a_o, a_1 \ldots a_p ; b_1 \ldots b_q , \xi)$. Since both $\{\underline{y}_t\}$ and $\{\underline{x}_t\}$ are observable, we have, in fact, bivariate observations, and 'sample size' shall be interpreted as the number of points in time that the bivariate process $(\underline{y}_t , \underline{x}_t)$ is observed. However, the model (2.6.1) is essentially univariate and admits a treatment analogous to that of the univariate models in the preceding sections. Therefore the problem is treated at this stage. As usual, let the generating functions of the a's and b's be denoted by $A(z)$ and $B(z)$ respectively.

Notice that $\sum_{k=0}^{p} a_k \underline{x}_{t-k}$ can be considered as the *linear regression function* of $\underline{y}_t$ on $(\underline{x}_t , \underline{x}_{t-1}, \ldots \underline{x}_{t-p})$. Therefore we may hope that we do not need any restriction on $A(z)$.

THEOREM 2.6.1 If in the model (2.6.1) $\theta = \mathbb{C}^{p+1} \times S_q^{MA} \times Z$ then the sample size $p + q + 1$ is second-order informative for $(A(z), B(z), \sigma^2)$.

PROOF. Compare the proof of theorem 2.4.2 for the ARMA(p,q) case. First we prove the sample size $p + 1$ to be second-order informative for $A(z)$. Let

$$\left.\begin{aligned} \gamma_s &:= E\{ \underline{x}_{t+s} \ \bar{\underline{x}}_t \} \\ \tilde{\gamma}_s &:= E\{ \underline{y}_{t+s} \ \bar{\underline{x}}_t \} \end{aligned}\right\} \quad s,t = 0, \pm 1, \ldots$$

Because $\{\underline{x}_t\}$ and $\{\underline{\varepsilon}_t\}$ are mutually orthogonal processes it follows from (2.6.1) that

$$\tilde{\gamma}_s = \sum_{k=0}^{p} a_k \gamma_{s-k} \quad , \qquad s = 0, \pm 1, \ldots$$

Put the first p+1 of these equations into matrix form

$$(2.6.2) \qquad \begin{pmatrix} \tilde{\gamma}_o \\ \tilde{\gamma}_1 \\ \vdots \\ \tilde{\gamma}_p \end{pmatrix} = R_x^{(p+1)} \begin{pmatrix} a_o \\ a_1 \\ \vdots \\ a_p \end{pmatrix} \quad ,$$

where $R_x^{(p+1)}$ denotes the covariance matrix of $(\underline{x}_o \ \ldots \ \underline{x}_p)$. Suppose the sample size p+1 is not second-order informative for $A(z)$. Then (2.6.2) implies that $R_x^{(p+1)}$ must be singular, and so there exist numbers $c_o, \ldots, c_p$ not all zero such that

$$\sum_{k=0}^{p} c_k \underline{x}_{p-k} = 0 \qquad \text{a.s.}$$

Since $\{\underline{x}_t\}$ is weakly stationary this implies

$$\sum_{k=0}^{p} c_k \underline{x}_{t-k} = 0 \qquad \text{a.s} \ , \ t = 0, \pm 1, \ldots$$

Thus $\{\underline{x}_t\}$ satisfies a homogeneous difference equation and so it must have a spectral measure that is concentrated in a finite set. (appendix corollary A.1.3). This contradicts the absolute continuity and proves the sample size p+1 to be second-order informative for $A(z)$.

In the same way as in the proof of theorem 2.4.2 we can now prove the sample size p + q + 1 to be second-order informative for $(B(z), \sigma^2)$ conditional on $A(z)$ and application of the conditional identification theorem (theorem 1.2.3) completes the proof. □

REMARK.(2.6.2) are the *normal equations* of the regression problem (2.6.1).

CHAPTER III

INFORMATIVE SAMPLE SIZES IN MULTIVARIATE STATIONARY MODELS

3.1 INTRODUCTION

In this chapter we treat multivariate analogues of the problems of informative sample sizes of Chapter II. As in the univariate case all stochastic processes are supposed to be defined for $t = 0, \pm 1, \ldots$ and are allowed to have complex valued components. Thus we are interested now in m-variate weakly stationary observable processes $\{\underline{x}_t\}$ that satisfy a linear stochastic difference equation of the form

$$(3.1.1) \qquad \sum_{k=0}^{p} A_k \, \underline{x}_{t-k} = \sum_{j=0}^{q} B_j \, \underline{\varepsilon}_{t-j} , \qquad t = 0, \pm 1, \ldots$$

$$A_o = B_o = I_m$$

where $\{\underline{\varepsilon}_t\}$ is a (non-observable) m-variate white noise process (i.e. : $E\{\underline{\varepsilon}_t \, \underline{\varepsilon}_s^*\} := \delta_{t-s} \, \Sigma_\varepsilon$, $t = 0, \pm 1, \ldots, \Sigma_\varepsilon \geq 0$) with $E\{\underline{\varepsilon}_t\} = 0$, $\forall_t$. As in the univariate case the integers p and q are supposed to be known and we shall use the same terminology for the cases p=0, q=0 and $pq \geq 1$, respectively. The $m \times m$ matrices $A_1, \ldots, A_p$, $B_1, \ldots, B_q$, Σ_ε are supposed to be unknown and are allowed to have complex components. If $\xi \in Z$ is a parameter that characterizes the distribution of the process $\{\underline{\varepsilon}_t\}$, then we write $\Sigma_\varepsilon = \Sigma_\varepsilon(\xi)$ and we put $\Theta = (A_1 , A_2 , \ldots, A_p , B_1 , \ldots, B_q , \xi)$. Define the matrix generating functions

$$A(z) := \sum_{k=0}^{p} A_k \, z^k , \qquad B(z) = \sum_{j=0}^{q} B_j \, z^j , \qquad z \in \mathbb{C}$$

Matrix valued functions of this type are called *matrix polynomials*. If $A_p \neq 0$, then p is the degree of A(z) (similarly

for $B(z)$).

Let $\mathbb{C}(m)$ denote the set of all complex $m \times m$ matrices, and $\mathbb{C}^n(m)$ its n-fold cartesian product. Then we can write $(A_1, \ldots, A_p) \in \mathbb{C}^p(m)$ and in order to ensure the existence of a unique weakly stationary solution of (3.1.1), we must restrict the range of $(A_1, \ldots, A_p)$ to $S_p(m) \subset \mathbb{C}^p(m)$ where

$$S_p(m) := \{(A_1, \ldots, A_p) \mid \det A(z) \neq 0, \quad |z| = 1\} .$$

So θ should be chosen such that

$$\theta \subset S_p(m) \times \mathbb{C}^q(m) \times \mathbf{Z} .$$

In this case the process $\{\underline{x}_t\}$ has a spectral density matrix f_θ given by

$$(3.1.2)\quad f_\theta(e^{-i\lambda}) = \frac{1}{2\pi} A^{-1}(e^{-i\lambda}) B(e^{-i\lambda}) \Sigma_\varepsilon B^*(e^{-i\lambda}) A^{-1*}(e^{-i\lambda}),$$

$$\lambda \in (-\pi, \pi]$$

(see appendix). Here it will become clear why the multivariate case is essentially more difficult than the univariate case; in the latter case we could give conditions in terms of zeros of polynomials; however, one cannot expect to find conditions purely in terms of zeros of determinants of matrixpolynomials since the spectra depend explicitly on the matrix polynomials themselves. It turns out that the zeros of determinants only partly take over the role of zeros of polynomials in the univariate case.

Another source of possible trouble is the fact that for some matrixpolynomial $Q(z)$, $\det Q(z)$ may vanish identically, while $Q(z)$ does not. As a consequence, an m-variate weakly stationary process satisfying a homogeneous difference equation does not necessarily have a singular spectral measure (see theorem A.2.1 appendix).

In the univariate case the singularity was used in the proof of theorem 2.6.1.

Finally, we want to mention an intuitive argument for the separate treatment of univariate and multivariate models. In the univariate case we proved a sample to be informative when its dimensionality (i.e. the number of observed random variables) equals the number of unknown parameters (e.g. theorem 2.2.1 remark 1). However, in the multivariate case this number increases rapidly with m, and, as we shall see, the number of observed random variables may be smaller then the number of unknown parameters and still be informative for those parameters.

3.2 THE FUNDAMENTAL LEMMA AND SOME SPECIAL MATRIX THEORY

The role of absolute values of polynomials in the univariate case is in the multivariate case taken over by hermitian positive semi-definite quatratic forms. Further the zeros of determinants are important, and therefore we have to make sure that determinants do not vanish identically. Necessary and sufficient for this to be so is that the matrixpolynomial is nonsingular for at least one point $z_1 \in \mathbb{C}$. In order to obtain a generalization of lemma 2.1.1 we consider $m \times m$ matrices. Their elements are functions that are analytic on $|z| \leq \rho$ for some $\rho > 1$. We have

LEMMA 3.2.1 Let the $m \times m$ matrix $P(z)$ be analytic on $|z| < \rho$ for $\rho > 1$, and let $P(0)$ be nonsingular.
If Σ is an arbitrary hermitian positive definite $m \times m$ matrix, then there exists a matrix function $Q(z)$, analytic on $|z| < \rho$, and satisfying

a) $\det Q(z) \neq 0 \quad , \quad |z| < 1,$

b) $Q(e^{-i\lambda}) \; \Sigma \; Q^*(e^{-i\lambda}) = P(e^{-i\lambda}) \; \Sigma \; P^*(e^{-i\lambda}) \quad , \; \lambda \in (-\pi \; , \; \pi]$

Furthermore, $Q(z)$ is uniquely determined except for a right-multiplicative constant $m \times m$ matrix H with $H\Sigma H^* = \Sigma$.
In the special case that $P(z)$ is a polynomial of degree p,

the matrix $Q(z)$ is also a polynomial of degree p.

PROOF. Since Σ can be decomposed into TT^* and the (non singular) matrix T can be absorbed into $P(z)$ (leaving the degree invariant), it is no restriction to take $\Sigma = I_m$, the $m \times m$ unit matrix.

Existence. If $\det P(z) \neq 0$, $|z| < 1$ we can take $Q(z) = P(z)$. Thus suppose $\det P(z_o) = 0$ for some $z_o \in \mathbb{C}$ with $0 < |z_o| < 1$, and let $c_1, \ldots, c_k$ be an orthonormal basis for $\ker P(z_o)$. The matrix with columns $c_1, \ldots, c_k$ is denoted by C. Let $d_1, \ldots, d_{m-k}$ be an orthonormal basis for $[\ker P(z_o)]^{\perp}$, the orthogonal complement of $\ker P(z_o)$ and denote the matrix with columns $d_1, \ldots, d_{m-k}$ by D. Put $U = [C ; D]$; then we have

$$U^* U = U U^* = I_m \quad . \tag{3.2.1}$$

Since $P_o(z_o) C = 0$ we can write

$$P(z) C = (z - z_o) P_1(z) \quad , \quad z \in \mathbb{C} \ ,$$

where $P_1(z)$ is a $m \times k$ matrix function. Hence

$$P(z) U = [(z - z_o) P_1(z) \ ; \ P(z) D] \ , \ z \in \mathbb{C},$$

Post multiplying by U^* and using (3.2.1) yields

$$P(z) = [(z - z_o) P_1(z) \ ; \ P(z) D] U^* \ , \ z \in \mathbb{C} \ .$$

Put

$$Q_o(z) := [(z \bar{z}_o - 1) P_1(z) \ ; \ P(z) D] U^* \ , z \in \mathbb{C}.$$

Then we have

$$\det Q_o(z) = (z \bar{z}_o - 1)^k \det [P_1(z) \ ; \ P(z) D] \det U^* =$$

$$= \left(\frac{z \bar{z}_o - 1}{z - z_o} \right)^k \det P(z) \quad , \ z \neq z_o \ .$$

Thus, ignoring multiplicities, det $Q_o(z)$ has at least one zero less than det $P(z)$ in $|z| < 1$.
Furthermore

$$Q_o(e^{-i\lambda})Q_o^*(e^{-i\lambda}) = |e^{-i\lambda}\bar{z}_o - 1|^2 P_1(e^{-i\lambda})P_1^*(e^{-i\lambda}) + P(e^{-i\lambda})DD^*P^*(e^{-i\lambda})$$

$$= |e^{-i\lambda} - z_o|^2 P_1(e^{-i\lambda})P_1^*(e^{-i\lambda}) + P(e^{-i\lambda})DD^*P^*(e^{-i\lambda})$$

$$= P(e^{-i\lambda})P^*(e^{-i\lambda}) \qquad , \quad \lambda \in (-\pi , \pi] ,$$

so that $Q_o(z)$ satisfies condition b). The procedure can be repeated for every zero of det $P(z)$ that lies inside the unit circle and since there are only a finite number of them we can find in finitely many steps an analytic matrix $Q(z)$ satisfying a) and b). It follows from the construction that if $P(z)$ is a matrix polynomial of degree p, then $Q(z)$ is also a matrix polynomial of degree p.

Uniqueness Let $\tilde{Q}(z)$ be a matrix function, analytic on $|z| < \rho$ satisfying a) and b). Then we have to prove $\tilde{Q}(z) = Q(z)\,H$ for some unitary (constant) matrix H. Let z_o be a zero of det $P(z)$ with $|z_o| = 1$ (if any). Then also det $\tilde{Q}(z)$ and det $Q(z)$ have that zero and b) implies

$$\ker \tilde{Q}^*(z_o) = \ker Q^*(z_o) \ .$$

Thus we can write as before

$$\tilde{Q}^*(z) = [\,\{(z - z_o)\,\tilde{Q}_1(z)\}^* \ ; \ \tilde{Q}^*(z)\,D]\,U^* \qquad , \ z \in \mathbb{C}$$

and

$$Q^*(z) = [\,\{(z - z_o)\,Q_1(z)\}^* \ ; \ Q^*(z)\,D]\,U^* \qquad , \ z \in \mathbb{C}$$

where $\tilde{Q}_1(z)$ and $Q_1(z)$ are now $k \times m$ matrix-functions. Put

$$\tilde{q}(z) = \begin{bmatrix} \tilde{Q}_1(z) \\ D^* \tilde{Q}(z) \end{bmatrix} \quad , z \in \mathbb{C}$$

and

$$q(z) = \begin{bmatrix} Q_1(z) \\ D^* Q(z) \end{bmatrix} \quad , z \in \mathbb{C} \quad .$$

Then we have $\tilde{Q}(z) = W(z)\tilde{q}(z)$ and $Q(z) = W(z)q(z)$ where

$$W(z) := U \begin{bmatrix} (z - z_o) I_k & \vdots & 0 \\ \cdots\cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & I_{m-k} \end{bmatrix} \quad , z \in \mathbb{C} \quad .$$

Thus for all z we have

$$\tilde{q}(z) = \lim_{\eta \to z} W^{-1}(\eta)\tilde{Q}(z) \quad , \quad q(z) = \lim_{\xi \to z} W^{-1}(\xi)Q(z).$$

Hence

$$\tilde{q}(e^{-i\lambda})\ \tilde{q}^*(e^{-i\lambda}) = \lim_{\eta \to e^{-i\lambda}} W^{-1}(\eta)\ \tilde{Q}(e^{-i\lambda})\ \tilde{Q}^*(e^{-i\lambda})\,(W^{-1}(\eta))^* =$$

$$= \lim_{\xi \to e^{-i\lambda}} W^{-1}(\xi)\ Q(e^{-i\lambda})\ Q^*(e^{-i\lambda})\,(W^{-1}(\xi))^* =$$

$$= q(e^{-i\lambda})\ q^*(e^{-i\lambda}) \quad , \quad \lambda \in (-\pi, \pi] \quad .$$

Thus we have $\ker \tilde{q}^*(z_o) = \ker q^*(z_o)$ and the multiplicity of the zero z_o of $\det \tilde{q}(z)$ (or $\det q(z)$) is k units less than the multiplicity of z_o as a zero of $\det \tilde{Q}(z)$ (or $\det Q(z)$). In this way we can remove all zeros from the unitcircle and therefore it is no restriction to assume $\tilde{Q}(z)$ and $Q(z)$ non singular on $|z| = 1$.

Consider

$$V(z) := Q^{-1}(z)\ \tilde{Q}(z) \quad .$$

Using b) we have

$$V(e^{-i\lambda})\ V^{*}(e^{-i\lambda}) = Q^{-1}(e^{-i\lambda})\ \tilde{Q}(e^{-i\lambda})\ \tilde{Q}^{*}(e^{-i\lambda})\ Q^{-1*}(e^{-i\lambda}) =$$

$$= Q^{-1}(e^{-i\lambda})\ Q(e^{-i\lambda})\ Q^{*}(e^{-i}\ )\ Q^{-1*}(e^{-i}\ ) = I_m$$

or, equivalently, $V(z)$ is unitary on $|z| = 1$. Furthermore $V(z)$ is analytic on $|z| < \rho_o$ for some $\rho_o > 1$ and nonsingular there. Therefore both $V(z)$ and $V^{-1}(z)$ can be expanded into a power-series

$$V(z) = \sum_{k=0}^{\infty} M_k\ z^k \quad , \quad V^{-1}(z) = \sum_{k=0}^{\infty} N_k\ z^k \quad , \quad |z| < \rho_o \quad .$$

Since both series converge on the unit circle and $V(e^{-i\lambda})$ is unitary we have

$$V^{-1}(e^{-i\lambda}) = \sum_{k=0}^{\infty} N_k\ e^{-ik\lambda} = V^{*}(e^{-i\lambda}) = \sum_{k=0}^{\infty} M_k^{*}\ e^{\ ik\lambda}, \ \lambda \in (-\pi, \pi] .$$

As in the univariate case we obtain

$$N_o = M_o^{*}$$

$$N_k = M_k = 0, \qquad k \geq 1$$

by simply equating coefficients. It follows that $V(z) = N_o$ or equivalently $\tilde{Q}(z) = Q(z)\ N_o$ with N_o unitary. This proves the lemma. □

REMARK 1. It should be noted that if $p(0)$ is prescribed (e.g. I_m), then $Q(z)$ is uniquely determined.

REMARK 2. If $P(0) = Q(0) = I_m$, if follows from the construction given in the proof, that although det $P(z)$ and det $Q(z)$ have different zeros, they have the same nullspaces. Furthermore, if $P(z)$ is a polynomial of degree p, the coefficients $Q_1 \ldots Q_p$ of $Q(z)$ depend continuously on the coefficients $P_1 \ldots P_p$ of $P(z)$.

In the following section we shall use this lemma to solve the problem of minimum informative sample size for the m-variate MA process. Furthermore it is the basis for finding informative sample sizes for a range of models with MA errors.

REMARK. In m-variate models "sample size" must be interpreted as the number of points in time the observable m-variate process is observed; when the sample size is k, then mk random variables are observed.

We conclude this section with some special matrix theory. Let $P(z)$, $H(z)$ and $Q(z)$ be matrix polynomials. If $P(z) = H(z)\,Q(z)$ $z \in \mathbb{C}$, then $H(z)$ is called a *left factor (-divisor)* of $P(z)$ and $P(z)$ is called a *right multiple* of $H(z)$ (similar for right factors and left multiples). Two polynomials $P(z)$ and $\tilde{P}(z)$ have a common left factor if they can be written $P(z) = H(z)\,Q(z)$ and $\tilde{P}(z) = H(z)\,\tilde{Q}(z)$. The matrix polynomial $H(z)$ is called a greatest common left divisor (g.c.l.d.) for $P(z)$ and $Q(z)$ if any other common left factor has $H(z)$ as a right multiple. If $H_o(z)$ is another g.c.l.d. then we have $H_o(z) = H(z)\,U(z)$ where $U(z)$ is *unimodular* (Mc DUFFEE [10] p.35) A matrix polynomial $U(z)$ is called unimodular if it has constant determinant. A simple example is

$$U(z) = \begin{bmatrix} 1 & z \\ 0 & 2 \end{bmatrix} .$$

The following lemma will be useful .

LEMMA 3.2.2 Let $v, w \in \mathbb{C}^m$ be orthogonal ($v^* w = 0$). Then

$$\det (I_m + w\ v^*\ z) = 1 , \quad z \in \mathbb{C} .$$

PROOF. Using $v^* w = 0$ we have

$$(I_m + w\ v^*\ z)(I_m - w\ v^*\ z) = I_m - w\ v^*\ z\ I_m + $$

$$+ w\ v^* z I_m - w\ v^*\ w\ v^*\ z^2 = I_m$$

$$z \in \mathbb{C} .$$

Hence $(I_m + w\ v^*\ z)^{-1} = (I_m - w\ v^*\ z)$, and so

$$(3.2.2)\quad \det (I_m + w\ v^*\ z) = \{\det (I_m - w\ v^*\ z)\}^{-1}, \quad z \in \mathbb{C} .$$

Since both the determinants of $I_m + w\ v^*\ z$ and $I_m - w\ v^*\ z$ are either constant or a polynomial in z, (3.2.2) implies that they must be constant. Taking $z = 0$ it follows that they must be equal to 1. □

From (3.1.2) it can be seen that in the ARMA (p,q) case we can not expect to find second-order informative sample sizes if the matrix polynomials A(z) and B(z) have a common left factor. Therefore the factorization of matrix polynomials deserves some attention. However, it will be clear now that matrixpolynomials don't factorize as simply as scalar polynomials do; e.g. $\det A(z_o) = \det B(z_o) = 0$ for some $z_o \in \mathbb{C}$ does not imply that A(z) and B(z) have a (non trivial) common factor. As the following two lemmas show, the corresponding nullspaces of $A^*(z_o)$ and $B^*(z_o)$ play a role.

<u>LEMMA 3.2.3</u> Let $P, Q \in \mathbb{C}(m)$. Then $\ker P^* \cap \ker Q^* = \{0\}$ iff $r[P\ ;\ Q] = m$.

PROOF. We have

$$r\,[P\ ;\ Q] = m \iff c^*\,[P\ ;\ Q] \neq 0, \qquad c \in \mathbb{C}^m,\ c \neq 0 \iff$$

$$\iff c^*\,P \neq 0 \vee c^*\,Q \neq 0\ , \qquad c \in \mathbb{C}^m,\ c \neq 0 \iff$$

$$\iff P^*\,c \neq 0 \vee Q^*\,c \neq 0\ , \qquad c \in \mathbb{C}^m,\ c \neq 0 \iff$$

$$\iff \ker P^* \cap \ker Q^* = \{0\}\ .\quad \square$$

LEMMA 3.2.4 Two arbitrary m × m matrix polynomials A(z) and B(z) have a non-unimodular common left factor iff $r\,[A(z_o)\ ;\ B(z_o)] < m$ for some $z_o \in \mathbb{C}$.

PROOF. (Only if) Let H(z) be a non-unimodular common left factor. Then det H(z) is a polynomial and so there exist a $z_o \in \mathbb{C}$ with $\det H(z_o) = 0$.
Hence there exist a $v \in \mathbb{C}^m$ such that $v^*\,H(z_o) = 0$ and so $v^*\,[A(z_o)\ ;\ B(z_o)] = 0$. This implies that $r\,[A(z_o)\ ;\ B(z_o)] < m$.
(If) let $r\,[A(z_o)\ ;\ B(z_o)] < m$. Then by lemma 3.2.3 follows that the null spaces of $A^*(z_o)$ and $B^*(z_o)$ have non-null intersection. Let C be an m × c matrix the columns of which are an orthonormal basis for $\ker A^*(z_o) \cap \ker B^*(z_o)$, and let D be an m × (m-c) matrix with columns an orthonormal basis for $(\ker A^*(z_o) \cap \ker B^*(z_o))^{\perp}$.
Using the same partitioning as in the proof of the uniqueness part of lemma 3.2.1 we can write

$$U^*\,A(z) = \begin{bmatrix} (z - z_o)\,A_1(z) \\ \\ D^*A(z) \end{bmatrix},\quad U^*\,B(z) = \begin{bmatrix} (z - z_o)\,B_1(z) \\ \\ D^*B(z) \end{bmatrix},\quad z \in \mathbb{C}$$

where $U := [C\ ;\ D]$, $UU^* = U^*\,U = I_m$.

Hence

$$A(z) = U \begin{bmatrix} (z - z_o)\, I_c & 0 \\ \hdashline 0 & I_{m-c} \end{bmatrix} \begin{bmatrix} A_1(z) \\ \\ D^*\, A(z) \end{bmatrix} \quad , \; z \in \mathbb{C}$$

and

$$B(z) = U \begin{bmatrix} (z - z_o)\, I_c & 0 \\ \hdashline 0 & I_{m-c} \end{bmatrix} \begin{bmatrix} B_1(z) \\ \\ D^*\, B(z) \end{bmatrix} \quad , \; z \in \mathbb{C}$$

which proves the lemma . □

COROLLARY. The $m \times m$ matrix polynomials $A(z)$ and $B(z)$ have I_m as a g.c.l.d. iff $r\,[A(z)\ ;\ B(z)] = m$, for all $z \in \mathbb{C}$.

Let $D(z)$ be a rational $m \times m$ matrix function (i.e. a matrix with rational functions as elements). Then $D(z)$ can be written as $A^{-1}(z)\, B(z)$ where $A(z)$ and $B(z)$ are $m \times m$ matrix polynomials. We shall say that $D(z)$ is decomposed into a *right numerator* $B(z)$ and a *left denominator* $A(z)$. Clearly there also exists a decomposition $P(z)\, Q^{-1}(z)$ where $P(z)$ is called a *left numerator* and $Q(z)$ a *right denominator*, but we shall not consider such decompositions. The following lemma gives conditions under which the decomposition of a rational matrix function into a right numerator and a left denominator is unique.

LEMMA 3.2.5 Let the rational $m \times m$ matrix function $D(z)$ decompose into the left denominator $A(z) = \sum_{k=0}^{p} A_k\, z^k$ and the right numerator $B(z) = \sum_{k=1}^{q} B_k\, z^k$, with $A(0) = I_m$, $r\,[A(z)\ ,\ B(z)] = m$ and $r\,[A_p\ ,\ B_q] = m$. Then $A(z)$ and $B(z)$ are uniquely determined by $D(z)$.

PROOF. Suppose there exist an alternative left denominator $\tilde{A}(z)$

and right numerator $\tilde{B}(z)$ satisfying the conditions of the theorem. Since $A(0) = \tilde{A}(0) = I_m$ the function

$$C(z) := \frac{1}{z} (A(z) - \tilde{A}(z))$$

is a matrix polynomial of degree $\leq p-1$. Furthermore

$$C(z)\ D(z) = \frac{1}{z} (B(z) - \tilde{B}(z)) \ ,$$

and since $B(0) = \tilde{B}(0) = D(0)$ the matrix $C(z)\ D(z)$ is a polynomial of degree $\leq q-1$.
As $\det A(z)$ does not vanish identically since $A(0) = I_m$, $\det A(z)$ is a polynomial. Denote the degree of $\det A(z)$ by d. Put

$$K(z) \quad = (\det A(z))\ A^{-1}(z) \ .$$

Then $K(z)$ is a matrix polynomial in z, and so $S(z) := C(z)\ K(z)$ is a polynomial. If s denotes the degree of $S(z)$ we may write

$$S(z) = \sum_{k=0}^{s} S_k\ z^k \ , \quad S_s \neq 0 \ .$$

1) Suppose $d \leq s$. Since $S(z)\ B(z) = C(z)\ D(z) \det A(z)$ is a matrix-polynomial of degree at most $q-1 + d$ we must have

$$\sum_{k=0}^{s} S_k\ B_{j-k} = 0 \ , \qquad j = q + d \ , q + d + 1 \ ,...$$

Similarly for $S(z)\ A(z) = C(z) \det A(z)$ we obtain

$$\sum_{k=0}^{s} S_k\ A_{n-k} = 0 \ , \qquad n = p + d \ , p + d + 1 \ ,...$$

Choosing $j = g + s$ and $n = p + s$ yields

$$S_s\ B_q = S_s\ A_p = 0 \ ,$$

which contradicts $r\,[A_p \ , B_q] = m$.

2) If $d \geq s+1$ there exist at least one zero, say z_o , of det $A(z)$ such that $S(z_o) \neq 0$.
Since $(\det A(z))^{-1} S(z) B(z) = C(z) D(z)$ is a matrix-polynomial it follows that we must have

$$S(z_o)\, B(z_o) = 0 \quad . \tag{3.2.3}$$

We also have

$$S(z_o)\, A(z_o) = C(z_o) \det A(z_o) = 0 \quad . \tag{3.2.4}$$

Since (3.2.3) and (3.2.4) contradict the condition
$r\,[A(z)\ ,\ B(z)] = m$, the lemma is proved . □

The following lemma gives a usefull criterion for divisibility.

LEMMA 3.2.6 *(Generalized BÉZOUT - theorem)*

Let $G(z) = \sum_{k=0}^{g} G_k\, z^k$ be an arbitrary $m \times m$ matrix-polynomial and A an arbitrary $m \times m$ matrix. Then $G(z)$ is divisible by the binomial $z\, I_m - A$ on the right (left) iff

$$G_r(A) := \sum_{k=0}^{g} G_k\, A^k = 0 \qquad (G_l(A) := \sum_{k=0}^{g} A^k\, G_k = 0\ ).$$

PROOF. see GANTMACHER [15] p. 81 . □

REMARK. In general we have $G_r(A) \neq G_l(A)$ since A and G_k need not commute for $k = 0, 1, \ldots, g$. Therefore divisibility on the right (left) does not imply divisibility on the left (right).
As an application of the generalized BÉZOUT theorem we shall prove a lemma that is important for the ARMA(p,q) case.

LEMMA 3.2.7 Let A and $\tilde{A}$ be two arbitrary $m \times m$ matrices. Then

there exist $m \times m$ matrices C_0 , C_1 ..., C_m such that

$$A^{m+1} = \sum_{k=0}^{m} C_k A^k \quad \text{and} \quad \tilde{A}^{m+1} = \sum_{k=0}^{m} C_k \tilde{A}^k \ .$$

PROOF. Put

$$G(z) = \{\det(z\, I_m - A)\}(z\, I_m - \tilde{A}) \quad , \qquad z \in \mathbb{C} \ .$$

Since $\det(z\, I_m - A)$ is a scalar polynomial of degree exactly m, it follows that $G(z)$ is a matrix polynomial of degree exactly $m+1$, and so it can be written as

$$(3.2.5) \qquad G(z) = I_m\, z^{m+1} - \sum_{k=0}^{m} C_k\, z^k \ , \qquad z \in \mathbb{C} \ .$$

We also have

$$G(z) = (z\, I_m - \tilde{A})\ \det\,(z\, I_m - A) =$$

$$= (z\, I_m - \tilde{A})\ \{\det\,(z\, I_m - A)\}\ (z\, I_m - A)^{-1}\ (z\, I_m - A) =$$

$$= (z\, I_m - \tilde{A})\ \{\,\mathrm{Adj}(z\, I_m - A)\}\ (z\, I_m - A) \quad , \quad z \in \mathbb{C}$$

where $\mathrm{Adj}(M)$ denotes the adjoint of M.
Thus $G(z)$ is a matrix polynomial of degree not exceeding $m+1$ with I_m as leading coefficient matrix, that is divisible on the right by $z\, I_m - \tilde{A}$ as well as by $z\, I_m - A$. But then it follows from the generalized BÉZOUT theorem that we must have

$$G_r(A) = G_r(\tilde{A}) = 0.$$

By (3.2.5) this implies

$$A^{m+1} = \sum_{k=0}^{m} C_k A^k \ , \qquad \tilde{A}^{m+1} = \sum_{k=0}^{m} C_k \tilde{A}^k \ ,$$

and the lemma is proved. □

Finally we state a lemma that is the m-variate generalization of the theorem of FEJÉR and RIESZ (lemma 2.1.3) .

Lemma 3.2.8 If

$$f(\lambda) := \frac{1}{2\pi} \sum_{k=-q}^{q} \Gamma_k e^{-ik\lambda} \geq 0 \quad , \lambda \in (-\pi , \pi]$$

and $f(\lambda)$ has a determinant not identically zero, then $f(\lambda)$ can be represented in the form

$$f(\lambda) = \frac{1}{2\pi} \{ \sum_{k=0}^{q} A_k e^{-ik\lambda} \} \{ \sum_{k=0}^{q} A_k e^{-ik\lambda} \}^{*} \quad ,$$

where A_o is Hermitian positive definite and no zeros of $\det \sum_{k=0}^{q} A_k z^k$ lie inside the unit circle. The representation is unique.

PROOF. See HANNAN [20] p. 64 - 65 for the construction of a factorization. The uniqueness follows from lemma 3.2.1 □

REMARK. An explicit expression for $\sum_{k=0}^{q} A_k z^k$ as in the univariate case (2.2.3) is not known. However, it follows from the construction in the first part of HANNAN'S proof, that there exists a factorization such that $A_o \ldots A_q$ depend continuously on $\Gamma_o \ldots \Gamma_q$.

By remark 2 on the fundamental lemma 3.2.1 it then follows that also for the unique factorization $A_o \ldots A_q$ depend continuously on $\Gamma_o \ldots \Gamma_q$.

3.3 THE MINIMUM INFORMATIVE SAMPLE SIZE FOR MULTIVARIATE MA(q) PROCESSES

Suppose we are sampling from the m-variate MA(q) process

$$\underline{x}_t = \sum_{k=0}^{q} B_k \, \underline{\varepsilon}_{t-k} \quad , \qquad t = 0, \pm 1, \ldots \qquad B_o = I_m \, .$$

Similar to the univariate case we introduce

$$S_q^{MA}(m) \; : = \{(B_1 \, , \ldots , \, B_q) \mid \det B(z) \neq 0, \quad |z| < 1\} \, .$$

Let $\mathbb{C}^+(m)$ denote the set of all positive definite hermitian $m \times m$ matrices. Then Σ_ε is said to be *completely unknown* if the range of Σ_ε is equal to $\mathbb{C}^+(m)$, i.e. if $\Sigma_\varepsilon(Z) = \mathbb{C}^+(m)$.

THEOREM 3.3.1 If in the m-variate MA(q) case $\theta = S_p^{MA}(m) \times Z$ and $\Sigma_\varepsilon(Z) \subset \mathbb{C}_m^+$ then the sample size q+1 is second-order informative for $(B(z), \, \Sigma_\varepsilon)$.

PROOF. let Γ_s denote the covariance function of the process i.e.

$$\Gamma_s \; : = E \, \{\underline{x}_t \, \underline{x}_{t-s}^*\} \, , \qquad t \, , \, s = 0, \pm 1, \ldots$$

Since this function vanishes for $|s| \geq q+1$, in the same way as in the univariate case it is sufficient to prove that the spectral density matrix $f_\theta(\lambda)$ has a unique decomposition of the form

$$(3.3.1) \qquad f_\theta(\lambda) = \frac{1}{2\pi} B(e^{-i\lambda}) \, \Sigma_\varepsilon \, B^*(e^{-i\lambda}) \, , \qquad \lambda \in (-\pi, \, \pi \,].$$

Suppose there exist $\tilde{B}(z)$ and $\tilde{\Sigma}_\varepsilon$ such that

$$(3.3.2) \qquad f_\theta(\lambda) = \frac{1}{2\pi} \tilde{B}(e^{-i\lambda}) \, \tilde{\Sigma}_\varepsilon \, \tilde{B}^*(e^{-i\lambda}) \, , \qquad \lambda \in (-\pi, \, \pi \,].$$

As Σ_ε and $\tilde{\Sigma}_\varepsilon$ are nonsingular there exists a (nonsingular) $m \times m$ matrix V such that

$$\Sigma_\varepsilon = V\ \tilde{\Sigma}_\varepsilon\ V^* \quad .$$

Since $\det B(z) \neq 0$, $|z| < 1$ and $\det \tilde{B}(z) \neq 0$ for $|z| < 1$, it follows from lemma 3.2.1 that there exists a matrix H (which does not depend on z) such that

$$B(z)\ V\ H = \tilde{B}(z) \quad , \qquad z \in \mathbb{C}.$$

Substituting $z = 0$ we obtain

$$B(0)\ VH = VH = \tilde{B}(0) = I_m \ .$$

Hence $B(z) = \tilde{B}(z)$, $z \in \mathbb{C}$ and so (by (3.3.1) and (3.3.3)) $\Sigma_\varepsilon = \tilde{\Sigma}_\varepsilon$. This proves the sample size q+1 to be second-order informative. □

REMARK 1. It is noteworthy that the distribution of a sample of size q+1 is in fact an m(q+1)-dimensional distribution, and, since the total number of unknown parameters equals $m^2q + \frac{1}{2}m(m+1)$, for $m > 1$ one hase identification even though the number of observations is less than the number of unknown parameters.

REMARK 2. That the theorem does not hold in general if Σ_ε is allowed to be singular, can be seen as follows. Let Σ_ε be singular and $v \in \mathbb{C}^m$ such that $\Sigma_\varepsilon\ v = 0$. Let $w \in \mathbb{C}^m$ be orthogonal to v. Then by lemma 3.2.2 we have

$$\det\ (I_m + w\ v^*\ z) = 1 \quad , \quad z \in \mathbb{C} \ .$$

We also have

$$(3.3.3) \qquad (I_m + w\ v^*\ z)\ \Sigma_\varepsilon = \Sigma_\varepsilon\ (I_m + v\ w^*\ z) = \Sigma_\varepsilon \ , \quad z \in \mathbb{C}.$$

Put $B_q = 0$ and $\tilde{B}(z)$ $(I_m + w\ v^*\ z)$. Then $\tilde{B}(z)$ is a matrixpolynomial of degree at most q, and we have by (3.3.3)

$$\tilde{B}(e^{-i\lambda})\Sigma_\varepsilon \tilde{B}^*(e^{-i\lambda}) = B(e^{-i\lambda})(I_m + w\ v^* e^{-i\lambda})\Sigma_\varepsilon (I_m + v\ w^* e^{-i\lambda})B^*(e^{-i\lambda})$$

$$= B(e^{-i\lambda})\Sigma_\varepsilon\ B^*(e^{-i\lambda}) \quad , \quad \lambda \in (-\pi\ ,\ \pi] \ .$$

Thus $(B(z)\ ,\ \Sigma_\varepsilon)$ generates the same covariances as $(\tilde{B}(z)\ ,\ \Sigma_\varepsilon)$.

The following theorem shows that under the additional condition that Σ_ε is completely unknown, the sample size q+1 is the minimum second-order informative sample size. As in the univariate case we shall prove a much stronger result on nowhere local identifiability. Local identifiability here refers to the *Euclidean* topology on $\mathbb{C}^{m^2(q+1)}$ restricted to $S_q^{MA}(m) \times \mathbb{C}^+(m)$ i.e. by considering $(B_1 \ldots B_q\ ,\ \Sigma_\varepsilon)$ as vector in $\mathbb{C}^{m^2(q+1)}$.
As in the univariate case we need a lemma.

LEMMA 3.3.2 let $D := \{(B_1 \ldots B_q) \mid \det B(z) \neq 0\ ,\ |z| \leq 1\}$
Then D is a dense subset of $S_q^{MA}(m)$.

PROOF. see DEISTLER, DUNSMUIR and HANNAN [8] , lemma 1. □

THEOREM 3.3.3 If in the m-variate MA(q) case $\theta = S_q^{MA}(m) \times Z$ and Σ_ε is completely unknown, then the sample size q is nowhere locally second-order informative for $(B(z),\ \Sigma_\varepsilon)$.

PROOF. Since D is dense in $S_q^{MA}(m)$ by lemma 3.3.2, $D \times \mathbb{C}^+(m)$ is dense in $S_q^{MA}(m) \times \mathbb{C}^+(m)$ and so by lemma 1.9.2 it is sufficient to prove the theorem for $\theta = D \times Z$.
Let $(B(z)\ ,\ \Sigma_\varepsilon)$ be such that B(z) is nonsingular on $|z| \leq 1$.

Then the theorem is proved if we can find a sequence $(B^{(n)}(z)\ ,\ \Sigma_\varepsilon^{(n)})$ converging to $(B(z)\ ,\ \Sigma_\varepsilon)$ (convergence of coefficients) such that for all n one has $\det B^{(n)}(z) \neq 0$ for $|z| \leq 1$, $(B^{(n)}(z)\ ,\ \Sigma_\varepsilon^{(n)})$ is not identical to $(B(z)\ ,\ \Sigma_\varepsilon)$, and $(B^{(n)}(z)\ ,\ \Sigma_\varepsilon^{(n)})$ generates the same covariances Γ_s $(s = 0, \pm 1, \ldots \pm (q-1))$ as $(B(z)\ ,\ \Sigma_\varepsilon)$.

As in the univariate case (theorem 2.2.4) we introduce the covariance generating function

$$G(z) := \sum_{s=-q}^{q} \Gamma_s z^{-s} \quad , \qquad z \neq 0.$$

Similar to the univariate case we obtain

$$G(z) = B(z)\ \Sigma_\varepsilon\ B^*(\frac{1}{\bar{z}}) \quad , \qquad z \neq 0.$$

Note, that

$$G(e^{-i\lambda}) = B(e^{-i\lambda})\ \Sigma_\varepsilon\ B^*(e^{-i\lambda}) \in \mathbb{C}^+(m) \quad , \qquad \lambda \in (-\pi\ ,\pi\ ]$$

since $B(z)$ is nonsingular on $|z| = 1$. Thus we may define $\delta > 0$ by

$$\delta = \min_{\substack{\lambda\in(-\pi,\pi] \\ a\in\mathbb{C}^m, a\neq 0}} \frac{a^*\ G(e^{-i\lambda})\ a}{a^*\ a}\ .$$

Let $n > \frac{1}{\delta}$ and put

$$G^{(n)}(z) = G(z) + \frac{1}{4n}\ I_m\ z^q + \frac{1}{4n}\ I_m\ z^{-q} \qquad , z \neq 0\ .$$

Then we have

$$G^{(n)}(e^{-i\lambda}) = G(e^{-i\lambda}) + \frac{1}{2n}\ I_m \cos q\ \lambda \qquad , \lambda \in (-\pi\ ,\pi]$$

and so for all vectors $a \neq 0$ and $\lambda \in (-\pi\ ,\pi]$

$$a^* G^{(n)}(e^{-i\lambda})a = a^*a\left\{\frac{a^*G(e^{-i\lambda})a}{a^* a} + \frac{1}{2n}\cos q\ \lambda\right\} \geq$$

$$\geq a^*a\ (\delta - \tfrac{1}{2}\ \delta) = \tfrac{1}{2}\ \delta\ a^*a > 0.$$

Thus $G^{(n)}(e^{-i\lambda}) \in \mathbb{C}^+(m)$. Furthermore

$$(3.3.4)\quad \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{is\lambda}\ G^{(n)}(e^{-i\lambda})d\lambda = \begin{cases} \Gamma_s & ,|s|= 0,\ 1,\ldots\ q-1 \\ \Gamma_s+\frac{1}{4n}\ I_m & ,|s|= q \\ 0 & ,|s|= q+1\ ,\ 1+2,\ldots \end{cases}$$

Hence by lemma 3.2.8 it follows that $G^{(n)}(e^{-i\lambda})$ can be decomposed into

$$B^{(n)}(e^{-i\lambda})\ \Sigma_\varepsilon^{(n)}\ B^{(n)*}(e^{-i\lambda}) \quad , \qquad \lambda \in (-\pi\ ,\pi]$$

where $B^{(n)}(z)$ is a matrix polynomial of degree $\leq q$ with $\det B^{(n)}(z) \neq 0$, $|z| \leq 1$. Since Σ_ε is completely unknown, $\Sigma_\varepsilon^{(n)}$ may be chosen to obtain $B^{(n)}(0) = I_m$. Finally, it follows from (3.3.4) and the remark following lemma 3.2.8 that we must have

$$(3.3.5)\quad \lim_{n\to\infty} B^{(n)}(z)\ \Sigma_\varepsilon^{(n)\frac{1}{2}} = B(z)\ \Sigma_\varepsilon^{\frac{1}{2}} \quad , \quad z \in \mathbb{C}$$

where $\Sigma_\varepsilon^{(n)\frac{1}{2}}$ and $\Sigma_\varepsilon^{\frac{1}{2}}$ denote the (unique) hermitian positive definite square roots of $\Sigma_\varepsilon^{(n)}$ and Σ_ε respectively. Taking $z = 0$ we obtain

$$\lim_{n\to\infty} \Sigma_\varepsilon^{(n)} = \Sigma_\varepsilon \ ,$$

and so with (3.3.5)

$$\lim_{n\to\infty} B^{(n)}(z) = B(z) \quad , \qquad z \in \mathbb{C} \quad .$$

Thus we found a sequence $(B^{(n)}(z)\ ,\ \Sigma_\varepsilon^{(n)})$ converging to $(B(z)\ ,\ \Sigma_\varepsilon)$ with the desired properties.
This completes the proof. □

3.4 INFORMATIVE SAMPLES FROM MULTIVARIATE AR(p) PROCESSES

For the case that we are sampling from a process $\{\underline{x}_t\}$ satisfying the m-variate vectorial difference equation

$$(3.4.1) \qquad \sum_{k=0}^{p} A_k\, \underline{x}_{t-k} = \underline{\varepsilon}_t\ , \qquad t = 0, \pm 1, \ldots\ , A_o = I_m$$

we introduce

$$S_p^{AR}(m) := \{(A_1\ ,\ldots,\ A_p) \mid \det A(z) \neq 0\ , |z| \leq 1\}\ .$$

we have

THEOREM 3.4.1 If in the m-variate AR(p) case $\theta = S_p^{AR}(m) \times Z$ and $\Sigma_\varepsilon(z) \subset \mathbb{C}^+(m)$, then the sample size p+1 is second-order informative for $(A(z)\ ,\ \Sigma_\varepsilon)$.

PROOF. Let Γ_s denote the covariance function of $\{\underline{x}_t\}$ i.e.

$$\Gamma_s := E\{\underline{x}_t\ \underline{x}_{t-s}^*\}\ , \qquad t,s = 0, \pm 1, \ldots$$

As in the univariate case (theorem 2.3.1) we may write down the YULE-WALKER equations

$$(3.4.2) \qquad \sum_{k=0}^{p} A_k\, \Gamma_{s-k} = \begin{cases} \Sigma_\varepsilon\ , & s = 0 \\ 0\ , & s = 1, 2, \ldots, p\ . \end{cases}$$

As $\Gamma_{-s} = \Gamma_s^*$, these equations can be written in matrix form as

$$(3.4.3) \qquad (A_1, \ldots, A_p\ ,\ \Sigma_\varepsilon)\ W + (\Gamma_o, \ldots, \Gamma_p) = 0\ ,$$

where W is the m(p+1) × m(p+1) matrix given by

$$(3.4.4)\quad W = \begin{bmatrix} \Gamma_1^* & \Gamma_0 & \cdots & \Gamma_{p-1} \\ \Gamma_2^* & \Gamma_1^* & \cdots & \Gamma_{p-2} \\ \vdots & & & \vdots \\ \Gamma_p^* & \Gamma_{p-1}^* & \cdots & \Gamma_0^* \\ -I_m & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \Gamma_1^* & & & \\ \Gamma_2^* & & & \\ \vdots & & R^{(p)} & \\ \Gamma_p^* & & & \\ -I_m & 0 & \cdots & 0 \end{bmatrix} .$$

Suppose the sample size p+1 is not second-order informative for $(A(z), \Sigma_\varepsilon)$. Then there exist pairs $(A^{(1)}(z), \Sigma_\varepsilon^{(1)}) \neq (A^{(2)}(z), \Sigma_\varepsilon^{(2)})$ that generate the same Γ_s for $s = 0, \pm 1, \ldots \pm p$.
It follows from (3.4.3) that in that case W must be singular and by (3.4.4) this is equivalent to singularity of $R^{(p)}$.
Hence there exist vectors $c_j \in \mathbb{C}^m$, $j = 0, 1, \ldots p-1$ not all zero, such that

$$(3.4.5)\qquad \sum_{j=0}^{p-1} c_j^* \underline{x}_{t-j} = 0 \quad (a.s) \quad , t = 0, \pm 1, \ldots \quad .$$

Since the process $\{\underline{x}_t\}$ has the spectral density matrix

$$f(\lambda) := \frac{1}{2\pi} A^{-1}(e^{-i\lambda}) \, \Sigma_\varepsilon \, A^{-1*}(e^{-i\lambda}) \quad , \lambda \in (-\pi, \pi]$$

relation (3.4.5) implies

$$\left[\sum_{j=0}^{p-1} c_j^* e^{-ij\lambda}\right] f(\lambda) \left[\sum_{j=0}^{p-1} c_j^* e^{-ij\lambda}\right]^* = 0$$

a.e. w.r.t. Lebesgue measure. Since $f(\lambda)$ is a nowhere singular matrix it follows that

$$\sum_{j=0}^{p-1} c_j^* e^{-ij\lambda} = 0 \quad , \lambda \in (-\pi, \pi]$$

and so $c_j = 0$, $j = 0, 1, \ldots, p-1$. Thus we have a contradiction and the theorem is proved. □

REMARK. As in the MA-case the theorem does not hold in general when Σ_ε is allowed to be singular. To see this, take $v, w \in \mathbb{C}^m$ orthogonal such that $\Sigma_\varepsilon v = 0$ and put $A_p = 0$.
Let $\tilde{A}(z) := (I_m - w\, v^*\, z)\, A(z)$. Then we have

$$\tilde{A}^{-1}(e^{-i\lambda})\Sigma_\varepsilon \tilde{A}^{-1*}(e^{-i\lambda}) = A^{-1}(e^{-i\lambda})(I_m + wv^* e^{-i\lambda})\Sigma_\varepsilon (I_m + vw^* e^{i\lambda})A^{-1*}(e^{-i\lambda})$$

$$= A^{-1}(e^{-i\lambda})\Sigma_\varepsilon A^{-1*}(e^{-i\lambda}) \quad , \lambda \in (-\pi, \pi]$$

and so $(A(z), \Sigma_\varepsilon)$ generates the same covariances as $(\tilde{A}(z), \Sigma_\varepsilon)$.

3.5 INFORMATIVE SAMPLES FROM MULTIVARIATE ARMA(p,q) PROCESSES; PREDICTABILITY.

So far the multivariate generalizations of the results in Chapter II where straightforward, except for some minor difficulties. However, the multivariate generalizations of theorems 2.4.2 and 2.5.2 for the ARMA case have given some real trouble, since it turned out that the 'obvious' generalizations are not true. To be more precise, suppose we are sampling from the model (3.1.1) with $p\, q \geq 1$. Put

$$S_{p,q}^{ARMA}(m) = \{(A_1 \ldots A_p, B_1 \ldots B_q) \mid \det A(z) \neq 0, z \leq 1; \det B(z) \neq 0, z < 1$$

$$r[A_p, B_q] = m \text{ and } r[A(z), B(z)] = m \ \forall z \ \}.$$

For the existence of second-order informative sample sizes for $(A(z), B(z), \Sigma_\varepsilon)$, the condition $r[A(z), B(z)] = m$ is clearly necessary as can be seen from (3.1.2) and lemma 3.2.4. In contrast to the univariate case we cannot drop the condition

$r\,[A_p\ ,\ B_q] = m$ either; if $r\,[A_p,\ B_q] < m$, there exists $v \in \mathbb{C}^m$ with $v^*A_p = v^*B_q = 0$, $v \neq 0$ and choosing $w \in \mathbb{C}^m$, $w \neq 0$ such that $w^*v = 0$, by lemma 3.2.2 we see that $I_m + w\ v^*z$ is a unimodular left factor that does not increase the degree of $A(z)$ and $B(z)$. Thus the 'obvious' generalization of theorem 2.4.2 is: If $\theta = S_{p,q}^{ARMA}(m) \times Z$ and $\Sigma_\varepsilon(Z) \subset \mathbb{C}^+(m)$, then the sample size $p+q+1$ is second-order informative for $(A(z),\ B(z), \Sigma_\varepsilon)$. If Γ_s ,$s = 0, \pm 1, \ldots$ denotes the covariance function of the observable process, then a natural attempt to prove this, would be to write down the covariance equations similar to (2.4.3)

$$(3.5.1) \qquad \sum_{k=0}^{p} A_k\ \Gamma_{s-k} = 0 \qquad , s = q+1,\ q+2\ ,\ldots,$$

and to prove the equations for $s = q+1,\ q+2,\ \ldots\ q+p$ to be uniquely solvable for $A_1 \ldots A_p$. In matrix notation we have (compare (2.4.4))

$$[A_1 \ldots A_p]\ \tilde{R}_q + [\Gamma_{q+1}\ ,\ldots,\ \Gamma_{q+p}] = 0,$$

where $\tilde{R}_q$ is now the $mp \times mp$ matrix given by

$$(3.5.2)\quad \tilde{R}_q = \begin{bmatrix} \Gamma_q & \Gamma_{q+1} & \cdots & \Gamma_{q+p-1} \\ \Gamma_{q-1} & \Gamma_q & \cdots & \Gamma_{p+p-2} \\ \vdots & \vdots & & \vdots \\ \Gamma_{q-p+1} & \Gamma_{q-p+2} & \cdots & \Gamma_q \end{bmatrix} .$$

However, proving that $\tilde{R}_q$ is nonsingular in a similar way as in the univariate case is impossible because we do not have the multivariate analogue of lemma 2.4.1 for vectorvalued functions. Nevertheless HANNAN stated in 1975 ([19]) that $\tilde{R}_q$ is nonsingular,

but the proof given there is incorrect as was pointed out in TIGELAAR [34]. The example we shall present shows, in fact, that the statement is false. Before doing so we consider the problem of finding second-order informative sample sizes for the spectral measure, or, equivalently, second-order predictive sample sizes for the future. The 'obvious' generalization of theorem 2.5.2 is, that if $\theta = S_p^{AR}(m) \times \mathbb{C}^q \times Z$, then the sample size p+q+1 is second-order informative for the spectral measure. It turns out that the same counterexample proves this statement to be false, and it is instructive to see where the 'natural' attempt to prove it goes wrong. To see this, let $(A(z), B(z), \Sigma_\varepsilon) \neq (A(z), B(z), \tilde{\Sigma}_\varepsilon)$ and let $\{\Gamma_s \; s = 0, \pm 1, \ldots\}$ and $\{\tilde{\Gamma}_s \; s = 0, \pm 1, \ldots\}$ be the corresponding covariance functions. Then we have to prove the implication

$$[(\Gamma_0 \ldots \Gamma_{p+q}) = (\tilde{\Gamma}_0, \ldots \tilde{\Gamma}_{p+q})] \Rightarrow [\{\Gamma_s, s=0, \pm 1, \ldots\} = \{\tilde{\Gamma}_s, s=0, \pm 1, \ldots\}] .$$

The m-variate analogues of (2.5.2) and (2.5.3) are

$$(3.5.3) \quad \begin{cases} \sum_{k=0}^{p} A_k \, \Gamma_{s-k} = 0 \;, & s = q+1, q+2, \ldots \\ \sum_{k=0}^{p} \tilde{A}_k \, \tilde{\Gamma}_{s-k} = 0 \;, & s = q+1, q+2, \ldots \end{cases}$$

and

$$(3.5.4) \qquad \sum_{k=0}^{p} \tilde{A}_k \, \Gamma_{s-k} = 0 \;, \qquad s = q+1, q+2, \ldots q+p \;.$$

When we try to prove the validity of (3.5.4) for s = p+q+1 by the method of the univariate proof we must substitute

$$\Gamma_s = - \sum_{k=1}^{p} A_k \, \Gamma_{s-k} \;, \qquad s = q+1, q+2, \ldots$$

into

$$\sum_{k=0}^{p} \tilde{A}_k \, \Gamma_{q+p+1-k} \;,$$

to obtain

$$(3.5.5)\quad -\sum_{j=0}^{p}\tilde{A}_j\sum_{k=1}^{p}A_k\Gamma_{q+p+1-j-k}=-\sum_{k=1}^{p}\sum_{j=0}^{p}\tilde{A}_jA_k\Gamma_{q+p+1-j-k}.$$

Since in general the matrices $\tilde{A}_j$ and A_k do not commute, it is possible to put the last expression equal to the zero matrix as in the univariate case. It is, however, not necessary that the matrices $\tilde{A}_j$ and A_k commute for all j and k. To see this, let $f(\lambda)$ and $\tilde{f}(\lambda)$ denote the spectral density matrices corresponding to Γ_s and $\tilde{\Gamma}_s$ respectively.
Substituting for $\tilde{\Gamma}_s$ in (3.5.3) and (3.5.5) its spectral representation

$$\tilde{\Gamma}_s = \int_{-\pi}^{\pi} e^{is\lambda}\, f(\lambda)d\lambda\ , \qquad s = 0, \pm 1,\ldots,$$

we obtain

$$\int_{-\pi}^{\pi} e^{is\lambda}\tilde{A}(e^{-i\lambda})\, f(\lambda)d\lambda = 0\ ,\ s = q+1,\ q+2,\ \ldots,$$

and (using $\Gamma_s = \tilde{\Gamma}_s$, $s = 0, \ldots, q+p$)

$$\int_{-\pi}^{\pi} e^{i(q+p+1)\lambda}\tilde{A}(e^{-i\lambda})(I_m-A(e^{-i\lambda}))\tilde{f}(\lambda)d\lambda\ .$$

If in the last expression $\tilde{A}(e^{-i\lambda})$ and $I_m - A(e^{-i\lambda})$ may be interchanged, we obtain

$$\int_{-\pi}^{\pi} e^{i(p+q+1)\lambda}(I_m-A(e^{-i\lambda}))\ \tilde{A}(e^{-i\lambda})\tilde{f}(\lambda)d\lambda =$$

$$= -\sum_{k=1}^{p}A_k\int_{-\pi}^{\pi} e^{i(q+p+1-k)\lambda}\tilde{A}(e^{-i\lambda})\ \tilde{f}(\lambda)\ d\lambda = 0.$$

So it would be sufficient that the polynomials $A(z)$ and $\tilde{A}(z)$ commute on $|z| = 1$. However, they do not in general and a challenging problem arises. Before we try to find informative sample sizes we now present the counterexample. In fact we only prove the *existence* of a counterexample, and therefore we state the following theorem.

THEOREM 3.5.1 If in the bivariate ARMA(1,1) case $\theta = S^{ARMA}_{1,1}(2) \times Z$ and Σ_ε is completely unknown, then the sample size 3 is not second-order informative for the spectral measure (and so not for $(A(z), B(z), \Sigma_\varepsilon)$).

PROOF. Let ε and δ be positive numbers yet to be chosen. Consider the matrices

$$G(\lambda) := \begin{bmatrix} |1-\delta e^{-i\lambda}|^2 - 2\varepsilon\delta + \varepsilon e^{-i\lambda} + \varepsilon e^{i\lambda} & \varepsilon e^{i\lambda} - \varepsilon\delta + \varepsilon e^{-i\lambda} \\ \varepsilon e^{-i\lambda} - \varepsilon\delta + \varepsilon e^{i\lambda} & 1 + \varepsilon e^{-i\lambda} + \varepsilon e^{i\lambda} \end{bmatrix}, \tag{3.5.6}$$

and

$$\tilde{G}(\lambda) := \begin{bmatrix} 1 + \varepsilon e^{-i\lambda} + \varepsilon e^{i\lambda} - 2\delta\varepsilon + \delta^2 & \varepsilon e^{i\lambda} + \varepsilon e^{-i\lambda} - \delta e^{-i\lambda} - \delta\varepsilon \\ \varepsilon e^{-i\lambda} + \varepsilon e^{i\lambda} - \delta e^{i\lambda} - \delta\varepsilon & 1 + \varepsilon e^{-i\lambda} + \varepsilon e^{i\lambda} \end{bmatrix}. \tag{3.5.7}$$

Both $G(\lambda)$ and $\tilde{G}(\lambda)$ tend to I_2 (uniformly in λ) as $\varepsilon, \delta \to 0$ and so they are positive definite for all $\lambda \in (-\pi, \pi]$ if ε and δ are sufficiently small.

Introduce

$$A := \begin{bmatrix} \delta & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{A} := \begin{bmatrix} 0 & \delta \\ 0 & 0 \end{bmatrix},$$

and put $A(z) = I_2 - Az$, $\tilde{A}(z) = I_2 - \tilde{A}z$.
Then $\det A(z) = 1 - \delta z$, $\det \tilde{A}(z) = 1$, $z \in \mathbb{C}$ and so ε and δ may be chosen in such a way that

$$\begin{cases} \det A(z) \neq 0 \quad , |z| \leq 1 \\ \det \tilde{A}(z) \neq 0 \quad , |z| \leq 1 \\ G(\lambda) \text{ is positive definite for all } \lambda \\ \tilde{G}(\lambda) \text{ is positive definite for all } \lambda \end{cases}$$

hold simultaneously. Thus $G(\lambda)$ and $\tilde{G}(\lambda)$ are hermitian positive definite trigonometric matrix polynomials of degree 1. Hence by lemma 3.2.8 there exist matrix polynomials $B(z)$ and $\tilde{B}(z)$ both of degree 1 and hermitian positive definite matrices Σ and $\tilde{\Sigma}$ such that $B(0) = \tilde{B}(0) = I_2$,

$$G(\lambda) = B(e^{-i\lambda})\ \Sigma\ B^*(e^{-i\lambda}) \quad , \qquad \lambda \in (-\pi\ ,\ \pi] \tag{3.5.8}$$

$$\tilde{G}(\lambda) = \tilde{B}(e^{-i\lambda})\ \tilde{\Sigma}\ \tilde{B}^*(e^{-i\lambda}) \quad , \qquad \lambda \in (-\pi\ ,\ \pi] \tag{3.5.9}$$

and $\det B(z) \neq 0$, $|z| < 1$ $\det \tilde{B}(z) \neq 0$, $|z| < 1$.
Note, that Σ, $\tilde{\Sigma} \in \Sigma_\varepsilon(Z)$ since Σ_ε is completely unknown.
Let $B(z) = I_2 + Bz$ and $\tilde{B}(z) = I_2 + \tilde{B}z$. We shall first prove that $(A, B) \in S^{ARMA}_{1,1}(2)$ and $(\tilde{A}, \tilde{B}) \in S^{ARMA}_{1,1}(2)$.
Comparing coefficients of $e^{-i\lambda}$ in (3.5.6) and (3.5.8) we obtain

$$\begin{bmatrix} \varepsilon - \delta & \varepsilon \\ \varepsilon & \varepsilon \end{bmatrix} = \Sigma\ B \quad ,$$

and so $r[B] = 2$. Similarly from (3.5.7) and (3.5.9) it follows that $r[\tilde{B}] = 2$, and therefore

$$r[A, B] = r[\tilde{A}, \tilde{B}] = 2.$$

Furthermore $r[\tilde{A}(z)] = 2$, $z \in \mathbb{C}$ and so $r[\tilde{A}(z), \tilde{B}(z)] = 2$, $z \in \mathbb{C}$.
Remains to prove that $r[A(z), B(z)] = 2$, $z \in \mathbb{C}$.
If $r[A(z), B(z)] < 2$, then by lemma 3.2.4 $A(z)$ and $B(z)$ have a non unimodular left factor in common, and this can only be the case if $A(z) = B(z)$, $z \in \mathbb{C}$ since both have degree 1 and $A(0) = B(0) = I_2$. But then $B = A$ and so $r[B] = 1$ contradicting $r[B] = 2$.
Next we shall prove that $(A(z), B(z), \Sigma)$ and $(\tilde{A}(z), \tilde{B}(z), \tilde{\Sigma})$ generate the same covariances Γ_s for $s = 0, 1, 2$. Let f and $\tilde{f}$ denote the spectral density matrices corresponding to $(A(z), B(z), \Sigma)$ and $(\tilde{A}(z), \tilde{B}(z), \tilde{\Sigma})$, respectively.
Straightforward calculation yields

$$f(\lambda)=\frac{1}{2\pi}\begin{bmatrix} 1+\varepsilon e^{-i\lambda}+\delta\varepsilon\frac{e^{-2i\lambda}}{1-\delta e^{-i\lambda}} & +\varepsilon e^{i\lambda}+\frac{\delta\varepsilon e^{2i\lambda}}{1-\delta e^{i\lambda}} & \varepsilon e^{-i\lambda}+\varepsilon e^{i\lambda}+\delta\varepsilon\frac{e^{-2i\lambda}}{1-\delta e^{-i\lambda}} \\ \varepsilon e^{-i\lambda}+\varepsilon e^{i\lambda}+\delta\varepsilon\frac{e^{2i\lambda}}{1-\delta e^{i\lambda}} & & 1+\varepsilon e^{-i\lambda}+\varepsilon e^{i\lambda} \end{bmatrix}$$

and

$$\tilde{f}(\lambda)=\frac{1}{2\pi}\begin{bmatrix} 1+\varepsilon e^{-i\lambda}+\delta\varepsilon e^{-2i\lambda}+\varepsilon e^{i\lambda}+\delta\varepsilon e^{2i\lambda} & \varepsilon e^{i\lambda}+\varepsilon e^{-i\lambda}+\delta\varepsilon e^{-2i\lambda} \\ \varepsilon e^{-i\lambda}+\varepsilon e^{i\lambda}+\delta\varepsilon e^{2i\lambda} & 1+\varepsilon e^{-i\lambda}+\varepsilon e^{i\lambda} \end{bmatrix}.$$

Calculation of the Fourier coefficients of f (the covariances) gives

$$\Gamma_o = I_2\ ,\ \Gamma_1 = \varepsilon\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},\quad \Gamma_k = \varepsilon\ \delta^{k-1}\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\quad k = 2, 3,\ldots$$

and for $\tilde{f}$

$$\tilde{\Gamma}_o = I_2 \ , \ \tilde{\Gamma}_1 = \varepsilon \begin{bmatrix} 1 & 1 \\ & \\ 1 & 1 \end{bmatrix} , \quad \tilde{\Gamma}_2 = \varepsilon\ \delta \begin{bmatrix} 1 & 1 \\ & \\ 0 & 0 \end{bmatrix} , \quad \tilde{\Gamma}_k = \begin{bmatrix} 0 & 0 \\ & \\ 0 & 0 \end{bmatrix}, k=3,4\ldots$$

Thus for k =0, 1, 2 (=p+q) we have $\Gamma_k = \tilde{\Gamma}_k$ and since $\varepsilon\ \delta > 0$ we have $\Gamma_k \neq \tilde{\Gamma}_k$ for k = 3, 4,... Hence the sample size 3 is not second-order informative for the spectral measure. □

REMARK. Since p=q=1 we have $\tilde{R}_q = \Gamma_1$ (see (3.5.2)) which is singular and so we have indeed a counterexample for HANNAN'S lemma in [19].

As a first step in obtaining informative sample sizes we shall restate and prove HANNAN'S classical result on the ARMA(p,q)-case in the terminology and notation introduced in the preceding sections.

THEOREM 3.5.2 (HANNAN 1968) If in the m-variate ARMA(p,q) case $\theta = S_{p,q}^{ARMA}(m) \times Z$ and $\Sigma_\varepsilon(z) \subset \mathbb{C}^+(m)$, then the whole process $\{\underline{x}_t\}$ is second-order informative for $(A(z), B(z), \Sigma_\varepsilon)$.

PROOF. The proof is complete if we can prove the factorization (3.1.2) of the spectral density matrix to be unique. Suppose we have an alternative factorization into $\tilde{A}(z)$, $\tilde{B}(z)$ and $\tilde{\Sigma}$. Then we have from lemma 3.2.1

$$A^{-1}(z)\ B(z)\ \Sigma^{\frac{1}{2}} = \tilde{A}^{-1}(z)\ \tilde{B}(z)\ \tilde{\Sigma}^{\frac{1}{2}} \quad , \ |z| \leq 1$$

and as $A(0) = B(0) = I_m$, this implies $\Sigma = \tilde{\Sigma}$ and $A^{-1}(z)\ B(z) = \tilde{A}^{-1}(z)\ \tilde{B}(z)$, $|z| \leq 1$. Since both $(A(z), B(z))$ and $(\tilde{A}(z), \tilde{B}(z))$ satisfy the conditions of lemma 3.2.5 it follows that we must have $A(z) = \tilde{A}(z)$, $B(z) = \tilde{B}(z)$. □

REMARK. In HANNAN'S paper [18] the matrix Σ_ε is allowed to be singular but we already showed in the MA-case and the AR-case that the result is then in general false.

Intuitively one can feel that there must be some *finite* sample size that is second order informative for $(A(z), B(z), \Sigma_\varepsilon)$ (or the spectral measure) and that its minimum value depends on m. We have

THEOREM 3.5.3 If in the m-variate ARMA(p,q) case $\Theta = S_p^{AR}(m) \times \mathbb{C}^q(m) \times Z$ then the sample size q+(m+1)p is second-order informative for the spectral measure, or, equivalently, second-order predictive for the future.

PROOF. Let $\{\underline{x}_t\}$ be the observable process.

Putting

$$\underline{Y}_t = \begin{bmatrix} \underline{x}_t \\ \underline{x}_{t-1} \\ \vdots \\ \underline{x}_{t-p+1} \end{bmatrix}, \qquad t = 0, \pm 1, \ldots ,$$

we obtain an mp-variate process $\{\underline{Y}_t\}$ satisfying

$$(3.5.10)\quad \underline{Y}_t + \begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ -I_m & 0 & \cdots & 0 \\ 0 & -I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -I_m \; 0 \end{bmatrix} \underline{Y}_{t-1} = \begin{bmatrix} \sum_{k=0}^{q} B_k \underline{\varepsilon}_{t-k} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad t=0,\pm 1,\ldots$$

Thus $\{\underline{Y}_t\}$ is an mp-variate ARMA(1,q) process that is equivalent

to $\{\underline{x}_t\}$ in the sense that prediction of $\{\underline{x}_t\}$ is the same problem as prediction of $\{\underline{y}_t\}$. Therefore we may restrict ourselves to the case p = 1 without loss of generality. The fact that the error process in (3.5.10) has a singular covariance matrix is not important since we did not assume $\Sigma_\varepsilon \in \mathfrak{C}^+(m)$. Furthermore it is important to note that a sample of size s from the process $\{\underline{y}_t\}$ corresponds to a sample of size s + p - 1 from the process $\{\underline{x}_t\}$ (*sample size* must as before be interpreted as the number of points in time the process under consideration is observed).

So, consider the m-variate ARMA(1,q) model

$$\underline{x}_t - A\,\underline{x}_{t-1} = \sum_{k=0}^{q} B_k\,\underline{\varepsilon}_{t-k}\,, \qquad t = 0, \pm 1, \ldots$$

with covariance function Γ_s, s = 0, ± 1,... and let an alternative specification be denoted by

$$\underline{x}_t - \tilde{A}\,\underline{x}_{t-1} = \sum_{k=0}^{q} \tilde{B}_k\,\underline{\varepsilon}_{t-k}\,, \qquad t = 0, \pm 1, \ldots$$

with covariance function $\tilde{\Gamma}_s$, s = 0, ± 1,... Then we have to prove the implication

$$[\,\Gamma_s = \tilde{\Gamma}_s,\ s = 0, \pm 1, \ldots \pm(q+m)] \Rightarrow [\Gamma_s = \tilde{\Gamma}_s,\ s = 0, \pm 1, \ldots]\,.$$

From (3.5.2) we obtain

$$\Gamma_s = A\,\Gamma_{s-1}\,, \quad \tilde{\Gamma}_s = \tilde{A}\,\tilde{\Gamma}_{s-1}\,, \qquad s = q+1, q+2, \ldots,$$

or, equivalently,

$$\Gamma_s = A^{s-q}\,\Gamma_q\,, \quad \tilde{\Gamma}_s = \tilde{A}^{s-q}\,\Gamma_q\,, \quad s = q+1, q+2, \ldots \tag{3.5.11}$$

Suppose we have $\Gamma_s = \tilde{\Gamma}_s$, $s = 0, \pm 1, \ldots \pm (q+m)$.
By lemma 3.2.6 there exist $m \times m$ matrices $C_o, \ldots, C_m$ such that

$$A^{m+1} = \sum_{k=0}^{m} C_k A^k \quad \text{and} \quad \tilde{A}^{m+1} = \sum_{k=0}^{m} C_k \tilde{A}^k \ .$$

Hence, using (3.5.11) and $\Gamma_s = \tilde{\Gamma}_s$ $s = 0, \pm 1, \ldots, \pm (q+m)$

$$\Gamma_{m+q+1} - \tilde{\Gamma}_{m+q+1} = (A^{m+1} - \tilde{A}^{m+1})\Gamma_q = \sum_{k=0}^{m} C_k (A^k - \tilde{A}^k)\Gamma_q =$$

$$= \sum_{k=0}^{m} C_k (\Gamma_{q+k} - \tilde{\Gamma}_{q+k}) = 0$$

and so, by induction $\Gamma_s = \tilde{\Gamma}_s$, $s = 0, \pm 1, \ldots$
Thus we proved for an m-variate ARMA(1,q) process the sample size q+m+1 to be second-order informative for the spectral measure; hence for an m-variate ARMA(p,q) process the sample size q+mp+1+p-1 = q+(m+1)p is second-order informative for the spectral measure. This completes the proof. □

REMARK. The informative sample size in the theorem has less intuitive appeal than the previous results, since only for m=p=1 it corresponds to the result of theorem 2.5.2. For $m \geq 2$, the number q+(m+1)p - (p+q+1) = mp-1 can be interpreted a penalty for allowing the coefficients to be elements of a non commutative ring. It is, however, not certain that this penalty is the minimal one.

As an immediate consequence of theorems 3.5.2 and 3.5.3 we have

<u>THEOREM 3.5.4</u> If in the m-variate ARMA(p,q) case $\theta = S_{p,q}^{ARMA}(m) \times Z$ and $\Sigma_\varepsilon(z) \subset \mathbb{C}^+(m)$ then the sample size q + (m+1)p is second-

order informative for $(A(z), B(z), \Sigma_\varepsilon)$.

PROOF. let $(A_1(z), B_1(z), \Sigma_{\varepsilon,1}) \neq (A_2(z). B_2(z), \Sigma_{\varepsilon,2})$ correspond to θ_1 and θ_2 respectively. If F_θ denotes the spectral measure and R_θ the covariance matrix of a sample of size $q + (m+1)p$, then it follows from theorem 3.5.2 that we must have $F_{\theta_1} \neq F_{\theta_2}$. Hence by theorem 3.5.3 $R_{\theta_1} \neq R_{\theta_2}$. This proves the theorem. □

REMARK. In the counterexample presented in theorem 3.5.1 we found that in the case $m = 2$, $p = q = 1$ the sample size 3 is not second-order informative for $(A(z), B(z), \Sigma_\varepsilon)$. Theorem 3.5.4 implies that the sample size 4 is second-order informative for $(A(z), B(z), \Sigma_\varepsilon)$ and so we found in that particular case the minimum second-order informative sample size.

CHAPTER IV

DYNAMIC SIMULTANEOUS EQUATIONS WITH MA-ERRORS

4.1 INTRODUCTION

Consider the m-variate ARMA(p,q) model

$$(4.1.1) \qquad \sum_{k=0}^{p} A_k \underline{x}_{t-k} = \sum_{k=0}^{q} B_k \underline{\varepsilon}_{t-k} , \qquad t = 0, \pm 1, \ldots$$

As most econometric models of this type contain one or more *identities*, i.e., equations with known coefficients and zero errors, we cannot apply the theory of chapter III to such models (exept for theorem 3.5.3). Moreover, most model specifications are such that $A_o \neq I_m$. In econometrics such models are called systems of simultaneous equations. If $p > 0$, among the equations there are difference equations, and the model is then called dynamic.

In the sequel we shall frequently use partitioned matrices and vectors. If necessary for the sake of clarity, we shall indicate the dimensions of the partition as follows:

$$\underset{(k_1)\ (k_2)}{\begin{bmatrix} \cdots & \vdots & \cdots \end{bmatrix}}\begin{matrix}(n_1)\\ \\ (n_2)\end{matrix} \quad \text{or} \quad \underset{}{\begin{bmatrix} \cdots \end{bmatrix}}\begin{matrix}(n_1)\\ \\ (n_2)\end{matrix} .$$

We shall make the following assumptions.

a) The last m_o equations in (4.1.1) are identities , $(0 \leq m_o \leq m-1)$, i.e., $\underline{\varepsilon}_t$ can be written as

$$\underline{\varepsilon}_t = \begin{bmatrix} \underline{\tilde{\varepsilon}}_t \\ 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ (m_o) \end{matrix}, \qquad t = 0, \pm 1, \ldots$$

We shall assume that $\{\tilde{\varepsilon}_t\}$ is $(m-m_o)$-variate white noise with $E\{\tilde{\varepsilon}_t\} = 0$ and $\Sigma_{\tilde{\varepsilon}} := E\{\underline{\tilde{\varepsilon}}_t \underline{\tilde{\varepsilon}}_t^*\} > 0$ which is supposed to be unknown. Hence

$$\Sigma_\varepsilon = \begin{bmatrix} \Sigma_{\tilde{\varepsilon}} & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix}.$$

$$\qquad\qquad (m-m_o) \quad (m_o)$$

We shall use a similar partitioning for A_k and B_k,

$$A_k = \begin{bmatrix} A_k^{(11)} & \vdots & A_k^{(12)} \\ \cdots\cdots & \vdots & \cdots\cdots \\ A_k^{(21)} & \vdots & A_k^{(22)} \end{bmatrix} = \begin{bmatrix} A_k^{(1)} \\ \cdots\cdots \\ A_k^{(2)} \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix}, \quad k=0,1,\ldots p.$$

$$\qquad\qquad (m-m_o) \quad (m_o)$$

The matrices B_k are supposed to be of the form

$$B_k = \begin{bmatrix} B_k^{(o)} & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix}, \qquad k = 0, 1, \ldots, q,$$

$$\qquad\qquad (m-m_o) \quad (m_o)$$

with $B_o^{(o)} = I_{m-m_o}$.

For $A(z) := \sum_{k=0}^{p} A_k z^k$ and $B(z) := \sum_{j=0}^{q} B_j z^j$ we can now write

$$(4.1.3) \qquad A(z) = \begin{bmatrix} A^{(11)}(z) & \vdots & A^{(12)}(z) \\ \cdots\cdots & \vdots & \cdots\cdots \\ A^{(21)}(z) & \vdots & A^{(22)}(z) \end{bmatrix} = \begin{bmatrix} A^{(1)}(z) \\ \cdots\cdots \\ A^{(2)}(z) \end{bmatrix} \begin{matrix} (m-m_o) \\ , \\ (m_o) \end{matrix} \qquad z \in \mathbb{C},$$

$$\begin{matrix} (m-m_o) & (m_o) \end{matrix}$$

$$(4.1.4) \qquad B(z) = \begin{bmatrix} B^{(o)}(z) & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ , \\ (m_o) \end{matrix} \qquad z \in \mathbb{C} .$$

$$\begin{matrix} (m-m_o) & (m_o) \end{matrix}$$

Since the last m_o rows of $A(z)$ correspond to the identities, $A^{(2)}(z)$ is a *known* matrixpolynomial.

b) A_o is nonsingular. In econometrics the model is then said to be *complete*. Obviously we need this assumption if we want $A(z)$ to be nonsingular on $|z| \leq 1$.

c) $A_o^{(22)}$ is nonsingular. This assumption makes it possible to interpret the identities as *definitions* of the last m_o components of $\underline{x}_t$ and enables us to substitute them into the first $m-m_o$ equations.

d) degree $\{A^{(2)}(z)\} := p_o \leq \max(0, p-1)$.
This assumption states, roughly speaking, that for $p \geq 1$ the identities do not contain components of $\underline{x}_{t-\ell}$ for the maximum *time-lag* $\ell=p$. Of course the assumption is not restrictive if $p=0$.

e) $A^{(22)}(z)$ is a *proper* matrix polynomial (i.e. a matrix-

polynomial with nonsingular leading coefficient matrix).

Let $\xi \in Z$ be a parameter characterizing the distribution of the process $\{\underline{\tilde{\varepsilon}}_t\}$. Then we can write $\Sigma_\varepsilon = \Sigma_\varepsilon(\xi)$ and $\Sigma_{\tilde{\varepsilon}} = \Sigma_{\tilde{\varepsilon}}(\xi)$.

We can take $\theta := (A_o^{(1)},\ldots,A_p^{(1)} ; B_1^{(o)},\ldots,B_q^{(o)} ; \xi)$.
Putting

$$S = \{(A_o^{(1)},\ldots,A_p^{(1)}) \mid \det A(z) \neq 0,\ |z| = 1\} ,$$

then the parameterspace θ is supposed to satisfy
$\theta \subset S \times \mathbb{C}^{q(m-m_o)} \times Z$.
In econometrics (4.1.1) is called the *structural form* of the model. Premultiplying with A_o^{-1} yields the so called *reduced form*,

$$(4.1.5) \qquad \sum_{k=0}^{p} P_k \underline{x}_{t-k} = \sum_{k=0}^{q} Q_k \underline{\eta}_{t-k} , \qquad t = 0, \pm 1,\ldots,$$

where $P_o = I_m$, $P_k := A_o^{-1} A_k$, $k = 1, 2, \ldots, p$,

$$Q_k := A_o^{-1} B_k A_o , \qquad k = 0, 1, 2, \ldots q,$$

and $\underline{\eta}_t := A_o^{-1} \underline{\varepsilon}_t , \qquad t = 0, \pm 1,\ldots$.

Let $P(z) := A_o^{-1} A(z)$ and $Q(z) := A_o^{-1} B(z) A_o$ denote the generating functions of (P_k) and (Q_k) and let $\Sigma_\eta := A_o^{-1} \Sigma_\varepsilon A_o^{-1*}$.
A sample is called (second-order) informative for the structural form if it is informative for $(A(z), B(z), \Sigma_\varepsilon)$, and it is called informative for the reduced form if it is informative for $(P(z), Q(z), \Sigma_\eta)$.
It should be noted that in order to find informative sample sizes for the reduced form we cannot apply theorem 3.5.4 to (4.1.5) since Σ_η is allowed to be singular. We can however apply theorem 3.5.3 to find informative sample sizes for the spectral

measure and we may hope that *prior knowledge* on $(A(z), B(z), \Sigma_\varepsilon)$ enables us to identify it.
In section 4.2 we shall find informative sample sizes for the reduced form.
In section 4.3 we shall consider the structural form.

4.2 INFORMATIVE SAMPLES FOR THE REDUCED FORM

Consider the model (4.1.1) under the assumptions a) - e). The spectral density matrix f_θ of the process $\{\underline{x}_t\}$ is given by

$$(4.2.1) \qquad f_\theta(\lambda) = A^{-1}(e^{-i\lambda})\, B(e^{-i\lambda})\, \Sigma_\varepsilon\, B^*(e^{-i\lambda})\, A^{-1*}(e^{-i\lambda}), \quad \lambda \in (-\pi, \pi] ,$$

or, equivalently,

$$(4.2.2) \qquad f_\theta(\lambda) = P^{-1}(e^{-i\lambda})\, Q(e^{-i\lambda})\, \Sigma_\eta\, Q^*(e^{-i\lambda})\, P^{-1*}(e^{-i\lambda}), \quad \lambda \in (-\pi, \pi] .$$

Although the sample size q+(m+1)p is second-order informative for the spectral measure by theorem 3.5.3, it is not necessarily second-order informative for $(P^{-1}(z)\, Q(z), \Sigma_\eta)$ since Σ_η is singular (see the MA case § 3.3). But even if prior knowledge enables us to identify $P^{-1}(z)\, Q(z)$ we have a problem, since the conditions of lemma 3.2.5 under which rational matrix functions uniquely decompose into a left denominator and a right numerator are not realistic for most simultaneous equation models. To see this, note that the condition $r\,[P_p, Q_q] = m$ is equivalent to $r\,[A_p, B_q] = m$ and since B_q has zeros in the last m_o rows (and columns) this implies that $A_p^{(2)}$ has rank m_o, which is almost never fulfilled in practice. Therefore we shall prove the following modification of Lemma 3.2.5 for partitioned matrices.

LEMMA 4.2.1 Let the rational m × m matrix function D(z) decompose as

$$D(z) = A^{-1}(z)\, B(z),$$

with $A(0) = A_o$, and where A(z) is of the form (4.1.3) satisfying assumptions b) - e), B(z) is of the form (4.1.4) and $r\,[A(z), B(z)] = m$, for all $z \in \mathbb{C}$. Moreover, suppose degree $\{A^{(12)}(z)\} \leq p_o$ and $r\,[A_p^{(11)}, B_q^{(o)}] = m - m_o$.
Then A(z) and B(z) are uniquely determined by A_o, $A^{(2)}(z)$ and D(z).

PROOF. For p = 0 nothing remains to prove so let $p \geq 1$.
Suppose there exist an alternative decomposition

$$D(z) = \tilde{A}^{-1}(z)\, \tilde{B}(z),$$

satisfying the conditions of the theorem, such that $\tilde{A}(0) = A(0)$ and $\tilde{A}^{(2)}(z) = A^{(2)}(z)$ for all $z \in \mathbb{C}$. Then the funtion

$$C(z) := \frac{1}{z}\,(A(z) - \tilde{A}(z))$$

is a polynomial of degree $\leq p-1$, and can be partitioned as

$$C(z) = \begin{bmatrix} C^{(1)}(z) & \vdots & C^{(2)}(z) \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix} \qquad (4.2.3)$$

$$\qquad\qquad (m-m_o) \qquad (m_o)$$

where degree $\{C^{(2)}(z)\} \leq p_o - 1$.
We also have

$$C(z)\, D(z) = \frac{1}{z}\,(B(z) - \tilde{B}(z)),$$

which is a polynomial of degree $\leq$ q-1 since $B(0) = \tilde{B}(0) = A_o^{-1} D(0)$. Let d denote the degree of the polynomial det A(z). As in the proof of Lemma 3.2.5 we introduce the matrixpolynomial

$$S(z) := C(z)\ A^{-1}(z)\ \det A(z),$$

with degree $\{S(z)\} = s$. From (4.2.3) it follows that S(z) can be partioned into

$$(4.2.4) \qquad \begin{bmatrix} S^{(1)}(z) & \vdots & S^{(2)}(z) \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix}$$

$$\qquad\qquad (m-m_o) \qquad (m_o)$$

Suppose $d \leq s$. Let $S^{(1)}(z) = \sum_{k=0}^{s} S_k^{(1)} z^k$ and $S^{(2)}(z) = \sum_{k=0}^{s} S_k^{(2)} z^k$. We shall first prove that $S_s^{(1)} \neq 0$. If $S_s^{(1)} = 0$ then we must have $S_s^{(2)} \neq 0$.

From S(z) A(z) = C(z) det A(z) we obtain using (4.1.3), (4.2.3) and (4.2.4)

$$S^{(1)}(z)\ A^{(12)}(z) + S^{(2)}(z)\ A^{(22)}(z) = C^{(2)}(z) \det A(z).$$

The first term on the left has degree $\leq s-1+p_o$, since $S_s^{(1)} = 0$ and degree $\{A^{(12)}(z)\} \leq p_o$. The second term has degree $s + p_o$ since $S_s^{(2)} \neq 0$ and $A^{(22)}(z)$ is proper by assumption e). Hence $C^{(2)}(z) \det A(z)$ has degree $s+p_o$. But we also have

$$\text{degree } \{C^{(2)}(z) \det Z(z)\} \leq p_o-1+d,$$

and since $d \leq s$ we have a contradiction, proving $S_s^{(1)} \neq 0$.

Since $S(z)B(z) = C(z)D(z)\det A(z)$ is a matrix polynomial of degree at most q-1+d, and has the same degree as $S^{(1)}(z)\ B^{(o)}(z)$, we must have

$$\sum_{k=0}^{s} S_k^{(1)}\ B_{\ell-k}^{(o)} = 0 \quad , \qquad \ell=q+d,\ q+d+1,\ \ldots \quad .$$

Similarly for $S(z)\ A(z) = C(z)\ \det A(z)$ we obtain

$$\sum_{k=0}^{s} [S_k^{(1)},\ S_k^{(2)}]\ A_{n-k} = 0\ ,\ n = p+d,\ p+d+1,\ \ldots \quad .$$

Choosing $\ell=q+s$ and $n=p+s$ yields

$$(4.2.5) \qquad S_s^{(1)}\ B_q^{(o)} = 0\ ,\ [S_s^{(1)},\ S_s^{(2)}]\ A_p = 0.$$

From degree $\{A^{(2)}(z)\} \leq p_o \leq p-1$ it follows that the second relation in (4.2.5) is equivalent to

$$S_s^{(1)}\ A_p^{(11)} = 0 \quad ,$$

which together with the first relation in (4.2.5) contradicts

$$r\ [A_p^{(11)},\ B_q^{(o)}] = m-m_o.$$

Now suppose $d \geq s+1$. Then we obtain in a similar way as in Lemma 3.2.5 a contradiction to $r\ [A(z),\ B(z)] = m$. This proves the Lemma. □

REMARK 1. For $m_o = 0$ and $A_o = I_m$ we obtain the result of Lemma 3.2.5, thus we have a genuine generalization.

REMARK 2. If A(z) in (4.1.1) does not satisfy degree$\{A^{(12)}(z)\} \leq p_o$, then one can always, by means of substitutions, obtain an equivalent model specification that does satisfy this condition, since $A^{(22)}(z)$ is proper. Therefore it is a rather natural condition for identifiability.

Put

$$S^{DSE} = \left\{ (A_o^{(1)} \ldots A_p^{(1)}; B_1^{(o)} \ldots B_q^{(o)}) \,\middle|\, \det A(z) \neq 0,\ |z| \leq 1; \right.$$

$$\det B^{(o)}(z) \neq 0,\ |z| < 1\ ;\ r\,[A_p^{(11)}, B_q^{(o)}] = m - m_o\ ;$$

$$\left. r\,[A(z),\ B(z)] = m\ \forall z \text{ and degree } \{A^{(12)}(z)\} \leq p_o \right\}.$$

Since $A_o^{(22)}$ is nonsingular we can (by substitution of the m_o identities into the first $m-m_o$ equations) rewrite (4.1.1) as

$$(4.2.6) \qquad \sum_{k=0}^{p} \tilde{A}_k \underline{x}_{t-k} = \sum_{k=0}^{q} B_k \underline{\varepsilon}_{t-k}, \quad t = 0, \pm 1, \ldots,$$

where

$$\tilde{A}_o = \begin{bmatrix} \tilde{A}_o^{(11)} & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ A_o^{(21)} & \vdots & A_o^{(22)} \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix} .$$

$$\qquad\qquad (m-m_o) \quad (m_o)$$

Since the reduced form of (4.2.6) is equal to the reduced form of the original model (4.1.1), it is sufficient to prove the identifiability of the reduced form under the extra assumption that $A_o^{(12)} = 0$. In that case $A_o^{(11)}$ is nonsingular. Let $\tilde{S}^{DSE}$ be the set obtained from S^{DSE} by adding the condition $A_o^{(12)} = 0$.

LEMMA 4.2.2 Let S be a nonsingular m×m matrix, and A and B arbitrary m×m matrices. Then r [A,B] = m iff r [SA, SBS^{-1}] = m.

PROOF. Since S is nonsingular we have r [SA, SBS^{-1}] = r [A, BS^{-1}]. Moreover, {r [A, BS^{-1}] = m} $\iff$ {y* A $\neq$ 0 $\vee$ y*BS^{-1} $\neq$ 0 $\forall y \in \mathbb{C}^m \setminus 0$} $\iff$ {y*A $\neq$ 0 $\vee$ y*B $\neq$ 0 $\forall$ y $\in \mathbb{C}^m \setminus 0$} $\iff$ {r [A, B] = m}. □

LEMMA 4.2.3 Consider an arbitrary sample size for the model (4.1.1), and let $\theta = \tilde{S}^{DSE} \times Z$ with $\Sigma_{\tilde{\varepsilon}}$ completely unknown. Then the functions $\phi(\theta) := A_o$ and $\psi(\theta) := (P(z), Q(z), \Sigma_\eta)$ are second-order informationally independent.

PROOF. Put $\nu(\theta) = (\phi(\theta), \psi(\theta))$. We shall first prove that $\nu(\theta) = \phi(\theta) \times \psi(\theta)$. Let $\tilde{A}_o \in \psi(\theta)$ and $(P(z), Q(z), \Sigma_\eta) = (A_o^{-1} A(z), A_o^{-1} B(z) A_o, A_o^{-1} \Sigma_\varepsilon A_o^{-1*}) \in \psi(\theta)$ be arbitrary. Then we have to show that $(\tilde{A}_o, P(z), Q(z), \Sigma_\eta) \in \nu(\theta)$, or, equivalently, that

$$(4.2.7)\quad (\tilde{A}_o, \tilde{A}_o P_1, \ldots \tilde{A}_o P_p, \tilde{A}_o Q_1 \tilde{A}_o^{-1}, \ldots, \tilde{A}_o Q_q \tilde{A}_o^{-1}) \in \tilde{S}^{DSE},$$

and that the last m_o rows and columns of $\tilde{A}_o \Sigma_\eta \tilde{A}_o^*$ (= $A_o A_o^{-1} \Sigma_\varepsilon A_o^{-1} A_o$) are zero.
Since $A_o^{(12)} = 0$, the matrix A_o^{-1} has the form

$$A_o^{-1} = \begin{matrix} \\ (m-m_o) \\ \\ (m_o) \end{matrix} \begin{matrix} \\ \left[\begin{array}{c|c} & \\ & 0 \\ & \end{array}\right] \\ \\ \end{matrix},$$

$\qquad\qquad\qquad\qquad\qquad\qquad\quad (m-m_o) \quad (m_o)$

and since the last m_o rows of A_o and $\tilde{A}_o$ are equal (assumption a)), it follows that $\tilde{A}_o A_o^{-1}$ has the form

$$\tilde{A}_o\, A_o^{-1} = \left[\begin{array}{c:c} T & 0 \\ \hdashline 0 & I_{m_o} \end{array}\right] \begin{matrix} (m-m_o) \\ (m_o) \end{matrix} \qquad (\det T \neq 0).$$

$$\begin{matrix} (m-m_o) & (m_o) \end{matrix}$$

So

$$\tilde{A}_o\, Q(z)\, \tilde{A}_o^{-1} = \left[\begin{array}{c:c} T\, B^{(o)}(z)\, T^{-1} & 0 \\ \hdashline 0 & 0 \end{array}\right], \quad z \in \mathbb{C}$$

and

$$(4.2.8) \quad \tilde{A}_o\, \Sigma_\eta\, \tilde{A}_o = \left[\begin{array}{c:c} T\, \Sigma_{\tilde{\varepsilon}}\, T & 0 \\ \hdashline 0 & 0 \end{array}\right].$$

Furthermore Lemma 4.2.2 gives

$$r\,[\tilde{A}_o\, P(z),\ \tilde{A}_o\, Q(z)\, \tilde{A}_o^{-1}] = m, \quad z \in \mathbb{C},$$

and

$$r\,[T\, A_p^{(11)},\ T\, B_q^{(o)}\, T^{-1}] = m-m_o.$$

We also have

$$\tilde{A}_o\, P(z) = \left[\begin{array}{c:c} & T\, A^{(12)}(z) \\ \hdashline & \end{array}\right] \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix},$$

$$\begin{matrix} (m-m_o) & (m_o) \end{matrix}$$

and so degree $\{T\, A^{(12)}(z)\}$ = degree $\{A^{(12)}(z)\} \leq p_o$.

Finally, $\det\{\tilde{A}_o\ P(z)\}$ and $\det A(z)$ have the same zeros and $\det\{T\ B^{(0)}(z)\ T^{-1}\} = \det B^{(0)}(z)$, $z \in \mathbb{C}$. Thus we proved (4.2.7), and from (4.2.8) together with the fact that $\Sigma_{\tilde{\varepsilon}}$ is completely unknown it follows that $(\tilde{A}_o, P(z), Q(z), \Sigma_\eta) \in \nu(\theta)$.

We recall that two values ν_1 and ν_2 are second-order observationally equivalent if they generate the same covariance structure for the sample. From the second-order version of theorem 1.2.9 it follows that it is necessary and sufficient to prove the equivalence classes to be of the form $U_o \times V_o$, where $U_o \subset \Phi(\theta)$ and $V_o \subset \Psi(\theta)$. This follows immediately from the facts that $\nu(\theta) = \Phi(\theta) \times \Psi(\theta)$ and that the spectral measure (and hence the covariance structure of the sample) only depends on $\Psi(\theta)$. □

THEOREM 4.2.4 If in a dynamic system of simultaneous equations $\theta = S^{DSE} \times Z$, then the sample size $q + (m+1)p$ is second-order informative for the reduced form.

PROOF. Without loss of generality we can take $\theta = \tilde{S}^{DSE} \times Z$ and assume $\Sigma_{\tilde{\varepsilon}}$ completely unknown. Let θ_1 and $\theta_2 \in \theta$ correspond to $(A(z), B(z), \Sigma_\varepsilon)$ and $(\tilde{A}(z), \tilde{B}(z), \tilde{\Sigma}_\varepsilon)$, respectively, where $A(0) = \tilde{A}(0) = A_o$. Let R_θ denote the covariance matrix of a sample of size $q+(m+1)p$, and suppose $R_{\theta_1} = R_{\theta_2}$. From theorem 3.5.3 it follows that then the spectral measures also coincide i.e.

$$A^{-1}(e^{-i\lambda})B(e^{-i\lambda})\Sigma_\varepsilon\ B^*(e^{-i\lambda})A^{-1*}(e^{-i\lambda}) = \tilde{A}^{-1}(e^{-i\lambda})\tilde{B}(e^{-i\lambda})\tilde{\Sigma}_\varepsilon\ \tilde{B}^*(e^{-i\lambda})\tilde{A}^{-1*}(e^{-i\lambda}). \tag{4.2.9}$$

Putting $W(z) = A(z)\ \tilde{A}^{-1}(z)$, from (4.2.9) we obtain

(4.2.10)

$$B(e^{-i\lambda})\Sigma_\varepsilon B^*(e^{-i\lambda}) = W(e^{-i\lambda})\tilde{B}(e^{-i\lambda})\tilde{\Sigma}_\varepsilon \tilde{B}^*(e^{-i\lambda})W^*(e^{-i\lambda}),$$

$$\lambda \in (-\pi, \pi].$$

Let $W(z)$ be partitioned in the same way as $A(z)$ and $B(z)$ as

$$W(z) = \begin{bmatrix} W^{(11)}(z) & \vdots & W^{(12)}(z) \\ \cdots\cdots & \vdots & \cdots\cdots \\ W^{(21)}(z) & \vdots & W^{(22)}(z) \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix}$$

$$\qquad\qquad (m-m_o) \qquad\quad (m_o)$$

Since $B(z)$, $\tilde{B}(z)$, Σ_ε and $\tilde{\Sigma}_\varepsilon$ have zeros in the last m_o rows and columns and $\tilde{B}^{(o)}(e^{-i\lambda})\tilde{\Sigma}_\varepsilon$ is nonsingular a.e. it follows from (4.2.10) that $W^{(21)}(z) = 0$, $z \in \mathbb{C}$. (Since $A^{(2)}(z) = \tilde{A}^{(2)}(z)$, $z \in \mathbb{C}$ and $r[A^{(2)}(z)] = m_o$ for almost all $z \in \mathbb{C}$). It now follows from $W(z)\tilde{A}(z) = A(z)$ that we must have $W^{(22)}(z) = I_{m_o}$. But then it follows from (4.2.10) that

(4.2.11)

$$B^{(o)}(e^{-i\lambda})\Sigma_{\tilde{\varepsilon}} B^{(o)*}(e^{-i\lambda}) = W^{(11)}(e^{-i\lambda})\tilde{B}^{(o)}(e^{-i\lambda})\tilde{\Sigma}_{\tilde{\varepsilon}}\tilde{B}^{(o)*}(e^{-i\lambda})$$
$$W^{(11)}(e^{-i\lambda}).$$

As $W^{(11)}(z)$ is analytic on $|z| < \rho$ for a $\rho > 1$ (because $A(z)$ and $\tilde{A}^{-1}(z)$ are), it follows from (4.2.11) and the fundamental Lemma 3.2.1 that

$$\Sigma_{\tilde{\varepsilon}} = \tilde{\Sigma}_{\tilde{\varepsilon}} \quad \text{and} \quad B^{(o)}(z) = W^{(11)}(z)\tilde{B}^{(o)}(z), \quad z \in \mathbb{C}.$$

But then we also have $B(z) = W(z)\tilde{B}(z)$ and so $A^{-1}(z)B(z) = \tilde{A}^{-1}(z)\tilde{B}(z)$, $z \in \mathbb{C}$. Since $A_o = \tilde{A}_o$ and $A^{(2)}(z) = \tilde{A}^{(2)}(z)$ it

follows from Lemma 4.2.1 that $A(z) = \tilde{A}(z)$ and $B(z) = \tilde{B}(z)$, $z \in \mathbb{C}$. Thus we proved the sample size $q+(m+1)p$ to be second order informative for $(A(z), B(z), \Sigma_\varepsilon)$ *conditional on* A_o. But then the sample size $q+(m+1)p$ is also second order-informative for $(P(z), Q(z), \Sigma_\eta)$ conditional on A_o, and since A_o and $(P(z), Q(z), \Sigma_\eta)$ are second-order informationally independent by Lemma 4.2.2, it follows by theorem 1.2.8 that the sample size $q+(m+1)p$ is second-order informative for $(P(z), Q(z), \Sigma_\eta)$. □

We shall show by an example that we cannot drop the condition degree $\{A^{(12)}(z)\} \leq p_o$. It is a slight modification of an example from KOOPMANS, RUBIN and LEIPNIK (see THEIL [32], p. 494).

EXAMPLE 4.2.5 Consider the following simple two-equation system

$$(4.2.12)\quad \begin{cases} \underline{x}_{t1} + \beta_1 \underline{x}_{t-1,1} + \beta_2 \underline{x}_{t-1,2} = \underline{\varepsilon}_{t1} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t = 0, \pm 1, \ldots \\ \underline{x}_{t2} + \gamma \underline{x}_{t1} = 0 \end{cases}$$

where γ is a known non-zero constant. Thus we have $p=1$, $q=0$, $p_o=0$ and

$$A(z) = \left[\begin{array}{c:c} 1 + \beta_1 z & \beta_2 z \\ \hdashline \gamma & 1 \end{array}\right] .$$

Hence

$$(4.2.13)\ldots\quad A_o^{-1} A(z) = \begin{bmatrix} 1 + \beta_1 z & \beta_2 z \\ -\gamma\beta_1 z & 1-\gamma\beta_2 z \end{bmatrix} .$$

When we lag the second equation in (4.2.12) and add the result to the first equation, we obtain a system with coefficient ge-

nerating function

$$\tilde{A}(z) := \begin{bmatrix} 1+(\beta_1+\gamma)z & (\beta_2+1)z \\ \gamma & 1 \end{bmatrix} .$$

The corresponding reduced form is

$$(4.2.14) \qquad \tilde{A}_o^{-1}\,\tilde{A}(z) = \begin{bmatrix} 1+(\beta_1+\gamma)z & (\beta_2+1)z \\ -\gamma(\beta_1+\gamma)z & 1-\gamma(\beta_2+1)z \end{bmatrix} .$$

Clearly, (4.2.13) and (4.2.14) are indistinguishable.
Note, that if we replace the second equation by $\underline{x}_{t2} + \gamma\underline{x}_{t1} = \underline{\varepsilon}_{t2}$, then the argument does not apply, because in that case the bivariate AR(1) model would become a bivariate ARMA(1,1) model.

4.3 INFORMATIVE SAMPLES FOR THE STRUCTURAL FORM.

Intuitively it is clear that in general under the conditions of theorem 4.2.4 we do not have identifiability of the structural form, or, equivalently, of the functions

$$\delta(\Theta) := (A(z), B(z)) \quad , \quad \Theta \in \theta$$

and Σ_ε. On the other hand in most systems of simultaneous equations there is *prior knowledge*, which is usually restricted to $A(z)$, such as *zero-restrictions* and A_o having unit diagonal elements. Therefore we may hope that certain functions of $\delta(\Theta)$ can be identified.
Throughout this section we assume $\theta = \Delta \times \mathbf{Z}$ where $\Delta := \delta(\theta)$ is an open subset of S^{DSE}.
Let

$$\Psi(\delta(\Theta)) := (P(z), Q(z)) \quad , \quad \Theta \in \theta$$

and let $\delta : \Delta \to \Lambda$ be an arbitrary function. Since the sample size $q+(m+1)p$ is second-order informative for $\Psi \circ \delta$ by theorem 4.2.4, a sufficient condition for this sample size to be second-order informative for $\Phi \circ \delta$, is the existence of a function $\nu : \Psi(\Delta) \to \Lambda$, such that $\Phi \circ \delta = \nu \circ \Psi \circ \delta$. We shall see that, if $\Sigma_{\tilde{\varepsilon}}$ is completely unknown, this condition is also necessary.

REMARK. In order to obtain necessary conditions, one sometimes meets the condition that the $\underset{\sim}{\tilde{\varepsilon}}_t$ are i.i.d. and that the class of distributions of $\underset{\sim}{\tilde{\varepsilon}}_t$ is *closed under nonsingular linear transformations*, i.e. if $\underset{\sim}{\tilde{\varepsilon}}_t$ has distribution P_ξ and $T \in \mathbb{C}(m)$ is nonsingular, then there exists $\tilde{\xi} \in Z$ such that $P_\xi \, T^{-1} = P_{\tilde{\xi}}$. Since we consider second-order identifiability, it is more natural to assume $\Sigma_{\tilde{\varepsilon}}$ completely unknown.

Let $R_\Theta^{(n)}$ denote the covariance matrix of a sample of size n, and $\Sigma_\eta(\Theta) = A_o^{-1} \Sigma_\varepsilon A_o^{-1*}$. We have

LEMMA 4.3.1

$$\left.\begin{array}{l} R_{\Theta_1}^{(n)} \neq R_{\Theta_2}^{(n)} \\ \\ \Sigma_\eta(\Theta_1) = \Sigma_\eta(\Theta_2) \end{array}\right\} \Rightarrow \Psi(\delta(\Theta_1)) \neq \Psi(\delta(\Theta_2)) \quad \Theta_1, \Theta_2 \in \Theta \quad .$$

PROOF. Follows immediately from the fact that the spectral measure, and so $R_\Theta^{(n)}$, is uniquely determined by $\Psi(\delta(\Theta))$ and Σ_η. □

THEOREM 4.3.2 If $\Sigma_{\tilde{\varepsilon}}$ is completely unknown, then the sample size $n \geq q + (m+1)p$ is second-order informative for $\Phi \circ \delta$ iff there exists a function ν such that $\Phi \circ \delta = \nu \circ \Psi \circ \delta$.

PROOF. The "if" part being trivial, we shall only prove the "only if" part. Suppose the sample size $q+(m+1)p$ is second-order informative for $\Phi \circ \delta$. Let $\delta_1 = (A(z), B(z))$ and

$\delta_2 = (\tilde{A}(z), \tilde{B}(z))$ be such that $\Phi(\delta_1) \neq \Phi(\delta_2)$. Let further $\xi_1 \in Z$ be arbitrary and put

$$\theta_1 = (A_o^{(1)} \ldots A_p^{(1)}, B_1^{(o)} \ldots B_q^{(o)}, \xi_1).$$

Since $\tilde{A}_o A_o^{-1}$ can be written as

$$\tilde{A}_o A_o^{-1} = \begin{bmatrix} T & \vdots & U \\ \ldots & \vdots & \ldots \\ 0 & \vdots & I_{m_o} \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix} \qquad (\det T \neq 0),$$

$$\begin{matrix} & (m-m_o) & (m_o) \end{matrix}$$

it follows that

$$\tilde{A}_o A_o^{-1} \Sigma_\varepsilon A_o^{-1*} \tilde{A}_o^{*} = \begin{bmatrix} T \underset{\sim}{\Sigma}_\varepsilon T^{*} & \vdots & 0 \\ \ldots & \vdots & \ldots \\ 0 & \vdots & 0 \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix} .$$

$$\begin{matrix} & (m-m_o) & (m_o) \end{matrix}$$

Since T is nonsingular and $\Sigma_{\tilde{\varepsilon}}$ is completely unknown, there exists $\xi_2 \in Z$ such that $T \Sigma_{\tilde{\varepsilon}} T^{*} = \Sigma_{\tilde{\varepsilon}}(\xi_2)$. Let $\tilde{\Sigma}_\varepsilon$ correspond to $\tilde{\underset{\sim}{\Sigma}}_\varepsilon :=$ $\Sigma_{\underset{\sim}{\varepsilon}}(\xi_2)$. Then we have $A_o^{-1} \Sigma_\varepsilon A_o^{-1*} = \tilde{A}_o^{-1} \tilde{\Sigma}_\varepsilon \tilde{A}_o^{-1*}$. Putting

$$\theta_2 = (\tilde{A}_o^{(1)} \ldots \tilde{A}_p^{(1)}, \tilde{B}_1^{(o)} \ldots \tilde{B}_q^{(o)}, \xi_2) \quad ,$$

we obtain $\Sigma_\eta(\theta_1) = \Sigma_\eta(\theta_2)$.
We also have $R_{\theta_1}^{(n)} \neq R_{\theta_2}^{(n)}$ for $n \geq q+(m+1)p$, since the sample size $q+(m+1)p$ is second-order informative for $\Phi \circ \delta$. Thus by Lemma 4.3.1 it follows that we must have $\Psi(\delta(\theta_1)) \neq \Psi(\delta(\theta_2))$.
Thus we proved the implication

$$\Phi(\delta_1) \neq \Phi(\delta_2) \Rightarrow \Psi(\delta_1) \neq \Psi(\delta_2), \qquad \delta_1, \delta_2 \in \Delta$$

which in turn implies the existence of a function ν such that $\phi \circ \delta = \nu \circ \Psi \circ \delta$. □

The practical value of this theorem is limited, since the condition is difficult to verify. More tractable conditions can be obtained by considering the equations separately. We shall make the following assumption.

ASSUMPTION. The matrix A_o has unit diagonal elements.

First we introduce some notations and conventions. Let a'_{ki} denote the i^{th} row of the matrix A_k (k=0,...p). Clearly, if a'_{oi} is identifiable for all i=1,...,m-m_o then A_o is identifiable and therefore (A(z), B(z), Σ_ε), since the reduced form is identifiable by theorem 4.2.4. Therefore we do not care about the error part for the moment and define

DEFINITION 4.3.3 A sample (size) is said to be second-order informative for the i^{th} equation, if it is second-order informative for $[a_{oi} \dots a_{pi}]$.

Let i be fixed. For notational convenience we re-order the equations in such a way, that a'_{oi} takes the form

$$a'_{oi} = \underset{(1)}{(\,1\,,} \ \underset{(m_i)}{\alpha'_{oi}\,,} \ \underset{(m-m_i-1)}{0\,)} \quad .$$

Thus $(1, \alpha'_{oi})'$ is the (m_i+1)-vector of coefficients of the non-lagged variables that occur in the i^{th} equation. Then A_o can be partitioned as

$$A_o = \begin{matrix} \\ \\ \\ \end{matrix} \begin{matrix} \overset{(m_i+1)}{\overbrace{}} & \overset{(m-m_i-1)}{\overbrace{}} & \\ \left[\begin{array}{c:c} 1,\ \alpha'_{oi} & 0 \\ \hdashline A_{oi}^{rest} & A_{oi} \end{array}\right] & & \begin{matrix} (1) \\ (m-1) \end{matrix} \end{matrix} \quad .$$

Let k_{ih} denote the number of variables with lag h occurring in the i^{th} equation (h=1,2,...,p) and let α_{hi} denote the k_{ih}-vector of coefficients. Then (after re-ordering) the matrix A_h can be written as

$$A_h = \begin{bmatrix} \alpha'_{hi} & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ A^{rest}_{hi} & \vdots & A_{hi} \end{bmatrix} \begin{matrix} (1) \\ \\ (m-1) \end{matrix}, \quad h=1,2,\ldots,p.$$

$$(k_{ih}) \quad (m-k_{ih})$$

Let the matrix P_h be partitioned as

$$P_h = \begin{bmatrix} \pi_{hi} & \vdots & \hat{\pi}_{hi} \\ \cdots\cdots & \vdots & \cdots\cdots \\ \Pi^{rest}_{hi} & \vdots & \hat{\Pi}^{rest}_{hi} \end{bmatrix} \begin{matrix} (1) \\ \\ (m_i) \\ \\ (m-m_i-1) \end{matrix}, \quad h=1,2,\ldots,p.$$

$$(k_{ih}) \quad (m-k_{ih})$$

Thus π'_{hi} is the k_{ih}-vector of coefficients in the i^{th} equation from the reduced form, of the variables that occur in the i^{th} structural equation, etc. The relation $A_o P_h = A_h$ can now be written as

$$(4.3.1) \quad \begin{bmatrix} 1 & \alpha'_{oi} & \vdots & 0 \\ \cdots\cdots & & \vdots & \cdots \\ A^{rest}_{oi} & & \vdots & A_{oi} \end{bmatrix} \begin{bmatrix} \pi_{hi} & \vdots & \hat{\pi}_{hi} \\ \cdots\cdots & \vdots & \cdots\cdots \\ \Pi_{hi} & \vdots & \hat{\Pi}_{hi} \\ \cdots\cdots & \vdots & \cdots\cdots \\ \Pi^{rest}_{hi} & \vdots & \hat{\Pi}^{rest}_{hi} \end{bmatrix} = \begin{bmatrix} \alpha'_{hi} & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots \\ A^{rest}_{hi} & \vdots & A_{hi} \end{bmatrix},$$

which implies

$$\pi_{hi} + \alpha'_{oi}\,\Pi_{hi} = \alpha'_{hi}\,, \qquad h=1,\ldots,p, \tag{4.3.2}$$

and

$$\pi_{hi} + \alpha'_{oi}\,\hat{\Pi}_{hi} = 0\,, \qquad h=1,\ldots,p. \tag{4.3.3}$$

LEMMA 4.3.4 $r\,[\hat{\Pi}_{1i},\ldots,\hat{\Pi}_{pi}] + m - m_i - 1 = r\,[A_{oi},\ldots,A_{pi}$

PROOF. From (4.3.1) we obtain

$$\begin{matrix}(1)\\(m_i)\\(m-m_i-1)\end{matrix}\begin{bmatrix}\hat{\pi}_{hi}\\\cdots\cdots\\\hat{\Pi}_{hi}\\\cdots\cdots\\\hat{\Pi}^{rest}_{hi}\end{bmatrix} = A_o^{-1}\begin{matrix}\\\\\\\\\end{matrix}\begin{bmatrix}0\\\\\cdots\\\\A_{hi}\end{bmatrix}\begin{matrix}(1)\\\\\\\\(m-1)\end{matrix}\,, \qquad h=1 \tag{4.3.4}$$

We also have

$$\begin{matrix}(1)\\(m_i)\\(m-m_i-1)\end{matrix}\begin{bmatrix}0\\\cdots\cdots\\0\\\cdots\cdots\\I_{m-m_i-1}\end{bmatrix} = A_o^{-1}\begin{bmatrix}0\\\cdots\cdots\\A_{oi}\end{bmatrix}\begin{matrix}(1)\\\\\\(m-1)\end{matrix}\,. \tag{4.3.5}$$

Combining (4.3.4) and (4.3.5) yields

$$\left[\begin{array}{c:c}0 & \hat{\pi}_{hi}\\\cdots\cdots & \cdots\cdots\\0 & \hat{\Pi}_{hi}\\\cdots\cdots & \cdots\cdots\\I_{m-m_i-1} & \hat{\Pi}^{rest}_{hi}\end{array}\right] = A_o^{-1}\left[\begin{array}{c:c}0 & 0\\\cdots\cdots & \cdots\cdots\\A_{oi} & A_{hi}\end{array}\right]\,, \qquad h=1,\ldots \tag{4.3.6}$$

From (4.3.3) it follows that $\hat{\pi}_{hi}$ is linearly expressible in the rows of $\hat{\Pi}_{hi}$ and so (4.3.6) implies

$$(4.3.7) \quad r[\hat{\Pi}_{hi}] + m-m_i-1 = r[A_{oi}, A_{hi}], \quad h=1,\ldots,p,$$

which implies the result. □

We can now prove the *rank condition for the reduced form*.

THEOREM 4.3.5 Let the sample size n be second-order informative for the reduced form of a dynamic system of simultaneous equations. Then a sufficient condition in order to be second-order informative for the i^{th} structural equation is $r[\hat{\Pi}_{1i},\ldots,\hat{\Pi}_{pi}] = m_i$, $\Theta \in \Theta$. If $q=0$ and $\Sigma_{\underset{\sim}{\varepsilon}}$ is completely unknown, then this condition is also necessary.

PROOF. If $r[\hat{\Pi}_{1i},\ldots,\hat{\Pi}_{pi}] = m_i$, $\Theta \in \Theta$, then the equations (4.3.3) can be solved uniquely for α_{oi}, and so $\alpha_{1i},\ldots,\alpha_{pi}$ are uniquely determined by (4.3.2). Hence by theorem 4.3.2 the sample size n is second-order informative for $\alpha_{oi},\ldots,\alpha_{pi}$ or equivalently for the i^{th} structural equation.
Suppose $q=0$, that $\Sigma_{\underset{\sim}{\varepsilon}}$ is completely unknown and that $r[\hat{\Pi}_{1i},\ldots\hat{\Pi}_{pi}] < m_i$. Then there exists $s \in \mathbb{C}^{m_i}$, $s \neq 0$ such that $s' \Pi_{hi} = 0$, $h=1,2,\ldots,p$. Since the function Φ defined by $\Phi(\delta(\Theta)) = (\alpha_{oi},\ldots,\alpha_{pi})$, $\Theta \in \Theta$ is continuous and Δ is open, $\Phi(\Delta)$ is also open. Hence there exists $\varepsilon > 0$ such that $(\tilde{\alpha}_{oi},\ldots,\tilde{\alpha}_{pi}) \in \Phi(\Delta)$, where

$$\tilde{\alpha}_{oi} := \alpha_{oi} + \varepsilon s ,$$

and

$$\tilde{\alpha}'_{hi} := \alpha_{hi} + \varepsilon s' \Pi_{hi} , \qquad h=1,2,\ldots,p.$$

Since also $(\tilde{\alpha}_{oi},\ldots,\tilde{\alpha}_{pi})$ solves (4.3.2) and (4.3.3), and $q=0$, it follows from theorem 4.3.2 that the sample size n is not

second-order informative for $(\alpha_{oi},\ldots,\alpha_{pi})$. □

COROLLARY 4.3.6 (*order condition*) If q=0 and $\Sigma_{\tilde{\varepsilon}}$ is completely unknown, a necessary condition for identifiability of the i^{th} structural equation is

$$mp \geq m_i + \sum_{h=1}^{p} k_{ih}.$$

PROOF. Follows at once from the fact that $[\hat{\Pi}_{1i},\ldots\hat{\Pi}_{pi}]$ is an $m_i \times \sum_{h=1}^{p}(m-k_{ih})$ matrix which can only have rank m_i if $m_i \leq \sum_{h=1}^{p}(m-k_{ih})$. □

In practice the following theorem is more efficient than theorem 4.3.5, since it does not require calculation of the reduced form.

THEOREM 4.3.7 (*rank condition for the structural form*) Let the sample size n be second-order informative for the reduced form of a dynamic system of simultaneous equations. Then a sufficient condition for that sample size to be second-order informative for the i^{th} structural equation is $r\,[A_{oi},\ldots,A_{pi}] = m-1$, $\theta\in\theta$. If q=0 and Σ_{ε} is completely unknown, then the condition is also necessary.

PROOF. Follows immediately from (4.3.7) and theorem 4.3.5. □

4.4 THE NON-HOMOGENEOUS CASE

In this section we shall treat the case where the system of simultaneous equations is allowed to contain a deterministic component $\mu_t \in \mathbb{C}^m$. Consider the system

$$(4.4.1)\qquad \sum_{h=0}^{p} A_h \underline{x}_{t-h} + \mu_t = \sum_{h=0}^{q} B_h \underline{\varepsilon}_{t-k}\,, \qquad t=0, \pm 1,\ldots\ .$$

We shall suppose that the homogeneous system satisfies assumptions a) - e) and that $\mu_t = Cu_t \quad t=0,\pm1,\ldots$ where $u_t \in \mathbb{C}^k$ is a known sequence of non-random regressors. Furthermore, we assume that

$$C = \begin{bmatrix} C^{(1)} \\ \cdots \\ C^{(2)} \end{bmatrix} \begin{matrix} (m-m_o) \\ \\ (m_o) \end{matrix} \quad .$$
$$(k)$$

With $C^{(2)}$ known and $C^{(1)}$ unknown. In econometrics u_t is called the vector of *exogenous* variables (and $\underline{x}_t$ the vector of *endogenous* variables). First of all we have to define the process $\{\underline{x}_t\}$ properly.

DEFINITION 4.4.1 The sequence $\{u_t\}$ is said to be *non-exponentially increasing* as $t\to-\infty$, if for all $\rho>1$ we have

$$\lim_{t\to-\infty} \rho^t u_t = 0.$$

From now on we shall assume that $\{u_t\}$ is non-exponentially increasing as $t\to-\infty$. Note that then also $\{\mu_t\}$ is non-exponentially increasing.
Suppose $\det A(z) \neq 0$, $|z| \leq 1$. Then $A^{-1}(z)$ can be expanded into a power series

$$A^{-1}(z) = \sum_{j=0}^{\infty} V_j z^j , \quad |z| \leq \rho_o$$

for a $\rho_o > 1$, and so we have

$$\lim_{j\to\infty} V_j \rho_o^j = 0.$$

Hence, $\|V_j\| \leq M_1 \rho_o^{-j}$ for some $M_1 > 0$. For $1<\rho<\rho_o$ we also have $\|\mu_{t-j}\| \leq M_2 \rho^j$.
So

$$\|V_j\ \mu_{t-j}\| \le \|V_j\| \|\mu_{t-j}\| \le M_1 M_2\ (\frac{\rho}{\rho_o})^j \quad ,$$

which implies that $\sum_{j=0}^{\infty} V_j\ \mu_{t-j}$ is convergent.

Put

$$(4.4.2) \qquad \underline{Y}_t = \underline{x}_t - \sum_{j=0}^{\infty} V_j\ \mu_{t-j} \quad , \qquad t=0, \pm 1, \ldots \quad .$$

Then $\{\underline{Y}_t\}$ satisfies the homogeneous system of equations. Therefore we define the process $\{\underline{x}_t\}$ by (4.4.2), where $\{\underline{Y}_t\}$ is the unique weakly stationary solution of the homogeneous system. Then it is easily seen that $\{\underline{x}_t\}$ is the unique *covariance stationary* solution of (4.4.1).

Let

$$\Theta = (A_o^{(1)}; A_1^{(1)} \ldots A_p^{(1)};\ B_1^{(o)} \ldots B_q^{(o)};\ \xi\ ;\ C^{(1)}) \quad ,$$

and

$$\Theta = S^{DSE} \times Z \times W \quad ,$$

where W (the range of $C^{(1)}$) is some subset of $(\mathbb{C}^{m-m_o})^k$.
At first we consider the reduced form i.e.

$$(4.4.3) \qquad \sum_{h=0}^{p} P_h \underline{x}_{t-h} + A_o^{-1}\ C\ u_t = \sum_{h=0}^{q} Q_h \underline{\eta}_{t-h} \quad , \quad t=0, \pm 1, \ldots \quad .$$

A sample is called second-order informative for the reduced form if it is second-order informative for $(P(z),\ Q(z),\ \Sigma_\eta,\ A_o^{-1}C)$.
We shall assume that the model is *non-collinear* i.e. that there exists $n_o \in \mathbb{N}$ such that

$$r\left[\sum_{t=0}^{n_o} u_t\ u_t^*\right] = k.$$

THEOREM 4.4.1 If in a non-homogeneous dynamic system of simultaneous equations $\theta = S^{DSE} \times Z \times W$, then the sample size max $(n_o+1+p,\ q+(m+1)p)$ is second-order informative for the reduced form.

PROOF. Since the covariance function of the process $\{\underline{x}_t\}$ is generated by the spectral measure of the process $\{\underline{y}_t\}$ defined by (4.4.2), it follows at once from theorem 4.2.4 that the sample size $q+(m+1)p$ is second-order informative for $(P(z),\ Q(z),\ \Sigma_\eta)$.

Let $\underline{s}_t := \sum_{h=0}^{p} P_h\, \underline{x}_{t-h}$, $t=0\pm1,\ldots$ Then it follows from (4.4.3) that a sample of size n_o+1 from the process $\{\underline{s}_t\}$ is second-order (even first-order) informative for $A_o^{-1}C$. Hence a sample of size n_o+1+p from $\{\underline{x}_t\}$ is second-order informative for $A_o^{-1}C$ conditional on $P(z)$. By theorem 1.2.3 then the sample size max $(n_o+1+p,\ q+(m+1)p)$ is second-order informative for $A_o^{-1}C$. This proves the theorem. □

As in §4.3 we shall now consider the i^{th} structural equation. After re-ordering the equations we may partition the matrix C in a similar way as A_h,

$$C = \begin{bmatrix} \gamma_i & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots \\ C_i^{rest} & \vdots & C_i \end{bmatrix} \begin{matrix} (1) \\ \\ (m-1) \end{matrix}$$

$$(c_i) \qquad (k-c_i)$$

So c_i is the number of exogenous variables in the i^{th} equation. Now the relation $A_o\ (A_o^{-1}C) = C$ plays exactly the same role as the relation $A_o\ P_h = A_h$ did in §4.3. Therefore we can immediately generalize the results of §4.3 to the non-homogeneous case and obtain the following theorems.

THEOREM 4.4.2 (*order condition*). If $q=0$ and $\Sigma_{\underset{\sim}{\varepsilon}}$ is completely

unknown, a necessary condition for identifiability of the i^{th} structural equation is

$$mp \geq m_i + c_i + \sum_{h=1}^{p} k_{ih}.$$

THEOREM 4.4.3 (*rank condition for the structural form*). Let the sample size n be second-order informative for the reduced form of a non-homogeneous dynamic system of simultaneous equations. Then a sufficient condition for that sample size to be second-order informative for the i^{th} structural equation is $r\,[A_{oi},\ldots,A_{pi}, C_i] = m-1$, $\theta \in \Theta$. If $q=0$ and $\Sigma_{\tilde{\varepsilon}}$ is completely unknown, then the condition is also necessary.

The proofs of these theorems are similar to the proofs of corollary 4.3.6 and theorem 4.3.7 and are omitted.

To illustrate the results, we shall give an example from econometrics.

EXAMPLE 4.4.4 Consider the following system of simultaneous equations

$$\begin{cases} \underline{x}_{t1} = \delta_{01} + \delta_{11}\,\underline{x}_{t-1,3} + \delta_{21}\,\underline{x}_{t3} + \underline{\varepsilon}_{t1} + \alpha_1\,\underline{\varepsilon}_{t-1,1} \\ \underline{x}_{t2} = \delta_{02} + \delta_{12}\,\underline{x}_{t3} + \underline{\varepsilon}_{t2} + \alpha_2\,\underline{\varepsilon}_{t-1,2} \\ \underline{x}_{t3} = \underline{x}_{t1} + \underline{x}_{t2} - v_t \end{cases}$$

where

$\underline{x}_{t1}$ = consumption in period t ,

$\underline{x}_{t2}$ = net investments in period t ,

$\underline{x}_{t3}$ = national income in period t ,

v_t = other expenses in period t .

In the notation of the preceding sections we have $p=q=1$, $m=3$, $m_o=1$, $p_o=0$ and

$$A(z) = \left[\begin{array}{cc:c} 1 & 0 & -(\delta_{21}+\delta_{11}z) \\ 0 & 1 & -\delta_{12} \\ \hdashline -1 & -1 & 1 \end{array}\right] \quad , \; B^{(o)}(z) = \begin{bmatrix} 1+\alpha_1 z & 0 \\ 0 & 1+\alpha_2 z \end{bmatrix} .$$

First of all we see that degree $\{A^{(12)}(z)\} = 1 > p_o$, so we cannot apply theorem 4.4.1. Therefore we substitute $\underline{x}_{t-1,3} = \underline{x}_{t-1,1} + \underline{x}_{t-1,2} - v_{t-1}$ into the first equation and obtain a model with coefficient generating function

$$\tilde{A}(z) := \left[\begin{array}{cc:c} 1-\delta_{11}z & -\delta_{11}z & -\delta_{21} \\ 0 & 1 & -\delta_{12} \\ \hdashline -1 & -1 & 1 \end{array}\right] .$$

Note, that there is a 1-1 correspondence between the coefficients of $A(z)$ and $\tilde{A}(z)$. Hence, identifiability of $\tilde{A}(z)$ is equivalent to identifiability of $A(z)$. It should also be noted that the substitution introduces v_{t-1} into the first equation. Therefore we have to take

$$u_t = \begin{bmatrix} 1 \\ v_t \\ v_{t-1} \end{bmatrix} \quad , \; C = \begin{bmatrix} -\delta_{01} & 0 & \delta_{11} \\ -\delta_{02} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} .$$

Note, that the model is non-collinear iff the sequence $\{v_t\}$ does not satisfy a first-order linear difference equation. We shall suppose that $n_o=2$, i.e. that $r\left[\sum_{t=0}^{2} u_t u_t^*\right] = 3$. In order that $\theta \subset S^{DSE} \times Z \times W$, we must have

1) $\det \tilde{A}(z) \neq 0, \ |z| \leq 1$ or, equivalently

$$|\delta_{11}| < |1 - \delta_{12} - \delta_{21}| .$$

2) $\det B^{(o)}(z) \neq 0, \ |z| < 1$ or, equivalently

$$|\alpha_1| \leq 1 \wedge |\alpha_2| \leq 1.$$

3) $$r[\tilde{A}_p^{(11)}, B_q^{(o)}] = r \begin{bmatrix} -\delta_{11} & -\delta_{11} & \vdots & \alpha_1 & 0 \\ & & \vdots & & \\ 0 & 0 & \vdots & 0 & \alpha_2 \end{bmatrix} = m-m_o = 2.$$

Hence, $(\delta_{11} \neq 0 \vee \alpha_1 \neq 0) \vee \alpha_2 \neq 0$.

4) $$r[\tilde{A}(z), B(z)] = r \begin{bmatrix} 1-\delta_{11}z & -\delta_{11}z & -\delta_{21} & \vdots & 1+\alpha_1 z & 0 & 0 \\ 0 & 1 & -\delta_{12} & \vdots & 0 & 1+\alpha_2 z & 0 \\ -1 & -1 & 1 & \vdots & 0 & 0 & 0 \end{bmatrix} = 3$$

for all z. This is the case iff

$$\alpha_1 \delta_{11} \neq 1-\delta_{12}-\delta_{21} \vee \alpha_2 \delta_{11} \neq 1-\delta_{12}-\delta_{21} .$$

Thus by theorem 4.4.1 we obtain that the sample size $\max(n_o+1+p, q+(m+1)p) = \max(4,5) = 5$ is second-order informative for the reduced form, if the conditions 1) - 4) are satisfied.

Finally we shall consider the rank condition for the structural form (theorem 4.4.3).

For the first equation we have

$$r[A_{o1}, A_{11}, C_1] = r \begin{bmatrix} 1 & \vdots & 0 & \vdots & 0 \\ & \vdots & & \vdots & \\ -1 & \vdots & 0 & \vdots & 1 \end{bmatrix} = 2 = m-1 \quad ,$$

and for the second equation

$$r\,[A_{o2},A_{12},C_2] = r\begin{bmatrix} 1 & -\delta_{11} & -\delta_{11} & 0 & 0 & \delta_{11} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} = 2 = m-1 \ .$$

Thus the sample size 5 is second order-informative for both equations if the coefficients satisfy conditions 1) - 4).

4.5 THE NON MIXED CASE

Since the results in this chapter are based on the theorems on the m-variate ARMA(p,q) model, we may expect sharper results in the cases p=0 or q=0. In fact we have for the system (4.4.1) whith p=0 that the sample size $\max(n_o+1,q+1)$ is second-order informative for the reduced form. In the case q=0 we find the second-order informative sample size $\max(n_o+1+p,p+1)=n_o+1+p$ for the reduced form. The proofs for these statements go along the same lines as the proof for the mixed case and shall be omitted. In the case of a large number of equations, which is generally the case in econometrics, the obtained reduction of the informative sample size can be important.

APPENDIX

A.1 UNIVARIATE SPECTRAL THEORY AND STOCHASTIC DIFFERENCE EQUATIONS

Let Λ be a finite dimensional Euclidean space and let $\mathfrak{B}(\Lambda)$ denote the Borel field of Λ. Suppose we have a σ-finite measure μ on $(\Lambda, \mathfrak{B}(\Lambda))$ and a stochastic process $\{\underline{z}\{s\},\ s \in \mathfrak{B}(\Lambda)\}$ such that

$$\text{(A.1.1)} \qquad \underline{z}\{s_1 \cup s_2\} = \underline{z}\{s_1\} + \underline{z}\{s_2\} \text{ if } s_1 \cap s_2 = \emptyset ,$$

$$\text{(A.1.2)} \qquad E\{\underline{z}\{s_1\}\overline{\underline{z}\{s_2\}}\} = \mu\{s_1 \cap s_2\} , \quad s_1, s_2 \in \mathfrak{B}(\Lambda) .$$

Such processes are called *random measures with orthogonal increments*. From (A.1.1) and A.1.2) it follows that for disjoint s_j with $\sum_{j=1}^{\infty} \mu\{s_j\} < \infty$, we have

$$\underline{z}\{\bigcup_{j=1}^{\infty} s_j\} = \sum_{j=1}^{\infty} \underline{z}\{s_j\} \quad \text{a.s.} ,$$

where the convergence of the series on the right is in mean square.

Let $L^2(\mu)$ denote the (Hilbert) space of functions $f : \Lambda \to \mathbb{C}$ that are square integrable with respect to μ. We shall give a brief exposition how we may define stochastic integrals of the type

$$\int_{\Lambda} f(\lambda)\underline{z}\{d\lambda\},\ f \in L^2(\mu), \quad \text{or} \quad \int_{\Lambda} f\, d\underline{z} \text{, for short.}$$

We follow GRENANDER and ROSENBLATT [16] p. 25-27. If $s \in \mathfrak{B}(\Lambda)$ with $\mu\{s\} < \infty$, then we define

$$\int_{\Lambda} 1_s\, d\underline{z} \ := \underline{z}\{s\},$$

where 1_{s_n} stands for the indicatorfunction of s. Similarly, for *simple functions* $f = \sum_{k=1} \alpha_k 1_{s(k)}$ we put

$$(A.1.3) \qquad \int_\Lambda \sum_{k=1}^{n} \alpha_k 1_{s(k)} d\underline{z} = \sum_{k=1}^{n} \alpha_k \underline{z}\{s(k)\}.$$

It is not difficult to see that the integral (A.1.3) does not depend on the specific representation of f. From (A.1.1) and (A.1.2) it follows, that for disjoint s(k) we have

$$(A.1.4) \qquad E\{\left|\int_\Lambda f d\underline{z}\right|^2\} = \sum_{k=1}^{n} |\alpha_k|^2 \mu\{s(k)\} = \int_\Lambda |f|^2 d\mu \ .$$

For arbitrary $f \in L^2(\mu)$ we can find a sequence of simple functions f_n $n = 1, 2, \ldots$, such that f_n converges to f in the norm of $L^2(\mu)$, i.e.

$$\lim_{n\to\infty} \int_\Lambda |f-f_n|^2 d\mu = 0.$$

Then we have

$$E\{\left|\int_\Lambda f_n d\underline{z} - \int_\Lambda f_m d\underline{z}\right|^2\} = \int_\Lambda |f_n - f_m|^2 d\mu \to 0, \ n,m\to\infty \ .$$

Hence there exists a random variable $\underline{y}$ with

$$\text{l.i.m.}_{n\to\infty} \int_\Lambda f_n d\underline{z} = \underline{y},$$

and

$$E\{|\underline{y}|^2\} = \lim_{n\to\infty} E\{\left|\int_\Lambda f_n d\underline{z}\right|^2\} = \int_\Lambda |f|^2 d\mu \ .$$

It is not difficult to see that $\underline{y}$ does not depend on the particular sequence f_n. Therefore we may define

$$\int_{\Lambda} f d\underline{z} \; := \; \underset{n\to\infty}{\text{l.i.m.}} \int_{\Lambda} f_n d\underline{z}.$$

The integral is now defined for all $f \in L^2(\mu)$ and has the following properties .

$$\text{(A.1.5)} \qquad \int_{\Lambda} (\alpha f+\beta g) d\underline{z} = \alpha \int_{\Lambda} f d\underline{z} + \beta \int_{\Lambda} g d\underline{z} \;, \qquad f,g \in L^2(\mu) \;, \quad \alpha,\beta \in \mathbb{C}$$

$$\text{(A.1.6)} \qquad \underset{n\to\infty}{\text{l.i.m.}} \int_{\Lambda} f_n d\underline{z} = \int_{\Lambda} f d\underline{z} \iff \lim_{n\to\infty} \int_{\Lambda} |f-f_n|^2 d\mu = 0 \qquad f, f_n \in L^2(\mu) \;,$$

$$\text{(A.1.7)} \qquad E\{\int_{\Lambda} f d\underline{z} \overline{\int_{\Lambda} g d\underline{z}}\} = \int_{\Lambda} f\bar{g} d\mu \;, \qquad f,g \in L^2(\mu) \;,$$

in particular

$$E\{|\int_{\Lambda} f d\underline{z}|^2\} = \int_{\Lambda} |f|^2 d\mu \;, \qquad f \in L^2(\mu) \;.$$

The proofs are straightforward and will be omitted (see e.g. [33]). Finally, we define $\int_s f d\underline{z}$ as $\int_{\Lambda} 1_s f d\underline{z}$ for all $s \in \mathfrak{B}(\Lambda)$ and $f \in L^2(\mu)$.

Consider the stochastic process $\{\underline{z}_1\{s\}, s \in \mathfrak{B}(\Lambda)\}$ defined by

$$\underline{z}_1\{s\} \; := \; \int_s f d\underline{z} \;,$$

for some fixed $f \in L^2(\mu)$. Then $\underline{z}_1\{s\}$ is also a random measure with orthogonal increments since (A.1.1) is trivially satisfied and by (A.1.7)

$$E\{\underline{z}_1\{s_1\}\overline{\underline{z}_1\{s_2\}}\} = \int_\Lambda 1_{s_1 \cap s_2} |f|^2 \, d\mu =$$

$$= \int_{s_1 \cap s_2} |f|^2 \, d\mu = \mu_1\{s_1 \cap s_2\} ,$$

where μ_1 is the measure on $(\Lambda, \mathcal{B}(\Lambda))$ which is absolutely continuous w.r.t. μ with density $|f|^2$. Using the usual differential formalism we write $d\mu_1 = |f|^2 \, d\mu$, and in analogy to this formalism we shall write $d\underline{z}_1 = f d\underline{z}$.

From now on we specialize to the case where $\Lambda = (-\pi, \pi]$ and for μ we take a *finite* measure F. Thus we have $F\{(-\pi,\pi]\} < \infty$ and in particular we have $e^{it\lambda} \in L^2(F)$ for all $t \in \mathbb{R}$. Consequently we can define a stochastic process $\{\underline{x}_t, \; t=0,\pm 1,\ldots\}$ by

$$\underline{x}_t := \int_{(-\pi,\pi]} e^{it\lambda} \, \underline{z}\{d\lambda\} , \qquad t = 0, \pm 1, \ldots . \tag{A.1.8}$$

By (A.1.7) we obtain

$$E\{\underline{x}_t \, \overline{\underline{x}}_s\} = \int_{(-\pi,\pi]} e^{i(t-s)\lambda} F\{d\lambda\} , \qquad t,s = 0, \pm 1, \ldots \tag{A.1.9}$$

Hence the process $\{\underline{x}_t\}$ is weakly stationary. It can be shown that every weakly stationary process can be represented in the form (A.1.8) and that this representation is unique (see e.g. GRENANDER and ROSENBLATT [16] p. 27-29). It is then called the *spectral representation* of the process and (A.1.9) is then called the spectral representation of the product moment function

(or the covariance function if $E\{\underline{x}_t\} = 0$, $t = 0, \pm 1, \ldots$). The measure F is called the *spectral measure* of $\{\underline{x}_t\}$, and $\underline{z}\{.\}$ is called the random measure with orthogonal increments *associated* with $\{\underline{x}_t\}$. A theorem due to HERGLOTZ ([11] p. 634-635) states that a finite measure F on $(-\pi, \pi]$ is uniquely determined by the sequence of its *Fouriercoefficients* γ_s i.e. by

$$\gamma_s = \int_{(-\pi,\pi]} e^{is\lambda} F\{d\lambda\}, \quad s = 0, \pm 1, \ldots .$$

Hence by (A.1.9) we have a 1-1 correspondence between covariance functions of weakly stationary processes and finite measures concentrated on $(-\pi, \pi]$. We shall use the notations F_x and $\underline{z}_x\{.\}$ for the spectral measure and random measure with orthogonal increments associated with the weakly stationary process $\{\underline{x}_t\}$, and use a similar notation when other weakly stationary processes are involved.

A linear operation that transforms the process $\{\underline{x}_t\}$ into $\{\underline{y}_t\}$ such that (in mean square)

$$\underline{y}_t = \sum_{k=0}^{\infty} c_k \underline{x}_{t-k}, \qquad t = 0, \pm 1, \ldots$$

is called a (time invariant) *linear filter*. It follows from (A.1.5) and (A.1.6) that

$$\underline{y}_t = \int_{(-\pi,\pi]} e^{it\lambda} \left(\sum_{k=0}^{\infty} c_k e^{-ik\lambda} \right) \underline{z}_x\{d\lambda\}, \quad t = 0, \pm 1, \ldots$$

where the convergence of $\sum_{k=0}^{\infty} c_k e^{-ik\lambda}$ is in the norm of $L^2(F_x)$.

The function $\varphi(\lambda) := \sum_{k=0}^{\infty} c_k e^{-ik\lambda}$ is called the *frequency response function* of the linear filter. The following theorem on linear filters is fundamental in the theory of stochastic linear difference equations.

THEOREM A.1.1 Let $\{\underline{x}_t\}$ be weakly stationary and let $\underline{Y}_t := \sum_{k=0}^{\infty} c_k \underline{x}_{t-k}$ $t = 0, \pm 1, \ldots$. If $\varphi(\lambda) = \sum_{k=0}^{\infty} c_k e^{-ik\lambda}$, $\lambda \in (-\pi, \pi]$ then $1/\varphi \in L^2(F_y)$.

Furthermore there exist weakly stationary processes $\{\tilde{\underline{x}}_t\}$ and $\{\underline{u}_t\}$, mutually orthogonal, with

$$\left.\begin{aligned} \underline{x}_t &= \tilde{\underline{x}}_t + \underline{u}_t \\ \underline{Y}_t &= \sum_{k=0}^{\infty} c_k \tilde{\underline{x}}_{t-k} \\ \tilde{\underline{x}}_t &= \int_{(-\pi,\pi]} \frac{e^{it\lambda}}{\varphi(\lambda)} \underline{z}_y\{d\lambda\} \end{aligned}\right\} \qquad t = 0, \pm 1, \ldots \ .$$

PROOF. Define $1/\varphi$ constant, (for measurability) on the set

$$E_o = \{\lambda \mid \varphi(\lambda) = 0\}.$$

Since $d\underline{z}_y = \varphi d\underline{z}_x$, we have $dF_y = |\varphi|^2 \, dF_x$ and so $F_y\{E_o\} = 0$. Hence,

$$\int_{(-\pi,\pi]} \frac{1}{|\varphi|^2} \, dF_y = \int_{E_o^c} \frac{1}{|\varphi|^2} \, dF_y = \int_{E_o^c} dF_x = F_x\{E_o^c\} < \infty \ .$$

Thus $1/\varphi \in L^2(F_y)$ and we may define

$$\tilde{\underline{x}}_t = \int_{(-\pi,\pi]} \frac{e^{it\lambda}}{\varphi(\lambda)} \underline{z}_y\{d\lambda\} \ , \qquad t = 0, \pm 1, \ldots \ .$$

We have

$$\tilde{\underline{x}}_t = \int_{E_o^c} \frac{e^{it\lambda}}{\varphi(\lambda)} \underline{z}_y\{d\lambda\} = \int_{E_o^c} \frac{e^{it\lambda}}{\varphi(\lambda)} \varphi(\lambda) \underline{z}_x\{d\lambda\} =$$

$$= \int_{E_o^c} e^{it\lambda}\, \underline{z}_x\{d\lambda\} \quad , \qquad t = 0, \pm 1, \ldots \quad .$$

If $E_o = \emptyset$ we can take $\underline{u}_t = 0$, $t = 0, \pm 1, \ldots$ and the theorem is proved. If $E_o \neq \emptyset$ we put

$$\underline{u}_t = \int_{E_o} e^{it\lambda}\, \underline{z}_x\{d\lambda\} \quad , \qquad t = 0, \pm 1, \ldots \quad .$$

Then $\{\underline{u}_t\}$ is weakly stationary, orthogonal w.r.t. $\{\tilde{\underline{x}}_t\}$ and $\underline{u}_t + \tilde{\underline{x}}_t = \underline{x}_t$ a.s. $t = 0, \pm 1, \ldots$ Furthermore,

$$\sum_{k=0}^{\infty} c_k \underline{u}_{t-k} = \int_{E_o} \varphi(\lambda)\, e^{it\lambda}\, \underline{z}_x\{d\lambda\} = 0 \quad , \quad t = 0, \pm 1, \ldots \quad .$$

Hence

$$\sum_{k=0}^{\infty} c_k \tilde{\underline{x}}_{t-k} = \sum_{k=0}^{\infty} c_k \underline{x}_{t-k} - \sum_{k=0}^{\infty} c_k \underline{u}_{t-k} = \sum_{k=0}^{\infty} c_k \underline{x}_{t-k} = \underline{y}_t \quad . \square$$

Consider the homogeneous difference equation

$$\text{(A.1.10)} \qquad \sum_{k=0}^{\infty} a_k \underline{x}_{t-k} = 0 \quad , \qquad t = 0, \pm 1, \ldots \quad ,$$

where $\sum_{k=0}^{\infty} |a_k| < \infty$. We are interested in nonzero weakly stationary solutions. Let $A(z) = \sum_{k=0}^{\infty} a_k z^k$, $z \in \mathbb{C}$. By the bounded convergence theorem of Lebesgue, it follows that $A(z)$ is continuous on $|z| \leq 1$, and so for all finite measures F on $(-\pi, \pi]$ we have

$$A(e^{-i\lambda}) \in L^2(F) \quad .$$

THEOREM A.1.2 The stochastic difference equation (A.1.10) has

a nonzero weakly stationary solution iff the set $E_o := \{\lambda \,|\, A(e^{-i\lambda}) = 0\}$ is nonempty. All weakly stationary solutions $\{\underline{x}_t\}$ satisfy $F_x\{E_o^c\} = 0$.

PROOF. ($\Rightarrow$) Let $\{\underline{x}_t\}$ be a nonzero weakly stationary solution. Then we have

$$\int_{(-\pi,\pi]} e^{it\lambda} A(e^{-i\lambda}) \, \underline{z}_x\{d\lambda\} = 0, \qquad t = 0, \pm 1, \ldots \ .$$

Hence

$$\int_{(-\pi,\pi]} |A(e^{-i\lambda})|^2 F_x\{d\lambda\} = 0,$$

which implies that $A(e^{-i\lambda}) = 0$ F_x-a.e., or equivalently $F_x\{E_o^c\}=0$. Thus, since $\{\underline{x}_t\}$ is a nonzero solution,

$$F_x\{E_o\} = F_x\{(-\pi,\pi]\} = E\{|\underline{x}_t|^2\} > 0,$$

and so E_o must be nonempty.

($\Leftarrow$) If $\lambda_o \in E_o$ and $\underline{z}$ is an arbitrary random variable with $0 < E\{|\underline{z}|^2\} < \infty$, then a weakly stationary solution of (A.1.10) is given by

$$\underline{x}_t = \underline{z}\, e^{-it\lambda_o}, \qquad t = 0, \pm 1, \ldots \qquad \square$$

COROLLARY A.1.3 If $A(z)$ is analytic on $|z| < \rho$ for some $\rho > 1$, there are at most finitely many zeros on $|z| = 1$ and so every weakly stationary solution has a spectral measure concentrated on a finite set.

Next we consider the non homogeneous difference equation that corresponds to (A.1.10),

$$\text{(A.1.11)} \qquad \sum_{k=0}^{\infty} a_k \underline{x}_{t-k} = \underline{y}_t \ , \qquad t = 0, \pm 1, \ldots$$

where $\{\underline{y}_t\}$ is a given weakly stationary process. We have

THEOREM A.1.4 The stochastic difference equation (A.1.11) has a weakly stationary solution $\{\underline{x}_t\}$ iff $A^{-1}(e^{-i\lambda}) \in L^2(F_y)$, where $A^{-1}(e^{-i\lambda})$ is defined to be ∞ on the set $E_o := \{\lambda \,|\, A(e^{-i\lambda}) = 0\}$. Every weakly stationary solution $\{\underline{x}_t\}$ can be written in the form

$$\underline{x}_t = \int_{(-\pi,\pi]} \frac{e^{it\lambda}}{A(e^{-i\lambda})} \underline{z}_y\{d\lambda\} + \underline{u}_t \ , \quad t = 0, \pm 1, \ldots .$$

where the process $\{\underline{u}_t\}$ is a weakly stationary solution of the corresponding homogenous equation, which is orthogonal to $\{\underline{x}_t\}$.

PROOF. ($\Leftarrow$) If $A^{-1}(e^{-i\lambda}) \in L^2(F_y)$ then we must have $F_y\{E_o\} = 0$. Define

$$\underline{x}_t = \int_{(-\pi,\pi]} \frac{e^{it\lambda}}{A(e^{-i\lambda})} \underline{z}_y\{d\lambda\} \ , \qquad t = 0, \pm 1, \ldots .$$

Then

$$\sum_{k=0}^{\infty} a_k \underline{x}_{t-k} = \int_{(-\pi,\pi]} e^{it\lambda} \frac{A(e^{-i\lambda})}{A(e^{-i\lambda})} \underline{z}_y\{d\lambda\} =$$

$$= \int_{E_o^c} e^{it\lambda} \underline{z}_y\{d\lambda\} = \int_{(-\pi,\pi]} e^{it\lambda} \underline{z}_y\{d\lambda\} = \underline{y}_t \ , \quad t = 0, \pm 1, \ldots .$$

($\Rightarrow$) If the weakly stationary proces $\{\underline{x}_t\}$ satisfies (A.1.11),

it follows from theorem A.1.1 that we have $A^{-1}(e^{-i\lambda}) \in L^2(F_y)$, and that $\underline{x}_t$ is of the form stated in the theorem. □

It follows from theorems A.1.2 and A.1.4 that if $A(z) \neq 0$, $|z| = 1$, then the process $\{\underline{u}_t\}$ vanishes identically (a.s.) and

$$\underline{x}_t = \int_{(-\pi,\pi]} \frac{e^{it\lambda}}{A(e^{-i\lambda})} \underline{z}_y\{d\lambda\} \quad , \qquad t = 0, \pm 1, \ldots$$

is the unique weakly stationary solution of (A.1.11). The spectral measure is then absolutely continuous w.r.t. F_y and is given by

$$F_x\{d\lambda\} = \frac{1}{|A(e^{-i\lambda})|^2} F_y\{d\lambda\} \quad , \qquad \lambda \in (-\pi,\pi] .$$

If in addition F_y is absolutely continuous with spectral density f_y, then F_x is absolutely continuous with spectral density

$$f_x(\lambda) = \frac{f_y(\lambda)}{|A(e^{-i\lambda})|^2} \quad , \qquad \lambda \in (-\pi,\pi] .$$

In particular, if $\{\underline{y}_t\}$ is a moving average

$$\underline{y}_t = \sum_{k=0}^{\infty} b_k \underline{\varepsilon}_{t-k} \quad , \qquad t = 0, \pm 1, \ldots \quad ,$$

with $\{\underline{\varepsilon}_t\}$ white noise and $E\{|\varepsilon_t|^2\} = \sigma^2$, we find

$$f_x(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})} \right|^2 , \qquad \lambda \in (-\pi,\pi] ,$$

where the function $B(e^{-i\lambda})$ is defined as the limit in $L^2(\mu)$ norm (μ = Lebesgue measure),

$$B(e^{-i\lambda}) := \lim_{n\to\infty} \sum_{k=0}^{n} b_k e^{-ik\lambda}, \qquad \lambda \in (-\pi,\pi] .$$

It is easily seen to exist iff $\sum_{k=0}^{\infty} |b_k|^2 < \infty$.

A.2 SOME MULTIVARIATE SPECTRAL THEORY

As in the univariate case, let Λ denote a finite dimensional Euclidean space and $\mathfrak{B}(\Lambda)$ its Borelfield. Suppose we have a σ-finite measure μ taking values in the set of Hermitian positive semi definite $m \times m$ matrices. If $\{\underline{z}\{s\},\ s \in \mathfrak{B}(\Lambda)\}$ is an m-variate stochastic process with

$$\text{(A.2.1)} \qquad \underline{z}\{s_1 \cup s_2\} = \underline{z}\{s_1\} + \underline{z}\{s_2\} \qquad \text{if} \quad s_1 \cap s_2 = \emptyset ,$$

$$\text{(A.2.2)} \qquad E\{\underline{z}\{s_1\}\ \underline{z}^*\{s_2\}\} = \mu\{s_1 \cap s_2\} , \qquad s_1,\ s_2 \in \mathfrak{B}(\Lambda) ,$$

then $\underline{z}\{.\}$ is called an m-variate random measure with orthogonal increments.

Consider the real measure $\tilde{\mu} := \mathrm{tr}\mu$. Obviously the components of μ are absolutely continuous w.r.t. $\tilde{\mu}$ and so there exists a semi positive definite matrix valued function $\tilde{f}$ with $d\mu = \tilde{f} d\tilde{\mu}$. If φ is some $n \times m$ matrix function on Λ such that $\varphi\, \tilde{f}\, \varphi^*$ is integrable w.r.t. $\tilde{\mu}$ (componentwise) then we shall say that $\varphi \in L^2(\mu)$, and we denote: $\varphi d\mu \varphi^* = \varphi \tilde{f} \varphi^* d\tilde{\mu}$. Similar: $\varphi d\mu \Psi^* := \varphi \tilde{f} \Psi^* d\tilde{\mu}$. In a way similar to the univariate case we may now introduce stochastic integrals w.r.t. the random measure $\underline{z}$ for functions in $L^2(\mu)$. The properties corresponding to (A.1.5) - (A.1.7) are

$$\text{(A.2.3)} \qquad \int_\Lambda (Af + Bg)\, d\underline{z} = A\int_\Lambda f d\underline{z} + B\int_\Lambda g d\underline{z} \qquad \begin{array}{l} f,\ g \in L^2(\mu) \\ A,\ B \in \mathbb{C}(m) \end{array} ,$$

$$(A.2.4)\qquad \underset{n\to\infty}{\text{l.i.m}} \int_\Lambda f_n d\underline{z} = \int_\Lambda f d\underline{z} \iff \lim_{n\to\infty} \int_\Lambda (f-f_n)\, d\mu\, (f-f_n)^* = 0$$

$$f,\ f_n \in L^2(\mu)\ ,$$

$$(A.2.5)\qquad E\{\int_\Lambda f d\underline{z} (\int_\Lambda g d\underline{z})^*\} = \int_\Lambda f d\mu g\ , \qquad f,\ g \in L^2(\mu)\ .$$

As in the univariate case it can be shown that every m-variate weakly stationary process $\{\underline{x}_t\}$ has a spectral representation

$$\underline{x}_t = \int_{(-\pi,\pi]} e^{it\lambda}\ \underline{z}_x\{d\lambda\}\ , \qquad t = 0, \pm 1, \ldots\ ,$$

and

$$E\{\underline{x}_t \underline{x}_s^*\} = \int_{(-\pi,\pi]} e^{i(t-s)\lambda}\ F_x\{d\lambda\}\ ,\ \ t,s = 0, \pm 1, \ldots\ ,$$

where $\underline{z}_x$ is now an m-variate random measure with orthogonal increments and the spectral measure F_x takes semi positive definite m×m matrices as values. If tr F_x is absolutely continuous w.r.t. Lebesgue measure, F_x is said to be so and we can write $dF_x = f_x d\lambda$. The m×m matrixfunction f_x is called the spectral density matrix of the process. For details on multivariate spectral theory we refer to HANNAN [20].

Consider the process $\{\underline{y}_t\}$ obtained from $\{\underline{x}_t\}$ by the linear filter

$$\underline{y}_t = \sum_{k=0}^{\infty} C_k \underline{x}_{t-k}\ , \qquad t = 0, \pm 1, \ldots\ ,$$

where C_o, C_1, ... are n×m matrices. Then $\{\underline{y}_t\}$ is an n-variate weakly stationary process. Putting $C(z) := \sum_{k=0}^{\infty} C_k z^k$ we can write

$$\underline{Y}_t = \int_{(-\pi,\pi]} e^{it\lambda}\, C(e^{-i\lambda})\underline{z}_x\{d\lambda\} \quad , \qquad t = 0, \pm 1, \ldots \ ,$$

and by (A.2.5) it follows that

$$(A.2.6) \qquad F_y\{d\lambda\} = C(e^{-i\lambda})F_x\{d\lambda\}C^*(e^{-i\lambda}) \quad .$$

The multivariate generalizations of theorems A.1.1, A.1.2 and A.1.4 are now straightforward. As an example we shall prove the generalization of theorem A.1.2.

THEOREM A.2.1 Consider the m-variate stochastic difference equation

$$\sum_{k=0}^{\infty} A_k \underline{x}_{t-k} = 0 \quad , \qquad t = 0, \pm 1, \ldots \quad ,$$

where $A_0, A_1, \ldots$ are m×m matrices such that $\sum_{k=0}^{\infty} \|A_k\| < \infty$. Put $A(z) := \sum_{k=0}^{\infty} A_k z^k$. Then there exist a non-zero weakly stationary solution iff the set $E_o := \{\lambda \,|\, \det A(e^{-i\lambda}) = 0\}$ is nonempty. All weakly stationary solutions $\{\underline{x}_t\}$ satisfy $F_x\{E_o^c\} = 0$.

PROOF. ($\Rightarrow$) Let $\{\underline{x}_t\}$ be a nonzero weakly stationary solution. Then we can write

$$\int_{(-\pi,\pi]} e^{it\lambda} A(e^{-i\lambda})\underline{z}_x\{d\lambda\} = 0 \quad , \qquad t = 0, \pm 1, \ldots \ .$$

Hence

$$\int_{(-\pi,\pi]} A(e^{-i\lambda})F_x\{d\lambda\}A^*(e^{-i\lambda}) = 0 \quad .$$

Putting $\Phi_x = \operatorname{tr} F_x$ and $dF_x = \tilde{f}_x d\Phi_x$ this can equivalently be written

$$\int_{(-\pi,\pi]} A(e^{-i\lambda})\tilde{f}_x(\lambda)A^*(e^{-i\lambda})\Phi_x\{d\lambda\} = 0.$$

Since $A(e^{-i\lambda})\tilde{f}_x(\lambda)A^*(e^{-i\lambda})$ is semi positive definite this implies

$$A(e^{-i\lambda})\tilde{f}_x(\lambda)A^*(e^{-i\lambda}) = 0 \quad , \qquad \text{a.e.-}\Phi_x.$$

Thus for almost all $\lambda \in E_o^c$ (a.e.-Φ_x) we must have $\tilde{f}_x(\lambda) = 0$, and so

$$F_x\{E_o^c\} = \int_{E_o^c} F_x\{d\lambda\} = \int_{E_o^c} \tilde{f}_x(\lambda)\Phi_x\{d\lambda\} = 0 \quad .$$

Hence

$$F_x\{E_o\} = F_x\{(-\pi,\pi]\} = E\{\underline{x}_t\underline{x}_t^*\} \neq 0 \quad ,$$

which implies that E_o is nonempty.

($\Leftarrow$) If $\lambda_o \in E_o$ and $\underline{z} \in \mathbb{C}^m$ is some arbitrary random vector with $0 < E\{\|\underline{z}\|^2\} < \infty$, then a weakly stationary solution is given by

$$\underline{x}_t := \underline{z}\, e^{it\lambda_o} \quad , \qquad t = 0, \pm 1, \ldots \quad . \qquad \square$$

REMARK. If $A(z)$ is analytic on $|z| < \rho$ for some $\rho > 1$, the spectral measure is, in contrast to the univariate case, not necessarily concentrated on a finite set since $\det A(e^{-i\lambda})$ may vanish identically without $A(z)$ being identical to the $m \times m$ zero matrix. As a concequence, a weakly stationary process satisfying a homogeneous difference equation can have an absolutely continuous spectral measure.

Finally we consider the m-variate ARMA(p,q) model (3.1.1).

Putting $\underline{Y}_t = \sum_{j=0}^{q} B_j \underline{\varepsilon}_{t-j}$, $t = 0, \pm 1, \ldots$, we can write

$$\underline{Y}_t = \int_{(-\pi,\pi]} e^{it\lambda} B(e^{-i\lambda})\, \underline{z}_\varepsilon\{d\lambda\}, \quad t = 0, \pm 1, \ldots .$$

Hence $\underline{z}_y\{d\lambda\} = B(e^{-i\lambda})\underline{z}_\varepsilon\{d\lambda\}$, $\lambda \in (-\pi,\pi]$. If $\det A(z) \neq 0, |z|=1$ it follows from the multivariate generalization of theorem A.1.4 and theorem A.2.1 that the unique weakly stationary solution of (3.1.1) is

$$\underline{x}_t = \int_{(-\pi,\pi]} A^{-1}(e^{-i\lambda})\underline{z}_y\{dy\} = \int_{(-\pi,\pi]} A^{-1}(e^{-i\lambda})B(e^{-i\lambda})\underline{z}_\varepsilon\{d\lambda\},$$

$$t = 0, \pm 1, \ldots ,$$

and has a spectral density matrix given by

$$f_x(\lambda) = A^{-1}(e^{-i\lambda})B(e^{-i\lambda})\Sigma_\varepsilon\, B^*(e^{-i\lambda})A^{-1*}(e^{-i\lambda}), \quad \lambda\in(-\pi,\pi] .$$

REFERENCES

1. ANDERSON, T.W., (1958), *The statistical analysis of time series*, Wiley, New York.

2. ASH, R.B., (1972), *Real analysis and probability*, Ac. press New York, San Francisco London.

3. BOSE, B.C., (1950-1951), *Least Squares Aspects of the analysis of variance*, Mimeographed series, No. 9 North Carolina University.

4. DEISTLER, M., (1975), *z-Transform and identification of linear econometric models with autocorrelated errors*, Metrika 12, 13-25.

5. DEISTLER, M., (1976), *Linear models with autocorrelated errors: structural identifiability in the absence of minimality assumptions*, Res. Memo. Dept. econometrics, University of Bonn.

6. DEISTLER, M., (1978), *The structural identifiability of linear models with autocorrelated errors in the case of cross-equation restrictions*, Journal of econometrics 8, 23-31.

7. DEISTLER, M. & H.G. SEIFERT, (1978), *Identifiability and consistent estimability in econometric models*, Econometrica 6, 969-980.

8. DEISTLER, M., W. DUNSMUIR & E.J. HANNAN, (1978), *Vector linear time series models: corrections and extensions*, Adv. Appl. Prob. 10, 360-372.

9. DOOB, J.L., (1953), *Stochastic processes*, Wiley, New York.

10. MACDUFFEE, C.C., (1956), *The theory of matrices*, Chelsea publ. comp. New York.

11. FELLER, W., (1971), *An introduction to probability theory and its applications*, Vol. II, Wiley, New York.

12. FERGUSON, T.S., (1967), *Mathematical statistics, a decision theoretic approach*. Ac. press, New York, London.

13. FISHER, F., (1966), *The identification problem in econometrics*, Mc.Graw Hill, New York.

14. GENUGTEN, B.B. v.d., (1977), *Identifiability in statistical inference*, Statistica Neerlandica 2, 69-89.

15. GANTMACHER, F.R., (1959), *The theory of matrices*, Vol. I Chelsea publ. comp. New York.

16. GRENANDER, U. & M. ROSENBLATT, (1957), *Statistical analysis of stationary time series*, Wiley, New York.

17. HANNAN, E.J., (1971), *The identification problem for multiple equation systems with moving average errors*, Econometrica, 39. 5.

18. HANNAN, E.J., (1969), *The identification of vector mixed autoregressive moving average systems*, Biometrika, 57.

19. HANNAN, E.J., (1975), *The estimation of ARMA models*, Ann. Math. Stat. 4. 975-981.

20. HANNAN, E.J., (1970), *Multiple time series*, Wiley, New York.

21. HANNAN, E.J., (1980), *Personal communication with TIGELAAR*.

22. KADANE, J.B., (1975), *The role of identification in Bayesian theory*, Studies in Bayesian econometrics and statistics, 175-191.

23. KOOPMANS, T.C. & O. RIERSOL, (1950), *The dientification of structural characteristics*, Ann. Math. Stat. 21, 165-181.

24. LINDLEY, D.V., (1971), *Bayesian statistics: a review*, Philadelphia, SIAM.

25. MARAVALL, A., (1979), *Identification in dynamic shock error models*, Springer, Berlin.

26. MORALES, J.A., (1971), *Bayesian full information structural analysis*, Springer, Berlin.

27. RIERSOL, O., (1950), *Identifiability of a linear relation between variables which are subject to error*, Econometrica 18, 375-389.

28. RIESZ, F., (1923), *Über die Randwerte einer analytischen Funktion*, Math. Zeit. vol. 18, 87-95.

29. ROTHENBERG, T., (-) *The value of structural information: a Bayesian approach*, Unpublished.

30. ROZANOV, YU.A., (1967), *Stationary random processes*. Holden-Day, San Francisco.

31. SCHÖNFELD, P., (1975), *A survey of recent concepts of identification*, CORE discussion paper no. 7515, Leuven.

32. THEIL, H., (1971), *Principles of Econometrics*, Wiley, New York.

33. TIGELAAR, H.H., (1975), *Spectraalanalyse en stochastische differentie vergelijkingen*, Reeks ter discussie 75.001.

34. TIGELAAR, H.H., (1980), *Personal communication with HANNAN*.

35. TOLSTOV, G.P., (1962), *Fourier series*, Prentice-Hall Inc. New Jersey.

36. WESTCOTT, M., (1970), *Identifiability in Linear processes*, Zeitschr. Wahrsch. Theorie u. verw. Geb. 16, 39-46.

SUBJECT INDEX

autoregressive process 37

closed under nonsingular linear transformations 110

complete econometric model 97

conditional identifiability 11

conditional identification theorem 12

conditional informative 11

decision rule 20

endogenous variables 117

error in variables model 17

exogenous variables 117

Fourier coefficients 128

frequency response function 128

future of a process 29

generalized Bezout theorem 73

identifiability 7

identifiable hull 13

ignorability 18 , 20

informational class 8

informational independence 13

informativeness 7

informative sample size 26

left denominator 71

left factor 68

linear filter 128

local identifiability 35

matrix polynomial 61

minimum informative sample size 27

minimum predictive sample size 29

mixed autoregressive moving average process 37

moving average process 37

m-variate random measure with orthogonal increments 134

non collinearity 118

non exponentially increasing sequence 117

nowhere local identifiability 35

observational equivalence 6

opinion 22

order condition 116 , 119

pay-off function 19

predictability 28

predictive sample size 29

prior measure 22

probabilistic prediction problem 29

proper matrix polynomial 97

random measure with orthogonal increments 124

rank condition for reduced form 115

rank condition for structural form 116 , 120

reduced form 98

right multiple 68

right numerator 71

r-th order informativeness 32
r-th order predictive sample (size) 32
S_p^{AR}, S_q^{MA} 41
$S_{p,q}^{ARMA}$ 49
$\tilde{S}_{p,q}^{ARMA}$ 57
$S_p^{AR}(m)$ 49
$S_q^{MA}(m)$ 76
$S_{p,q}^{ARMA}(m)$ 83
S^{DSE}, $\tilde{S}^{DSE}$ 103
spectral density matrix 135
spectral measure 128
spectral representation 127
statistical prediction problem 28
statistical statement 6
structural form 98

t-informativeness 33

uniform identifiability 24
unimodular matrix 68

weak observational equivalence 32
white noise 37 , 61

YULE-WALKER equations 42 , 81