TABLE I
MONTE CARLO INPUTS

Case	1	2	3	4
Amplitude	2.0	4.0	4.0	6.0
Rate	0.02	0.02	0.2	0.2

TABLE II
MONTE CARLO STATISTICS FOR MODEL PARAMETER ESTIMATES

Case	1	2	3	4
$\tilde{a}_1$	−0.008	−0.01	−0.027	−0.019
$\tilde{a}_1$	0.012	0.013	0.040	0.027
$\tilde{a}_3$	0.005	−0.007	−0.024	−0.016
$\tilde{a}_4$	0.002	−0.002	0.004	0.001
$\hat{\sigma}_{a_1}$	0.019	0.020	0.031	0.021
$\hat{\sigma}_{a_2}$	0.036	0.032	0.046	0.031
$\hat{\sigma}_{a_3}$	0.025	0.024	0.029	0.020
$\hat{\sigma}_{a_4}$	0.014	0.014	0.029	0.007
$\bar{\sigma}_{a_1}$	0.018	0.017	0.011	0.007
$\bar{\sigma}_{a_2}$	0.030	0.028	0.018	0.012
$\bar{\sigma}_{a_3}$	0.030	0.029	0.018	0.008
$\bar{\sigma}_{a_4}$	0.018	0.017	0.011	0.007

TABLE III
MONTE CARLO STATISTICS FOR JUMP PARAMETER ESTIMATES

Case	1	2	3	4
$\tilde{c}$	−0.140	−0.328	−0.057	0.052
$\tilde{\lambda}$	0.016	0.003	0.006	0.006
$\hat{\sigma}_c$	0.936	0.390	0.163	0.087
$\hat{\sigma}_\lambda$	0.016	0.010	0.019	0.014
$\hat{P}_{FA}$	0.0015	0.001	0.008	0.0005
$\hat{P}_{MD}$	0.929	0.214	0.076	0.007
P_{FA}	0.0016	0.001	0.009	0.0006
P_{MD}	0.828	0.152	0.049	0.003

IV. CONCLUSIONS

A joint system model parameter estimation-jump input detection and estimation procedure has been presented. In order to apply this joint estimator-detector, the structure of the jump process must be known, but the parameters of the jump (i.e., the amplitude and rate) may be estimated. Analysis of performance of this algorithm is difficult, since there is two-way coupling between the decision-directed jump input detector/estimator and the system model parameter estimator. Consequently, Monte Carlo studies have been performed to assess empirically the performance. These simulation studies indicate that the algorithm performs well, yielding parameter estimates with negligible bias and satisfactory detection performance. Furthermore, a self-stabilizing effect is observed, indicating that the problem of simultaneously identifying system model parameters and detecting and estimating jump inputs is robust.

REFERENCES

[1] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[2] A. Makowski, W. S. Levine, and M. Asher, "The nonlinear mmse filter for partially observed systems driven by non-Gaussian white noise, with applications to failure estimation," in *23rd IEEE Conf. Decision and Control*, 1984, pp. 644–650.

[3] H. Robbins, "The empirical bayes approach to statistical decision problems," *Rev. Inst. Int. Statist.*, vol. 31, pp. 195–208, 1963.

[4] L. D. Davisson and S. C. Schwartz, "Analysis of a decision-directed receiver with unknown priors," *IEEE Trans. Inform. Theory*, vol. IT-16, no. 3, pp. 270–276, May 1970.

[5] A. Katopis and S. C. Schwartz, "Decision-directed learning using stochastic approximation," in *Proc. 3rd Ann. Pittsburgh Conf. Modeling and Simulation*, Univ. Pittsburgh, Pittsburgh, PA, Apr. 1972, pp. 473–481.

[6] W. C. Stirling and M. Morf, "A new decision-directed algorithm for nonstationary priors," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 10, pp. 928–930, Oct. 1984.

[7] A. Segall, "Recursive estimation from discrete-time point processes," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 4, pp. 422–431, July 1976.

[8] S. C. Schwartz and A. Katopis, "Modified stochastic approximation to enhance unsupervised learning," in *Proc. 1977 Conf. Decision and Control*, 1977, pp. 1067–1079.

[9] D. Kazakos and L. D. Davisson, "An improved decision-directed detector," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 1, pp. 113–116, Jan. 1980.

Some Asymptotic Properties of Multivariable Models Identified by Equation Error Techniques

PAUL VAN DEN HOF AND PETER JANSSEN

Abstract—**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. Assuming that the input is unit variance white noise, it is shown that the Markov parameters of the system are estimated asymptotically unbiased over a certain interval around $t = 0$.**

I. INTRODUCTION

In system identification literature, there is a growing interest in considering situations where the process at hand is not necessarily contained in the chosen model set. This interest is motivated by the fact 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 behavior. The way in which the original process is approximated by the model now is dictated by the applied identification method, and the chosen model set. Equation error techniques are rather popular, mainly due to their computational simplicity. Since, in many situations, the performance of an identified model is judged upon its ability to simulate the process under study, it is important to analyze the simulation behavior of an approximate model obtained by equation error techniques. By considering the Markov parameters of the identified model, we will focus on properties in the time domain. For an analysis in the frequency domain, see [1].

Manuscript received May 5, 1986; revised September 15, 1986.

P. Van den Hof was with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. He is now with the Laboratory for Measurement and Control, Department of Mechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands.

P. Janssen is with the Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands.

IEEE Log Number 8611556.

0018-9286/87/0100-0089$01.00 © 1987 IEEE

Earlier work on this subject has been published in 1976 by Mullis and Roberts [2], 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 [3], but restricted to a full polynomial parametrization. The previous work will be extended to a general class of multivariable models, while the basic instrumental variable method will also be considered.

We will consider a discrete-time, linear, time-invariant process:

$$y(t)=H(z)u(t)+\xi(t) \tag{1}$$

with $y(t)$, $\xi(t) \in R^q$, $u(t) \in R^p$, $t \in Z$, $H(z)$ a $(q \times p)$-transfer matrix, z the forward shift operator, and $\xi(t)$ a random signal that is uncorrelated with the input signal $u(t)$; $u(t)$ and $\xi(t)$ are supposed to be jointly wide-sense stationary and ergodic.

The class of models that we use is a parametrized set of linear, time-invariant, discrete-time multivariable I/O models, given by the following general MFD (matrix fraction description) form:

$$P(z;\ \theta)y(t)=Q(z;\ \theta)u(t)+\epsilon(t;\ \theta) \tag{2}$$

where $P(z; \theta) := [p_{ij}(z)]_{q\times q}$ and $Q(z; \theta) := [q_{ij}(z)]_{q\times p}$ are $(q \times q)$, resp. $(q \times p)$-polynomial matrices; for ease of notation, the explicit dependency of the polynomial entries on the parameter θ has been omitted.

The polynomials are specified by

$$p_{ij}(z)=\delta_{ij}z^{\nu_j}-\alpha_{ij\nu_{ij}}z^{\nu_{ij}-1}-\cdots-\alpha_{ijr_{ij}}z^{r_{ij}-1}] \qquad 1 \leqslant i,\ j \leqslant q \tag{3}$$

$$\text{where } \delta_{ij}=\begin{cases}1 & \text{for } i=j\\ 0 & \text{for } i\neq j\end{cases}$$

$$q_{ij}(z)=\beta_{ij\mu_{ij}}z^{\mu_{ij}-1}+\cdots+\beta_{ijs_j}z^{s_j-1} \qquad 1 \leqslant i \leqslant q,\ 1 \leqslant j \leqslant p. \tag{4}$$

The integer indexes ν_i, ν_{ij}, μ_{ij}, r_{ij}, and s_j determine the structure of the model set (2); without loss of generality they are restricted to $\nu_i \geqslant 0$, $\nu_{ij} \geqslant 0$, $\mu_{ij} \geqslant 0$, $r_{ij} \geqslant 1$, $s_j \geqslant 1$.

If $\nu_{ij} < r_{ij}$, then $p_{ij}(z) = \delta_{ij}z^{\nu_j}$, and if $\mu_{ij} < s_j$, then $q_{ij}(z) = 0$.

As a restriction on the model set we will require that

$$\nu_j \geqslant \nu_{ij} \qquad \text{for } 1 \leqslant i,\ j \leqslant q. \tag{5}$$

This means that the leading column coefficient matrix of $P(z)$ is equal to the identity matrix.

In this paper we will consider the situation where the vector θ of unknown parameters consists of all the coefficients α and β occurring in the polynomial matrices $P(z; \theta)$ and $Q(z; \theta)$. The equation error

$$\epsilon(t;\ \theta)=P(z;\ \theta)y(t)-Q(z;\ \theta)u(t) \tag{6}$$

is dependent on the parameter vector θ, but is not parametrized itself.

The model set (2)–(5) 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, and that it admits both one-step-ahead and more general k-step-ahead prediction models.

II. Asymptotic Results for Equation Error Methods

A common equation error method for obtaining an estimate of θ in (2) is the simple least-squares estimator; in the asymptotic case (infinite number of samples), this corresponds to minimizing

$$V(\theta)=E\epsilon(t;\ \theta)^T\epsilon(t;\ \theta) \tag{7}$$

with respect to θ, under assumption of stationarity and ergodicity of the input and output signals. (E denotes the expectation-operator.) As a direct result of the linearity in the parameters of (6), the asymptotic least squares estimator $\hat{\theta}$ satisfies

$$E\epsilon_i(t;\ \hat{\theta})z^{l-1}\phi_{ij}(t)=0 \qquad r_{ij} \leqslant l \leqslant \nu_{ij},\ 1 \leqslant i,\ j \leqslant q \tag{8}$$

$$E\epsilon_i(t;\ \hat{\theta})z^{l-1}u_j(t)=0 \qquad s_j \leqslant l \leqslant \mu_{ij},\ \begin{matrix}1 \leqslant i \leqslant q\\ 1 \leqslant j \leqslant p\end{matrix} \tag{9}$$

with $\phi_{ij}(t) = y_j(t)$.

For a basic instrumental variable method, equivalent expressions (8) and (9) are valid, with $\phi_{ij}(t)$, however, now being a filtered or delayed input signal (see [4]).

We now define $\hat{P}(z) := P(z; \hat{\theta})$ and $\hat{Q}(z) := Q(z; \hat{\theta})$. The output signal $\hat{y}(t)$ of the estimated model, when excited by the original input signal, is given by

$$\hat{P}(z)\hat{y}(t)=\hat{Q}(z)u(t) \qquad -\infty<t<\infty. \tag{10}$$

If we assume the input signal $u(\cdot)$ to be a zero-mean stationary white noise sequence with unit variance (i.e., $Eu(k)u^T(l) = \delta(k - l)\cdot I$), and we denote

$$M(k) := Ey(t)u^T(t-k) \qquad -\infty<k<\infty \tag{11a}$$

$$\hat{M}(k) := E\hat{y}(t)u^T(t-k) \qquad -\infty<k<\infty \tag{11b}$$

then $M(k)$ is the kth Markov parameter of the process $H(z)$ and $\hat{M}(k)$ the kth Markov parameter associated with the estimated model.

Proposition 1: Let $\hat{P}(z)$ and $\hat{Q}(z)$ be as defined above and let $\hat{\theta}$ be an asymptotic estimator fulfilling (9). Then the Markov parameters of the estimated model and the process satisfy

$$\hat{P}_{i*}(z)M_{*j}(k)=\hat{Q}_{ij}(z)\delta(k) \qquad 1-\mu_{ij} \leqslant k \leqslant 1-s_j$$
$$1 \leqslant i \leqslant q,\ 1 \leqslant j \leqslant p \tag{12}$$

and

$$\hat{P}(z)\hat{M}(k)=\hat{Q}(z)\delta(k)I \qquad -\infty<k<\infty \tag{13}$$

where $\hat{P}_{i*}(z)$, resp. $M_{*j}(k)$ denote the ith row vector of $\hat{P}(z)$, resp. the jth column vector of $M(k)$.

Proof: The result follows directly from (6), (9), and (10).

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. 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.

Theorem 1: Consider a multivariable I/O model as defined in (2)–(5). If this model is used for identifying a linear time invariant system by an equation error technique, fulfilling (9), 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

$$\hat{M}_{ij}(t)=M_{ij}(t) \qquad \text{for } -\tilde{\mu}_j+\tilde{r}_i \leqslant t \leqslant 1-s_j+\nu_i \tag{14}$$

under the condition that the Markov parameters of the original process satisfy

$$M_{ij}(t)=0 \qquad \text{for } -\tilde{\mu}_j+\tilde{r}_i \leqslant t \leqslant \gamma_{ij} \tag{15}$$

where

$$\tilde{r}_j := \min_i r_{ij} \qquad 1 \leqslant j \leqslant q \tag{16}$$

$$\tilde{\mu}_j := \max_i \mu_{ij} \qquad 1 \leqslant j \leqslant p \tag{17}$$

$$\gamma_{ij} := \max\ [\nu_i-\mu_{ij},\ \max_l\ (\nu_{li}-1-\mu_{lj})]. \tag{18}$$

Proof: The proof of this theorem is given in the Appendix.

Theorem 1 has been stated in a general setting, due to the generality of the chosen model set (2)–(5). Its full implications will become clear if it is applied to specific parametrizations, with specific restrictions on the structure indexes. If condition (15) is fulfilled, the result shows that apparently a Padé type of approximation is involved, where the length of the matching interval is determined by the chosen parametrization and the chosen structure indexes of the model.

Note that by (18), γ_{ij} can be interpreted as the maximal difference between the degrees of $p_{li}(z)$ and $q_{lj}(z)$ minus 1:

$$\gamma_{ij} = \max_l [\text{degr } (p_{li}(z)) - \text{degr } (q_{lj}(z))] - 1. \tag{19}$$

When modeling causal systems by MFD model sets, the degrees of the polynomials $q_{ij}(z)$ are usually chosen equal to the ith row degree of $P(z)$:

$$\mu_{ij} - 1 = \max_l [\text{degr } p_{il}(z)] \qquad 1 \leqslant j \leqslant p, \ 1 \leqslant i \leqslant q. \tag{20}$$

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 $Q(z)$. As a result, the corresponding model sets will fulfill $\gamma_{ij} \leqslant -1$.

We can now present the following result as a direct consequence of Theorem 1.

Proposition 2: Consider the situation as described in Theorem 1. If the original system is causal and if $\gamma_{ij} \leqslant -1$, then

$$\hat{M}_{ij}(t) = M_{ij}(t) \qquad \text{for } t \leqslant 1 - s_j + \nu_i, \ 1 \leqslant i \leqslant q, \ 1 \leqslant j \leqslant p. \tag{21}$$

Proof: The result follows directly from Theorem 1 and the observation that $\hat{M}_{ij}(t) = 0$ for $t \leqslant \nu_i - \tilde{\mu}_j$.

A remarkable result is that the estimated model will necessarily be causal too, notwithstanding the fact that the applied model set possibly admits noncausal models.

For most common MFD models, $s_j = 1$ and the above result (21) can be formulated as

$$\hat{M}_{ij}(t) = M_{ij}(t), \qquad t \leqslant \nu_i, \ 1 \leqslant i \leqslant q, \ 1 \leqslant j \leqslant p. \tag{22}$$

The asymptotic matching of Markov parameters, as given in Theorem 1, finds its source in the correlation result (9). Bearing this in mind, the number of Markov parameter entries that is forced to be matched by the equation error method will be equal to the number of β-parameters in the model. If the transfer matrix $[\hat{P}(z)]^{-1}\hat{Q}(z)$ of the model (2) is proper, the sequence of matched Markov parameter entries lies completely in the causal range ($t \geqslant 0$); if not, a noncausal part will also be involved. In [5] the results of this section are applied to a number of commonly used parametrizations, both in forward and backward time shift operator.

Note that for the result of Theorem 1 use has only been made of (9). If the instrumental variables $\phi_{ij}(t)$ are chosen to be delayed input signals, (9) will apparently be fulfilled over a larger interval and the asymptotic matching of Markov parameters will hold on a (possibly) larger interval than in the original results.

An utmost consequence of this result will be illustrated by a simple example.

Example 2.1: Consider a causal SISO-system. Suppose that we use the model set

$$y(t) = -\sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{m} \beta_j u(t-j) + \epsilon(t; \theta). \tag{23}$$

The instrumental variable vector is chosen, according to Wouters [6], to consist of delayed input signals $u(t - l)$ $(0 \leqslant l \leqslant m + n)$.

Application of the foregoing results shows that

$$\hat{M}(t) = M(t) \qquad \text{for } 0 \leqslant t \leqslant m+n. \tag{24}$$

All the 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, a result which is, in fact, equivalent to a Padé approximation. As a consequence of this, use of this IV method can result in unstable models, a phenomenon which is inherent to Padé approximations.

III. 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.

In the asymptotic equation error results, all β-parameters are, in fact, determined in such a way that the Markov parameters of the process are matched over a certain range. The α-parameters determine how the start-sequence of Markov parameters of the model is extended to infinity. As a result, the choice of a parametrization can have a severe effect on the emphasis that is adjusted to the start-sequence, resp. the extension sequence of Markov parameters; the specific set of structure indexes determines the length of the matching interval for the various input–output transmittances within the model.

It has been illustrated in [7] 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. This effect especially occurs when the system impulse response is small in the start-sequence, and increases outside this interval, a very common situation in practice (e.g. because of time delays). A similar result will probably hold for MIMO systems. The results in this paper, presented in terms of Markov parameters, can, of course, also be formulated in terms of step responses.

V. Conclusions

For a general class of linear multivariable models, asymptotic properties are derived for the Markov parameters of approximate models, when identified by equation error identification methods.

The 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 indexes of the model. 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 noncausal models.

For basic IV techniques, it has been shown that the method of Wouters [6] is asymptotically equivalent to Padé approximation in model reduction.

Appendix
Proof of Theorem 1

From (13) we obtain

$$z^{\nu_i}\hat{M}_{ij}(t) = -\hat{P}^*_{i*}(z)\hat{M}_{*j}(t) + \hat{Q}_{ij}(z)\delta(t) \qquad -\infty < t < \infty \tag{A-1}$$

where

$$P^*(z) = P(z) - \text{diag } [z^{\nu_1}, z^{\nu_2}, \cdots, z^{\nu_q}].$$

Since $\hat{Q}_{ij}(z)\delta(t) = 0$ for $t \leqslant -\tilde{\mu}_j$, it follows directly that $z^{\nu_i}\hat{M}_{ij}(t) = 0$ for $t \leqslant -\tilde{\mu}_j$. Combining this with condition (15) results in

$$z^{\nu_i}\hat{M}_{ij}(t) = z^{\nu_i}M_{ij}(t) \qquad \text{for } -\tilde{\mu}_j + \tilde{r}_i - \nu_i \leqslant t \leqslant -\tilde{\mu}_j. \tag{A-2}$$

In order to prove the remaining part of (14), we observe with (12) that

$$z^{\nu_i}M_{ij}(t) = -\hat{P}^*_{i*}(z)M_{*j}(t) + \hat{Q}_{ij}(z)\delta(t) \qquad 1 - \mu_{ij} \leqslant t \leqslant 1 - s_j. \tag{A-3}$$

Under condition (15) this relation will also hold on for $1 - \tilde{\mu}_j \leqslant t \leqslant$

$-\mu_{ij}$. Using (A-1)–(A-3) on the extended interval $1-\tilde{\mu}_j \leqslant t \leqslant 1-s_j$, it follows by induction that

$$z^{\nu_i}\hat{M}_{ij}(t)=z^{\nu_i}M_{ij}(t) \qquad \text{for } 1-\tilde{\mu}_j \leqslant t \leqslant 1-s_j. \tag{A-4}$$

In order to proof this, consider that (A-4) holds for $t = 1 - \tilde{\mu}_j$. For the inductive continuation we assume that $z^{\nu_i}\hat{M}_{ij}(t) = z^{\nu_i}M_{ij}(t)$ for $1-\tilde{\mu}_j \leqslant t \leqslant s^*$, where $1-\tilde{\mu}_j \leqslant s^* \leqslant -s_j$. As a result of (A-3)

$$z^{\nu_i}M_{ij}(s^*+1)=-P^*_{i*}(z)M_{*j}(s^*+1)+Q_{ij}(z)\delta(s^*+1). \tag{A-5}$$

Using the inductive assumption and (A-2), it follows that

$$P^*_{i*}(z)M_{*j}(s^*+1)=P^*_{i*}(z)\hat{M}_{*j}(s^*+1). \tag{A-6}$$

Combination of (A-1), (A-5), and (A-6) finally gives

$$z^{\nu_i}\hat{M}_{ij}(s^*+1)=z^{\nu_i}M_{ij}(s^*+1),$$

providing the inductive argument leading to (A-4).

References

[1] B. Wahlberg and L. Ljung, "Design variables for bias distribution in transfer function estimation," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 134–144, 1986.

[2] C. T. Mullis and R. A. Roberts, "The use of second-order information in the approximation of discrete time linear systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 226–238, 1976.

[3] Y. Inouye, "Approximation of multivariable linear systems with impulse response and autocorrelation sequences," *Automatica*, vol. 19, pp. 265–277, 1983.

[4] T. Söderström and P. G. Stoica, *Instrumental Variable Methods for System Identification* (Lecture Notes in Control and Information Sciences, vol. 57). Berlin, Germany: Springer, 1983.

[5] P. M. J. Van den Hof and P. H. M. Janssen, "Some asymptotic properties of multivariable models identified by equation error techniques," in *Proc. Conf. Decision Contr.*, Athens, Greece, Dec. 10–12, 1986.

[6] W. R. Wouters, "On-line identification in an unknown stochastic environment," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-2, pp. 666–668, 1972.

[7] A. A. H. Damen, R. L. Moses, P. M. J. Van den Hof, and Y. Tomita, "Bounds on the simulation power of equation error estimates," submitted to 1986 Conf. Decision Contr., Athens, Greece, Dec. 1986.