Some asymptotic properties of multivariable models identified by equation error techniques
Van den Hof, P.M.J.; Janssen, P.H.M.

Published: 01/01/1985

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author’s version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. Jan. 2019
Some Asymptotic Properties of Multivariable Models Identified by Equation Error Techniques

By
P.M.J. Van den Hof and P.H.M. Janssen

EUT Report 85-E-153
ISBN 90-6144-153-6
ISSN 0167-9708
November 1985
(138) Nicola, V.F.
*A SINGLE SERVER QUEUE WITH MIXED TYPES OF INTERRUPTIONS:*
Application to the modelling of checkpointing and recovery in a transactional system.

(139) Arts, J.G.A. and W.F.H. Merck
*TWO-DIMENSIONAL MHD BOUNDARY LAYERS IN ARGON-CESIUM PLASMAS.*

(140) Willems, F.M.J.
*COMPUTATION OF THE WYNER-ZIV RATE-DISTORTION FUNCTION.*

*INTERRUPTION OF A DRY-TYPE TRANSFORMER IN NO-LOAD BY A VACUUM CIRCUIT-BREAKER.*

(142) Fronczak, J.
*DATA COMMUNICATIONS IN THE MOBILE RADIO CHANNEL.*

(143) Stevens, M.P.J. en M.P.M. van Loon
*EEN MULTIFUNCTIELE I/O-DONGSTEEN.*

(144) Dijk, J. and A.P. Verlijsdonk, J.C. Arnbak
*DIGITAL TRANSMISSION EXPERIMENTS WITH THE ORBITAL TEST SATELLITE.*
EUT Report 84-E-144. 1984. ISBN 90-6144-144-7

(145) Woert, M.J.M. van
*KNHIHALISATIE VAN PROGRAMMBLOKE LOGIC ARRAYS.*

(146) Jochems, J.C. en P.M.C.M. van den Eijnden
*TOEGANG-TOEGANG IN SEQUENTIELE CIRCUITS.*

(147) Bosendaal, L.T. en M.P.J. Stevens, P.M.C.M. van den Eijnden
*DE REALISATIE VAN EEN MULTIFUNCTIELE I/O-CONTROLLER MET BEHULP VAN EEN GATE-ARRAY.*

(148) Eijnden, P.M.C.M.
*A COURSE ON FIELD PROGRAMMABLE LOGIC.*

(149) Beezem, P.A.
*MILLIMETER-WAVE ANTENNA MEASUREMENTS WITH THE HP8510 NETWORK ANALYZER.*
SOME ASYMPTOTIC PROPERTIES OF MULTIVARIABLE MODELS IDENTIFIED BY EQUATION ERROR TECHNIQUES

by

P.M.J. Van den Hof

and

P.H.M. Janssen

EUT Report 85-E-153
ISBN 90-6144-153-6
ISSN 0167-9708
Coden: TEUEDE

Eindhoven
November 1985
ISBN 90-6144-153-6
SISO 656 UDC 519.71,001.3 UGI 650
Trefw.: systeemidentificatie.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>2</td>
</tr>
<tr>
<td>2. The system and the model set</td>
<td>3</td>
</tr>
<tr>
<td>3. Equation error methods; some asymptotic results</td>
<td>5</td>
</tr>
<tr>
<td>4. Main result</td>
<td>9</td>
</tr>
<tr>
<td>5. Discussion</td>
<td>16</td>
</tr>
<tr>
<td>6. Conclusions</td>
<td>18</td>
</tr>
<tr>
<td>Appendix</td>
<td>20</td>
</tr>
<tr>
<td>References</td>
<td>23</td>
</tr>
</tbody>
</table>
SOME ASYMPTOTIC PROPERTIES OF MULTIVARIABLE MODELS
IDENTIFIED BY EQUATION ERROR TECHNIQUES

P.M.J. Van den Hof
P.H.M. Janssen

Dept. Electrical Engineering
Eindhoven University of Technology
The Netherlands

ABSTRACT

In this paper some interesting properties are derived for simple equation error identification techniques - least squares and basic instrumental variable methods-, applied to a class of linear time-invariant time-discrete multivariable models. The system at hand is not supposed to be contained in the chosen model set. An analysis of the approximating model is performed in the time-domain, relating the Markov parameters of the original system to the Markov parameters of the identified model. The results are asymptotic in the sense that the number of data samples is supposed to be infinite; the input signals are supposed to be stationary zero mean white noise sequences with unit variance.

The asymptotic results are derived for a general class of linear multivariable models in I/O form (matrix fraction descriptions), incorporating most models currently used in system identification. In terms of approximation of systems, the results can also be applied in model reduction.

Mailing address of the authors:
Eindhoven University of Technology
Department of Electrical Engineering
P.O. Box 513
5600 MB Eindhoven
The Netherlands
Tel. (40)-473280
1. INTRODUCTION

In system identification literature there is a growing interest in identification methods that give reliable results in situations where the process at hand is not necessarily contained in the chosen model set. This aspect is considered to be a valuable robustness property [1]. Its importance is indicated by realizing that in many practical situations of system identification, a model will be required that is of restricted complexity, approximating the essential characteristics of the - possibly very complex - process, rather than a very sophisticated model that exactly models the process behaviour.

If the problem of system identification, or rather approximate modelling, is considered in this context, it comes very close to the problem of model reduction. An important item now becomes: which criterion has to be used to approximate the original process, resp. the higher order model, and which model set has to be chosen.

It has been recognized that these choices highly determine the performance of the model, when used for specific purposes such as simulation, prediction, (minimum variance) control etc. [2], [3]. In many situations the performance of an identified model is judged upon its ability to simulate the process under consideration. However, output error methods, being most appropriate if the simulation behaviour of the model is concerned, are much more complex than equation error methods. It is therefore important to analyse the simulation behaviour of an equation error model. A frequency domain analysis of this aspect of approximate models, identified by prediction error methods, is given in [4],[5]. By considering the Markov parameters of the identified model, we will focus on properties of the approximate model in the time domain.

In this paper an analysis is made of simple least squares and instrumental variable methods applied to equation error type models. Earlier work on this subject has been published in 1976 by Mullis and Roberts [6], who established a connection between results in model reduction and asymptotic results in least squares system identification. An extension to
the multivariable case has been worked out by Inouye [7], but restricted to a very special parametrization (full polynomial ARMA form). The previous work will be extended to a general class of multivariable models, containing both canonical and pseudo-canonical ones, and to simple instrumental variable methods. The analysis of the relation between the original process and the identified model will be carried out in the time domain, by means of the Markov parameters.

In section 2 the class of models is defined and notations are introduced. Asymptotic equation error results are presented in section 3. In section 4 the main theorem is introduced and applied to a number of different parametrizations. A discussion on the results follows in section 5.

2. THE SYSTEM AND THE MODEL SET

We consider a discrete-time multivariable system having a p-dimensional input signal \( u(t) \) and a q-dimensional output signal \( y(t) \) at time instant \( t \in \mathbb{Z} \). At the outset we will keep the discussion quite general and therefore only require that the input- and output signals are jointly wide-sense stationary and ergodic. Later on we will specify our results for white input signals with unit variance matrix, and for linear systems.

Consider the problem of modelling the system approximately by using a parametrized set of linear time-invariant discrete-time multivariable I/O models, given by the following general MFD (Matrix Fraction Description)-form:

\[
P(z; \theta)y(t) = Q(z; \theta)u(t) + \varepsilon(t; \theta) \tag{1}
\]

where
- \( z \) denotes the forward shift operator \( zy(t) := y(t+1) \);
- \( P(z; \theta) := [P_{ij}(z)]_{q \times q} \) and \( Q(z; \theta) := [q_{ij}(z)]_{q \times p} \)
- are \( (q \times q) \), resp. \( (q \times p) \)-polynomial matrices where for ease of notation the explicit dependency of the polynomial entries on the parameter \( \theta \) has been omitted.
- \( P(z; \theta) = P_d(z) + P^*(z; \theta) \tag{2} \)

where
We will now consider the situation where the vector \( \theta \) of unknown parameters consists of all the coefficients \( \alpha \) and \( \beta \) occurring in the polynomial matrices \( P(z;\theta) \) and \( Q(z;\theta) \). The residual

\[
c(t;\theta) = P(z;\theta)y(t) - Q(z;\theta)u(t)
\]

is dependent on the parameter vector \( \theta \), but is not parametrized itself. It is called an equation error and can be computed from the available input and output samples; the residual is linear-in-the-parameter vector \( \theta \).

Methods for estimating \( \theta \) in this context are often called equation error methods \([1]\) and will be considered in the next section.
Remark 2.1
The model set (1) is very general and encompasses most uniquely identi­fiable MFD-forms, currently used in the identification of multivariable systems. For most forms it will follow that \( r_{ij} = 1, s_j = 1 \). Note that the model set is not necessarily restricted to causal models.
In section 4 it will be illustrated how all specific forms fit in the general context.

3. EQUATION ERROR METHODS; SOME ASYMPTOTIC RESULTS

Equation error methods are very popular in system identification. The main reason for this is the simplicity of the corresponding identifi­cation algorithm, due to the linearity-in-the-parameters of the model.
In this section some asymptotic results will be presented for least squares and basic instrumental variable estimators.

A common equation error method for obtaining an estimate of \( \theta \) in (1) is the simple least-squares estimator, minimizing

\[
V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t;\theta)^T \varepsilon(t;\theta)
\]

with respect to \( \theta \). \( N \) denotes the number of data samples. Being inter­ested only in asymptotic results, the asymptotic analogon of this problem will be considered, minimizing

\[
V(\theta) = \mathbb{E} \varepsilon(t;\theta)^T \varepsilon(t;\theta)
\]

with respect to \( \theta \), under assumption of stationary and ergocidity of the input and output signals. (\( \mathbb{E} \) denotes the expectation-operator).

Remark 3.1
In our theoretical analysis we do not impose the condition that the para­meter vector \( \hat{\theta} \) minimizing \( V(\theta) \) is unique. With respect to the identifi­cation algorithm, however, one would like to use model sets which guaran­tee uniqueness. Examples of these will be given in section 4. For the analysis, there is no objection to the non-uniqueness of \( \hat{\theta} \).

With respect to the determination of \( \hat{\theta} \), minimizing \( V(\theta) \), we can now state the following:
Proposition 3.1

The asymptotic least squares estimator \( \tilde{\theta} \) satisfies

\[
E \varepsilon_i(t; \tilde{\theta}) z_i^{l-1} y_j(t) = 0 \quad r_{ij} < l < v_{ij} \quad 1 < i, j < q
\]  
(12a)

and

\[
E \varepsilon_i(t; \tilde{\theta}) z_i^{l-1} u_j(t) = 0 \quad s_{ij} < l < \mu_{ij} \quad 1 < i < q \quad 1 < j < p
\]  
(12b)

Proof

\( V(\theta) \) is quadratic in \( \theta \). A standard necessary and sufficient condition for \( \tilde{\theta} \) to be a minimum of \( V(\theta) \) is given by:

\[
\frac{\partial}{\partial \theta_s} [V(\theta)] \bigg|_{\theta = \tilde{\theta}} = 0 \text{ for each component } \theta_s \text{ of } \theta.
\]  
(13)

Equivalently,

\[
E[\varepsilon(t; \theta)^T] \frac{\partial}{\partial \theta_s} \varepsilon(t; \theta) \bigg|_{\theta = \tilde{\theta}} = 0 \text{ for each component } \theta_s.
\]  
(14)

Since \( \varepsilon(t; \theta) \) is a linear function in all the components of \( \theta \):

\[
\alpha_{ij} \quad (1 < i < q, \quad r_{ij} < l < v_{ij})
\]

and

\[
\beta_{ij} \quad (1 < i < q, \quad 1 < j < p, \quad s_{ij} < l < \mu_{ij})
\]

the result follows immediately.

Now we define

\[
\tilde{P}(z) = P(z; \tilde{\theta})
\]  
(15a)

and

\[
\tilde{Q}(z) = Q(z; \tilde{\theta})
\]  
(15b)

The output signal \( y(t) \) of the estimated model, when excited by the original input signal, is given by:

\[
\tilde{P}(z) y(t) = \tilde{Q}(z) u(t) \quad -\infty < t < \infty
\]  
(16)

Defining the correlation functions

\[
\Phi_{yu}(k) := E y(t) u^T(t-k) \quad -\infty < k < \infty
\]  
(17a)

\[
\Phi^*_{yu}(k) := E y(t) u^T(t-k) \quad -\infty < k < \infty
\]  
(17b)

\[
\Phi_{uu}(k) := E u(t) u^T(t-k) \quad -\infty < k < \infty
\]  
(17c)

the following general result can be established:
Proposition 3.2

Let \( \hat{P}(z) \) and \( \hat{Q}(z) \) be as defined in (15a) and (15b), and let \( \hat{\theta} \) be the asymptotic least squares estimator. Then the correlation functions \( \Phi_{yu}(\cdot) \), \( \Phi_{yu}^{-1}(\cdot) \) and \( \Phi_{uu}(\cdot) \) are related according to

\[
\hat{P}_i(z)[\Phi_{yu}]_{j}(-k) = \hat{Q}_i(z)[\Phi_{uu}]_{j}(-k) \quad s_j - 1 < k < \mu_{ij} - 1, \quad 1 < i < q, 1 < j < p; \tag{18}
\]

\[
\hat{P}_i(z)[\Phi_{yu}]_{j}(-k) = \hat{Q}_i(z)[\Phi_{uy}]_{j}(-k) \quad r_{ij} - 1 < k < \nu_{ij} - 1 \tag{19}
\]

and

\[
\hat{P}(z) \Phi_{yu}^{-1}(-k) = \hat{Q}(z)\Phi_{uu}(-k) \quad \Rightarrow k \gg 1 \tag{20}
\]

where \( \hat{P}_i(z), [\Phi_{yu}]_{j}(-k) \) denote the \( i \)-th row of \( \hat{P}(z) \) resp. the \( j \)-th column of \( \Phi_{yu}^{-1}(-k) \) etc.

**Proof**

The \( i \)-th component of the residual can be written:

\[
\varepsilon_i(t;\hat{\theta}) = \hat{P}_i(z)y(t) - \hat{Q}_i(z)u(t) \tag{21}
\]

Substituting this into (12a) and (12b) of proposition 3.1 simply leads to (19) and (18).

(20) follows directly from (16).

Our attention will be focussed on results (18) and (20) of proposition 3.2.

It shows that on a certain interval the expressions for \( \Phi_{yu}^{-1}(-k) \) and \( \Phi_{yu}^{-1}(-k) \) are similar. The consequence of this property will be studied in the next section, especially in the case where the input signal is unit variance white noise. First we want to establish that some parts of propositions 3.1 and 3.2 also hold for certain instrumental variable (IV) methods.
Instrumental variable methods form a useful technique for estimating the parameters in equation error models. Many variants of this technique have been proposed (for an overview see [8]). In this discussion only the most elementary one, the basic IV-method, will be considered.

Asymptotically, the basic IV-method for estimating $\theta$ amounts to the solution of the set of equations

$$E Z(t)\varepsilon(t;\theta) = 0$$  \hspace{1cm} (22)

where $Z(t)$ is a $n_0 \times q$ matrix consisting of properly chosen instrumental variables [8], and $n_0$ is the number of parameters in the model. In our analysis it will be assumed that (22) indeed has a solution. Typical choices for $Z(t)$ are [8]:

$$Z(t) = \text{diag}[z_1(t), \ldots, z_q(t)]$$  \hspace{1cm} (23)

where the length of vector $z_i(t)$ corresponds to the number of parameters in the $i$-th equation of (1),

and $z_i(t)$ contains $z^{k-1}_j(t)$, $r_{ij} < k < v_{ij}$; $1 < j < q$

$$z^{k-1}_j(t)$$

and $z^{k-1}_j(t)$, $s_j < k < \mu_{ij}$; $1 < j < p$

with $\phi_j(t)$ a filtered or delayed input signal.

Because of (23), equation (22) now becomes:

$$E z_i(t) \varepsilon_i(t;\theta) = 0 \hspace{1cm} 1 < i < q$$  \hspace{1cm} (25)

Substituting the second part of (24) gives:

$$E z^{k-1}_j(t) \varepsilon_i(t;\theta) = 0 \hspace{1cm} 1 < j < p, \hspace{1cm} 1 < i < q, \hspace{1cm} s_j < k < \mu_{ij}$$  \hspace{1cm} (26)

which is exactly the same expression as (12b) in Proposition 3.1.

From this it immediately follows that, for this specific IV estimator, Proposition 3.2 concerning the correlations $\phi_{yu}(\cdot)$ and $\phi_{yu}^\prime(\cdot)$ is also valid.
4. MAIN RESULT

In Proposition 3.2 we have obtained a general relationship involving \( \Phi_{yu}(.) \), \( \Phi_{yu}^-(.) \), \( \Phi_{uu}(.) \) and the estimated model \( P(z) \) and \( Q(z) \). In this section we will further elaborate on this relation for the special case where the input signal \( u(.) \) is zero mean stationary white noise with unit variance (i.e. \( E u(k)u'(l) = \delta(k-l).I \)).

For ease of notation we will write:

\[
M(k) := \Phi_{yu}(k) \quad -\infty < k < \infty \tag{27}
\]
\[
\hat{M}(k) := \Phi_{yu}^-(k) \quad -\infty < k < \infty \tag{28}
\]

Because of the restrictions on the input signal, \( \hat{M}(k) \) is the k-th Markov parameter associated with the model (16).

Under assumption of linearity and time-invariance of the original process, \( \hat{M}(k) \) is the k-th Markov parameter of the process. If we consider a linear and time-invariant process with a disturbance on the output signal that is uncorrelated to the input signal, \( \hat{M}(k) \) is the k-th Markov parameter of the process.

Since \( \Phi_{uu}(k) = \delta(k)I \) \tag{29}

we can reformulate the results (18) and (20) of Proposition 3.2 to

\[
\hat{P}_i(z)M_j(t) = \hat{Q}_{ij}(z)\delta(t) \quad 1-\mu_{ij} < t < 1-\sigma_j \tag{30}
\]

\[
1 < i < q, \ 1 < j < p
\]

and

\[
\hat{P}(z)\hat{M}(t) = \hat{Q}(z)\delta(t)I \quad -\infty < t < \infty \tag{31}
\]

Note that the Markov parameters of the process satisfy the same relationship as the Markov parameters of the identified model, however on a restricted interval. In this section it will be shown that, as a result of this, under some conditions, the two sequences of Markov parameters are equal on a restricted interval.

First we will present a general property of the model set (1)-(8).
Proposition 4.1

Consider a multivariable I/O model as defined in (1)-(8). Let
\[ \hat{\mu}_j = \max \mu_{ij} \]
Then for the Markov parameters of this model holds:
\[ \hat{M}_{ij}(t) = 0 \quad \text{for } t < \hat{\nu}_i - \hat{\mu}_j \]
and
\[ \hat{M}_{ij}(t) = \hat{\mu}_j \quad \text{for } t < \hat{\nu}_i \hat{\mu}_j + 1 \]
This result is inherent to the chosen model set (1)-(8).

Proof: The proof of this proposition is added in the appendix.

We can now present the main result in the following theorem.

Theorem 4.1

Consider a multivariable I/O model as defined in (1)-(8). If this model is used for identifying a linear time invariant system by equation error techniques, as discussed in section three, if the input signal is zero mean, stationary unit variance white noise and if the number of data samples tends to infinity, then the Markov parameters \( \hat{M}(t) \) of the identified model satisfy:
\[ \hat{M}_{ij}(t) = M_{ij}(t) \quad \text{for } \hat{\nu}_j + \hat{r}_i < t < 1 - \hat{\nu}_j + \hat{\nu}_i \]
under the condition that the Markov parameters of the original process satisfy,
\[ M_{ij}(t) = 0 \quad \text{for } \hat{\nu}_j + \hat{r}_i < t < \gamma_{ij} \]
where
\[ \hat{r}_i := \min \hat{r}_{ij} \quad 1 < j < q \]
\[ \hat{\mu}_j := \max \mu_{ij} \quad 1 < j < p \]
\[ \gamma_{ij} := \max[\nu_i - \mu_{ij}, \max(\nu_{ij} - 1 - \mu_{ij})] \]

Proof: The proof of this theorem is given in the appendix.
As a combined result of (33) and Theorem 4.1 it follows that

\[ N_{ij}(t) = 0 \quad \text{for} \ t < \gamma_{ij} \quad (39) \]

Note that by (38), \( \gamma_{ij} \) can be interpreted as the maximal difference between the degrees of \( p_{\lambda i}(z) \) and \( q_{\lambda j}(z) \) minus 1:

\[
\gamma_{ij} = \max_{l} \left[ \deg(p_{\lambda i}(z) - q_{\lambda j}(z)) \right] - 1 \quad (40)
\]

Theorem 4.1 has been stated in a very general setting, due to the generality of the chosen model set (1). Its full implications will become clear if it is applied to specific parametrizations, with specific restrictions on the structure indices. If condition (35) is fulfilled, the result shows that apparently a Padé type of approximation is involved, where the length of the fit on Markov parameters is determined by the chosen parametrization and the chosen structure indices of the model.

We will now specify the result for a more practical situation and illustrate it finally for a number of identifiable parametrizations that are commonly used in the identification of causal systems.

When modelling causal systems by MFD model sets the degrees of the polynomials \( q_{ij}(z) \) are usually chosen equal to the \( i \)-th row degree of \( p(z) \):

\[
\mu_{ij} = \max_{l} \left[ \deg p_{\lambda l}(z) \right] \quad 1 < j < p \quad 1 < i < q \quad (41)
\]

This choice is motivated by the fact that for causal MFD's the row degrees of \( P(z) \) have to be greater than or equal to the corresponding row degrees of \( Q(z) \) (see [9], lemma 6.3.10). As a result, the corresponding model sets will fulfil \( \gamma_{ij} < -1 \). In the presentation of different parametrizations later on in this section, this condition is always shown to be satisfied.

We can now present the following result as a direct consequence of Theorem 4.1.
Proposition 4.2

Consider the situation as described in Theorem 4.1. If the original system is causal and if $y_{ij} < -1$ (the latter condition is shown to be fulfilled for all parametrizations presented later on in this section), then

$$\hat{M}_{ij}(t) = M_{ij}(t) \quad \text{for} \quad t < 1 - s_j + v_i \quad 1 \leq i \leq q \quad 1 \leq j \leq p \quad (42)$$

Proof: The result follows directly from Theorem 4.1 and eq. (39).

Remark 4.1

As a direct result of Proposition 4.2, it has to be noted that the estimated model will be causal, notwithstanding the fact that the applied model set could possibly admit non-causal models. This result is quite remarkable.

Remark 4.2

For most common MFD-models, $s_j = 1$ and the result of Proposition 4.2 can be formulated as

$$\hat{M}_{ij}(t) = M_{ij}(t), \quad t < v_i \quad 1 \leq i \leq q, \quad 1 \leq j \leq p \quad (43)$$

In table 1 it is illustrated how the results of this section work out for a number of - commonly used - parametrizations of the original model set (1). The parametrizations will not be described in detail; for a thorough description the reader is referred to the literature. A general and up-to-date account on the use of identifiable parametrizations for multivariable linear systems is given in [10].

Table 1 shows the structure specification, condition and results of Theorem 4.1, for a number of MFD forms: the canonical observability form [11]; the pseudo-canonical (overlapping) form [12],[13],[14]; the Hermite form [15],[16]; the diagonal form [17],[18] and the generalized pseudo-canonical form [19].
<table>
<thead>
<tr>
<th>MFD Structure specification</th>
<th>CANONICAL OBSERVABILITY FORM</th>
<th>PSEUDO-CANONICAL FORM</th>
<th>HERMITE FORM</th>
<th>DIAGONAL FORM</th>
<th>GENERALIZED PSEUDO-CANONICAL FORM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_{ij} = \min(v_i + 1, v_j), i &gt; j )</td>
<td>( v_{ij} = v_j )</td>
<td>( v_{ij} = v_i ) if ( v_i = v_j )</td>
<td>( v_{ij} = \max(v_i + 1, v_j) )</td>
<td>( v_{ij} = 0 ) if ( i &gt; j )</td>
<td>( v_{ij} = m_j + n_j - 1 )</td>
</tr>
<tr>
<td>( \mu_{ij} = v_i + 1 )</td>
<td>( r_{ij} = 1; s_j = 1 )</td>
<td>( r_{ij} = 1; s_j = 1 )</td>
<td>( r_{ij} = 1; s_j = 1 )</td>
<td>( r_{ij} = 1; s_j = 1 )</td>
<td></td>
</tr>
<tr>
<td>Auxiliary variables</td>
<td>( x_j = 1; \tilde{v}_j = v_m + 1 )</td>
<td>( x_j = 1; \tilde{v}_j = v_m + 1 )</td>
<td>( x_j = 1; \tilde{v}_j = v_m + 1 )</td>
<td>( x_j = 1; \tilde{v}_j = v_m + 1 )</td>
<td></td>
</tr>
<tr>
<td>( \gamma_{ij} = -1 )</td>
<td>( \gamma_{ij} = v_i - \mu_{ij} )</td>
<td>( \gamma_{ij} = \text{see (38)} )</td>
<td>( \gamma_{ij} = -1 )</td>
<td>( \gamma_{ij} = -1 )</td>
<td></td>
</tr>
<tr>
<td>Condition (35): ( M_{ij}(t) = 0 ) for</td>
<td>( -v_m &lt; t &lt; -1 )</td>
<td>( -v_m &lt; t &lt; v_i - \mu_{ij} )</td>
<td>( -v_m &lt; t &lt; \gamma_{ij} )</td>
<td>( -v_m &lt; t &lt; -1 )</td>
<td>( -v_m &lt; t &lt; -1 )</td>
</tr>
<tr>
<td>Theorem 4.1: ( \tilde{M}<em>{ij}(t) = M</em>{ij}(t) )</td>
<td>( -v_m &lt; t &lt; v_i )</td>
<td>( -v_m &lt; t &lt; v_i )</td>
<td>( -v_m &lt; t &lt; v_i )</td>
<td>( -v_m &lt; t &lt; v_i )</td>
<td>( -v_m &lt; t &lt; v_i )</td>
</tr>
<tr>
<td>Result (42): ( M_{ij}(t) = 0 ) for</td>
<td>( t &lt; -1 )</td>
<td>( t &lt; v_i - \mu_{ij} )</td>
<td>( t &lt; \gamma_{ij} )</td>
<td>( t &lt; -1 )</td>
<td>( t &lt; -1 )</td>
</tr>
</tbody>
</table>

\( v := \max(v_i) \)

Table 1- Structure specifications for a number of multivariable parametrizations in MFD-form; condition and results for the asymptotic fit of Markov parameters.
<table>
<thead>
<tr>
<th>MFD Structure specification</th>
<th>PRESCRIBED MAXIMAL LAG FORM</th>
<th>FULL POLYNOMIAL FORM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( S(n_1, \ldots, n_q, \ldots, n_p) )</td>
<td>( S(n,m) )</td>
</tr>
<tr>
<td>( v_i = \max(n_k, m_k) )</td>
<td>( v_i = \max(n, m) )</td>
<td></td>
</tr>
<tr>
<td>( v_{ij} = v_j = s )</td>
<td>( v_{ij} = v_j = s )</td>
<td></td>
</tr>
<tr>
<td>( \mu_{ij} = s + 1 )</td>
<td>( \mu_{ij} = s + 1 )</td>
<td></td>
</tr>
<tr>
<td>( r_{ij} = s-n_j +1; s_j = s-m_j +1 )</td>
<td>( r_{ij} = s-n+1; s_j = s-m+1 )</td>
<td></td>
</tr>
</tbody>
</table>

| Auxiliary variables         | \( \tilde{r}_j = s-n_j +1; \tilde{\mu}_j = s +1 \) | \( \tilde{r}_j = s-n_j +1; \tilde{\mu}_j = s +1 \) |
|                            | \( \gamma_{ij} = -1 \)                             | \( \gamma_{ij} = -1 \) |

| Condition (35):            | \(- n_i < t < -1 \)                                | \(- n < t < -1 \) |
| \( M_{ij}(t)=0 \) for             |                                                  |                     |
| Theorem 4.1: \( \tilde{M}_{ij}(t)=M_{ij}(t) \) | \(- n_i < t < m_j \)                             | \(- n < t < m \) |

| Result (42): \( \tilde{M}_{ij}(t)=0 \) for | \( t < -1 \)                                    | \( t < -1 \) |

Table -2- MFD structure specifications for two multivariable parameterizations in ARMAX-form; condition and results for the asymptotic fit of Markov parameters.
Although the analysis and discussion has, until now, been based on MFD-model sets, defined in the forward shift-operator \( z \), the obtained results can also be used for ARMAX model sets, which are defined in the backward shift operator \( z^{-1} \). By transforming the ARMAX-model to a corresponding MFD-form, the results of this section can still be applied.

In Table 2, the results are listed for two commonly used ARMAX-model forms: the "prescribed maximum lag"-form \( S(n_1, \ldots, n_q, m_1, \ldots, m_p) \), \cite{20,21,22} and the full polynomial form \( S(n,m) \), \cite{23,24}.

**Remark 4.3**

With respect to the determination of \( \gamma_{ij} \), \cite{38,40}, it can be shown that for all parametrizations, except the Hermite form, presented in Tables 1 and 2, \( \gamma_{ij} \) is given by \( \gamma_{ij} = v_{i-j} \) \( \gamma_{ij} \) (44)

This means that in these situations the maximum in eq. (40) always occurs at \( \lambda = i \).

**Remark 4.4**

The asymptotic fit of Markov parameters for the full polynomial ARMAX form has been presented before by Mullis and Roberts \cite{4} for the single input single output (SISO) case, and by Inouye \cite{5}, for the multivariable (MIMO) case, both applying the least squares equation error technique.

**Remark 4.5**

The asymptotic matching of Markov parameters, as given in Theorem 4.1, finds its source in the correlation results (12b) of Proposition 3.1.

Bearing this in mind, it is easy to see that the number of Markov parameter entries that is forced to be matched by the equation error method, is equal to the number of \( \beta \)-parameters in the model.

Dependent on the causality of the model set, the sequence of matched Markov parameter entries lies in the causal (\( t > 0 \)), or partly in the non-causal range.
5. DISCUSSION

The analysis in this paper has been motivated by the consideration that in many practical situations the performance of an identified model is assessed by its ability to simulate the given process. A proper way to analyse the simulation behaviour of a model is to analyse its sequence of Markov parameters.

Considering Remark 4.5, we see that in the asymptotic equation error results, all $\beta$-parameters are in fact determined in such a way that the Markov parameters of the process are matched over a certain range. The degrees of freedom available in the $\alpha$-parameters determine how the start-sequence of Markov parameters of the model is extended to infinity. As a result, the choice for a parametrization can have a severe effect on the emphasis that is adjusted to the start-sequence c.q. the extension sequence of Markov parameters. Given a parametrization, the specific set of structure indices determines the length of the matching interval for the various input-output transmittances within the model.

An utmost consequence of the matching of Markov parameters can be recognized when we consider the instrumental variable estimator suggested by Wouters [25], for a causal SISO system. As a model set is chosen:

$$y(t) = \sum_{i=1}^{n} \alpha_i y(t-i) + \sum_{j=0}^{m} \beta_j u(t-j) + \varepsilon(t; \theta)$$  (45)

The instrumental variable is chosen to be a delayed input signal with delay $m+1$. Consequently, the asymptotic estimator satisfies

$$E \varepsilon(t; \theta) z^{-1} u(t) = 0 \quad \text{for } - (m+n)+1 < \lambda < 1$$  (46)

in correspondence with proposition 3.1.

Applying the analysis of section 4 shows that

$$\hat{M}(t) = M(t) \quad \text{for } 0 < t < m+n$$  (47)

In this case all parameters of the model are used to match the Markov parameters. Therefore the model is completely determined by the first $m+n+1$ Markov parameters of the system. This result is, in fact, equivalent to a Padé approximation.
In [4], [5] the performance of prediction error identification methods is studied, based on a frequency domain analysis. In this analysis it is formalized that equation error models in general emphasize the high frequency aspects of the model, at the expense of a worse low frequency behaviour. This statement is intuitively confirmed by the exact matching of the start sequence of Markov parameters.

It has been illustrated in [3] for some SISO examples that the asymptotic unbiasedness of the start-sequence of Markov parameters causes the equation error model to generate a bad sequence of Markov parameters in the extension to infinity. It is shown [3], [26] that this effect especially occurs when the system impulse response is small in the start-sequence (e.g. because of time delays), and increases outside this interval. A similar result will probably hold for MIMO systems but has not yet been analysed.

Being aware of the asymptotic unbiasedness of the start sequence of Markov parameters, equation error methods can also be used when an estimation of this start sequence is required, e.g. for application of approximate realization methods [27].

Equation error methods have also found their use in model reduction problems [6],[7],[28],[29]. The analysis in this paper can certainly also be applied to these kinds of problems. The asymptotic results for the model parameters of the reduced order model can be constructed based on a finite number of elements of the sequence of Markov parameters and the autocorrelation function of the outputs.

The results for theorem 4.1 have been specified for a very general class of models, showing the property of linearity-in-the-parameters, in a context of one-step-ahead-prediction models. Generalizations to k-step-ahead prediction models are straightforward as long as one keeps up with the linearity-in-the-parameters of the model. The results on the fit of the start-sequence of Markov parameters remains unchanged for equation error models based on k-step-ahead-prediction, if the same input signal components are involved in (1). However, a different result will be obtained for the extension of the Markov parameter sequence if \( k > 1 \).
We have to note that the results of section 4 cannot simply be generalized to situations of non-white input signals. Although the quantitites

$$\Phi_{yu}(t) \quad \text{and} \quad \Phi_{yu}^-(t)$$

satisfy the same relationship on a restricted interval (see proposition 3.2), they will not match in general, since the associated initial conditions for the recursive relations are different in the case of non-white input signals.

6. CONCLUSIONS

We have established a number of asymptotic results for the approximate modelling of multivariable systems by equation error methods, applying least squares and basis IV techniques. These methods, commonly used for the identification of systems, are analysed for situations where the system at hand is not necessarily contained in the chosen model set. The results have been obtained for a very general class of linear models, having the property of linearity in the parameters. This class of models covers all commonly used MFD (Matrix Fraction Description)-forms and ARMAX forms.

The main results of this paper are valid for general linear time-invariant systems corrupted by output noise that is not correlated with the input signal. Under the condition of white input noise, it has been shown that the Markov parameters of the system are estimated asymptotically unbiased over a certain interval around $t=0$. The position of the interval is dictated by the chosen structure indices of the model. The general results for this match between Markov parameters of the system and the model are specified for various parametrizations in MFD form (canonical observability form, pseudo-canonical form, Hermite form, diagonal form, generalized pseudo-canonical form), and in ARMAX form (prescribed maximal lag form, full polynomial form).

Moreover it has been shown that for causal systems the identified model is asymptotically causal, notwithstanding the fact that the applied model set might contain non-causal models.

Equation error methods can also be applied to model reduction problems; it is shown that the instrumental variable method of Wouters [25] in fact is equivalent to the Padé approximation in model reduction.
The analysis in this paper gives a deeper understanding in the essential properties of equation error methods when applied to problems of approximate modelling. The presented analysis does not claim to be complete; properties for the start-sequence of Markov parameters have been analysed, properties of the extension of this sequence have not been considered.
APPENDIX

Proof of Proposition 4.1

The Markov parameters $\hat{M}(t)$ of the model (1) satisfy:

$$ P(z)M(t) = Q(z)\delta(t)I \quad -\infty < t < \infty $$

with $I$ the pxp identity matrix.

Using (2) this set of p.q equations can be written as:

$$ z^{i-1}M_{ij}(t) = -P^{*}_i(z)\hat{M}_{ij}(t) + Q_{ij}(z)\delta(t) \quad -\infty < t < \infty \quad 1 < i < q, 1 < j < p $$

With restriction (8) ($\nu_{\lambda} \leq \nu_i^*, 1 < i, \lambda < q$) it follows directly that (A-2) is a recursive expression for calculating $\hat{M}_{ij}(t)$, in the sense that for every $i$ and $j$ the element $\hat{M}_{ij}(t)$ with the highest time index, occurs only in the left hand part of (A-2).

From (6) it follows that $Q_{ij}(z)\delta(t) = 0$ for $t < -\mu_{ij}$

Because of the recursiveness of (A-2) it follows that

$$ z^{i-1}M_{ij}(t) = 0 \quad \text{if } t \text{ satisfies } t < -\mu_{ij} $$

This condition on $t$ can be written as:

$$ t < -\max(\mu_{ij}, \tilde{\mu}_{j} - 1) \quad \text{with } \tilde{\mu}_{j} = \max_{s} \mu_{ij} $$

The result of the proposition follows directly.

Proof of Theorem 4.1

According to (30) and (2)

$$ z^{i-1}M_{ij}(t) = -P^{*}_i(z)\hat{M}_{ij}(t) + Q_{ij}(z)\delta(t) \quad 1 - \mu_{ij} < t < 1 - \mu_{j} $$

and according to (31)

$$ z^{i-1}M_{ij}(t) = -P^{*}_i(z)\hat{M}_{ij}(t) + Q_{ij}(z)\delta(t) \quad -\infty < t < \infty \quad 1 < i < q, 1 < j < p $$

Since (A-2) is shown to be recursive, the same holds for (A-6) and (A-7).
Using the result of Proposition 4.1, the Markov parameters of the model can be calculated according to
\[ v_i M_{ij}(t) = 0 \quad \text{for} \quad t < \tilde{\mu}_{ij} \]
\[ v_i M_{ij}(t) = -\hat{p}_{ij}(z) M_{ij}(t) + \hat{p}_{ij}(z) \delta(t) \quad \text{for} \quad t > \tilde{\mu}_{ij} \]  
\( (A-8) \)

Based on (A-6) a similar expression for \( M_{ij}(t) \) will be derived. For this purpose first (A-6) will be extended to the interval \( 1-\tilde{\mu}_{ij} < t < 1-s_j \).

Subsequently the initial conditions for the recursive equations will be analysed.

Extending (A-6) to the time interval \( 1-\tilde{\mu}_{ij} < t < 1-s_j \) requires that an extra set of equations has to be fulfilled; using (A-3) this set of equations can be written:
\[ v_i M_{ij}(t) = -\hat{p}_{ij}(z) M_{ij}(t) \quad \text{for} \quad 1-\tilde{\mu}_{ij} < t < \tilde{\mu}_{ij} \]  
\[ 1 < i < q, \ 1 < j < p \]  
\( (A-9) \)

Fulfillment of (A-9) puts restrictions on the process Markov parameters. A sufficient condition for (A-9) is that all Markov parameters that occur in (A-9) are supposed to be zero.

By inspection, it follows that (A-9) contains
\[ M_{ij}(t), \ 1-\tilde{\mu}_{ij} + \nu_i < t < \mu_{ij} + \nu_i \]  
\[ (A-10) \]

and \[ M_{kj}(t), \ -\tilde{\mu}_{ij} + r_{ik} < t < \mu_{ij} + \nu_{ik} \]  
for all \( 1 < i, k < q, \ 1 < j < p \).

Combining these sets, it follows that (A-9) contains for all \( i=1, \ldots, q \) the following \( M_{kj}(t) \):
\[ M_{kj}(t), \ -\tilde{\mu}_{ij} + \min_{i} r_{ik} < t < \max_{i} \left[ -\mu_{kj} + \nu_{k} , \max_{i} \left( -\mu_{ij} + \nu_{ik} \right) \right] \]  
\[ (A-11) \]

As a result, a sufficient condition for (A-9) is:
\[ M_{kj}(t) = 0 \quad \text{for} \quad -\tilde{\mu}_{ij} + r_{ik} < t < \gamma_{kj} \]  
\[ (A-12) \]

and under this condition (A-6) can thus be extended to
\[ z v_i M_{ij}(t) = -\hat{p}_{ij}(z) M_{ij}(t) + \hat{p}_{ij}(z) \delta(t) \quad 1-\tilde{\mu}_{ij} < t < 1-s_j \]  
\[ 1 < i < q, \ 1 < j < p \]  
\( (A-13) \)
In order to run the recursive relation (A-13) from $t=1-\tilde{\mu}_j$ until $t=1-s_j$, the Markov parameter entries occurring in the right-hand side of (A-13) have to be available. Some of these entries can be computed by (A-13). These entries that occur in the right-hand side of (A-13) are:

$$M_{kj}(t) \text{ for } r_{ik} \tilde{\mu}_j < t < v_{ik} - s_j \text{ for all } 1 \leq i, k \leq q, \quad 1 \leq j < p$$  \hspace{1cm} (A-14)

The entries that can be computed from (A-13) are:

$$M_{kj}(t) \text{ for } 1-\tilde{\mu}_j + v_k < t < 1-s_j + v_k \text{ for all } 1 \leq k \leq q, \quad 1 \leq j < p$$  \hspace{1cm} (A-15)

Comparing (A-14) and (A-15) the entries that have to be explicitly prescribed as "initial" conditions to (A-13) are:

$$M_{kj}(t) \text{ for } r_{ik} \tilde{\mu}_j < t < \min[\tilde{\mu}_j + v_k, \max_{i}(v_{ik}) - s_j]$$  \hspace{1cm} (A-16)

By inspection it follows that this set of conditions is a subset of the set of equations (A-12).

As a result, it follows that under condition (A-12) the "initial" conditions for running the recursive equations (A-7) and (A-13) for $t=1-\tilde{\mu}_j, \ldots, 1-s_j$ are equal. Since $M_{kj}(t)$ and $\tilde{M}_{kj}(t)$ follow the same recursive relation on the aforementioned time interval, it follows that

$$M_{kj}(t) = \tilde{M}_{kj}(t) \quad \text{for } r_{ik} \tilde{\mu}_j < t < 1-s_j + v_k \quad 1 \leq k \leq q, \quad 1 \leq j < p$$  \hspace{1cm} (A-17)


REFERENCES

Theory and practice of recursive identification. Cambridge, Mass.:  
MIT Press. MIT Press series in signal processing, optimization, and  
control, Vol. 4.

What does system identification have to offer? In: Identification  
and System Parameter Estimation. Proc. 6th IFAC Symp., Washington,  
Pergamon, 1983. P. 77-84.

Equation error versus output error methods in system identification.  
(Submitted for publication).

Design variables for bias distribution in transfer function esti­
mation. In: Proc. 23rd IEEE Conf. on Decision and Control, Las  

Frequency domain properties of identified transfer functions. Re­
port LiTH-ISY-I-0668, University of Linköping, Sweden.

The use of second order information in the approximation of discrete  
Vol. ASSP-24, pp. 226-238.

Approximation of multivariable linear systems with impulse response  

Instrumental variable methods for system identification. Berlin:  
Springer. Lecture notes in control and information sciences,  
Vol. 57.


Invariants and canonical forms for systems; structural and parametric identification. Automatica, Vol. 17, pp. 117-133.


Choice of models for the identification of linear multivariable  
1326-1330.

On the recursive identification of multi-input, multi-output systems.  

Structural identification of linear multivariable systems using  
overlapping forms: a new parametrization. Int. J. Control, Vol. 40,  
pp. 971-987.


The properties of the parametrization of ARMAX systems and their  
1187-1207.

pp. 492-561.

The identification of vector mixed autoregressive moving-average  

Generalized Yule-Walker equations and testing the orders of multi-  
variate time series. Int. J. Control, 1983, Vol. 37, pp. 1159-  
1166.

On-line identification in an unknown stochastic environment. IEEE  

