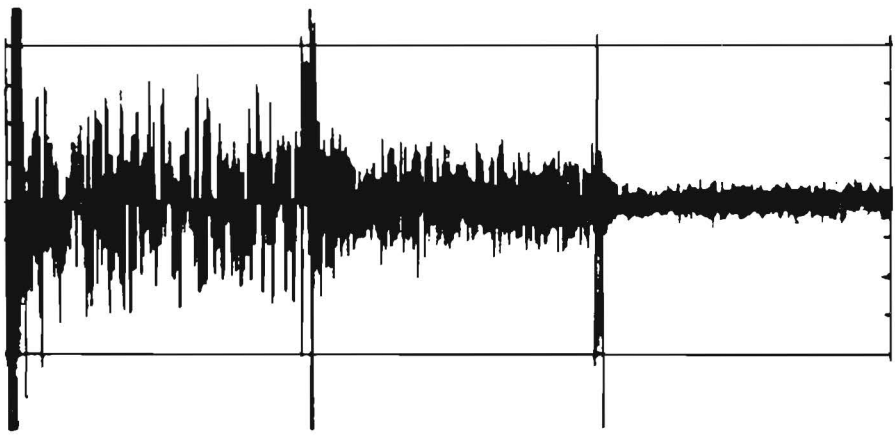
System identification : on the variety and coherence in parameter- and order estimation methods

SYSTEM IDENTIFICATION

ON THE VARIETY AND COHERENCE IN
PARAMETER- AND ORDER ESTIMATION METHODS



AD VAN DEN BOOM

DISSERTATIE DRUKKERIJ
wibro
HELMOND
TELEFOON 04920-23981

SYSTEM IDENTIFICATION

ON THE VARIETY AND COHERENCE IN PARAMETER- AND ORDER ESTIMATION METHODS

SYSTEM IDENTIFICATION

ON THE VARIETY AND COHERENCE IN PARAMETER- AND ORDER ESTIMATION METHODS

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE HOGESCHOOL EINDHOVEN, OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. S. T. M. ACKERMANS, VOOR EEN COMMISSIE AANGEWEZEN DOOR HET COLLEGE VAN DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP DINSDAG 28 SEPTEMBER 1982 TE 16.00 UUR

DOOR

ADRIANUS JOHANNES WILHELMUS VAN DEN BOOM

GEBOREN TE GELDROP

Dit proefschrift is goedgekeurd door de promotoren

prof. dr. ir. P. Eykhoff

en

prof. ir. H.B. Verbruggen

aan Johannie

CIP-gegevens

Boom, Adrianus Johannes Wilhelmus van den

System identification: on the variety and coherence in parameter- and order estimation methods / Adrianus Johannes Wilhelmus van den Boom. - [S.l. : s.n.]. - Fig.
Proefschrift Eindhoven. - Met lit. opg., reg.
ISBN 90-9000352-5
SISO 656 UDC 519.71.001.3
Trefw.: regeltechniek / meettechniek.

CONTENTS

SUMMARY

This study concerns the coherence and variety in parameter- and order estimation methods, which are basic in system identification. For the estimation of the parameters of dynamical systems, several methods have been proposed in the last decade on a rather ad-hoc basis. These methods are all attempts to ensure the consistence of the estimates, which for convenient parametrizations is usually not achieved with common least squares estimators.

The present study aims to present a coherent picture of this field. Therefore three basic components of existing estimators are recognised. These three basic components are: filtering, model extension and use of an (extra) instrumental variable signal. A general scheme is presented containing these three basic components.

It is shown that the existing estimators like Generalized Least Squares, Extended Matrix Method, Approximate Maximum Likelihood, Implicit Quasi Linearization, Prior Knowledge Fitting, Instrumental Variable Estimator and Suboptimal Instrumental Variable Estimator are special cases of this general scheme. The advantage of such a presentation is twofold: it gives a better understanding of the interrelations of the existing estimators, and computer programs for such estimators can be designed in such a way that one program can represent all estimators considered.

Based on these concepts, several estimators are proposed for situations where both input- and output signals are noise corrupted. These estimators have in common that two of the basic components are combined to obtain consistence.
No additional knowledge of noise covariance is needed or assumptions concerning equal colouring, as in the existing literature, are made. It is indicated that the choice of the instrumental variable quantity, which is one of the two basic components for these estimators, can be improved if extra measurements of the input- or of the output signal or signals, which are related to the input or output, can be made available. In such a way, existing information concerning the process, which is usually at hand in practical situations, can be

exploited easily. The algorithms that have been proposed are simple and fast. Experiments with simulated processes show the usefulness of these estimators.

The present study also includes an extensive discussion of order testing methods from a point of view of a potential user. Order testing methods are of prime importance in system identification as usually the order of the desired model is unknown, e.g. for control purposes. Furthermore, if high order models are wanted, containing detailed information of the process under study, then order tests can be used to decide whether the measured signals available contain sufficient information for producing these models.

In the given discussion, the close relations between different order testing methods are shown.

The above mentioned estimation and order testing methods have been incorporated within an extensive interactive computer package SATER. Special attention has been given to the interactive aspects of this package and its modular design. This package is useful for research and educational purposes.

CHAPTER ONE:

PRELIMINARIES

System identification covers, by definition, all possible methods which provide (aggregated) knowledge of a (partly) unknown system based on observations. It follows directly from this rather broad definition that numerous activities of model building are included as, generally, the knowledge of processes is concentrated in their corresponding models. These models may be of widely varying structure, ranging from exclusively verbal to strictly formal mathematical. Also the class of possible processes to be described by models is, in principle, unlimited and of a strongly varying nature.

For a systematic presentation of system identification as a coherent science, this extreme variety of possible processes, of possible models and of possible methods is still prohibitive at present. Even for the class of mathematical models, such a coherent picture is not yet well-established. In this context, a characterization of the field of estimation by "a bag of tricks" (cf. Eykhoff, Van den Boom and Van Rede, 1981) has been appropriate for the past decade; see also fig. 1.1. The introduction of "template functions", provides a powerful tool for a more systematic classification of the field if mathematical models are considered. The fundamentals of this classification are indicated in fig. 1.2; cf. op. cit.

The aim of the present study is to provide a more concise classification and ordering of the elements in the blocks II and III of fig. 1.2, as well as an extension of the concepts. This ordering is based on three aspects: a) choice of the measurables (Ω in fig. 1.2), denoted as model extension, b) interpretation of the template function (Z in fig. 1.2) as filtering b1) or correlation b2). This will be explained in detail in chapters 3 and 4. This classification provides a better insight into the relations among the different estimators in blocks II and III of fig. 1.2.

In chapter 2 some general notions pertinent to model building in a technical sense are given as an introduction to chapters 3 and 4,

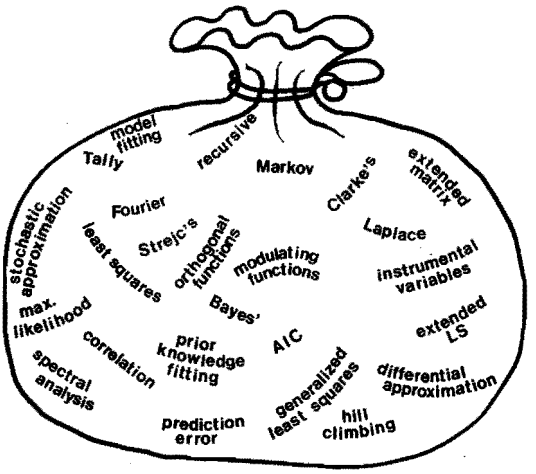


Fig. 1.1 A "bag of tricks"

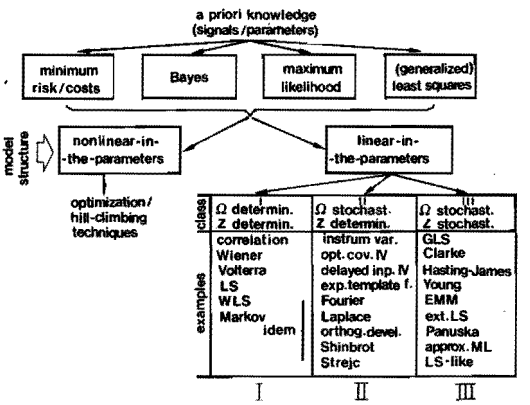


Fig. 1.2 The "bag of tricks" ordered

where estimators are discussed for which only one set of measurables (either input of output) are contaminated by noise.

In chapter 5 the results of chapters 3 and 4 are used to propose several estimators for the situation where all measurables of the process are noise corrupted. In chapter 6 results of the estimators proposed in previous chapters are given, mainly based on simulations, to make an evaluation of the quality of the estimators possible.

Finally, in chapter 7, the problem of order testing is discussed and several - practical - order tests are compared.

CHAPTER TWO:

A RECAPITULATION OF BASIC CONCEPTS: MODELLING, PARAMETRIZATION, ORDER, IDENTIFIABILITY AND IDENTIFICATION PROTOCOL

2.1 Introduction

The aim of this chapter is to discuss briefly the principles of model building in relation to identification. We will start with a few remarks on model building in a general setting, i.e. the meaning of the concept of models used by human beings, but we shall restrict ourselves, quite early on, to model building in an engineering sense. Aspects like parametrization, order and identifiability will be reviewed, in order to provide an adequate basis for the following chapters.

2.2 Some general notions

Modelling is one of man's oldest activities. The image that man forms of his surrounding world, based on observations, is the result of "model building". In fact, all notions about what is often referred to as "REALITY" or "NATURE" or "TRUTH", are models of varying complexity. In communicating with others, e.g. by using words, referring to some consensus of these words, an individualistic interpretation of this consensus cannot be avoided. These interpretations are personal images or models. This personal interpretation can lead to misunderstanding, but on the other hand it introduces variability, which can result in evolution of the consensus itself.

An important aspect of modelling should be stressed here, i.e. its intended use. The construction, the form and the complexity of a model should be mainly determined by those aspects of the "real system" or object which are relevant, or are believed to be relevant, for the intended use of the model.

The appearance and the form of a model and of the studied object are not equal. Models for weather forecasting may consist, for example, of a very complex set of non-linear differential equations, which can be solved only by very large computers, or may consist of a few simple principles in the mind of a farmer.

In general, the validity of a model will be limited. If a model has become too restricted, e.g. due to increased demands, it has to be replaced by a (more) complicated one, explaining more aspects of the object.

Scientific theories are, in fact, also models which are valid until they can be "falsified". Extension or adaptation of the existing theories usually follows such a falsification, or sometimes a new theory is proposed with widely different aspects, which remains valid until, in turn, it too can be falsified; cf. Popper (1959). The falsified model often can serve as a good approximation of the newly developed model, under certain restrictions.

2.3 Technical modelling and parametrization

As we are interested in engineering methods of model building and identification which are suited for algorithmization, we shall restrict ourselves here to models which can be treated mathematically. These models are not only useful for the description of industrial objects but also for objects that are not necessarily technical, such as a variety of bio-medical, social and economic objects.

The first principal decision that has to be made with respect to modelling, concerns the way of parametrization of the model, i.e. the form of the mathematical description of the input-output relationship. Niederlinski and Hajdasinski (1979) formulate three important objectives for a convenient parametrization:

1) universality, i.e. it should be applicable to all objects in the class of interest.
2) limited number of (un)known parameters. This is related to the principle of parsimony, formulated by William of Ockham

(1285 - ± 1349) and known as Ockham's razor: "Non sunt multiplicanda entia praeter necessitatem"; it is applicable to model building as well.

3) identifiability of (unknown) parameters of interest.

Parzen (1974) makes a distinction between structural and synthetic models. The parameters of a structural model have a natural structural interpretation: they will rely on physical laws. These parameters provide explanation of the object which generates the data. Synthetic models are not based on physical laws. Their parameters need not be physically meaningful. Their interest is in the use for simulation, for prediction of future behaviour, for interpretation of past behaviour, and for (optimizing- and adaptive) control. In the literature, the structural model is sometimes also called generic model or explanatory model, whereas the synthetic model may be called non-generic model or input-output model; cf. also Richalet (1981), and Hajdasinski, Eykhoff, Damen and Van den Boom (1982).

Besides this distinction among models, a characterization may be based on the following list of adjectives; cf. also Hajdasinski et al. (1982) for further explanation:

time-continuous	time-discrete
time-invariant	time-variant
linear dynamics	non-linear dynamics
single-input single-output (SISO)	multi-input multi-output (MIMO)
lumped parameters	distributed parameters
parametric	non-parametric
deterministic	non-deterministic
single layer	hierarchical
causal	non-causal
one dimensional	more dimensional
non-fuzzy	fuzzy
non-verbal	verbal

Note that a model can be characterized by several of these descriptors; even a combination of the two opposing descriptors on the same line is possible (e.g. a model may be partly time-continuous, partly time-discrete).

A crucial choice which has to be made concerns the linearity of the models, in the sense of whether the output quantity is a linear or a non-linear dynamic function(al) of the input signal. For linear models the theory and practice of model building and estimation is far more developed than for non-linear models. This is partly due to the fact that a coherent and complete description of non-linear systems does not exist. A rather general description like Volterra series expansion has, for many practical cases, the drawback of having an excessive number of parameters. In many cases, depending on the intended use, it is sufficient to have a model only in a certain working point. In this case, linearizing can yield a simpler and more useful model. For an extensive review of non-linear models, see the survey paper by Haber and Keviczky (1976).

Other types of simplification may occur when systems with distributed parameters are to be modelled by lumped models, when time-continuous systems are modelled by time-discrete models and time-variant systems by time-invariant models. For these types of model simplifications and the general aspect of the construction of lower order models, the survey paper by Gwinner (1976) gives a good introduction.

In the following chapters we shall concentrate our discussion on models which are linear-in-the-parameters. The parametrization of these models may cover many linear, as well as some non-linear systems for single-input single-output models, as indicated in table 2.1

From the point of view of applicability of the available parameter-estimation methods, the property of linearity-in-the-parameters of models is important. The model error, i.e. some difference between object and model behaviour can then be expressed as linear-in-the-parameters, so that the gradient of a quadratic performance criterion with respect to the parameters can be evaluated quite easily. For this reason much attention will be given to these types of models in

LINEAR DYNAMICS	- linear difference or differential equations - ARMA, (ARMAX, transfer function) - impulse response, Markov parameters, Hankel matrix - Laguerre polynomials - state space models (canonical forms and others)		
NON LINEAR DYNAMICS	- non linear difference or differential equations - Volterra kernels - Chebychev polynomials - dispersion models - GMDH models - Hammerstein models - Wiener models - catastrophe models	Volterra Smets Rajbman c.s Ivakhnenko Hammerstein Wiener	(1959) (1960) (1980) (1968) (1930) (1958)

Table 2.1 Parametrization of models

the following chapters. In this context the concept of generalized models is valuable, as it gives us the possibility of obtaining linearity-in-the-parameters in a flexible way. In fig. 2.1, three distinct types of models are shown, a) the output error model, b) the input error model and c) the generalized model.

It will be clear that these types of models are often synthetic models. Within the field of control engineering they are widespread, because they can be used satisfactorily for a variety of applications. Their way of parametrization can be chosen within broad limits, depending on the particular situation at hand. A few possibilities are; cf. Eykhoff (1974)

output error model:

$$\varepsilon_k = y_k - F_o(u_k;\theta'_k) \tag{2.1}$$

e.g. - moving average model
impulse response model
Hankel model

$$F_o = \sum_j \theta'_j u_{k-j} \tag{2.2}$$

- transfer function model

$$F_o = H(z^{-1})u_k = \frac{[b_o+B(z^{-1})]}{[1+A(z^{-1})]} u_k \qquad (2.3)$$

input error model:

$$\epsilon_k = u_k - F_i(y_k;\theta_k) \qquad (2.4)$$

e.g. - autoregressive model

$$F_i = \sum_j \theta_j y_{k-j} \qquad (2.5)$$

- inverse transfer function model

$$F_i = \frac{[1+A(z^{-1})]}{[b_o+B(z^{-1})]} y_k \qquad (2.6)$$

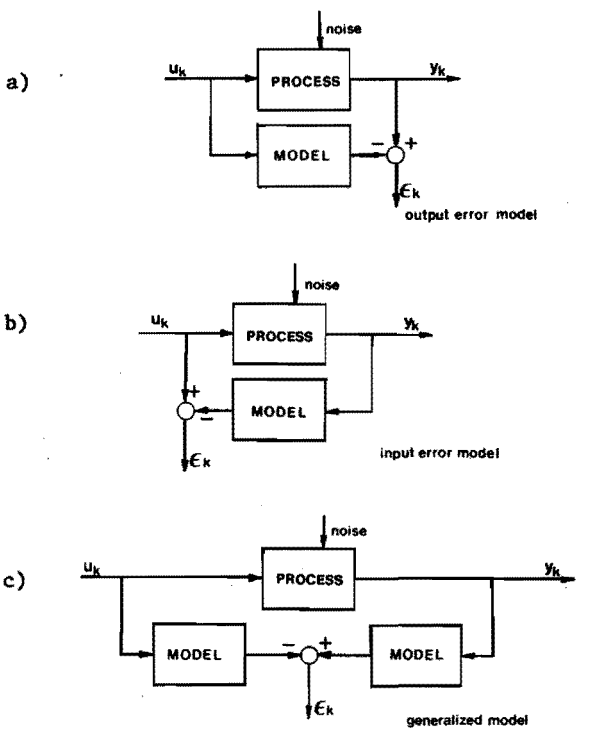


Fig. 2.1 Distinct types of model errors

generalized model:

$$\varepsilon_k = -F_o(u_k;\theta'_k) + F_1(y_k;\theta_k) \tag{2.7}$$

e.g. - autoregressive - moving average model (ARMA)

$$F_o = \sum_j \theta'_j u_{k-j} \quad ; \quad F_1 = \sum_j \theta_j y_{k-j} \tag{2.8}$$

- state space models

A few remarks can be made:

- a purely autoregressive model and a purely moving average (or Hankel, or impulse response) model usually needs a (very) large number of parameters for an adequate representation of the dynamics of the system, in the sense that the resulting error is small or negligible.
- transfer function models or inverse transfer function models may have a smaller number of parameters but these models are not linear-in-the-parameters.

First we consider the synthetic input-output models. Several variants have been proposed in the identification literature:

Moving Average (MA) models:

deterministic

$$y_k = [b'_o + B'(z^{-1})]u_k \tag{2.9}$$

stochastic; cf. Levin (1960)

$$y_k = [b'_o + B'(z^{-1})]u_k + e_k \tag{2.10}$$

where e_k is the modelled noise. In eq. (2.9) and eq. (2.10) $[b'_o + B'(z^{-1})]$ is a polynomial operator, which may have arbitrary length. These models are also called impulse response models or Markov models. They usually need a very large parameterset, which is often a drawback, so that the following models have been proposed.

Autoregressive Moving Average (ARMA) models:

deterministic

$$[1 + A(z^{-1})]y_k = [b_o + B(z^{-1})]u_k \tag{2.11}$$

stochastic

$$[1 + A(z^{-1})]y_k = [b_o + B(z^{-1})]u_k + e_k \tag{2.12}$$

Here $[1+A(z^{-1})]$ and $[b_o+B(z^{-1})]$ are polynomial operators, with memory length of resp. q and p. The correct choice of q and p refers to the problem of order testing, see chapter 7. If we extend this stochastic model by taking into account the dynamics of the noise, we have several possibilities: we may use MA, AR (autoregressive) or ARMA modelling for the noise.

$$\text{MA} \qquad e_k = [1+C(z^{-1})]\,\xi_k$$

$$\text{AR} \qquad e_k = \frac{1}{[1+D(z^{-1})]}\,\xi_k \;\rightarrow\; e_k = -[D(z^{-1})]e_k + \xi_k \qquad (2.13)$$

$$\text{ARMA} \qquad e_k = \frac{[1+C(z^{-1})]}{[1+D(z^{-1})]}\,\xi_k \;\rightarrow\; e_k = -[D(z^{-1})]e_k + [1+C(z^{-1})]\xi_k$$

where ξ_k is a (conceptual) white noise input sequence.
This leads to the following models of process- and noise dynamics which have been proposed in the literature:

ARMAX model; cf. Åström and Bohlin (1965)

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k + [1+C(z^{-1})]\xi_k \qquad (2.14)$$

Clarke's model; cf. Clarke (1967)

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k + \frac{1}{[1+D(z^{-1})]}\,\xi_k \qquad (2.15)$$

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k - [D(z^{-1})]e_k + \xi_k \qquad (2.16)$$

Model of Talmon and Van den Boom (1973)

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k + \frac{[1+C(z^{-1})]}{[1+D(z^{-1})]}\,\xi_k \qquad (2.17)$$

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k - [D(z^{-1})]e_k + [1+C(z^{-1})]\xi_k \qquad (2.18)$$

This model is a generalization of the two previous models. It incorporates the advantages of both models, i.e. by properly choosing the degrees of the AR and MA parts, one can model the MA part of the noise by the MA part of the noise model and the AR part of the noise by the AR part of the noise model. Purely MA noises need not be modelled by AR models as with Clarke's model. The Talmon and Van den Boom model gives therefore more flexibility to arrive at a minimal parameterset.

An important observation is that the above models are linear-in-the-parameters, provided that the signals e_k and ξ_k are available. It is obvious that this will not be the case, as only input and output samples u_k and y_k are available. The models which will be used in practice will then need an estimate of these signals e_k and ξ_k. These estimates can be obtained by making use of previous estimates of process- and noise parameters. Therefore, the above models are linear-in-the-process-parameters A and B but are non-linear-in-the-noise-parameters C and D. This will cause the appropriate estimation methods to need iterations or recursions to handle these non-linearities. This will be explained further in chapters 3 and 4. In fig. 2.2 a diagram is given for the most general model, i.e. that of Talmon and Van den Boom, where it is used for the generation of the generalized model errors $\hat{e}$ and $\hat{\xi}$.

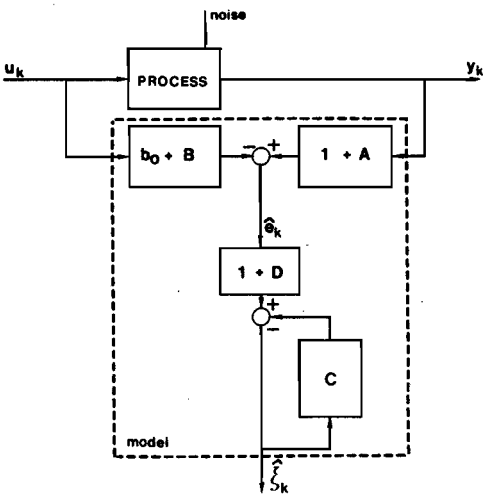


Fig. 2.2 The model of Talmon and Van den Boom

The above given models (2.12), (2.14), (2.15) and (2.17) are generalized models. Also output error models have been proposed in the past. They do not have the attractive property that the model error is linear in the parameters. We will give some examples:

- Transfer function model

$$y_k = \frac{[b_o+B(z^{-1})]}{[1+A(z^{-1})]} u_k + e_k \qquad (2.19)$$

- Box-Jenkins model; cf. Box and Jenkins (1976)

$$y_k = \frac{[b_o+B(z^{-1})]}{[1+A(z^{-1})]} u_k + \frac{[1+C(z^{-1})]}{[1+D(z^{-1})]} \xi_k \qquad (2.20)$$

Output error models have been used by Dugard and Landau (1980) using the Model Reference Adaptive System (MRAS) techniques.

Ljung (1979) proposed a model which contains the Talmon and Van den Boom model and the Box-Jenkins model as special cases

$$[1+A(z^{-1})]y_k = \frac{[b_o+B(z^{-1})]}{[1+F(z^{-1})]} u_k + \frac{[1+C(z^{-1})]}{[1+D(z^{-1})]} \xi_k \qquad (2.21)$$

Next we will consider <u>state space modelling</u>. The general expression is

$$\left.\begin{aligned} \underline{x}_{k+1} &= A\underline{x}_k + B\underline{u}_k \\ \underline{y}_k &= C\underline{x}_k \end{aligned}\right] \qquad (2.22)$$

where $\underline{u}_k$ is the input vector, $\underline{x}_k$ is the state vector and $\underline{y}_k$ is the output vector, and the triplet (A,B,C) is called the realization of the dynamical multivariable system. For the state vector, we may look for a minimal set, i.e. the realization with the lowest possible order. An infinite number of state vectors can be found; also an infinite number of triplets (A,B,C). The realization (A,B,C) is not unique, as the T-equivalent realization (TAT^{-1}, TB, CT^{-1}), T being a transformation matrix, gives the same transfer function matrix H(z) or impulse response matrices M_k. The problem of selection of a suitable state space realization will not be discussed here. In Goodwin and Payne (1977) a review is given of the construction of several canonical state space models; see also Denham (1974) and Hajdasinski, Eykhoff, Damen and Van den Boom (1982).

The impulse response matrix or Markov matrix for the k^{th} time instant can be constructed by:

$$M_k = C\, A^{k-1} B \qquad (2.23)$$

resulting in the following model

$$y_k = \sum_j M_j u_{k-j} \qquad (2.24)$$

The matrices M_j can be brought into a Hankel matrix, which is a key

for obtaining a state space realization from the Markov parameters; cf. Ho and Kalman (1966) for the deterministic case.

The transfer function matrix is found by:

$$H(z) = C(zI-A)^{-1}B \qquad (2.25)$$

The above relations apply, in principle, for SISO as well as for MIMO systems. There is a rapidly increasing literature concerning aspects of the choice of parametrization, especially for MIMO systems. In general, the Markov parameters using the Hankel matrix are widely accepted. In figure 2.3 the relation between different parametrizations is given; cf. Hajdasinski and Damen (1979).

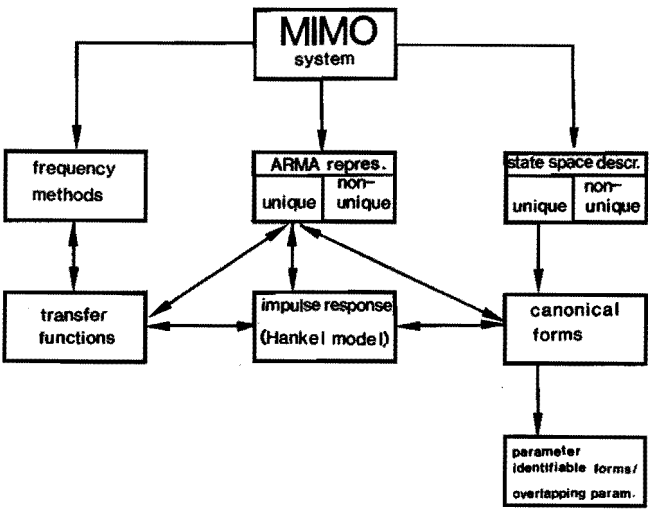


Fig. 2.3 Relations between different parametrizations

From the relations given above it will be evident that the calculation of the Markov parameters and the transfer function from a given realization in state space is straightforward. The calculation of a realization in state space from the transfer function or Markov parameters, however, is rather involved and is the subject of the realization theory; cf. Silverman (1971), Ho and Kalman (1966).

If uncertainties are present in the measured signals, then the transfer function or the Markov parameters resulting from an estimation procedure will be available only as approximations. Then the algorithms for construction of a realization in the state space have to be modified; cf. Hajdasinski and Damen (1979), Van Zee (1981) and Damen and Hajdasinski (1982).

For the parametrization of SISO models, an ARMA representation is appropriate, as its interpretation is very close to that of a transfer function, consisting of a quotient of two finite polynomials, which are relatively prime. From a historical point of view, transfer functions have been used extensively in control engineering for stability considerations and design. Their parameters are closely related to the physically meaningful parameters of generic models, which can be advantageous for interpretation of results.
State space models are of a more general nature. Their parameters may only be indirectly related to the physical parameters of the system. They provide useful insight into properties of controllability and observability of the overall system. Their parametrization may be very compact, depending on the realization chosen, which is primarily important for MIMO systems.

In the following chapters, we are primarily interested in estimation methods which yield consistent estimates for a model having a parametrization with a limited number of parameters. We have already seen that ARMA models usually have such a limited number of parameters. It was also mentioned that we will need to extend these models to stochastic models, i.e. we need an adequate description of the noise. For this purpose we will make use of ARMAX models, which have a moving average, or an autoregressive-, or a mixed autoregressive moving average parametrization of the noise colouring. This type of modelling of the noise characteristics is motivated by the spectral factorization theorem. This theorem gives a unique factorization for a noise e_k, being wide sense stationary and rational. The spectrum of this noise can be interpreted by considering the noise as an output of a time invariant finite dimensional filter H(z) driven by a white noise input

$$W = H(z)\ \Sigma_1\ H^T(z^{-1}) \qquad (2.26)$$

Σ_1 being the covariance of the white noise input. It should be noted that many different filters H(z) can produce the same spectrum W. Having this spectrum W, the spectral factorization theorem states that a unique spectral factorization of W can be found satisfying:

1) $W = G(z)\ \Sigma_2\ G^T(z^{-1})$

2) G(z) has all its poles inside the unit circle $|z| = 1$

3) $G^{-1}(z)$ has all its poles inside the unit circle $|z| = 1$

4) $\lim_{z \to \infty} G(z) = I$

Proof cf. Youla (1961), Åström (1970).

This theorem is very useful, as it provides a motivation for modelling stationary and rational noise sequences by a stable, minimum-phase ARMA description. The inverse of the model is then also stable and minimum-phase and causally invertible; cf. Gevers and Kailath (1973). We will frequently need this property in the following chapters.

2.4 The notion of order

For noise-free single-input single-output systems (SISO) with linear and lumped dynamics, the notion of system order can be defined quite easily. For a state space description in canonical form, the order is defined as the number of independent states. For transfer function types of parametrization, the order is the number of poles of the system, provided that no pole-zero cancellation occurs.

The notion of the desired order of a model is a more questionable one however, due to the fact that the model is something that is constructed as an image of an unknown process. It need not cover all aspects of the process itself, so that the model may very well be of a lower complexity. Therefore a class of models of interest has to be specified and within this class the most suitable member has to be found, according to a certain predefined sense. Model validation, which will be dealt with later in this chapter, then has to be used in deciding whether this class can be accepted, or whether another, richer class is to be chosen.

Let us assume that we have two model structures M_1 and M_2, with p_1 and p_2 independent components in the parameter vector $\underline{\theta}$ respectively; where $p_1 < p_2$ and $M_1 < M_2$, i.e. M_1 is a subset of M_2. If in the class of models M_j (j=1,2) the loss function $V_j(\hat{\underline{\theta}}_j)$ is minimal, then the model M_j is considered to be the best. Here $\hat{\underline{\theta}}_j$ is an estimate in M_j. If $V_1(\hat{\underline{\theta}}_1)$ and $V_2(\hat{\underline{\theta}}_2)$ are equal then the extra degrees of freedom in M_2 do not contribute to the model in the sense of Ockham's principle of parsimony. The decrease of the test quantity for increasing model order may be obscured by noise, especially if the extra degree of freedom gives only a slightly smaller value for the loss function in the noiseless case. This can lead to a selection of a lower model order. For selection of a proper model order, many order test methods have been developed; cf. Van den Boom and Van den Enden (1973) and chapter 7 of this thesis.

For noise-free MIMO systems the notion of order can be given in an analogous way. For a minimal realization in state space, the order can be defined as the number of independent states. An alternative for definition of the order is to use the realizability index r of the Markov parameters; cf. Hajdasinski and Damen (1979). This is defined as:

$$M_{r+j} = \sum_{i=1}^{r} a(i)\, M_{r+j-i} \qquad \forall_{j>0} \tag{2.27}$$

This means that r Markov parameters M_i, $1 \leq i \leq r$ are sufficient to construct a minimal realization; cf. Ho and Kalman (1966).

With respect to a suitable model order for MIMO systems, the same remarks apply as for SISO systems; cf. also Hajdasinski, Eykhoff, Damen and Van den Boom (1982).

2.5 The concept of identifiability

Another fundamental problem is the identifiability concept. So far in literature this seems to be more of theoretical than of engineering interest. Nevertheless, one should not circumvent this aspect as it determines whether an application of parameter estimation methods might be successful, given experimental conditions such as structure specification, available data, etc.

So far, several authors have studied the identifiability problem and consequently several (closely related) definitions have been introduced. Most definitions are based on consistency of the estimators: the "true" process parameter $\underline{\theta_t}$ is said to be identifiable if the sequence of estimates $\underline{\hat{\theta}_N}$ converges to $\underline{\theta_t}$ in some probabilistic sense, where N is the number of observations.

Åström and Bohlin (1965) use in this respect, consistence with probability, Staley and Yue (1970) convergence in a mean squares sense, Tse and Anton (1972) consistence in probability. Tse (1978) introduces a measure of identifiability based on the following. For a certain identification method, the corresponding identification error is ε_N, where N is the number of observations. The quantity ε_N is then a probabilistic function of $\underline{\hat{\theta}_N} - \underline{\theta_t}$. By bounding ε_N above by $\bar{\varepsilon}_N$, and below by $\underline{\varepsilon}_N$, identifiability conditions can be derived by studying the asymptotic behaviour of $\bar{\varepsilon}_N$ and $\underline{\varepsilon}_N$ for $N \to \infty$. A resolvability function is introduced which describes these bounds completely. In this way a quantitative measure of identifiability is established, which measures the degree of resolvability between parameters. The asymptotic behaviour of this function gives necessary and sufficient conditions for global identifiability.

Another attempt at studying (global) identifiability was made by Bellman and Åström (1970). Here a model is said to be (globally) identifiable if the identification criterion has a unique global minimum. This is interesting as the notion of "true" parameters is not used in this definition. The identifiability property is therefore an attribute of the specified model. This definition leaves much more freedom for the actual system being studied, as it allows also lower order models.

For analysis of identifiability conditions for systems (with feedback control), Ljung, Gustavsson and Söderström (1974) introduce, rather formally, the following quadruplet of notions:

a) the experimental condition X, referring to the manner in which the signals are determined

b) the stochastic system S, given by the general form:

$$y_k = G_S(z^{-1})u_k + H_S(z^{-1})e_k \tag{2.28}$$

where the output vector y_k and the random variable e_k are vectors of dimension n_y and the input vector u(k) has dimension n_u.

c) the model structure $\mu(\theta)$. The model structure is obtained by parametrizing the functions $G(z^{-1})$ and $H(z^{-1})$ in a suitable manner. A model $\mu(\underline{\theta})$ is then given by

$$y_k = G_\mu(z^{-1})u_k + H_\mu(z^{-1})e_k \tag{2.29}$$

d) the identification method J. The parameter estimates at time N for given S, μ, J and X are denoted by $\hat{\underline{\theta}}(N;S,\mu,J,X)$.

With this quadruplet of notions the following identifiability notions are given:

I The system S is said to be system identifiable [SI(μ,J,X)] under μ,J and X, if:

$$\hat{\underline{\theta}}(N;S,\mu,J,X) \to D_T(S,\mu) \quad \text{w.p.1 as } N \to \infty \tag{2.30}$$

where

$$D_T(S,\mu) = \{\hat{\underline{\theta}} \mid G_\mu(z) = G_S(z) \text{ and } H_\mu(z) = H_S(z) \text{ at every } z\} \tag{2.31}$$

i.e. the set of parameter estimates $\hat{\underline{\theta}}$, for which the transfer for process and model are equivalent. This set may contain numerous parameters, including e.g. models with pole-zero cancellations.

II The system S is said to be strongly system identifiable [SSI(J,X)], under J and X, if it is SI(μ,J,X) for all μ such that $D_T(S,\mu)$ is non empty.

III The system S is said to be parameter-identifiable [PI(μ,J,X)] under μ, J and X if it is SI(μ,J,X) and the set $D_T(S,\mu)$ consists of only one element.

Note that a system may be system identifiable, but not parameter identifiable for a certain type of model. An example is when too high order models are used so that pole-zero cancellation in $\mu(\underline{\theta})$ may occur; cf. also chapter 7.

For the class of prediction error estimators, Ljung (1976,1979) derives conditions for consistence and hence for identifiability. We will consider consistence for different identification methods in chapters 3 and 4.

2.6 Identification protocol and model validation

After having touched upon some of the crucial aspects of system identification, it is now possible to discuss the relation of those aspects within the identification protocol. Three main phases can be distinguished in the protocol:

A) preparation
B) estimation
C) validation

In the preparatory phase A, the prerequisites of the estimation phase should be checked:

A1) Check of time invariancy. For proper model building, one should choose between time-invariant and time-variant models. Based on possible a priori knowledge, by careful (visual) inspection of the data or applying time series analysis routines (off-line) for detection (and correction) of trends or drift, one can decide for time (in)variancy of the models to be used.

A2) Check of linearity. The use of linear models greatly simplifies the estimation. Also here, based on a priori knowledge, and analysis of the measured data (e.g. Rajbman's linearity measure, cf. Rajbman and Chadeev (1980)) one should decide whether linear models are admissible. If the process is expected to be non-linear, it is worthwhile considering the possibility of linearizing in a certain working point, if the use of the model is restricted to the vicinity of this working point.

A3) Check or choice of input signals. The frequency content of the input signals should cover the frequency range of the process. A "high" frequency process is hard to identify by a "low frequency" signal. For reliable results "persistently

exciting" input signals are required; cf. Ljung (1971), i.e. the input signals should excite all of the (relevant) modes of the process. In general a "white" noise input signal is very attractive as it covers the whole frequency range, but is physically not realistic. In many cases, it is not permitted or quite impossible (e.g. in some biological processes) to influence the signals of the process in order to improve the "identification quality" of the input signals. For these cases it is good to keep in mind that (as always) the identification results are dictated, c.q. limited, by the type of input signals used; cf. Rooijakkers (1982). In some experimental circumstances it is possible, within some margins, to select the (type of) input signals. In such a case one can choose an "optimal" test signal. For a survey of this topic see Mehra (1974, 1981).

- Another point is that the type of (optimal) input signals used for identification should cope with the type of input signals of the model when applied within the context of the intended use of the identification results. This could otherwise cause problems if a linearized or a (deliberately) lower order model is used.
- Data should be checked for damage such as outlyers, missing data points, parts with excessive noise, pertinent disturbances etc.
- The choice of the sampling rate is another point of interest, as it can also be a trade-off between accuracy/reliability and costs/technical limitations. Aliasing effects should be avoided; cf. also Goodwin and Payne (1977).

A4) Check of correlation of noise. The prerequisite that the disturbing noise and input signals are uncorrelated should not be discarded, as it can cause inconsistent estimates.

A5) Choice of model. The model should be chosen with its intended use in mind. Usually this will be diagnosis or control. This is a very important point and it should be stressed here, as the intended use completely dictates, in principle, the extent and form of the model. A good understanding of the intended use is a great help in choosing a model structure. For control, several types of synthetic models may be adequate (even

lower order models). The choice then depends on mathematical attractiveness of the description and suitability for estimation routines. Also the possibility of incorporating available a priori knowledge in the model can play a role.

In the estimation phase B, a choice should be made of existing estimation routines and order tests. The availability of software for those routines can play a role. The use of a general - interactive - computer package for selection of the appropriate routines can be very helpful; cf. Lemmens and Van den Boom (1977), and chapter 6.

The model validation phase C is perhaps the most difficult one. It determines whether a model should be accepted or not. A model which has been chosen in the preparatory phase A can be rejected at this stage. If the model is rejected, one should proceed to the first phase and start the whole procedure again. In investigating model (in)adequacy several aspects are relevant:

C1) Cross-validation. The confrontation of the obtained results from one set of data with the results from another, independent set of data is worthwhile. Also the use of different parametrizations and the check of their consistence can give information.

C2) Check of residuals. The residuals should usually be white and not contain signals such as peaks, sines etc.

C3) Consistence with a priori knowledge. The confrontation of the properties of the model with possibly available knowledge of the process can give insight. Also other types of input signals can be applied to the model and the resulting output signals can be compared with known behaviour of the process in similar circumstances. Usually, if possible, it is wise to investigate the sensitivity of the identification results to a change of the type of input signal.

2.7 Conclusions

In this chapter, we have discussed the most important aspects of modelling in relation to identification. We have given several types of parametrization of synthetic models, which will be the basis for the discussion in the following chapters. We have discussed aspects of choosing a suitable parametrization and aspects of model validation. It will become clear that for almost all aspects that we have reviewed, we have not given strict, hard and fast rules on how to proceed, but rather we have presented how these aspects are interrelated and what possibilities exist. The main conclusion can therefore be that modelling and validation cannot be mechanized completely, but that good engineering intuition and experience are needed for handling practical problems. Nevertheless, the majority of papers on modelling and identification deal with the pure estimation phase (phase B in paragraph 2.4). This aspect in the whole identification protocol is often the least cumbersome, as nowadays, good estimation packages exist. A practical experience is therefore that a relatively small amount, say approximately 15 per cent of the time devoted to, or the costs spent on modelling, estimation and validation is for the estimation itself, which still means, of course, that the utmost care has to be given to the estimation phase.

CHAPTER THREE:

EXPLICIT LEAST SQUARES ESTIMATORS

3.1 Introduction

The basic principle of the least squares method was introduced in 1795 by K.F. Gauss for the estimation of the parameters of planetary orbits. In the last few decades the method has become popular for the estimation of parameters of earthly dynamical systems based on observations of input- and output signals. The broad application was a.o. stimulated by the availability of digital computers for the - sometimes - excessive amount of number crunching.
At this moment several methods based on the least squares principle are available. Some methods, like the instrumental variable method, were originally proposed outside the framework of the least squares principle, but finally it turned out that they also belong to this class.

This chapter is set out as follows. In paragraph 3.2 the weighted least squares estimator will analysed and the appropriate signal- and process definitions will be given. In paragraph 3.3 the weighting matrix will be considered in more detail and, as a consequence of this, we will distinguish between the "correlative" and the "filtering" type of weighting matrix. This distinction is important as it will yield two elements of the concept of three basic operations by which the estimation methods will be classified later. The combination of these two basic operations results in the set of (explicit) instrumental variable estimators, including the Tally estimator. Also the third basic operation of estimation schemes will be dealt with: model extension. In paragraph 3.4 the combined application of the three basic operations are reviewed and a general scheme for explicit estimators will be given. Known explicit schemes like the implicit quasi linearization (IQL) and the suboptimal instrumental variable estimator (SIV) fit into this scheme. In chapter 4, this concept of three basic operations will be used for classification of the recursive estimators. Schemes like the approx-

imate maximum likelihood (AML), the extended matrix method (EMM) and combined schemes like IV-AML will also fit into this general scheme.

In paragraph 3.5 the maximum likelihood estimator will be dealt with briefly, due to its relation with the AML scheme. In paragraph 3.6 the accuracy of the estimators is discussed, based on the Cramér-Rao results.

3.2 The ordinary least squares method (LS), the weighted least squares method (WLS)

Consider the following conceptual description of the process P. We assume that P can be described by the following input-output relation, cf. figure 3.1

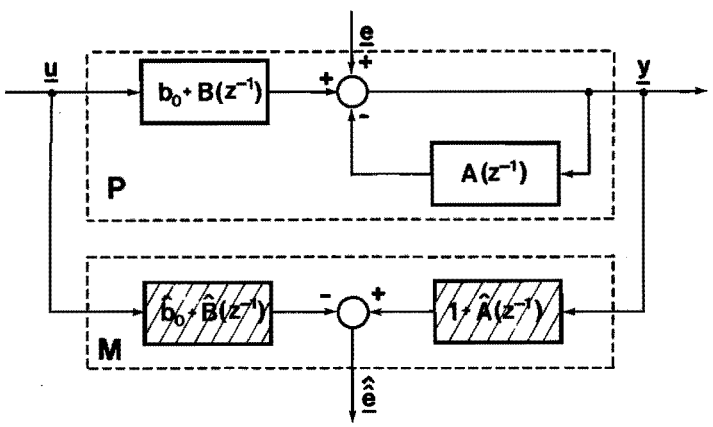


Fig. 3.1 Process, model and disturbing noise

$$\underline{y} = \Omega(u,y)\underline{\theta}_c + \underline{e} \tag{3.1}$$

with

$$\left.\begin{aligned} \underline{y}^T &= (y_{q+1}, \ldots, y_N) \\ \underline{e}^T &= (e_{q+1}, \ldots, e_N) \end{aligned}\right] \tag{3.2}$$

$$\underline{\theta}_t^T = (b_o,\ldots,b_p, -a_1,\ldots,-a_q)_t \tag{3.3}$$

$$\Omega(u,y) = \left[\begin{array}{ccc|ccc} u_{q+1} & \cdots & u_{q+1-p} & y_q & \cdots & y_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ u_N & \cdots & u_{N-p} & y_{N-1} & \cdots & y_{N-q} \end{array}\right] \tag{3.4}$$

Here y_k is the output sample at the k-th instant of time. The vector $\underline{e}$ is denoted as the equation error, and $\underline{\theta}_t$ denote the true parameters of the dynamical system. A corresponding output noise n_k, i.e. the noise considered is concentrated at the output of the process, can be defined implicitly as:

$$y_k = x_k + n_k \tag{3.5}$$

where x_k is the undisturbed output. The relation between the equation error e_k and this corresponding output noise n_k can be given; cf. eq. (3.16).

The input-output relation can also be written in polynomial form:

$$[1+A(z^{-1})]_t y_k = [b_o+B(z^{-1})]_t u_k + e_k \tag{3.6}$$

where the polynomials are defined as:

$$\left.\begin{aligned} [1+A]_t &= [1+a_1 z^{-1}+ a_2 z^{-2} + \ldots + a_q z^{-q}]_t \\ [b_o+B]_t &= [b_o+b_1 z^{-1} + b_2 z^{-2} + \ldots + b_p z^{-p}]_t \end{aligned}\right] \tag{3.7}$$

For notational simplificiation, further on we will omit the argument z^{-1} in the expressions for the polynomials.

For asymptotic stability of the signals involved, it is assumed that the roots of $z^q[1+A]_t$ and $z^p[b_o+B]_t$ all lie inside the unit circle of the complex z-plane. We assume here that a part of the relationship between the measured input-output samples may not be representable by the dynamical part of the difference equation (3.6), but that some uncertainty can be admitted. This may be interpreted in different ways:

a) system noise: insufficient modelling of the dynamical part (non-linearities, too low order, etc.);

b) measurement noise: contaminated measurements of the output signal, poorly observed input signals.

The parametrization of the quantity e_k can be done along the lines indicated in chapter 2, where MA, AR and ARMA descriptions were given.

If a delay is present in the system, it is wise to shift the input samples in time so that the description (3.6) is valid. For the determination of this delay, in case it is unknown a priori, see chapter 7.

The generalized model M which is built will be fed by the measured signals u_k and y_k as indicated in figure 3.1. Based on N observations of input and output signals, an estimate $\underline{\hat{\theta}}$ of $\underline{\theta}_t$ using a least squares criterion is found. The input-output description of the model is written as:

$$\underline{y} = \Omega(u,y)\underline{\theta} + \underline{\hat{\hat{e}}} \qquad (3.8)$$

where $\underline{\theta}$ represents the parametrization of the model set. It is assumed that $\underline{\theta}_t$ lies inside this modelset.

A quadratic error criterion is defined, based on the estimated equation error $\underline{\hat{\hat{e}}}$; which will be called residual.

$$V = \frac{1}{N-q}\,\underline{\hat{\hat{e}}}^T\underline{\hat{\hat{e}}} \qquad (3.9)$$

Minimizing this with respect to $\underline{\theta}$, we obtain the least squares estimator in an explicit form:

$$\underline{\hat{\theta}}_{LS} = \left[\Omega^T(u,y)\Omega(u,y)\right]^{-1}\Omega^T(u,y)\underline{y} \qquad (3.10)$$

In figure 3.2 a schematic diagram of this estimator is given, where the signals, the choice of model involved and the quantity $\underline{\hat{\hat{e}}}$ used for the criterion (3.9) are shown.

In many cases it will be desirable that the estimators are unbiased, at least asymptotically. This aspect is given much attention in literature. It seems reasonable that an estimation algorithm should aim at the "right" parameter value, but in the case of some adaptive

control schemes, it is not always necessary to provide the controller with unbiased estimates. In such cases, simple and fast estimators, which may have bias, e.g. LS estimators combined with minimum variance control, can often be used fruitfully. Moreover the concept of "right" parameter values is doubtful. We can only think of "best" parameter values within a certain class of models, given a certain minimization criterion. Comparison of estimated model parameters with process parameters is only feasible in simulated experiments where model-to-model estimations are being performed.

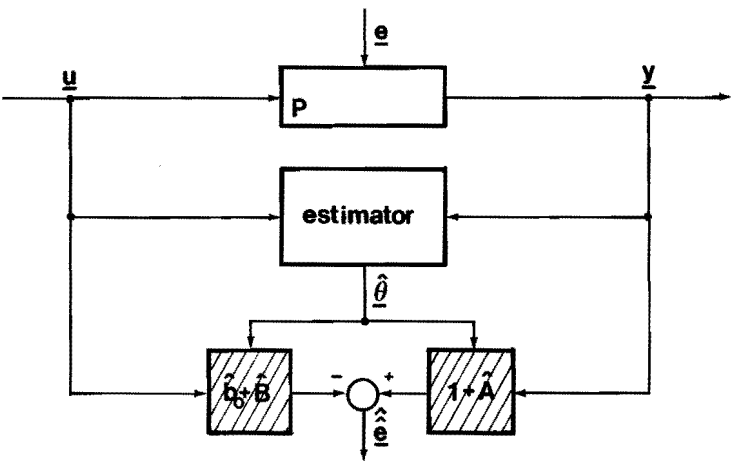


Fig. 3.2 Least squares estimator

The asymptotical bias can be investigated by taking the probability limit of $\hat{\underline{\theta}}$ and using Slutsky's theorems, cf. appendix V:

$$\operatorname*{plim}_{N\to\infty}\left[\hat{\underline{\theta}}\right] = \underline{\theta}_t + \operatorname*{plim}_{N\to\infty}\left[\left[\Omega^T(u,y)\Omega(u,y)\right]^{-1}\Omega^T(u,y)\,\underline{e}\right] =$$

$$= \underline{\theta}_t + \left\{\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)\Omega(u,y)\right]\right\}^{-1}\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)\underline{e}\right] \qquad (3.11)$$

If

$$\begin{aligned} &\text{I} \quad \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)\Omega(u,y)\right] = \Gamma \qquad \Gamma \text{ is non-singular} \\ &\text{II} \quad \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)\underline{e}\right] = \underline{0} \end{aligned} \qquad (3.12)$$

then plim $[\hat{\underline{\theta}}] = \underline{\theta}_t$, so the estimator is asymptotically unbiased (consistent). Condition I assures that the measured signals $\underline{u}$ and $\underline{y}$ contain sufficient degrees of freedom to make the estimation meaningful. This condition is related to the requirement of persistently exciting input measurables. Condition II gives insight into the required colouring of $\underline{e}$ for obtaining consistent estimates:

$$\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)\underline{e}\right] = \operatorname*{plim}_{N\to\infty}\frac{1}{N-q}\begin{bmatrix} \sum_{q+1}^{N} u_i e_i \\ \vdots \\ \sum_{q+1}^{N} u_{i-p} e_i \\ \sum_{q}^{N-1} y_{i-1} e_i \\ \vdots \\ \sum_{q}^{N-1} y_{i-q} e_i \end{bmatrix} \triangleq \begin{bmatrix} \Psi_{ue}(0) \\ \vdots \\ \Psi_{ue}(p) \\ \Psi_{ye}(1) \\ \vdots \\ \Psi_{ye}(q) \end{bmatrix} \tag{3.13}$$

If $\underline{u}$ and $\underline{e}$ are independent, which also implies that the two signals do not both have a non-zero mean value:

$$\Psi_{ue}(i) = 0 \qquad \forall\ i \tag{3.14}$$

The same holds for $\underline{x}$, the undisturbed part of the output signal of the process:

$$\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)\underline{e}\right] = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Psi_{xe}(1) + \Psi_{ne}(1) \\ \vdots \\ \Psi_{xe}(q) + \Psi_{ne}(q) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Psi_{ne}(1) \\ \vdots \\ \Psi_{ne}(q) \end{bmatrix} \tag{3.15}$$

From this relation it can be observed directly that the estimation of the AR process parameters will cause complications, as the right-hand

term of eq. (3.15) is usually not zero, except for very specific colouring of e_k, which will be investigated soon. If a pure MA parametrization of the process dynamics is used, then consistence of the estimates is guaranteed, if (3.14) is fulfilled. This implies that the various techniques for obtaining consistent estimates, which will be dealt with in this chapter and chapter 4, are not needed. This is an interesting advantage for MA models. The drawback is that a greater number of parameters is necessary for MA models, leading often to approximated models, with, in practice, a limited number of parameters. This will also cause inconsistence of the estimated parameters.

A relationship exists between the process noise e_k, which can be seen from figure 3.1, and an equivalent output noise n_k.

$$e_k = [1+A]n_k \tag{3.16}$$

This can be interpreted as an input-output relationship where e_k is the input signal and n_k the output signal:

$$e_k \longrightarrow \boxed{[1+A]^{-1}} \longrightarrow n_k \tag{3.17}$$

If $\{e_k\}$ is a white noise sequence, then $\Psi_{ne}(i) = 0;\ i > 0$. This means that in this case

$$\underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q}\, \Omega^T(u,y)\underline{e}\right] = \underline{0} \tag{3.18}$$

so that

$$\underset{N\to\infty}{\text{plim}} \left[\underline{\hat{\theta}}\right] = \underline{\theta}_t \tag{3.19}$$

The requirement imposed on n_k is rather severe: the output noise n_k is an autoregressive (AR) type of filtering of a white noise sequence, using the AR parameters of the process as AR parameters of the noise filter. This will, of course, seldom occur in practical situations, hence the simple least squares estimator (3.10) is usually asymptotically biased.

By using a weighting on the measurements, we can arrive at the weighted least squares estimator. Define the weighted error criterion:

$$V = \frac{1}{N-q}\, \underline{\hat{e}}^T W \underline{\hat{e}} \tag{3.20}$$

where W is an appropriate weighting matrix of dimension (N-q)x(N-q)

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1(N-q)} \\ \vdots & & \vdots \\ w_{(N-q)1} & \cdots & w_{(N-q)(N-q)} \end{bmatrix} \tag{3.21}$$

Minimizing (3.20) with respect to $\underline{\theta}$, yields the weighted least squares estimator

$$\underline{\hat{\theta}}_{WLS} = [\Omega^T(u,y)W\Omega(u,y)]^{-1}\,\Omega^T(u,y)W\underline{y} \tag{3.22}$$

It will be obvious that the properties of this estimator highly depend on the choice of the weighting matrix W.
For the estimator (3.22), the probability limit can be given:

$$\operatorname*{plim}_{N\to\infty}[\underline{\hat{\theta}}_{WLS}] = \underline{\theta}_t + \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)W\Omega(u,y)\right]^{-1} \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q}\,\Omega^T(u,y)W\underline{e}\right] \tag{3.23}$$

where the second term of the right hand side gives the asymptotical bias. In the next paragraph, it will be shown that this can be made zero by the proper choice of W.

3.3 Development of the concept of three basic operations

In this paragraph we will introduce three basic operations related to the least-squares estimator. The first two operations, filtering and correlative weighting, will be derived from the weighting matrix, as already encountered in the previous paragraph. The third basic operation, model extension, will also be discussed.

In the forthcoming paragraphs the possible combinations of these three basic operations will be discussed and the existing, explicit estimation methods will then appear to be constructed by using one or more of these three operations.

3.3.1 The filtering type of weighting matrix

Consider the (N-q)x(N-q) weighting matrix

$$W = R^{-1} \tag{3.24}$$

where

$$R = E\{\underline{e}\ \underline{e}^T\} \tag{3.25}$$

We will also assume an ARMA parametrization for the signal $\underline{e}$ in the following way:

$$e_k = \frac{[1+C]_t}{[1+D]_t}\xi_k \tag{3.26}$$

where ξ_k represents a white noise sequence, and

$$\left.\begin{aligned}[1+C]_t &= [1+c_1 z^{-1} + \ldots\ldots + c_s z^{-s}]_t \\ [1+D]_t &= [1+d_1 z^{-1} + \ldots\ldots + d_r z^{-r}]_t\end{aligned}\right] \tag{3.27}$$

In appendix I it is shown that a good approximation for R^{-1} can be given by (if there are no poles and zeros on the unit circle):

$$R^{-1} = D_t'^T D_t' \sigma_\xi^{-2} \tag{3.28}$$

where the matrix D_t' is related to a finite polynomial $[1+D']_t$ of the pure autoregressive parameters, approximating the ARMA modelling of the noise filter:

$$[1+D']_t \approx \frac{[1+D]_t}{[1+C]_t} \tag{3.29}$$

The matrix D_t' is then a (N-q)x(N-q) lower triangular band matrix.

$$D_t' = \begin{bmatrix} 1 & & & & \\ d_1' & \ddots & & & \\ \vdots & & \ddots & & \\ d_\nu' & & & \ddots & \\ & \ddots & & & 1 \end{bmatrix} \tag{3.30}$$

containing the AR parameters of the filter defined by (3.29).
The matrix D_t' causes a filtering of the signals in the matrix $\Omega(u,y)$, yielding $\Omega(\tilde{u},\tilde{y})$:

$$\Omega(\tilde{u},\tilde{y}) = D_t'\Omega(u,y) \tag{3.31}$$

with

$$\left.\begin{aligned}\tilde{u}_k &= u_k + d_1' u_{k-1} + \ldots\ldots\ldots + d_\nu' u_{k-\nu} \\ \tilde{y}_k &= y_k + d_1' y_{k-1} + \ldots\ldots\ldots + d_\nu' y_{k-\nu}\end{aligned}\right] \tag{3.32}$$

This filtering can only be performed if the AR noise parameters are known a priori. The estimator (3.22) can now be written as:

$$\hat{\underline{\theta}} = [\Omega^T(\tilde{u},\tilde{y})\Omega(\tilde{u},\tilde{y})]^{-1}\Omega^T(\tilde{u},\tilde{y})\tilde{\underline{y}} \tag{3.33}$$

For this weighting the probability limit is:

$$\underset{N\to\infty}{\text{plim}}\ [\hat{\underline{\theta}}] = \underline{\theta}_t + \{\underset{N\to\infty}{\text{plim}}[\frac{1}{N-q}\ \Omega^T(\tilde{u},\tilde{y})\Omega(\tilde{u},\tilde{y})]\}^{-1}\underset{N\to\infty}{\text{plim}}[\frac{1}{N-q}\Omega^T(\tilde{u},\tilde{y})\underline{\xi}] \tag{3.34}$$

Now

$$\underset{N\to\infty}{\text{plim}}[\frac{1}{N-q}\ \Omega^T(\tilde{u},\tilde{y})\underline{\xi}] = \underline{0} \tag{3.35}$$

Consequently, the estimator (3.33) is consistent.

Analogously to the common least squares estimator, a schematic diagram of this estimator can be given; cf. fig. 3.3, where the known filters F_1 are used to perform the filtering given by eq. (3.32).

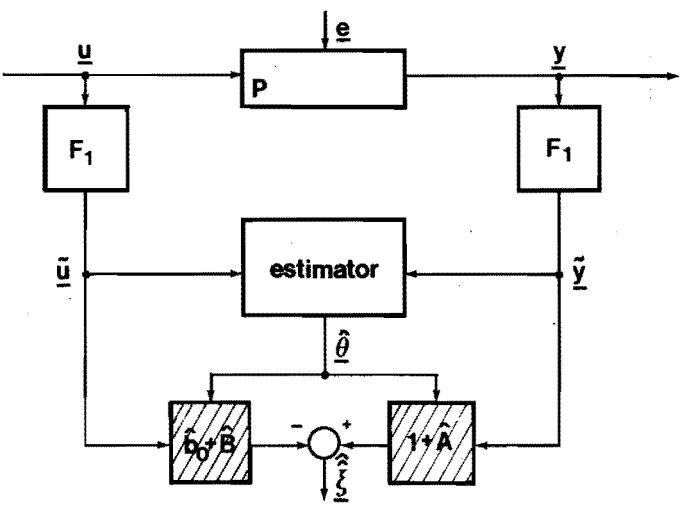


Fig. 3.3 Schematic diagram of a least squares estimator with weighted filtering

This estimator, using known noise parameters, is usually referred to as the Markov estimator. For models with only MA parameters, it can easily be proved that this estimator is unbiased for all N and that it yields a minimum variance estimate; cf. Goldberger (1964), Eykhoff (1974).

So far we have assumed that the filter parameters are known, so that the filtering will yield white residuals $\hat{\underline{\xi}}$. The variant where an approximation or estimation of the AR filter parameters $\underline{d}'_t$ is used

also belongs to the class of filtering type of weighting. This is necessary when exact knowledge about these parameters is lacking. These estimates can be obtained in an iterative estimation scheme where the available data are used several times successively, usually off-line. The results from a previous iteration are then used in the next iteration. In the case of filtering, the estimation results of noise parameters in a previous run are then used as filter parameters in the next run, which yields new estimates of noise parameters.

Clarke (1967) describes a method where the estimates $\hat{\underline{d}}'$of $\underline{d}'_t$ are used to filter the measurables $\underline{u}$ and $\underline{y}$ along the lines given in (3.32). This method is known as the generalized least squares estimator (GLS). The outline of the method is as follows:

ith iteration:

a) from the previous iteration i-1, the estimate $\hat{\underline{\theta}}^{i-1}$ and the filtered signals $\tilde{\underline{y}}^{i-1}$ and $\tilde{\underline{u}}^{i-1}$ are available.

b) perform (3.33) using $\tilde{\underline{u}}^{i-1}$ and $\tilde{\underline{y}}^{i-1}$ and yielding $\hat{\underline{\theta}}^{i}$.

c) generate the sequence

$$\hat{\underline{\xi}}^{i} = \tilde{\underline{y}}^{i-1} - \Omega(\tilde{u}^{i-1},\tilde{y}^{i-1})\ \hat{\underline{\theta}}^{i} \qquad (3.36)$$

d) estimate $\underline{d}'^{i}$ by:

$$\hat{\underline{d}}'^{i} = -\left[\hat{\underline{\underline{\xi}}}^{iT}\hat{\underline{\underline{\xi}}}^{i}\right]^{-1}\hat{\underline{\underline{\xi}}}^{iT}\hat{\underline{\xi}}^{i} \qquad (3.37)$$

where

$$\hat{\underline{\underline{\xi}}}^{i} = \begin{bmatrix} \hat{\xi}^{i}_{2} & & \\ | & \phi & \\ \hat{\xi}^{i}_{N} & & \hat{\xi}^{i}_{N-\nu} \end{bmatrix} \qquad (3.38)$$

e) filter $\tilde{\underline{u}}^{i-1}$ and $\tilde{\underline{y}}^{i-1}$ by $[1+\hat{D}'^{i}]$ yielding $\tilde{\underline{u}}^{i}$ and $\tilde{\underline{y}}^{i}$.

f) go to a) and proceed until convergence of the estimates occurs.

For starting this scheme the ordinary least-squares estimator can be used. Usually a first order model for the filter $[1+\hat{D}']$ is used,

as from step e) in the iterative scheme after a few iterations, a higher order modelling of the noise filter occurs.

Another version of this algorithm also exists, where no higher order modelling of the noise occurs after several iterations. In this version, in step e, the unfiltered signals u and y are filtered by the filters defined by the estimated AR noise parameters of the previous iteration. A higher order noise filter model can therefore then be necessary for obtaining reasonable noise whitening. The first version is capable of modelling higher order noise filters after some iterations, or modelling filters which are not purely autoregressive. Hastings-James and Sage (1969) proposed a recursive (and approximate) form of the GLS, which will be dealt with in chapter 4.

Stoica and Söderström (1977) proposed a GLS variant where MA noise parameters are estimated. Introduce:

$$\left.\begin{aligned} [1+\hat{C}^{i-1}]\tilde{u}_k^{i-1} &= \tilde{u}_k^{i-2} \\ [1+\hat{C}^{i-1}]\tilde{y}_k^{i-1} &= \tilde{y}_k^{i-2} \end{aligned}\right] \qquad (3.39)$$

The MA noise parameter estimates are being used here for AR filtering of the measured signals, whereas in the scheme of Clarke the AR noise parameter estimates are being used for MA filtering of the measured signals; cf. eq. (3.32). Starting from

$$[1+A]y_k = [b_o+B]u_k + [1+C]\hat{\xi}_k \qquad (3.40)$$

the following model will be formed by AR filtering of the signals $\tilde{u}^{i-2}$ and $\tilde{y}^{i-2}$; cf. eq. (3.39):

$$[1+\hat{A}^i]\tilde{y}_k^{i-1} = [\hat{b}_o^i+\hat{B}^i]\tilde{u}_k^{i-1} + \hat{\tilde{\xi}}_k^{i-1} \qquad (3.41)$$

with

$$\hat{\tilde{\xi}}_k^{i-1} = \frac{[1+C]_t}{\prod_{j=1}^{i-1}[1+\hat{C}^j]}\,\xi_k \qquad (3.42)$$

If $\hat{\tilde{\xi}}_k^{i-1}$ has nearly white properties, the ARMA process parameters will be estimated consistently, allowing proper estimation of the MA noise parameters.

A separate estimator is being used in all these variants to obtain the desired filter parameters using the estimates $\hat{\xi}_k^i$, generated by the estimator of the process parameters. Estimation schemes, which use in one iteration (or in one recursion step, as will be indicated in chapter 4) different estimators for different parameters will be denoted by bootstrap estimators. The results of one estimator is hereby used by the other estimator.

Due to the fact that only the output signal $\hat{\xi}_k^i$ of the (conceptual) noise filter is available, the corresponding schemes of the estimator are very simple; cf. fig. 3.4 for the two cases where AR or MA noise filter modelling is used. Here $\bar{\xi}_k^i$ denotes the corresponding input of the estimated noise filter.

$$\text{AR:} \quad [1+\hat{D}^i]\hat{\xi}_k^i = \bar{\xi}_k^i \tag{3.43a}$$

$$\text{MA:} \quad \hat{\xi}_k^i = [1+\hat{C}^i]\bar{\xi}_k^i \tag{3.43b}$$

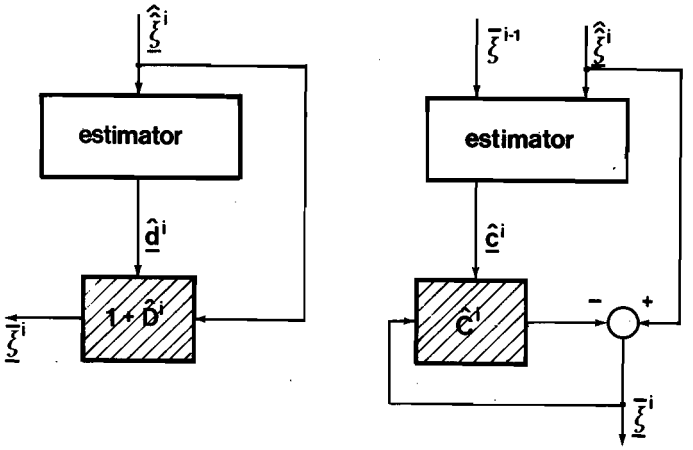


Fig. 3.4 LS estimators for AR or MA noise parameters.

A straightforward extension of the concept of GLS is to combine the parametrization of the noise, as used in the schemes considered. Such an ARMA modelling for GLS has not yet been proposed. For another (recursive), estimator the extended matrix method (EMM), such

an ARMA modelling for the noise has been used; cf. paragraph 4.4.4.

3.3.2 The correlative type of weighting matrix

Consider the (N-q)x(N-q) weighting matrix

$$W = \frac{1}{N-q} ZZ^T \tag{3.44}$$

The loss function is then

$$V = \frac{1}{(N-q)^2} \hat{\underline{e}}^T ZZ^T \hat{\underline{e}} \tag{3.45}$$

The estimator with this type of weighting is then, cf. White (1971) and Van den Boom (1976):

$$\hat{\underline{\theta}} = (\Omega^T ZZ^T \Omega)^{-1} \Omega^T ZZ^T \underline{y} \tag{3.46}$$

if the matrix $\frac{\Omega^T Z}{N-q}$ is non-singular for all N not too small, then

$$\hat{\underline{\theta}} = (Z^T \Omega)^{-1} Z^T \underline{y} \tag{3.47}$$

It can easily be shown that the minimum of this weighted loss function is zero. To this end recall $\hat{\underline{e}}$ and substitute (3.47)

$$\hat{\underline{e}} = \underline{y} - \Omega\hat{\underline{\theta}} = \underline{y} - \Omega(Z^T\Omega)^{-1} Z^T \underline{y} \tag{3.48}$$

Substitute this in (3.45):

$$V = \frac{1}{(N-q)^2} [\underline{y}^T - \underline{y}^T Z(\Omega^T Z)^{-1} \Omega^T] ZZ^T [\underline{y} - \Omega(Z^T\Omega)^{-1} Z^T \underline{y}] =$$

$$= \frac{1}{(N-q)^2} (\underline{y}^T Z - \underline{y}^T Z)(Z^T \underline{y} - Z^T \underline{y}) = 0 \tag{3.49}$$

As V is a quadratic scalar, this means that

$$\frac{1}{N-q} Z^T \hat{\underline{e}} = \underline{0} \tag{3.50}$$

For asymptotic unbiasedness the following two conditions should be fulfilled:

$$\left.\begin{aligned} &\text{I} \quad \operatorname*{plim}_{N\to\infty} \left[\frac{1}{N-q} Z^T \Omega\right] = \Gamma \qquad \Gamma \text{ non-singular} \\ &\text{II} \quad \operatorname*{plim}_{N\to\infty} \left[\frac{1}{N-q} Z^T \underline{e}\right] = \underline{0} \end{aligned}\right] \tag{3.51}$$

These two conditions are of the same form as already seen in (3.12) and (3.35), but they are also the conditions that Z has to fulfill in order to be an instrumental variable (IV); cf. Wong (1966), and Wong

and Polak (1967). These two conditions for Z indicate that z_k should be a quantity that is correlated with the process signals u_k (and hence y_k) but uncorrelated with the disturbances e_k.

A usual choice for z_k is the output of a (fixed) filter whose input is the process input u_k. The parameters of such a filter can be chosen rather freely. Often (delayed) estimates of the process parameters are used to generate the instrumental variable, which can then be seen as estimates of the undisturbed output signal. This requires that a few iterations have to be done.

Wouters (1972) proposed the use of delayed process inputs and Gersch (1970) the use of delayed process outputs. Stoica and Söderström (1979) used a combined scheme where both delayed inputs and outputs were used. In appendix II we investigate the feasibility of the use of such IV quantities and conclude that the delay should be chosen with care, as too large a delay can give numerical problems and too small a delay can violate the consistence demand for the choice of IV quantities.

Another possibility for choosing IV quantities, is making use of an extra signal, if available, which fulfills condition (3.51). Such a signal consists of an extra measurement with another independent measurement noise or a signal which constitutes the main characteristics of the measured signal, e.g. (an average over) signals measured previously over comparable "batches" or a normalized reference signal (e.g. in biomedical applications). This class can be widely extended to standard deterministic signals which are already known as a basis for analysis, e.g. sine/cosine, orthogonal functions or Walsh functions. For a more detailed description of this class of estimators, using so-called template functions, cf. Eykhoff (1980) and Eykhoff, Van den Boom and Van Rede (1981).

In fig. 3.5, the IV estimator is shown schematically. Here the IV filters Φ_1 and Φ_2 are introduced, which produce the above mentioned IV quantities using the input- or output signal or external signals. The estimator block is more complicated than in the previous figures as it processes the input-output measurables and the IV quantities.

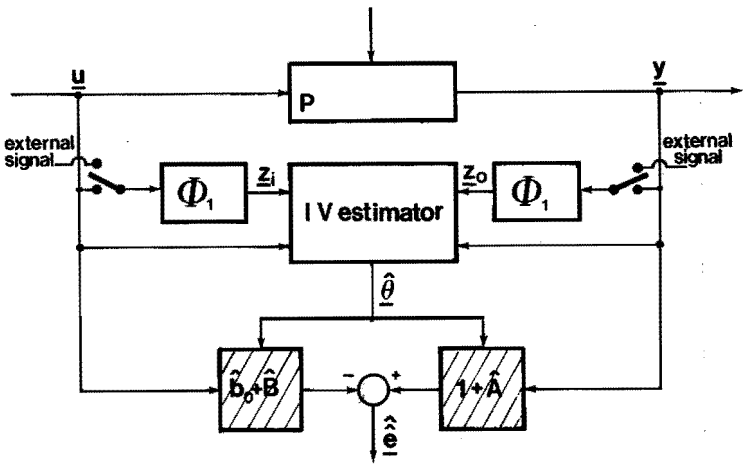


Fig. 3.5 Simple IV estimator

3.3.3 The model extension

As already mentioned, the estimator (3.10) will usually be biased due to the non-white equation error $\underline{e}$. By enlarging the model, the bias problem can be tackled as follows. Consider the following modelling:

$$[1+A]y_k = [b_o+B]u_k + \frac{1}{[1+D]}\hat{\hat{\xi}}_k \qquad (3.52)$$

$\hat{\hat{\xi}}_k$ being a white noise sequence. Multiplication with $[1+D]$ yields:

$$[1+A']y_k = [b_o'+B']u_k + \hat{\hat{\xi}}_k \qquad (3.53)$$

with

$$\left.\begin{aligned}[1+A'] &= [1+A][1+D] \\ [b_o'+B'] &= [b_o+B][1+D]\end{aligned}\right] \qquad (3.54)$$

Estimation of the enlarged parameter vector $(b_o',\ldots,b_{p'}',-a_1',\ldots,-a_{q'}')_t$ will be consistent if (3.53) has a white residual error $\hat{\hat{\xi}}_k$. An estimate of the original parameter vector $(b_o,\ldots,b_{p'},-a_1,\ldots,-a_{q'})_t$ can be found by determining the common factors of the estimated polynomials $[1+\hat{A}']$ and $[\hat{b}_o'+\hat{B}']$. This can give problems as, due to the uncertainties in the estimated parameters, the "common" polynomials in $[1+\hat{A}']$ and $[\hat{b}_o'+\hat{B}']$ are not exactly equal but close to each other. The problem of determining the common factors in this

context has been studied by several authors; cf. Söderström (1975) and Stoica (1976).

The above described method of estimation of process parameters using an over-parametrized model is of limited interest in practice, due to the (excessive) amount of computation involved. For unknown model orders, however, the use of over-parametrized models and determination of the common factors give a good and reliable order testing method; cf. chapter 7.

If we consider the model (3.53) in somewhat more detail, we come to the following schematic diagram; cf. eq. (3.54) and fig. 3.6a, where M_1 and M_2 represent the MA and AR parameters and M_3 the common factors.

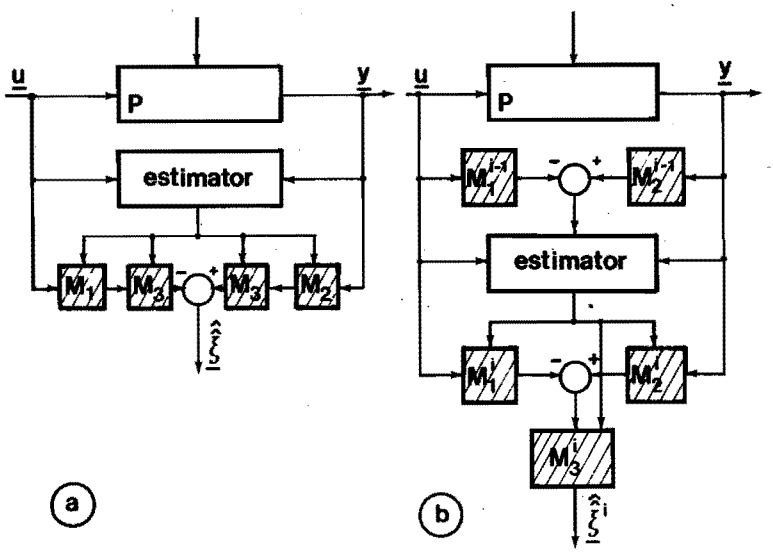


Fig. 3.6 LS Estimators with model extension

If we start from fig. 3.6a, it is easy to arrive at fig. 3.6b, as we are dealing with linear filters. Nevertheless there is a difference. In the case of fig. 3.6a an extended ARMA model of the process parameters is constructed, whereas, in the case of fig. 3.6b, a model of the noise is constructed. This means that these noise signals should

also be fed to the estimator block as indicated. These noise signals are not available, so they should be generated using estimation results of prior iterations. For a recursive estimation scheme, this is quite straightforward, as during the recursion, the estimates of these noise signals can be made available. This will be dealt with in detail in chapter 4.

A more general representation of the dynamical characteristics of the noise is the ARMA model

$$[1+D]e_k = [1+C]\xi_k \qquad (3.55)$$

This means that the model block M_3 in fig. 3.6b has to be split into two parts, as will be indicated in some of the forthcoming schemes.

3.4 General scheme for explicit estimators

In this paragraph we will present a comprehensive view of existing explicit estimation methods based on the classification according to the three basic operations. In figure 3.7 a general diagram is given, which presents the different explicit methods as particular cases. In table 3.1 these methods are summarized. They are subdivided into levels of complexity 0,1 and 2 according to the number of basic operations involved. The estimators belonging to the levels 0 and 1 have been mainly dealt with in the previous paragraph. The estimator which utilizes model extension - the over-parametrized least squares - is listed in the table for completeness, but is not important as an explicit estimator. For a recursive scheme, this principle is attractive, yielding the important Extended Matrix Method estimator (EMM).

The estimates with combine two of the three basic operations discussed in the previous paragraph will be dealt with in this paragraph, where they represent existing explicit estimation schemes. These combinations are:

- filtering and instrumental variable
- filtering and model extension

The third combination, instrumental variable and model extension, has not been proposed as an explicit scheme. We will meet this combination among the implicit estimators in the next chapter.

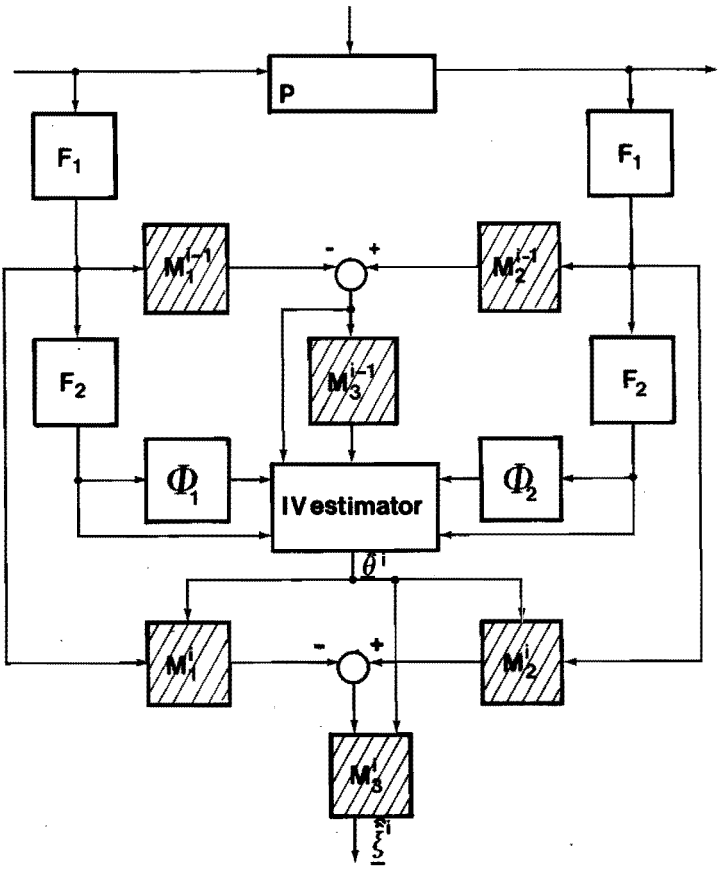


Fig. 3.7 General diagram for explicit estimators

3.4.1 Filtering and instrumental variable combined

Wong (1966) showed that the estimator

$$\hat{\underline{\theta}} = [\Omega^T(u,x)R^{-1}\Omega(u,y)]^{-1}\,\Omega^T(u,x)R^{-1}\underline{y} \qquad (3.56)$$

is optimal with respect to its behaviour of the covariance for large N. This estimator, nevertheless, is only of theoretical importance, as the undisturbed output signal x_k is not available, and the covariance matrix R is usually not known. Therefore an approximated but useful version is given by estimator (3.57) where a model output z_k is used as an estimate for the undisturbed process output x_k.

	Method	Model			signal filters		Instr. Var. filters		
		M_1	M_2	M_3	F_1	F_2	Φ_1	Φ_2	
0	Least Squares	$[\hat{b}_o+\hat{B}]$	$[1+\hat{A}]$	-	-	-	-	-	
1	Generalized Least Squares	a) $[\hat{b}_o+\hat{B}]$	$[1+\hat{A}]$	-	$[1+\hat{D}]$	-	-	-	Separate estimator for $[1+\hat{D}]$
		b) $[b_o+\hat{B}]$	$[1+\hat{A}]$	-	$[1+\hat{C}]^{-1}$	-	-	-	" for $[1+\hat{C}]$
1	Instrumental Variable	$[\hat{b}_o+\hat{B}]$	$[1+\hat{A}]$	-	-	-	a) fixed b) $\dfrac{[\hat{b}_o+\hat{B}]}{[1+\hat{A}]}$ c) delay d) - e) delay	- - - - - delay delay	
1	Least Squares, overparametrized	$[\hat{b}'_o+\hat{b}']$	$[1+\hat{A}']$	-	-	-	-	-	common factors in M_1 and M_2
2	IQL/AML	a) $[\hat{b}^i_o+\hat{B}^i]$	$[1+\hat{A}^i]$	$[1+\hat{C}^i]$	$[1+\hat{C}^{i-1}]^{-1}$	-	-	-	
		b) $[\hat{b}^i_o+\hat{B}^i]$	$[1+\hat{A}^i]$	$[1+\hat{C}^i]$ $[1+D^i]$	$[1+\hat{C}^{i-1}]^{-1}$	$[1+\hat{D}^{i-1}]$	-	-	
2	Suboptimal IV	$[\hat{b}^i_o+\hat{B}^i]$	$[1+\hat{A}^i]$	-	$[1+\hat{D}^{i-1}]$	-	$\dfrac{[\hat{b}_o+\hat{B}^{i-1}]}{[1+\hat{A}^{i-1}]}$	-	separate estimator for $[1+\hat{D}]$

Table 3.1

This estimator is called the suboptimal IV estimator.

$$\underline{\hat{\theta}} = [\Omega^T(u,z)\hat{D}^T\hat{D}\Omega(u,y)]^{-1}\ \Omega^T(u,z)\hat{D}^T\hat{D}\underline{y} \qquad (3.57)$$

Here also, estimates $\underline{\hat{d}}'$ of the AR parameters of the noise filter are used to filter the signals z_k, u_k and y_k as an approximation for the weighting function R^{-1}.

3.4.2 Filtering and model extension combined

In paragraph 3.3.1 we have argued that iteration of the solution can improve the estimates substantially. If we consider an MA description after the i-th iteration:

$$[1+\hat{A}^i]y_k = [\hat{b}_o^i+\hat{B}^i]u_k + [1+\hat{C}^i]\hat{\xi}_k^i \qquad (3.58)$$

we see that the product of the estimated terms $[1+\hat{C}^i]$ and $\hat{\xi}_k^i$ causes a non-linearity in the parameters.

Schultz (1968) used a quasilinearization technique to estimate the denominator polynomial of the transfer function $\frac{[b_o+B]_t}{[1+A]_t}$ in a noise-free experiment, which is a nonlinear estimation problem. Fuhrt (1973a, 1973b) applied this technique to the nonlinear estimation problem of (3.58). He presented his method, the Implicit Quasi-Linearization (IQL), within the framework of maximum likelihood estimators, but the method can also be presented within the scope of least squares estimators. The error is linear in the parameters a_i and b_i but nonlinear in the noise parameters c_i. Applying Gauss-Newton technique for minimalization of $\hat{\xi}_k^{i+1}$ is then linearizing this equation around the solution of the i-th iteration. This yields an expression for the i+1-st iteration:

$$[1+\hat{C}^i]\hat{\xi}_k^i + [\hat{C}^{i+1}-\hat{C}^i]\hat{\xi}_k^i + [1+\hat{C}^i][\hat{\xi}_k^{i+1}-\hat{\xi}_k^i] =$$

$$= [1+\hat{A}^{i+1}]y_k - [\hat{b}_o^{i+1}+\hat{B}^{i+1}]u_k \qquad (3.59)$$

This can be rewritten:

$$\hat{\xi}_k^{i+1} = \frac{[1+\hat{A}^{i+1}]}{[1+\hat{C}^i]}y_k - \frac{[\hat{b}_o^{i+1}+\hat{B}^{i+1}]}{[1+\hat{C}^i]}u_k - \frac{[1+\hat{C}^{i+1}]}{[1+\hat{C}^i]}\hat{\xi}_k^i + \hat{\xi}_k^i \qquad (3.60)$$

Introduce:

$$\left.\begin{aligned} \tilde{y}_k^i &= \frac{1}{[1+\hat{C}^i]}\, y_k \\ \tilde{u}_k^i &= \frac{1}{[1+\hat{C}^i]}\, u_k \\ \tilde{\xi}_k^i &= \frac{1}{[1+\hat{C}^i]}\, \hat{\xi}_k^i \end{aligned}\right] \qquad (3.61)$$

Equation (3.61) can then be rewritten as:

$$\tilde{y}_k^i = [\hat{b}_o^{i+1}+\hat{B}^{i+1}]\tilde{u}_k^i - [\hat{A}^{i+1}]\tilde{y}_k^i + [\hat{C}^{i+1}]\tilde{\xi}_k^i + \tilde{\xi}_k^i - \hat{\xi}_k^i + \hat{\xi}_k^{i+1} \qquad (3.62)$$

Now define:

$$\Omega(\tilde{u}^i,\tilde{y}^i,\tilde{\xi}^i) = \left[\begin{array}{ccc|ccc|ccc} \tilde{u}_1^i & & \phi & & & & & & \\ & \searrow & & & & & & & \\ & & \tilde{u}_1^i & & & & & & \\ \vdots & & \vdots & & & & & & \\ \tilde{u}_N^i & & \tilde{u}_{N-p}^i & \tilde{y}_{N-1}^i & & \tilde{y}_{N-q}^i & \tilde{\xi}_{N-1}^i & & \tilde{\xi}_{N-s}^i \end{array}\right] \qquad (3.63)$$

and

$$\tilde{\underline{w}}^i = \tilde{\underline{y}}^i - \tilde{\underline{\xi}}^i + \hat{\underline{\xi}}^i \qquad (3.64)$$

so that (3.62) can be written in matrix form:

$$\tilde{\underline{w}}^i = \Omega(\tilde{u}^i,\tilde{y}^i,\tilde{\xi}^i)\ \hat{\underline{\theta}}^{i+1} + \hat{\underline{\xi}}^{i+1} \qquad (3.65)$$

Minimizing

$$V = \frac{1}{N}\, \hat{\underline{\xi}}^{i+1\,T}\, \hat{\underline{\xi}}^{i+1} \qquad (3.66)$$

yields

$$\hat{\underline{\theta}}^{i+1} = [\Omega^T(\tilde{u}^i,\tilde{y}^i,\tilde{\xi}^i)\ \Omega(\tilde{u}^i,\tilde{y}^i,\tilde{\xi}^i)]^{-1}\ \Omega^T(\tilde{u}^i,\tilde{y}^i,\tilde{\xi}^i)\tilde{\underline{w}}^i \qquad (3.67)$$

Fuhrt (1973a) shows results which are good for a variety of simulated conditions. Due to the linearizing around the solution of the i-th iteration, convergence is not guaranteed. Therefore he proposes a start up of his method by using Clarke's method, as explained in a preceeding section.

This method is rather time consuming due to the filtering operations combined with the extended matrix $\Omega(\tilde{u}^i, \tilde{y}^i, \tilde{\xi}^i)$. Van den Hoven (1978) extended this principle of quasi-linearization to the more general models of the form:

$$[1+\hat{A}^i]y_k = [\hat{b}_o+\hat{B}^i]u_k + \frac{[1+\hat{C}^i]}{[1+\hat{D}^i]}\hat{\xi}^i_k \tag{3.68}$$

Linearizing yields:

$$\hat{\xi}^{i+1}_k = [1+\hat{A}^{i+1}]\frac{[1+\hat{D}^i]}{[1+\hat{C}^i]}y_k - [\hat{b}^{i+1}_o+\hat{B}^{i+1}]\frac{[1+\hat{D}^i]}{[1+\hat{C}^i]}u_k -$$

$$- [1+\hat{C}^{i+1}]\frac{1}{[1+\hat{C}^i]}\hat{\xi}^i_k + [1+\hat{D}^{i+1}]\frac{1}{[1+\hat{D}^i]}\hat{\xi}^i_k \tag{3.69}$$

Define the following filtered variables:

$$\left.\begin{aligned}
\tilde{y}^i_k &= \frac{[1+\hat{D}^i]}{[1+\hat{C}^i]}y_k \\
\tilde{u}^i_k &= \frac{[1+\hat{D}^i]}{[1+\hat{C}^i]}u_k \\
\tilde{\xi}^i_k &= \frac{1}{[1+\hat{C}^i]}\hat{\xi}^i_k \\
\tilde{\tilde{\xi}}^i_k &= \frac{1}{[1+\hat{D}^i]}\hat{\xi}^i_k
\end{aligned}\right] \tag{3.70}$$

Equation (3.69) can then be written as:

$$\tilde{y}^i_k = [\hat{b}^{i+1}_o+\hat{B}^{i+1}]\tilde{u}^i_k - [\hat{A}^{i+1}]\tilde{y}^i_k + [\hat{C}^{i+1}]\tilde{\xi}^i_k - [\hat{D}^{i+1}]\tilde{\tilde{\xi}}^i_k + \tilde{\xi}^i_k - \tilde{\tilde{\xi}}^i_k + \hat{\xi}^{i+1}_k \tag{3.71}$$

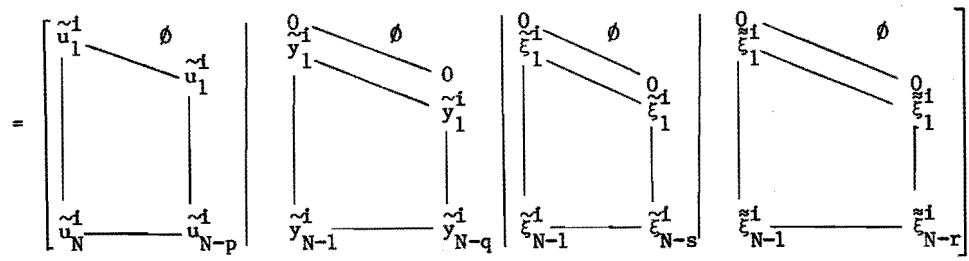
in matrix notation:

$$\underline{\tilde{w}}^i = \Omega(\tilde{u}^i, \tilde{y}^i, \tilde{\xi}^i, \tilde{\tilde{\xi}}^i)\,\underline{\hat{\theta}}^{i+1} + \underline{\hat{\xi}}^{i+1} \tag{3.72}$$

$$\underline{\tilde{w}}^i = \underline{\tilde{y}}^i - \underline{\tilde{\xi}}^i + \underline{\tilde{\tilde{\xi}}}^i$$

where

$$\Omega(\tilde{u}^i, \tilde{y}^i, \tilde{\xi}^i, \tilde{\xi}^i) =$$



(3.73)

Minimizing the loss function:

$$V = \frac{1}{N} \hat{\underline{\xi}}^{i+1^T} \hat{\underline{\xi}}^{i+1} \tag{3.74}$$

yields:

$$\hat{\underline{\theta}}^{i+1} = [\Omega^T(\tilde{u}^i, \tilde{y}^i, \tilde{\xi}^i, \tilde{\xi}^i)\ \Omega(\tilde{u}^i, \tilde{y}^i, \tilde{\xi}^i, \tilde{\xi}^i)]^{-1}\ \Omega^T(\tilde{u}^i, \tilde{y}^i, \tilde{\xi}^i, \tilde{\xi}^i)\underline{\tilde{w}} \tag{3.75}$$

3.5 Relation with the Maximum Likelihood Estimator (MLE)

Although an extensive treatment of the maximum likelihood estimator lies outside the scope of this chapter, it is interesting to discuss the relationship between the Maximum Likelihood Estimator (MLE) and the Least Squares Estimators (LSE). The principle of the MLE was already pointed out by Gauss (1809), also stating that MLE and LSE are two basic methods, closely related but independent.

The MLE maximizes a likelihood function of a random variable $\underline{x}$ with respect to $\underline{\theta}$, whose probability density function is $p(\underline{x}; \underline{\theta})$; $\underline{\theta}$ being an unknown parameter. The likelihood function is essentially this probability density function. Fisher (1922) investigated the MLE and showed its favourable properties like consistence, asymptotic efficiency and asymptotic normality under very general conditions.
Åström and Bohlin (1965) used the MLE for the estimation of parameters of dynamic systems and showed the usefulness and reliability of the method.

As log likelihoodfunction we choose:

$$L = \ln\{p(\underline{\hat{\xi}}|\underline{u},\underline{\theta})\} = \sum_{k=1}^{N} \ln\{p(\hat{\xi}_k|\underline{u},\underline{\theta})\} \qquad (3.76)$$

where $\underline{\hat{\xi}}$ is the model residual and is assumed to be an independent random variable so that the probability density function $p(\underline{\hat{\xi}}|\underline{u},\underline{\theta})$ can be written as a product of the probability densities $p(\hat{\xi}_1|\underline{u},\underline{\theta})$. This can be achieved by proper modelling of the noise parameters. Maximization of L leads to the following equation:

$$\left.\frac{\partial L}{\partial \underline{\theta}}\right|_{\underline{\theta}=\underline{\hat{\theta}}} = \frac{\partial}{\partial \underline{\theta}} \sum_{k=1}^{N} \ln\{p(\hat{\xi}_k|\underline{u},\underline{\theta})\}\Big|_{\underline{\theta}=\underline{\hat{\theta}}} = \underline{0} \qquad (3.77)$$

This equation is valid for any probability density function. In the literature the solution of (3.77) is worked out for a normal density, leading to:

$$L(\underline{\theta}) = \ln\{(\frac{1}{\sigma_{\hat{\xi}}\sqrt{2\pi}})^N \prod_{k=1}^{N} \exp(\frac{-\frac{1}{2}\hat{\xi}_k^2}{\sigma_{\hat{\xi}}^2})\} =$$

$$= -\frac{1}{2\sigma_{\hat{\xi}}^2} \cdot \sum_{k=1}^{N} \hat{\xi}_k^2 - N\ln\sigma_{\hat{\xi}} - \frac{N}{2}\ln 2\pi \qquad (3.78)$$

An estimate of $\sigma_{\hat{\xi}}^2$ can be obtained by putting

$$\left.\frac{\partial L}{\partial \sigma_{\hat{\xi}}}\right|_{\sigma_{\hat{\xi}} = \hat{\sigma}_{\hat{\xi}}} = 0 \qquad (3.79)$$

yielding

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^{N} \hat{\xi}^2 \qquad (3.80)$$

Maximizing eq. (3.78) with respect to $\underline{\theta}$ is equivalent to minimizing

$$V(\underline{\theta}) = \frac{1}{N} \sum_{k=1}^{N} \hat{\xi}^2 \qquad (3.81)$$

where from

$$\left.\frac{\partial V}{\partial \underline{\theta}}\right|_{\underline{\theta}=\underline{\hat{\theta}}} = \underline{0} \qquad (3.82)$$

the expressions for the estimator $\underline{\hat{\theta}}$ follow.

For a modelling of the noise with an autoregressive model [1+D], the residual $\hat{\xi}$ is nonlinear in these parameters. This will also be

the case for a modelling with pure MA parameters. Due to such a non-linearity, linearization of the loss function V will be necessary:

$$V(\underline{\theta}+\Delta\underline{\theta}) = V(\underline{\theta})+V'^{T}(\underline{\theta})\Delta\underline{\theta} + \tfrac{1}{2}\Delta\underline{\theta}^{T}V''(\underline{\theta})\Delta\underline{\theta} + \ldots. \tag{3.83}$$

where

$$V'^{T}(\underline{\theta}) = \left(\frac{\partial V}{\partial\theta_1},\ldots\ldots,\frac{\partial V}{\partial\theta_n}\right) \tag{3.84}$$

$$V''(\underline{\theta}) = \begin{bmatrix} \frac{\partial^2 V}{\partial\theta_1\partial\theta_1} & \cdots & \frac{\partial^2 V}{\partial\theta_1\partial\theta_n} \\ \vdots & & \vdots \\ \frac{\partial^2 V}{\partial\theta_n\partial\theta_1} & \cdots & \frac{\partial^2 V}{\partial\theta_n\partial\theta_n} \end{bmatrix} \tag{3.85}$$

Now several gradient methods can be used in order to find iteratively the estimated parameter vector $\hat{\underline{\theta}}$. For a review of these possible methods; cf. Eykhoff (1974). Usually a Newton-Raphson algorithm is used; cf. Åström and Bohlin (1965). The expression for the i-th iteration is:

$$\hat{\underline{\theta}}^{i+1} = \hat{\underline{\theta}}^{i} - [V''(\hat{\underline{\theta}}^{i})]^{-1} V'(\hat{\underline{\theta}}^{i}) \tag{3.86}$$

It can easily be shown that this reduces to the general expression (3.10) for least squares estimation if only process parameters are involved. In this case, the algorithm (3.86) converges in one step.

From this short review of the maximum likelihood estimator it will be clear that this is a method which, in principle, is suited for any specific distribution but which, in practice, is discussed only for normally distributed signals. However, in this limited case it coincides with the least squares estimator. To this end it is necessary to choose such a type of model from which independent residuals are obtained. Generally, the minimization of the quadratic loss function leads to a non-linear problem due to the noise parameters encountered. Iterative solutions are then necessary, such as Newton-Raphson schemes or the scheme as proposed by Fuhrt or Clarke, to find the minima of this loss function.

It is a pity that a confusion can be noticed in the literature with respect to the use of the terms maximum likelihood estimators, approximate maximum likelihood estimators and (weighted) least squares estimators. Usually these terms are used for situations where noise modelling is used, leading to the filtering concepts as treated before. Then no special attention is paid to the requirements of gaussian distribution with respect to the maximum likelihood method. It would give a clearer picture if the term (weighted) least squares estimators were used for cases where no special attention is paid to the distribution of the signal, i.e. when dealing with a distribution free method.

3.6 The accuracy of the least squares and maximum likelihood estimators

For a process, which can be modelled by a (limited) number of MA parameters, useful expressions for the covariance of the estimate exist; cf. Eykhoff (1974).

Starting from

$$\underline{y} = U\underline{b}_t + \underline{e} \tag{3.87}$$

The weighted least squares estimate is

$$\underline{\hat{b}} = [U^T W U]^{-1} U^T W \underline{e} \tag{3.88}$$

If $E\{\underline{e}\} = \underline{0}$ and $\underline{e}$ and $\underline{u}$ are mutually uncorrelated:

$$E\{\underline{\hat{b}}\} = \underline{b}_t \tag{3.89}$$

$$\text{cov } \underline{\hat{b}} = [U^T W U]^{-1} U^T W E\{\underline{e}\,\underline{e}^T\} W U [U^T W U]^{-1} \tag{3.90}$$

For a Markov estimator the weighting matrix is the inverse covariance matrix of the noise:

$$W = [E\{\underline{e}\,\underline{e}^T\}]^{-1} = R^{-1} \tag{3.91}$$

then

$$\text{cov } \underline{\hat{b}} = [U^T R^{-1} U]^{-1} \tag{3.92}$$

The importance of this estimator lies in the fact that it yields an unbiased minimum variance estimate. For this reason it is also the basis for the concept of the filtering type of weighting using

$$D_t'^T D_t' = R^{-1} \tag{3.93}$$

For models where ARMA parametrization has to be used, no expressions for the covariance for finite N are known to us, nor expressions for the expectation of the estimate for finite N. Nevertheless the ideas with respect to filtered weighting, corresponding to a Markov type of weighting matrix, are for this type of modelling widely accepted, as they also yield consistent estimates.

For consistent estimators, a lower bound for the covariance of the estimates can be given. This is the Cramér-Rao lower bound (CRLB) due to Cramér (1946) and Rao (1945). Introducing the Fisher information matrix J:

$$J = -E\left\{\frac{\partial^2 L}{\partial \underline{\theta} \partial \underline{\theta}^T}\right\}\Bigg|_{\underline{\theta}=\underline{\theta}_t} \tag{3.94}$$

where L is the log likelihood function, we find the Cramér-Rao inequality:

$$\text{cov } \underline{\hat{\theta}} = E\{(\underline{\hat{\theta}}-\underline{\theta}_t)(\underline{\hat{\theta}}-\underline{\theta}_t)^T\} \geq J^{-1} \tag{3.95}$$

Åström and Bohlin (1965) and Åström (1970) used this result for situations where an ARMA model is used for the process dynamics and a MA model for the noise dynamics. Costongs (1979) used the Cramér-Rao lower bound formulas for the general model where the process dynamics and the noise dynamics are modelled by ARMA models of the following form:

$$\frac{[1+C]}{[1+D]}\lambda\hat{\xi}_k = [1+A]y_k - [b_o+B]u_k \tag{3.96}$$

where $\hat{\xi}_k$ is assumed to have normal N(0,1) statistics. Introduce the variable $\hat{\xi}_k$ defined by:

$$\frac{[1+C]}{[1+D]}\hat{\xi}_k = [1+A]y_k - [b_o+B]u_k \tag{3.97}$$

The logarithm of the likelihood function is; cf. eq. (3.78):

$$L = \text{constant} - N\ln\lambda - \frac{1}{2\lambda^2}\sum_{k=1}^{N}\hat{\xi}_k^2 \tag{3.98}$$

Taking the expectation of the second partial derivatives ; cf. Appendix IV:

$$E\left\{\frac{\partial^2 L}{\partial \lambda^2}\right\}\Bigg|_{\lambda=\lambda_t} = -\frac{2N}{\lambda^2} \triangleq -J_{\lambda\lambda} \tag{3.99}$$

$$E\left\{\frac{\partial^2 L}{\partial\theta_i\,\partial\lambda}\right\}\Bigg|_{\lambda=\lambda_t;\ \theta_i=\theta_{i,t}} = 0 \qquad (3.100)$$

$$E\left\{\frac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right\}\Bigg|_{\theta_i=\theta_{i,t};\theta_j=\theta_{j,t}} = -\frac{1}{\lambda^2}\sum_{k=1}^{N} E\left\{\frac{\partial\xi_k}{\partial\theta_i}\cdot\frac{\partial\xi_k}{\partial\theta_j}\right\} \qquad (3.101)$$

$$J = \begin{bmatrix} J_{bb} & J_{ba} & J_{bc} & J_{bd} & J_{b\lambda} \\ J_{ab} & J_{aa} & J_{ac} & J_{ad} & J_{a\lambda} \\ J_{cb} & J_{ca} & J_{cc} & J_{cd} & J_{c\lambda} \\ J_{db} & J_{da} & J_{dc} & J_{dd} & J_{d\lambda} \\ J_{\lambda b} & J_{\lambda a} & J_{\lambda c} & J_{\lambda d} & J_{\lambda\lambda} \end{bmatrix} \qquad (3.102)$$

From eq. (3.100) it follows that

$$J_{a\lambda} = J_{b\lambda} = J_{c\lambda} = J_{d\lambda} = J_{\lambda a} = J_{\lambda b} = J_{\lambda c} = J_{\lambda d} = 0$$

and $J_{\lambda\lambda}$ is given in eq. (3.99). Eq. (3.101) determines 16 submatrices in eq. (3.102). In appendix IV the computation of these submatrices is performed for normal distributions. As a result of this it can be stated that J can be computed if all the parameters of the polynomials $[1+A]_t$, $[b_o+B]_t$, $[1+C]_t$, $[1+D]_t$ and λ_t are known and if the autocorrelation of the input signal u_k is known. Consequently, in practical cases, this lower bound is not applicable without modification. From various experiments, cf. Costongs (1979), it has been shown that

a) the use of estimated parameters instead of the true parameters does not affect the reliability of the approximated lower bound very much;
b) the existing (recursive) estimation schemes, which will be dealt with in chapter 4, give estimation results with a quality which is very close to the (real) CRLB;
c) the assumption of gaussian data does not seem to be very crucial. Simulation with rectangular distributions gave comparable results.

These observations make the approximate CRLB a very useful tool in interpreting the estimation results of these methods.

From appendix IV it follows further that:

$$J = N\begin{bmatrix} \{\frac{1}{\lambda^2}r_{22}(i-j)\} & \{\frac{1}{\lambda^2}r_{12}(i-j)\} & \emptyset & \emptyset & 0 \\ \{\frac{1}{\lambda^2}r_{21}(i-j)\} & \{\frac{1}{\lambda^2}r_{11}(i-j)+r_{33}(i-j)\} & \{r_{35}(i-j)\} & \{r_{34}(i-j)\} & 0 \\ \emptyset & \{r_{53}(i-j)\} & \{r_{55}(i-j)\} & \{r_{54}(i-j)\} & 0 \\ \emptyset & \{r_{43}(i-j)\} & \{r_{45}(i-j)\} & \{r_{44}(i-j)\} & 0 \\ 0 & 0 & 0 & 0 & \frac{2N}{\lambda^2} \end{bmatrix} \qquad (3.103)$$

where the correlation functions are defined as:

$$\left.\begin{aligned} r_{ij}(\tau) &= \frac{1}{N}\sum_{k=1}^{N} x_{i,k}x_{j,k+\tau} \qquad i,j = \{1,2\} \\ r_{ij}(\tau) &= E\{x_{i,k}x_{j,k+\tau}\} \qquad i,j = \{3,4,5\} \end{aligned}\right\} \qquad (3.104)$$

with:

$$\left.\begin{aligned} x_{1,k} &= \frac{[b_o+B][1+D]}{[1+A][1+C]}u_k \\ x_{2,k} &= \frac{[1+C]}{[1+D]}u_k \\ x_{3,k} &= \frac{1}{[1+A]}\xi_k \\ x_{4,k} &= \frac{1}{[1+D]}\xi_k \\ x_{5,k} &= \frac{1}{[1+C]}\xi_k \end{aligned}\right\} \qquad (3.105)$$

In order to gain insight into the accuracy of the different estimated parameters for various signal-to-noise ratios, we will assume that the input signal is a white noise signal and we will consider the different parameters separately. The matrix inversion in

$$\text{cov } \underline{\hat{\theta}} \geq J^{-1} \tag{3.106}$$

causes interference of the different parameters.

a. The A-parameters

The covariance, when estimating only A-parameters, is given by:

$$\text{cov } \underline{\alpha} \geq \frac{1}{N\{\frac{1}{\lambda^2} r_{11}(i-j) + r_{33}(i-j)\}_{ij}} \tag{3.107}$$

We can distinguish two cases

a.1

$$r_{33}(i-j) \ll \frac{1}{\lambda^2}\, r_{11}(i-j) \tag{3.108}$$

This is the case for high S/N ratios, i.e. for rather undisturbed output signals. Inequality (3.108) can be rewritten:

$$\text{cov } \alpha \geq \frac{\lambda^2}{N\{r_{11}(i-j)\}_{ij}} \tag{3.109}$$

which means that the cov $\underline{\alpha}$ will increase for increasing λ^2

a.2

$$r_{33}(i-j) \gg \frac{1}{\lambda^2}\, r_{11}(i-j) \tag{3.110}$$

This is the case for low S/N ratios, i.e. very disturbed outputs. Eq. (3.107) is then:

$$\text{cov } \underline{\alpha} \geq \frac{1}{N\{r_{33}(i-j)\}_{ij}} \tag{3.111}$$

which is constant for all λ^2. In this case the component of the process output in the signal y_k is no longer dominating for identification of the A-parameters, but the corrupting noise.

b. The B-parameters

The covariance, when estimating only B-parameters, is given by:

$$\text{cov } \underline{\beta} \geq \frac{\lambda^2}{N\{r_{22}(i-j)\}_{ij}} \tag{3.112}$$

which is increasing for increasing λ^2.

c. The C- and D-parameters

The covariances, when estimating only C-parameters, or only D-parameters are given by resp.:

$$\text{cov } \underline{\gamma} \geq \frac{1}{N\{r_{55}(i-j)\}_{ij}} \tag{3.113}$$

$$\text{cov } \underline{\delta} \geq \frac{1}{N\{r_{44}(i-j)\}_{ij}} \tag{3.114}$$

These lower bounds are constant for all λ^2.

In figure 3.8 this behaviour for the different covariances is sketched. In Talmon and Van den Boom (1973), this behaviour has already been presented, but based on simulation results. In chapter 6 we will present, in more detail, comparable results obtained from simulations with different estimation methods and confronted with the real and the approximated Cramér-Rao lower bounds. From these simulations it can be concluded that, if all A, B, C, D parameters are estimated simultaneously, the above mentioned behaviour for separate estimation of the parameters will also occur.

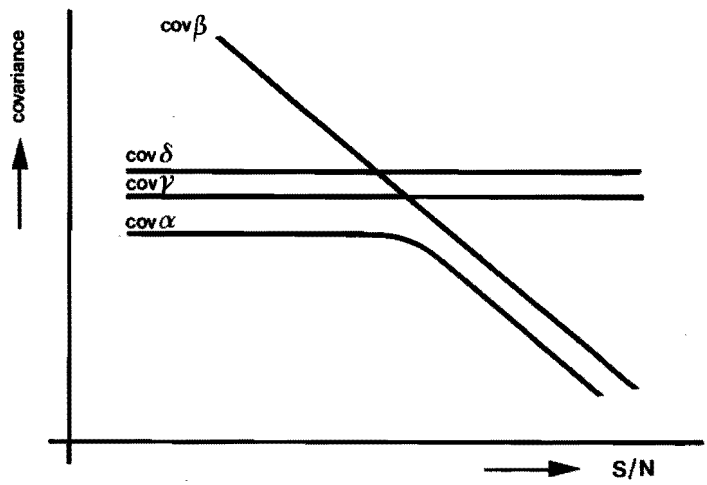


Fig. 3.8 Typical behaviour for the covariances of the different parameter estimates as a function of the noise level.

3.7 Conclusions

In this chapter we discussed the explicit least squares estimation schemes and met the consistency problem of the common least squares estimator. Two basic conditions for consistence are mentioned; these conditions also coincide with the definitions which characterize the class of instrumental variable estimators. Therefore a close link between the instrumental variable estimators and the least squares estimators can be found, and for several instrumental variable estimators proposed in the past the corresponding weighted loss function in the least squares sense is given.

For the construction of consistent least squares estimators, three different approaches can be distinguished.

A. the correlative weighting, leading to the instrumental variables estimators.
B. the filtered weighting, leading to filtered observation data.
C. the model extension, leading to an extended parameter set.

These three basic approaches can be used separately for tackling the consistence problem, but two or even three approaches can be combined within one scheme. This will be more apparent when we consider the recursive schemes in the next chapter.

For the correlative weighting an instrumental variable has to be generated. This can be done by making use of the measured input or output signals or by making use of an extra measurement. This gives the possibility of introducing extra (a priori) information.

For the filtered type of weighting, a priori knowledge of the noise characteristics has to be available. This can be obtained by applying the type C approach, which is model extension, as already seen in the methods of Fuhrt and Söderström. Now recursive methods are interesting as they deliver this wanted information during the estimation process.

The model extension is an efficient method for obtaining white residuals by filtering the correlated noise through the extended model.

The expression for the Cramér-Rao lower bound for the variance of the estimated parameters was given. This lower bound is derived within the context of the maximum likelihood estimator. Due to the close relationship of this estimator with the least squares estimator, this lower bound is applicable for Gaussian conditions for the least squares estimators as well.

For a general situation where ARMA process parameters and ARMA noise parameters are estimated, expressions are given which give insight into the behaviour of the covariance for different parameters if different signal-to-noise ratios are considered. This typical behaviour shows a maximum value for the lower bounds for all parameters, except for the MA parameters of the process, whose variances continue to increase for decreasing signal-to-noise ratios.

CHAPTER FOUR:

RECURSIVE LEAST SQUARES ESTIMATORS

4.1 Introduction

In the previous chapter the basic idea of least squares estimation has been considered. No attention was paid to requirements and problems which arise when applying these techniques in practical situations, like, for example, the need for permanent availability of an updated estimation result for control purposes, based on the already processed data samples, and the necessary a priori knowledge of statistical properties of the disturbances for obtaining consistent estimates of the process parameters.

In this respect, recursive estimators, as discussed in this chapter, are more practically applicable. Some of them are mathematically equivalent to explicit estimators; others are approximate versions of explicit estimators, the approximation being made mostly for computational reasons. For example, the recursive variant of the weighted least squares estimator usually starts without knowledge of the noise parameters but during the estimation, estimates of noise parameters of increasing reliability become available for the necessary filtering. Thus, during the estimation process, this estimator shifts from a (non-consistent) common least squares to a (consistent) Markov type of weighted least squares estimator.

The study of these types of estimators, however, is rather complicated as usually both model parameters and noise parameters have to be estimated simultaneously. For reliable estimation of the noise parameters, the corresponding input-output signals of the noise filter would have to be available. But actually only approximated (=estimated) signals are available, based on previous estimates of the process parameters and previous estimates of the noise parameters.

Due to this complexity, it is rather difficult to derive the properties of these estimators for a finite number of samples. Therefore

in the literature the attention is focused on the investigation of the asymptotic convergence properties of such recursive estimators. For some estimators (which will be dealt with also in this chapter) the convergence cannot always be guaranteed; examples show that divergence may exist for such schemes. In such cases, the behaviour of the estimators for finite sample length has mostly been studied through practical examples and simulations.

This chapter is set out as follows. In paragraph 4.2 some general notions and definitions related to recursive estimators are given, whereas the general classification of recursive estimators, based on the three basic operations which we have already met in chapter 3 will be given in paragraph 4.3. In paragraph 4.4 we will discuss, in some detail, the different estimators which appear in the classification of paragraph 4.3. Finally, in paragraph 4.5, we will comment on the convergence properties of the estimators.

4.2 The concept of recursive estimation

In paragraph 3.2 we met the expression for the explicit estimator

$$\underline{\hat{\theta}} = [\Omega^T(u,y)\Omega(u,y)]^{-1}\Omega^T(u,y)\underline{y} \tag{4.1}$$

The corresponding estimate is based on N samples, and is calculated after these N samples have been acquired. When collecting these samples, such a type of explicit estimator gives no intermediate estimation results. We can construct a recursive estimator which, after the acquisition of the new measurements (u_k, y_k), updates the previous estimate $\underline{\hat{\theta}}_{k-1}$, with $k \leq N$. This implicit or recursive estimator yields results identical to the explicit form (4.1), if initiated appropriately; cf. Eykhoff (1974).

$$\left.\begin{aligned} \underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}(y_k - \underline{\omega}_k^T\underline{\hat{\theta}}_{k-1}) \\ P_k &= P_{k-1} - P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}\underline{\omega}_k^T P_{k-1} \end{aligned}\right] \tag{4.2}$$

where $\underline{\hat{\theta}}_k$ is the estimate after k samples and

$$P_{k-1} = [\Omega_{k-1}^T(u,y)\Omega_{k-1}(u,y)]^{-1} \tag{4.3}$$

with:

$$\Omega_{k-1}(u,y) = \left[\begin{array}{ccc|ccc} u_{q+1} & \cdots & u_{q+1-p} & y_q & \cdots & y_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ u_{k-1} & \cdots & u_{k-p-1} & y_{k-2} & \cdots & y_{k-q-1} \end{array}\right] \qquad (4.4)$$

$$\underline{\omega}_k^T = (u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q}) \qquad (4.5)$$

so that $\underline{\omega}_k^T$ is the last row of $\Omega_k(u,y)$.

If we consider the expression for the recursive estimator in more detail, we notice the term $y_k - \underline{\omega}_k^T \hat{\underline{\theta}}_{k-1}$ which is usually referred to as the prediction error. This prediction error is the difference between the next measurement y_k and its one-step-ahead prediction based on the model using the previous parameter estimates. The prediction error is used to evaluate the gradient direction for obtaining the next model setting. Dependent on the extent of the model used (leading to extended expressions for (4.4), cf. (4.9) and (4.10)), we have the choices $\hat{e}_k$ and $\hat{\xi}_k$.

$$\left.\begin{aligned} \hat{e}_k &= [1+\hat{A}^{k-1}]y_k - [\hat{b}_o^{k-1}+\hat{B}^{k-1}]u_k \\ \hat{\xi}_k &= [1+\hat{D}^{k-1}]\hat{e}_k - [\hat{C}^{k-1}]\hat{\xi}_k \end{aligned}\right] \qquad (4.6)$$

where $[1+\hat{A}^{k-1}]$, $[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$, $[1+\hat{C}^{k-1}]$ and $[1+\hat{D}^{k-1}]$ are the polynomials containing the estimated parameters after the $k-1^{st}$ recursion step.

We shall also need the residuals $\hat{\hat{e}}_k$ and $\hat{\hat{\xi}}_k$ which can be computed in the k-th recursion when the k-th estimate has become available:

$$\left.\begin{aligned} \hat{\hat{e}}_k &= [1+\hat{A}^{k}]y_k - [\hat{b}_o^{k}+\hat{B}^{k}]u_k \\ \hat{\hat{\xi}}_k &= [1+\hat{D}^{k}]\hat{\hat{e}}_k - [\hat{C}^{k}]\hat{\hat{\xi}}_k \end{aligned}\right] \qquad (4.7)$$

To complete the notions related to the errors involved, we will recall the equation error e_k as discussed in chapter 3:

$$\left.\begin{aligned} e_k &= [1+A]_t y_k - [b_o+B]_t u_k \\ \xi_k &= [1+D]_t e_k - [C]_t \xi_k \end{aligned}\right] \qquad (4.8)$$

Taking $\hat{e}_k$ as prediction error, i.e. modelling only the ARMA (process) part of the model, yields the result of eq. (4.2). Extension of the model with noise parameters yields the following:

$$\Omega_{k-1}(u,y,\hat{\xi},\hat{e}) =$$

$$= \left[\begin{array}{ccc|ccc|ccc|ccc} u_{q+1} & \cdots & u_{q+1-p} & y_q & \cdots & y_1 & \hat{\xi}_q & \cdots & \hat{\xi}_{q-s} & \hat{e}_q & \cdots & \hat{e}_{q-r} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ u_{k-1} & \cdots & u_{k-p-1} & y_{k-2} & \cdots & y_{k-q-1} & \hat{\xi}_{k-2} & \cdots & \hat{\xi}_{k-s-1} & \hat{e}_{k-2} & \cdots & \hat{e}_{k-r-1} \end{array}\right] \qquad (4.9)$$

$$\underline{\omega}_k^T = (u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q},\hat{\xi}_{k-1},\ldots,\hat{\xi}_{k-s},\hat{e}_{k-1},\ldots,\hat{e}_{k-r}) \qquad (4.10)$$

Applying a minimalization procedure to minimize in the k-th recursion, based on the k-1st model setting, the criterion

$$V_k = \tfrac{1}{2}\hat{\xi}_k^2 \qquad (4.11)$$

using

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} - Q_k \left.\frac{\partial V}{\partial \underline{\theta}}\right|_{\underline{\hat{\theta}}_{k-1}} \qquad (4.12)$$

we find

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} + Q_k \underline{\tilde{\omega}}_k \hat{\xi}_k \qquad (4.13)$$

where the gradient direction is:

$$\left.\frac{\partial V}{\partial \underline{\theta}}\right|_{\underline{\hat{\theta}}_{k-1}} = -\underline{\tilde{\omega}}_k \hat{\xi}_k \qquad (4.14)$$

with

$$\underline{\tilde{\omega}}_k^T = -\frac{\partial \hat{\xi}_k}{\partial \underline{\theta}^T} = (\tilde{u}_k,\ldots,\tilde{u}_{k-p},\tilde{y}_{k-1},\ldots,\tilde{y}_{k-q},\tilde{\hat{\xi}}_{k-1},\ldots,\tilde{\hat{\xi}}_{k-s},\tilde{\hat{e}}_{k-1},\ldots,\tilde{\hat{e}}_{k-r}) \qquad (4.15)$$

where

$$\left.\begin{aligned}
\tilde{u}_k &= \frac{[1+\hat{D}]}{[1+\hat{C}]}\, u_k \\
\tilde{y}_k &= \frac{[1+\hat{D}]}{[1+\hat{C}]}\, y_k \\
\tilde{\hat{e}}_k &= \frac{1}{[1+\hat{D}]}\, \hat{e}_k \\
\tilde{\hat{\xi}}_k &= \frac{1}{[1+\hat{C}]}\, \hat{\xi}_k
\end{aligned}\right] \qquad (4.16)$$

For the choice of Q_k, various possibilities have been proposed, ranging from a scalar quantity, as in the stochastic approximation schemes, to a matrix quantity meant for orthogonalization of the scheme such as the Newton-Raphson variants where Q_k can be interpreted as the (matrix) second derivative of the criterion function.

4.3 General classification of recursive estimators

In this paragraph we will classify the various recursive estimation schemes as presented in the literature. We will use the concept of three basic operations for this classification. We will present diagrams similar to those in the previous chapter, although the interpretation is within the context of the recursive character of these estimators.

A block representing the recursive estimator is used. It makes a recursion from step k-1 to step k using the necessary (filtered) input-ouput signals, the IV quantities, the signals gained from the model extension and the prediction error, cf. fig. 4.1. The prediction error quantity is denoted by a dot in the forthcoming figures.

The diagrams of the various estimation schemes consist, apart from the recursive estimator block as shown in fig. 4.1, of the following elements, cf. fig. 4.2:

- the model blocks M_1, M_2, M_3 and M_4 having the values of the previously estimated parameters as model setting. The model has, in its most extended version, the form of the model of Talmon and Van den Boom; cf. chapter 2

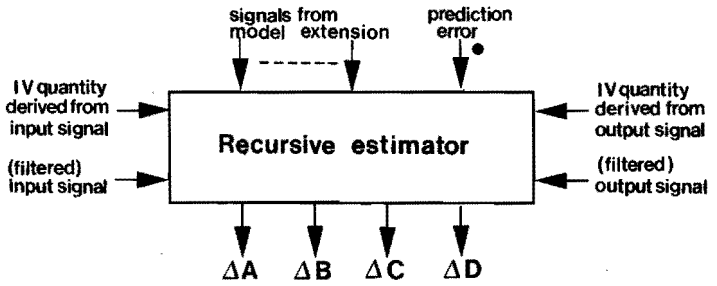


Fig. 4.1 Recursive estimator block

- the filters F_1 and F_2 for the additional noise whitening filtering as already seen in chapter 3. The parameters of these filters may be tuned according to any available a priori information of the noise colouring, but usually they will follow from recently, i.e. in previous recursion steps, estimated noise parameters, e.g. from the model extension part.
- the IV filters Φ_1 and Φ_2. They generate the IV quantity using the input- or the output signal or both, or using extra signals, which are additionally available. The possibility of using additional signals is not indicated in fig. 4.2.

From fig. 4.2 it can also quite easily be seen how the noise is handled in the estimator. In chapter 3 we have already seen that the correlation between the resulting error and the (shifted) input-output samples should be zero. This can be done in two ways:

a) operations which affect the noise colouring.
   Here we have two possibilities of making the resulting error white: a1) add extra filtering and/or a2) through the model extension resulting also in additional filtering. Depending on the choice of prediction error among ($\hat{\xi}_k^1$, $\hat{\xi}_k^2$, $\hat{\xi}_k^3$) the path for the equation error contained in y_k through F_1, M_2, M_3 and M_4 should be noise-whitening.
b) substitute the process signals by a related signal which contains no (or less) noise components. This is the IV approach and is realized by the filters Φ_1 and/or Φ_2.

The approaches listed under a) and b) may be combined, resulting in a family of estimators.

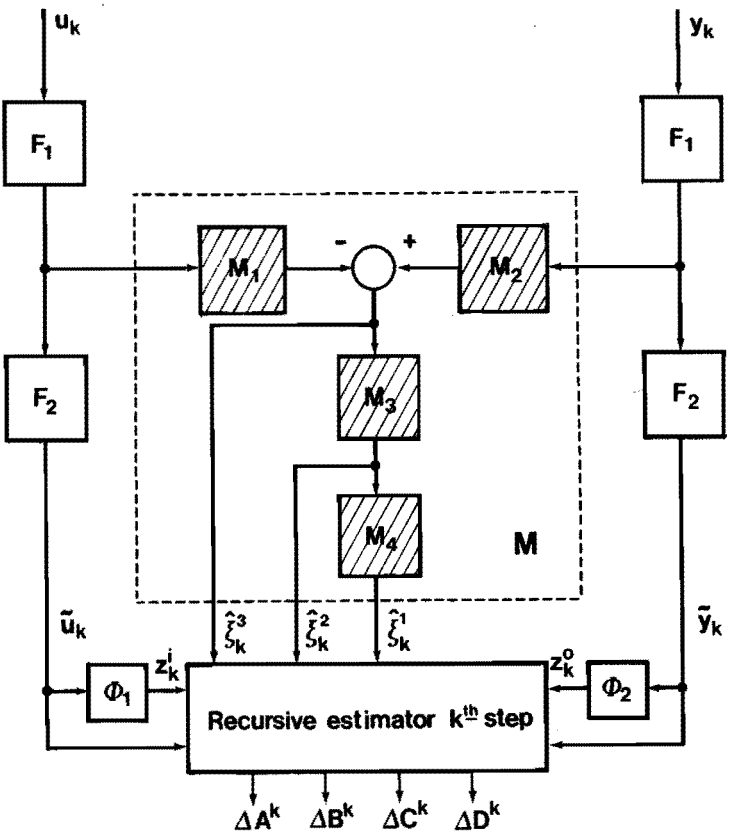


Fig. 4.2 The general recursive estimator

Fig. 4.3 shows the relationships between different recursive estimators. They will be discussed in more detail in the next paragraph. It follows from this diagram that estimators can be distinguished at four levels of complexity. The simplest estimator, LS, is very efficient with respect to computer time, but gives biased results, particularly for low S/N ratios. In cases where this is unwanted, the estimators of level 1 are of interest. They are more expensive with respect to computation time, but yield better results for low S/N ratios. Only in very specific cases, as will be shown in the sequel, will these estimators diverge. The estimators of level 2 and 3 are

	Method	Model			
		M_1	M_2	M_3	M_4
0	Least Squares	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-
1a	GLS	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-
1b	OLS	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-
1b	EMM a)	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	-
	b)	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	$[1+\hat{C}^{k-1}]^{-1}$
	c)	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$
1b	EECM	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$
1c	IV	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-
2ab	AML/IQL	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	$[1+\hat{C}^{k-1}]^{-1}$
2ab	Appr. Markov	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$
2ac	Subopt. IV	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-
2ac	IV-AML IV:	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-
	AML:			$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$
2bc	IVEMM	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$
3abc	general	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$

Table 4.1 Choice of Model, Signal Filters and Instrumental Variable

Signal filters		Instr. Variable Filters		prediction error	remarks
F_1	F_2	Φ_1	Φ_2		
-	-	-	-	$\hat{\xi}^3_k$	
$[1+\hat{D}^{k-1}]$	-	-	-	$\hat{\xi}^3_k$	separate est. for $[1+\hat{D}]$
-	-	-	-	$\hat{\xi}^3_k$	
-	-	-	-	$\hat{\xi}^2_k$	common factors in M_1, M_2
-	-	-	-	$\hat{\xi}^1_k$	$\xi^2_k = \xi^3_k$
-	-	-	-	$\hat{\xi}^1_k$	
-	-	-	-	$\hat{\xi}^3_k$	separate est. for $[1+\hat{D}]$ and $[1+\hat{C}]$
-	-	a) fixed	-	$\hat{\xi}^3_k$	
		b) $\dfrac{[\hat{b}_o^{k-1}+B^{k-1}]}{[1+\hat{A}^{k-1}]}$	-	$\hat{\xi}^3_k$	
		c) delay	-	$\hat{\xi}^3_k$	
		d) -	delay	$\hat{\xi}^3_k$	
		e) delay	delay	$\hat{\xi}^3_k$	
$[1+\hat{C}^{k-1}]^{-1}$	-	-	-	$\hat{\xi}^3_k$	
$[1+\hat{C}^{k-1}]^{-1}$	$[1+\hat{D}^{k-1}]$	-	-	$\hat{\xi}^2_k$	
$[1+\hat{D}^{k-1}]$	-	$\dfrac{[\hat{b}_o^{k-1}+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$	-	$\hat{\xi}^3_k$	separate est. for $[1+\hat{D}]$
$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$ $[1+\hat{C}^{k-1}]^{-1}$	$\dfrac{[\hat{b}_o^{k-1}+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$	-	$\hat{\xi}^3_k$	
-	-	see choices a-e under 1c		$\hat{\xi}^1_k$	
$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	see choices a-e under 1c		$\hat{\xi}^1_k$	

Filters for different Estimators; see also fig. 4.2

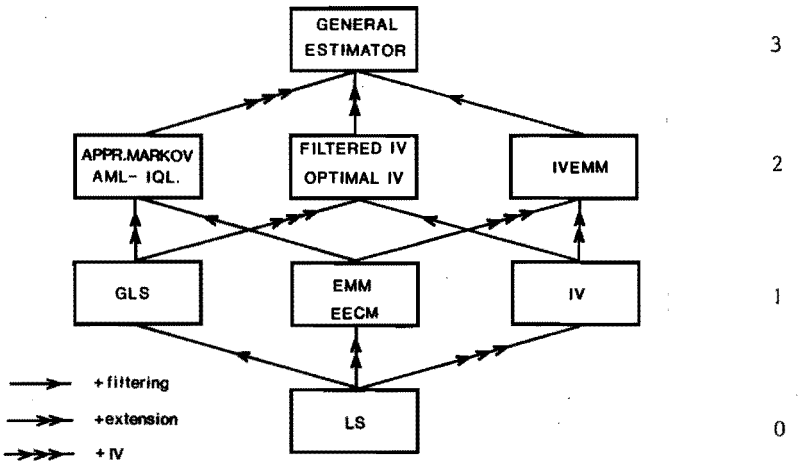


Fig. 4.3 Relation between recursive estimators

even more complex and time consuming. With respect to noise sensitivity in relation to consistence, these estimators are superior to the estimators of level 0 and 1. Those estimators are not easy to apply as they have to be started by estimators of class 1 which, in their turn, have to be started by the LS estimator, except some variants of the IV estimator.

In the next paragraph we will show that all these estimators fit very well into the general scheme of figure 4.2. It can be observed that some schemes have a bootstrap nature, as they use separate estimators for obtaining the necessary information for the filtering, such as GLS and suboptimal IV, or for the necessary compensation as in EECM. These extra estimators are usually simple LS estimators for obtaining the AR or ARMA parameters of the model of the noise.

In table 4.1 a summary of the important aspects of the different recursive estimators is given. Here the particular choices for the model parts M_1, M_2, M_3 and M_4, the filters F_1 and F_2 and the IV filters Φ_1 and Φ_2 are indicated.

4.4 Details of the different recursive estimators

In this paragraph we shall discuss, in more detail, the recursive estimators which have been summarized already in table 4.1. We will consider the schemes in order of increasing complexity.

4.4.1 The recursive least squares estimator

This estimator is the recursive form of the LS estimator (3.10). The expressions are:

$$\left.\begin{aligned} \underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}\hat{e}_k \\ P_k &= P_{k-1} - P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}\underline{\omega}_k^T P_{k-1} \\ \hat{e}_k &= y_k - \underline{\omega}_k^T \underline{\hat{\theta}}_{k-1} \end{aligned}\right] \qquad (4.17)$$

where

$$\underline{\hat{\theta}}_k^T = (\hat{b}_o,\ldots,\hat{b}_p,-\hat{a}_1,\ldots,-\hat{a}_q)_k \qquad (4.18)$$

$$\underline{\omega}_k^T = (u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q}) \qquad (4.19)$$

The expression is mathematically equivalent to the implicit estimator (3.10) if initiated properly; cf. Eykhoff (1974). The diagram of this estimator is shown in fig. 4.4

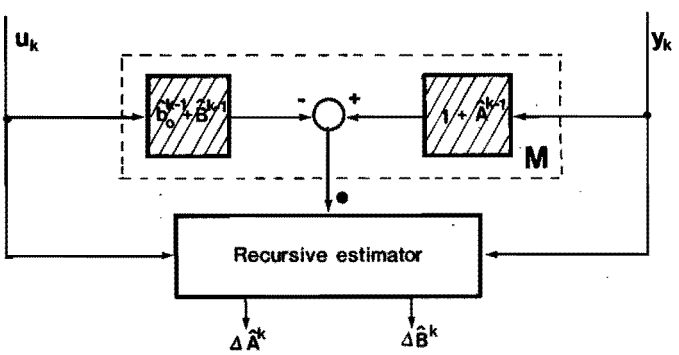


Fig. 4.4 The recursive least squares estimator

4.4.2 The Generalized Least Squares estimator (GLS)

Based on the Markov type of weighting, a recursive algorithm has been proposed by Hastings-James and Sage (1969) where filtering of the measurables is performed by a MA filter consisting of the AR parameters of the noise:

$$[1+A]y_k = [b_o+B]u_k + \frac{1}{[1+D]_t}\hat{\xi}_k \tag{4.20}$$

$$[1+A]\tilde{y}_k = [b_o+B]\tilde{u}_k + \hat{\xi}_k \tag{4.21}$$

where

$$\left.\begin{aligned}\tilde{y}_k &= [1+D]_t y_k \\ \tilde{u}_k &= [1+D]_t u_k\end{aligned}\right] \tag{4.22}$$

The estimator for the ARMA process parameters is then:

$$\left.\begin{aligned}\underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\tilde{\omega}}_k(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{\omega}}_k)^{-1}\hat{\xi}_k \\ P_k &= P_{k-1} - P_{k-1}\underline{\tilde{\omega}}_k(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{\omega}}_k)^{-1}\underline{\tilde{\omega}}_k^T P_{k-1} \\ \hat{\xi}_k &= \tilde{y}_k - \underline{\tilde{\omega}}_k^T\underline{\hat{\theta}}_{k-1}\end{aligned}\right] \tag{4.23}$$

where

$$\underline{\tilde{\omega}}_k^T = (\tilde{u}_k,\ldots,\tilde{u}_{k-p},\tilde{y}_{k-1},\ldots,\tilde{y}_{k-q}) \tag{4.24}$$

Usually, as the knowledge of $[1+D]_t$ is lacking, we will use a separate scheme to estimate these noise parameters. In the k-th recursion, we generate the residual, using the newly available estimate $\underline{\hat{\theta}}_k$

$$\hat{e}_k = y_k - \underline{\omega}_k^T\underline{\hat{\theta}}_k \tag{4.25}$$

This leads to the estimator for the AR noise parameters:

$$\left.\begin{aligned}\underline{\hat{d}}_k &= \underline{\hat{d}}_{k-1} - Q_{k-1}\underline{\hat{e}}_k(1+\underline{\hat{e}}_k^T Q_{k-1}\underline{\hat{e}}_k)^{-1}\hat{\xi}_k \\ Q_k &= Q_{k-1} - Q_{k-1}\underline{\hat{e}}_k(1+\underline{\hat{e}}_k^T Q_{k-1}\underline{\hat{e}}_k)^{-1}\underline{\hat{e}}_k^T Q_{k-1} \\ \hat{\xi}_k &= \hat{e}_k - \underline{\hat{e}}_{k-1}^T\underline{\hat{d}}_{k-1}\end{aligned}\right] \tag{4.26}$$

In figure 4.5 the schematic diagram for this bootstrap estimator is given.

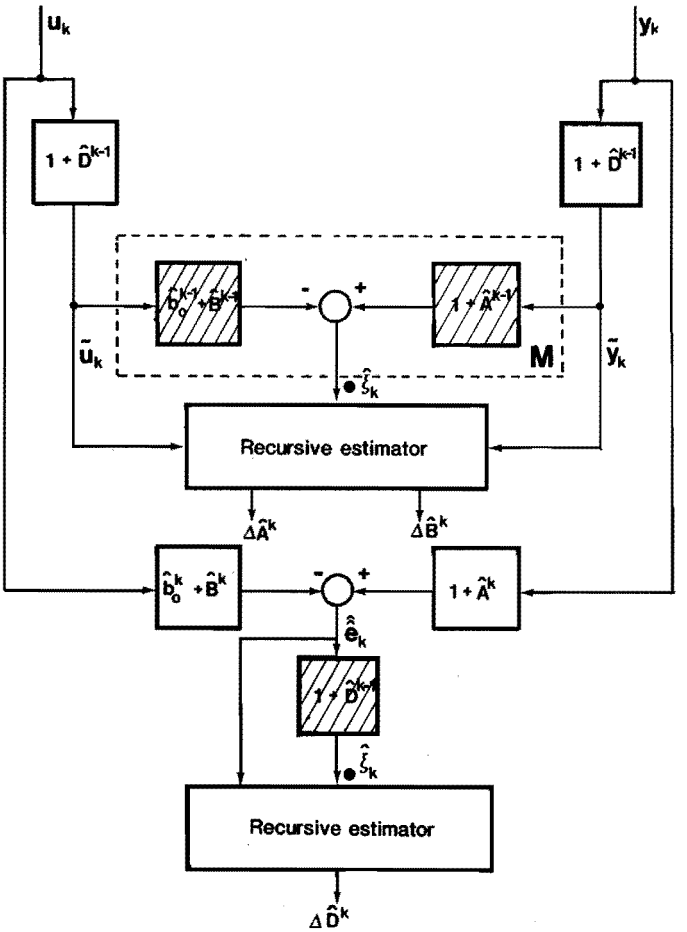


Fig. 4.5 The GLS estimator

4.4.3 Over-parametrized Least Squares estimator (OLS)

If we model the noise with an AR model, and if we estimate ARMA process parameters with an over-parametrized model, we have:

$$[1+A']y_k = [b_o'+B']u_k + \hat{\xi}_k \qquad (4.27)$$

where

$$\left.\begin{aligned}[1+A'] &= [1+A][1+D] \\ [b_o'+B'] &= [b_o+B][1+D]\end{aligned}\right\} \qquad (4.28)$$

If we make the model order sufficiently large, we can attain a white prediction error so that consistent estimation may be performed. The problem is then to calculate the "common" factors in the polynomials $[\hat{b}_o'+\hat{B}']$ and $[1+\hat{A}']$ as, due to the disturbing noise, these factors are not exactly equal; cf. Söderström (1975) and Stoica (1976). Hsia (1975) proposes a method, the so-called Multi Stage Least Squares Estimator (MSLS), where in the first stage the polynomials $[1+\hat{A}']$ and $[\hat{b}_o'+\hat{B}']$ are estimated, in the second stage the parameters $[1+\hat{A}]$ and $[\hat{b}_o+\hat{B}]$ are estimated from the the estimated over-parametrized model $[\hat{b}_o'+\hat{B}']$ and $[1+\hat{A}']$, and in the third stage the parameters $[1+\hat{D}]$ from the parameters found in the two previous stages. Another approach is to use an over-parametrized MA model for the process. The estimated MA parameters of such a model are unbiased if we take a model of sufficient length; cf. Van den Boom and Melis (1968). Hsia (1981) gives a method to derive the ARMA parameters from these estimated MA parameters. Also the Ho-Kalman algorithm can be used to find the ARMA realization from the estimated MA parameters, as these are equivalent to the Markov parameters; cf. Hajdasinski (1980). For these methods we can give the following schemes:

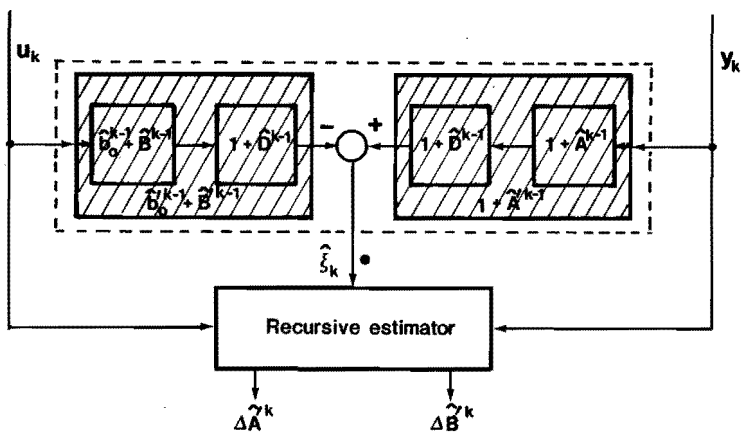


Fig. 4.6a Over-parametrized Least Squares

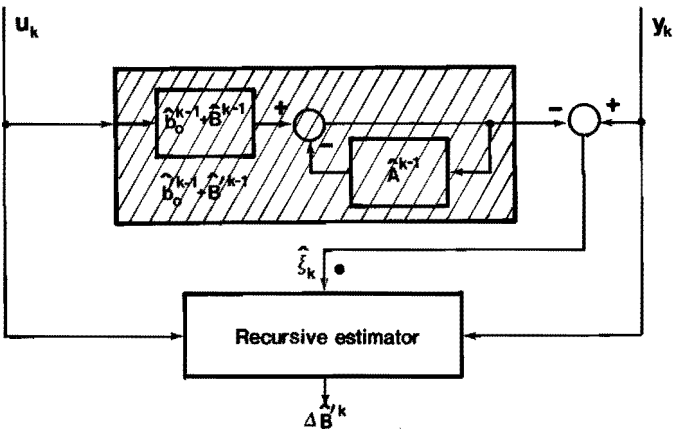


Fig. 4.6b Over-parametrized Least Squares

4.4.4 The Extended Matrix Method (EMM)

Another way of obtaining white residuals using the concept of model extension, is the extended matrix method EMM, where a single stage estimator is used for the estimation of process and noise parameters together. The vector of parameters $\underline{\theta}$ and the vector of measurables $\underline{\omega}$ is extended to contain the noise quantities as well. The method was originally proposed by Panuska (1969) within the context of Stochastic Approximation (SA) for the estimation of the A, B, D parameters, and has been generalized by Talmon (1971) and Talmon and Van den Boom (1973) for the more general case of A, B, C, D parameters and within the context of prediction error estimators. The algorithm is:

$$\left.\begin{aligned} \underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}\,\hat{\xi}_k \\ P_k &= P_{k-1} - P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}\underline{\omega}_k^T P_{k-1} \\ \hat{\xi}_k &= y_k - \underline{\omega}_k^T\underline{\hat{\theta}}_{k-1} \end{aligned}\right\} \qquad (4.29)$$

where

$$\left.\begin{aligned}\underline{\theta}_k^T &= (\hat{b}_o,\ldots,\hat{b}_p,-\hat{a}_1,\ldots,-\hat{a}_q,\hat{c}_1,\ldots\hat{c}_s,-\hat{d}_1,\ldots-\hat{d}_r)_k \\ \underline{\omega}_k^T &= (u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q},\hat{\xi}_{k-1},\ldots,\hat{\xi}_{k-s},\hat{e}_{k-1},\ldots,\hat{e}_{k-r})\end{aligned}\right] \tag{4.30}$$

The noise signals $\hat{e}$ and $\hat{\xi}$ can be obtained using prior estimated parameters:

$$\hat{e}_k = y_k-(u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q}) \begin{bmatrix} \hat{b}_o \\ \vdots \\ \hat{b}_p \\ -\hat{a}_1 \\ \vdots \\ -\hat{a}_q \end{bmatrix}_{k-1} \tag{4.31}$$

$$\hat{\xi}_k = \hat{e}_k- (\hat{\xi}_{k-1},\ldots,\hat{\xi}_{k-s},\hat{e}_{k-1},\ldots,\hat{e}_{k-r}) \begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_s \\ -\hat{d}_1 \\ \vdots \\ -\hat{d}_r \end{bmatrix}_{k-1} \tag{4.32}$$

The schematic diagram of this estimator is given in fig. 4.7.

Indeed, it would be possible to replace, at each recursion step, all previously generated estimates $\hat{e}_i$ and $\hat{\xi}_i$, i = 1, ..., k-1 by new updates based on the most recent parameter estimates. The quality of the sequences $\hat{e}_i$ and $\hat{\xi}_i$ would improve in such a case, but this would give rise to an enormous increase in computational effort. The schemes that have been used successfully in the past use the samples $\hat{e}_k$ and $\hat{\xi}_k$ (which were generated as prediction error in the recursion from k-1 to k) as samples of measurables in

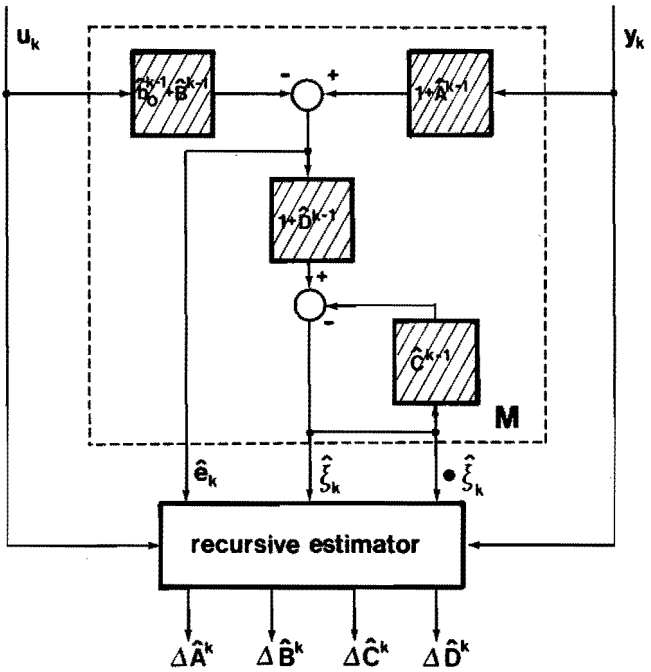


Fig. 4.7 The Extended Matrix Estimator (EMM)

the $\underline{\omega}$ vectors of the succeeding recursions. The starting procedure of this estimator is then as follows:

a. First apply common least squares to a limited number of samples N_o and generate $\hat{e}_i$, $i=1,\ldots,N_o$, where $N_o > p+q+1$.
b. Extend the parameter vector with one parameter $\hat{d}_i$, $i=1,\ldots,r$ per recursion step and generate $\hat{\xi}_i$ until all r AR noise parameters are estimated.
c. Extend the parameter vector with one parameter $\hat{c}_i$, $i=1,\ldots,$ s per recursion step until all s MA noise parameters are estimated.
d. Continue recursion with all $p+q+r+s+1$ parameters.

This EMM method proved to be a reliable method for the majority of cases, like the GLS method, although in exceptional cases divergence has been reported. This will be dealt with in section 4.5.

In the above given recursion step, two calculations take place: calculation of the prediction error and calculation of the new parameter estimate. With this new estimate the residual of this recursion step can be calculated and the residuals of the successive recursion steps can be transferred to the vector of measurables - instead of the prediction errors. There is a small increase of computational burden. In this case we have

$$\left.\begin{aligned} \hat{\underline{\theta}}_k &= \hat{\underline{\theta}}_{k-1} + P_{k-1}\hat{\underline{\omega}}_k(1+\hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k)^{-1}\,\hat{\xi}_k \\ P_k &= P_{k-1} - P_{k-1}\hat{\underline{\omega}}_k(1+\hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k)^{-1}\hat{\underline{\omega}}_k^T P_{k-1} \end{aligned}\right] \qquad (4.33)$$

with

$$\left.\begin{aligned} \hat{\underline{\omega}}_k^T &= (u_k,\cdots,u_{k-p},y_{k-1},\cdots,y_{k-q},\hat{\xi}_{k-1},\cdots,\hat{\xi}_{k-s},\hat{e}_{k-1},\cdots,\hat{e}_{k-r}) \\ \hat{\xi}_k &= y_k - \hat{\underline{\omega}}_k^T\hat{\underline{\theta}}_{k-1} \end{aligned}\right] \qquad (4.34)$$

and for the residual we find:

$$\hat{\hat{e}}_k = y_k - (u_k,\cdots,u_{k-p},y_{k-1},\cdots,y_{k-q})\begin{bmatrix}\hat{\underline{b}}\\ -\hat{\underline{a}}\end{bmatrix}_k \qquad (4.35)$$

$$\begin{aligned} \hat{\hat{\xi}}_k &= \hat{\hat{e}}_k - (\hat{\xi}_{k-1},\cdots,\hat{\xi}_{k-s},\hat{\hat{e}}_{k-1},\cdots,\hat{\hat{e}}_{k-r})\begin{bmatrix}\hat{\underline{c}}\\ -\hat{\underline{d}}\end{bmatrix}_k = \\ &= y_k - \hat{\underline{\omega}}_k^T\hat{\underline{\theta}}_k = y_k - \hat{\underline{\omega}}_k^T[\hat{\underline{\theta}}_{k-1} + P_{k-1}\hat{\underline{\omega}}_k(1+\hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k)^{-1}\hat{\xi}_k] = \\ &= \hat{\xi}_k - \hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k(1+\hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k)^{-1}\hat{\xi}_k = (1+\hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k)^{-1}\hat{\xi}_k \end{aligned} \qquad (4.36)$$

This formula shows that a very efficient calculation of the residual from the prediction error can be performed as the scalar

$$(1+\hat{\underline{\omega}}_k^T P_{k-1}\hat{\underline{\omega}}_k)^{-1}$$

is available during the recursion. If only MA noise parameters have to be estimated, this is a minor increase in computational burden. If AR noise parameters are estimated, eq. (4.35) has to be performed during each recursion step. This modification of the EMM algorithm is of interest, as its convergence can be proved, without the need of monitoring the estimates, i.e. checking the stability of the estimated model; cf. Solo (1979).

4.4.5 The Equation Error Compensation Method (EECM)

Another possibility of model extension is with the use of separate estimators for process and noise parameters. This scheme was proposed by Talmon (1971), who reported good results with this estimator. The estimated parameters of the recursion block for the noise parameters are used for compensation of the prediction error $\hat{e}_k$ in the process parameter block:

$$[1+A]y_k = [b_o+B]u_k + \hat{\hat{e}}_k \tag{4.37}$$

Now we have for the k^{th} recursion:

step a) first an estimate of the process (PR) parameters is made:

$$\left.\begin{aligned} \underline{\hat{\theta}}_k^{PR} &= \underline{\hat{\theta}}_k^{PR} + P_{k-1}\underline{\omega}_k^{PR}(1+\underline{\omega}_k^{PR^T}P_{k-1}\underline{\omega}_k^{PR})^{-1}\hat{\xi}_k^* \\ P_k &= P_{k-1} - P_{k-1}\underline{\omega}_k^{PR}(1+\underline{\omega}_k^{PR^T}P_{k-1}\underline{\omega}_k^{PR})^{-1}\underline{\omega}_k^{PR^T}P_{k-1} \end{aligned}\right] \tag{4.38}$$

where

$$\left.\begin{aligned} \underline{\omega}_k^{PR^T} &= (u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q}) \\ \underline{\hat{\theta}}_k^{PR^T} &= (\hat{b}_o,\ldots,\hat{b}_p,-\hat{a}_1,\ldots,-\hat{a}_q)_k \end{aligned}\right] \tag{4.39}$$

$$\hat{\xi}_k^* = y_k - \underline{\omega}_k^{PR^T}\underline{\hat{\theta}}_{k-1}^{PR} - \hat{e}_k^* \tag{4.40}$$

where $\hat{e}_k^*$ is the compensation based on k-1 samples resulting from the k-1-th iteration. See also step c.

step b) using $\underline{\hat{\theta}}_k^{PR}$ construct:

$$\hat{\hat{e}}_k = y_k - \underline{\omega}_k^{PR^T}\underline{\hat{\theta}}_k^{PR} \tag{4.41}$$

with this $\hat{\hat{e}}_k$ and with $\hat{\hat{e}}$ and $\hat{\hat{\xi}}$ obtained from previous recursion, an estimate of the noise (N) parameters $\underline{\hat{\theta}}_k^N$ can be made:

$$\left.\begin{aligned} \underline{\hat{\theta}}_k^N &= \underline{\hat{\theta}}_{k-1}^N + Q_{k-1}\underline{\omega}_k^{N^T}(1+\underline{\omega}_k^{N^T}Q_{k-1}\underline{\omega}_k^N)^{-1}\hat{\xi}_k \\ Q_k &= Q_{k-1} - Q_{k-1}\underline{\omega}_k^{N^T}(1+\underline{\omega}_k^{N^T}Q_{k-1}\underline{\omega}_k^N)^{-1}\underline{\omega}_k^{N^T}Q_{k-1} \end{aligned}\right] \tag{4.42}$$

where

$$\left.\begin{aligned} \underline{\omega}_k^{N^T} &= (\hat{\xi}_{k-1},\ldots,\hat{\xi}_{k-s},\hat{\hat{e}}_{k-1},\ldots,\hat{\hat{e}}_{k-r}) \\ \underline{\hat{\theta}}_k^{N^T} &= (\hat{c}_1,\ldots,\hat{c}_s,-\hat{d}_1,\ldots,-\hat{d}_r)_k \end{aligned}\right] \qquad (4.43)$$

$$\hat{\xi}_k = \hat{e}_k - \underline{\omega}_k^{N^T}\,\underline{\hat{\theta}}_{k-1}^{N} \qquad (4.44)$$

In this recursion we use the residual $\hat{\hat{e}}_k$, as $\underline{\hat{\theta}}_k^{PR}$ is available, but we have to use the prediction error $\hat{\xi}_k$, because $\underline{\hat{\theta}}_k^N$ is available after execution of eq. (4.42).

Here $\hat{\xi}_k$ is the residual, as explained in the previous paragraph. The use of the previous residual $\hat{\hat{\xi}}_{k-1}$ instead of the previous prediction error $\hat{\xi}_{k-1}$ is not strictly necessary but it can be obtained very easily by a scalar multiplication, see the previous paragraph. In the diagram we will denote the generation of the residual $\hat{\hat{\xi}}_k$ from the prediction error $\hat{\xi}_k$ by a block R.

step c) Construct (for use in the next recursion)

$$\hat{\hat{\xi}}_k = \hat{\hat{e}}_k - \underline{\omega}_k^{N^T}\,\underline{\hat{\theta}}_k^{N} \qquad (4.45)$$

i.e. the prediction error $\hat{\xi}_k$ is here replaced by the residual $\hat{\hat{\xi}}_k$ for the next recursion. Using

$$\underline{\omega}_{k+1}^{N^T} = (\hat{\xi}_k,\ldots,\hat{\xi}_{k+1-s},\hat{\hat{e}}_k,\ldots,\hat{\hat{e}}_{k+1-r}) \qquad (4.46)$$

a prediction $\hat{e}^*_{k+1}$ for the next recursion is made:

$$\hat{e}^*_{k+1} = \underline{\omega}_{k+1}^{N^T}\,\underline{\hat{\theta}}_k^{N} \qquad (4.47)$$

The diagram of this estimator is given in fig. 4.8.

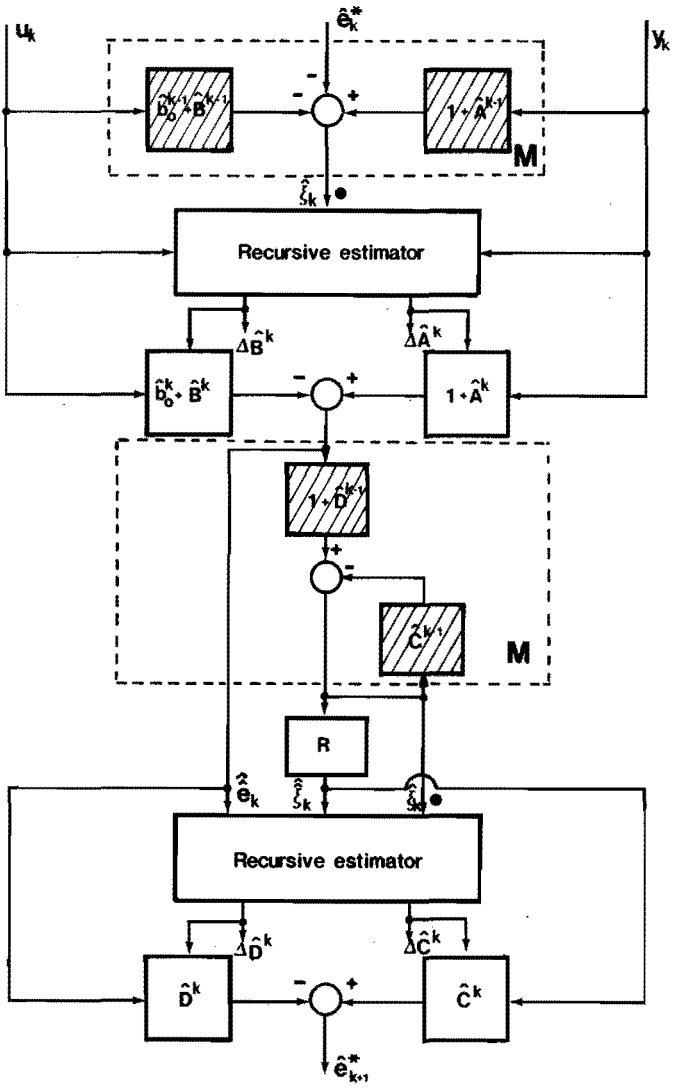


Fig. 4.8 The Equation Error Compensation Method

4.4.6 The Instrumental Variable Method (IV)

In chapter 3 we introduced the basic concepts of the instrumental variable estimators. We reviewed some important (practical) instrumental variable estimators, which were non-recursive. In this paragraph we will extend this list of instrumental variables by including the recursive variants.

By writing the explicit expression of the IV estimator in a recursive way, we obtain:

$$\left.\begin{aligned} \hat{\underline{\theta}}_k &= \hat{\underline{\theta}}_{k-1} + P_{k-1}\underline{z}_k(1+\underline{\omega}_k^T P_{k-1}\underline{z}_k)^{-1}\,(y_k - \underline{\omega}_k^T\hat{\underline{\theta}}_{k-1}) \\ P_k &= P_{k-1} - P_{k-1}\underline{z}_k(1+\underline{\omega}_k^T P_{k-1}\underline{z}_k)^{-1}\,\underline{\omega}_k^T P_{k-1} \end{aligned}\right] \qquad (4.48)$$

In this expression P_{k-1} is a non-symmetrical matrix. The vector $\underline{\omega}_k$ contains the measurables $\{u_k\}$ and $\{y_k\}$, as with the least squares estimators, and the vector $\underline{z}_k$ contains the IV quantity.

Because of the recursive character of the estimator, we can generate the IV quantity along with the recursion, so that estimation results can be used for this purpose.

Due to the non-symmetrical form of the matrices P, and the presence of an additional signal $\underline{z}_k$, we have a slightly modified recursion block, which will be denoted "IV recursion". The schematic diagram for this IV estimator is given in figure 4.9.

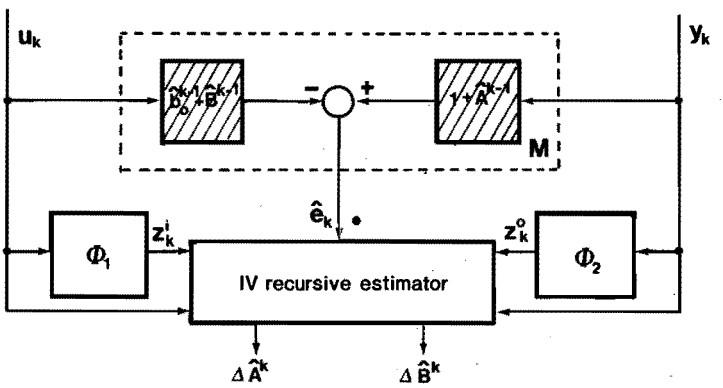


Fig. 4.9 The Instrumental Variable Estimator (IV)

Possible choices for simple IV estimators are:

	Φ_1	Φ_2	Literature
a)	fixed filter within bandwidth of process	-	Young (1968) Finnigan and Rowe (1973)
b)	model of process	-	Young (1965) Wong (1966) Smets (1970)
c)	delay - delay	- delay delay	Wouters (1972) Gersch (1970) Stoica + Söderström (1979)
d)	template function	-	Eykhoff (1980)

Finnigan and Rowe (1973) showed that the use of a fixed stable filter provides IV quantities satisfying the consistency conditions. Pandya (1972) and Pandya and Pagurek (1973) proposed a bootstrap estimator where separate estimators are used for the estimation of the parameters and for the estimation of the incidental parameters (i.e. the clean output signal or the noise signal). Their results are to be compared with those of Wong (1966) and Young (1965) where models of the process are used for the estimation of the clean outputs.

For a particular realization of the data, these instrumental variable estimators can show instabilities. This is caused, when (usually slightly) unstable estimates are delivered. These estimates are used for the generation of the IV quantity, being therefore also unstable. To avoid this, a Jury stability test was used by Wong (1966) and Young (1965). Furthermore, delayed estimates $\hat{\underline{\theta}}_{k-\tau}$ were used for the generation of the IV quantity. Smets (1970) could not notice that a removal of the delay caused problems.

When setting up a recursive scheme, it is always a difficult task to decide whether newly available estimates have to be used to improve

all old quantities which were generated by previous estimates. Also in the case of generation of an instrumental variable using an adaptive model, recalculations of previous samples can be done when a new estimate becomes available. Usually, this decision is a trade-off between accuracy and speed of calculation. Smets (1970) uses the newly up-dated estimate for calculation of only the next IV sample. Sometimes it is suggested to recalculate as many previous samples of the IV quantity as parameters to be estimated. When an ARMA type of model is used, the extra amount of recalculations can be reasonable. With such an ARMA model previous u_k and z_k are used to generate z_{k+1}. As the previous z_k all depend on previous u_k, all samples z depend on the input signal u. Rowe (1970) calls this "feed forward information" leading to the calculation of an instrumental variable. He suggests a closer connection of the IV quantity with the output quantity and uses a few previous output samples to start the ARMA filtering calculations for each recursion step. In principle there is little difference between this approach and the use of input signals only.

4.4.7 The Approximate Maximum Likelihood estimator and the Implicit Quasi Linearization scheme (AML/IQL)

The (IQL) methods, which are based on linearization of the loss function V, have also been presented in recursive form. Fuhrt (1973a, 1973b), who modelled the disturbing noise by MA parameters, used the same matrix lemma as Eykhoff (1974) to derive the recursive estimator from the one-shot expression as given in eq. (3.67):

$$\left.\begin{aligned} \underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\tilde{\omega}}_k(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{\omega}}_k)^{-1}(\tilde{w}_k - \underline{\tilde{\omega}}_k^T\underline{\hat{\theta}}_{k-1}) \\ P_k &= P_{k-1} - P_{k-1}\underline{\tilde{\omega}}_k(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{\omega}}_k)^{-1}\underline{\tilde{\omega}}_k^T P_{k-1} \end{aligned}\right] \qquad (4.49)$$

where

$$\underline{\tilde{\omega}}_k^T = (\tilde{u}_k, \ldots, \tilde{u}_{k-p-1}, \tilde{y}_{k-1}, \ldots, \tilde{y}_{k-q}, \tilde{\hat{\xi}}_{k-1}, \ldots, \tilde{\hat{\xi}}_{k-s}) \qquad (4.50)$$

This vector is obtained by filtering $\underline{\omega}_k$ as will be indicated soon.

$$\tilde{w}_k = \tilde{y}_k - \tilde{\xi}_k + \tilde{\hat{\xi}}_k \qquad (4.51)$$

The recursive procedure is as follows:

1) calculate $\{\hat{\xi}_k, \hat{\xi}_{k-1}, \ldots, \hat{\xi}_{k-s}\}$ using the previous estimate $\underline{\hat{\theta}}_{k-1}$:

$$\hat{\xi}_i = y_i - \underline{\omega}_i^T \underline{\hat{\theta}}_{k-1} \qquad i = k-s, \ldots, k \tag{4.52}$$

2) filter the signals $\{y_k, \ldots, y_{k-q}\}, \{u_k, \ldots, u_{k-p}\}$ and $\{\hat{\xi}_k, \ldots\ldots, \hat{\xi}_{k-s}\}$ by the filter $[1+\hat{C}^{k-1}]^{-1}$.

This filter is the most recent estimate of the MA noise parameters.

3) construct $\tilde{\underline{\omega}}_k$ and $\tilde{w}_k$
4) use the recursion formula (4.49)

In step 1 many previous samples of the "prediction" error have to be recalculated. Further, many previous samples of the three signals, u, y and ξ are filtered in step 2. Moreover, in step three a compensation action takes place, leading to enlarged matrices to be handled. For these reasons this method is rather time consuming.

The schematic diagram of this estimator is shown in fig. 4.10. Here it is depicted according to the original formulation of the estimator by Fuhrt. By shifting some of the filters $[1+\hat{C}^{k-1}]^{-1}$ and combining the results of the two summation points we can rearrange this diagram drastically. The prediction error in (4.49) can be rewritten (and therefore be simplified):

$$\tilde{w}_k - \tilde{\underline{\omega}}_k^T \underline{\hat{\theta}}_{k-1} = \tilde{y}_k - \tilde{\hat{\xi}}_k + \hat{\xi}_k - \tilde{\underline{\omega}}_k^T \underline{\hat{\theta}}_{k-1} = \tilde{y}_k - \tilde{y}_k + \tilde{\underline{\omega}}_k^T \underline{\hat{\theta}}_{k-1} + \hat{\xi}_k - \tilde{\underline{\omega}}_k^T \underline{\hat{\theta}}_{k-1} = \hat{\xi}_k \tag{4.53}$$

This leads to the simplified diagram of fig. 4.11.

Söderström (1973) used also a Taylor expansion of the loss function and a Newton-Raphson technique. The minimization of the loss function

$$V_k(\underline{\theta}) = \frac{1}{k} \sum_{i=1}^{k} \hat{\xi}_i^2(\underline{\theta}) \tag{4.54}$$

is considered. A relation between $V_k(\underline{\theta})$ and $V_{k-1}(\underline{\theta})$ can be given:

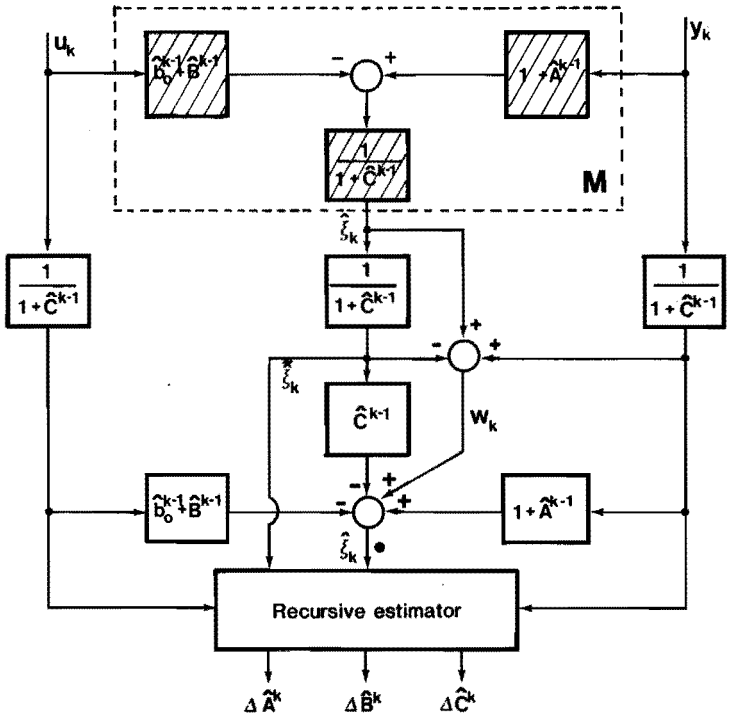


Fig. 4.10 The Implicit Quasi-Linearization Estimator (IQL)

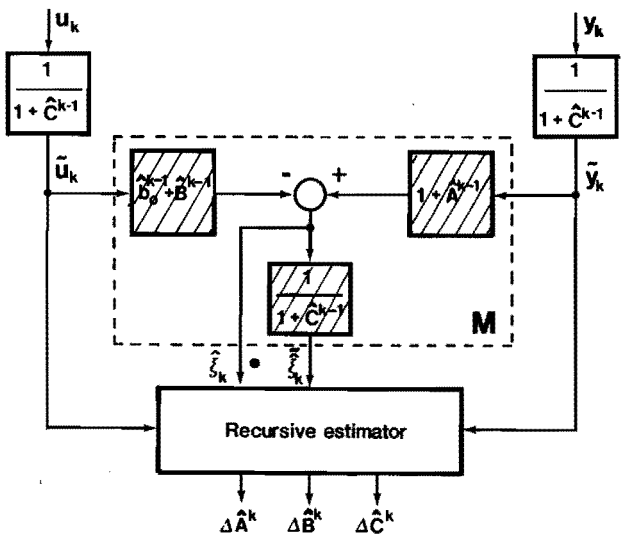


Fig. 4.11 The IQL/AML estimator (rearranged scheme)

$$V_k(\underline{\theta}) = \frac{k-1}{k} V_{k-1}(\underline{\theta}) + \frac{1}{k} \hat{\xi}_k^2(\underline{\theta}) \tag{4.55}$$

Take the derivatives of this expression with respect to $\underline{\theta}$:

$$V_k'(\underline{\theta}) = \frac{k-1}{k} V_{k-1}'(\underline{\theta}) + \frac{2}{k} \hat{\underline{\xi}}_k'(\underline{\theta}) \hat{\xi}_k(\underline{\theta}) \tag{4.56}$$

$$V_k''(\underline{\theta}) = \frac{k-1}{k} V_{k-1}''(\underline{\theta}) + \frac{2}{k} \hat{\underline{\xi}}_k'(\underline{\theta}) \hat{\underline{\xi}}_k'^{T}(\underline{\theta}) + \frac{2}{k} \hat{\xi}_k''(\underline{\theta}) \hat{\xi}_k(\underline{\theta}) \tag{4.57}$$

As approximation we take in the point $\hat{\underline{\theta}}_{k-1}$ which minimized $V_{k-1}(\underline{\theta})$:

$$\left.\begin{aligned} &\underline{V}_{k-1}'(\hat{\underline{\theta}}_{k-1}) = \underline{0} \qquad \text{(due to the minimum)} \\ &\hat{\xi}_k(\hat{\underline{\theta}}_{k-1})\, \hat{\xi}_k''(\hat{\underline{\theta}}_{k-1}) = \emptyset \\ &V_{k-1}''(\hat{\underline{\theta}}_{k-1}) = V_{k-1}''(\hat{\underline{\theta}}_{k-2}) \quad (\hat{\underline{\theta}}_{k-2} \text{ close to } \hat{\underline{\theta}}_{k-1}) \end{aligned}\right\} \tag{4.58}$$

If we substitute $\hat{\underline{\theta}}_{k-1}$ in (4.56) and (4.57), taking care of (4.58) we get:

$$V_k'(\hat{\underline{\theta}}_{k-1}) = \frac{2}{k} \hat{\underline{\xi}}_k'(\hat{\underline{\theta}}_{k-1}) \hat{\xi}_k(\hat{\underline{\theta}}_{k-1}) \tag{4.59}$$

$$V_k''(\hat{\underline{\theta}}_{k-1}) = \frac{k-1}{k} V_{k-1}''(\hat{\underline{\theta}}_{k-2}) + \frac{2}{k} \hat{\underline{\xi}}_k'(\hat{\underline{\theta}}_{k-1}) \hat{\underline{\xi}}_k'^{T}(\hat{\underline{\theta}}_{k-1}) \tag{4.60}$$

Define

$$P_k^{-1} = \frac{k}{2} V_k''(\hat{\underline{\theta}}_{k-1}) \tag{4.61}$$

$$\tilde{\underline{\omega}}_k = -\hat{\underline{\xi}}_k'(\hat{\underline{\theta}}_{k-1}) \tag{4.62}$$

Equation (4.61) can be rewritten

$$P_k^{-1} = P_{k-1}^{-1} + \tilde{\underline{\omega}}_k \tilde{\underline{\omega}}_k^{T} \tag{4.63}$$

which can be written as

$$P_k = P_{k-1} - P_{k-1} \tilde{\underline{\omega}}_k (1 + \tilde{\underline{\omega}}_k^{T} P_{k-1} \tilde{\underline{\omega}}_k)^{-1} \tilde{\underline{\omega}}_k^{T} P_{k-1} \tag{4.64}$$

This is one part of the already known type of recursive formulae; for obtaining the other part, remember the Newton Raphson formula:

$$\hat{\underline{\theta}}_k = \hat{\underline{\theta}}_{k-1} - [V_k''(\hat{\underline{\theta}}_{k-1})]^{-1} V_k'(\hat{\underline{\theta}}_{k-1}) \tag{4.65}$$

Substitute (4.61), (4.62) and (4.59) in (4.65)

$$\hat{\underline{\theta}}_k = \hat{\underline{\theta}}_{k-1} + P_k \tilde{\underline{\omega}}_k \hat{\xi}_k(\hat{\underline{\theta}}_{k-1}) \tag{4.66}$$

which can then be rewritten using (4.64)

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\tilde{\omega}}_k \,(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{\omega}}_k)^{-1}\,\hat{\xi}_k \tag{4.67}$$

Söderström (1973) applied these formulae on a model of the form:

$$[1+A]y_k = [b_o+B]u_k + [1+C]\hat{\xi}_k \tag{4.68}$$

He used a state space description associated with this modelling and solved the state equation to obtain $\hat{\hat{\xi}}_k$. It can easily be seen in eq. (4.62) that the vector $\underline{\tilde{\omega}}_k$ contains the derivatives of the prediction error $\hat{\xi}_k$ with respect to the parameters. These derivatives are given by:

$$\left.\begin{aligned} [1+\hat{C}]\frac{\partial\hat{\xi}_k}{\partial\hat{a}_i} &= y_{k-i} \\ [1+\hat{C}]\frac{\partial\hat{\xi}_k}{\partial\hat{b}_i} &= -u_{k-i} \\ [1+\hat{C}]\frac{\partial\hat{\xi}_k}{\partial\hat{c}_i} &= -\hat{\xi}_{k-i} \end{aligned}\right] \tag{4.69}$$

These equations can also be brought into a state space description as a solution for the derivatives. The diagram of this method is also shown in figure 4.11.

From the considerations given above, it can be concluded that Fuhrt's IQL method and Söderströms AML method are very closely related to each other.

The minor differences are:

- the amount of old samples that will be recalculated;
- filtering is performed as AR filtering by Fuhrt (or its approximated MA version) while Söderström obtains the filtered quantities by solving a state space equation;
- small differences of approximation.

Goedheer (1976) generalized the approximate maximum likelihood estimator for models where ARMA modelling for the noise is used. His algorithm, the approximate Markov estimator, has the same form as Söderström's estimator, cf. eq. (4.64) and (4.67) but the parameter

vector $\underline{\theta}_k$ is extended with AR noise parameters and consequently the vector $\underline{\tilde{\omega}}_k$ is extended accordingly:

$$\underline{\tilde{\omega}}_k^T = (\tilde{u}_k,\ldots,\tilde{u}_{k-p},\tilde{y}_{k-1},\ldots,\tilde{y}_{k-q},\tilde{\tilde{\xi}}_{k-1},\ldots,\tilde{\tilde{\xi}}_{k-s},\tilde{\tilde{\xi}}_{k-1},\ldots,\tilde{\tilde{\xi}}_{k-r}) \tag{4.70}$$

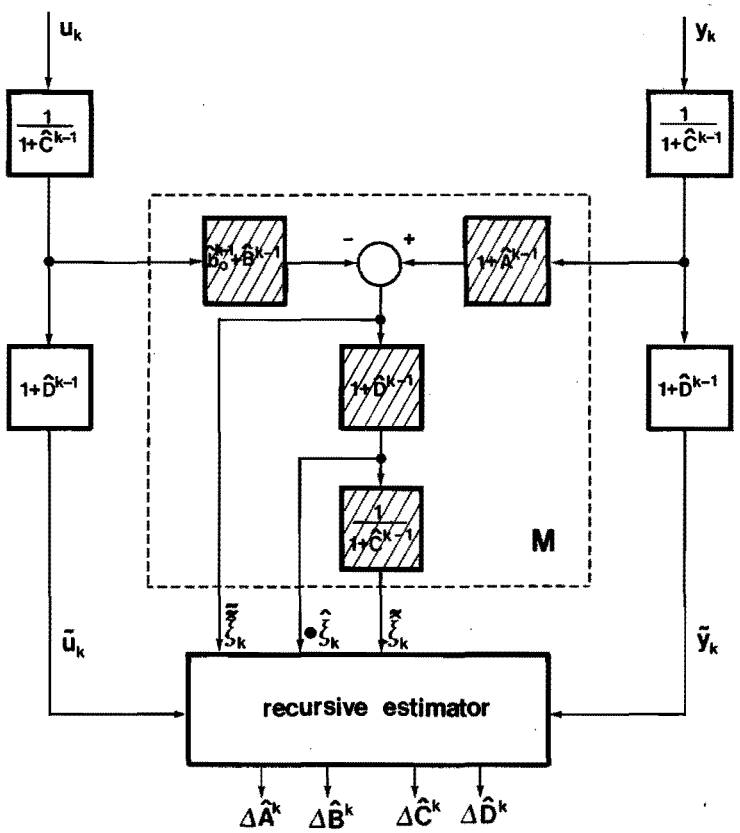


Fig. 4.12 The approximate Markov estimator

The filtered quantities in this vector are again found by taking the derivatives of the prediction error:

$$\frac{\partial\hat{\xi}_k}{\partial a_i} = \frac{[1+\hat{D}]}{[1+\hat{C}]} y_{k-i} = \tilde{y}_{k-i}$$

$$
\left.
\begin{aligned}
-\frac{\partial \hat{\xi}_k}{\partial b_i} &= \frac{[1+\hat{D}]}{[1+\hat{C}]} u_{k-i} = \tilde{u}_{k-i} \\
-\frac{\partial \hat{\xi}_k}{\partial c_i} &= \frac{1}{[1+\hat{C}]} \hat{\xi}_{k-i} = \tilde{\xi}_{k-i} \\
\frac{\partial \hat{\xi}_k}{\partial d_i} &= \frac{1}{[1+\hat{D}]} \hat{\xi}_{k-i} = \tilde{\xi}_{k-i}
\end{aligned}
\right| \qquad (4.71)
$$

The diagram of this estimator is given in figure 4.12

4.4.8 The Suboptimal IV estimator (SIV)

The above given simple IV estimators yield, in general, consistent estimates. No special extra filtering or extension of matrices, as we have seen in previous paragraphs, is necessary. Nevertheless more complex IV schemes have been proposed to improve the quality of the estimate. Young (1965) proposed, on a rather heuristic basis, the use of fixed pre-filters to filter the input- and output signal, prior to feeding them into the model and the IV generating filters. These filters were not updated during the recursion, and their bandpass was chosen within the bandpass of the system itself.

We have already seen in chapter 3 that Wong (1966) proposed an "optimal" estimator:

$$
\underline{\hat{\theta}} = [\Omega^T(u,x)R^{-1}\Omega(u,y)]^{-1}\, \Omega^T(u,x)R^{-1}\underline{y} \qquad (4.72)
$$

Here the pre-filtering of the different signals is determined by the noise characteristics. Smets (1970) used this principle for a recursive estimator with fixed, a priori known, filters $\frac{[1+C]_t}{[1+D]_t}$, according to diagram 4.13, in order to realize the weighting by the inverse covariance of the noise. An estimate $\hat{x}_k$ of the clean output x_k was generated using an adaptive model. Young (1976) presented, in fact, the same scheme and called it "refined IV algorithm". The complication with these latter schemes is that knowledge of the noise spectrum should be available.

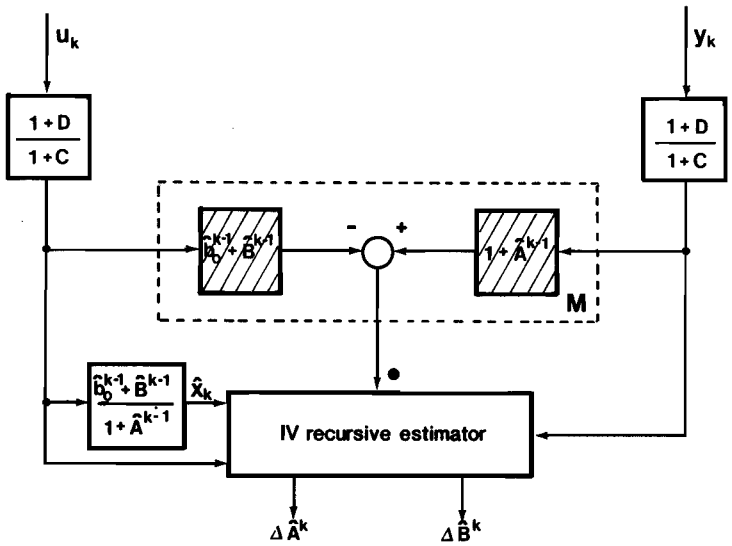


Fig. 4.13 The Suboptimal IV estimator

As we have already seen in previous paragraphs, this will not usually be the case. To gain this knowledge, along with the estimation of the process parameter, a separate estimator, which is capable of estimating noise parameters from the residuals, can be used. A variety of possibilities for this purpose has been mentioned already in the previous paragraphs. Young and Hastings-James (1970) proposed the use of an AML estimator for this purpose.

4.4.9 The IVEMM estimator

Smets (1970) proposed an extended IV algorithm along the lines of the extended matrix method. Here the IV estimator using an adaptive model is extended, i.e. the vectors $\underline{\omega}_k$ and $\underline{z}_k$ contain samples of $\hat{\underline{e}}_k$ and $\hat{\underline{\xi}}_k$ as already seen with the EMM.

The expression for the estimator is also given in eq. (4.48), with the appropriate extensions of $\underline{\omega}_k$ and $\underline{z}_k$. The diagram of the estimator is given in figure 4.14.

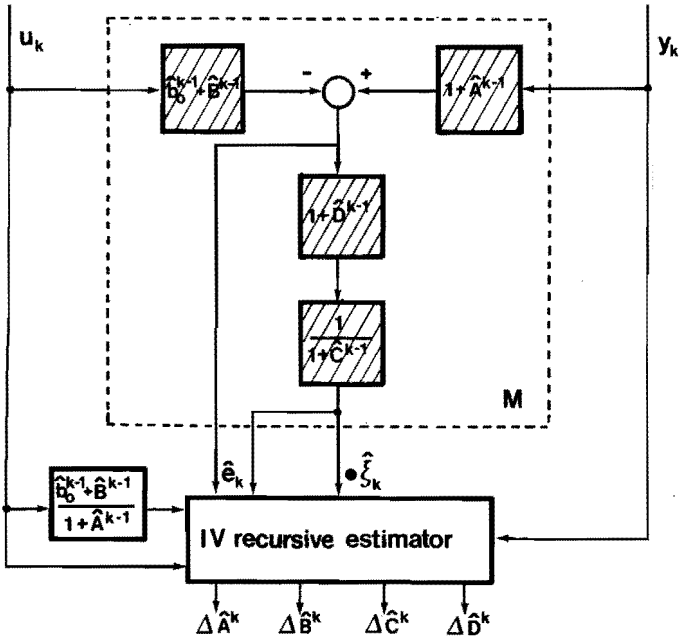


Fig. 4.14 The IV Extended Matrix (IVEMM) estimator

4.4.10 The general estimator

By comparing the diagrams of the approximate Markov estimator, the suboptimal IV estimator and the IV extended matrix estimator, we may propose a general diagram for these types of advanced estimators, leading to an even more versatile estimator: the general estimator. This estimator is, in fact,

- the IV variant of the approximate Markov estimator or
- the extended model variant of the optimal IV estimator or
- the filtering variant of IVEMM.

We have already given the diagram of the general estimator in fig. 4.2. For the choices of the model blocks M_i, $i = 1,2,3,4$ and the filters F_i and Φ_i, $i = 1,2$ we refer to table 4.1. This estimator and its properties have not yet been proposed in the literature. From the general framework within which we have presented the current estimation schemes, we can conclude that the various estimators are

very much related and can be considered as special cases of this general estimator.

The main distinction is in the extent of the parameter vector to be estimated. If all A,B,C,D parameters are estimated, the availability of the noise parameters can be used to apply weighted loss functions with optimality properties like the AML, Markov estimator and the (sub)optimal IV estimator. The weighting of these estimators is identical. It can also be seen that the EMM estimator is a simplified version of the AML and the approximated Markov estimator. The filtering is then indeed not essential for obtaining a white prediction error, but can help to prevent divergence as we will see in section 4.5. In the same way, the IVEMM is the simplified version of the general estimator.

For the design of computer programs for estimation algorithms, it is very helpful to use the concepts of the general estimator, as the different functions of the general estimator, indicated by the blocks M_i, F_i and Φ_i, can be incorporated in one program, which allows the use of various estimation schemes.

4.4.11 The Stochastic Approximation algorithm (SA)

From a historical and a computational point of view, the class of stochastic approximation algorithms is of interest. It can be seen as a simplified version of the recursion algorithm that we have already met in the preceeding paragraphs. Tsypkin (1966) gives a review of this class of methods and analyses their convergence properties. For our type of application, the estimation of process parameters, the requirement of independent noise is usually not met. This can be compared completely with the conditions for consistent estimation for the common least squares estimator. The basic idea is the following: For some convex loss function V e.g.

$$V = (y_k - \underline{\omega}_k^T \hat{\underline{\theta}}_k)^2 \qquad (4.73)$$

the gradient of V with respect to the parameter vector is computed.

The recursion formula for updating the parameter estimate is then:

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} - \rho_k \nabla V \Big|_{\underline{\hat{\theta}}_{k-1}} \tag{4.74}$$

where ρ_k is some positive number. In Tsypkin (1966) the conditions can be found for convergence of this algorithm. They dictate the choice of ρ_k, and usually a choice like $\rho_k = k^{-1}$ will satisfy. Comparison of those stochastic approximation algorithms with the types of weighted least squares methods already met, shows that the main difference is the choice of the scalar ρ_k instead of the matrix P_k for the weighted least squares algorithms. In this sense the stochastic approximation algorithms can be seen as a simplified version of the weighted least squares algorithms. In principle all modifications as displayed for weighted least squares estimation could be of interest for the stochastic approximation algorithm as well. Nevertheless, in the more recent literature, the stochastic approximation does not play an important role. The extra amount of computations involved with handling the matrix P_k instead of a scalar ρ_k is for the modern, very fast and relatively large (even for mini's) computers usually not prohibitive.

An interesting development within the stochastic approximation algorithm is worth mentioning, as it also influenced the development of the weighted least squares estimators. This is an idea attributed to Panuska (1969), who observed the inconsistence properties of the stochastic approximation algorithm when dealing with coloured equation errors. He proposed to estimate the MA parameters of the noise along with the process parameters, introducing an enlarged parameter vector $\underline{\theta}_k$ and corresponding vector $\underline{\omega}_k$. His results were promising. Valis and Gustavsson (1969) compared this method with Åström's maximum likelihood algorithm and found that more passes through the data were necessary in order to obtain estimates having nevertheless a worse variance than the maximum likelihood estimator.

4.5 Convergence aspects of recursive estimators

In the previous paragraphs we reviewed several recursive estimators and concluded that these estimators are based on one or more of the three principles for prediction error whitening: filtering, model

extension or use of IV quantities. We have also seen that several of these estimators are approximate versions of (possibly more favourable) variants, where for example filtering with filters, containing recently estimated parameters is performed over a larger amount of past data than in the more currently used versions, where only the last sample is filtered with the newly obtained filter to avoid excessive compuation times. Because of this, the signals contained in the vector $\underline{\tilde{\omega}}_k$ are not statistically stationary, which makes a convergence analysis rather hard. Nevertheless it is important to know, or at least to have an idea, under which experimental conditions an estimator may have consistence properties.

For studying the convergence aspects of prediction error estimators, an analysis method has been proposed by Ljung (1977a,b), based on the concept of an ordinary differential equation (ODE), associated with the recursive estimation algorithm. We will give here only a brief outline of this method.
The recursive estimator is in its general form:

$$\left.\begin{aligned} \underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\underline{\omega}_k^+ \,(1+\underline{\omega}_k^{+T}P_{k-1}\underline{\omega}_k^+)^{-1}\hat{\varepsilon}_k \\ P_k &= P_{k-1} - P_{k-1}\underline{\omega}_k^+ \,(1+\underline{\omega}_k^{+T}P_{k-1}\underline{\omega}_k^+)^{-1}\underline{\omega}_k^{+T}P_{k-1} \end{aligned}\right] \qquad (4.75)$$

where $\hat{\varepsilon}_k$ is the one step ahead prediction error, depending on all previous estimates $\underline{\hat{\theta}}_i$. It is specified for the different methods in the previous section. It contains the (filtered) measurables of the (extended) model. The quantity $\underline{\omega}_k^+$ is any of the vectors of (filtered or extended) measurements met in the previous paragraphs. If we write:

$$P_k^{-1} = \sum_{i=1}^{k} \underline{\omega}_i^+ \,\underline{\omega}_i^{+T} = kR_k \qquad (4.76)$$

then it follows:

$$R_k - R_{k-1} = \frac{1}{k}\,(\underline{\omega}_k^+ \,\underline{\omega}_k^{+T} - R_{k-1}) \qquad (4.77)$$

and for $\underline{\hat{\theta}}_k$

$$\underline{\hat{\theta}}_k - \underline{\hat{\theta}}_{k-1} = R_{k-1}^{-1}\,\left[(k-1)+(k-1)\underline{\omega}_k^{+T}R_{k-1}^{-1}\underline{\omega}_k^+\right]^{-1}\,\underline{\omega}_k^+\,\hat{\varepsilon}_k \qquad (4.78)$$

Now the following idealisations are made:

a) the term $\left[(k-1)+(k-1)\underline{\omega}_k^{+T} R_{k-1}^{-1} \underline{\omega}_k^{+}\right]^{-1}$ will be approximated by k^{-1}

b) the prediction error $\hat{\varepsilon}_k$, which is dependent on all previous estimates $\underline{\hat{\theta}}_i$, is thought to be dependent only on the previous estimate $\underline{\hat{\theta}}_{k-1}$. The same holds for the vector $\underline{\omega}_k^{+}$. These new quantities will be denoted by $\bar{\hat{\varepsilon}}_k$ and $\underline{\bar{\omega}}_k^{+}$.

c) introduce

$$\left.\begin{aligned} \underline{f}(\underline{\theta}) &= E\{\underline{\bar{\omega}}_k^{+}\, \bar{\hat{\varepsilon}}_k\} \\ G(\underline{\theta}) &= E\{\underline{\bar{\omega}}_k^{+}\, \underline{\bar{\omega}}_k^{+T}\} \end{aligned}\right] \qquad (4.79)$$

This means that we will replace terms in the above given equations by their expectations:

$$\left.\begin{aligned} \underline{\hat{\theta}}_k - \underline{\hat{\theta}}_{k-1} &= \frac{1}{k} R_{k-1}^{-1}\, \underline{f}(\underline{\theta}_{k-1}) \\ R_k - R_{k-1} &= \frac{1}{k}\left[G(\underline{\theta}_{k-1}) - R_{k-1}\right] \end{aligned}\right] \qquad (4.80)$$

d) instead of the above difference equation which is time variant, we consider the time invariant difference equation

$$\left.\begin{aligned} \underline{\hat{\theta}}_k - \underline{\hat{\theta}}_{k-1} &= R_{k-1}^{-1} \underline{f}(\underline{\theta}_{k-1}) \\ R_k - R_{k-1} &= G(\underline{\theta}_{k-1}) - R_{k-1} \end{aligned}\right] \qquad (4.81)$$

The ODE associated with this recursive estimator is then

$$\left.\begin{aligned} \frac{d\underline{\hat{\theta}}(\tau)}{d\tau} &= R^{-1}(\tau)\, \underline{f}(\underline{\hat{\theta}}(\tau)) \\ \frac{dR(\tau)}{d\tau} &= G(\underline{\hat{\theta}}(\tau)) - R(\tau) \end{aligned}\right] \qquad (4.82)$$

with

$$\left.\begin{aligned} \underline{f}(\underline{\hat{\theta}}) &= E\{\underline{\bar{\omega}}^{+}(\underline{\hat{\theta}})\, \bar{\hat{\varepsilon}}(\underline{\hat{\theta}})\} \\ G(\underline{\hat{\theta}}) &= E\{\underline{\bar{\omega}}^{+}(\underline{\hat{\theta}})\, \underline{\bar{\omega}}^{+T}(\underline{\hat{\theta}})\} \end{aligned}\right] \qquad (4.83)$$

where $\bar{\hat{\varepsilon}}_k$ and $\underline{\bar{\omega}}^{+}$ are obtained by using a constant parameter $\underline{\theta}$ instead of the sequence of estimates that become available during the estimation. So the quantities $\bar{\hat{\varepsilon}}_k$ and $\underline{\bar{\omega}}_k^{+}$ may be considered as stationary stochastic processes.

For stability analysis of the ODE choose a Lyapunov function:

$$V = E\{\hat{\underline{e}}^T\hat{\underline{e}}\} \tag{4.84}$$

The relation with the function $\underline{f}(\underline{\theta})$ is:

$$\underline{f}(\underline{\theta}) = E\{\tilde{\underline{\omega}}(\underline{\theta})\hat{\underline{e}}^T(\underline{\theta})\} = -\tfrac{1}{2}\frac{d}{d\underline{\theta}}\{E\{\hat{\underline{e}}^T\hat{\underline{e}}\}\} = -\tfrac{1}{2}\frac{dV}{d\underline{\theta}} \tag{4.85}$$

Then it follows that:

$$\frac{dV}{d\tau} = \left(\frac{dV}{d\underline{\theta}}\right)^T \frac{d\underline{\theta}}{d\tau} = \left(\frac{dV}{d\underline{\theta}}\right)^T R^{-1}(\tau)\,\underline{f}(\underline{\theta}(\tau)) =$$

$$= -\tfrac{1}{2}\left(\frac{dV}{d\underline{\theta}}\right)^T R^{-1}(\tau)\left(\frac{dV}{d\underline{\theta}}\right) \leq 0 \tag{4.86}$$

It can be concluded that the ODE has stationary stability points to which all solutions converge. An important observation is that the quantities $\hat{\underline{e}}_k$ and $\bar{\underline{\omega}}_k^+$ have to be stationary. If the previous estimate is used to generate these quantities this assumption might be violated as the previous estimate might cause even an unstable process. Therefore the obtained estimates have to be monitored and converted into stable estimates before being used to generate $\hat{\underline{e}}_k$ and $\bar{\underline{\omega}}_k^+$. This is a weak point of the prediction error algorithms. Ljung's analysis method using the ODE requires this stability.

The relation between the recursive estimation algorithm and the ODE has been worked out by Ljung (1977b). He showed that convergence with probability one of the estimation algorithm is associated with stability of the ODE:

- only stable stationary points of the ODE are possible convergence points of the estimation algorithm. This implies that if the estimated parameter $\hat{\underline{\theta}}_k$ converges to $\underline{\theta}^*$ then

  a) $$\underline{f}(\underline{\theta}^*) = \underline{0} \tag{4.87}$$

  b) $$G^{-1}(\underline{\theta}^*)\left.\frac{d\underline{f}(\underline{\theta})}{d\underline{\theta}}\right|_{\underline{\theta} = \underline{\theta}^*} \tag{4.88}$$

  has all eigenvalues in the left half plane. This can be understood by linearizing the ODE around $\underline{\theta}^*$ where $\underline{f}(\underline{\theta}^*) = \underline{0}$. This yields:

$$R = G(\underline{\theta}^*) \tag{4.89}$$

$$H(\underline{\theta}^*) = \left.\frac{d\underline{f}(\underline{\theta})}{d\underline{\theta}}\right|_{\underline{\theta} = \underline{\theta}^*} \tag{4.90}$$

$$\underline{\Delta\theta} = \underline{\theta} - \underline{\theta}^*$$

leading to

$$\underline{\dot{\Delta\theta}} = G^{-1}(\underline{\theta}^*)\, H(\underline{\theta}^*)\, \underline{\Delta\theta} \tag{4.91}$$

- global asymptotic stability of the solution $\underline{\theta}^*$ of the ODE implies that the estimate $\hat{\underline{\theta}}_k$ converges to $\underline{\theta}^*$ with probability one, as $k \to \infty$.

- the trajectories of the ODE can be interpreted as the "asymptotic paths" of the algorithm.

After this review of the (commonly accepted) convergence analysis of Ljung, we may review the consistence properties of the estimators of levels of complexity 0 and 1, as presented in the recent literature.

Level 0 The LS estimator. In chaper 3 we have already seen that consistence only occurs for a specific colouring of the error. The equation error has to be a white noise signal. This leads to the requirement:

$$\frac{[1+C]_t}{[1+D]_t} = 1 \tag{4.92}$$

This requirement is a severe limitation and is, in fact, responsible for the development of the estimators of the other classes.

Level 1 The GLS estimator: Söderström (1974) showed that for low signal-to-noise ratios this algorithm can converge to other local minima instead of to the global solution. For high signal-to-noise ratios this algorithm gives, in general, reliable results.

The EMM estimator: This method has been used with success by a wide variety of authors, in spite of the fact that consistence is not guaranteed in all cases. Ljung, Söderström and Gustavsson (1975) constructed counter-exam-

ples to convergence and showed also the possibility for divergence. Goedheer (1976) gives simulation results, showing also the possibility of divergence for this method. A nice explanation for this phenomenon can be given, based on the particular choice of the experimental conditions and the structure of the EMM algorithm. This explanation fits very well with the remedy which Ljung, Söderström and Gustavsson (1975) proposed to overcome the divergence problems of the EMM. The explanation is based on the following: create an experimental condition where in the input (and/or output) signal one frequency, or a small frequency band, is dominant. If the inverse noise filter (in the diagram of EMM) for this particular frequency gives a phase shift between 90° and 270° then the gradient of the loss function with respect to the parameters (this is the product $\underline{\omega}_k\hat{\xi}_k$ in eq. (4.29)) will be directed in the wrong direction. So therefore divergence can be expected.
In this reasoning, the term $P_{k-1}(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}$ is not taken into account.
This term gives an orthogonalization to the adjustment scheme. Using simulations, Goedheer (1976) investigated the influence of the cross terms in the P_{k-1} matrix for the divergence effect for this type of experimental conditions and found that the presence of these cross terms indeed improved the behaviour of the adjustment, but the resulting quality of the adjustment remained very poor. From this type of simulation and the counter examples given by Ljung, Söderström and Gustavsson (1975) we can conclude that we may expect divergence, or poor convergence, only for very restricted types of experimental conditions.

Based on the observation that for a stable solution the matrix $G^{-1}(\underline{\theta}^*)H(\underline{\theta}^*)$ in eq. (4.91) should have stable eigenvalues, i.e. in the left half plane of the s-plane, Ljung (1977a,b) shows that this can be guaranteed for this estimator if $[1+\hat{C}]^{-1}-\frac{1}{2}$ is positive real.
It is worth noticing that stable noise processes exist

where this condition is not fulfilled. Based on the concept of extra phase-shift between the signals u_k and ξ_k, it is easy to understand the remedy to prevent divergence, as proposed by Ljung, Söderström and Gustavsson (1975), which they concluded from the already mentioned theoretical analysis. They propose the use of an extra filtering of the signals contained in the extended vector of measurables $\underline{\omega}_{k-1}$ by the inverse (estimated) noise filter. In figure 4.15 a diagram of this modified EMM estimator is drawn for the case of A, B, C parameters. In this diagram it can easily be seen that a phase shift between $\tilde{u}_k$ and $\hat{\xi}_k$, due to the inverse noise filter, does not occur.

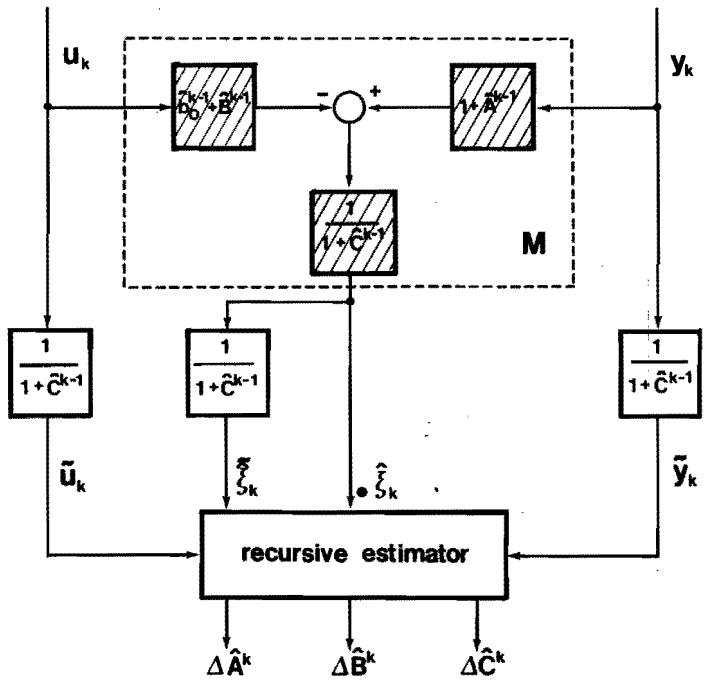


Fig. 4.15 Modified EMM estimator

If we shift the filters properly, we arrive at the same diagram as for the AML/IQL estimator (cf. fig. 4.11).

This means that this modified EMM estimator belongs to class 2 estimators as it contains two extra operations (model extension and filtering) to obtain consistent estimates. If we extend the interpretation of phase shift between $\underline{\omega}_k$ and $\hat{\xi}_k$ caused by the inverse noise filter for the common EMM estimator, for cases where the phase shift is not between 90° and 270° but between 270° and 90°, then we may expect fast convergence to occur when the phase shift is zero and slower convergence when there is a component in the error signal $\hat{\xi}_k$ with a phase shift of 90° or 270°. This component will not contribute to the adjustment and the in-phase component will be therefore smaller.

An important contribution to the discussion of the convergence of the extended matrix method has been given by Solo (1979), who considered a MA modelling of the noise. He analyses an interesting version of the EMM, already introduced by Young (1974), where the prediction error is replaced by the residuals, which are obtained by using the estimated ARMA parameters of the process of the same recursion step, cf. paragraph 4.4.5. For this type of modified EMM estimator, Solo proved convergence by using the martingale convergence theorem, provided that the following condition is fulfilled:

$[1+\hat{C}]^{-1} - \frac{1}{2}$ is positive real.

The main difference for this type of EMM-scheme is that no monitoring of estimation results is necessary, as in the case where past prediction errors are used in the extended vector of measurables.

The IV estimator: Söderström, Ljung and Gustavsson (1978) showed that convergence occurs for the IV estimator if, again, the matrix $G^{-1}(\underline{\theta}^*)\underline{f}'(\underline{\theta}^*)$ is stable, if a fixed model $\frac{[\bar{b}_o+\bar{B}]}{[1+\bar{A}]}$ is used to generate the IV quantity. The estimates need to be monitored in this case as well. Solo (1981)

shows that, using residuals instead of prediction errors, the convergence of the IV algorithm can be guaranteed provided that $[1+A]^{-1} - \frac{1}{2}$ is positive real.
This means that for some stable systems, which do not fulfill this condition, divergence may occur.
Until now, it is an open question under what conditions a recursive IV estimator with an adaptive model has convergence properties.

For the more complex estimation schemes, only a thorough convergence analysis has been done for the AML estimator; cf. Åström and Söderström (1974). It can be proved that the matrix $G^{-1}(\underline{\theta}^*)f'(\underline{\theta}^*)$ for this estimator is always asymptotically stable so that no restrictive conditions for noise dynamics and/or process dynamics may be expected.

4.6 Conclusion

In this chapter we have presented a general set-up for simple and more complicated least squares estimators. We showed that within this general framework the recursive estimators can be classified. This framework is based on the distinction of three basic operations related to estimation: correlative weighting, filtered weighting and model extension. Based on these three main operations, a general estimator could be proposed, such that the existing recursive estimators are special cases of this general estimator. From this general concept, the relationship between the existing estimators can then be shown. A short discussion of existing literature is given concerning the convergence analysis of these estimators; this is still a popular topic in the very recent literature.

CHAPTER FIVE:

ESTIMATORS FOR NOISE CORRUPTED INPUT-OUTPUT MEASUREMENTS

5.1 Introduction

In the preceeding chapters we discussed the estimation of model parameters in the case where uncertainties can be modelled as an equation error, which can be treated as disturbances on the output signal. In this case measurement data of the undisturbed input signal are assumed to be available.

Often it will occur that both input- and output measurements are disturbed to such an extent that the previously described estimation algorithms will produce biased results. Such disturbances can be inherently due to the measuring principles used. For instance in biological systems often only signals with a bad signal-to-noise ratio can be obtained, e.g. when measurements have to be restricted to non-invasive ones only. For such cases it is of interest to modify the existing algorithms in such a way that also disturbed input-output data can be processed.

The problem of estimating parameters using disturbed input-output measurements has received rather little attention in literature, compared to many papers devoted to the simpler problem of output noise contamination. A few solutions have been proposed but, unfortunately, considerable a priori knowledge is assumed or restricting conditions are imposed on the disturbances. Akashi and Moustafa (1975) use an EMM approach for estimation of process and noise parameters, but they assume equal colouring for both the input and output noises.

With this restrictive assumption the EMM approach, as discussed in the previous chapter, can be used, assuming also that $b_0 = 0$. An estimator for the case that $b_0 \neq 0$ is also proposed, but then it is assumed that the covariance matrix of the noise is known. Roosdorp (1974) proposes a combination of two IV estimators, which have to be performed successively. The technique is rather involved

and works well only for first order processes with white disturbances.

However, $b_0 \neq 0$ is possible in this approach. Söderström (1979) suggests an IV approach, which, again, is useful for white disturbances. For this case, he proposes a different type of parametrization using innovations form, and shows that the parameters of this parametrization can be identified. Klijn (1977) also assumes white disturbances and gives an implicit quadratic expression for an estimator. This expression can be obtained from the common least squares estimator, if the matrix $\Omega^T\Omega$ is replaced by $\Omega^T\Omega - \Sigma$, where Σ is a diagonal matrix containing the noise variances.

Rogers and Steiglitz (1970) propose a maximum likelihood approach, but the solution for non-white noise is numerically prohibitive, so that they only present simulation results for white disturbances. A promising method has been proposed recently, cf. van den Dungen (1978) and van den Dungen and Eykhoff (1981). This method is called the Least Squares Like method (LSL) and does not impose limiting assumptions on the colouring of the noises. On the other hand, it assumes that two independent measurements of both input- and output signals are available. This may be a practical limitation, depending on the circumstances. Using these independent measurements, IV-like estimators can be constructed, where the extra available measurements serve as IV signals. Simulation results show a good performance of this estimator.

In the following paragraphs we will propose two estimators for the case of input-output disturbances, where <u>different noise colouring</u> is allowed. Each of these estimators can be presented in two variants. These estimators are involved IV-estimators, where each estimator contains two IV-operations and an extra operation which is either model extension or filtering. For a suitable choice of the IV quantities an extra measurement of either the input signal or the output signal (or measurements of signals which are related to the input- or output signals) will be appropriate. This type of extra necessary measurements is, in many cases, practically feasible and less restrictive than the a priori information needed for the schemes discussed in the chapters before. The design of these estimators is a direct result of the concept of designing estimation schemes, consisting

of three basic operations, as introduced in chapters 3 and 4.

5.2 Problem formulation

Assume for the process description, the situation of fig. 5.1.

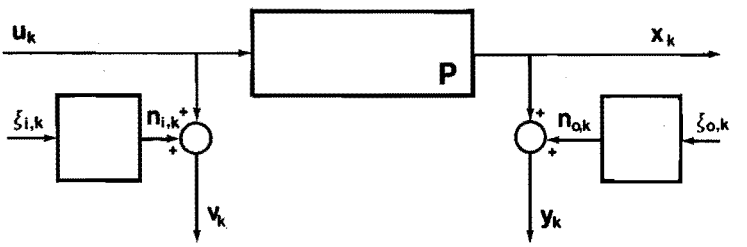


Fig. 5.1 Corrupted input and output measurements

Analogously to the formulation in the previous chapters, we can write:

$$\left.\begin{aligned} \underline{x} &= -X\underline{a}_t + U\underline{b}_t = \Omega(u,x)\underline{\theta}_t \\ \underline{y} &= \underline{x} + \underline{n}_o \\ \underline{v} &= \underline{u} + \underline{n}_i \end{aligned}\right] \qquad (5.1)$$

where we assume that $\underline{n}_o$ and $\underline{n}_i$ are mutually uncorrelated, zero mean, stationary noise sequences.

The vectors $\underline{x}$, $\underline{y}$, $\underline{n}_o$, $\underline{n}_i$, $\underline{v}$, $\underline{g}$, $\underline{e}$ and $\underline{f}$ are signal vectors, all defined like $\underline{y}$:

$$\underline{y}^T = (y_{q+1}, \ldots, y_N) \qquad (5.2)$$

$$\underline{\theta}_t^T = (\underline{b}^T, -\underline{a}^T)_t = (b_o, \ldots, b_p, -a_1, \ldots, -a_q)_t \qquad (5.3)$$

$$\Omega(u,x) = [U|X] = \left[\begin{array}{ccc|ccc} u_{q+1} & \cdots & u_{q+1-p} & x_q & \cdots & x_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ u_N & \cdots & u_{N-p} & x_{N-1} & \cdots & x_{N-q} \end{array}\right] \qquad (5.4)$$

The matrix $\Omega(v,y)$ has dimensions equal to $\Omega(u,x)$:

$$\Omega(v,y) = [V|Y] \qquad (5.5)$$

The matrix N_o has dimensions equal to Y, whereas the matrix N_i has dimensions equal to U (and V).

The set of equations (5.1) can be rewritten:

$$\underline{y} = \Omega(v,y)\underline{\theta}_t + \underline{n}_o + N_o\underline{a}_t - N_i\underline{b}_t \tag{5.6}$$

Introducing:

$$\left.\begin{aligned} \underline{e} &= \underline{n}_o + N_o\underline{a}_t \\ \underline{f} &= -N_i\underline{b}_t \\ \underline{g} &= \underline{e} + \underline{f} \end{aligned}\right] \tag{5.7}$$

we have:

$$\underline{y} = \Omega(v,y)\underline{\theta}_t + \underline{g} \tag{5.8}$$

In this description we have an equation error g_k which consists of two components. The quantity e_k is well known from previous chapters; it is the equation error due to the output disturbances. It is essential for our forthcoming considerations that e_k is independent of the measured input signal v_k. Further we have f_k, which is the equation error due to the input disturbances; f_k has to be independent of the measured output signal.

It is illustrative to consider the situation of fig. 5.2 where an input disturbance is not measured but is fed into the process; i.e. the case of incomplete input observation.

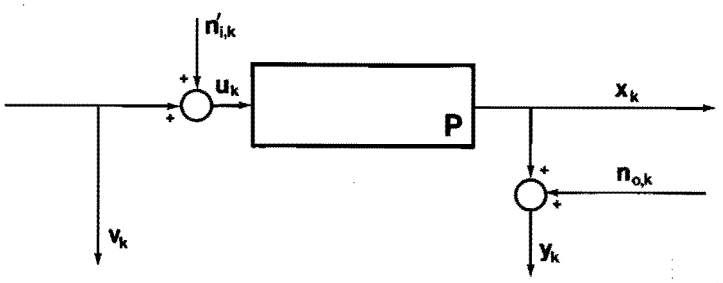


Fig. 5.2 Incomplete input observation

This situation can be handled by treating this type of input disturbance as coloured output disturbance:

$$\underline{y} = \Omega(v,y)\underline{\theta}_t + \underline{n}_o + N_o\underline{a}_t + N_i'\underline{b}_t \qquad (5.9)$$

The equation error

$$\underline{g} = \underline{n}_o + N_o\underline{a}_t + N_i'\underline{b}_t \qquad (5.10)$$

is independent of the input measurement v_k, if v_k and $n_{i,k}$ are mutually independent, so that this type of input disturbance can be treated as coloured output disturbance.

It would seem attractive to model the input disturbance n_i, as in fig. 5.1, in such a way that it can be treated, as in fig. 5.2, as output disturbance, so that existing estimation schemes suited for output disturbances could also be used for input-output disturbances. Therefore consider fig. 5.3.

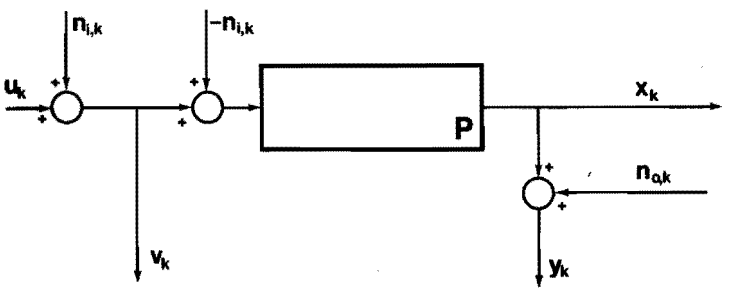


Fig. 5.3 Input measurement noise modelled as output noise

If the disturbance $-n_{i,k}$ is modelled as an output disturbance as explained in eq. (5.10), then the resulting equation error e_k is not independent of the input measurement v_k as can be seen from fig. 5.3. This violation of the independence requirement of output disturbance and input measurement will cause inconsistence of the resulting estimates.

Now we return to the scheme given in fig. 5.1. In the description (5.8) we have now an equation error $\underline{e} + \underline{f}$, which is generated from

two independent white noise sources $\underline{\xi_i}$ and $\underline{\xi_o}$ through shaping filters. There is the possibility of modelling the noise shaping filters in different ways, using AR, MA or ARMA descriptions. It will turn out that AR descriptions will be convenient:

$$\left.\begin{aligned} e_k &= \frac{1}{[1+D_o]_t}\,\xi_{o,k} \\ f_k &= \frac{1}{[1+D_i]_t}\,\xi_{i,k} \end{aligned}\right] \qquad (5.11)$$

so that the input and output noises are derived from the mutually independent white sources $\xi_{o,k}$ and $\xi_{i,k}$ in the following way:

$$\left.\begin{aligned} n_{o,k} &= \frac{1}{[1+A]_t[1+D_o]_t}\,\xi_{o,k} \\ n_{i,k} &= -\frac{1}{[b_o+B]_t[1+D_i]_t}\,\xi_{i,k} \end{aligned}\right] \qquad (5.12)$$

where the polynomials are defined as:

$$\left.\begin{aligned} [1+A]_t &= [1 + a_1 z^{-1} + \ldots\ldots + a_q z^{-q}]_t \\ [b_o+B]_t &= [b_o + b_1 z^{-1} + \ldots\ldots + b_p z^{-p}]_t \\ [1+D_o]_t &= [1 + d_{o,1} z^{-1} + \ldots\ldots + d_{o,r_o} z^{-r_o}]_t \\ [1+D_i]_t &= [1 + d_{i,1} z^{-1} + \ldots\ldots + d_{i,r_i} z^{-r_i}]_t \end{aligned}\right] \qquad (5.13)$$

For asymptotic stability of the signals involved, it is assumed that the roots of $z^q[1+A]_t$, $z^p[b_o+B]_t$, $z^{r_o}[1+D_o]_t$ and $z^{r_i}[1+D_i]_t$ all lie inside the unit circle of the complex z-plane.

After this notational introduction of the situation with corrupted input and output measurements, the estimation task is formulated as: "Find a (recursive) algorithm such that the process parameters $\underline{\theta_t}$ will be estimated consistently; if it is necessary to estimate (some of) the noise parameters for this goal, they should also be included".

5.3 Estimators for disturbed input-output data

In this paragraph we will propose two different types of estimators which are suited for corrupted input-output data as specified in the previous paragraph.

In chapter 3 and chapter 4 we have seen that three basic operations for obtaining consistent estimates can be distinguished. In principle, each of these basic operations is capable of converting the inconsistent common least squares estimator into a consistent estimator of higher complexity. In these cases we were dealing with a coloured equation error, which was obtained by filtering a single white noise source.

From eq. (5.6), (5.7) and (5.11) we can conclude that the resulting equation error g_k originates from two independent white noise sources, ξ_i and ξ_o, which are both filtered by different filters:

$$g_k = e_k + f_k = \frac{1}{[1+D_o]_t}\,\xi_{o,k} + \frac{1}{[1+D_i]_t}\,\xi_{i,k} \qquad (5.14)$$

The main idea of the proposed estimators will be a combination of two different basic operations within one estimator to obtain consistent results. One operation has to be tuned to the input noise f_k and the other operation to the output noise e_k.

5.3.1 The IOIVEMM approach

With this estimator we combine, within one estimator, the model extension and the instrumental variable approach. Each of these two operations will be tuned to one of the noises, which means that we have two ways of applying this algorithm.

The estimation of noise parameters in the case of input-output disturbances is more complicated compared to the case of only output disturbances. This is due to the fact that no good estimates of the two equation errors e_k and f_k can be made available separately. Only the quantity $g_k = e_k + f_k$ can be used, but this implies that, for estimation of output noise parameters, the quantity f_k acts as a disturbance.

For consistent estimation of the output noise parameters, we have to aim at a white residual due to e_k. This implies that, because of the different colouring of input- and output noises, the residual due to f_k will be coloured.

If a least squares type of estimator is applied then, because of the symmetric form of this estimator, this will lead to quadratic terms in g_k. This will then cause inconsistent estimates. An IV type of estimator allows us to make use of an extra signal which, in this case, may not be correlated with the input disturbance.
Again e_k, or a good estimate of it, could be used for this but, unfortunately, this signal is not available. Signals which are correlated with e_k, but uncorrelated with f_k, are available, e.g. y_k. Then an extra quantity is introduced implicitly, viz. x_k, whose correlation with f_k is zero, as $y_k = x_k + n_{o,k}$.
This is indeed true but, for finite data sequences, the sample correlation
function is not equal to zero. Moreover, as an estimate $\hat{g}_k$ of g_k has to be used, this estimate $\hat{g}_k$ will be correlated with the input signal, which is also the case for x_k, leading to a non-zero correlation of these two signals. This introduces an extra error cause to the estimation result. The use of extra output measurements can be advantageous. The IV quantity, for estimating output noise parameters, could then be:

$$y_{1,k} - y_{2,k} = n_{o1,k} - n_{o2,k}$$

leading also to a vanishing sample covariance of $n_{o2,k}$ and f_k.

5.3.1a First variant of the IOIVEMM approach

In this section we will propose an IVEMM estimator, where the IV part is used to tackle the input disturbance and the model extension to tackle the output disturbance. For explanatory reasons, we will first derive an explicit version of this estimator. As with all EMM estimators, a recursive version is more appropriate. This will be dealt with at the end of this section.

Combining (5.8) and (5.11) yields:

$$\underline{y} = \Omega(v,y)\underline{\theta}_t + \underline{g}$$

$$g_k = e_k + f_k = \frac{1}{[1+D_o]_t}\xi_{o,k} + \frac{1}{[1+D_i]_t}\xi_{i,k} \tag{5.15}$$

Pre-multiply this equation by $[1+D_o]_t$

$$[1+D_o]_t g_k = \xi_{o,k} + \frac{[1+D_o]_t}{[1+D_i]_t}\xi_{i,k} \tag{5.16}$$

In vector-matrix notation:

$$\underline{g} = -G\underline{d}_{ot} + \underline{\xi}_o + \frac{[1+D_o]_t}{[1+D_i]_t}\underline{\xi}_i \tag{5.17}$$

where

$$\left.\begin{aligned} \underline{g}^T &= (g_{q+1},\ldots,g_N) \\ \underline{\xi}_o^T &= (\xi_{o,q+1},\ldots,\xi_{o,N}) \\ \underline{\xi}_i^T &= (\xi_{i,q+1},\ldots,\xi_{i,N}) \\ \underline{d}_{ot}^T &= (d_{o,1},\ldots,d_{o,r_o})_t \end{aligned}\right] \tag{5.18}$$

$$G = \begin{bmatrix} g_q & \cdots & g_{q+1-r_o} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ g_{N-1} & \cdots & g_{N-r_o} \end{bmatrix} \tag{5.19}$$

Combining (5.8) and (5.17) yields:

$$\underline{y} = \Omega(v,y)\underline{\theta}_t - G\underline{d}_{ot} + \underline{\xi}_o + \frac{[1+D_o]_t}{[1+D_i]_t}\underline{\xi}_i \tag{5.20}$$

Applying the matrix extension principle:

$$\underline{y} = \Omega(v,y,g)\begin{bmatrix} \underline{\theta}_t \\ -\underline{d}_{ot} \end{bmatrix} + \underline{\xi}_o + \frac{[1+D_o]_t}{[1+D_i]_t}\underline{\xi}_i \tag{5.21}$$

where

$$\Omega(v,y,g) = [\Omega(v,y)|G] = [V|Y|G] \tag{5.22}$$

The quantity $\underline{g}$ will not be available, so that an iterative or recursive scheme has to be used to generate an estimate of this signal

based on previous estimates and using eq. (5.15) where $\underline{\theta}_t$ is substituted by an estimate $\underline{\hat{\theta}}$.

Pre-multiplying eq. (5.21) with Z_1^T yields:

$$Z_1^T \underline{y} = Z_1^T \, \Omega(v,y,g) \begin{bmatrix} \underline{\theta}_t \\ -\underline{d}_{ot} \end{bmatrix} + Z_1^T \, \underline{\xi}_o + Z_1^T \frac{[1+D_o]_t}{[1+D_i]_t} \underline{\xi}_i \qquad (5.23)$$

Now we will take as IVEMM estimator:

$$\underline{\hat{\theta}}' = \begin{bmatrix} \underline{\hat{\theta}} \\ -\underline{\hat{d}}_o \end{bmatrix} = \left[Z_1^T \, \Omega(v,y,g) \right]^{-1} Z_1^T \underline{y} \qquad (5.24)$$

This estimator will be consistent if

$$\left.\begin{array}{ll} \text{I} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_1^T \, \Omega(v,y,g) \right] \text{ is non-singular} \\ \text{II} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_1^T \, \underline{\xi}_o \right] = \underline{0} \\ \text{III} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_1^T \frac{[1+D_o]_t}{[1+D_i]_t} \underline{\xi}_i \right] = \underline{0} \end{array}\right] \qquad (5.25)$$

From the discussion in previous chapters concerning the choice of the IV matrix, we know that several possibilities exist: the use of delayed signals, use of fixed filters or use of varying filters based on previous estimates. We will give here an example of such a choice:

$$Z_1 = \left[\hat{U} \mid z^{-\tau} Y \mid z^{-\tau'} Y \right] =$$

$$\left[\begin{array}{ccc|ccc|ccc} \hat{u}_{q+1} & \cdots & \hat{u}_{q+1-p} & y_{q-\tau} & \cdots & y_{1-\tau} & y_{q-\tau'} & \cdots & y_{q+1-r_o-\tau'} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \hat{u}_N & \cdots & \hat{u}_{N-p} & y_{N-1-\tau} & \cdots & y_{N-q-\tau} & y_{N-1-\tau'} & \cdots & y_{N-r_o-\tau'} \end{array}\right] \qquad (5.26)$$

In this matrix Z_1, all submatrices are IV quantities: the matrix $\hat{U}$ and both shifted submatrices Y. The entries to the submatrix $\hat{U}$ can be constructed using the output signal y as the input to a fixed or varying filter, e.g. previous inverse model of the process. This filter needs to have a delay in order to fulfill condition II. The submatrix $z^{-\tau'}Y$ is suited as IV quantity for the noise ξ_i, as

it is not dependent on ξ_i. The submatrix $z^{-\tau}Y$ is an IV matrix for the output signal. The reason for the introduction of the time shifts $z^{-\tau}$ and $z^{-\tau'}$ is that condition I has to be fulfilled, which is not the case if $\tau=\tau'=0$. A choice of $\tau=0$ and $\tau'=q+1$ satisfies condition I, but problems can occur as explained in appendix II. An alternative choice, especially when $r_o < q$, can be $\tau'=0$ and $\tau = r_o+1$. The choice of a suitable IV matrix is for this type of application somewhat more difficult as three conditions (5.25) have to be fulfilled, where two similar conditions appear for the case of only output noise. The extra complication is now that for the model extension an IV approach also has to be followed (i.e. the choice of the last submatrix $z^{-\tau'}Y$ instead of G).
The use of the choice of eq. (5.26) for an IV quantity is only feasible close to the correct parameter value $\underline{\theta}_t$. This is due to the use of $\hat{g}_k$ instead of g_k. Far from the true parameter $\underline{\theta}_t$, this $\hat{g}_k$ will contain remnants which are related to u_k. Also in $z^{-\tau'}Y$, which is used as an IV quantity for g_k, the component x_k is related to u_k. The correlation of both signals will result in an extra term, yielding a bias for the estimates. This may even lead to divergence, as has been noticed in simulations.

If a second independent measurement of the input signal or the output signal is available, the choice of a proper IV matrix is easier with respect to the three consistence conditions (5.25). We will consider two cases:

1) Two independent measurements of the input signal are available and one measurement of the output signal:

$$\left.\begin{aligned} v_{1,k} &= u_k + n_{i1,k} \\ v_{2,k} &= u_k + n_{i2,k} \\ y_k &= x_k + n_{o,k} \end{aligned}\right] \qquad (5.27)$$

The disturbances $n_{i1,k}$, $n_{i2,k}$ and $n_{o,k}$ may have different colouring but are mutually independent.
A possible choice is now:

$$\left.\begin{aligned} \Omega(v_1,y,\hat{g}) &= [V_1|Y|\hat{G}_1] \\ Z_1 &= [V_2|\hat{Y}_2|\hat{G}_2] \end{aligned}\right| \qquad (5.28)$$

$$\left.\begin{aligned}\hat{\underline{g}}_1 &= \underline{y} - \Omega(v_1,y)\hat{\underline{\theta}}^* \\ \hat{\underline{g}}_2 &= y - \Omega(v_2,y)\hat{\underline{\theta}}^*\end{aligned}\right]$$

where $\hat{\underline{\theta}}^*$ is an estimate of $\underline{\theta}_t$ from a previous iteration, and $\hat{\underline{y}}_2$ is generated by using $\underline{v}_2$.

2) Two independent measurements of the output signal are available and one measurement of the input signal:

$$\left.\begin{aligned}v_k &= u_k + n_{i,k} \\ y_{1,k} &= x_k + n_{o1,k} \\ y_{2,k} &= x_k + n_{o2,k}\end{aligned}\right] \qquad (5.29)$$

Also here, the disturbances $n_{i,k}$, $n_{o1,k}$ and $n_{o2,k}$ may have different colouring, but are mutually independent.
A possible choice is now:

$$\left.\begin{aligned}\Omega(v,y_1,\hat{g}) &= [V \mid Y_1 \mid \hat{G}] \\ Z_1 &= [\hat{U}_2 \mid z^{-\tau} Y_2 \mid z^{-\tau'}(Y_1 - Y_2)] \\ \hat{\underline{g}} &= \underline{y}_1 - \Omega(v,y_1)\hat{\underline{\theta}}^*\end{aligned}\right] \qquad (5.30)$$

where $\hat{\underline{u}}_2$ is generated from $\underline{y}_2$.
This is an interesting choice as the conditions (5.25) can be fulfilled quite easily.

The requirement for a second independent measurement of the input- or output signal can be weakened slightly if we recognize its purpose, which is removal of those components which are related to u_k, in the IV quantity. For this we do not exclusively need measurements of the input-or output signal, but also signals which are closely related to the input- and output signals can be used, e.g. an electrical signal, which controls a valve, instead of the flow itself. Depending on the circumstances, these signals may be readily available, which extends the applicability of the method of eq. (5.30) considerably.

As mentioned already, an iterative or recursive version of this esti-

mator has to be chosen, as the quantity $\underline{g}$ is not available. The formulation of the recursion for the k-th step is as follows:

$$\left.\begin{aligned} \hat{\underline{\theta}}'_k &= \hat{\underline{\theta}}'_{k-1} + P_{k-1}\,\underline{z}_k(1+\underline{\omega}_k^T P_{k-1}\underline{z}_k)^{-1}(y_k - \underline{\omega}_k^T \hat{\underline{\theta}}'_{k-1}) \\ P_k &= P_{k-1} - P_{k-1}\,\underline{z}_k\,(1+\underline{\omega}_k^T P_{k-1}\underline{z}_k)^{-1}\,\underline{\omega}_k^T P_{k-1} \end{aligned}\right] \qquad (5.31)$$

where, for the case of one input- and two output measurements:

$$\left.\begin{aligned} \underline{\omega}_k^T &= (v_k,\ldots,v_{k-p},\ y_{1,k-1},\ldots,y_{1,k-q},\ \hat{g}_{k-1},\ldots,\hat{g}_{k-r_o}) \\ \underline{z}_k^T &= (\hat{u}_k,\ldots,\hat{u}_{k-p},\ y_{2,k-1-\tau},\ldots,y_{2,k-q-\tau},z_{k-1-\tau},\ldots,z_{k-r_o-\tau}) \\ \hat{u}_{k-1} &= \frac{[1+\hat{A}^{k-1}]}{[\hat{b}_o+\hat{B}^{k-1}]}\,y_{2,k-1} \\ \hat{g}_{k-1} &= [1+\hat{A}^{k-1}]\,y_{1,k-1} - [\hat{b}_o+\hat{B}^{k-1}]v_{k-1} \\ z_{k-1} &= y_{1,k-1} - y_{2,k-1} \end{aligned}\right] \qquad (5.32)$$

A schematic diagram of this estimator is given in fig. 5.4.

In this figure the three IV generating filters Φ_1, Φ_2, Φ_3 for the respective IV quantities in the IV matrix Z_1 are shown. In eq. (5.26) and (5.30) $\Phi_1 = z^{-\tau'}$, $\Phi_2 = \frac{[1+\hat{A}^{k-1}]}{[\hat{b}_o+\hat{B}^{k-1}]}$ and $\Phi_3 = z^{-\tau}$ are chosen.

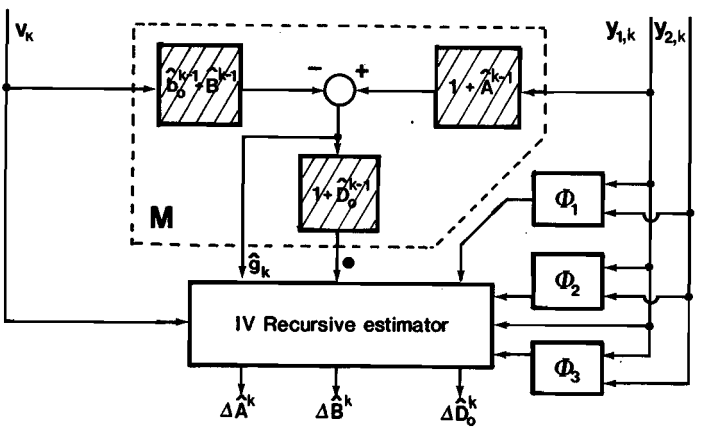


Fig. 5.4 IVEMM estimator for noisy input-output data

5.3.1b Second variant of the IOIVEMM approach

A second variant of the IVEMM estimator can be constructed when the rôles of the two operations are interchanged. The model extension will be used now to take care of the input disturbance, and the IV is used to take care of the output disturbance.

Analogous to eq. (5.20) we can write:

$$\underline{y} = \Omega(v,y)\underline{\theta}_t - G\underline{d}_{it} + \frac{[1+D_i]_t}{[1+D_o]_t}\underline{\xi}_o + \underline{\xi}_i \tag{5.33}$$

where

$$\underline{d}_{it}^T = (d_{i,1},\ldots,d_{i,r_i})_t \tag{5.34}$$

Equation (5.33) can be rewritten:

$$\underline{y} = \Omega(v,y,g)\begin{bmatrix}\underline{\theta}_t \\ -\underline{d}_{it}\end{bmatrix} + \frac{[1+D_i]_t}{[1+D_o]_t}\underline{\xi}_o + \underline{\xi}_i \tag{5.35}$$

An explicit IVEMM estimator can be given now

$$\underline{\hat{\theta}}'' = \begin{bmatrix}\underline{\hat{\theta}} \\ -\underline{\hat{d}}_i\end{bmatrix} = [Z_2^T\,\Omega(v,y,g)]^{-1} Z_2^T\,\underline{y} \tag{5.36}$$

This estimator is consistent if

$$\begin{aligned}
&\text{I} \quad \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q} Z_2^T\,\Omega(v,y,g)\right] \text{ is non-singular} \\
&\text{II} \quad \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q} Z_2^T\,\underline{\xi}_i\right] = \underline{0} \\
&\text{III} \quad \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N-q} Z_2^T \frac{[1+D_i]_t}{[1+D_o]_t}\,\underline{\xi}_o\right] = \underline{0}
\end{aligned} \tag{5.37}$$

An example of a possible choice for Z_2 is:

$$Z_2 = [z^{-\tau}v \,|\, \hat{y} \,|\, z^{-\tau'}z] =$$

$$\left[\begin{array}{ccc|ccc|ccc}
v_{q+1-\tau} & \cdot & v_{q+1-p-\tau} & \hat{y}_q & \cdots & \hat{y}_1 & z_{q-\tau'} & \cdots & z_{q+1-r_i-\tau'} \\
\cdot & & \cdot & \cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot & \cdot & & \cdot \\
v_{N-\tau} & \cdots & v_{N-p-\tau} & \hat{y}_{N-1} & \cdots & \hat{y}_{N-q} & z_{N-1-\tau'} & \cdots & z_{N-r_i-\tau'}
\end{array}\right] \tag{5.38}$$

Here the submatrix Z contains IV quantities for proper estimation of $\underline{d}_{it}$, see e.g. eq. (5.41) and (5.42).
The recursive estimator is given also by eq. (5.31), and (5.32), except the expression for the IV quantity, which is in this case:

$$\underline{z}_k^T = (v_{k-\tau},\cdots,v_{k-p-\tau},\hat{y}_{k-1},\cdots,\hat{y}_{k-q},z_{k-1-\tau'},\cdots,z_{k-r_1-\tau'}) \tag{5.39}$$

The expression for the estimated model output is:

$$\hat{\underline{y}}_{k-1} = \frac{[\hat{b}_o+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]} v_{k-1} \tag{5.40}$$

A schematic diagram for this estimator is given in fig. 5.5 for the case of two input- and one output measurements.

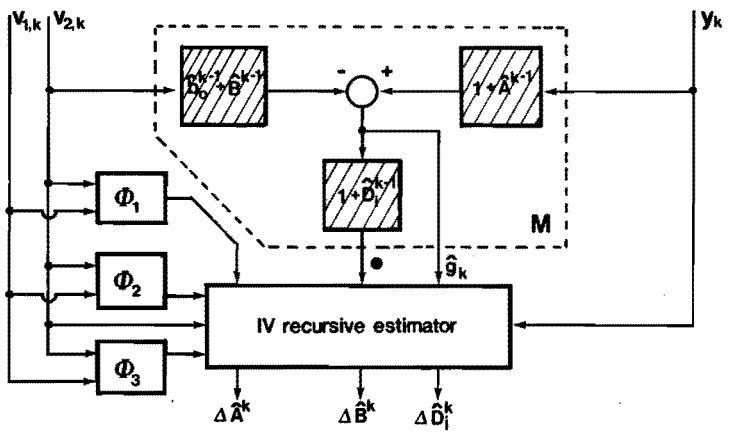


Fig. 5.5 IVEMM estimator for noisy input-output data

In eq. (5.38) $\Phi_1 = z^{-\tau'}$, $\Phi_2 = \frac{[\hat{b}_o+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$ and $\Phi_3 = z^{-\tau}$ is chosen.

For multiple measurements, as indicated in eq. (5.27) or eq. (5.29), we may have the following choices for the IV matrix:

1) two inputs v_1 and v_2, one output $\underline{y}$:

$$\left.\begin{aligned} \Omega(v_1,y,\hat{g}_1) &= [V_1|Y|\hat{G}_1] \\ Z_2 &= [z^{-\tau}V_2|\hat{Y}_2|z^{-\tau'}(V_1-V_2)] \\ \hat{\underline{g}}_1 &= \underline{y} - \Omega(v_1,y)\underline{\theta}^* \end{aligned}\right] \tag{5.41}$$

where $\hat{\underline{y}}_2$ is generated from $\underline{v}_2$.

2) one input $\underline{v}$, two outputs $\underline{y}_1$ and $\underline{y}_2$

$$\left.\begin{aligned} \Omega(v,y_1\hat{g}_1) &= [V|Y_1|\hat{G}_1] \\ Z_2 &= [\hat{U}_2|Y_2|\hat{G}_2] \\ \hat{\underline{g}}_2 &= \underline{y}_2 - \Omega(v,y_2)\hat{\underline{\theta}}^* \end{aligned}\right\} \qquad (5.42)$$

5.3.2 Filtering and IV-approach combined

The next combination of two basic operations is the combination of filtering (to obtain white residuals) and the IV-approach. For the same reasons as in the previous paragraph, we can find two alternatives for this scheme as well.

5.3.2a First variant

For the first variant we choose the IV operation to tackle the input disturbance and the filtering to take care of the output disturbance.

Starting with eq. (5.8) we can apply filtering of the signals involved by using a filter $[1+D_o]_t$

$$\left.\begin{aligned} \tilde{\underline{y}}_k &= [1+D_o]_t y_k \\ \tilde{\underline{v}}_k &= [1+D_o]_t v_k \end{aligned}\right\} \qquad (5.43)$$

Then (5.4) can be rewritten

$$\tilde{\underline{y}} = \Omega(\tilde{v},\tilde{y})\ \underline{\theta}_t + \underline{\xi}_o + \tilde{\underline{f}} \qquad (5.44)$$

where

$$\tilde{\underline{f}} = \frac{[1+D_o]_t}{[1+D_i]_t}\,\underline{\xi}_i \qquad (5.45)$$

Now, take as IV estimator for the process parameters:

$$\hat{\underline{\theta}} = [Z_3^T\ \Omega(\tilde{v},\tilde{y})]^{-1} Z_3^T\ \tilde{\underline{y}} \qquad (5.46)$$

This estimator is consistent if:

$$\left.\begin{array}{lll} \text{I} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_3^T \, \Omega(\tilde{v},\tilde{y})\right] & \text{is non-singular} \\ \text{II} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_3^T \, \underline{\xi}_o\right] = \underline{0} & \\ \text{III} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_3^T \, \tilde{\underline{f}}\right] = \underline{0} & \end{array}\right] \qquad (5.47)$$

As possible choices there are:

$$\left.\begin{array}{l} Z_3 = [\hat{U} \mid Y] \\ Z_3 = [\tilde{\hat{U}} \mid \tilde{Y}] \end{array}\right] \qquad (5.48)$$

where the submatrix $\hat{U}$ can be constructed using as input, the output signal y and a fixed or varying filter, e.g. previous inverse model of the process; cf. eq. (5.32) for the recursive case.

For proper application of this algorithm the parameters of the AR filter $[1+D_i]_t$ are needed. They can be obtained by using a separate estimator. For this purpose we use the result of eq. (5.17):

$$\underline{g} = -G\underline{d}_{ot} + \underline{\xi}_o + \frac{[1+D_o]_t}{[1+D_i]_t} \underline{\xi}_i \qquad (5.49)$$

The idealized IV estimator for the AR parameters of the output noise would be:

$$\underline{\hat{d}}_o = -[Z_4^T G]^{-1} Z_4^T \underline{g} \qquad (5.50)$$

In practice, the precise signals $\underline{g}$ (and G) are not available, so therefore they will be estimated. The estimates $\hat{\underline{g}}$ will be obtained by using eq. (5.32); cf. also eq. (5.57). The resulting estimator will then be:

$$\underline{\hat{d}}_o = -[Z_4^T \hat{G}]^{-1} Z_4^T \hat{\underline{g}} \qquad (5.51)$$

This estimator is consistent if the following three conditions are fulfilled:

$$\left.\begin{array}{lll} \text{I} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_4^T \, G\right] & \text{is non-singular} \\ \text{II} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_4^T \, \underline{\xi}_o\right] = \underline{0} & \\ \text{III} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_4^T \frac{[1+D_o]_t}{[1+D_i]_t} \underline{\xi}_i\right] = \underline{0} & \end{array}\right] \qquad (5.52)$$

A possible choice for Z_4 is:

$$Z_4 = Y \tag{5.53}$$

Y having proper dimensions (N-q) x r_o. As both estimators (5.46) and (5.50) use each other's results, an iterative or a recursive scheme has to be developed.

For a recursive scheme the formulation is as follows:

<u>step k</u>

a) Filter the new samples v_k, y_k using the estimates $\underline{\hat{d}}_o^{k-1}$ of the previous recursion step.

$$\left.\begin{aligned} \tilde{y}_k &= [1+\hat{D}_o^{k-1}]\, y_k \\ \tilde{v}_k &= [1+\hat{D}_o^{k-1}]\, v_k \end{aligned}\right] \tag{5.54}$$

b) Form the vectors $\underline{\tilde{\omega}}_k$ and $\underline{\tilde{z}}_{3,k}$:

$$\left.\begin{aligned} \underline{\tilde{\omega}}_k &= (\tilde{v}_k,\ldots,\tilde{v}_{k-p},\ \tilde{y}_{k-1},\ldots,\tilde{y}_{k-q}) \\ \underline{\tilde{z}}_{3,k} &= (\tilde{u}_k,\ldots,\tilde{u}_{k-p},\tilde{y}_{k-1},\ldots\tilde{y}_{k-q}) \end{aligned}\right] \tag{5.55}$$

c) Apply the recursive algorithm

$$\left.\begin{aligned} \underline{\hat{\theta}}_k &= \underline{\hat{\theta}}_{k-1} + P_{k-1}\,\underline{\tilde{z}}_{3,k}(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{z}}_{3,k})^{-1}(\tilde{y}_k - \underline{\tilde{\omega}}_k^T\underline{\hat{\theta}}_{k-1}) \\ P_k &= P_{k-1} - P_{k-1}\,\underline{\tilde{z}}_{3,k}\,(1+\underline{\tilde{\omega}}_k^T P_{k-1}\underline{\tilde{z}}_{3,k})^{-1}\underline{\tilde{\omega}}_k^T\, P_{k-1} \end{aligned}\right] \tag{5.56}$$

d) Generate

$$\hat{\tilde{g}}_k = [1+\hat{A}^k]y_k - [b_o^k+\hat{B}^k]v_k \tag{5.57}$$

e) Form

$$\left.\begin{aligned} \underline{\hat{\tilde{g}}}_k &= (\hat{\tilde{g}}_{k-1},\ldots,\hat{\tilde{g}}_{k-r_o}) \\ \underline{z}_{4,k} &= (y_{k-1},\ldots,y_{k-r_o}) \end{aligned}\right] \tag{5.58}$$

f) Apply the recursive algorithm

$$\left.\begin{aligned} \underline{\hat{d}}_{o,k} &= \underline{\hat{d}}_{o,k-1} + Q_{k-1}\underline{z}_{4,k}(1+\underline{\hat{\tilde{g}}}_k^T Q_{k-1}\underline{z}_{4,k})^{-1}(\hat{\tilde{g}}_k - \underline{\hat{\tilde{g}}}_k^T\underline{\hat{d}}_{o,k-1}) \\ Q_k &= Q_{k-1} - Q_{k-1}\underline{z}_{4,k}(1+\underline{\hat{\tilde{g}}}_k^T Q_{k-1}\underline{z}_{4,k})^{-1}\underline{\hat{\tilde{g}}}_k^T\, Q_{k-1} \end{aligned}\right] \tag{5.59}$$

A schematic diagram of this recursive estimator is given in fig. 5.6 for the case of one input- and two output measurements. In this diagram the IV generating filters Φ_1 and Φ_2 are shown.

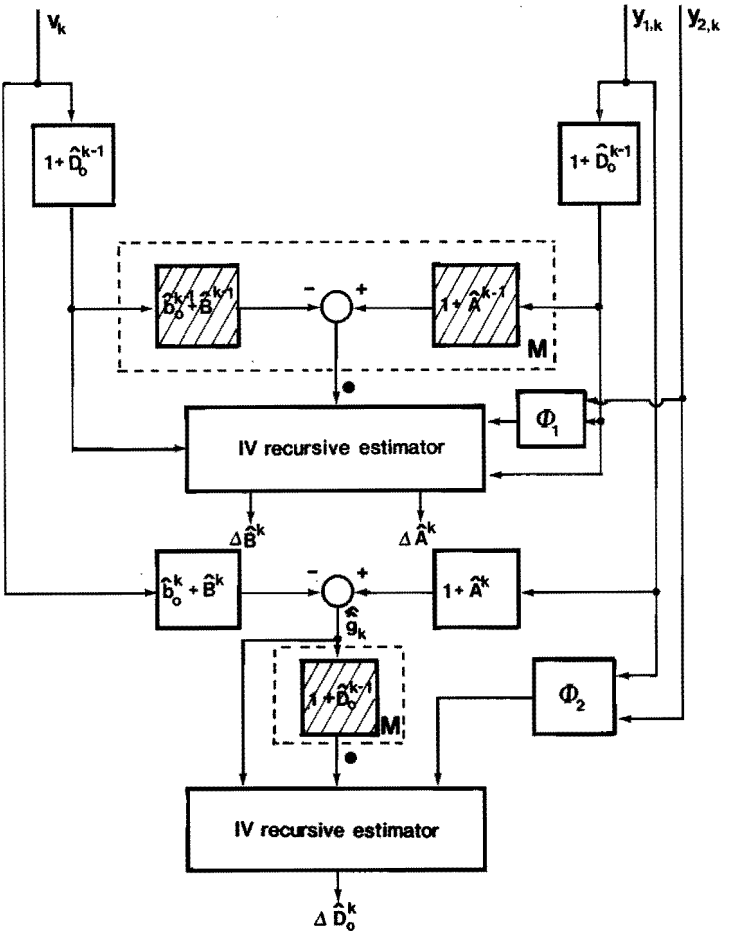


Fig. 5.6 First variant of filtering - IV combination

The choices for these filters are resp. $\Phi_1 = \frac{[1+\hat{A}^{k-1}]}{[\hat{b}_o+\hat{B}^{k-1}]}$ in eq. (5.48) and $\Phi_2 = 1$ in eq. (5.53).

Using this estimator, only the AR parameters of the output noise will be estimated. If for some reason one is interested in the AR parameters of the input noise, a similar approach can be followed, leading to an extra, separate estimator for these parameters. For details see the next section, where these parameters are estimated separately.

If multiple measurements are available we may have the following choices for the IV matrix:

1) two inputs, one output

$$\left.\begin{aligned} \Omega(\tilde{v}_1,\tilde{y}) &= [\tilde{V}_1|\tilde{Y}] \\ Z_3 &= [\tilde{V}_2|\tilde{Y}] \end{aligned}\right] \qquad (5.60)$$

2) one input, two outputs

$$\left.\begin{aligned} \Omega(\tilde{v},\tilde{y}_1) &= [\tilde{V}|\tilde{Y}_1] \\ Z_3 &= [\tilde{U}_2|\tilde{Y}_2] \end{aligned}\right] \qquad (5.61)$$

where $\hat{U}_2$ is generated using $\underline{y}_2$.

5.3.2b Second variant

A second variant can be constructed if we interchange the rôles of the filtering and the IV operation. The measurables will now be filtered by the filter $[1+D_i]_t$:

$$\left.\begin{aligned} \tilde{\tilde{y}}_k &= [1+D_i]_t y_k \\ \tilde{\tilde{v}}_k &= [1+D_i]_t v_k \end{aligned}\right] \qquad (5.62)$$

Then (5.8) can be rewritten:

$$\tilde{\tilde{y}} = \Omega(\tilde{\tilde{v}},\tilde{\tilde{y}})\underline{\theta}_t + \tilde{\tilde{e}} + \underline{\xi}_i \qquad (5.63)$$

where

$$\tilde{\tilde{e}} = \frac{[1+D_i]_t}{[1+D_o]_t}\underline{\xi}_o \qquad (5.64)$$

The IV estimator for the process parameters is

$$\hat{\underline{\theta}} = [Z_5^T\ \Omega(\tilde{\tilde{v}},\tilde{\tilde{y}})]^{-1} Z_5^T\ \tilde{\tilde{y}} \qquad (5.65)$$

This estimator is consistent, if:

$$\left.\begin{aligned} &\text{I} \quad \operatorname*{plim}_{N\to\infty} \left[\frac{1}{N-q} Z_5^T\ \Omega(\tilde{\tilde{v}},\tilde{\tilde{y}})\right] \quad \text{is non-singular} \\ &\text{II} \quad \operatorname*{plim}_{N\to\infty} \left[\frac{1}{N-q} Z_5^T\ \underline{\xi}_i\right] = \underline{0} \\ &\text{III} \quad \operatorname*{plim}_{N\to\infty} \left[\frac{1}{N-q} Z_5^T\ \tilde{\tilde{e}}\right] = \underline{0} \end{aligned}\right] \qquad (5.66)$$

The recursive version proceeds along the lines indicated for the first variant. We will not explain this procedure in detail here. The schematic diagram is given in fig. 5.7, for the case of two input- and one output measurements.

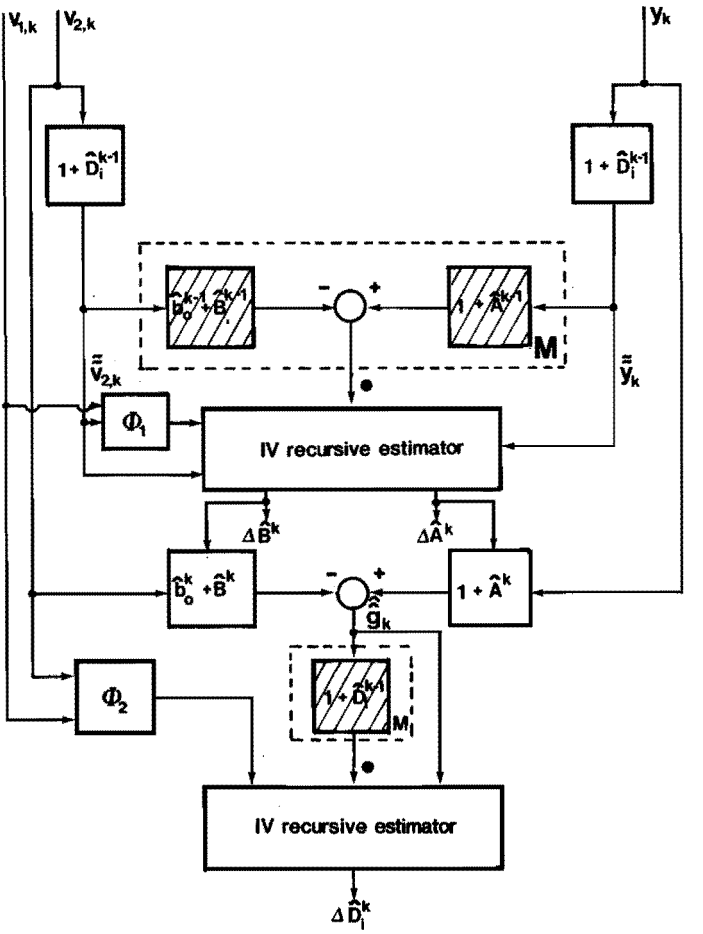


Fig. 5.7 Second variant of filtering-IV combination

Some possible choices are

$$\left.\begin{aligned} z_5 &= [v \mid \hat{Y}] \\ z_5 &= [\tilde{\tilde{v}} \mid \tilde{\tilde{Y}}] \end{aligned}\right] \qquad (5.67)$$

the IV estimator for the AR parameters of the input noise is

$$\underline{\hat{d}}_i = -[Z_6\hat{\underline{G}}]^{-1} Z_6^T\hat{\underline{\xi}} \tag{5.68}$$

where $\hat{\underline{\xi}}$ is again generated using previous estimates $\hat{\underline{\theta}}$. This estimator is consistent if the following conditions are fulfilled:

$$\left.\begin{array}{lll} \text{I} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_6^T G\right] & \text{is non-singular} \\ \text{II} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_6^T \underline{\xi}_i\right] = \underline{0} & \\ \text{III} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N-q} Z_6^T \frac{[1+D_o]_t}{[1+D_i]_t} \underline{\xi}_o\right] = \underline{0} & \end{array}\right\} \tag{5.69}$$

A possible choice for Z_6 is

$$Z_6 = V \tag{5.70}$$

The IV generating filters Φ_1 and Φ_2 in this diagram are chosen as resp. $\Phi_1 = \dfrac{[\hat{b}_o+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$ in eq. (5.67) and $\Phi_2 = 1$ in eq. (5.70).

5.4 Conclusions

In this chapter we have proposed several related estimators for situations where both input- and output measurements are corrupted by mutually independent additive disturbances.
These estimators are involved IV estimators,, and they combine the IV approach with the filtering approach or with the model extension approach. Therefore also AR noise parameters are estimated along with the process parameters. The interesting feature of these estimators is the fact that no rather restrictive assumptions, such as equal colouring for the input- and the output disturbance, need to be made.

The selection of a proper IV quantity may cause problems for the convergence of the estimator. If additional measurements of the input signal or the output signal are available, or a signal which is closely related to one of those signals but independent of the dis-

turbances, then proper selection of the IV quantity is possible. This is, in many practical situations, a less restrictive experimental condition than the necessary a priori knowledge of existing estimation schemes.

CHAPTER SIX:

EXPERIMENTAL RESULTS

6.1 Introduction

In previous chapters we have investigated the asymptotic properties of the estimators, viz. bias and variance. Often, for practical applications, it is important to know - or at least to have an idea - whether these asymptotic properties are also attained for finite, and often for very small, sample size. At this moment a theory which describes this small sample behaviour is lacking and its development and potential result cannot be predicted, but in any case the development of such a theory will need considerable effort. Therefore the experimenter who applies parameter estimation methods in a practical situation has to have experience with these methods. This experience should be gained from other (previous) experiments and from simulation activities which are related to the problem at hand. Many simulation runs, whose conditions are based on those of the practical experiment under study, can then give insight into which estimation properties may be expected. The availability of an interactive package, as described in the next paragraph, will be of great help.

In this chapter we will discuss several properties of estimation schemes, and we will draw conclusions which are based on practical experience. The bias and variance properties will be investigated in paragraphs 6.3 and 6.4, whereas the divergence aspects are considered in paragraph 6.5. In paragraph 6.6 we will present simulation results of estimators for input-output corrupted measurements as dealt with in chapter 5.

6.2 The interactive program package SATER

Interactive program packages become more and more standard tools in control research and development. The availability of such a package has several advantages: no tedious programming work, the time

available can be devoted to the control or estimation problem, students can gain experience by using these programs, several (estimation) methods can be compared in a universal framework, etc.; cf. Lemmens and Van den Boom (1977). Many interactive packages which have been developed recently, are more or less machine or application oriented. With the development of SATER this drawback was avoided as much as possible; cf. Van den Boom and Lemmens (1977).

The SATER package has primarily been developed to make the well-known estimation and order test routines available to a large group of users. The package can be used to become familiar with these techniques using simulated data or (for analysis of real processes) using recorded data.
Due to the interactive character of the program package, it is desirable that the computer has quick response times. Usually this can only be achieved satisfactorily using a dedicated minicomputer. In our system we now use a PDP 11/60 computer with RSX 11, 32k core partitions and a RK05 back-up memory. Other peripherals are a visual display unit TEKTRONIX 4014, an LA36 Decwriter and an LPS analog to digital converter.

As an interactive program contains a large number of program statements (e.g. about 10 000 FORTRAN), it cannot be loaded into the core at once, so the core memory has to be loaded by the proper software parts from the back-up memory. This back-up memory is also used for the storage of the internal datasets (like signal samples, Bode plots, estimated parameters etc.), for the messages and questions which appear on the screen and for the tables which determine the relationships between the operations and datasets, as will be explained later. In this paragraph we will only highlight the main features of the SATER system. For details cf. Lemmens (1979) and Bollen and Van den Boom (1982).

The software of the SATER system consists of three main parts. The supervisor, the application programs and the service routines as indicated in figure 6.1.

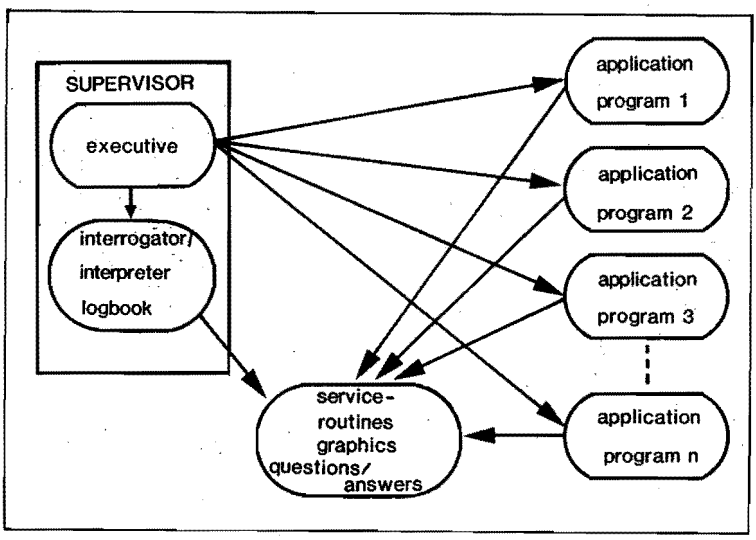


Fig. 6.1 Structure of the SATER System

The supervisor can be divided into an executive, an interrogator/ interpreter and a logbook updating device. The task of the executive is to start the right application programs indicated by the demands of the user. It checks the availability of the necessary datasets and suggests the creation of the missing datasets using other application programs. By means of the interrogator/interpreter it asks for the necessary data and parameters to run the wanted application program. After having finished that application program, it updates the logbook. This logbook contains the information concerning available datasets and how they have been created.

The application programs form the largest parts of the SATER system, but in fact, they do not contribute to the basic framework of the SATER system. The number of these application programs can be extended at will, depending on the interest of a group of users. In our SATER version, various parameter estimation routines, order test routines and basic control theory routines like Nyquist and Bode plots, as well as root loci have been implemented. Other application programs are sampling of continuous signals and file I/O.

The service routines consist of three main parts: the I/O subsystem, the question/answer subsystem (which displays the questions, accepts the answers and checks them) and the graphical subsystem (which faci-

litates the graphical display of data and the acquirement of data coordinates from the screen of the visual display).

The application programs available in the package constitute a set of algorithms/operations, operating on input data sets. The data set produced by one operation may be a necessary input data set for another operation. So a graph may be drawn representing the relationships between all data sets and all operations of the package. An example of a possible graph of this type is given in figure 6.2.

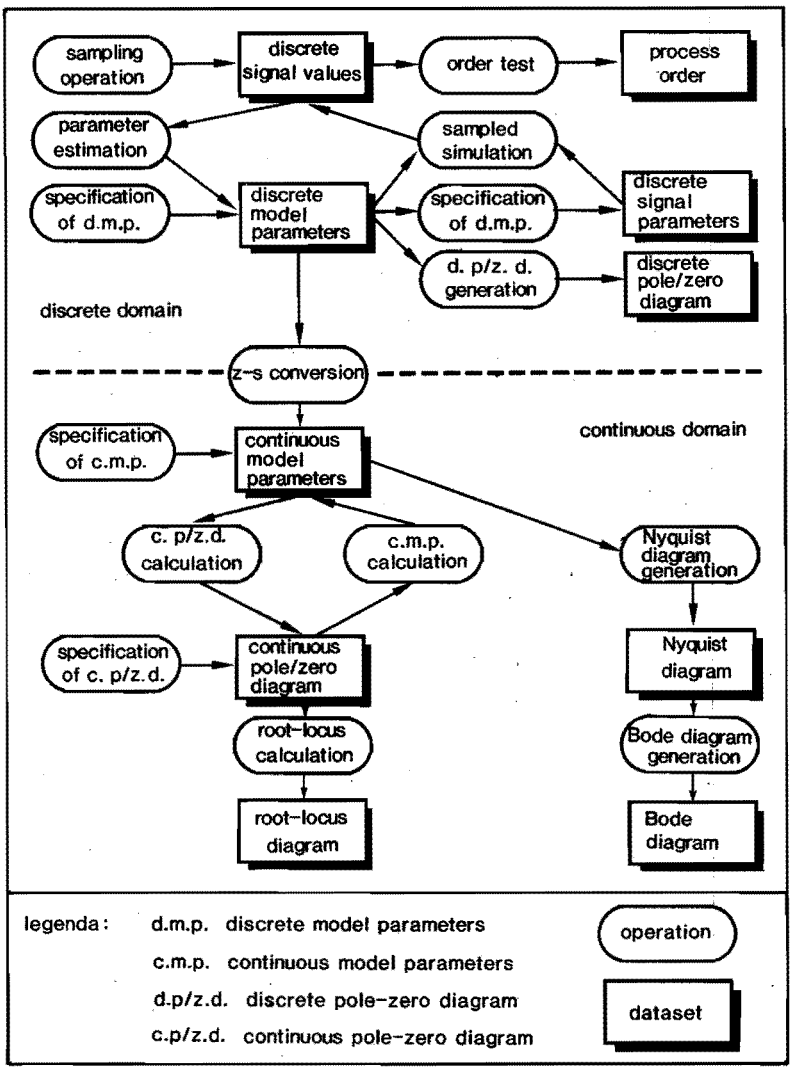


Fig. 6.2 Relations between datasets and operations

Tables are used to describe the relations in the above graph: one table for the datasets where all operations producing the same datasets have been gathered and a table of operations indicating which datasets are needed for each operation. These tables are used by the executive. In many cases more paths in the graph can be found from one dataset to another. The executive can detect these paths and can give suggestions to the user as to how to arrive at the desired dataset.

A program package like the one described here is not a rigid and fixed construction. It will be subject to continuous change, as new numerical programs become available and old programs become obsolete. By updating the tables, the structure can be modified easily.

All messages and questions that can be displayed on the screen are contained in one file on disk, the message file. This is done for two reasons: it contributes to the flexibility if messages and questions can be changed easily, without changing the program software; furthermore, it saves memory if those memory consuming texts are resident on disk.

The way in which the interaction between the machine and user has been realized, highly determines the usefulness and flexibility of the program package. Numerical programs, converting input data into output data, usually have to be controlled by a set of additional parameters. These parameters determine aspects such as, for example, which part of the input data has to be used, weighting factors of the estimation algorithm etc.
There are two ways of collecting these data: the question and answer method and the command language. The question and answer method, where questions are successively put to the user and answers gathered and checked for their correctness, is a straightforward method which is specially suited for non-experienced users. On the other hand, the interpretive command language puts obstacles in the path of the user because commands and the syntax of the command language have to be learned. Experienced users can mostly operate faster with the latter type of interaction.

The SATER system is a sort of intermediate between these two possibilities. It has a few commands: for asking for further explanation, for return to some starting points and for closing the sessions. The parameters of the numerical programs are clustered in blocks of related questions, which have pre-programmed default answers. It is important that these default answers are correct answers for the majority of uses. So the user should only change those default answers to suit his own needs if he is unable to make use of them.
If his answer is beyond predefined limits, the system will not accept it and will ask for a new answer. It is not possible to force the system to execute a numerical program when the values of the control parameters are beyond the predefined limits. All these aspects are covered by a question-answer subsystem. By calling a few subroutines, the questioning, gathering of data and checking for correctness is done automatically. Another aspect of interaction is the choice of the operation that has to be performed or the datasets that have to be created. For operations the system accepts a code (which is the number of the operation) and for datasets a name which is a standard name for the user (e.g. NYQUIST DIAGRAM). Synonyms are also allowed (e.g. POLAR-FIGURE). If the user is not familiar with the codes of the operations he can give the name of the wanted dataset. The system responds with a list of all operations and their codes producing this dataset.
If the necessary input datasets are missing, the system will detect this and give a summary of all operations capable of producing the missing datasets. For the recognition of dataset names a vocabulary is used, containing all relevant words and their synonyms; cf. Escher (1980).
For inexperienced users a HELP facility is provided which gives additional explanation or suggestions if a question is not well understood. By typing a "?" instead of the expected answer, the HELP mechanism is activated and a message containing helpful information is displayed. The contents of all HELP messages is also stored separately on disk so that changes and additions can be made very easily.

In the following paragraphs we will proceed with some examples of experimental results, obtained by using SATER.

6.3 Bias for finite sample size

The behaviour of the bias is investigated for a first order process; corrupted by first order ARMA noise e_k:

$$\left.\begin{aligned} y_k + 0.7\, y_{k-1} &= u_k + e_k \\ e_k - 0.5\, e_{k-1} &= \xi_k + 0.3\, \xi_{k-1} \end{aligned}\right] \qquad (6.1)$$

where $\{u_k\}$ and $\{\xi_k\}$ are white and gaussian, with N(0,1) and $N(0, \sigma_\xi^2)$ resp., and are mutually independent. For $\sigma_\xi^2 = 0.25$ and $\sigma_\xi^2 = 4$ the following results were obtained for the means of the estimates and the 95 per cent confidence interval using 10 runs EMM.

N	$\hat{b}_o$	$\hat{a}_1$	$\hat{c}_1$	$\hat{d}_1$
125	0.999 ± 0.026	0.684 ± 0.019	0.317 ± 0.116	-0.429 ± 0.107
250	0.995 ± 0.013	0.692 ± 0.010	0.231 ± 0.050	-0.543 ± 0.049
500	0.992 ± 0.013	0.700 ± 0.011	0.272 ± 0.029	-0.513 ± 0.036
1000	1.004 ± 0.013	0.698 ± 0.008	0.285 ± 0.033	-0.508 ± 0.024
2000	0.998 ± 0.009	0.702 ± 0.005	0.298 ± 0.026	-0.501 ± 0.023
true	1.000	0.700	0.300	-0.500

Table 6.1 Means and 95 per cent confidence intervals;

$\sigma_\xi^2 = 0.25$, S/N = 8.38 dB

125	1.018 ± 0.102	0.474 ± 0.118	-0.055 ± 0.171	-0.593 ± 0.134
250	0.950 ± 0.089	0.546 ± 0.163	0.105 ± 0.165	-0.640 ± 0.224
500	0.997 ± 0.036	0.566 ± 0.154	0.170 ± 0.104	-0.460 ± 0.172
1000	1.024 ± 0.057	0.663 ± 0.049	0.207 ± 0.095	-0.529 ± 0.041
2000	0.996 ± 0.050	0.680 ± 0.038	0.232 ± 0.087	-0.530 ± 0.037
4000*	0.994 ± 0.052	0.706 ± 0.018	0.302 ± 0.027	-0.498 ± 0.023
8000**	0.993 ± 0.021	0.705 ± 0.031	0.305 ± 0.033	-0.497 ± 0.012
true	1.000	0.700	0.300	-0.500

Table 6.2 Means and 95 per cent confidence intervals;

$\sigma_\xi^2 = 4$, S/N = -3.66 dB

The means and confidence intervals for N = 4000 and N = 8000 were obtained from 6 resp. 3 runs EMM; this is denoted in tables by * and **. From these tables it follows that the hypothesis that the true parameters can be accepted as mean values with a risk of 5 per cent is quite acceptable in all cases except for the encircled ones, which occur for relatively small sample sizes.
This is a remarkable result if we consider the actual signal-to-noise ratios. As for the given experimental condition:

$$\sigma_y^2 = 1.9608\ \sigma_u^2 + 1.1397\ \sigma_\xi^2 \qquad (6.2)$$

we find for $\sigma_\xi^2 = 0.25$: S/N = 8.38 dB

$\sigma_\xi^2 = 4$: S/N = -3.66 dB

In order to gain more experience, we present in fig. 6.3 and fig. 6.4 the adjustments of 10 independent runs of 2000 samples each, starting value $\hat{\underline{\theta}} = \underline{0}$ for each run, for the cases $\sigma_\xi^2 = 0.25$ (fig. 6.3) and $\sigma_\xi^2 = 4$ (fig. 6.4). The true parameter values are also given with the figures. It can be observed that for $\sigma_\xi^2 = 0.25$ reasonably fast and smooth adjustments for the process parameters occur in all 10 cases. The noise parameters which were estimated after 20 recursions of the process parameters show a much slower convergence. However, it can be observed that the convergence takes place from both sides of the actual parameter values.

From fig. 6.4 it is clear that much slower convergence occurs for this lower S/N ratio. In fig. 6.5 and fig. 6.6 it is shown that for this S/N ratio correct convergence will occur for increased sample lengths (N = 4000 and N = 8000 respectively).

For low S/N, the observed output signal y_k is mainly determined by the noise ξ_k. This noise is transferred to the output through a system which consists of two poles, determined by the parameters a_1 and d_1 and one zero. As these parameters are estimated separately from each other in the EMM algorithm, the true numerical values could be interchanged for small sample sizes, i.e. $\hat{a}_1$ estimates d_1 and $\hat{d}_1$ estimates a_1. This can be observed a few times in the fig. 6.4 and 6.5 and very nicely in the third adjustment of fig.

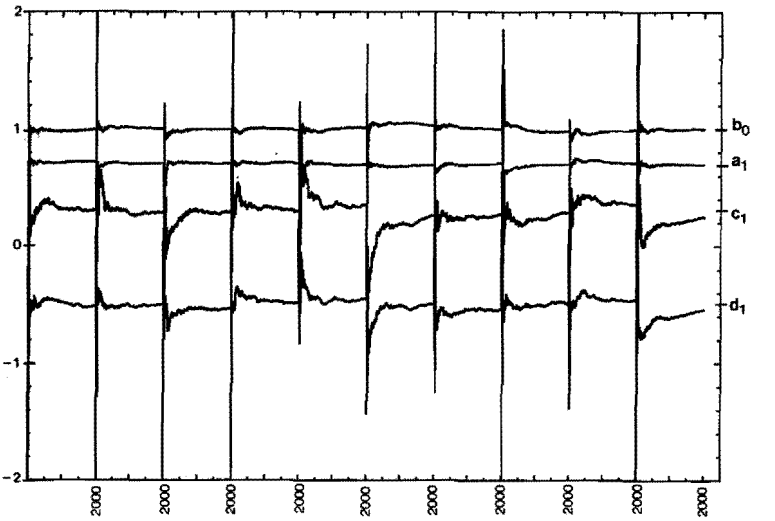


Fig. 6.3 Parameter-adjustment of 10 independent runs;
S/N = 8.38 dB, 2000 samples each

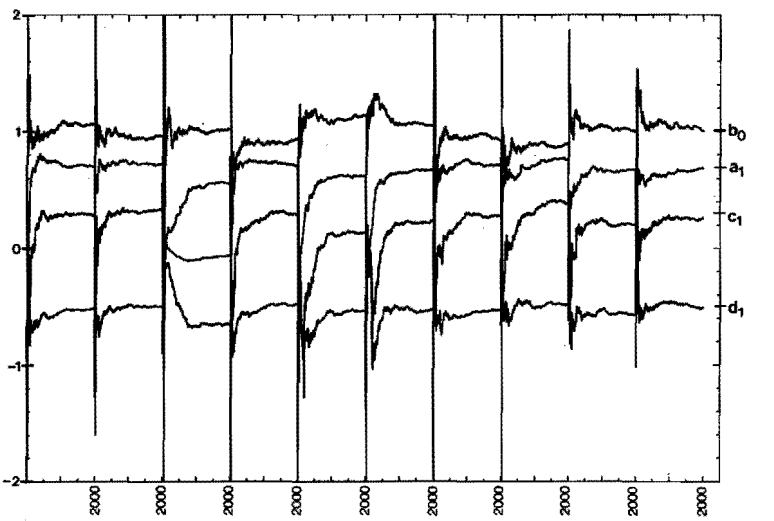


Fig. 6.4 Parameter-adjustment of 10 independent runs;
S/N = -3.66 dB, 2000 samples each

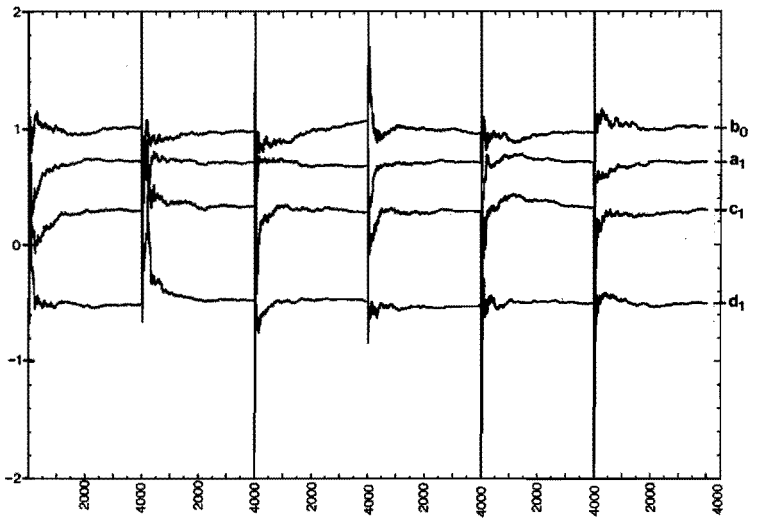


Fig. 6.5 Parameter-adjustment of 6 independent runs;
S/N = -3.66 dB, 4000 samples each

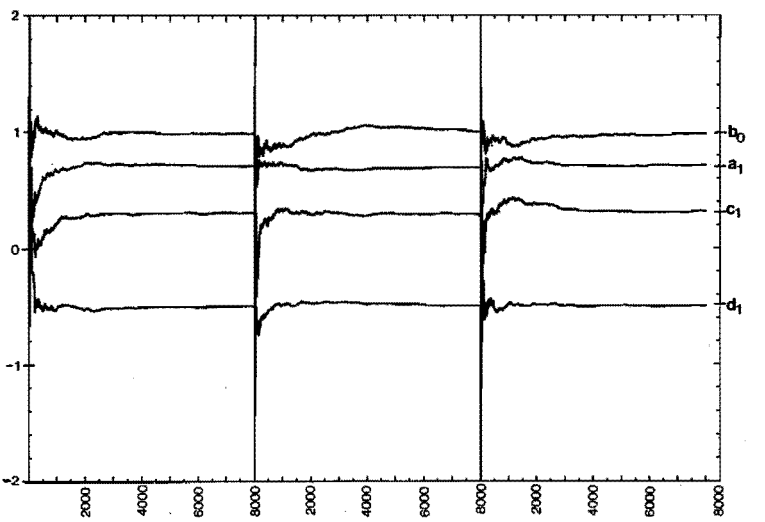


Fig. 6.6 Parameter-adjustment of 3 independent runs;
S/N = -3.66 dB, 8000 samples each

6.7, where three adjustments are shown for $\sigma_\xi^2 = 16$, i.e. S/N ratio = -9.68 dB. In such a case, the presence of the small influence of the input signal in the output, nevertheless, finally forces the interchange of the roles of $\hat{a}_1$ and $\hat{d}_1$. The adjustment of c_1 is hereby affected.

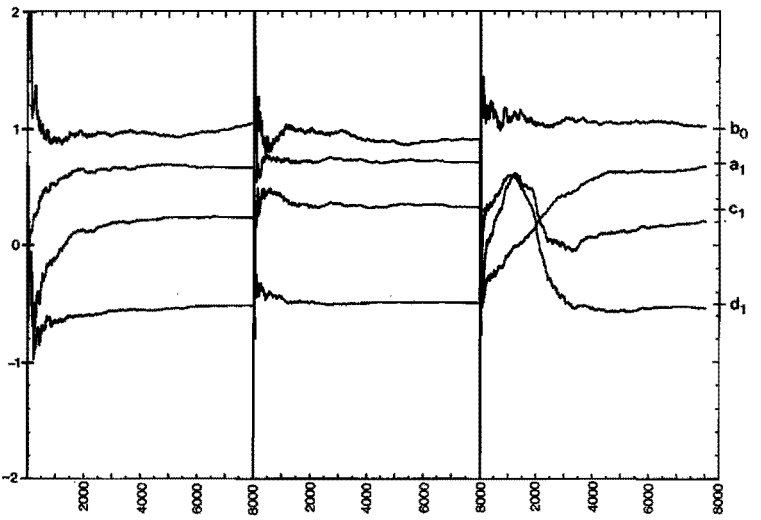


Fig. 6.7 Parameter-adjustment of 3 independent runs; S/N = -9.68 dB

From comparison of fig. 6.3, 6.4 and 6.7 it is clear that the convergence depends on the S/N ratio. This convergence can be accellerated by imposing a weighting on the observables, but this will increase the variance of the estimates; cf. Goedheer (1976).

A further example, of different character, concerns the following process:

$$y_k - 0.6y_{k-1} = u_k + e_k \qquad (6.3)$$

and the disturbing noise dynamics of the previous example

$$e_k - 0.5e_{k-1} = \xi_k + 0.3\xi_{k-1} \qquad (6.4)$$

The signals $\{u_k\}$ and $\{\xi_k\}$ are white and uniformly distributed, between $(-1,1)$ and $(-\lambda,\lambda)$ respectively. For $\lambda = 0.25$ and $\lambda=1$ the following results were obtained for the means and the 95 per cent confidence intervals, using 10 runs EMM.

N	$\hat{b}_o$	$\hat{a}_1$	$\hat{c}_1$	$\hat{d}_1$
125	0.999 ± 0.018	-0.615 ± 0.035	0.220 ± 0.089	-0.558 ± 0.084
250	1.002 ± 0.012	-0.617 ± 0.021	0.264 ± 0.061	-0.510 ± 0.062
500	1.003 ± 0.007	-0.609 ± 0.013	0.293 ± 0.064	-0.506 ± 0.052
1000	1.002 ± 0.005	-0.604 ± 0.010	0.287 ± 0.040	-0.513 ± 0.036
2000	1.002 ± 0.003	-0.606 ± 0.008	0.292 ± 0.022	-0.505 ± 0.017
true	1.000	-0.600	0.300	-0.500

Table 6.3 Means and 95 per cent confidence intervals;
$\lambda = 0.25$, S/N = 5.97 dB

N	$\hat{b}_o$	$\hat{a}_1$	$\hat{c}_1$	$\hat{d}_1$
125	1.007 ± 0.059	-0.712 ± 0.063	0.262 ± 0.108	-0.437 ± 0.122
250	1.026 ± 0.054	-0.652 ± 0.061	0.301 ± 0.106	-0.454 ± 0.098
500	1.006 ± 0.029	-0.674 ± 0.029	0.252 ± 0.053	-0.462 ± 0.048
1000	1.996 ± 0.030	-0.624 ± 0.023	0.277 ± 0.039	-0.504 ± 0.033
2000	0.995 ± 0.012	-0.619 ± 0.020	0.287 ± 0.029	-0.500 ± 0.028
true	1.000	-0.600	0.300	-0.500

Table 6.4 Means and 95 per cent confidence intervals;
$\lambda = 1$, S/N = -6.07 dB

From these tables it follows that the hypothesis that the true parameters can be accepted as mean values with a risk of 5 per cent is acceptable in all cases, except the encircled ones. As a practical result from these tables and the two tables of the previous example, we may find that the small sample behaviour of the estimates, with respect to the bias behaviour, is quite good. However, it is dependent on the signal-to-noise ratio.

As a third example, we consider the bias for different types of models for the following second order process:

$$\left.\begin{aligned} y_k - 1.5y_{k-1} + 0.7y_{k-2} &= u_{k-1} + 0.5u_{k-2} + e_k \\ e_k - 0.5e_{k-1} &= \xi_k + 0.3\,\xi_{k-1} \end{aligned}\right] \qquad (6.5)$$

The process is identical to that proposed by Åström (1968) as test process; the noise filter is identical to that propsed by Talmon and Van den Boom (1973). The corresponding pure MA viz. AR noise models are:

$$\left.\begin{aligned} &\text{MA:}\quad e_k = \xi_k + 0.8\xi_{k-1} + 0.4\xi_{k-2} + 0.2\xi_{k-3} + \ldots. \\ &\text{AR:}\quad e_k - 0.8e_{k-1} + 0.24e_{k-2} - 0.072e_{k-3} + \ldots. = \xi_k \end{aligned}\right] (6.6)$$

The input signal u_k and disturbing noise ξ_k were here independent, white noise sequences, uniformly distributed between -1 and +1. This implies that S/N = -3.7 dB. We will consider the following cases for the model:

a)	proc: ARMA-model;	no noise param:			fig. 6.8
b)	proc: ARMA-model;	MA-noise param:	1MA		fig. 6.9
			2MA		fig. 6.10
c)	proc: ARMA-model;	AR-noise param:	1AR		fig. 6.11
			2AR		fig. 6.12
d)	proc: ARMA-model;	ARMA-noise param:	1MA,1AR		fig. 6.13

In these figures the development of the average adjustment as function of N is shown. The average was taken over 10 runs. The true parameter values are also given; these figures give then an indication of the bias.

According to paragraph 3.2, it can be observed that in case a) the process parameters will be estimated with bias. The inclusion of MA-noise parameters decreases the bias, and the parameters of the pure MA-noise model are estimated satisfactorily. Nevertheless, even for two additional MA noise parameters, a remaining bias can be observed. This is due to the fact that the remaining residual has been coloured by the non-modelled part of the pure MA-noise model.

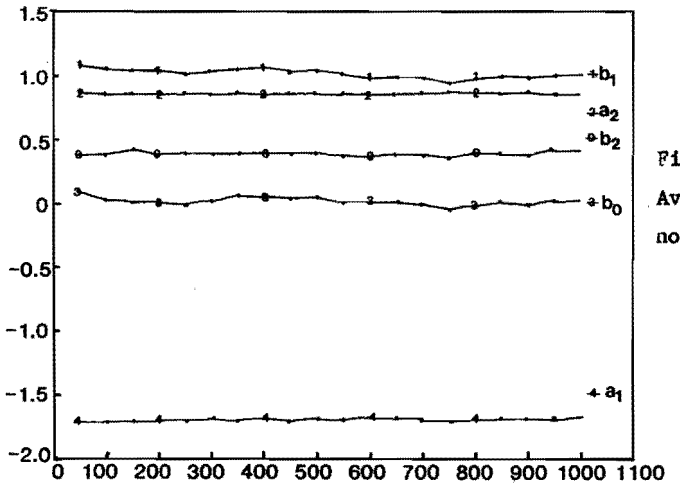


Fig. 6.8
Average adjustment;
no noise parameters

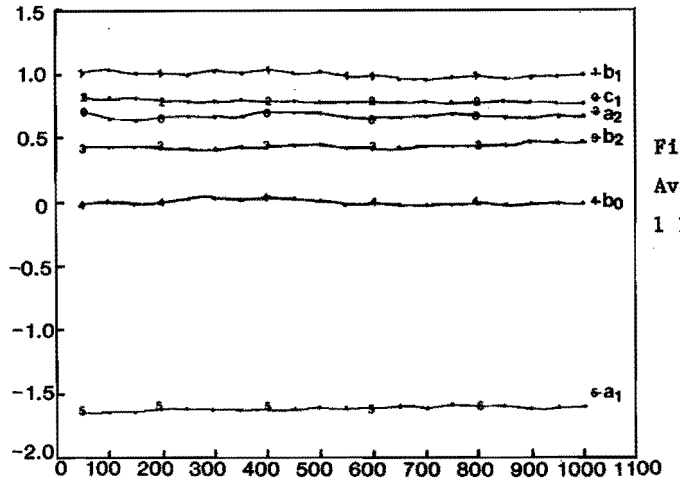


Fig. 6.9
Average adjustment;
1 MA noise parameter

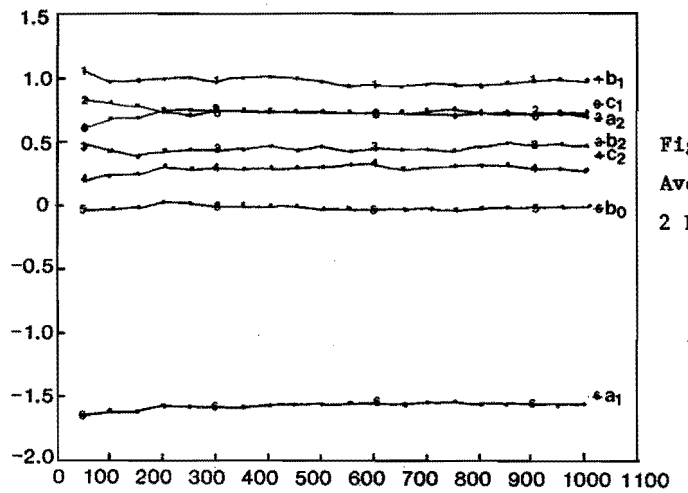


Fig. 6.10
Average adjustment;
2 MA noise parameters

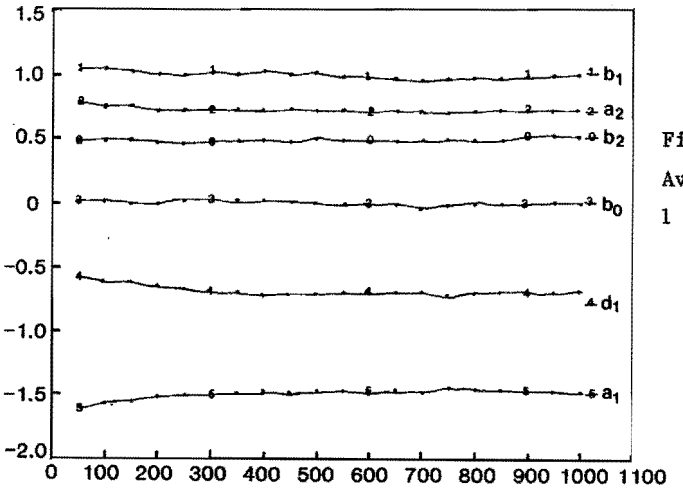


Fig. 6.11
Average adjustment;
1 AR noise parameter

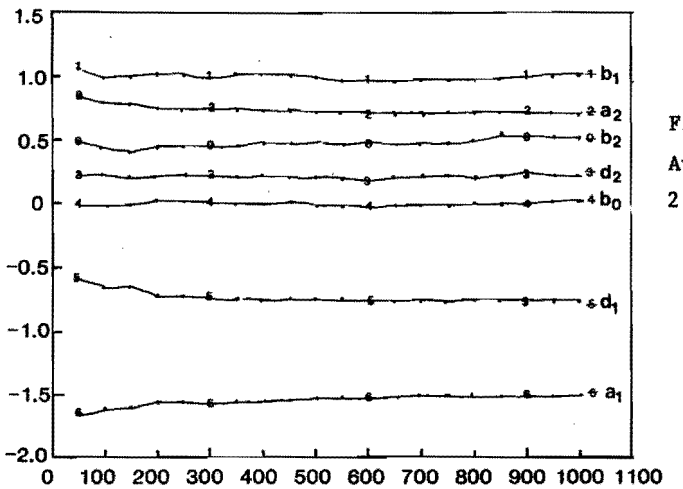


Fig. 6.12
Average adjustment;
2 AR noise parameters

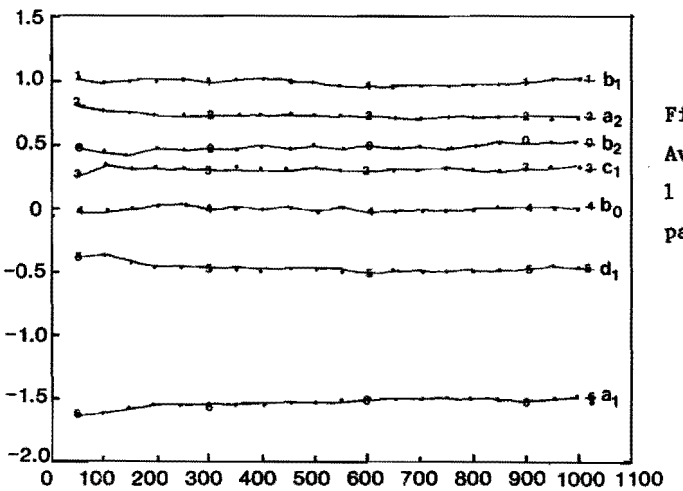


Fig. 6.13
Average adjustment;
1 MA and 1 AR noise
parameter

The inclusion of pure AR-noise parameters is more successful with respect to bias. This is because the non-modelled part of the pure AR-noise model is of less importance than in the pure MA case. This would suggest a preference for a low order AR-noise model in this case. But for another noise filter (e.g. the inverse of the considered one) a low order MA-noise filter would be preferable. Therefore ARMA noise models are more generally applicable. The results of such a noise modelling are very attractive for this specific example, as can be seen from fig. 6.13.

6.4 The variance of the estimators

For the second order process and the noise process of the previous paragraph, the behaviour of the variance is investigated. For the input signal u_k and the disturbing noise ξ_k, white gaussian noise was taken with N(0,1) and $N(0,\lambda^2)$ statistics, where λ ranged as 256^{-1}, 64^{-1}, 16^{-1}, 4^{-1}, 1, 4, 16, 64.
This implies a S/N ratio ranging from 44.3 dB to -39.7 dB, which covers the majority of practical disturbance ratios. The estimation method was again EMM and N = 400. In the figures 6.14, 6.15 and 6.16 the estimates of the standard deviations (s.d.), obtained from 10 independent estimation runs, with their 95 per cent confidence intervals, are given for all estimated parameters. In the figures, the results for the different parameters are shifted over one decade in order to make comparison possible. Besides the observed standard deviations, the range of the 10 corresponding approximated Cramér-Rao bounds are given which are calculated, based on the estimated parameters.

From the figures we can observe the following:

a) the standard deviations for the different parameter estimates behave qualitatively as indicated in paragraph 3.6.

b) the standard deviations for the a parameters change in behaviour between $\lambda = 1$ and $\lambda = 4$ which is in accordance with eq. (3.108) and (3.110).

c) the observed standard deviations are close to the Cramér-Rao bound; the deviation from the Cramér-Rao bound can be explained as follows:
   - EMM, as applied, is a fast, and therefore, approximate algorithm; no recalculation of all previous prediction errors are performed.
   - 10 independent runs are used for the calculation of the standard deviation; i.e. 10 independent realizations of the input signal, which also contributes to the variance, due to the increase of degrees of freedom in the experiment.

These results suggest that EMM is a very attractive estimator, both with respect to speed and results.

Bosch (1978) compared the behaviour of different estimators of the SATER package, viz. EMM, GLS, PKF (=Tally) and three IV estimators. These three IV estimators are

1) IV-1 use of model output as instrumental variable
2) IV-2 use of shifted inputs
3) IVEMM

These results are given in fig. 6.17, 6.18 and 6.19. Note that all curves are shifted with respect to each other over one decade each time, as indicated. Bosch used white uniformly distributed signals with almost the same S/N ratios as in the previous experiment and N = 200. From these experiments we can conclude the following:

a) the standard deviations of the different parameters show, the behaviour as qualitatively indicated in paragraph 3.6., for all estimators.
b) the use of uniformly distributed signals seems not to influence the results which, also quantitatively, can be compared with the results of the previous experiment.
c) GLS and EMM give highly comparable results. We have already seen in paragraph 6.3 that the simulated noise of this example can be very well modelled by a second order AR-model so that GLS is also appropriate.

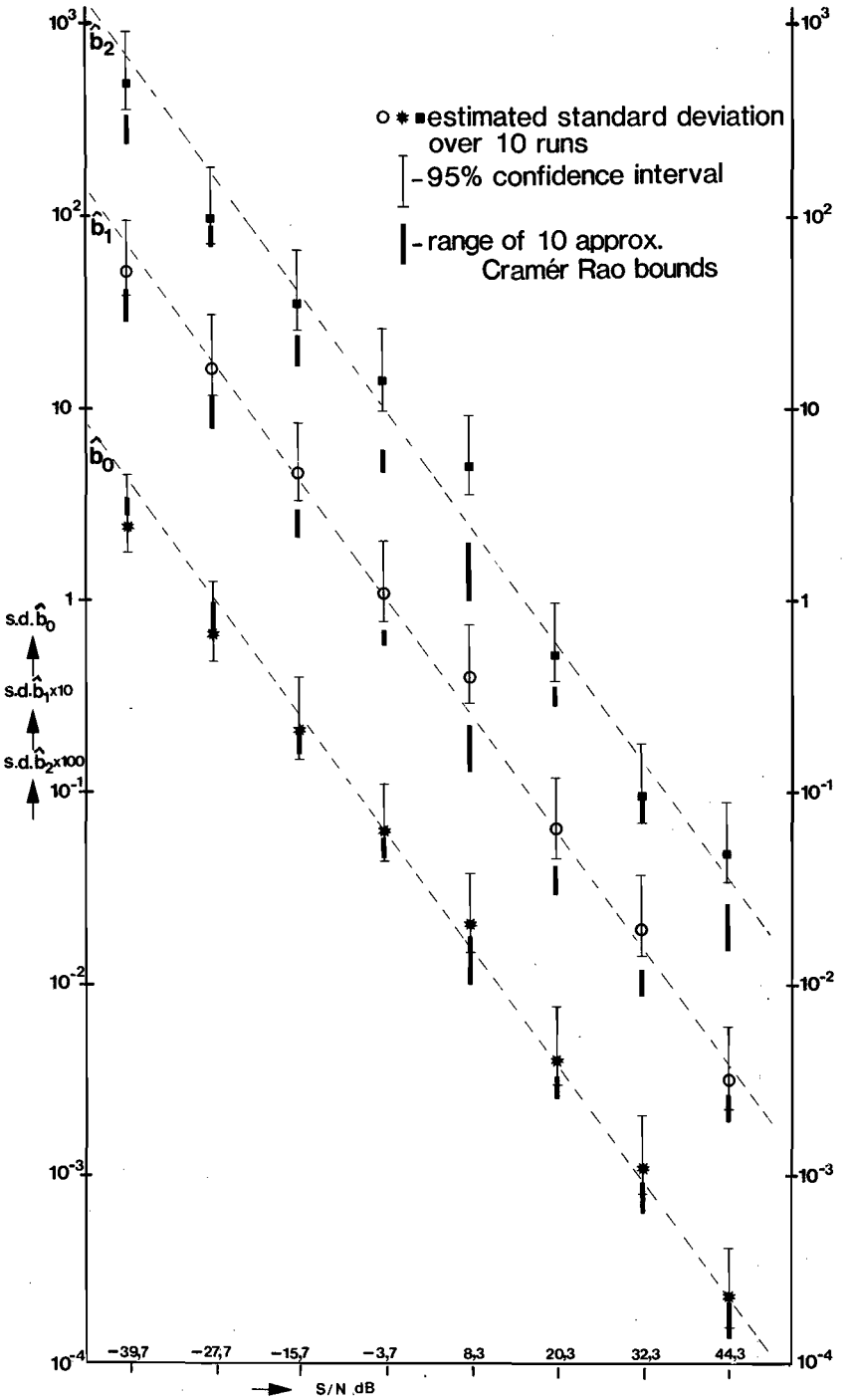


Fig. 6.14 Standard deviation of the MA process parameters

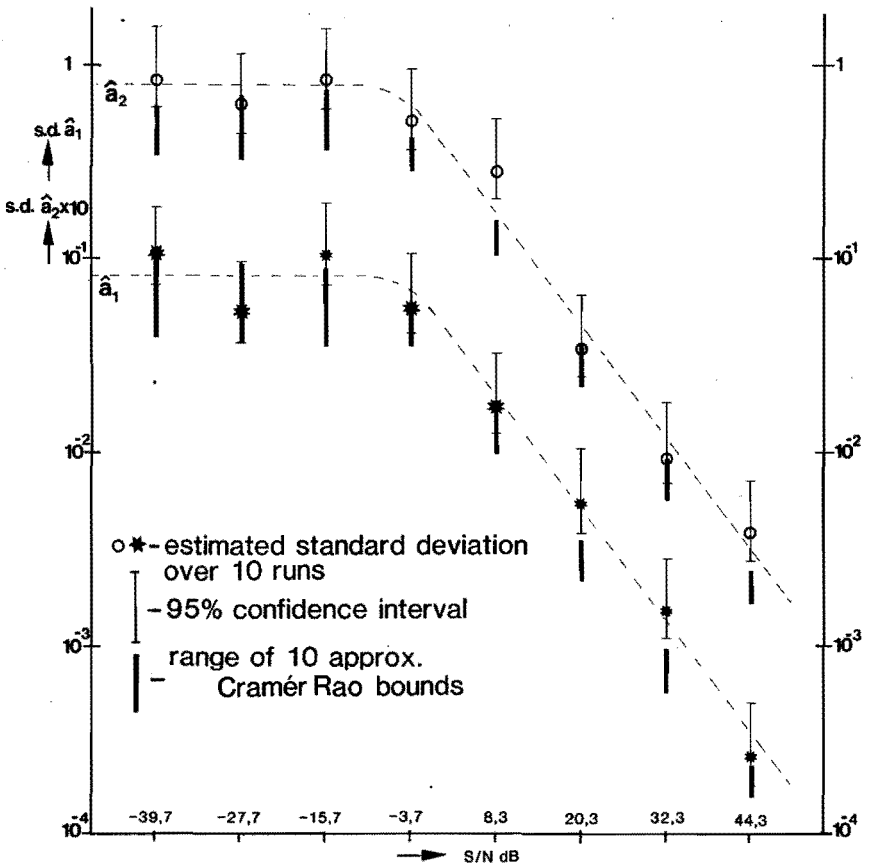


Fig. 6.15 Standard deviations of the the AR process parameters

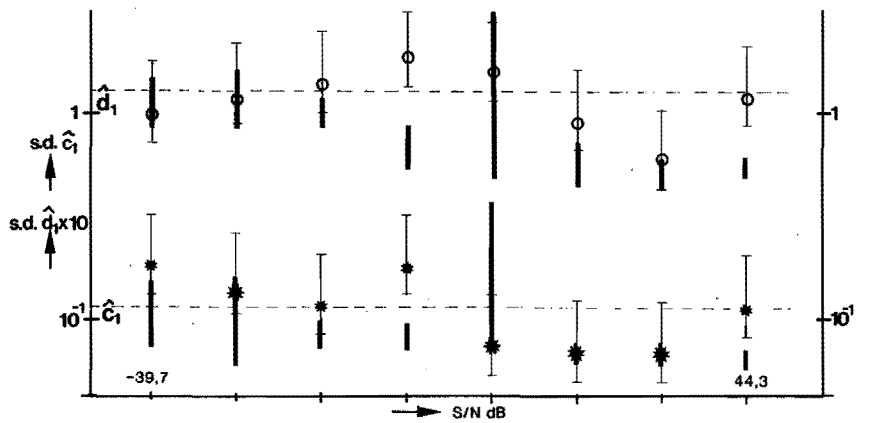


Fig. 6.16 Standard deviations of the ARMA noise parameters

d) the IV methods show a worse behaviour for the A parameters. This is due to the fact that IV is not very sensitive to noise, so that the estimation of A parameters due to the noise as information carrier for bad S/N ratios, is worse. The use of model extension shows improvement. This is because the model extension acts as an inverse noise filter, as with the suboptimal IV estimator.

e) The Tally estimator is only useful in a part of the whole S/N range. The reasons for this are explained in appendix II.

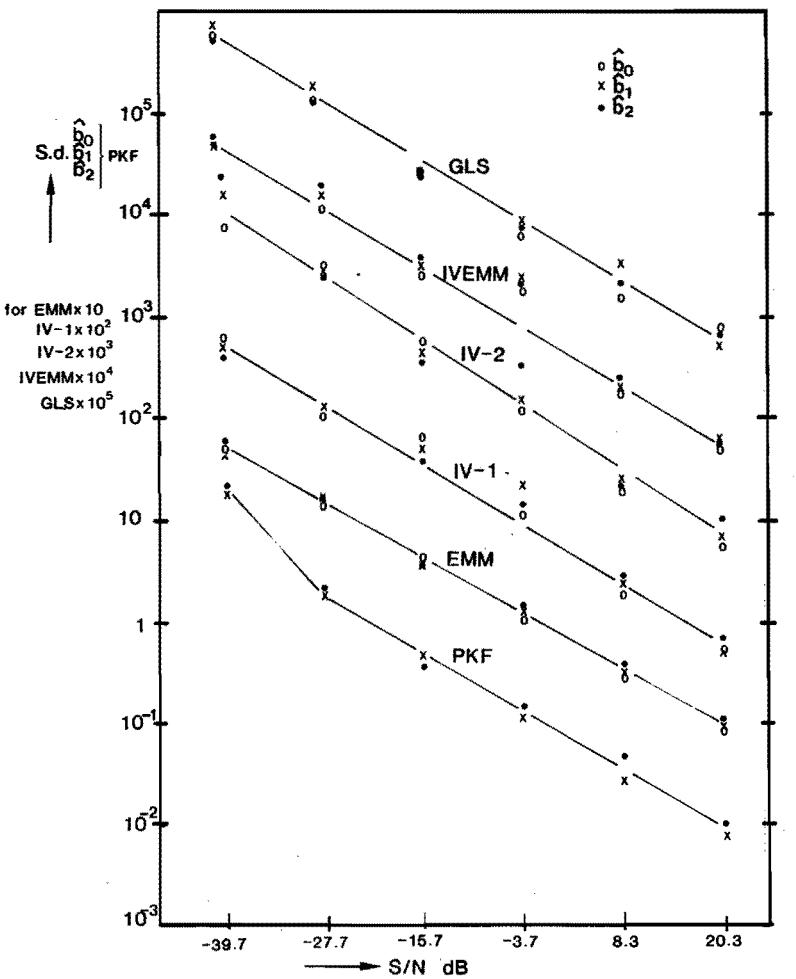


Fig. 6.17 Standard deviations of MA process parameters for different estimators

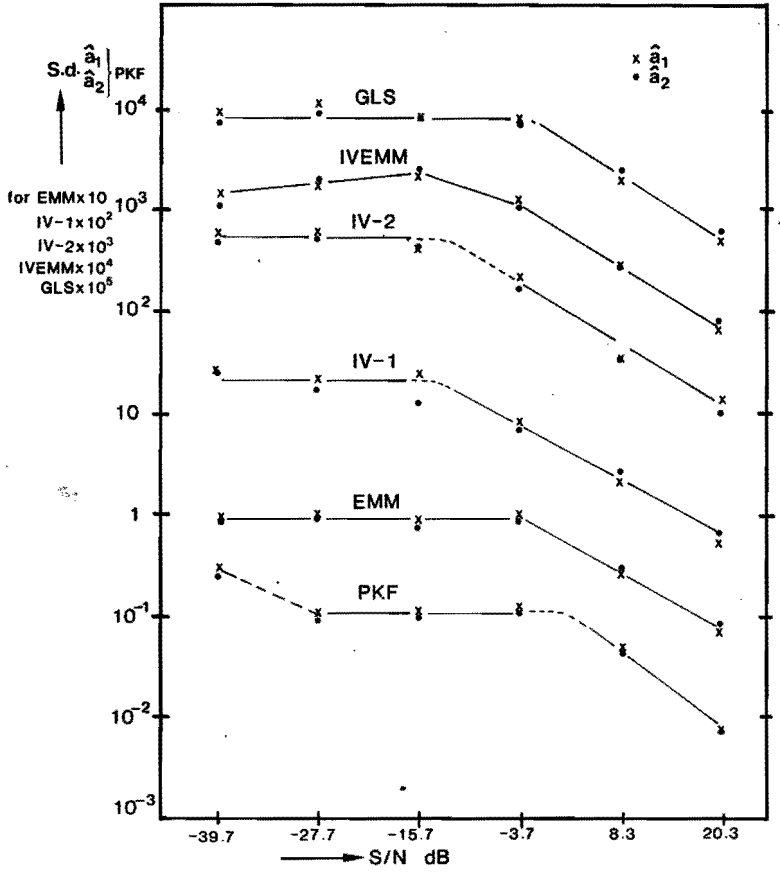


Fig. 6.18 Standard deviation of AR process parameters for different estimators

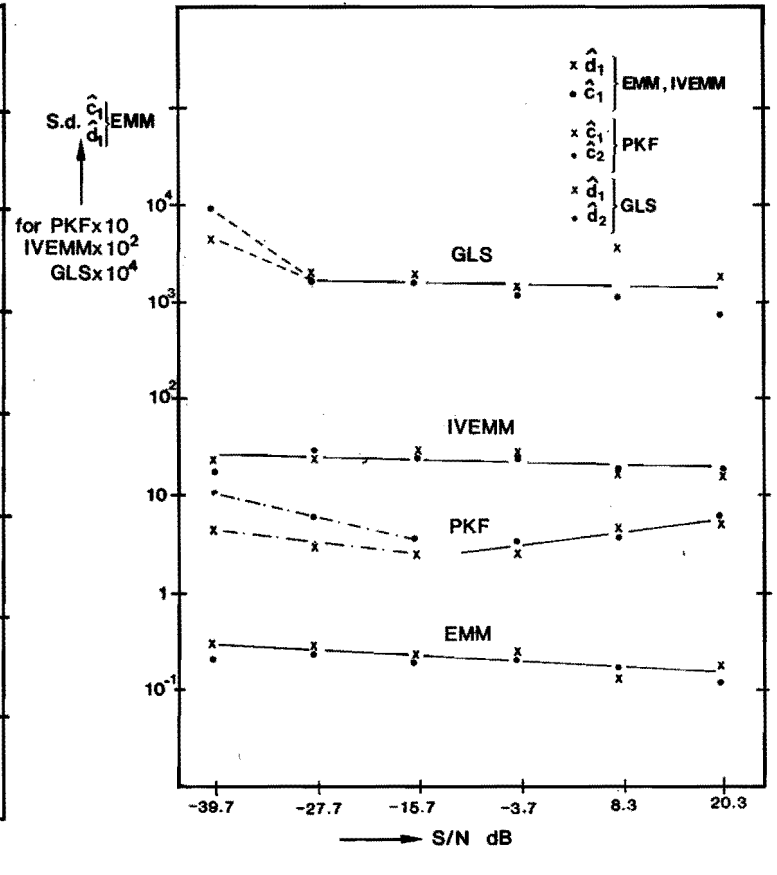


Fig. 6.19 Standard deviations of noise parameters for different estimators

6.5 Divergence of the estimators

In order to illustrate the discussion in paragraph 4.5 concerning the possibility of divergence of the EMM estimator, the following experiment was performed. Various types of input signals were applied to a simulated process; the disturbance was filtered by a noise filter; process and noise filter are given by:

$$\left.\begin{array}{ll} \text{process:} & y_k - 0.2y_{k-1} = u_k + e_k \\ \text{noise filter:} & e_k - 1.9e_{k-1} - 0.93e_{k-2} = \xi_k \end{array}\right] \quad (6.7)$$

The noise filter was excited by a white noise sequence, uniformly distributed between -0.05 and 0.05. The input signal was a white noise sequence, uniformly distributed between -1 and 1, or a sine (of different frequencies) or combinations of both. The following results were obtained; cf. table 6.5. All cases are with N = 2000 and a weighting factor $\rho = 1$, unless otherwise stated. It should be mentioned that the transient response is not discarded in all cases. This yields good initial estimates. The following remarks can be made

- case 1 gives good estimates and a smooth adjustment, as can be expected.
- cases 2 and 3 give also good estimates, although the adjustment of the d parameters is slower. The results are somewhat weaker than those of case 1, although the signal-to-noise ratio is actually improved.
- cases 4 and 5 give good initial estimates, due to the transient phenomenon, but exhibit a slow drift of the process parameters and an adjustment of the noise parameters towards wrong values; cf. fig. 6.20. The choice of these values can be understood as follows:
  For the chosen frequency (= 1/8) the transfer of a second order noise filter can be found:

$$z^{-1} = e^{-j\omega T} = \frac{1-j}{\sqrt{2}} \quad ; \qquad z^{-2} = -j \qquad (6.8)$$

case	input	$\hat{b}_o$	$\hat{a}_1$	$\hat{d}_1$	$\hat{d}_2$	remarks
1	white noise	0.998	-0.185	-1.862	0.900	smooth adjustment
2	white noise + sin 0.785k	0.992	-0.172	-1.844	0.868	smooth adjustment d-parameters slow
3	white noise + 4sin 0.785k	0.985	-0.196	-1.809	0.832	"
4	sin 0.785k	1.090	-0.017	-1.446	0.448	see fig. 6.20
5	sin 0.785k N = 10000	1.177	-0.107	-1.411	0.497	"
6	sin 0.785k N = 10000 ρ = 0.991	1.015	-0.299	-.1.357	0.486	see fig. 6.21
7	sin 0.785k	0.787	-0.520	--	--	no noise par. estimated
8	sin 0.3k	0.945	-0.192	-1.906	0.930	smooth adjustment
9	sin 1.2k	1.058	-0.223	-1.504	0.594	slow adjustment
10	sin 1.2k ρ = 0.991	0.936	-0.198	-1.639	0.638	see fig. 6.22
11	sin 5k	1.046	-0.224	-1.547	0.628	slow adjustment
true parameters		1.000	-0.200	-1.900	0.930	

Table 6.5 Influence of input signal

$$H_{NOISE}(z^{-1}) =$$

$$= \frac{1}{1+d_1 z^{-1}+d_2 z^{-2}} = \frac{1}{1+\frac{d_1}{\sqrt{2}} - j\left(\frac{d_1}{\sqrt{2}}+d_2\right)} = \frac{1+\frac{d_1}{\sqrt{2}} + j\left(\frac{d_1}{\sqrt{2}}+d_2\right)}{\left(1+\frac{d_1}{\sqrt{2}}\right)^2 + \left(\frac{d_1}{\sqrt{2}}+d_2\right)^2} \qquad (6.9)$$

For $d_1 = -\sqrt{2}$, the transfer is imaginary, i.e. there is a 90^0 phase shift; and if $d_1 < -\sqrt{2}$, the real part is negative. The noise filter in this example has a negative real part for this frequency, due to the fact that $d_1 = -1.9$.

This would, therefore, according to the theory of paragraph 4.5, lead to a divergence of the estimate, if the noise filter is fixed. But in this algorithm there is freedom to give different settings to the noise parameters. If $\hat{d}_1 \approx -\sqrt{2}$ is taken as estimate, then there is a 90^0 phase shift for the adjustment caused by frequency 1/8, leading to non-divergence, but this also implies a very slow adjustment if no other components are present in the input signal. As the estimates of the noise filter are not fixed at the value $-\sqrt{2}$ but are wandering around this value, a slow divergence may occur.

- in case 6, the weighting factor ρ is smaller than one. This implies an artificial increase of the variance of the estimator. In this case this is sufficient for an unstable behaviour of the estimates. Here also, the $\hat{d}_1$ parameter is oscillating around the value $-\sqrt{2}$; cf. fig 6.21.
- case 7 shows the necessity of estimating noise parameters.
- in case 8 no sufficient phase shift is obtained in order to make the estimates unstable.
- in cases 9 and 10 a different value for the estimate for the noise parameters is achieved. For this frequency, the following can be found:

$$z^{-1} = 0.362 - 0.932j \quad ; \quad z^{-2} = -0.737 - 0.675j \qquad (6.10)$$

$$H_{NOISE} = \frac{1}{1+0.362d_1 - 0.737d_2 + j(-0.932d_1 - 0.675d_2)} \qquad (6.11)$$

The transfer is purely imaginary if

$1 + 0.362\hat{d}_1 - 0.737\hat{d}_2 = 0$

This equation is, in a good approximation, fulfilled for the estimates found in cases 9 and 10.

- case 11 shows behaviour comparable to the two previous cases.

As a summarizing practical conclusion, we find that divergence may occur in very specific cases, where poor information contents of the input signal coincides with unfavourable noise filter dynamics.

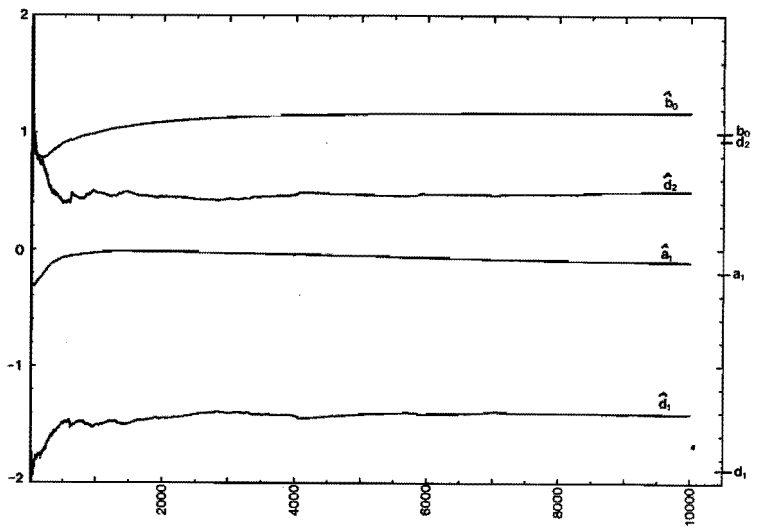


Fig. 6.20 Divergence of estimates
input signal: sin 0.785k; ρ = 1

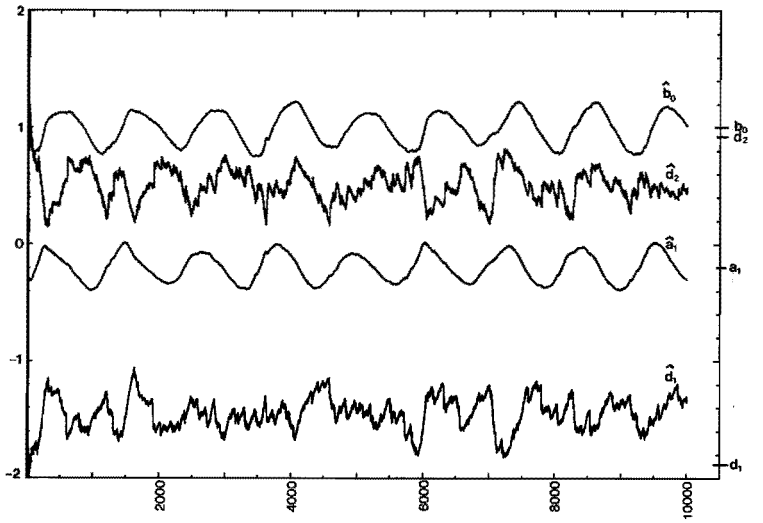


Fig. 6.21 Non-convergence of estimates
input signal: sin 0.785k; ρ = 0.991

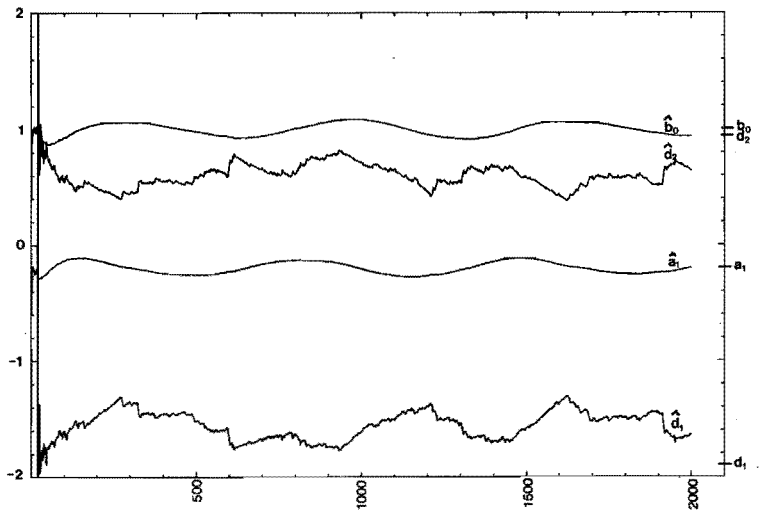


Fig. 6.22 Non-convergence of estimates
input signal: sin 1.2k; ρ = 0.991

6.6 Estimators for input- and output corrupted data

The following system is considered:

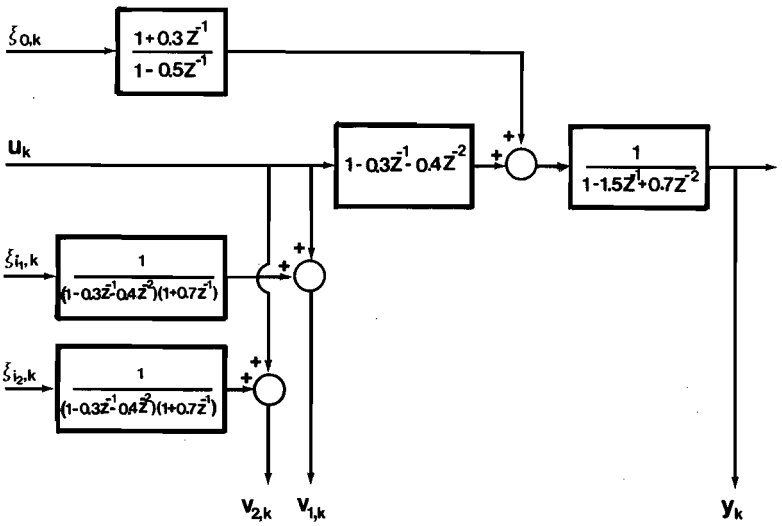


Fig. 6.23 Block diagram of the disturbed input-output experiment

In this experiment a second, independent measurement $v_{2,k}$ of the input is available as indicated in eq. (5.41). This measurement serves as instrumental variable, and its colouring is therefore not essential.

It is assumed that the white noises $\{\xi_{o,k}\}$, $\{\xi_{i1,k}\}$ and $\{\xi_{i2,k}\}$ are mutually independent. The input $\{u_k\}$ is a white signal, with a uniform distribution between (-1,+1). The disturbing noises are uniformly distributed between $(-\lambda,+\lambda)$. For N = 2900 and λ = 0.25 we find for the different estimators:

	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{d}_{i1,1}$	$\hat{c}_{o,1}$	$\hat{d}_{o,1}$
IOIVEMM	0.990	-0.302	-0.406	-1.508	0.703	0.702	—	—
IVEMM	0.901	-0.241	-0.445	-1.501	0.697	—	-0.878	-0.896
IV1	0.908	-0.393	-0.565	-1.677	0.775	—	—	—
IV2	0.900	-0.242	-0.450	-1.505	0.693	—	—	—
true	1.000	-0.300	-0.400	-1.500	0.700	0.700	—	—

where IOIVEMM — IVEMM for input-output corrupted measurement

IVEMM — see chapter 4

IV1 — IV, with use of model output as instruments

IV2 — IV, with use of delayed inputs as instruments

The methods IVEMM, IV1 and IV2 use only $v_{1,k}$ as input measurement. The results show that IOIVEMM gives better estimates for the MA process parameters than the other methods, and the input noise parameters are estimated well.

For worse S/N ratios we have the following results; ($\lambda = 0.5$, $N = 2900$)

	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{d}_{11,1}$	$\hat{c}_{o,1}$	$\hat{d}_{o,1}$
IOIVEMM	0.989	-0.331	-0.411	-1.539	0.719	0.685	—	—
IVEMM	0.727	-0.154	-0.474	-1.486	0.688	—	-0.735	-0.892
IV1	0.754	-0.281	-0.599	-1.700	0.760	—	—	—
IV2	0.737	-0.151	-0.486	-1.508	0.676	—	—	—
true	1.000	-0.300	-0.400	-1.500	0.700	0.700	—	—

Here we also find better estimation of the MA process parameters and a reasonable estimation of the AR noise and AR process parameters. The adjustments of IOIVEMM and IVEMM are shown in fig. 6.24 and fig. 6.25 resp.

For $\lambda = 0.5$, two other cases are considered, where the first order colouring $[1+0.7z^{-1}]$ of the input noise is changed; $N = 2900$

a) v_{i1} coloured by $[1+0.7z^{-1}]^{-1}$, v_{i2} by $[1+0.4z^{-1}]^{-1}$

b) v_{i1} coloured by $[1+0.4z^{-1}]^{-1}$, v_{i2} by $[1+0.7z^{-1}]^{-1}$

		$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{d}_{11,1}$	$\hat{c}_{o,1}$	$\hat{d}_{o,1}$
a)	IOIVEMM	0.970	-0.247	-0.418	-1.489	0.692	0.649	—	—
b)	IOIVEMM	1.000	-0.282	-0.386	-1.498	0.696	0.294	—	—
b)	IVEMM	0.798	-0.223	-0.415	-1.505	0.704	—	-0.354	-0.630
	true	1.000	-0.300	-0.400	-1.500	0.700	a) 0.700 b) 0.400		

The results of the adjustments are shown in fig. 6.26, 6.27 and 6.28 resp. It may be concluded that an improvement of the estimates is obtained, especially for $\hat{b}_0$, and that the various input-noise parameters are estimated reasonably.

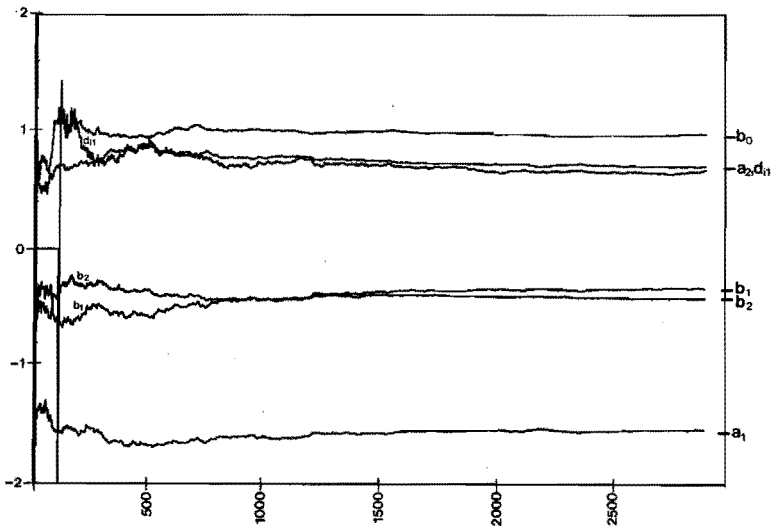


Fig. 6.24 Adjustment using IOIVEMM (λ = 0.5)

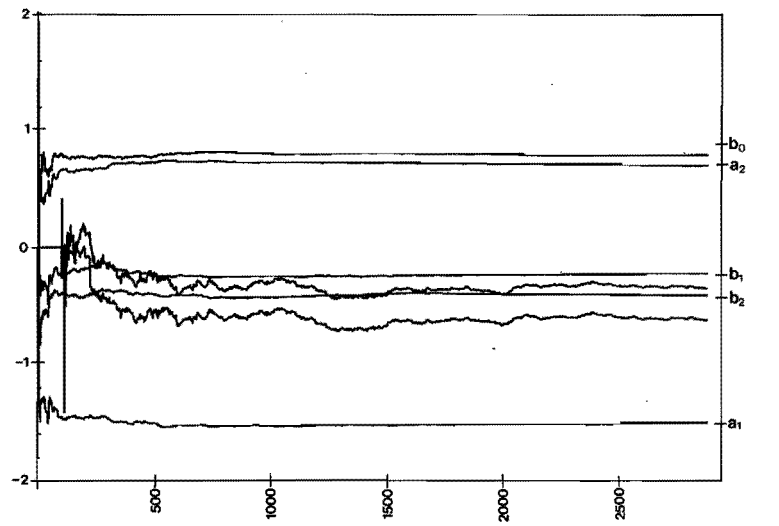


Fig. 6.25 Adjustment using IVEMM (λ = 0.5)

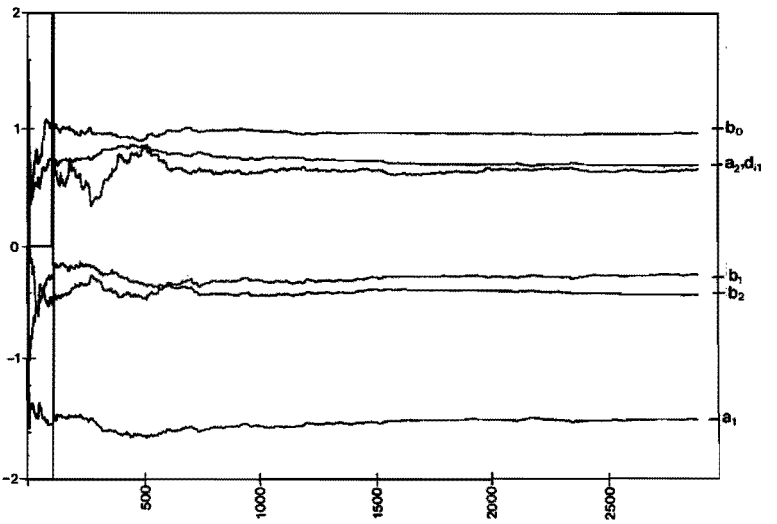


Fig. 6.26 Adjustment using IOIVEMM (λ = 0.5); different input noise colouring

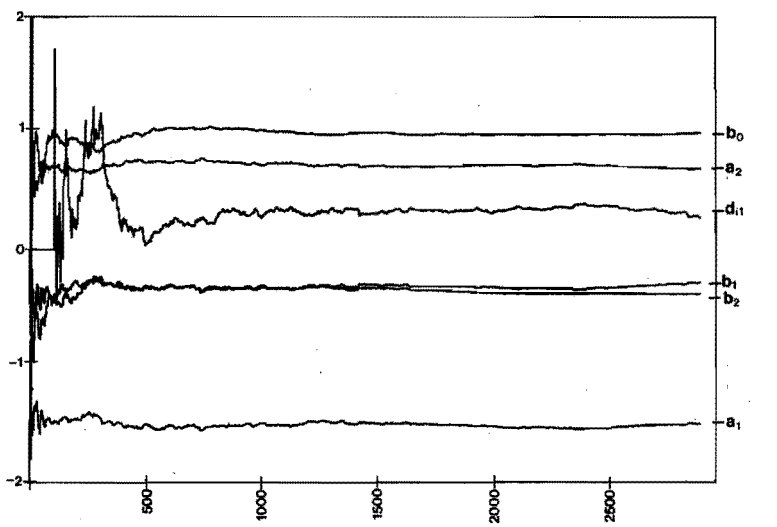


Fig. 6.27 Adjustment using IOIVEMM (λ = 0.5); different input noise colouring

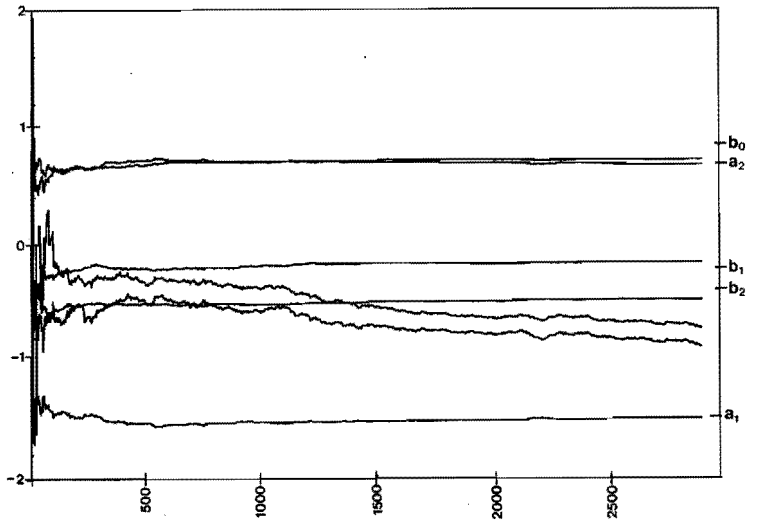


Fig. 6.28 Adjustment using IVEMM (λ = 0.5); different input noise colouring

As a general conclusion of the results of this paragraph, it can be stated that the proposed algorithm IOIVEMM provides better results, at the expense of an extra measurement and slightly more computational effort.

6.7 Concluding remarks

In this chapter we have presented several results which support the theory of chapters 3, 4 and 5. We will not repeat the conclusions in detail as they have already been drawn in the previous paragraphs of this chapter, but we will restrict ourselves to the following summarizing remarks:

a) Based on simulations, it is shown that for small sample sizes the asymptotic properties of the estimators are reasonably achieved. The dependence of the convergence on the S/N ratio is clearly noticeable.

b) The Cramér-Rao lower bounds are approximately reached by simple estimators such as EMM and GLS. Unless divergence of these estimators occurs, which rarely happens, these estimators are therefore attractive in terms of quality and cost. EMM is often favourable due to its flexibility of noise modelling.

c) Divergence can often be expected or predicted from the experimental conditions and monitored during execution of the estimation algorithm. The use of a more sophisticated algorithm like AML is then appropriate.

d) A consistent estimation of process parameters in cases of corrupted input-ouput measurements is feasible. The results are weaker than for the non-corrupted input case, but one should realize that there is a worse overall signal-to-noise condition. Usually many measurables are necessary for a good convergence of the estimates. The algorithm, however, is simple and fast.

e) The availability of extra information, e.g. an extra measurement of the input- or output signal or quantities that are related to the undisturbed process signals, replaces the necessity of a priori information about the disturbances or assumptions about colouring of the disturbances as assumed in the literature.

CHAPTER SEVEN:

ORDER TESTS

7.1 Introduction

In this chapter we will study, in some detail, order testing methods for SISO systems. Based on own experience and on reports in literature we will give emphasis to those methods which have proved to be practically applicable.

As already mentioned in chapter two, the notion of desired order of a model is difficult to define satisfactorily in general terms. This is due to the fact that a model itself always remains a subject for discussion because it has been constructed based on past observations of the process and not on exact theoretical knowledge. If a predescribed set of models is defined, o.a. based on physical reasoning and/or previous observations of the process and on intended future use of the model, then the problem of finding a suitable model order within this set is more tractable. The models within the given set can now be examined with respect to various aspects like

a) are the assumptions of the estimation algorithm met?;
   e.g. white residuals, minimum of loss function, no correlation between residuals and input signal, good resemblance between process- and model output etc.
b) is the principle of parsimony met?; i.e. is the resulting model not over-parametrized.

From a careful analysis of the estimation algorithm, one can deduce quantities, like the ones listed under a), that will show a pronounced behaviour around the "true" order of the system. This is the basis of order testing methods. Such quantities are used as indicators for the desired model order. For a noise free case, the situation is straightforward: for the correct model order, the loss function is minimal and equal to zero; the process- and model responses coincide, the product moment matrix of the measurables still has full rank, and the residuals are zero. In a stochastic case, these indicators will be less pronounced, but still useful: the loss

function will be minimal, the process- and model responses are "close", the residuals are white and there is no correlation between residuals and input signal.

The parsimony principle can cause some problems because the over-fitting of models may not be very easily detectable, due to noise effects. Two effects are relevant in this respect, demonstrating the ambiguous rôle of the noise.

I Depending on the (lack of) noise modelling in the estimation algorithm, the noise dynamics may be (partly) modelled as part of the model of the process.
These noise parameters are estimated as a pole-zero pair, which is more or less cancelling. If this cancelling is obvious, then no problem will arise in recognizing these parameters as being induced by the noise dynamics. In a poorly defined situation (small sample size, bad S/N ratio, non-stationarity of parameters, existence of non-linearities in the process), this cancellation will be hard to detect and will lead to a too high order of the model.

II It may happen that a less important part of the process, e.g. a pole with a small residual value, giving a relatively small contribution to the process output, will cause only a small decrease of the loss function, when taken into account in the model. This small effect may very well be blurred by the disturbing noise if the S/N ratio is bad. This will cause a too low order determination of the model.

From these considerations we may conclude that for a good interpretation of the diagrams and test quantities, a good understanding of what is going on in the estimation and order testing algorithm is necessary. This implies that the choice of a model order is often a subjective decision, which is not suited for complete algorithmization. A fast and on-line order test is therefore not usually feasible.

This chapter is set out as follows. In paragraphs 7.2 to 7.5 the different families of tests based on the aspects listed under a) and

b) will be discussed in detail, and the close ties between these tests will be shown. In paragraph 7.6 a review will be given of the stochastic order tests as proposed in the literature.

7.2 The loss functions

Least squares estimators deal with the minimization of a quadratic loss function for a given order. The idea of using the loss functions for the determination of the parameters can be extended to the selection of the model order within the chosen model set. This approach provides a family of loss functions tests. Usual test quantities are:

$$V_1 = \frac{1}{N} \underline{\hat{e}}^T \underline{\hat{e}} \qquad (7.1)$$

$$V_2 = \frac{1}{N} \underline{\hat{\xi}}^T \underline{\hat{\xi}}$$

where $\underline{\hat{e}}$ and $\underline{\hat{\xi}}$ are the prediction errors; $\underline{\hat{e}}$ is used only if the process dynamics is being modelled, and $\underline{\hat{\xi}}$ if the noise dynamics is also being modelled. The magnitude of these signals $\underline{\hat{e}}$ and $\underline{\hat{\xi}}$ is dependent on the model order, which is illustrated in fig. 7.1.

For too low order models, we can observe that V will decrease for increasing order, as not enough degrees of freedom have been inserted in the model; cf. Van den Boom and Van den Enden (1973).
This can be explained as follows. Suppose that $\hat{\xi}_k$ is a non-white sequence. If this is observed one has to decide whether the model, which generates this sequence of non-white prediction errors is acceptable. If there is a remaining colouring in the prediction error, it implies that information is still contained in the prediction error, which is not used for modelling. The decision not to look for another - possibly more extensive - model is then governed by some prejudice, because it is arbitrary as to which type of remaining colouring is acceptable and which type is not. Therefore, one should then proceed by looking for models which yield white prediction errors. This can be done by parametrizing the non-white error $\hat{\xi}_k$ by a suitable parameterset, e.g.

$$\hat{\xi}_k = \xi_k + \hat{c}_1\bar{\xi}_{k-1} + \hat{c}_2\bar{\xi}_{k-2} + \ldots \hat{c}_n\bar{\xi}_{k-n} + \ldots \qquad (7.2)$$

where $\bar{\xi}_k$ is a white sequence. Then:

$$E\{V_{\hat{\xi}}\} = E\{\frac{1}{N}\,\underline{\hat{\xi}}^T\underline{\hat{\xi}}\} = E\{\frac{1}{N}\,(\underline{\bar{\xi}}^T+\hat{c}_1 z^{-1}\underline{\bar{\xi}}^T+\ldots)(\underline{\bar{\xi}} + \hat{c}_1 z^{-1}\underline{\bar{\xi}} +\ldots)\} =$$

$$= (1+\hat{c}_1^2 + \hat{c}_2^2 + \ldots)\, E\{\frac{1}{N}\,\underline{\bar{\xi}}^T\underline{\bar{\xi}}\} > E\{\frac{1}{N}\,\underline{\bar{\xi}}^T\underline{\bar{\xi}}\} = E\{V_{\bar{\xi}}\} \qquad (7.3)$$

This result is important as it provides a motivation to look for models with a white prediction error, as they lead to a smaller value of the loss function. Moreover, from eq. (7.3), the relation between the loss function test and the whiteness test, viz. paragraph 7.4, is apparent.

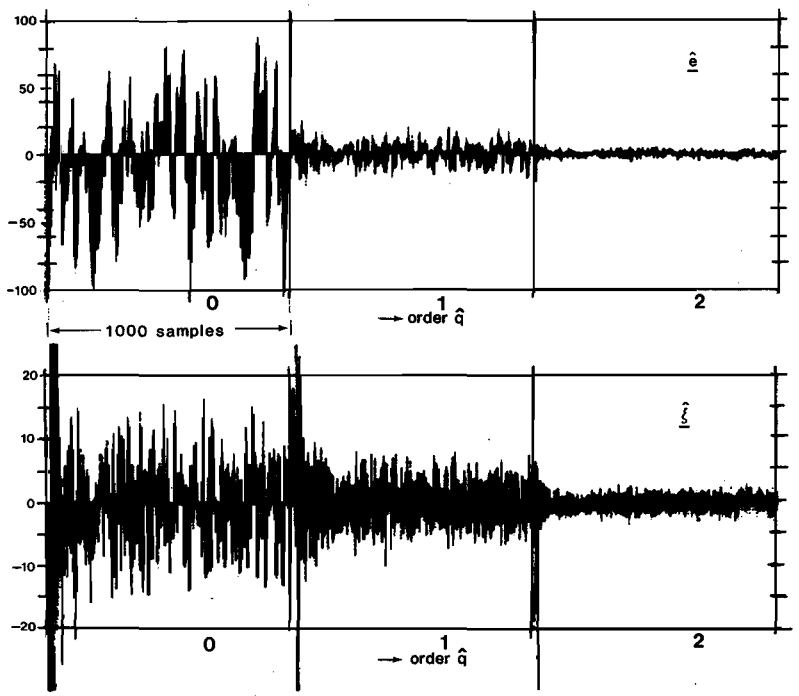


Fig. 7.1 $\underline{\hat{e}}$ and $\underline{\hat{\xi}}$ for different model order.

For over-parametrized models, it has been observed that the loss functions are still - slightly - decreasing. This has led to several adaptations of the loss function as a test quantity, such as Average

Information Criterion (AIC) and several variants of it. They will be dealt with in paragraph 7.6.

A typical behaviour characteristic of the loss function is shown in fig. 7.2, where several cases are given: situation a and b show comparable behaviour; the difference is the worse S/N ratio for situation a. Situation c refers to a process where a part is of lower importance, as discussed in paragraph 7.1. In such a case, dependent on the intended use of the model, one might choose a lower order model, according to the principle of parsimony.

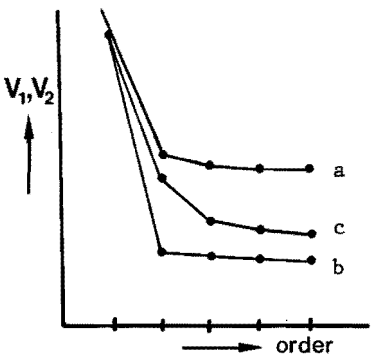


Fig. 7.2 Typical behaviour of loss functions

Now a simple order test can be constructed using the by-products of the estimation algorithm for the calculation of the loss function. Van der Sommen (1971), Gustavsson (1972), and Van den Boom and Van den Enden (1973) give very satisfactory results for these tests, using V_1 even for bad S/N ratios. In the last reference, good results are reported using the test quantity V_2 for determination of the order of the process polynomials $[1+\hat{A}]$ and $[\hat{b}_o+\hat{B}]$, as well as the order of the noise polynomials $[1+\hat{C}]$ and $[1+\hat{D}]$. In fig. 7.3 and 7.4 results are shown using these test quantities V_1 and V_2 for determining process order for different S/N ratios. The process was identical to that proposed by Åström (1968); the noise filter was identical to that proposed by Talmon and Van den Boom (1973):

$$\left.\begin{aligned} &\text{process:} \quad y_k - 1.5y_{k-1} + 0.7y_{k-2} = u_{k-1} + 0.5u_{k-2} + e_k \\ &\text{noise:} \quad\quad e_k - 0.5e_{k-1} = \xi_k + 0.3\xi_{k-1} \end{aligned}\right\} \quad (7.4)$$

The signals u_k and ξ_k were white noise sequences (500 samples) which were uniformly distributed between (-1,1) resp. $(-\lambda,\lambda)$. This results in the following S/N ratios at the output

$$\left.\begin{aligned} &\lambda = \tfrac{1}{4} \quad\quad S/N = 8.3 \text{ dB} \\ &\lambda = 1 \quad\quad S/N = -3.7 \text{ dB} \\ &\lambda = 4 \quad\quad S/N = -15.7 \text{ dB} \end{aligned}\right\} \quad (7.5)$$

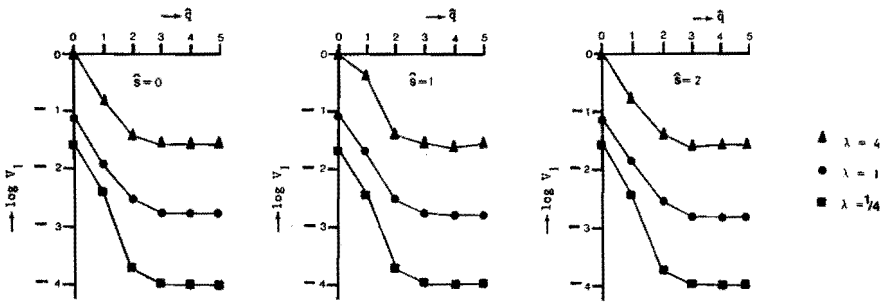


Fig. 7.3 Test of process order using V_1

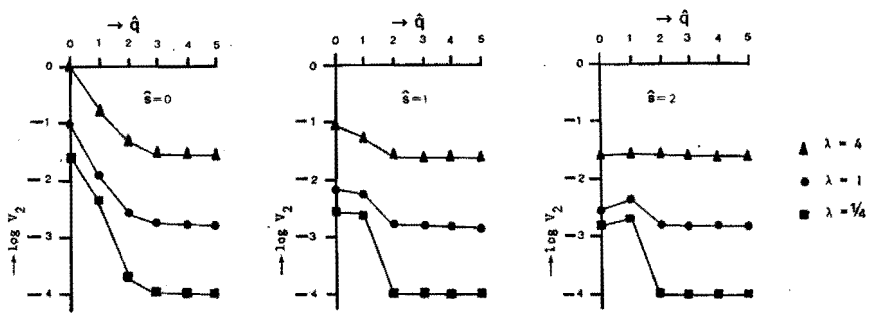


Fig. 7.4 Test of process order using V_2

The parameter estimation method was EMM, which enables the use of a noise model which has an order ($=\hat{s}$) different from the process model ($=\hat{q}$). The results are very satisfactory even for bad S/N ratios.

In many practical applications a time delay is an inherent part of the process. Often this time delay parameter is not well-known or may even be unknown. In such cases the time delay parameter has to be estimated, which can be done in a way analogously to the order tests using the loss functions. If the time delay is not treated separately, then a model has to be used which does not assume that the leading MA parameters are zero. This would imply that a substantial amount of MA parameters, which are in fact zero, has to be estimated. Therefore, the quality of the other parameters will deteriorate considerably. A better approach is to calculate the loss function V_1 as a function of the delay parameter, i.e. the time shift of the input sequence. It has been shown that V_1 will attain a minimum for the correct value of the delay parameter, if the data sequence is sufficiently large; cf. White (1971). A problem may arise when the time delay is not some integer of the sample period, as the minimum of V_1 will be less pronounced in such a case. Therefore a high sample frequency is advisable if a time delay is expected; cf. Rooijakkers (1982).

Woodside (1971) proposed a powerful method for calculating the loss function V_1 using the matrix of observables Q, without performing a parameter estimation. We will consider the noise-free case first. The loss function V_1 can be expressed as a ratio of the determinants of data product-moment matrices related with models of order $\hat{q}+1$ and $\hat{q}$ resp. This can be derived from the matrix $\Omega(u,y;\hat{q}+1)$, which, for this purpose, is constructed in a way which is slightly different from the matrix $\Omega(u,y)$ in previous chapters:

$$\Omega(u,y;\hat{q}+1) = \left[\begin{array}{c|cc|c|cc} u_{k+1} & y_k & u_k & \cdots & y_{k-\hat{q}} & u_{k-\hat{q}} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ u_{k+N} & y_{k+N-1} & u_{k+N-1} & \cdots & y_{k+N-\hat{q}-1} & u_{k+N-\hat{q}-1} \end{array}\right]$$

$$= [\underline{u}_{k+1} | \underline{y}_k | \Omega(u,y;\hat{q})] \qquad (7.6)$$

Now define:

$$Q(u,y;\hat{q}+1) = \Omega^T(u,y;\hat{q}+1)\Omega(u,y;\hat{q}+1) \qquad (7.7)$$

The matrices $H(u,y;\hat{q}+1)$ and $Q(u,y;\hat{q})$ can now be found from the following equation, giving an implicit definition for $H(u,y;\hat{q}+1)$ and $\underline{t}$:

$$Q(u,y;\hat{q}+1) = \left[\begin{array}{c|c} & \\ \hline & H(u,y;\hat{q}+1) \end{array}\right] = \left[\begin{array}{c|c} \underline{y}^T\underline{y} & \underline{t}^T \\ \hline \underline{t} & Q(u,y;\hat{q}) \end{array}\right] \quad (7.8)$$

The matrix $H(u,y;\hat{q}+1)$ can be constructed by removing the first row and column from $Q(u,y;\hat{q}+1)$; and $Q(u,y;\hat{q})$ can be constructed by removing the first row and column from $H(u,y;\hat{q}+1)$.

It can be recognized that:

$$\underline{t} = \Omega^T(u,y;\hat{q})\ \underline{y} \quad (7.9)$$

so that

$$\hat{\underline{\theta}} = Q^{-1}(u,y;\hat{q})\ \underline{t} \quad (7.10)$$

In the parametervector $\hat{\underline{\theta}}$, the parameters appear in a different sequence than in the previous chapters due to the different ordering of the observations in the matrix $\Omega(u,y;\hat{q})$. Using the result V.3 of Appendix V:

$$\det H(u,y;\hat{q}+1) = [\underline{y}^T\underline{y} - \underline{t}^T Q^{-1}(u,y;\hat{q})\underline{t}].\det Q(u,y;\hat{q}) \quad (7.11)$$

The loss function V_1, related to a model of order $\hat{q}$, can be written as (based on residuals):

$$\begin{aligned} V_1 &= \frac{1}{N}\hat{\underline{e}}^T\hat{\underline{e}} = \frac{1}{N}[\underline{y}^T - \hat{\underline{\theta}}^T\Omega^T(u,y;\hat{q})][\underline{y} - \Omega(u,y;\hat{q})\ \hat{\underline{\theta}}] = \\ &= \frac{1}{N}[\underline{y}^T\underline{y} - \hat{\underline{\theta}}^T\underline{t} - \underline{t}^T\hat{\underline{\theta}} + \hat{\underline{\theta}}^T\Omega^T(u,y;\hat{q})\Omega(u,y;\hat{q})\hat{\underline{\theta}}] = \\ &= \frac{1}{N}[\underline{y}^T\underline{y} - 2\underline{t}^T Q^{-1}(u,y;\hat{q})\underline{t} + \\ &\qquad + \underline{t}^T Q^{-1}(u,y;\hat{q})Q(u,y;\hat{q})Q^{-1}(u,y;\hat{q})\underline{t}] = \\ &= \frac{1}{N}[\underline{y}^T\underline{y} - \underline{t}^T Q^{-1}(u,y;\hat{q})\underline{t}] \end{aligned} \quad (7.12)$$

This is exactly the term in brackets in eq. (7.11), yielding:

$$V_1 = \frac{1}{N}\frac{\det H(u,y;\hat{q}+1)}{\det Q(u,y;\hat{q})} \quad (7.13)$$

In noise-free cases the use of this expression for the calculation of V_1 may give numerical problems for over-parametrized models, as the

determinants become zero, due to the linear dependence of the samples, as can be verified by eq. (2.11). The collapse of the determinants can, therefore, be the basis for other order tests: they will be discussed in paragraph 7.3.

For cases where input and output samples are disturbed by noise, the decrease of the quantity V_1 will be less pronounced than in the noise-free case, due to the fact that the minimum of V_1 is now dictated by the variance of the noise. This contribution of the noise to V_1 may be larger than the contribution originating from a model mismatch.

In such a case it is worthwhile using estimates $\hat{Q}(u,y;\hat{q})$ and $\hat{H}(u,y;\hat{q}+1)$ of the non-corrupted matrices $H(u,x;\hat{q}+1)$ and $Q(u,x;\hat{q})$

$$\hat{Q}(u,x;\hat{q}) = Q(u,y;\hat{q}) - \hat{\sigma}^2 R \qquad (7.14)$$

The matrix $\hat{H}(u,y;\hat{q}+1)$ is a submatrix of $\hat{Q}(u,x;\hat{q}+1)$ as defined in eq. (7.8). In eq. (7.14) R is the noise covariance matrix, which needs to be known. This enhancement of the Q-matrix can be combined within a parameter estimation scheme where estimated noise filter parameters become available, e.g. EMM, GLS, AML, and IVEMM. This test received considerable appreciation, due to its reliable results.

So far we have not considered loss function tests, which are put in a statistical framework. Åström (1968) describes a test based on the statistical independence of the quantities $V_2(\hat{q}_2)$ and $V_2(\hat{q}_1) - V_2(\hat{q}_2)$, where $V_2(\hat{q}_1)$ and $V_2(\hat{q}_2)$ are loss functions based on models with order $\hat{q}_1$ and $\hat{q}_2$ respectively and assuming normal residuals. If $\hat{q}_2 > \hat{q}_1 > q$, then $V_2(\hat{q}_2)$ and $V_2(\hat{q}_1) - V_2(\hat{q}_2)$ are independent random variables with χ^2 distributions and $N-(2\hat{q}_2+1)$ and $2(\hat{q}_2-\hat{q}_1)$ degrees of freedom respectively. The test quantity

$$t = \frac{V_2(\hat{q}_1)-V_2(\hat{q}_2)}{V_2(\hat{q}_2)} \cdot \frac{N-(2\hat{q}_2+1)}{2(\hat{q}_2-\hat{q}_1)} \qquad (7.15)$$

has then $F(2\hat{q}_2-2\hat{q}_1,\ N-2\hat{q}_2-1)$ distribution. For a risk level of 5 per cent and $N > 100$ we test for $t < 3$. In table 7.1 results are given for $\lambda = 1/4$, with $N = 200$, using data samples 301 to 500. Note that, as in previous examples, we are dealing here with uniformly distributed random variables. In practical situations we found no difference in results when testing with normally distributed noise.

$\hat{s}$	$\hat{q}_2$ / $\hat{q}_1$	1	2	3	4	5
0	0	610	6955	8088	6004	4963
	1	—	1859	1653	1091	846
	2	—	—	73·4	36·3	26·4
	3	—	—	—	0·00	2·09
	4	—	—	—	—	4·17
1	0	11·9	1352	883	687	544
	1	—	2402	1176	814	604
	2	—	—	−0·97	1·77	1·17
	3	—	—	—	4·55	2·25
	4	—	—	—	—	0·00
2	0	−26·2	718	506	375	320
	1	—	1988	1049	692	553
	2	—	—	6·09	3·02	4·45
	3	—	—	—	0·00	3·47
	4	—	—	—	—	6·94

t-values

Table 7.1 F-test, $\lambda = 1/4$

For $\hat{s} = 0$, we choose a third order model, although the increase in order from 4 to 5 gives some improvement. This is because for $\hat{s} = 0$, higher order models of the process are necessary for obtaining independent residuals. These higher order models have (almost) pole-zero cancelling pairs, which result in a contribution to the loss function of this pair which is less relevant. For $\hat{s} = 1$, we select $\hat{q} = 2$, although the same problem arises at the increase of model order from 3 to 4. For $\hat{s} = 2$ we have problems in selecting model order 2; here also orders 3 and 5 give some improvement. From the principle of parsimony we could neglect this slight decrease; but this depends highly on the intended use of the model.

As a conclusion we can state that the F-test yields somewhat less impressive results. Also here, there is the phenomenon that for improper modelling of the noise dynamics, too high order models of the process may be expected. The tendency of selecting far too high order models can be understood by realizing that a slight decrease of the loss function is qualified by the F-test as a significant improvement. If $V_2(\hat{q}_2) = 0.96V_2(\hat{q}_1)$, then $t \approx 4$, which is significant.

7.3 The rank of the data product moment matrix

Several tests are based on the behaviour of the data product moment matrix $Q(u,y;\hat{q})$. In the noise-free case the matrix $Q(u,x;\hat{q})$ will become singular for $\hat{q} > q$. The rank of this matrix is then $\hat{q}+q+1$, if the input signal has sufficient degrees of freedom and assuming $\hat{p} = \hat{q}+1$ and $p = q+1$; cf. Lee (1964).

An appropriate method for detecting this singularity for $\hat{q} > q$ is by observing the determinant of this matrix for several values of $\hat{q}$. In noisy cases, we have to deal with $\Omega(u,y;\hat{q})$ instead of $\Omega(u,x;\hat{q})$, so we will consider near-singularity. Therefore, take $\hat{q}_{max}$, the maximal order chosen for testing, and construct the matrix $Q(u,y;\hat{q}_{max})$.

The corresponding matrices $Q(u,y;\hat{q})$ for $\hat{q} < \hat{q}_{max}$ can then easily be obtained by removing appropriate rows and columns from $Q(u,y;\hat{q}_{max})$. The matrix $Q(u,y;\hat{q}_{max})$ can be decomposed into an upper triangular matrix Δ, according to Schur, using unitary transformations U_s; cf. Zurmühl (1950)

$$U_s^T \, Q \, U_s = \Delta \tag{7.16}$$

This can be done recursively, starting from $Q(u,y;\hat{q}_{min})$, which is related to Δ_{min}. By adding new rows and columns to $Q(u,y;\hat{q})$ a new Δ can be constructed, with previous Δ's as submatrices. As det $Q(u,y;\hat{q})$ is invariant for this transformation,

we find:

$$\det Q(u,y;\hat{q}) = \prod_{i=1}^{2\hat{q}+1} \delta_{ii} \tag{7.17}$$

where δ_{ii} are the diagonal elements of Δ. If we consider these elements in more detail, we can find a nice relation between this type of determinant and the loss function test of paragraph 7.2. From eq. (7.13) we have:

$$V_1(\hat{q}) = \frac{1}{N} \frac{\det H(u,y;\hat{q}+1)}{\det Q(u,y;\hat{q})} = \frac{1}{N} \frac{\prod_{i=1}^{2\hat{q}+2} \delta_{ii}}{\prod_{i=1}^{2\hat{q}+1} \delta_{ii}} = \frac{1}{N} \delta_{2\hat{q}+2,2\hat{q}+2} \tag{7.18}$$

This implies that the loss function $V_1(\hat{q})$ for different values of $\hat{q}$ can be found on alternate points of the main diagonal of Δ, as shown in eq. (7.19), while on the remaining points of the main diagonal * indicates non-relevant information in this respect.

$$\Delta = N \begin{bmatrix} * & \cdots & & & & & & & & & & & & & \\ & V_1(\hat{q}) & & & & & & & & & & & & & \\ & & * & & & & & & & & & & & & \\ & & & V_1(\hat{q}-1) & & & & & & & & & & & \\ & & & & * & & & & & & & & & & \\ & & & & & \ddots & & & & & & & & & \\ & & & & & & V_1(2) & & & & & & & & \\ & & & & & & & * & & & & & & & \\ & & & & \emptyset & & & & V_1(1) & & & & & & \\ & & & & & & & & & * & & & & & \\ & & & & & & & & & & V_1(0) & & & & \\ & & & & & & & & & & & * & & & \end{bmatrix} \qquad (7.19)$$

Instead of the determinant, the trace of $Q(u,y;\hat{q})$ can be investigated:

$$\text{trace } Q(u,y;\hat{q}) = \sum_{i=1}^{2\hat{q}+1} \lambda_i \qquad (7.20)$$

whereas

$$\det Q(u,y;\hat{q}) = \prod_{i=1}^{2\hat{q}+1} \lambda_i \qquad (7.21)$$

λ_i being the eigenvalues of $Q(u,y;\hat{q})$.

For near singularity, trace $Q(u,y;\hat{q})$ will hardly increase for increasing $\hat{q}$. This test can be performed without parameter estimation; cf. Unbehauen and Göhring (1973). Woodside (1971) calculated a determinant ratio DR for successive values of $\hat{q}$:

$$DR = \frac{\det Q(u,y;\hat{q})}{\det Q(u,y;\hat{q}+1)} \qquad (7.22)$$

or an enhanced determinant ratio EDR if the noise covariance matrix is available:

$$EDR = \frac{\det \hat{Q}(u,y;\hat{q})}{\det \hat{Q}(u,y;\hat{q}+1)} \tag{7.23}$$

where $\hat{Q}(u,y;\hat{q}) = Q(u,y;\hat{q}) - \sigma^2 R(n;\hat{q})$ (7.24)

The close similarity between DR, EDR and the test of eq. (7.18) is apparant. In DR and EDR one row and one column extra are used but they represent the input signal, which does not contribute to the singularity phenomenon. Van den Boom and Van den Enden (1973) calculated a relative determinant:

$$\text{rel.det } Q(u,y;\hat{q}) = \frac{\det Q(u,y;\hat{q})}{\text{max. content of } Q} \tag{7.25}$$

The use of a relative determinant instead of the determinant itself is important, as the determinant is a function of the "power" of the signals:

$$\det Q = f(\text{power}^{2\hat{q}+1}) \tag{7.26}$$

In paragraphs 7.1 and 7.2 we have already met the phenomenon that, for improper noise modelling, a too high order model of the process may be found. Therefore it is important to have separate order tests for the noise dynamics. Only Van den Boom and Van den Enden (1973) give tests for a separate detection of the noise order. They use the EMM estimator. From the following data product moment matrix:

$$Q(u,y,\hat{\xi},\hat{e}) = \frac{1}{N} \Omega^T(u,y,\hat{\xi},\hat{e})\Omega(u,y,\hat{\xi},\hat{e}) \tag{7.27}$$

the following sub-matrices Q_{11} and Q_{22} are interesting for separate order determination of process and noise dynamics:

$$Q(u,y,\hat{\xi},\hat{e}) = \left[\begin{array}{cc|cc} U^TU & U^TY & U^T\hat{\Xi} & U^T\hat{E} \\ Y^TU & Y^TY & Y^T\hat{\Xi} & Y^T\hat{E} \\ \hline \hat{\Xi}^TU & \hat{\Xi}^TY & \hat{\Xi}^T\hat{\Xi} & \hat{\Xi}^T\hat{E} \\ \hat{E}^TU & \hat{E}^TY & \hat{E}^T\hat{\Xi} & \hat{E}^T\hat{E} \end{array}\right] = \left[\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right] \tag{7.28}$$

During estimation, however, P_k is available:

$$P_k = Q^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{7.29}$$

where, from result V.4 of appendix V:

$$\left.\begin{aligned} C_{11} &= (Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}^T)^{-1} \\ C_{22} &= (Q_{22} - Q_{12}^T Q_{11}^{-1} Q_{12})^{-1} \end{aligned}\right] \tag{7.30}$$

As u_k and ξ_k are mutually independent:

$$Q_{12} \approx \frac{1}{N} \begin{bmatrix} \emptyset & \emptyset \\ Y^T \hat{\Xi} & Y^T \hat{E} \end{bmatrix} \tag{7.31}$$

$$C_{11} \approx \left[Q_{11} - \begin{bmatrix} \emptyset & \emptyset \\ \emptyset & M \end{bmatrix} \right]^{-1} \tag{7.32}$$

and for Q_{11}:

$$Q_{11} \approx C_{11}^{-1} + \begin{bmatrix} \emptyset & \emptyset \\ \emptyset & M \end{bmatrix} \tag{7.33}$$

where M is a submatrix, which is only important if the signal-to-noise ratio is bad. This will be indicated later in this section. If we neglect M for good signal-to-noise ratios, then C_{11} can be used as an indicator for the order of the process. Now, if Q_{11} is near singular, it has some eigenvalues which are small compared to the others. As the eigenvalues of C_{11} are, approximately, the inverse of the eigenvalues of Q_{11}, C_{11} also has eigenvalues which are strongly different in magnitude. So C_{11} can also be used as an indicator for near singularity, occuring for $\hat{q} > q$. This can be done in several ways. Van den Boom and Van den Enden defined a relative determinant:

$$\text{rel.det } C_{11} = \frac{\det C_{11}}{\text{max. content of } C_{11}} \tag{7.34}$$

motivated by the fact that the determinant is a function of the "power" of the signals:

$$\det C = f(\text{ power }^{2\hat{q}+1}) \tag{7.35}$$

In fig. 7.5 the behaviour of rel.det C_{11} is shown for $\hat{s} = 0,1,2$ leading to $\hat{q} = 2$ for $\lambda = 1/4$, and $\lambda = 1$ and a doubtful determination $\hat{q} = 2$ for $\lambda = 4$. In practice, we found also that

rel.det C_{22} is a useful indicator for the order of the noise dynamics; cf. fig. 7.6.

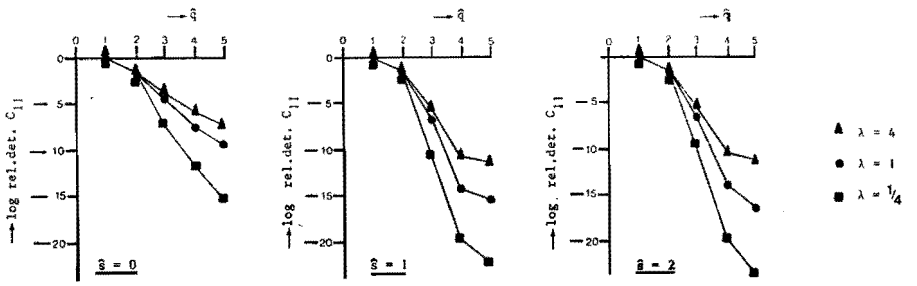


Fig. 7.5 Test of process order using rel.det. C_{11}

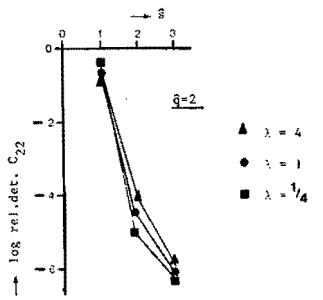


Fig. 7.6 Test of noise order using rel.det. C_{22}

Since det Q_{11} is small if $\hat{q} > q$, the quantity det C_{11} will be large. However, as the maximum content of C_{11} is increasing more rapidly than det C_{11}, then rel.det C_{11} will decrease for $\hat{q} > q$. The amount of increase can be defined by:

$$\Delta C_{11}(\hat{q}) = \frac{\det C_{11}(\hat{q})}{\det C_{11}(\hat{q}-1)} \qquad (7.36)$$

Now, by observing for which $\hat{q}$ a significant increase occurs, the order of the process can be found; cf. fig. 7.7. In this figure an increase of ΔC_{11} can be observed for $\hat{q} = 3$, corresponding to a model order $\hat{q} = 2$. The increase is dependent on the signal-to-

noise ratio, and is limited by the matrix M, which has been neglected; cf. eq. (7.33). If the signal-to-noise ratio is very large (λ = 1/64) this increase is more apparent; cf. fig. 7.7.

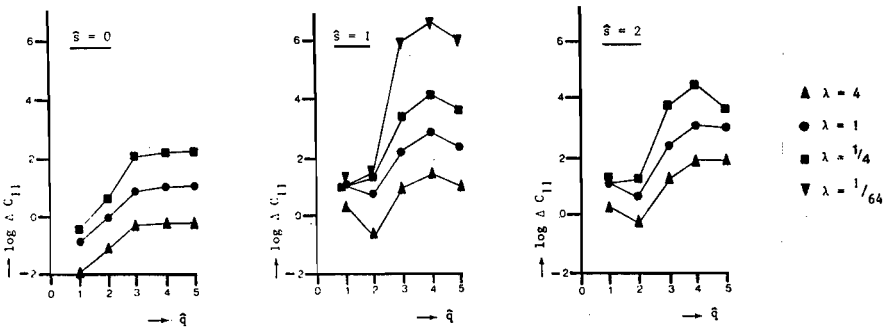


Fig. 7.7 Test of process order using ΔC_{11}

As mentioned before, the rank tests are obscured by the disturbing noise, which converts the singularity test into a test of ill-conditionedness. By making use of the instrumental variable estimator, one can make the data product moment matrix less sensitive for the disturbing noise. Wellstead (1978) reported an improved discriminatory power for this type of test. For a system where the input and output signals are disturbed by noise, the test quantity is:

$$Q(v,y,z;\hat{q}) = \frac{1}{N}\,\Omega^T(v,z;\hat{q})\,\Omega(v,y;\hat{q}) \qquad (7.37)$$

where z_k is the IV quantity. Then, due to the independence of the disturbance on the output and the IV quantity:

$$E\{Q(v,y,z;\hat{q})\} = E\{Q(v,x,z;\hat{q})\} \qquad (7.38)$$

For large N, the rank of $Q(v,y,z;\hat{q})$ is, therefore, equal to $\min(\hat{q}+q+1, 2\hat{q}+1)$ and is not dependent on the properties of the output noise. The input noise is not important in this respect, due to the necessary persistently exciting property of the input signal u_k, yielding sufficient degrees of freedom to make the rows u_k in the matrix $\Omega(u,x)$ independent. The disturbing noise cannot change this. For testing the rank condition of this instrumental variable product moment matrix, it is not necessary to execute an IV estimator, but an IV signal is needed, which can be generated using a (fix-

ed) auxiliary model. As this is an increase in computational burden, this method is only interesting where the previous methods are lacking in discriminatory power, i.e. for very bad signal-to-noise ratios.

Young, Jakeman and McMurtrie (1980) proposed the use of the inverse of the above-discussed IV product moment matrix, which is available during the estimation. We have discussed this idea before in the context of least squares estimators with extended models; cf. eq. (7.33) and Van den Boom and Van den Enden (1973).

So far we have discussed the behaviour of some quantities, when the model order is varying. These quantities showed a tendency towards a different behaviour for $\hat{q} > q$. In cases where bad signal-to-noise ratios occur, it is often difficult to notice this change in behaviour. Consequently, so far, the order test algorithms have been treated as a decision problem relying on human judgement, where ultimately the experimenter decides. Attempts to construct an automated model order testing algorithm have not been very successful so far. A necessary condition for such an automation is putting the order testing algorithm in a statistical framework; cf. paragraph 7.6.

A recent proposal for such a statistical treatment of the determinant test, as discussed above, has been given by Stoica (1981). In this approach, a Taylor series expansion is performed for the determinant of the product moment matrix $Q(\tilde{\underline{\psi}})$, where $\tilde{\underline{\psi}}$ is the vector containing all the sample covariances which appear in the matrix Q, and $\underline{\psi}$ the true covariances

$$\det Q(\tilde{\underline{\psi}}) = \det Q(\underline{\psi}) + \frac{d(\det Q(\underline{\psi}))}{d\underline{\psi}} (\tilde{\underline{\psi}}-\underline{\psi}) + \ldots \qquad (7.39)$$

Denoting

$$\underline{t}^T(\underline{\psi}) = \frac{d(\det Q(\underline{\psi}))}{d\underline{\psi}} \qquad (7.40)$$

we can write for the distribution of $\tilde{\underline{\psi}}-\underline{\psi}$:

$$\tilde{\underline{\psi}} - \underline{\psi} \sim \text{As } N(0,P) \qquad (7.41)$$

and for

$$\det Q(\tilde{\underline{\psi}}) - \det Q(\underline{\psi}) \sim \text{As } N(0, \underline{t}^T P \underline{t}) \tag{7.42}$$

Testing the null hypothesis H_o

$$H_o: \quad \det Q(\tilde{\underline{\psi}}) = 0 \tag{7.43}$$

yields a test quantity with a risk of 5 per cent:

$$\det Q(\tilde{\underline{\psi}}) \leq 1{,}95 \left[\underline{t}^T \hat{P} \underline{t}\right]^{\frac{1}{2}} \tag{7.44}$$

where $\hat{P}$ is an estimate of P. A drawback of this method is its limited applicability for small sequences.

A different approach to rank determination is by making use of the singular value decomposition (SVD) of the data product moment matrix $Q(u,y;\hat{q})$. This matrix is constructed for too high a model order $\hat{q}$. The SVD algorithm factorizes Q as:

$$Q(u,y;\hat{q}) = U_1 \Delta U_2^T \tag{7.45}$$

where U_1 and U_2 are orthonormal matrices and Δ is a diagonal matrix containing the singular values in a decreasing order; cf. Golub and Reinsch (1970). Inspection of the numerical values of the singular values obtained, gives a possibility for order test, as those singular values which are close to or identical to zero, compared to the others, do not contribute to the model behaviour. In modelling MIMO systems, the SVD has become popular for the factorization of the Hankel matrix, which contains the Markov parameters of the process. For more details cf. Hajdasinski and Damen (1979).

From a numerical point of view, the calculation of the singular values is attractive due to the robustness of the algorithm; cf. Klema and Laub (1980).

7.4 Whiteness of residuals and correlation of disturbances

From the discussions concerning the variety of estimation schemes in previous chapters, it will be clear that by proper modelling of the

process- and noise dynamics, $\hat{\xi}$ may tend to an acceptable estimate of the equation error. If the models are adequate, $\hat{\xi}$ has near white noise properties. This implies that estimation algorithms have to be used, which result in white prediction errors, such as EMM, IVEMM, GLS and AML/IQL.

The motivation to look for models with white prediction errors is twofold:

1) Estimation schemes, as mentioned above, will produce inconsistent estimates if the prediction errors are non-white, due to improper modelling.
2) Coloured prediction errors contain information which is not represented in the model. It is rather prejudicial to allow a part of the available information to be unmodelled.

The same reasoning holds for residuals, which can also be tested for whiteness.

For testing the "whiteness" of residuals or prediction errors, the sample autocorrelation function of $\hat{\hat{\xi}}$, resp. $\hat{\xi}$, can be used. We will show this here for prediction errors:

$$\hat{\Psi}_{\hat{\xi}\hat{\xi}}(\tau) = \frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_{\ell+i} \, \hat{\xi}_{\ell+i+\tau} \tag{7.46}$$

Laning and Battin (1956) show that the variance of the sample correlation function $\hat{\Psi}(\tau)$ of white gaussian noise, calculated by using N samples, is:

$$\text{var}\{\hat{\Psi}(\tau)\} \approx \frac{\Psi^2(0)}{N} \qquad \tau \neq 0 \tag{7.47}$$

Now define a normed correlation function:

$$\hat{r}(\tau) = \frac{\hat{\Psi}(\tau)}{\hat{\Psi}(0)} \tag{7.48}$$

$$\text{Then} \quad \text{var}\{\hat{r}(\tau)\} \approx \frac{1}{N} \tag{7.49}$$

This property can be used as an order test:

calculate $\hat{r}(\tau)$ for some range of τ, for different model orders and check whether $\hat{r}(\tau)$ corresponds to an impulse, with a majority of the values for $\tau > 0$ below the value given by eq. (7.49), orwithin a confidence bound based on eq. (7.49). An example is given in fig. 7.8. If the signals have ergodic properties we can compress the

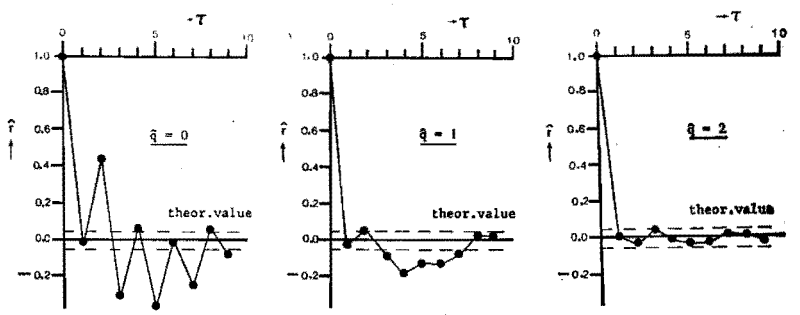


Fig. 7.8 Whiteness test

information as follows:

for each value of the model order calculate $\mathrm{var}\{\hat{r}(\tau)\}$ using all available $\tau > 0$. In this way the information contained in one plot of fig. 7.8 can be compressed into one point of fig. 7.9. For correct model order this point should be close to the theoretical value of (7.49). This also holds for model orders which exceed the true order.

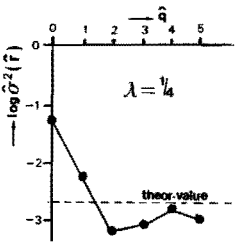


Fig. 7.9 Whiteness test, compression of information of fig. 7.8

It is not possible to use this test for estimation of noise order as well. This test can be regarded as an extension of the tests using the loss function V_1 and V_2; as

$$V_2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_i \hat{\xi}_i = \hat{\Psi}_{\hat{\xi}\hat{\xi}}(0) \qquad (7.50)$$

For safe results, stationarity of the signals $\hat{\xi}$ has to be guaranteed, otherwise N should be replaced by αN, $0 < \alpha < 1$, leading to a

decrease of accuracy and reliability of the test. As the successive samples of $\hat{\xi}_1$ are usually generated during execution of the recursive algorithm, this may cause problems in practical situations. This implies that the $\hat{\xi}_1$ in the sample record are all generated by making use of estimated parameters that are changing during that record, due to the recursive character of the estimation algorithms. For reliable results, this change may not be too pronounced. Otherwise, the resulting estimate of the total record has to be used for recalculation of previous $\hat{\xi}_1$, which implies an increase in computational burden.

A quick check, using this test, can be executed during the recursive estimation, when dealing with algorithms with extended models. In these algorithms, e.g. the EMM estimator, a submatrix of the P matrix deals with the sample covariances of the signal $\hat{\xi}$. For models with proper order, the order of magnitude of the elements on the main diagonal of this matrix, which denote $\hat{\Psi}_{\xi\xi}(0)$ should differ a factor $N^{-\frac{1}{2}}$ in magnitude with the off-diagonal terms, which denote $\hat{\Psi}_{\xi\xi}(\tau)$, $\tau > 0$. Usually, the number of MA-noise-parameters incorporated in extended models is limited, so that this check gives only partial insight.

In paragraph 7.2 the loss function V_2 was also used for construction of a test quantity t with Fisher statistics under the null hypothesis $\hat{q}_2 > \hat{q}_1 \geq q$. In an analogous way, the correlation function $\hat{\Psi}_{\hat{\xi}\hat{\xi}}(\tau)$ for $\tau \geq 0$ can also be used for construction of a Fisher statistic under the above null hypothesis.

An order test which is closely related to this whiteness test is the test of independence of the input signal u_k and the error signals $\hat{e}_k$ and $\hat{\xi}_k$. This can be done in an analogous way as with the whiteness test, by constructing the sample correlation function $\hat{\Psi}_{u\hat{\xi}}(\tau)$ and comparing its magnitude with the value

$$[\hat{\Psi}_{uu}(0) \quad \hat{\Psi}_{\hat{\xi}\hat{\xi}}(0)]^{\frac{1}{2}} N^{-\frac{1}{2}}. \tag{7.51}$$

Another possibility is the construction of a test quantity t, which,

under the null hypothesis H_o = [u_k, $\hat{\xi}_k$ are uncorrelated] has a Fisher statistic, as with the F-test of paragraph 7.1; cf. Bohlin (1971).

7.5 Over-parametrized models

An interesting class of order testing methods can be constructed by using a simple estimation algorithm, e.g. LS, in combination with over-parametrized models. This might seem to be quite time consuming, but applying a simple estimator in a form which is recursive in the order, cf. also Åström (1968), need not be expensive; cf. Hofman (1976).

The general form for the model was given by Talmon and Van den Boom (1973); cf. also chapter 2:

$$[1+A]y_k = [b_o+B]u_k + \frac{[1+C]}{[1+D]}\hat{\xi}_k \qquad (7.52)$$

The ARMA noise dynamics can be modelled by a pure AR model $[1+D']^{-1}$, which can be separated into two parts $[1+D'']^{-1}[1+D''']^{-1}$, where $[1+D'']$ represents the dominant poles of $[1+D']$. The resulting expression for the model is then

$$[1+A']y_k = [b_o'+B']u_k + [1+D''']^{-1}\hat{\xi}_k \qquad (7.53)$$

with

$$\begin{aligned}[1+A'] &= [1+A][1+D''] \\ [b_o'+B'] &= [b_o+B][1+D'']\end{aligned} \qquad (7.54)$$

The parameters $[A']$ and $[b_o'+B']$ of this model can be estimated by an ordinary least squares estimator. For over-parametrized models, $[1+D''']$ has little importance for increasing order; but nevertheless it cannot be neglected as it causes (slightly) biased estimates.

The new transfer function $\frac{[b_o'+B']}{[1+A']}$ has common factors for over-parametrized model orders, which can be detected quite easily. This effect we will call the pole-zero cancellation of the first kind

which means due to the noise effects. Apart from these noise effects, a second type of pole zero cancellation occurs due to the fact that a lower order process may be represented by an infinite number of higher order processes. The class of transfer functions of a particular order $\hat{q}$ of the following structure:

$$\frac{[b_o+B][1+D^*]}{[1+A][1+D^*]} = \frac{[b_o + b_1^* z^{-1} + \ldots + b_{\hat{q}}^* z^{-\hat{q}}]}{[1 + a_1^* z^{-1} + \ldots + a_{\hat{q}}^* z^{-\hat{q}}]} \qquad (7.55)$$

contains infinite elements, due to the fact that $[1+D^*]$ may be chosen rather arbitrarily. This implies that an estimation algorithm will have problems in finding stable estimates. During the estimation procedure, it can be observed that the estimates of all parameters are wandering in relation to each other. For one shot estimates this results in pole-zero plots with pole-zero cancelling pairs. This will be discussed later in this paragraph. For recursive estimators, this wandering behaviour is very pronounced, as may be seen from fig. 7.10 and fig. 7.11; the parameter adjustments are shown for

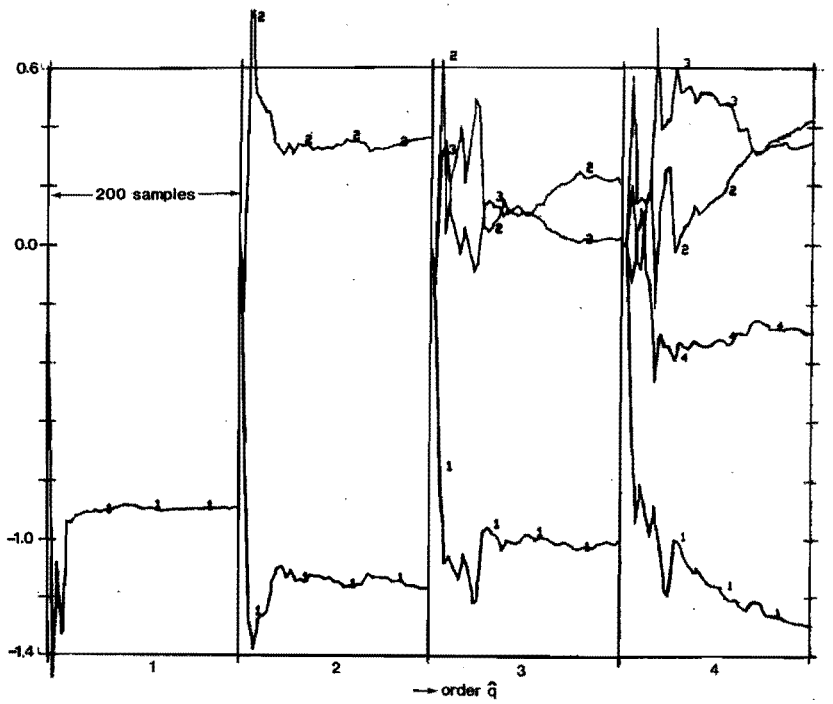


Fig. 7.10 Adjustment of AR parameters for different model orders

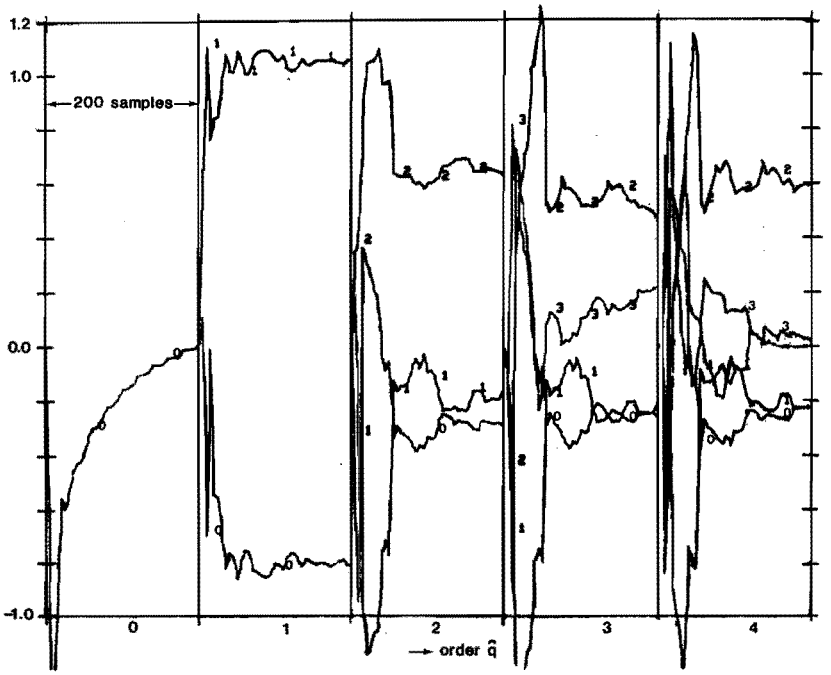


Fig. 7.11 Adjustment of the MA parameters for different model orders

different orders of the model, while the actual process order is 2.

From these figures, it can be seen that the adjustments of the estimates of the MA parameters are less smooth than those of the AR parameters. This indicates a bad S/N ratio; cf. also paragraph 3.6 and fig. 3.3 of chapter 3. It can also be seen that the adjustment of the third order model is smoother that that of the fourth order model. This is because a pole-zero cancellation of the first kind is occurring in the third order model. This pole-zero pair is not wandering, but is determined by the noise dynamics $[1+D'']$. This effect will also be apparent from the pole-zero plots later on in this paragraph.

From these considerations, we may conclude that with over-parametrized models the system is not parameter-identifiable (PI) but only

system-identifiable (SI) in the sense of Ljung's definitions, as given in chapter 2.

For the detection of common factors of $[1+A']$ and $[b_o+B']$, usually the pole-zero plot will be drawn. Due to the effects of the additive noise, no exact cancellation of the pole-zero pairs will occur; cf. fig. 7.12. It can be noticed that the non-cancelling poles and zeros for $\hat{q} \geq 2$ remain constant for the different model orders.

Apparently these belong to the wanted transfer function $\dfrac{[b_o+B]}{[1+A]}$.

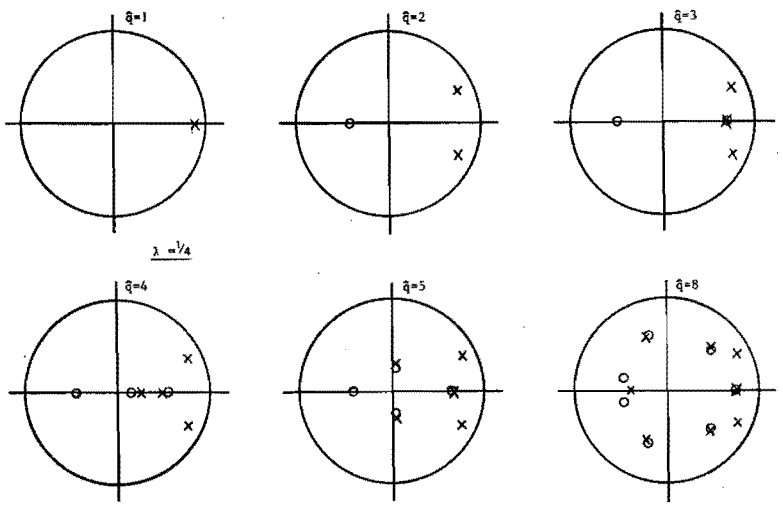


Fig. 7.12 Development of the pole-zero pattern (λ=1/4)

Furthermore it can be observed that one pole-zero cancelling pair in $z \approx 0.75$ remains fairly constant for $\hat{q} \geq 3$. This can be identified as the pole-zero cancellation of the first kind. The noise colouring as given by eq. (7.4) can be rewritten as

$$e_k = \frac{1}{1-0.8z^{-1} + 0.24z^{-2} - 0.072z^{-3} + \ldots\ldots} \xi_k \qquad (7.56)$$

This may be approximated by

$$e_k \approx \frac{1}{1-0.8z^{-1}} \xi_k \qquad (7.57)$$

which corresponds fairly to the observed pole-zero cancellation of the first kind. The other cancelling pairs are wandering for different orders and they can be identified as pole-zero cancellations of the second kind.

A comparable result can be obtained if a more sophisticated estimator, e.g. the EMM, is used in connection with over-parametrized models. The noise parameters will be identified for too low order models as ARMA noise parameters, but will shift, for over-parametrized models, to pole-zero cancelling pairs of the first kind. This results in pole-zero cancelling pairs close to $z = 0$ as estimated ARMA noise parameters; cf. Koenraads (1978).

Results of pole-zero cancellation tests are given by Van den Boom and Van den Enden (1973), Unbehauen and Göhring (1973). They are usually very satisfactory.

There are also methods which avoid the calculation of the poles and zeros of the transfer function. The following matrix may be constructed:

$$R = \begin{bmatrix} 1 & a'_1 & \cdots & a'_{\hat{q}} & \emptyset \\ \emptyset & 1 & a'_1 & \cdots & a'_{\hat{q}} \\ b'_o & b'_1 & \cdots & b'_{\hat{q}} & \emptyset \\ \emptyset & b'_o & b'_1 & \cdots & b'_{\hat{q}} \end{bmatrix} \qquad (7.58)$$

It is pointed out by Loonstra (1967) that det R=0 if $[1+A']$ and $[b_o+B']$ have common zeros. Due to the corrupted signals det R will become small for (nearly) cancelling pairs. In this way, however, the number of cancelling pairs present cannot be determined. Vogt and Bose (1970) offered an alternative method based on the theorem that if A is the companion matrix of f(z) and B the companion matrix of g(z) then f(z) and g(z) have common factors if det $f(B) \neq 0$ or det $g(A) \neq 0$. Cullen and Hall (1971) investigated the amount of elementary arithmetic operations to determine the common factors for

these methods and found that making use of a modified structure of R yields a minimum amount of

$(\frac{1}{3}\hat{q}^3+\hat{q}^2-\frac{1}{3}\hat{q} - 1)$ operations.

Söderström (1975) described, in a general way, the problem of determining common factors in two polynomials, where the parameters have some uncertainties. He proposed a method which, in fact, is an alternative and more efficient way to obtain a generalized least square model. The covariance matrix of the estimates of the coefficients of the transfer function is assumed to be known or, at least, an estimate of it is available. The algorithm for testing common factors of the two polynomials can be formulated as a minimization of a quadratic loss function under a complicated constraint.

7.6. Stochastical tests

In contrast to the above mentioned classes of order tests, which were proposed on a rather heuristical basis, another class of tests has been introduced in the literature, which has a more fundamental basis, e.g. information theory.

Akaike proposed two tests, the final prediction error (FPE), Akaike (1970), and the Average Information Criterion (AIC), Akaike (1973). We will not derive these tests here, but we will limit ourselves to a few comments.

The FPE is given by

$$\text{FPE} = \frac{N + \hat{q}}{N - \hat{q}} \det\left[\frac{1}{N} \sum_{i=1}^{N} \underline{\hat{\xi}}(\underline{\hat{\theta}}_N)\, \underline{\hat{\xi}}(\underline{\hat{\theta}}_N)^T\right] \tag{7.59}$$

where the expression between the parentheses is the estimate of the mean square one-step-ahead prediction error based on a maximum likelihood estimate $\underline{\hat{\theta}}_N$ of order $\hat{q}$ (= number of independent parameters in the model). If $\underline{\hat{\xi}}(\underline{\hat{\theta}}_N)$ is a (good estimate of) white noise then the close relation with the loss function will be apparent.

Akaike's Average Information Criterion (AIC) is defined as

$$AIC = 2\hat{q} - 2\log L(\underline{\hat{\theta}}(\hat{q})) \tag{7.60}$$

where $L(\underline{\hat{\theta}}(\hat{q}))$ is the likelihood function based on a MLE $\underline{\hat{\theta}}(\hat{q})$. The fact that the term $2\hat{q}$ appears in this criterion can be seen as an attempt to follow the principle of parsimony for models, as discussed in chapter 2. The relation of AIC with the loss function test is apparent as the second term of AIC in (7.60) is equivalent to the loss function V. The first term is introduced by Akaike to compensate for the (slight) decrease of V for over-fitted models.

Shibata (1976) studied the asymptotic distribution of AIC and found that asymptotically, there is a probability that this estimator of the order will deliver a too high value compared to the true order. Hipel (1981) gives and extensive survey of practical applications of these tests (FPE and AIC) and also concludes that there is a tendency to over-fit. Krolikowski (1982) gives a survey of the variants of AIC that have been proposed recently as attempts to obtain consistent estimators for the order:

$$\left.\begin{aligned} AIC &= 2(N)\hat{q} - \log L(\hat{\theta}(\hat{q})) \\ AIC &= f(N)\hat{q} - 2\log L(\hat{\theta}(\hat{q})) \end{aligned}\right] \tag{7.61}$$

where f is dependent on N.

This area of research is still in development and the construction of tests of order, which are theoretically well-founded and which deliver low order models, is still a paradise for theoreticians.

7.7 Conclusions

In this chapter we discussed several order tests which are of prime practical interest. We showed that there is a close relationship between the tests which are based on the loss-function and the (relative) determinant tests. Also the relation between the loss function test and the whiteness test has been shown.

Name	Test quantity	Test criterion	par. est. needed	proc.order P+ noise order N	input + output noise	overfitting necessary	experience
loss function V_1	$V_1 = \frac{1}{N} \hat{\underline{e}}^T \hat{\underline{e}}$	min. V_1	N Y (EMM)	N Y	N N	Y Y	good good
loss function V_2	$V_2 = \frac{1}{N} \hat{\underline{\xi}}^T \hat{\underline{\xi}}$	min. V_2	Y	Y	N	Y	good
signal errors	$\hat{e}$, $\hat{\xi}$	significant change	Y	Y	N	Y	good
whiteness	$\hat{r}(\tau)$	$\hat{r}(\tau) \rightarrow$ dirac	Y	N	N	Y	good
F-test	t	risk	Y/N	N	N	Y	doubtful
determinant	rel. det Q	significant change	N	N	N	Y	good
"	det. ratio	"	N	N	N	Y	good
"	IV det ratio	"	N	N	Y	Y	good
"	rel. det C_{11}	"	Y	Y	N	Y	good
"	rel. det C_{22}	"	Y	Y	N	Y	good
"	eigenvalues	$\lambda_{min}/\lambda_{max} \rightarrow 0$	N	N	N	Y	good
SVD	SVD	$\delta_{min}/\delta_{max} \rightarrow 0$	N	N	N	Y	good
trace	trace	no change for $\hat{q} > q$	N	N	N	Y	poor
pole zero	cancellation	different types of cancellation	Y	Y	N	Y	good
parameters	behaviour	stability	Y	N	N	Y	good
FPE	FPE	FPE	Y/N	N	N	Y	good
AIC	AIC	AIC	Y/N	N	N	Y	good

Table 7.2

Some tests are suited for separate discrimination of process and noise order, and some tests are suited for use in situations where input-output disturbances occur.

In order to facilitate the comparison, we have composed table 7.2 where the most important aspects of the different tests are listed.

As can be seen from the given examples, it is important that a good man-algorithm interaction exists and that for a good appreciation of the different order tests the aspect of experience cannot be excluded. Therefore, an interactive package like SATER, containing many of the afore-mentioned order tests, has proved its feasibility in practical situations.

CHAPTER EIGHT:

GENERAL CONCLUSIONS

In the present study we have investigated several aspects of estimators for SISO processes with disturbed outputs. The main conclusions are:

a) A general scheme has been presented. Its main feature is that three basic operations can be distinguished to ensure consistence of estimators, viz. filtering, model extension and use of an IV quantity. These three operations may be combined within one general estimator. Existing estimators like LS, GLS, EMM, IV, AML, IQL, IVEMM, sub-optimal IV and Tally, all fit into this scheme as special cases of this general estimator. The availability of such a general scheme is very attractive for various reasons: it gives better insight into the mutual relationships of existing estimators, it gives a nice connection between the LS- and IV-oriented estimators and computer programs for estimators can be designed in such a way that different estimators can be incorporated into one program.
As the AML/IQL schemes are part of this scheme, the relation with ML (for gaussian disturbances) is then trivial.

b) Expressions for the Cramér-Rao bounds using the Talmon and Van den Boom model, as introduced in chapter 2, are given. From these relations it can be concluded that qualitatively there exists a different behaviour for the variances of the different kinds of parameters as functions of S/N ratios. The qualitative behaviour has been confirmed by simulation, where the S/N ratio varied between extreme low and high values. Quantitatively these lower bounds are reasonably well approached by relatively simple estimators like EMM and GLS over the whole range of S/N values. Estimators like Tally and IV-oriented schemes showed less impressive results for bad S/N ratios.

c) The small sample behaviour was investigated experimentally with respect to bias and variance of the estimators. It was found that for high S/N values the bias was negligible for

small sample size (N = 200). Also the covariance of the estimates for small sample sizes was close to the Cramér-Rao bound. For low S/N ratios the convergence towards the true parameter values was evidently slower.

d) Relatively simple estimators like EMM, which we proposed earlier, and GLS proved to be reliable schemes with respect to bias and variance. In rare cases divergence may occur, but this can usually be expected/predicted from the experimental circumstances. The attractiveness of these schemes is also due to their simplicity and speed. In addition, the EMM algorithm is very flexible with respect to types of noise models that can be used.

e) Several estimators have been proposed in chapter 5 for cases where both input- and output signals are noise corrupted. These estimators allow the use of possible extra available information, e.g. an extra independent measurement of the input- or the output signal, or other signals which, as IV quantities, are related to the process signals. Such extra information is often practically more often available than a priori information such as known covariance of the noise as in schemes proposed in literature. It is also applicable more generally in practice than schemes that assume white disturbances, as also proposed in literature.

Simulations show that process- and noise parameters can be estimated without bias.

The behaviour is less favourable than in the case of only output disturbances which, of course, may be expected due to the worse overall signal-to-noise condition.

Based on observations of simulated experiments, the small sample behaviour is poor: many measurements are necessary for the convergence of the estimates.

f) An extensive review of order tests, which are practically applicable, has been presented and the close relations among these tests have been indicated. Also order tests which allow the distinct determination of the orders of process- and noise dynamics have been presented. Simulations show the applicability for these order tests for an extensive range of S/N ratio's.

g) An interactive program package SATER has been designed, which is described briefly in chapter 6. The availability of such an interactive package for identification purposes gives the experimenters much freedom in applying the different estimating- and order testing schemes.

Appendix I Approximation of the covariance matrix of the noise

Using the spectral factorization theorem for signals with rational spectra we find that the following unique parametrization is appropriate:

$$G(z^{-1}) = \frac{[1+C(z^{-1})]}{[1+D(z^{-1})]} \qquad (I\text{-}1)$$

where $G(z^{-1})$ is stable and minimum phase (if the spectrum has no singularities on $|z| = 1$), and hence causally invertible.

The following representations can also be given, which are degenerations of (I-1); where for notational simplicity the argument z^{-1} is dropped:

$$[1+D'] = \frac{[1+D]}{[1+C]} \qquad (I\text{-}2)$$

$$[1+C'] = \frac{[1+C]}{[1+D]} \qquad (I\text{-}3)$$

The polynomials $[1+C]$ and $[1+D]$ are of finite order, but $[1+C']$ and $[1+D']$ have, in principle, infinite length. We consider, however, stable and physically realizable systems so that for some ν, μ the tails

$$\sum_{i=\nu+1}^{\infty} c'_i z^{-i} \quad \text{and} \quad \sum_{i=\mu+1}^{\infty} d'_i z^{-i}$$

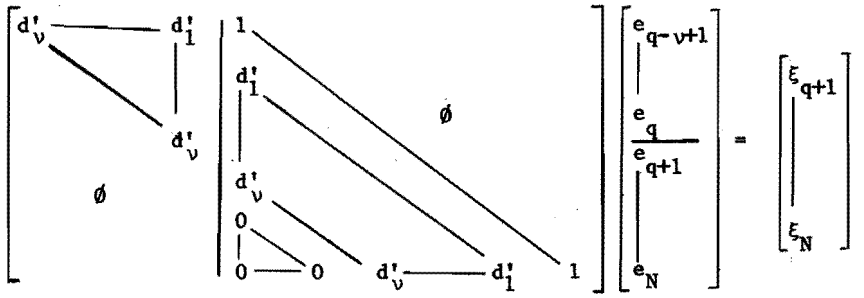
of the polynomials $[1+C']$ and $[1+D']$ do not give any measurable contribution to the transfer of the system.

From the descriptions

$$[1+D']e_k = \xi_k \qquad (I\text{-}4)$$

$$e_k = [1+C']\xi_k \qquad (I\text{-}5)$$

we construct the matrix notation:



(I-6)

which can be written as:

$$\left[D^{*\prime}\,\middle|\,D^{\prime}\right]\begin{bmatrix}\underline{e}^{*}\\ \underline{e}\end{bmatrix}=\underline{\xi} \tag{I-7}$$

where D' is a lower triangular matrix, or alternatively:

$$\begin{bmatrix}e_{q+1}\\ \vdots\\ e_N\end{bmatrix}=\begin{bmatrix}c'_\mu & \cdots & c'_1 & & & & \\ & \ddots & & & & & \\ & & c'_\mu & & & & \\ & \emptyset & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & \cdots & 0 & \\ & & & & \ddots & & \\ & & & & & c'_\mu & \cdots & c'_1\end{bmatrix}\begin{bmatrix}1 & & & & \\ c'_1 & \ddots & & & \\ \vdots & & \ddots & & \\ c'_\mu & & & \ddots & \\ & \ddots & & & 1\end{bmatrix}\begin{bmatrix}\xi_{q-\mu+1}\\ \vdots\\ \xi_q\\ \xi_{q+1}\\ \vdots\\ \xi_N\end{bmatrix}$$

or:

$$\underline{e}=\left[C^{*\prime}\,\middle|\,C^{\prime}\right]\begin{bmatrix}\underline{\xi}^{*}\\ \underline{\xi}\end{bmatrix} \tag{I-8}$$

where C' is a lower triangular matrix.

For the covariance matrix of the equation error we find:

$$\begin{aligned}\operatorname{cov}\underline{e}&=E\{\underline{e}\,\underline{e}^{T}\}=E\{(C^{*\prime}\underline{\xi}^{*}+C^{\prime}\underline{\xi})(\underline{\xi}^{*T}C^{*\prime T}+\underline{\xi}^{T}C^{\prime T})\}=\\&=C^{*\prime}E\{\underline{\xi}^{*}\underline{\xi}^{*T}\}C^{*\prime T}+C^{\prime}E\{\underline{\xi}\,\underline{\xi}^{T}\}C^{\prime T}=\\&=(C^{*\prime}C^{*\prime T}+C^{\prime}\,C^{\prime T})\sigma_\xi^2\end{aligned} \tag{I-9}$$

Neglecting, for $N \gg \mu$, the initial conditions, i.e. $\underline{\xi}^{*}$, we find

$$\operatorname{cov}\underline{e}\approx C^{\prime}C^{\prime T}\sigma_\xi^2 \tag{I-10}$$

For the autoregressive description we have:

$$\begin{aligned}&E\{(D^{*\prime}\underline{e}^{*}+D^{\prime}\underline{e})(\underline{e}^{*T}D^{*\prime T}+\underline{e}^{T}D^{\prime T})\}=\sigma_\xi^2 I\\&D^{*\prime}\operatorname{cov}\underline{e}^{*}\,D^{*\prime T}+D^{*\prime}E\{\underline{e}^{*}\underline{e}^{T}\}D^{\prime T}+D^{\prime}E\{\underline{e}\,\underline{e}^{*T}\}D^{*T}+D^{\prime}\operatorname{cov}\underline{e}\,D^{\prime T}=\\&=\sigma_\xi^2 I\end{aligned} \tag{I-11}$$

Neglecting, for $N \gg \nu$, the initial conditions:

$$D^{\prime}\operatorname{cov}\underline{e}\,D^{\prime T}\approx\sigma_\xi^2 I \tag{I-12}$$

and consequently

$$R^{-1}=(\operatorname{cov}\underline{e})^{-1}\approx D^{\prime T}D^{\prime}\sigma_\xi^{-2} \tag{I-13}$$

Appendix II Practical choices of the instrumental variable

For explicit estimation schemes, the generation of an instrumental variable such as the output of a model gives problems, as several iterations for obtaining reasonable model parameters are needed. Therefore other choices for the instrumental variable have been proposed. Wouters (1972) proposed the use of delayed inputs and outputs and Gersch (1970) the use of delayed outputs as an easy-to-implement instrumental variable. These choices can easily be motivated by the following reasoning.

Assume the description:

$$[1+A]_t y_k = [b_o+B]_t u_k + [1+C]_t \xi_k \qquad \text{(II-1)}$$

and

$$e_k = [1+C]_t \xi_k \qquad \text{(II-2)}$$

ξ_k being white noise.

The correlation of the equation error e_k with a shifted signal $y_{k-\ell}$ can be given:

$$E\{e_k y_{k-\ell}\} = E\{[1+C]_t \xi_k y_{k-\ell}\} = \qquad \text{(II-3)}$$

$$= E\{\xi_k y_{k-\ell}\} + c_1 E\{\xi_{k-1} y_{k-\ell}\} + \ldots + c_s E\{\xi_{k-s} y_{k-\ell}\}$$

For $\ell > s$, it can easily be seen that

$$E\{e_k y_{k-\ell}\} = 0 \qquad \text{(II-4)}$$

leading to

$$E\{[[1+A]_t y_k - [b_o+B]_t u_k] y_{k-\ell}\} = 0 \qquad \text{(II-5)}$$

Similarly:

$$E\{e_k u_{k-\ell}\} = 0 \qquad \forall \ell \qquad \text{(II-6)}$$

leading to:

$$E\{[[1+A]_t y_k - [b_o+B]_t u_k] u_{k-\ell}\} = 0 \qquad \text{(II-7)}$$

In principle, from equations (II-5) and (II-7), the parameters can be found. Because the exact cross- and autocorrelations, as they appear in these equations, are not known, they have to be approximated, assuming that the signals are ergodic:

$$\frac{1}{N}\sum_{k=1}^{N}\left[\left[\left[1+A\right]_t y_k - \left[b_o+B\right]_t u_k\right] y_{k-\ell}\right] = \tag{II-8}$$

$$= \frac{1}{N}\sum_{k=1}^{N}\xi_k y_{k-\ell} + \frac{c_1}{N}\sum_{k-1}^{N}\xi_{k-1} y_{k-\ell} + \ldots + \frac{c_s}{N}\sum_{k=1}^{N}\xi_{k-s} y_{k-\ell}$$

$$\frac{1}{N}\sum_{k=1}^{N}\left[\left[\left[1+A\right]_t y_k - \left[b_o+B\right]_t u_k\right] u_{k-\ell}\right] = \tag{II-9}$$

$$= \frac{1}{N}\sum_{k=1}^{N}\xi_k u_{k-\ell} + \frac{c_1}{N}\sum_{k=1}^{N}\xi_{k-1} u_{k-\ell} + \ldots + \frac{c_s}{N}\sum_{k=1}^{N}\xi_{k-s} u_{k-\ell}$$

Only for N tends to infinity are the right hand sides of equations (II-8) and (II-9) equal to zero, which, in practice, is assumed to be true.

An estimate of the process parameters can then be obtained by solving the equations

$$\frac{1}{N}\sum_{k=1}^{N}\left[\left[\left[1+\hat{A}\right] y_k - \left[\hat{b}_o+\hat{B}\right] u_k\right] y_{k-\ell}\right] = 0 \qquad \ell > s \tag{II-10}$$

$$\frac{1}{N}\sum_{k=1}^{N}\left[\left[\left[1+\hat{A}\right] y_k - \left[\hat{b}_o+\hat{B}\right] u_k\right] u_{k-\ell}\right] = 0 \qquad \forall\ \ell \tag{II-11}$$

Both equations lead to:

$$\underline{\hat{\theta}} = \left[\Omega^T(z_1,z_2)\ \Omega^T(u,y)\right]^{-1}\Omega^T(z_1,z_2)\underline{y} \tag{II-12}$$

where

a) $z_1 = u_k$; $z_2 = u_{k-p-1}$ cf. Wouters (1972).

b) $z_1 = u_k$; $z_2 = y_{k-p-1}$ if $p+1 > s$ cf. Gersch (1970).

The drawback of this method is the fact that the correlation functions, which are contained in the matrix, contain considerable time shifts. Usually this gives problems for finite data sequences. Let us consider, for example, gaussian noise. It can be shown, e.g. Laning and Battin (1956), that the variance of the approximated correlation function

$\hat{\Psi}_{xx}(\tau)$ based on N samples:

$$\Psi_{xx}(\tau) = \frac{1}{N}\sum_{i=1}^{N} x_i x_{i+\tau} \tag{II-13}$$

can be given by:

$$\mathrm{var}\left\{\Psi_{xx}(\tau)\right\} \approx \frac{\Psi_{xx}^2(0)}{N} \qquad \tau \neq 0 \tag{II.14}$$

This variance is fairly independent of τ for $\tau \neq 0$. The correlation functions most frequently encountered tend to small values for large shifts τ.

If we consider equations (II-8) and (II-9) we can have, in a practical example, e.g. the exact relation:

$$\Psi_2 \cdot a + \Psi_1 = \delta \qquad \text{(II-15)}$$

where Ψ_1, Ψ_2 are correlations based on N samples for small time shift, a is the parameter and δ is the right hand term of eq. (II-8) or (II-9). For a practical example with large time shift, we have:

$$\Psi_4 \cdot a' + \Psi_3 = \delta' \qquad \text{(II-16)}$$

where Ψ_3, Ψ_4 are correlations based on N samples for considerable time shift, so Ψ_3, Ψ_4 are one (or several) order(s) of magnitude smaller than Ψ_1,Ψ_2. However δ and δ' are of the same order of magnitude. So we have

$$a = -\frac{\Psi_1}{\Psi_2} + \frac{\delta}{\Psi_2} \quad \rightarrow \quad \hat{a} = -\frac{\Psi_1}{\Psi_2} \qquad \text{(II-17)}$$

$$a' = -\frac{\Psi_3}{\Psi_4} + \frac{\delta'}{\Psi_4} \quad \rightarrow \quad \hat{a}' = -\frac{\Psi_3}{\Psi_4} \qquad \text{(II-18)}$$

As δ and δ' are unknown, they will be approximated by zero. We see that the error in $\hat{a}'$ is larger than $\hat{a}$.

The method of Gersch can also be used for the estimation of the AR parameters of the process only if certain conditions are fulfilled:

We rewrite the input-output description of the process:

$$\underline{y} = - Y\underline{a}_t + U\underline{b}_t + \underline{e} \qquad \text{(II-19)}$$

$$\underline{e} = \Xi\underline{c}_t + \underline{\xi}$$

$$\underline{y} = - Y\underline{a}_t + U\underline{b}_t + \Xi\underline{c}_t + \underline{\xi} \qquad \text{(II-20)}$$

Premultiply eq. (II-20) with $Y'^T = z^{-\ell}Y^T$

$$Y'^T\underline{y} = -Y'^TY\underline{a}_t + Y'^TU\underline{b}_t + Y'^T\Xi\underline{c}_t + Y'^T\underline{\xi} \qquad \text{(II-21)}$$

so

$$\underline{a}_t = -(Y'^TY)^{-1}Y'^T\underline{y} + (Y'^TY)^{-1}Y'^TU\underline{b}_t + (Y'^TY)^{-1}Y'^T\Xi\underline{c}_t + (Y'^TY)^{-1}Y'^T\underline{\xi}$$

The estimator

$$\hat{\underline{a}} = -(Y'^TY)^{-1}Y'^T\underline{y} \qquad \text{(II-22)}$$

is consistent if:

$$\left.\begin{aligned}
&\text{a)}\ \operatorname*{plim}_{N\to\infty}\left[\tfrac{1}{N}(Y'^{T}Y)\right] \qquad \text{is non-singular}\\
&\text{b)}\ \operatorname*{plim}_{N\to\infty}\left[(Y'^{T}Y)^{-1}Y'^{T}\underline{Ub}_t\right] = \underline{0}\\
&\text{c)}\ \operatorname*{plim}_{N\to\infty}\left[(Y'^{T}Y)^{-1}Y'^{T}\underline{\Xi c}_t\right] = \underline{0}\\
&\text{d)}\ \operatorname*{plim}_{N\to\infty}\left[(Y'^{T}Y)^{-1}Y'^{T}\underline{\xi}\right] = \underline{0}
\end{aligned}\right\} \qquad \text{(II-23)}$$

ad a) For non-singularity of

$$\operatorname*{plim}_{N\to\infty}\left[\tfrac{1}{N}(Y'^{T}Y)\right] \qquad \text{(II-24)}$$

it is important that the time shift of Y' with respect to Y is not too large. This is dictated by the length p' of the MA filter $[b'_o+B']_t$ approximating the ARMA process

$$[b'_o+B']_t \approx \frac{[b_o+B]_t}{[1+A]_t} \qquad \text{(II-25)}$$

The MA filter $[b_o''+B'']$ with memory length p'' is used for shaping u_k from a white noise source w_k

$$u_k = [b''_o+B'']w_k \qquad \text{(II-26)}$$

So the length of the non zero part of the autocorrelation functions of y_k is p'+p''. The delay ℓ should be such that $\ell+q<p'+p''$.

ad b) Write eq. (II-26) in matrix notation:

$$U = WB'' \qquad \text{(II-27)}$$

So

$$\operatorname*{plim}_{N\to\infty}\left[(Y'^{T}Y)^{-1}Y'^{T}\underline{Ub}_t\right] = \operatorname*{plim}_{N\to\infty}\left[\tfrac{1}{N}(Y'^{T}Y)^{-1}\right]\operatorname*{plim}_{N\to\infty}\left[\tfrac{1}{N}Y'^{T}WB''\underline{b}_t\right] \qquad \text{(II-28)}$$

Now

$$\operatorname*{plim}_{N\to\infty}\left[\tfrac{1}{N}Y'^{T}W\right] = \begin{bmatrix} \Psi_{yw}(\ell+1) & \cdots & \Psi_{yw}(\ell+1-p+p'') \\ \vdots & & \vdots \\ \Psi_{yw}(\ell+q) & \cdots & \Psi_{yw}(\ell+q-p-p'') \end{bmatrix} \qquad \text{(II-29)}$$

This matrix equals zero if $\ell > p+p''$, as w_k is a white noise sequence.

ad c) This situation is identical to the condition $\ell > s$ in equation (II-10).

For possible estimation of the AR parameters of the process, ℓ should be chosen such that

$$\ell < p' + p'' - q \qquad \text{(II-30)}$$

and

$$\ell > \max(p+p'', s) \qquad \text{(II-31)}$$

Usually $p'-q>s$ so that the method can be used, but, however, with great care to avoid inconsistence as indicated by eq. (II-23).

Appendix III Relation between Tally estimator and instrumental variable estimator

The same equations (II-4) and (II-6) which led to the already mentioned instrumental variable variants of Wouters and Gersch, lead to an estimation method which is known in the literature as Prior Knowledge Fitting estimator (PKF) or Tally estimator; cf. Peterka and Halouskova (1970), Banyasz and Keviczky (1974) and Bosch (1978).
If we rewrite these equations as correlation functions, we have:

$$\left.\begin{aligned} \Psi_{eu}(-\ell) &= 0 \qquad \forall \ell \\ \Psi_{ey}(-\ell) &= 0 \qquad \ell > s \end{aligned}\right] \qquad \text{(III-1)}$$

For given model parameters $\underline{\theta}^*$ using N observations of input and output signals, we can compute estimates of these correlation functions, which will be denoted as $\tilde{\Psi}_{eu}(\ell;\underline{\theta}^*,\underline{u},\underline{y})$ and $\tilde{\Psi}_{ey}(\ell;\underline{\theta}^*,\underline{u},\underline{y})$.

An estimate $\hat{\underline{\theta}}$ is sought for which

$$\| \underline{\tilde{\Psi}}(\hat{\theta},u,y)-\underline{\Psi} \| \text{ is minimal} \qquad \text{(III-2)}$$

where $\underline{\tilde{\Psi}}$ and Ψ are vectors containing points of the above mentioned correlation functions for different ℓ.

As $\underline{\Psi} = \underline{0}$, eq. (III-2) can be rewritten:

$$\| \underline{\tilde{\Psi}}(\hat{\theta},u,y) \| \quad \text{is minimal.} \qquad \text{(III-3)}$$

We have already met the following equations cf. (II-8) and (II-9)

$$\tilde{\Psi}_{\hat{eu}}(-\ell) = \frac{1}{N}\Big[\sum_{k=1}^{N} y_k u_{k-\ell} + \sum_{j=1}^{q} \hat{a}_j \sum_{k=1}^{N} y_{k-j} u_{k-\ell} - \sum_{j=o}^{p} \hat{b}_j \sum_{k=1}^{N} u_{k-j} u_{k-\ell}\Big]$$

$$\tilde{\Psi}_{\hat{ey}}(-\ell) = \frac{1}{N}\Big[\sum_{k=1}^{N} y_k y_{k-\ell} + \sum_{j=1}^{q} \hat{a}_j \sum_{k=1}^{N} y_{k-j} y_{k-\ell} - \sum_{j=o}^{p} \hat{b}_j \sum_{k=1}^{N} u_{k-j} y_{k-\ell}\Big] \qquad \text{(III-4)}$$

Now construct

$$\underline{\tilde{\Psi}}^T = (\tilde{\Psi}_{\hat{eu}}(-1),\ldots,\tilde{\Psi}_{\hat{eu}}(-M),\ \tilde{\Psi}_{\hat{ey}}(-s-1),\ldots,\tilde{\Psi}_{\hat{ey}}(-M)) \qquad \text{(III-5)}$$

$$\underline{\tilde{\Phi}}^T = (\tilde{\Psi}_{yu}(-1),\ldots,\tilde{\Psi}_{yu}(-M),\ \tilde{\Psi}_{yy}(-s-1),\ldots,\tilde{\Psi}_{yy}(-M)) \qquad \text{(III-6)}$$

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\psi}_{uu}(-1) & \cdots & \tilde{\psi}_{uu}(p-1) & \tilde{\psi}_{yu}(0) & \cdots & \tilde{\psi}_{yu}(q-1) \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{\psi}_{uu}(-M) & \cdots & \tilde{\psi}_{uu}(p-M) & \tilde{\psi}_{yu}(-M+1) & \cdots & \tilde{\psi}_{yu}(q-M) \\ \tilde{\psi}_{uy}(-s-1) & \cdots & \tilde{\psi}_{uy}(-s-1+p) & \tilde{\psi}_{yy}(-s) & \cdots & \tilde{\psi}_{yy}(-s+q-1) \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{\psi}_{uy}(-M) & \cdots & \tilde{\psi}_{uy}(-M+p) & \tilde{\psi}_{yy}(-M+1) & \cdots & \tilde{\psi}_{yy}(-M+q) \end{bmatrix} \qquad \text{(III-7)}$$

M should be chosen such that all equations involved are useful, i.e. the auto- and cross- correlations involved should not be too small for M. Peterka suggests $M > 3q/2$.

Using the above given notation we have

$$\underline{\tilde{\psi}} = \tilde{\Psi}\,\underline{\hat{\theta}} + \underline{\tilde{\phi}} \qquad \text{(III-8)}$$

Minimization of $\underline{\tilde{\psi}}^T\underline{\tilde{\psi}}$ leads to an estimate $\underline{\hat{\theta}}$. Peterka and Halouskova (1970) use orthogonal triangularization for this minimization as already indicated by Peterka and Smuk (1969) and Smuk (1970). Note, for example, that (3.82) is an equation in the second order moments of the measured data. Differentiation of $\underline{\tilde{\psi}}^T\underline{\tilde{\psi}}$ with respect to $\underline{\hat{\theta}}$ yields:

$$\underline{\hat{\theta}} = -(\tilde{\Psi}^T\tilde{\Psi})^{-1}\tilde{\Psi}^T\underline{\tilde{\phi}} \qquad \text{(III-9)}$$

In this expression, fourth order moments of the measured data are involved. If $M = p+q+1$ is taken, then matrix $\tilde{\Psi}$ becomes a square matrix so that eq. (III-9) can be rewritten as:

$$\underline{\hat{\theta}} = -\tilde{\Psi}^{-1}\underline{\tilde{\phi}} \qquad \text{(III-10)}$$

At this point the close relationship with the instrumental variable method can be shown; cf. Bosch (1978). If the following instrumental variable matrix Z is used

$$Z^T = \begin{bmatrix} 0 & u_1 & \cdots & u_N \\ 0 & 0 & u_1 & u_{N-M} \\ 0 & 0 & y_1 & y_{N-s-1} \\ 0 & \cdots & 0 & y_1 \cdots y_{N-M} \end{bmatrix} \qquad \text{(III-11)}$$

then

$$\underline{\tilde{\Psi}} = \frac{1}{N} Z^T \underline{\hat{e}} \qquad \text{(III-12)}$$

and

$$\underline{\tilde{\Psi}}^T \underline{\tilde{\Psi}} = \frac{1}{N^2} \underline{\hat{e}}^T Z Z^T \underline{\hat{e}} \qquad \text{(III-13)}$$

which is the instrumental variable criterion, as already met in eq. (3.45). It can easily be shown that this choice of instrumental variable is a legitimate one, i.e. the conditions (3.51) are fulfilled:

$$\left.\begin{array}{lll} \text{I} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N} Z^T \Omega\right] = \Gamma & \Gamma \text{ is non-singular} \\ \text{II} & \underset{N\to\infty}{\text{plim}} \left[\frac{1}{N} Z^T \underline{e}\right] = \underline{0} & \end{array}\right] \qquad \text{(III-14)}$$

Ψ is a $(p+q+1) \times (2M-s)$ matrix.

By taking M not too large and using persistently exciting inputs u_k, condition I can be fulfilled. Condition II is assured as $\underline{\Psi} = \underline{0}$ is a priori information; cf. (III-1).

The difference between this Tally method and the already met instrumental variable method is only in the choice of M. For the Tally estimator, we will take, as suggested by Peterka, $M > 3q/2$ leading to $2M-s = 3q-s$ equations. As $p \sim q \sim s$, these are $2q$ equations. The instrumental variable estimator with delayed outputs (Gersch) takes $2p+q+1$ equations leading to square matrices and simpler programming. For high signal to noise ratios the use of more than $p+q+1$ equations is not necessary for noise suppression.

Appendix IV Derivation of the information matrix

The logarithm of the likelihood function is:

$$L = \text{constant} - N \ln \lambda - \frac{1}{2\lambda^2} \sum_{k=1}^{N} \hat{\xi}_k^2 \qquad \text{(IV-1)}$$

For the second derivatives with respect to the parameters and with respect to λ we find:

$$\frac{\partial^2 L}{\partial \lambda^2} = -\frac{3}{\lambda^4} \sum_{k=1}^{N} \hat{\xi}_k^2 + \frac{N}{\lambda^2} \qquad \text{(IV-2)}$$

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = -\frac{1}{\lambda^2} \sum_{k=1}^{N} \frac{\partial \hat{\xi}_k}{\partial \theta_i} \cdot \frac{\partial \hat{\xi}_k}{\partial \theta_j} - \frac{1}{\lambda^2} \sum_{k=1}^{N} \hat{\xi}_k \frac{\partial^2 \hat{\xi}_k}{\partial \theta_i \partial \theta_j} \qquad \text{(IV-3)}$$

$$\frac{\partial^2 L}{\partial \theta_i \partial \lambda} = \frac{2}{\lambda^3} \sum_{k=1}^{N} \hat{\xi}_k \frac{\partial \hat{\xi}_k}{\partial \theta_i} \qquad \text{(IV-4)}$$

If we keep in mind the model:

$$[1+A]y_k = [b_o+B]u_k + \lambda \frac{[1+C]}{[1+D]} \hat{\xi}_k \qquad \text{(IV-5)}$$

we find,

$$\frac{[1+C]}{[1+D]} \frac{\partial \hat{\xi}_k}{\partial a_i} = y_{k-i} = \frac{[b_o+B]}{[1+A]} u_{k-i} + \lambda \frac{[1+C]}{[1+A][1+D]} \hat{\xi}_{k-i} \qquad \text{(IV-6)}$$

$$\frac{[1+C]}{[1+D]} \frac{\partial \hat{\xi}_k}{\partial b_i} = -u_{k-i} \qquad \text{(IV-7)}$$

$$\frac{1}{[1+D]} \hat{\xi}_{k-i} + \frac{[1+C]}{[1+D]} \frac{\partial \hat{\xi}_k}{\partial c_i} = 0 \rightarrow [1+C] \frac{\partial \hat{\xi}_k}{\partial c_i} = -\hat{\xi}_{k-i} \qquad \text{(IV-8)}$$

$$-\frac{[1+C]}{[1+D]^2} \hat{\xi}_{k-i} + \frac{[1+C]}{[1+D]} \frac{\partial \hat{\xi}_k}{\partial d_i} = 0 \rightarrow \frac{-1}{[1+D]} \hat{\xi}_{k-i} + \frac{\partial \hat{\xi}_k}{\partial d_i} = 0 \qquad \text{(IV-9)}$$

This can be rewritten:

$$\frac{\partial \hat{\xi}_k}{\partial a_i} = \frac{[b_o+B][1+D]}{[1+A][1+C]} u_{k-i} + \frac{\lambda}{[1+A]} \hat{\xi}_{k-i} = \frac{[1+D]}{[1+C]} y_{k-i} \qquad \text{(IV-10)}$$

$$\frac{\partial \hat{\xi}_k}{\partial b_i} = - \frac{[1+D]}{[1+C]} u_{k-i} \qquad \text{(IV-11)}$$

$$\frac{\partial \hat{\xi}_k}{\partial c_i} = - \frac{\lambda}{[1+C]} \hat{\xi}_{k-i} \qquad \text{(IV-12)}$$

$$\frac{\partial \hat{\xi}_k}{\partial d_i} = \frac{\lambda}{[1+D]} \hat{\xi}_{k-i} \qquad \text{(IV-13)}$$

Take the mathematical expectations of eq. (IV-2), (IV-3) and (IV-4):

$$E\{\frac{\partial^2 L}{\partial \lambda^2}\} = - \frac{2N}{\lambda^2} \qquad \text{(IV-14)}$$

$$E\{\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\} = - \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial \theta_i} \cdot \frac{\partial \hat{\xi}_k}{\partial \theta_j}\} \qquad \text{(IV-15)}$$

$$E\{\frac{\partial^2 L}{\partial \theta_i \partial \lambda}\} = 0 \qquad \text{(IV-16)}$$

Introduce the variables:

$$x_{1,k} = \frac{[b_o+B][1+D]}{[1+A][1+C]} u_k \qquad \text{(IV-17)}$$

$$x_{2,k} = \frac{[1+D]}{[1+C]} u_k \qquad \text{(IV-18)}$$

$$x_{3,k} = \frac{1}{[1+A]} \hat{\xi}_k \qquad \text{(IV-19)}$$

$$x_{4,k} = \frac{1}{[1+D]} \hat{\xi}_k \qquad \text{(IV-20)}$$

$$x_{5,k} = \frac{1}{[1+C]} \hat{\xi}_k \qquad \text{(IV-21)}$$

These variables are stationary random processes, as the polynomials $[1+A]$, $[1+C]$ and $[1+D]$ have zeros inside the unit circle of the z-plane.

Introduce the correlation functions which we applied to the above mentioned variables $x_{1,k},\cdots,x_{5,k}$:

$$r_{ij}(\tau) = \frac{1}{N} \sum_{k=1}^{N} x_{i,k} x_{j,k+\tau} \qquad i,j = \{1,2\}$$

$$r_{ij}(\tau) = E\{x_{i,k}\, x_{j,k+\tau}\} \qquad i,j = \{3,4,5\} \qquad \text{(IV-22)}$$

Then we form the following sub-matrices:

$$(J_{aa})_{ij} = - E\{\frac{\partial^2 L}{\partial a_i \partial a_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial a_i} \frac{\partial \hat{\xi}_k}{\partial a_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{(\frac{[b_o+B][1+D]}{[1+A][1+C]} u_{k-i} + \frac{\lambda}{[1+A]} \hat{\xi}_{k-i})(\frac{[b_o+B][1+D]}{[1+A][1+C]} u_{k-j}$$

$$+ \frac{\lambda}{[1+A]} \hat{\xi}_{k-j})\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} x_{1,k-i} x_{1,k-j} + \lambda^2 E\{x_{3,k-i} x_{3,k-j}\} =$$

$$= \frac{N}{\lambda^2} \{r_{11}(i-j) + \lambda^2 r_{33}(i-j)\} \qquad \text{(IV-23)}$$

$$-(J_{ab})_{ij} = - E\{\frac{\partial^2 L}{\partial a_i \partial b_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial a_i} \frac{\partial \hat{\xi}_k}{\partial b_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{(\frac{[b_o+B][1+D]}{[1+A][1+C]} u_{k-i} + \frac{\lambda}{[1+A]} \hat{\xi}_{k-i})(-\frac{[1+D]}{[1+C]} u_{k-j})\} =$$

$$= - \frac{1}{\lambda^2} \sum_{k=1}^{N} x_{1,k-i} x_{2,k-j} = - \frac{N}{\lambda^2} r_{12}(i-j) \qquad \text{(IV-24)}$$

$$-(J_{ac})_{ij} = - E\{\frac{\partial^2 L}{\partial a_i \partial c_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial a_i} \frac{\partial \hat{\xi}_k}{\partial c_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{(\frac{[b_o+B][1+D]}{[1+A][1+C]} u_{k-i} + \frac{\lambda}{[1+A]} \hat{\xi}_{k-i})\,(-\frac{\lambda}{[1+C]} \hat{\xi}_{k-j})\} =$$

$$= - \frac{1}{\lambda^2} \sum_{k=1}^{N} \lambda^2 E\{x_{3,k-i} x_{5,k-j}\} = - N\, r_{35}(i-j) \qquad \text{(IV-25)}$$

$$(J_{ad})_{ij} = - E\{\frac{\partial^2 L}{\partial a_i \partial d_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial a_i} \frac{\partial \hat{\xi}_k}{\partial d_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{(\frac{[b_o+B][1+D]}{[1+A][1+C]} u_{k-i} + \frac{\lambda}{[1+A]} \hat{\xi}_{k-i})(\frac{\lambda}{[1+D]} \hat{\xi}_{k-j})\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} \lambda^2 \, E\{x_{3,k-i} x_{4,k-j}\} = N \, r_{34}(i-j) \qquad \text{(IV-26)}$$

$$(J_{bb})_{ij} = - E\{\frac{\partial^2 L}{\partial b_i \partial b_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial b_i} \frac{\partial \hat{\xi}_k}{\partial b_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} - \frac{[1+D]}{[1+C]} u_{k-i} \cdot - \frac{[1+D]}{[1+C]} u_{k-j} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} x_{2,k-i} x_{2,k-j} = \frac{N}{\lambda^2} r_{22}(i-j) \qquad \text{(IV-27)}$$

$$(J_{bc})_{ij} = - E\{\frac{\partial^2 L}{\partial b_i \partial c_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial b_i} \frac{\partial \hat{\xi}_k}{\partial c_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{- \frac{[1+D]}{[1+C]} u_{k-i} \cdot \frac{-\lambda}{[1+C]} \hat{\xi}_{k-j}\} = 0 \qquad \text{(IV-28)}$$

$$- (J_{bd})_{ij} = - E\{\frac{\partial^2 L}{\partial b_i \partial d_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial b_i} \frac{\partial \hat{\xi}_k}{\partial d_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{- \frac{[1+D]}{[1+C]} u_{k-i} \cdot \frac{\lambda}{[1+D]} \hat{\xi}_{k-j}\} = 0 \qquad \text{(IV-29)}$$

$$(J_{cc})_{ij} = - E\{\frac{\partial^2 L}{\partial c_i \partial c_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial c_i} \frac{\partial \hat{\xi}_k}{\partial c_j}\} =$$

$$= \frac{1}{\lambda^2} \sum_{k=1}^{N} \lambda^2 E\{\frac{1}{[1+C]} \hat{\xi}_{k-i} \cdot \frac{1}{[1+C]} \hat{\xi}_{k-j}\} = \sum_{k=1}^{N} E\{x_{5,k-i} x_{5,k-j}\} =$$

$$= N \, r_{55}(i-j) \qquad \text{(IV-30)}$$

$$- (J_{cd})_{ij} = - E\{\frac{\partial^2 L}{\partial c_i \partial d_j}\} = \frac{1}{\lambda^2} \sum_{k=1}^{N} E\{\frac{\partial \hat{\xi}_k}{\partial c_i} \frac{\partial \hat{\xi}_k}{\partial d_j}\} =$$

$$= - \frac{1}{\lambda^2} \sum_{k=1}^{N} \lambda^2 E\{\frac{1}{[1+C]} \xi_{k-i} \cdot \frac{1}{[1+D]} \xi_{k-j}\} = - \sum_{k=1}^{N} E\{x_{5,k-i} x_{4,k-j}\} =$$

$$= - N \, r_{54}(i-j) \qquad \text{(IV-31)}$$

$$(J_{dd})_{ij} = -E\left\{\frac{\partial^2 L}{\partial d_i \partial d_j}\right\} = \frac{1}{\lambda^2}\sum_{k=1}^{N} E\left\{\frac{\partial \hat{\xi}_k}{\partial d_i}\,\frac{\partial \hat{\xi}_k}{\partial d_j}\right\} =$$

$$= \frac{1}{\lambda^2}\sum_{k=1}^{N} \lambda^2 E\left\{\frac{1}{[1+D]}\,\hat{\xi}_{k-i}\cdot\frac{1}{[1+D]}\,\hat{\xi}_{k-j}\right\} = \sum_{k=1}^{N} E\{x_{4,k-i}x_{4,k-j}\} =$$

$$= N\, r_{44}(i-j) \qquad \text{(IV-32)}$$

Now construct the Fisher information matrix:

$$J = \begin{bmatrix} J_{bb} & J_{ba} & J_{bc} & J_{bd} & J_{b\lambda} \\ J_{ab} & J_{aa} & J_{ac} & J_{ad} & J_{a\lambda} \\ J_{cb} & J_{ca} & J_{cc} & J_{cd} & J_{c\lambda} \\ J_{db} & J_{da} & J_{dc} & J_{dd} & J_{d\lambda} \\ J_{\lambda b} & J_{\lambda a} & J_{\lambda c} & J_{\lambda d} & J_{\lambda\lambda} \end{bmatrix} \qquad \text{(IV-33)}$$

With the above results we have:

$$J = N\begin{bmatrix} \{\frac{1}{\lambda^2} r_{22}(i-j)\} & \{\frac{1}{\lambda^2} r_{12}(i-j)\} & \emptyset & \emptyset & 0 \\ \{\frac{1}{\lambda^2} r_{21}(i-j)\} & \{\frac{1}{\lambda^2} r_{11}(i-j)+r_{33}(i-j)\} & \{r_{35}(i-j)\} & \{r_{34}(i-j)\} & 0 \\ \emptyset & \{r_{53}(i-j)\} & \{r_{55}(i-j)\} & \{r_{54}(i-j)\} & 0 \\ \emptyset & \{r_{43}(i-j)\} & \{r_{45}(i-j)\} & \{r_{44}(i-j)\} & 0 \\ 0 & 0 & 0 & 0 & \frac{2N}{\lambda^2} \end{bmatrix}$$

(IV-34)

Appendix V Mathematical results

Result V.1: Slutsky's theorem

If the stochastic process $\{x_k\}$; $k = 1,2,\ldots$ in R^p converges in probability towards x and if the function g, which is a transformation from R^p into R^q, is continuous, then:

$$\operatorname*{plim}_{N\to\infty} [g(x_N)] = g(\operatorname*{plim}_{N\to\infty} [x_N]) = g(x)$$

If the elements of the matrices A_N and B_N have a probability limit, then:

$$\operatorname*{plim}_{N\to\infty} [A_N^{-1}B_N] = \{\operatorname*{plim}_{N\to\infty}[A_N]\}^{-1} \operatorname*{plim}_{N\to\infty}[B_N]$$

Proof: cf. Goldberger (1964).

Result V.2: Matrix inversion lemma

$$[A+BC]^{-1} = A^{-1} - A^{-1}B[I+CA^{-1}B]^{-1}CA^{-1}$$

Proof follows from multiplication by A+BC

Result V.3:

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

$$\det M = \det M_{22} \det(M_{11}-M_{12}M_{22}^{-1}M_{21})$$

Proof:

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -M_{22}^{-1}M_{21} & M_{22}^{-1} \end{bmatrix} = \begin{bmatrix} M_{11}-M_{12}M_{22}^{-1}M_{21} & M_{12}M_{22}^{-1} \\ 0 & I \end{bmatrix}$$

So $\det M.\det M_{22}^{-1} = \det(M_{11}-M_{12}M_{22}^{-1}M_{21})$

Result V.4:

If $$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

then $$M^{-1} = \begin{bmatrix} (M_{11}-M_{12}M_{22}^{-1}M_{21})^{-1} & -(M_{11}-M_{12}M_{22}^{-1}M_{21})^{-1}M_{12}M_{22}^{-1} \\ -(M_{22}-M_{21}M_{11}^{-1}M_{12})^{-1}M_{21}M_{11}^{-1} & (M_{22}-M_{21}M_{11}^{-1}M_{12})^{-1} \end{bmatrix}$$

The proof follows from $MM^{-1} = I$.

Appendix VI Notations, symbols and abbreviations

On the notation

As in many fields, also in identification a characteristic use of terminology has been developed through the many papers and other contributions. In this dissertation no attempt has been made to polish until textbook-presentation is approximated.

One term rather loosely used is "model". This has to be seen in the light of its context consisting of:

- a real (physical) process, most probably partially known, but certainly not fully known or even knowable (dimensionality and related limitations);
- a theoretical model that would adequately represent the real process for the intended use one has in mind when the identification task is tackled;
- an estimation model that should approach the theoretical model as well as possible, given the theoretical and practical limitations imposed by our (a priori) knowledge, experimental capabilities, etc.

From the text this distinction will be clear in the majority of cases.

For the sake of simplicity, it is often assumed that the theoretical model is available in the set of estimation models; it can be determined/approached by the choice of 'order' and by estimation of the parameter values.
In these cases the notation is such that $\hat{}$ indicates the 'replica' in the estimation model of the corresponding quantity in the theoretical model, e.g.

	theoretical model	estimation model
parameters	$\underline{\theta}_t$	$\underline{\theta}$, $\underline{\hat{\theta}}$
coloured noise	e	$\hat{e}$
white noise as the 'origin' of coloured noise	ξ	$\hat{\xi}$

Notations

$\forall$	for all
$\triangleq$	definition
$p(x)$	probability density function of the random variable x
$E\{x\}$	mathematical expectation of the random variable x
$\underset{k\to\infty}{\text{plim}}[x]$	probability limit of the random variable x
^{T} e.g. X^{T}	transpose of X
$^{-1}$ e.g. X^{-1}	inverse of X
det X	determinant of X
^{i} e.g. $\underline{\hat{\theta}}^{i}$	i-th operation; e.g. estimated vector, i-th iteration
^{i} e.g. $\underline{\tilde{u}}^{i}$	do. ; e.g. filtered signal u, i-th filtering
$_{k}$ e.g. u_k, y_k	k-th instant of time; e.g. samples of u,y at time k
$_{k}$ e.g. $\underline{\hat{\theta}}_{k}$	do. ; e.g. estimated parameter vector, k-th recursion.
$_{t}$ e.g. $\underline{\theta}_{t}$	"true" parameter vector.
$\sim$ $\approx$	filtering
$\vert$ e.g. $[X_1 \vert X_2]$	X_1 augmented with X_2
$\hat{}$, e.g. $\underline{\hat{\theta}}$	estimated parameter vector
$\hat{}$, e.g. $\underline{\hat{\xi}}$, $\underline{\hat{e}}$	prediction error
$\hat{\hat{}}$, e.g. $\underline{\hat{\hat{\xi}}}$, $\underline{\hat{\hat{e}}}$	residuals

Symbols

symbol	description	occurs in chapter
$\emptyset$	null matrix	
$\underline{a}$	parameter vector; AR param.	3,4,5,7
a	AR param.	3,4,5,6,7
$\underline{a}'$	parameter vector; AR param. (extended)	3,4,5,6,7

$[1+A]$	polynomial in z^{-1}; AR parameters	2,3,4,5,7
$[1+A']$	do., with model extension	3,4,7
A	system matrix state space model	2
$\underline{b}$	parameter vector; MA param.	3,4,5,7
b	MA param.	3,4,5,6,7
$\underline{b}'$	(extended) parameter vector; MA param.	3,4,7
$[b_o+B]$	polynomial in z^{-1}; MA param.	2,3,4,5,7
$[b_o'+B']$	do., with model extension	3,4,7
B	distribution matrix state space model	2
$\underline{c}$	parameter vector; MA noise param.	3,4,7
c	MA noise param.	3,4,6,7
$\underline{c}'$	(extended) parameter vector: MA noise param.	3,4
$[1+C]$	polynomial in z^{-1}; MA noise param.	2,3,4,5,7
$[1+C']$	do., extended	3,4
C	output matrix state space model	2
C	data product moment matrix	7
$\underline{d}$	parameter vector; AR noise param.	3,4,5,6,7
d	AR noise param.	3,4,5,6,7
$\underline{d}'$	(extended) parameter vector; AR noise param.	3,4
$\underline{d}_i$	parameter vector; AR noise param. (input noise)	5,6
$\underline{d}_o$	parameter vector; AR noise param. (output noise)	5,6
$[1+D]$	polynomial in z^{-1}; AR noise param.	2,3,4,5,7
$[1+D']$	do. ; AR noise param.	3,4
$[1+D_i]$	do. ; AR input noise param.	5
$[1+D_o]$	do. ; AR output noise param.	5
D'	noise whitening filter matrix	3
D	set of parameter estimates (identifiability conditions)	2
e_k	modelled noise sample at time k	2
$\underline{e}$	equation error, process noise (due to output noise)	3,4,5,7
$\underline{\hat{e}}$	prediction error, model error	3,4,5,7
$\underline{\hat{\hat{e}}}$	residual	3,4
E	matrix containing $\underline{e}$	3,4,7
$\hat{E}$	matrix containing $\underline{\hat{e}}$	3,4,7

$\underline{f}(\ )$	vector function (convergence discuss.)	4
$\underline{f}$	equation error (due to input noise)	5
F	matrix containing $\underline{f}$	5
F_o, F_i	function or functional of model, operating on process input viz. - output	2
$\underline{g}$	equation error (due to input- and output noise)	5
$\hat{\underline{g}}$	prediction error (due to input- and output noise)	5
G( )	matrix function (convergence discuss.)	4
G(z)	transfer function	
	- for noise spectral factorization	2
	- for stochastic system	2
H(z)	transfer function	
	- noise shaping filter	2
	- of (state space) model	2
H	matrix (convergence discuss.)	4
H	Hankel matrix	2
H	data product moment matrix	7
I	identity matrix	
J	Fisher information matrix	3
J	identification method (identifiability conditions)	2
k	instant of time	2,3,4,5,6,7
L	(log) likelihood function	3,4
M	Markov parameters, -matrix	
	impulse response matrix	2
M_k	do., for the k-th time instant	2
M	model structure	2,3,4,5
$\underline{n}$	(equivalent) output noise	3,4,7
n_k	(equivalent) output noise sample at time k	3,4,5
$\underline{n}_i$	input noise	5
$\underline{n}_o$	output noise	5
N	total number of observations	2,3,4,5,6,7
N_i	matrix containing $\underline{n}_i$	5
N_o	matrix containing $\underline{n}_o$	5
p	probability density function	3,4
P_k	'covariance matrix' of an estimate at time k	4,5,7

p	number of components in parameter vector	2
p	degree of polynomial $[b_o+B]$	3,4,5,7
q	degree of polynomial $[1+A]$	3,4,5,7
Q	gain constant in recursion	4
Q_k	"covariance matrix" of an estimate at time k	4
Q	data product moment matrix	7
r	degree of polynomial $[1+D]$	3,4,6,7
r_i	degree of polynomial $[1+D_i]$	5,6
r_o	degree of polynomial $[1+D_o]$	5,6
$r(\tau)$	correlation function	3,7
R	covariance matrix of noise	3,4
R	(convergence discuss.)	4
s	degree of polynomial $[1+C]$	3,4,6,7
S	stochastic system (identifiabilty conditions)	2
t	"true"	3,4,5
t	test quantity	7
T	transformation matrix realization	2
u_k	input sample at time k	2,3,4,5,6,7
$\underline{u}$	input vector	2,3,4,5,6,7
$\underline{u}_k$	input vector state space model at time k	2
U	matrix containing $\underline{u}$	3,4,5,7
$\underline{v}$	disturbed input vector	5,6
V	matrix containing $\underline{v}$	5
V_k	quadratic error criterion at time k	4
V	quadratic error criterion/loss function	3
V'	gradient of - -	3
V''	second derivative of - -	3
$\underline{w}$	error	3,4
W	spectrum of noise	2
W	weighting matrix	3
$\underline{x}$	undisturbed output signal	2,3,4,5,7
x_k	undisturbed output signal sample at time k	3,4,5,7
$\underline{x}_k$	state vector at time k	2
x	(auxil. var.)	3
X	experimental condition (identifiability conditions)	2

$\underline{y}$	disturbed output vector	3,4,5,6,7
y_k	disturbed output sample at time k	2,3,4,5,6,7
$\underline{y}_k$	output vector of state space model at time k	2
z	shift operator; Z-transform variable	2,3,4,5,6,7
$\underline{z}$	instrumental variable vector	3,4,5,7
z_k	instrumental variable sample at time k	3,4,5,6,7
Z	instrumental variable matrix	
	template function matrix	1,3,4,5,7
Γ	non-singular matrix	3,5
ε_k	model error sample at time k	2
ε	identification error for N observations	2
$\underline{\theta}$	parameter vector of model	2,3,4,5,7
$\underline{\theta}_t$	"true" parameter vector	3,4,5
$\underline{\hat{\theta}}$	estimated parameter vector	3,4,5,7
$\underline{\hat{\theta}}^{PR}$	estimated process parameter vector	4
$\underline{\hat{\theta}}^{N}$	estimated noise parameter vector	4
λ	"power level" of the white noise	3,4,6,7
μ	model structure	2
$\underline{\xi}$	input of noise filter; white noise	2,3,4,6,7
ξ_k	do.; at time k	2,3,4,6,7
$\underline{\xi}_i$	input of input noise filter;white noise	5
$\underline{\xi}_o$	input of output noise filter;white noise	5
$\underline{\hat{\xi}}$	prediction error	4,7
$\underline{\hat{\xi}}$	residual	3,4,7
Σ	covariance matrix of white noise signal	2
τ, τ'	delay	5,7
Φ	filter producing IV quantities	3
Ψ	(approx.) correlation function	3
$\underline{\omega}_k$	vector of measurables at time k	4
Ω	matrix containing measurables	1, def.in3,4
$\underline{\omega}^{PR}$	vector of measurables, process part	4
$\underline{\omega}^{N}$	vector of measurables, noise part	4

Abbreviations

AIC	average information criterion
AML	approximate maximum likelihood (estimator)
AR	auto regressive
ARMA	auto regressive-moving average
ARMAX	" " " with noise modelling
CRLB	Cramér-Rao lower bound
EECM	equation error compensation method
EMM	extended matrix method
FPE	final prediction error
GLS	generalized least squares (estimator)
GMDH	group method of data handling
IOIVEMM	IVEMM, suited for noise corrupted input-output measurements
IQL	implicit quasi-linearization (method) IV
IV	instrumental variable (estimator)
IVEMM	IV extended matrix method
LS	least squares (estimator)
LSL	least squares like (estimator)
MA	moving average
MIMO	multi input-multi output
MLE	maximum likelihood estimator
MRAS	model reference adaptive system
MSLS	multi-stage least squares (estimator)
PKF	prior knowledge fitting (estimator)
ODE	ordinary differential equation
OLS	over-parametrized least squares (estimator)
RLS	recursive least squares (estimator)
PI	parameter identifiable
SA	stochastic approximation (estimator)
SI	system identifiable
SIV	sub-optimal IV (estimator)
SSI	strongly system identifiable
SISO	single input-single output
WLS	weighted least squares (estimator)

REFERENCES

Akaike, H. (1970)
"Statistical predictor identification", Ann. Inst. Statist. Math., vol. 22, pp. 203-217.

Akaike, H. (1973)
"Information theory and an extension of the maximum likelihood principle", Proc. 2nd Int. Symp. on Information Theory, Tsahkadsor, Armenia, Ed. B.N. Petrov and F. Csaki, Akadémiai, Kiado; pp. 267-281.

Akashi, H. and K.A.F. Moustafa (1975)
"Parameter identification of systems with noise in input and output", Memoirs, Fac. of Eng., Kyoto University, 37. pp. 184-201.

Åström, K.J. and T. Bohlin (1965)
"Numerical identification of linear dynamic systems from normal operating records". Proc. IFAC Symp. on Self Adaptive Control Systems, Teddington, U.K., Ed. P.H. Hammond, pp. 96-111.

Åström, K.J. (1967)
"On the achievable accuracy in identification problems". Prepr. IFAC Symp. on Identification in Aut. Control Systems, Prague, paper 1.8.

Åström, K.J. (1968)
"Lectures on the identification problem - the least squares method". Report 6806 Sept. 1968, Lund Institute of Automatic Control, Lund, Sweden.

Åström, K.J. (1970)
"Introduction to Stochastic Theory". Academic Press, New York.

Åström, K.J. and T. Söderström (1974)
"Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model", IEEE Trans. Aut. Control, vol. AC-19, pp. 769-773.

Banyasz, Cs. and L. Keviczky (1975)
"Some experiments with the P.K.F. discrete time identification method". Problems of Control and Information Theory, vol. 4, pp. 259-274.

Bellman, R. and K.J. Åström (1970)
"On structural identifiability". Mathematical Biosciences, 7 pp. 329-339.

Bohlin, T. (1971)
"On the problem of ambiguities in maximum likelihood identification", Automatica, vol. 7, pp. 199-210.

Bohlin, T. (1978)
"Maximum power validation of models without higher-order fitting", Automatica, vol. 14, pp. 137-146.

Bollen, R. and A.J.W. van den Boom (1982)
"The interactive package SATER:
Part I: Description of the SATER system.
Part II: SATER programmer's manual.
Part III: SATER user's manual".
Internal Reports, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands. (to be published).

Boom, A.J.W. van den and J.H.A.M. Melis (1968)
"A comparison of some process parameter estimating schemes", TH-Report 68-E-63, EE Dept., Eindhoven University of Technology, Netherlands.

Boom, A.J.W. van den and A.W.M. van den Enden (1973)
"The determination of the orders of process- and noise dynamics". Proc. Third IFAC Symp. on Identification and System Parameter Estimation, the Hague/Delft, the Netherlands; also in Automatica, vol. 10, pp. 245-256.

Boom, A.J.W. van den (1976)
"On the relation between weighted least squares estimators and instrumental variable estimators". Proc. Fourth IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, USSR, pp. 1261-1271.

Boom, A.J.W. van den and W.J.M. Lemmens (1977).
"SATER, an interactive program package for education and research in parameter estimation-, control- and signal analysis techniques". IFAC Symp. on Trends in Aut. Control Education, Barcelona, pp. 160-169.

Boom, A.J.W. van den (1979)
"The interactive program package SATER for estimation and control in education and research". Journal A, vol. 20, pp. 92-102.

Bosch, J.P.L. (1978)
"Instrumental variable methods and Tally methods in the interactive package SATER" (in Dutch), M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Box, G.E.P. and G.W. Jenkins (1976)
"Time Series Analysis, Forecasting and Control", Holden Day, San Francisco.

Clarke, D.W. (1967)
"Generalised least squares estimation of the parameters of a dynamic model". Prepr. IFAC Symp. on Identification in Automatic Control Systems, Prague, paper 3.17.

Costongs, J.W. (1979)
"The accuracy of parameter estimation methods in the program package SATER" (in Dutch). M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Cramér, H. (1946)
"Mathematical Methods of Statistics", Princeton University Press

Cullen, C.G. and C.A. Hall (1971)
"On determining whether two polynomials are relatively prime", IEEE Trans. Aut. Control, vol. AC-16, pp. 369-370.

Damen, A.A.H. and A.K. Hajdasinski (1982)
"Practical tests with different approximate realizations based on the singular value decomposition of the Hankel matrix", Preprints, 6th IFAC Symp. on Identification and System Parameter Estimation, Washington D.C., pp. 809-814.

Denham, M.J. (1974)
"Canonical forms for the identification of multivariable linear systems". IEEE Trans. Aut. Control, vol. AC-19, pp. 646-656.

Dugard, L. and I.D. Landau (1980)
"Recursive output error identification algorithm theory and evaluation", Automatica, vol. 16, pp. 443-462.

Dungen, W.T.H.M. van den (1978)
"A generalisation of the least squares estimator" (in Dutch), M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Dungen, W.T.H.M. van den and P. Eykhoff (1981)
"A generalisation of least squares estimation". Prepr./ Proc. 8th IFAC World Congress, Kyoto, paper 251.

Escher, H.A. (1980)
"The recognition of the commands to the interactive program package SATER", M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Eykhoff, P. (1967)
"Process parameter and state estimation". Survey paper, Prepr. First IFAC Symp. on Identification, survey paper 2, Prague.

Eykhoff, P. (1974)
"System Identification, Parameter and State Estimation". John Wiley, London.

Eykhoff, P. (1980)
"System identification: approach to a coherent picture through template functions". Electronics Letters, vol. 16, pp. 502-504.

Eykhoff, P., A.J.W. van den Boom and A.A. van Rede (1981)
"System Identification methods - unification and information development using template functions". Prepr./Proc. 8-th IFAC World Congress, Kyoto.

Eykhoff, P. (1982)
"On the coherence among the multitude of system identification methods". Invited keynote speech, 6th IFAC Symp. on Identification and System Parameter Estimation, Washington D.C., pp. 737-748.

Finigan, B. and I.H. Rowe (1973)
"On the identification of linear discrete time system models using the instrumental variable method". Proc.

Third IFAC Symp. on Identification and System Parameter Estimation, the Hague/Delft, Netherlands, pp. 729-735.

Fisher, R.A. (1922)
"On the mathematical foundations of theoretical statistics". Phil. Trans. R. Soc., A-222 pp. 309-368.

Fuhrt, B. (1973a)
"New estimator for the identification of dynamic processes". IBK report, Institut "Boris Kidric", Vinca, Belgrade.

Fuhrt, B. (1973b)
"Maximum likelihood identification of Åström model by quasilinearisation". Proc. Third IFAC Symp. on Identification and System Parameter Estimation. The Hague/Delft, the Netherlands, pp. 737-740.

Gauss, K.F. (1809)
"Theoria Motus Corporum Coelestium". English version in "Theory of the Motion of the Heavenly Bodies". Dover, New York (1969).

Gersch, W. (1970)
"Estimation of the autoregressive parameters of a mixed autoregressive moving average time series". IEEE Trans. Aut. Control, vol. AC-15, pp. 583-588.

Gevers, M.R. and T. Kailath (1973)
"An innovations approach to least-squares estimation, part IV: Discrete time innovations representation and recursive estimators". IEEE Trans. Aut. Control, vol. AC-18, pp. 588-600.

Goedheer, L.D. (1976)
"Comparison of several identification methods and their mutual relations". M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Goldberger, A.S. (1964)
"Econometric Theory". Wiley, New York.

Golub, G.H. and C. Reinsch (1970)
"Singular value decomposition and least squares solutions", Numer. Math., vol. 14, pp. 403-420.

Goodwin, G.C. and R.L. Payne (1977)
"Dynamic System Identification: Experiment Design and Data Analysis". Academic Press, New York.

Gustavsson, I. (1972)
"Comparison of different methods for identification of industrial processes"
Automatica, vol. 8, pp. 127-142

Gwinner, K. (1976)
"Vereinfachung von Modellen dynamischer Systeme". Regelungstechnik, vol. 24, pp. 325-333.

Haber, R. and L. Keviczky (1976)
"Identification of nonlinear dynamic systems". Proc. Fourth IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, USSR.

Hajdasinski, A.K. and A.A.H. Damen (1979)
"Realisation of the Markov parameter sequences using the singular value decomposition of the Hankel matrix". TH-Report 79-E-95, EE Dept., Eindhoven University of Technology, The Netherlands.

Hajdasinski, A.K. (1980)
"Linear multivariable systems". TH-Report 80-E-106, EE Dept., Eindhoven University of Technology, Netherlands.

Hajdasinski, A.K., P. Eykhoff, A.A.H. Damen and A.J.W. van den Boom (1982)
"The choice and use of different model sets for system identification". Invited contribution, 6th IFAC Symp. on Identification and System Parameter Estimation, Washington D.C.; also as extended version: Internal Report, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Hammerstein, A. (1930)
"Nichtlineare Integralgleichungen nebst Anwendungen", Acta Math., 54, pp. 117-176.

Hastings-James, R. and M.W. Sage (1969)
"Recursive generalised least squares procedure for on line identification of process parameters". Proc. IEE, vol. 116, pp. 2057-2062.

Hipel, K.W. (1981)
"Geophysical model discrimination using the Akaike information criterion", IEEE Trans. Aut. Control, vol. AC-26, pp. 358-378.

Ho, B.L. and R.E. Kalman (1966)
"Effective construction of linear state variable models from input/output functions". Regelungstechnik, vol. 14, pp. 545-548.

Hofman, J.A.M. (1976)
"Testing methods for the determination of the order of linear processes" (in Dutch), M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Hoven, J.T. van den (1978)
"A maximum likelihood method for the interactive program package SATER" (in Dutch). M.Sc. Thesis, Group Measurement and Control, EE Dept. Eindhoven University of Technology, Netherlands.

Hsia, T.C. (1975)
"On multistage least squares approach to systems identification". Proc. IFAC 6th World Congres, Boston, paper 18.2.

Hsia, T.C. (1981)
"A two-stage least squares procedure for systems identification". IEEE Trans. Aut. Control, vol. AC-26, pp. 742-745.

Ivakhnenko, A.G. (1968)
"The group method of data handling - a rival of the method of stochastic approximation". Soviet Autom. Control, vol. 13, pp. 43-55; English translation of the Ukranian journal Avtomatika, vol. 13, pp. 57-72.

Klema, V.C. and A.J. Laub (1980)
"The singular value decomposition: its computation and some application", IEEE Trans. Aut. Control, vol. AC-25, no. 2, pp. 164-176.

Klijn, N.J.P. (1977)
"Identification of linear discrete processes with noise corrupted input and output signals". M.Sc. Thesis, Group

Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Koenraads, A. (1978)
"Methods for estimating the order of linear discrete processes in the interactive package SATER; Part I and II" (in Dutch), M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Krolikowski, A. (1982)
"Model structure selection in linear system identification - survey of methods with emphasis on the information theory approach", Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands. EUT Report 82-E-126.

Laning, J.H. and R.H. Battin (1956)
"Random Processes in Automatic Control". McGraw Hill, New York.

Lee, R.C.K. (1964)
"Optimal Estimation, Identification and Control", Res. Monograph no. 28, MIT Press, Cambridge, Mass.

Lemmens, W.J.M. and A.J.W. van den Boom (1977)
"Interactive computer programs for education and research; a survey". IFAC Symp. on Trends in Control Education, Barcelona; also in Automatica, vol. 15, pp. 113-121.

Lemmens, W.J.M. (1979)
"A review of the system programs of the interactive package SATER" (in Dutch), Internal Report, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Levin, M.J. (1960)
"Optimum estimation of impulsive response in the presence of noise". IRE Trans. Circuit Theory, vol. 7, pp. 50-56.

Ljung, L. (1971)
"Characterization of the concept of 'persistently exciting' in the frequency domain". Report 7119, Lund Institute of Technology, Lund, Sweden.

Ljung, L., I. Gustavsson and T. Söderström (1974)
"Identification of linear multivariable systems operating under linear feedback control". IEEE Trans. Aut. Control, vol. AC-19, pp. 836-840.

Ljung, L., T. Söderström and I. Gustavsson (1975)
"Counter-examples to general convergence of a commonly used recursive identification method". IEEE Trans. Aut. Control, vol. AC-20, pp. 551-575.

Ljung, L. (1976)
"Consistency of the least squares identification method". IEEE Trans. Aut. Control, vol. AC-21, no. 5, pp. 779-781.

Ljung, L. (1977a)
"On positive real functions and the convergence of some recursive schemes". IEEE Trans. Aut. Control, vol. AC-22, pp. 539-551.

Ljung, L. (1977b)
"Analysis of recursive stochastic algorithms". IEEE Trans. Aut. Control, vol. AC, vol. AC-22, pp. 551-575.

Ljung, L. (1979)
"Convergence of recursive estimators". Survey paper, Proc. 5th IFAC Symp. on Identification and System Parameter Estimation, Darmstadt, FRG, pp. 131-144.

Loonstra, F. (1967)
"Introduction to Algebra", (in Dutch), Wolters-Noordhoff, Groningen.

Mehra, R.K. (1974)
"Optimal input signals for parameter estimation in dynamic systems-survey and new results". IEEE Trans. Aut. Control, vol. AC-19, pp. 753-768.

Mehra, R.K. (1981)
"Choice of input signals". In: "Trends and Progress in System Identification", P. Eykhoff, editor, Pergamon Press, Oxford, pp. 305-366.

Niederlinski, A. and A.K. Hajdasinski (1979)
"Multivariable system identification - a survey". Proc. Fifth IFAC Symp. on Identification and System Parameter Estimation, Darmstadt, FRG, pp. 43-76.

Pandya, R.N. (1972)
"A class of bootstrap estimators for linear system identification". Int. J. of Control, vol. 15, pp. 1091-1104.

Pandya, R.N. and B. Pagurek (1973)
"Two-stage least squares estimators and their recursive approximation". Proc. Third IFAC Symp. on Identification and System Parameter Estimation, the Hague/Delft, the Netherlands, pp. 701-710.

Panuska, V. (1969)
"An adaptive recursive least squares identification algorithm". Proc. IEEE Symp. on Adaptive Processes, Pennsylvania State University, USA.

Parzen, E. (1974)
"Some recent advances in time series modelling". IEEE Trans. Aut. Control, vol. AC-19, pp.723-730.

Peterka, V. and K. Smuk (1969)
"On line estimation of dynamic model parameters from input-output data". Prepr. Fourth IFAC World Congress, Warsaw, paper 26.1.

Peterka, V. and A. Halouskova (1970)
"Tally estimate of Åström model for stochastic systems". Prepr. Second Symp. on Identification and Process Parameter Estimation, Prague, paper 2.3.

Popper, K. (1959)
"The Logic of Scientific Discovery". Hutchinson, London.

Rao, C.R. (1945)
"Information and accuracy attainable in the estimation of statistical parameters". Bull. Calcutta Math. Soc., vol. 37, pp. 81-91.

Rajbman, N.S. and V.M. Chadeev (1980)
"Identification of Industrial Processes". North Holland Publ. Co., Amsterdam.

Richalet, J. (1981)
"The model method". In: "Trends and Progress in System Identification". P. Eykhoff, editor, Pergamon Press, Oxford.

Rogers, A.E. and K. Steiglitz (1970)
"On system identification from noise obscured input and output measurements". Int. J. of Control, vol. 12, pp. 625-635.

Rooijakkers, J.C. (1982)
"Parameter estimation on in-vivo signals for determination of the aorta input impedance" (in Dutch), M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Roosdorp, W.B.M. (1974)
"On the identification of linear processes". M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Rowe, I.H. (1970)
"A bootstrap method for the statistical estimation of model parameters". Int. J. of Control, vol. 12, pp. 721-738.

Schulz, E.R. (1968)
"Estimation of pulse transfer function parameters by quasi-linearization", IEEE Trans. Aut. Control, vol. AC-13, pp. 424-426.

Shibata, R. (1976)
"Selection of the order of an autoregressive model by Akaike's information criterion", Biometrika, vol. 63, pp. 117-126.

Silverman, L.M. (1971)
"Realisation of linear dynamical systems". IEEE Trans. Aut. Control, vol. AC-16, pp. 554-567.

Smets, A.J. (1970)
"The instrumental variable method and related identification schemes". TH-Report 70-E-15, EE Dept., Eindhoven University of Technology, Netherlands.

Smets, H.B. (1960)
"Analysis and synthesis of nonlinear systems". IRE Trans. on Circuit Theory, vol. CT-7, pp. 459-469.

Smuk, K. (1970)
Comparison of some algorithms for least squares updating". Prepr. Second Symp. on Identification and Process Parameter Estimation, Prague, paper 2.10.

Söderström, T. (1972)
"On the convergence properties of the generalised least squares identification method". Report 7228, Lund Institute of Technology, Lund, Sweden.

Söderström, T. (1973)
"An on-line algorithm for approximate maximum likelihood identification of linear dynamic systems". Report 7308, Lund Institute of Technology, Lund, Sweden.

Söderström, T. (1974)
"Convergence properties of the generalised least squares identification method". Automatica, vol. 10, pp. 617-626.

Söderström, T. (1975)
"Test of pole-zero cancellation in estimated models", Automatica, vol. 11, pp. 537-540.

Söderström, T. (1977)
"On model structure testing in system identification". Int. J. of Control, vol. 26, pp. 1-18.

Söderström, T., L. Ljung and I. Gustavsson (1978)
"A theoretical analysis of recursive identification methods". Automatica, vol. 14, pp. 231-244.

Söderström, T. (1979)
"Some methods for identification of linear systems with noisy input-output data". Proc. Fifth IFAC Symp. on Identification and System Parameter Estimation, Darmstadt, pp. 357-364.

Söderström, T. (1981)
"On a method of model structure selection in system identification", Automatica, vol. 17, pp. 387-388.

Solo, V. (1979)
"The convergence of AML". IEEE Trans. Aut. Control, vol. AC-24, pp. 958-962.

Solo, V. (1981)
"The convergence of an instrumental-variable-like recursion". Automatica, vol. 17, pp. 545-547.

Sommen, F.L.M. van der (1971)
"The determination of the order of linear systems" (in Dutch), M.Sc. Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Staley, R.M. and P.C. Yue (1970)
"On system parameter identifiability". Information Sciences, vol. 2, pp. 127-138.

Stoica, P. (1976)
"The repeated least squares identification method", Journal A, vol. 17, pp. 151-156.

Stoica, P. and T. Söderström (1977)
"A method for the identification of linear systems using the generalized least squares principle", IEEE Trans. Aut. Control, vol. AC-22, pp. 631-634.

Stoica, P. and T. Söderström (1979)
"Consistency properties of an identification method using the instrumental variable principle". Revue Roumaine des Sciences Techniques, Serie Electrotechnique et Energetique, vol. 24, pp. 289-293.

Stoica, P. (1981)
"On a procedure for structural identification", Int. J. of Control, vol. 33, pp. 1177-1181.

Talmon, J.L. (1971)
"Approximated Gauss-Markov estimators and related schemes". Report 71-E-17, EE Dept., Eindhoven University of Technology, Netherlands.

Talmon, J.L. and A.J.W. van den Boom (1973)
"On the estimation of the transfer function parameters of process- and noise dynamics using a single-stage estimator". Proc. Third IFAC Conference on Identification and System Parameter Estimation, the Hague/Delft, Netherlands, pp. 711-720.

Tse, E. and J.J. Anton (1972)
"On the identifiability of parameters". IEEE Trans. Aut. Control, vol. AC-17, pp. 637-646.

Tse, E. (1978)
"A quantitative measure of identifiability". IEEE Trans. on Syst., Man and Cybern., vol. SMC-8, pp. 1-8.

Tsypkin, Ya. Z. (1966)
"Adaptation, learning and self-learning in control systems". Proc. Third IFAC World Congress, London.

Unbehauen, H. and B. Göhring (1973)
"Application of different statistical tests for the determination of the most accurate order of the model in parameter estimation". Proc. Third IFAC Symp. on Identifica-

tion and System Parameter Estimation, the Hague/Delft, Netherlands, pp. 917-928; also in: Automatica, vol. 10, pp. 233-244.

Valis, J. and I. Gustavsson (1969)
"Some computational results obtained by Panuska's method of stochastic approximations for identification of discrete time systems". Report 6915, Lund Institute of Technology, Lund, Sweden.

Vogt, W.G. and N.K. Bose (1970)
"A method to determine whether two polynomials are relatively prime", IEEE Trans. Aut. Control, vol. AC-15, pp. 379-380.

Volterra, V. (1959)
"Theory of functional and of integral and integro differential equations". Dover, New York.

Wellstead, P. (1978)
"An instrumental product moment test for model order estimation", Automatica, vol. 14, pp. 89-91.

White, R.C. (1971)
"Parameter estimation of a heat exchanger by the instrumental variable method". Report Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Wiener, N. (1958)
"Nonlinear problems in random theory". MIT Press Research Monographs, Cambridge 1958, 140.

Wilks, S.S. (1962)
"Mathematical Statistics", John Wiley, New York.

Wong, K.Y. (1966)
"Estimation of parameters of linear systems using instrumental variable method". Ph.D. Thesis, University of California, Berkeley.

Wong, K.Y. and E. Polak (1967)
"Identification of linear discrete time systems using the instrumental variable method". IEEE Trans. Aut. Control, vol. AC-12, pp. 707-718.

Woodside, C.M. (1971)
"Estimation of the order of linear systems", Automatica, vol. 7, pp. 727-733.

Wouters, W.R. (1972)
"On-line identification in an unknown stochastic environment". IEEE Trans. on Syst., Man and Cybern., vol. SMC-2, pp. 666-668.

Youla, D.C. (1961)
"On the factorization of rational matrices". IRE Trans. Inform. Theory, vol. IT-7, pp. 172-189.

Young, P.C. (1965)
"Process parameter estimation and self adaptive control", Proc. 2nd IFAC Symp. on The Theory of Self-Adaptive Control Systems (Ed. P. Hammond), Plenum Press, New York, pp. 118-149.

Young, P.C. (1968)
"The use of linear regression and related procedures for the identification of dynamic processes". Proc. of the 7th IEEE Symp. on Adaptive Processes, Univ. of California, Los Angeles.

Young, P.C. and R. Hastings James (1970)
"Identification and control of discrete dynamic systems subject to disturbances with rational spectral density". Proc. 9th IEEE Symp. on Adaptive Processes: Decision and Control, IEEE, New York.

Young, P.C. (1974)
"Recursive approaches to time series analysis". Bull Inst. Math. Appl., vol. 10, pp. 209-224.

Young, P.C. (1976)
"Some observations on instrumental variable methods of time series analysis". Int. J. of Control, vol. 23, pp. 593-612.

Young, P., A. Jakeman and R. McMurtrie (1980)
"An instrumental variable method for model order identification", Automatica, vol. 16, pp. 281-294.

Zee, G.A. van (1981)
"System identification for multivariable control". Dr. Thesis, Delft University of Technology, Delft, Netherlands.

Zurmühl, R. (1950)
"Matrizen", Springer Verlag, Berlin.

TEN SLOTTE

Het onderzoek, waarvan in dit proefschrift verslag gedaan wordt, is uitgevoerd in de vakgroep Meten en Regelen van de afdeling der Elektrotechniek van de Technische Hogeschool te Eindhoven.
Het onderzoek is een onderdeel van het onderzoekaccent "systeemidentificatie" van de vakgroep. Tot dit onderzoekaccent behoren ook o.m. het SATER project en het z.g. MIMO project, dat als uitvloeisel van- en aanvulling op het onderzoekproject SATER enige jaren geleden gestart is. Beide projecten maken momenteel deel uit van de werkgemeenschappen Theorie en Meten van de Stichting voor Meet- en Besturingstechnologie.

Aan deelonderwerpen hebben in het verleden vele afstudeerders en stageairs hun bijdrage geleverd. Alhoewel ik anderen daarmee ten onrechte te kort doe wil ik toch een paar namen noemen: Jan Talmon (EMM) en Ad van de Enden, Jan Hoffman, Anton Koenraads, ordetesters van formaat. Ook aan nauw verwante onderwerpen, die niet in dit proefschrift vermeld zijn, zijn door afstudeerders belangrijke bijdragen geleverd. Te noemen zijn de relatie tussen discrete modellen en continue processen - belangrijk voor toepassingen - en verder de toepassingen zelf: o.m. Wim Costongs met het dynamisch gedrag van een stationcar en Jan van Miltenburg en John Rooijakkers met hun schattingen aan het haemodynamisch gedrag van de aorta; dit laatste in plezierige samenwerking met Anton van Steenhoven, de deskundige op dit gebied bij de afdeling derWerktuigbouwkunde.

Ik denk met veel plezier terug aan de samenwerking met hen en alle anderen die niet vernoemd zijn.

Vermeld dienen te worden de activiteiten rond het ontwerp en de realisatie van het interaktieve programmapakket SATER, vanaf het eerste Algol SATER'tje op de PDP-8 naar het hudige pakket, een welhaast commercieel produkt. De werker van het eerste uur Pim Lemmens, die de SATER structuur ontwierp, Johan Vissers, die de omschakeling naar RSX realiseerde en Pim Bollen, die momenteel de laatste hand aan het geheel legt, ben ik veel dank verschuldigd voor hun inzet en samenwerking. De bijdragen van de vakgroep Systeem en Regeltechniek van de afdeling der Technische Natuurkunde, THE, en de vakgroep

Technopsychologie van de subfaculteit der Psychologie, Katholieke Hogeschool Tilburg, toen de ontwikkeling noodgedwongen wat trager ging, waren zeer welkom.

Voor de samenwerking met collega's - of liever vrienden - Ad Damen an Andrzej Hajdasinski, wat betreft discussies over onderwerpen van allerlei aard betreffende parameterschatting maar ook wat betreft het opzetten en het starten van het MIMO project prijs ik mij gelukkig. In dit vak kun je en mag je niet solitair opereren.

De vakgroep dank ik voor de mogelijkheid om de resultaten uiteindelijk op papier te zetten; Barbara Cornelissen voor haar doorzettingsvermogen en haar grote deskundigheid bij het uittypen en de verzorging van het verslag en Joop van Dinther voor het maken van de vele tekeningen op vakkundige wijze.

Tenslotte wil ik niet onvermeld laten, dat ik tijdens het schrijven van dit TEN SLOTTE grote moeite gehad heb om me neer te leggen bij de regel die bepaalt dat geen woorden van dank gericht dienen te worden aan promotoren.

LEVENSBERICHT

21-9-1943	geboren te Geldrop
	lagere school te Geldrop
1955-1961	middelbare school
17-6-1961	diploma gymnasium β aan het Gymnasium Augustinianum te Eindhoven
1961-1967	studie elektrotechniek aan de Technische Hogeschool te Eindhoven
29-6-1967	diploma elektrotechnisch ingenieur THE
per 15-8-1967	in dienst van de Technische Hogeschool te Eindhoven bij de vakgroep Meten en Regelen van de afdeling der Elektrotechniek.

STELLINGEN:

1. Omdat het merendeel van de benodigde rekenoperaties in de recursieve schattingsalgorithmen, zoals in dit proefschrift beschreven, in parallel uitvoerbare operaties opgesplitst kunnen worden, zijn deze algorithmen uitermate geschikt om door rekenmachines met parallel opererende processoren verwerkt te worden. Hierdoor neemt de on-line toepasbaarheid aanmerkelijk toe.

2. Indien enige in de praktijk voorhanden a-priori informatie omtrent het te schatten proces niet in het schattingsmodel en/of het schattingscriterium verwerkt is, is modelvalidatie m.b.t. deze informatie extra geboden. In dat geval zal automatisering van de schattingen doorgaans niet mogelijk zijn.

3. Bij het toepassen van iteratieve schattingsschema's is het, in tegenstelling tot wat steeds in de literatuur gesuggereerd wordt, niet nodig om, wanneer grote meetreeksen voorhanden zijn, voor de eerste iteraties alle meetwaarden te gebruiken. Dit kan tot een aanzienlijke reduktie in rekentijd leiden.
   Literatuur: bijv. Furht, B.P. (1973)
   "Maximum likelihood identification of Åström model by quasilinearization", Proc. Third IFAC Symp. on Identification and System Parameter Estimation, the Hague/Delft

4. Het gebruik van interaktieve programmatuur is onontbeerlijk bij het onderwijs in de signaalanalyse, systeemtheorie en parameterschatting.
   Literatuur: Lemmens, W.J.M. and A.J.W. van den Boom (1979)
   "Interactive programs for education and research; a survey", Automatica, vol 15, pp. 113-121.

5. Door de invoering van de tweede fase opleiding aan de Technische Hogescholen zal een nieuw type ingenieur, nl. met een hoge initiele specialisatiegraad en een te vrezen geringere breedte, zijn intrede in de beroepspraktijk doen.

6. Een evaluatie van de werkelijke effecten (zowel bedoelde als onbedoelde) van opeenvolgende door de overheid opgelegde bezuinigingsronden is in een volwassen demokratie geboden als verantwoording achteraf aan de bevolking.

7. De bedreigde cactussoorten dienen door een gezamenlijke inspanning van succulentenverenigingen voor ondergang behoed te worden.

Literatuur: Lyons, G. (1980)
"At long last, protection for endangered cacti", U.S. Cactus and Succulent Journal, vol. 52, p. 229.

8. De verkeersveiligheid rondom scholen wordt in gevaar gebracht door ouders die hun kinderen per auto naar school brengen.

9. Het projekteren van de evolutionaire tijdschaal op een schaal van 0 tot 24 uur suggereert een catastrophe om 24 uur.

10. Door de sterk wisselende regelingen t.a.v. de woningisolatie-subsidie zijn in het verleden dikwijls huiseigenaren in de kou gezet.

Stellingen behorende bij het proefschrift van A.J.W. van den Boom, Eindhoven, 28 september 1982.