A LINEAR MULTI-STAGE METHOD FOR MULTIVARIATE ARMAX PROCESS IDENTIFICATION - PART I: THE BASIC FORMS OF THE ESTIMATOR

S.D. Fassois[1] and *J.E. Lee*[2]

Report UM-MEAM-90-04

[1] Assistant Professor, Mechanical Engineering and Applied Mechanics

[2] Graduate Research Assistant, Mechanical Engineering and Applied Mechanics

A LINEAR MULTI-STAGE METHOD FOR MULTIVARIATE ARMAX PROCESS IDENTIFICATION PART I: THE BASIC FORMS OF THE ESTIMATOR

S.D. Fassois* and **J.E. Lee****

Department of Mechanical Engineering
and Applied Mechanics
The University of Michigan
Ann Arbor, Michigan 48109-2125

ABSTRACT

A novel and pragmatic method for the identification of multivariate (MIMO) ARMAX processes from input/output observations is proposed. At the core of the method is a parameter estimator that is based on the replacement of the original nonquadratic Prediction Error (PE) estimation problem by an appropriate sequence of simpler quadratic PE and/or linear subproblems, for which uniquely-determined closed-form solutions can be obtained. This is achieved by using basic ARMAX process properties, a truncated ARX representation, and appropriate filtering operations. By offering low computational complexity and mathematically guaranteed algorithmic stability, requiring no initial guess parameter values, and circumventing the local extrema/wrong convergence problems, this parameter estimation technique overcomes some of the main limitations and drawbacks of the currently available methods that have thus far hindered multivariate ARMAX identification in practice.

For the practically important case of unknown process structure, a computationally efficient structure and parameter estimation procedure that overcomes the prohibitive computational requirements of alternative approaches, as it does not need to store the original data or repeat the full estimation calculations as the model order is varied in search of a statistically adequate representation, is introduced within the context of the proposed method.

The paper is divided into two parts: The basic forms of the proposed parameter estimator are derived in the first part, whereas in the second (Fassois and Lee, 1990b) the computationally efficient structure and parameter estimation procedure is developed, and the performance of the method evaluated via numerical simulations.

* Assistant Professor, Mechanical Engineering and Applied Mechanics
** Research Assistant, Mechanical Engineering and Applied Mechanics

1. INTRODUCTION

The problem of determining stochastic linear multivariate (multiple-input multiple-output) models from noise-corrupted observations is of particular importance in many branches of engineering and science. Stochastic linear systems may be represented in a number of ways by using infinitely or finitely parametrized models, such as impulse response models (IR's), matrix fraction descriptions (MFD's), state-space models (SS's), and vector difference equation models (VDE's), which are often refered to as AutoRegressive Moving Average models with eXogeneous inputs (ARMAX's) (El-Sherief, 1982; Ljung, 1987; Gevers and Wertz, 1987).

In this paper the problem of *multivariate* ARMAX process identification is considered. Multivariate ARMAX models offer a convenient finite-dimensional representation of the wide class of multiple-input multiple-output linear systems characterized by rational transfer functions and subject to stochastic disturbances having rational, Hermitian, and positive definite spectra. In addition, they are also well-suited for parameter estimation, and transformations to other model forms may be easily established.

Although multivariate ARMAX models have a superficial resemblance to their univariate counterparts, their structure is much more complicated and gives rise to important identifiability issues that have been studied by a number of mathematicians associated with systems theory (Akaike, 1974; Kalman, 1974; Hannan, 1976; Priestley, 1981; Deistler, 1983; Hannan and Kavalieris, 1984). This complication is due to the fact that there exist many different multivariate ARMAX models giving rise to the same input/output relationship, and hence the same Gaussian likelihood. The set of all possible ARMAX models may be in fact divided into *likelihood equivalence classes*, and the identifiability problem then resolved by selecting a set of models composed of a unique representative from each class. Such a set is known as a (uniquely) *identifiable form* or *parametrization*, and a number of identifiable parametrizations that can be classified as either *canonical* or *pseudo-canonical* (overlapping parametrizations) have been constructed (Hannan, 1976; Kashyap and Rao, 1976; Gevers and Wertz, 1987). Canonical parametrizations are characterized by the fact that they contain a (unique) representation for *any* given system; a fact that is not true for pseudo-canonical forms (see Section 3 of the present).

Although identifiability may be thus resolved, the problem of multivariate ARMAX process identification is still far from being trivial. This is due to difficulties associated with both subproblems into which the (full) identification problem may be decomposed, namely *parameter* and *model structure estimation.*

Indeed, parameter estimation in multivariate ARMAX models is much more complicated than in their univariate counterparts (Pandit and Wu, 1983; Brockwell and Davis, 1987), much more prone to difficulties related to local extrema/wrong convergence and algorithmic instabilities, and the availability of good initial guess parameter values is of critical importance (Jones, 1985). Furthermore the computational complexity of multivariate estimation methods is typically excessive, to the point that identification becomes unrealistic for many applications (Kashyap and Nashburg, 1974; Keviczky et al., 1987; Stoica and Söderström, 1983).

On the other hand structure estimation is accomplished by determining the system's Kronecker indices in the case of a canonical, or the autoregressive (AR), moving-average (MA), and exogeneous (X) orders in the case of a pseudo-canonical representation. This can be achieved by using MFD representations and rank tests of covariance matrices or linear dependence tests on the rows of the estimated Hankel matrix (Tse and Weinert, 1975). Such procedures are, however, not only time-consuming but also numerically sensitive, and this may create significant problems as the accurate estimation of the Kronecker indices is of critical importance for the consistent estimation of the process parameters (Caines and Rissanen, 1974; Gevers and Wertz, 1987). Alternatively, procedures based on statistical decision theory criteria, such as the Final Prediction Error (FPE) criterion, Akaike's Information Criterion (AIC), and the BIC, have been also proposed (Akaike, 1974; 1977). These, however, require the successive fitting of a possibly large number of candidate models to the data, so that in addition to the difficulties associated with parameter estimation, an enormous computational complexity that makes them prohibitive for many applications is required (Hannan and Kavalieris, 1984).

Because of all these difficulties and limitations, multivariate ARMAX process identification methods are far from being standard in practice, and have been thus far applied to a relatively small number of actual problems. Furthermore, the multivariate identification problem has been significantly less extensively studied in the literature when compared to its univariate counterpart. The main available methods are briefly reviewed in the sequel.

The primary method for multivariate ARMAX process identification is based on the Maximum Likelihood (ML) principle [and becomes essentially equivalent to a general Prediction Error (PE) method if the Gaussianity assumption is removed], combined with an appropriate structure estimation procedure. In early approaches the parameter and structure estimation subproblems were in fact separated, with the latter solved by using rank tests of covariance matrices or linear dependence tests on the rows of the estimated process Hankel matrix, and preceding the former. More recently, however, combined structure and parameter estimation methods that involve the extension of the likelihood function with a penalty term that is a function of model complexity, have been introduced. Structure and parameter estimation is then achieved by optimizing this combined criterion. The aforementioned FPE, AIC, and BIC criteria have all been proposed within this context.

The ML method has been discussed in a number of publications, including Kashyap and Nashburg (1974) who considered the estimation of multivariate ARMAX processes in the so-called canonical form I by using both the exact and conditional ML methods, and Keviczky and Banyasz (1978) who discussed ML estimation for ARMAX processes in the fully-parametrized pseudo-canonical form. Deistler (1983) is an additional reference in this context. The ML method is characterized by optimal asymptotic properties, namely consistency and efficiency (Hannan and Kavalieris, 1984; Keviczky et al., 1987), but the estimation procedure is very complicated as the maximization of the likelihood function has to be performed simultaneously with respect to all unknown parameters, unlike the univariate case where this can be relaxed and the innovations variance estimated after the rest of the parameters have been obtained (Brockwell and Davis, 1987). As a consequence, the computational complexity becomes prohibitive for many applications, especially in the case of unknown model structure where a possibly large number of models has to be successively fitted. In addition, the problem of algorithmic instability becomes much more severe in the multivariate case, especially in conjunction with processes having MA roots close to the unit circle, and the likelihood function may be (even asymptotically) characterized by several local maxima that are much smaller than the global maximum and therefore lead to completely erroneous estimation results (Brockwell and Davis, 1987). As a consequence "good" initial parameter values are of crucial importance, and towards this end, the use of suboptimum procedures, such as the initial fitting of univariate models to each scalar component of the model, has been recommended in the literature (Jones, 1985). With respect to structure estimation the general comments made earlier regarding the difficulties associated with the estimation of the process Kronecker indices or the AR, MA, and X orders, are fully applicable here as well.

In an apparent attempt to alleviate some of the aforementioned difficulties, Kashyap and Nashburg (1974) also proposed the Limited Information Estimates based on their so-called canonical forms II and III. The main idea is on the ad-hoc decomposition of the multivariate estimation problem into a set of simpler univariate subproblems that are independently solved.

Other alternative approaches that have been discussed in the literature include the Ordinary Least Squares (OLS) method in conjunction with canonical form I (Kashyap and Nashburg, 1974) and the fully-parametrized pseudo-canonical form (Keviczky and Banyasz, 1978), as well as versions of the Generalized Least Squares (GLS) method in conjunction with canonical form I (Kashyap and Nashburg, 1974). In Keviczky and Banyasz (1978) and Goodwin et al. (1978) the GLS method is used in conjunction with AutoRegressive AutoRegressive models with eXogeneous inputs (ARARX models). In addition to being unable to provide estimates of the MA parameters, the OLS method is known to yield highly inconsistent estimates, whereas problems of poor performance at high noise-to-signal (N/S) ratios, the possibility of slow or wrong convergence, and a relatively high computational complexity due to a possibly large number of iterations, are associated with GLS.

Instrumental Variable (IV) methods for the estimation of the AR and X parameters have been also proposed (Stoica and Söderström 1982; 1983). Such a method was used in conjunction with canonical form I by Kashyap and Nashburg (1974), whereas IV methods based on the fully-parametrized pseudo-canonical form and a canonical form characterized by a diagonal AR polynomial matrix, were discussed by Stoica and Söderström (1982). The same authors (Stoica and Söderström, 1983) proposed optimal IV methods, and a similar procedure refered to as Instrumental Variable-Approximate Maximum Likelihood (IV-AML) was discussed by Jakeman and Young (1979) for both the fully-parametrized pseudo-canonical form and the canonical form characterized by a diagonal AR polynomial matrix. Both the optimal IV and IV-AML methods, however, require "bootstrapping" procedures that combine, in an iterative manner, the optimal IV scheme with an algorithm for the estimation of the noise dynamics. In the multivariate case the practical implementation of such methods is thus cumbersome (Stoica and Söderström, 1983), and problems of slow or wrong convergence and algorithmic instability may be encountered. It is in addition known that IV methods do not, in general, achieve the asymptotic accuracy of ML (or general PE) methods (Stoica and Söderström, 1983).

Kashyap and Nashburg (1974) have also discussed a method based on canonical form I and the multivariate extension of a simple, but statistically inefficient, estimation scheme proposed by Durbin (1960). Finally, Hannan and Kavalieris (1984) proposed a combined structure and parameter estimation method based on a three-stage scheme also inspired by Durbin's method and using the BIC criterion. This method estimates the model structure (Kronecker indices) by using inexpensively computed parameter estimates, and once the structure has been thus determined, statistically efficient parameter estimates are obtained through nonlinear techniques based on the Maximum Likelihood principle.

In this paper a novel and pragmatic method for the identification of multivariate ARMAX processes, that is based on exclusively *linear* techniques, and overcomes the difficulties of alternative approaches, namely:

- the excessive computational complexity,
- the occurence of algorithmic instabilities, and the approaches' inability to estimate processes characterized by MA roots close to the unit circle,
- the existence of local extrema in the identification criterion and the associated problems of wrong convergence,
- the need for initial guess parameter values,
- the sensitivity and tremendous computational complexity problems associated with structure estimation,

is introduced. The proposed method is based on a novel *linear multi-stage* parameter estimator inspired by the newly developed univariate Suboptimum Maximum Likelihood algorithm (Fassois and Lee, 1990a), the fully-parametrized pseudo-canonical form, and the BIC criterion.

At the core of the proposed method is the *main form* of the linear multi-stage parameter estimator that is based on the replacement of the original nonquadratic Prediction Error (PE) estimation problem by an appropriate sequence of simpler quadratic and/or PE subproblems, for which uniquely-determined closed-form solutions can be obtained. This is achieved by using basic ARMAX process properties, a truncated ARX representation, and appropriate filtering operations. In addition, a *guaranteed-stability form* of the estimator that overcomes the very important problem of algorithmic instabilities, and allows for the identification of all types of processes, including those characterized by MA roots close to the unit circle, is developed.

For the practically important case of unknown process structure, a novel and computationally very efficient procedure for *structure and parameter estimation*, that is based on a very fast correlation-type form of the linear multi-stage parameter estimator, the fully-parametrized ARMAX pseudo-canonical form, and the BIC, is developed within the context of the proposed method. This procedure uses the second-order moments of the measured data as a (pseudo) sufficient statistic, and is also based on exclusively linear techniques. The numerically sensitive calculations required for the estimation of the Kronecker indices and the tremendous computational complexity of alternative methods are circumvented, as the proposed procedure needs to estimate only the AR, MA, and X orders and achieves remarkably low computational and memory storage requirements since the original observations need not be stored in the memory, and only a small portion of the total estimation calculations has to be repeated as the model order is varied in search of a statistically adequate representation.

This paper is divided into two parts: Parameter estimation is primarily treated in Part I, where the main and guaranteed-stability forms of the estimator are derived. The combined structure and parameter estimation problem is treated in Part II (Fassois and Lee, 1990b), where the corresponding procedure is developed, and, in addition, the performance characteristics of the proposed linear multi-stage method are examined via numerical simulations. The presentation in Part I of the paper is organized as follows: The exact problem statement, along with some necessary background material on multivariate ARMAX processes, is presented in Section 2, and the derivation of the main form of the linear multi-stage parameter estimator given in Section 3. The derivation of the guaranteed-stability form of the estimator is discussed in Section 4, and the conclusions are finally summarized in Section 5.

2. MULTIVARIATE ARMAX PROCESSES: DEFINITIONS AND STRUCTURAL CONSIDERATIONS

In this section the multivariate ARMAX processes are defined, the corresponding identification problem posed, and certain issues pertaining to identifiability discussed.

The working assumption throughout the paper is that there exists an actual multivariate stochastic system S, that belongs to the class of discrete-time linear systems whose input-output relationships are characterized by rational transfer functions and the disturbances acting on them are wide-sense stationary stochastic signals with rational spectral densities, which is to be identified based on input/output observations. The system S is thus amenable to a multivariate ARMAX description of the form[1]:

$$S: \qquad \mathbf{A}^{o}(B)\cdot\mathbf{y}[t] \;=\; \mathbf{B}^{o}(B)\cdot\mathbf{x}[t] \;+\; \mathbf{C}^{o}(B)\cdot\mathbf{w}[t] \tag{1}$$

with $\{\mathbf{y}[t]\}$ representing the observable noise-corrupted s-dimensional output sequence, $\{\mathbf{x}[t]\}$ the observable m-dimensional input sequence, and $\{\mathbf{w}[t]\}$ an unobservable s-dimensional zero-mean white noise (innovations) process with covariance matrix $\mathbf{\Sigma}^{o}$. $\mathbf{A}^{o}(B), \mathbf{B}^{o}(B)$, and $\mathbf{C}^{o}(B)$ represent polynomial matrices in the backshift operator B (defined such that $B \cdot \mathbf{y}[t] = \mathbf{y}[t-1]$), and are refered to as the *Autoregressive, Exogeneous,* and *Moving Average* polynomials, respectively. Those may be expressed as:

$$\mathbf{A}^{o}(B) \;\stackrel{\Delta}{=}\; \mathbf{I}_s + \mathbf{A}^{o}(1)\cdot B + \cdots + \mathbf{A}^{o}(na)\cdot B^{na} \qquad [s \times s] \tag{2}$$

$$\mathbf{B}^{o}(B) \;\stackrel{\Delta}{=}\; \mathbf{B}^{o}(1)\cdot B + \cdots + \mathbf{B}^{o}(nb)\cdot B^{nb} \qquad [s \times m] \tag{3}$$

$$\mathbf{C}^{o}(B) \;\stackrel{\Delta}{=}\; \mathbf{I}_s + \mathbf{C}^{o}(1)\cdot B + \cdots + \mathbf{C}^{o}(nc)\cdot B^{nc} \qquad [s \times s] \tag{4}$$

where $\mathbf{I}_s$ stands for the identity matrix of order s, $\mathbf{A}^{o}(j)$, $\mathbf{B}^{o}(j)$, $\mathbf{C}^{o}(j)$ are coefficient matrices, and the quantities in the brackets indicate vector/matrix dimensions. The nonnegative integers na, nb, and nc that represent the polynomial orders are known as the autoregressive, exogeneous, and moving average order, respectively.

The system of Eqs. (1)-(4) is also assumed to satisfy the following standard assumptions:

A1. The sequence $\{\mathbf{w}[t]\}$ is an s-dimensional, zero-mean and uncorrelated (white) process that is wide-sense stationary and second-order ergodic with a positive-definite covariance matrix $\mathbf{\Sigma}^{o}$.

A2. The observable input $\{\mathbf{x}[t]\}$ is an m-dimensional, zero-mean, wide-sense stationary and second-order ergodic process with a positive-definite covariance matrix, and is persistently exciting of sufficiently high order.

A3. The input $\{\mathbf{x}[t]\}$ and the white noise $\{\mathbf{w}[t]\}$ are mutually independent.

[1]The superscript o is used to indicate the actual system and distinguish it from any given model.

A4. All zeros of the determinant of $\mathbf{A}^o(B)$ lie outside the unit circle (*"stationarity assumption"*).

A5. All zeros of the determinant of $\mathbf{C}^o(B)$ lie outside the unit circle (*"invertibility assumption"*).

A6. The Smith form of $[\mathbf{A}^o(B), \mathbf{B}^o(B), \mathbf{C}^o(B)]$ is $[\mathbf{I}, \mathbf{0}]$, or, equivalently, the greatest common left factor of $\mathbf{A}^o(B), \mathbf{B}^o(B)$ and $\mathbf{C}^o(B)$ is a unimodular matrix (i.e. a matrix with nonzero constant determinant).

The problem of multivariate ARMAX process identification may be then posed as follows:

"Given a set of N input $\{\mathbf{x}[t]\}$ and noise-corrupted output $\{\mathbf{y}[t]\}$ vector observations, identify a model of the form:

$$M: \quad \mathbf{A}(B,\boldsymbol{\theta})\cdot\mathbf{y}[t] \;=\; \mathbf{B}(B,\boldsymbol{\theta})\cdot\mathbf{x}[t] \;+\; \mathbf{C}(B,\boldsymbol{\theta})\cdot\mathbf{e}[t/\boldsymbol{\theta}] \qquad E\{\mathbf{e}[t/\boldsymbol{\theta}]\cdot\mathbf{e}^T[t/\boldsymbol{\theta}]\} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) \qquad (5)$$

that "best" fits the data."

In Eq.(5) $\boldsymbol{\theta}$ denotes the 4-tuple $\boldsymbol{\theta} \stackrel{\Delta}{=} (\mathbf{A}, \mathbf{B}, \mathbf{C}, \boldsymbol{\Sigma})$ that includes all parameters to be estimated and $\mathbf{e}[t/\boldsymbol{\theta}]$ the one-step-ahead prediction error corresponding to the model structure M.

As is however well-known, this identification problem is ill-posed, because of the lack of *identifiability* within the general ARMAX model structure M defined by Eq.(5). This is true even under the assumption of existence of a system S, defined by Eqs.(1)-(4) and subject to assumptions A1-A6, actually generating the observations, and is due to the fact that there exist many ARMAX models giving rise to the same input/output relationships (Hannan, 1976; Kashyap and Rao, 1976). All multivariate ARMAX models may be thus divided into *equivalence classes* according to an appropriate *equivalence relationship*. By considering the equivalence relationship giving rise to the same Gaussian likelihood function, *likelihood equivalence classes* may be constructed, and a set consisting of a *unique* representative from each such class is then called an *identifiable form* or *structure*. Since the Gaussian likelihood depends only on the spectra of the signals $\{\mathbf{x}[t]\}$ and $\{\mathbf{y}[t]\}$, the likelihood equivalence classes in the Gaussian case[2] are identical to *spectra equivalence classes*, and an identifiable form can be in this case constructed by selecting elements from each equivalence class that yield distinct auto and cross spectra $\mathbf{S}_{yy}(\omega)$, $\mathbf{S}_{yx}(j\omega)$, where ω represents frequency and j the imaginary unit.

A number of such identifiable forms, that can be classified as either *canonical* or *pseudo-canonical* (overlapping parametrizations), have been constructed (Hannan, 1976; Kashyap and Rao, 1976; Gevers and Wertz, 1987). Canonical forms have the advantage that any given system will *always* have a (unique) representative within the canonical form. Their utilization in identification becomes, however, problematic because of the fact that estimates of a set of structural parameters, known as *Kronecker indices*, need to be obtained by very time-consuming and sometimes

[2]It should be however clarified that Gaussianity is not restrictive in this context. Identifiability can be always defined based on spectra; just in the non-Gaussian case the term *second-order identifiability* should be more appropriately used.

numerically sensitive procedures (Gevers and Wertz, 1987). Moreover, the accurate estimation of the Kronecker indices is very critical for obtaining accurate parameter estimates, and, as is well-known, inconsistent parameter estimates are obtained if the indices are wrongly estimated (Caines and Rissanen, 1974). For these reasons the use of pseudo-canonical forms has been often recommended in identification. Pseudo-canonical forms may be indeed advantageous in the sense that instead of estimating the full set of structural indices one only needs to determine the system orders, and the given system may be then almost surely identified within the pseudo-canonical form. Their main disadvantage is that they may contain a few parameters more than canonical forms, but this is not viewed to be a serious drawback since it has been observed that the difference in the number of parameters between pseudo-canonical and certain canonical forms may be quite small (Kashyap and Rao, 1976).

3. THE MAIN FORM OF THE ESTIMATOR

In this section the main form of the linear multi-stage method for parameter estimation of multivariate ARMAX processes represented in the fully-parametrized pseudo-canonical form, also refered to as pseudo-canonical form II (Kashyap and Rao, 1976), is presented.

The fully-parametrized ARMAX pseudo-canonical form is characterized by full polynomial matrices $\mathbf{A}(B,\boldsymbol{\theta})$, $\mathbf{B}(B,\boldsymbol{\theta})$, and $\mathbf{C}(B,\boldsymbol{\theta})$, and an also full innovations covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ [see Eq.(5)]. This parametrization yields identifiability provided that the system S to be identified is described by Eqs.(1)-(4), and is subject to assumptions A1-A6, as well as the additional assumption:

A7. $Rank[\mathbf{A}^{o}(na) \quad \mathbf{B}^{o}(nb) \quad \mathbf{C}^{o}(nc)] = s$

This fact has been proven by Hannan [see for instance Hannan (1976)], who has additionally shown that there exist equivalence classes with no representative in the fully-parametrized pseudo-canonical form. Those correspond to systems that do not satisfy A7, but, as it has been pointed out, those equivalence classes are not expected to occur, and even more relevantly, it is most unlikely that the maximum of the likelihood, for any fixed orders na, nb, nc, will be at a point not included in this form (Hannan, 1976). As a consequence A7 does not represent a real practical limitation for identification, whereas it significantly simplifies the structure determination problem as only three structural indices, namely the orders na, nb, and nc, need to be determined, and the complicated and numerically sensitive procedures required for structure determination within canonical forms are circumvented.

The main form of the proposed linear multi-stage method is based on the replacement of the original nonquadratic Prediction Error (PE) estimation problem by an appropriate sequence of simpler quadratic PE and/or linear subproblems, for which uniquely-determined closed-form solutions may be obtained. More specifically, the estimation procedure is based on the observation that after separating the estimation of the parameter matrices from that of the innovations covariance, the original PE problem can be reduced into a quadratic one for a fixed MA polynomial matrix. An initial estimate of that matrix is then obtained by using a truncated multivariate ARX model structure, fundamental properties of multivariate ARMAX models, and PE-type estimators. By using appropriate filtering operations the problem is transformed into a form that is quadratic in the AR and X matrices, which are subsequently estimated by using a PE-type procedure. The final MA polynomial matrix and innovations covariance estimates are then obtained by using earlier derived estimates and ARMAX model properties.

This estimation method overcomes the severe problems associated with the ML and general PE approaches, while also circumventing the limitations (algorithmic instabilities and the need for a possibly large number of iterations) of optimal Instrumental Variable (IV) schemes. Also, unlike the GLS algorithm, the proposed method does not start with highly inconsistent estimates that have to be subsequently refined, and, in contrast to the method of Hannan and Kavalieris (1984), it does not depend on innovations sequence estimates. The main advantages of the proposed method may be thus summarized as: (a) the use of exclusively linear techniques and the resulting low computational complexity, (b) uniquely-determined estimates and the elimination of the local extrema/wrong convergence problems, and (c) no need for initial guess parameter values. Furthermore, a guaranteed-stability form that enables the method to also overcome the severe instability problems of alternative techniques and estimate all types of processes, including those characterized by MA roots close to the unit circle, is also introduced in Section 4 of the paper.

The main form of the linear multi-stage parameter estimation method consists of the following four stages:

Stage 1: Estimation of a truncated ARX representation.

Based on the invertibility assumption A5, and by introducing the impulse response matrices $\{\mathbf{H}_y^\circ(j)\}_{j=0}^{\infty}$ and $\{\mathbf{H}_x^\circ(j)\}_{j=1}^{\infty}$, the system representation S of Eq.(1) can be put into the following form:

$$S: \qquad \mathbf{H}_y^\circ(B)\cdot\mathbf{y}[t] = \mathbf{H}_x^\circ(B)\cdot\mathbf{x}[t] + \mathbf{w}[t] \qquad (6)$$

with:

$$\mathbf{H}_y^\circ(B) \triangleq [\mathbf{C}^\circ(B)]^{-1}\cdot\mathbf{A}^\circ(B) = \sum_{j=0}^{\infty}\mathbf{H}_y^\circ(j)\cdot B^j \qquad [s\times s] \qquad (7)$$

$$\mathbf{H}_x^\circ(B) \triangleq [\mathbf{C}^\circ(B)]^{-1}\cdot\mathbf{B}^\circ(B) = \sum_{j=1}^{\infty}\mathbf{H}_x^\circ(j)\cdot B^j \qquad [s\times m] \qquad (8)$$

and $\mathbf{H}_y^\circ(0) \triangleq \mathbf{I}_s$, with the middle expressions in (7) and (8) providing Matrix Fraction Descriptions of $\mathbf{H}_y^\circ(B)$ and $\mathbf{H}_x^\circ(B)$, respectively. The algorithm then starts by estimating the matrix $\mathcal{H}_1 \in$

$\Re^{s \times p(s+m)}$ defined as:

$$\mathcal{H}_1 \triangleq \left[\mathbf{H}_y(1)\ \mathbf{H}_y(2) \cdots \mathbf{H}_y(p) \vdots \mathbf{H}_x(1)\ \mathbf{H}_x(2) \cdots \mathbf{H}_x(p) \right] \tag{9}$$

within the following ARX model structure:

$$M_1: \qquad \sum_{j=0}^{p} \mathbf{H}_y(j) \cdot \mathbf{y}[t-j] = \sum_{j=1}^{p} \mathbf{H}_x(j) \cdot \mathbf{x}[t-j] + \epsilon[t/\mathcal{H}_1] \tag{10}$$

which is a truncated version of the system equation (6) of finite order p. In this model structure $\mathbf{H}_y(0) \triangleq \mathbf{I}_s$, and $\epsilon[t/\mathcal{H}_1]$ represents the associated one-step-ahead prediction error. By manipulating equation (10) the following alternative expression may be derived:

$$M_1: \qquad \mathbf{y}[t] = \left[\mathbf{I}_s \otimes \mathbf{u}_1^T[t]\right] \cdot \boldsymbol{h}_1 + \epsilon[t/\mathcal{H}_1] \qquad [s \times 1] \tag{11}$$

in which:

$$\mathbf{u}_1[t] \triangleq \left[-\mathbf{y}^T[t-1]\ -\mathbf{y}^T[t-2] \cdots -\mathbf{y}^T[t-p] \vdots \mathbf{x}^T[t-1] \cdots \mathbf{x}^T[t-p]\right]^T \tag{12}$$

$$\begin{aligned} \boldsymbol{h}_1 &\triangleq \operatorname{col}\left[\left[\mathbf{H}_y(1) \cdots \mathbf{H}_y(p) \vdots \mathbf{H}_x(1) \cdots \mathbf{H}_x(p)\right]^T \right] \\ &\triangleq \left[\ \mathbf{h}_1^T(1)\ \mathbf{h}_1^T(2)\ \cdots\ \mathbf{h}_1^T(s)\ \right]^T \end{aligned} \tag{13}$$

where $\mathbf{u}_1 \in \Re^{p(s+m)\times 1}$, $\boldsymbol{h}_1 \in \Re^{sp(s+m)\times 1}$, and $\otimes$ denotes Kronecker product (Magnus and Neudecker, 1988), the superscript ^{T} transpose, and the $p(s+m)$-dimensional vector $\mathbf{h}_1^T(i)$ represents the i-th row of the matrix $\mathcal{H}_1$. The model structure M_1 of Eq.(11) consists of a number of scalar forms that have a common regression vector $\mathbf{u}_1[t]$, and except for the vector $\boldsymbol{h}_1$, the prediction error covariance matrix is also unknown. The estimation of this type of models within the framework of the Conditional Maximum Likelihood method (or the Prediction Error method in the non-Gaussian case) is known to have an interesting solution according to which the estimator of the parameter vector $\boldsymbol{h}_1$ may be decomposed into a set of closed-form estimators of the form [Theorem 6.a.1 of Kashyap and Rao (1976)]:

$$\hat{\mathbf{h}}_1(i) = \left[\sum_{t=1}^{N} \mathbf{u}_1[t] \cdot \mathbf{u}_1^T[t] \right]^{-1} \cdot \left[\sum_{t=1}^{N} \mathbf{u}_1[t] \cdot y_i[t] \right] \qquad (i = 1, 2, \cdots, s) \tag{14}$$

where $y_i[t]$ represents the i-th element of the output vector $\mathbf{y}[t] = \left[y_1[t] \cdots y_i[t] \cdots y_s[t]\right]^T$. The collection of the above linear least-squares (LLS) estimators thus forms the desired estimator of the matrix $\mathcal{H}_1$. □

Stage 2: Initial estimation of the MA polynomial matrix $\mathbf{C}^{\circ}(B)$.

Based on the definition of the impulse response sequence $\{\mathbf{H}_y^{\circ}(j)\}$ [Eq.(7)], the following expression is derived:

$$\sum_{j=0}^{\min(i,nc)} \mathbf{C}^{\circ}(j) \cdot \mathbf{H}_y^{\circ}(i-j) = \mathbf{A}^{\circ}(i) \qquad i = 0, 1, 2, \cdots \tag{15}$$

in which $\mathbf{A}^o(i) \triangleq \mathbf{0}$ for $i > na$. By collecting together these matrix expressions for $i = r - nc + 1, \cdots, r$ $[r \geq \max(na, nc) + nc]$, and substituting the theoretical values of the impulse response samples $\mathbf{H}_y^o(j)$ by their corresponding estimates obtained in Stage 1, the following estimator expression can be derived for the MA matrices:

$$\begin{bmatrix} \hat{\mathbf{H}}_y^T(r-nc) & \hat{\mathbf{H}}_y^T(r-nc-1) & \cdots & \hat{\mathbf{H}}_y^T(r-2nc+1) \\ \hat{\mathbf{H}}_y^T(r-nc+1) & \hat{\mathbf{H}}_y^T(r-nc) & \cdots & \hat{\mathbf{H}}_y^T(r-2nc+2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{H}}_y^T(r-1) & \hat{\mathbf{H}}_y^T(r-2) & \cdots & \hat{\mathbf{H}}_y^T(r-nc) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{C}}^T(1) \\ \hat{\mathbf{C}}^T(2) \\ \vdots \\ \hat{\mathbf{C}}^T(nc) \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{H}}_y^T(r-nc+1) \\ -\hat{\mathbf{H}}_y^T(r-nc+2) \\ \vdots \\ -\hat{\mathbf{H}}_y^T(r) \end{bmatrix} \tag{16}$$

□

<u>Stage 3</u>: Estimation of the AR and X matrices.

The estimation of the AR and X matrices is based on the observation that a PE criterion is quadratic in them for known MA matrices. Indeed, for a given estimate $\hat{\mathbf{C}}(B)$ of the MA matrix polynomial, the following model structure may be constructed by replacing $\mathbf{C}(B,\theta)$ by $\hat{\mathbf{C}}(B)$ in the original model structure of Eq.(5):

$$M_2: \quad \hat{\mathbf{C}}^{-1}(B)\cdot\mathbf{y}[t] = \hat{\mathbf{C}}^{-1}(B)\cdot\sum_{j=1}^{nb}\mathbf{B}(j,\theta_2)\cdot\mathbf{x}[t-j] - \hat{\mathbf{C}}^{-1}(B)\cdot\sum_{j=1}^{na}\mathbf{A}(j,\theta_2)\cdot\mathbf{y}[t-j] + \epsilon'[t/\theta_2] \tag{17}$$

In this structure θ_2 represents the AR and X parameter vector defined as:

$$\theta_2 \triangleq \mathrm{col}[\mathbf{A}(1)\,\mathbf{A}(2)\,\cdots\,\mathbf{A}(na)\,\mathbf{B}(1)\,\cdots\,\mathbf{B}(nb)] \qquad [(s^2\cdot na + m\cdot s\cdot nb)\times 1] \tag{18}$$

and $\epsilon'[t/\theta_2]$ the associated one-step-ahead prediction error.

By using Theorem 2 of Chapter 2 of Magnus and Neudecker (1988), the following expressions may be obtained:

$$\hat{\mathbf{C}}^{-1}(B)\cdot\mathbf{y}[t] = [\,\mathbf{y}^T[t]\otimes\hat{\mathbf{C}}^{-1}(B)\,]\cdot\mathrm{col}[\mathbf{I}_s] \tag{19}$$

$$\hat{\mathbf{C}}^{-1}(B)\cdot\mathbf{A}(j)\cdot\mathbf{y}[t-j] = [\,\mathbf{y}^T[t-j]\otimes\hat{\mathbf{C}}^{-1}(B)\,]\cdot\mathrm{col}[\mathbf{A}(j)] \tag{20}$$

$$\hat{\mathbf{C}}^{-1}(B)\cdot\mathbf{B}(j)\cdot\mathbf{x}[t-j] = [\,\mathbf{x}^T[t-j]\otimes\hat{\mathbf{C}}^{-1}(B)\,]\cdot\mathrm{col}[\mathbf{B}(j)] \tag{21}$$

and by defining the following vector and matrix forms of the filtered observations:

$$\mathbf{y}_F[t] \triangleq [\,\mathbf{y}^T[t]\otimes\hat{\mathbf{C}}^{-1}(B)\,]\cdot\mathrm{col}[\mathbf{I}_s] \qquad [s\times 1] \tag{22}$$

$$\mathbf{Y}_F[t-j] \triangleq \mathbf{y}^T[t-j]\otimes\hat{\mathbf{C}}^{-1}(B) \qquad [s\times s^2] \tag{23}$$

$$\mathbf{X}_F[t-j] \triangleq \mathbf{x}^T[t-j]\otimes\hat{\mathbf{C}}^{-1}(B) \qquad [s\times m\,s] \tag{24}$$

the model structure M_2 of Eq.(17) may be rewritten as:

$$M_2: \quad \mathbf{y}_F[t] = \sum_{j=1}^{nb}\mathbf{X}_F[t-j]\cdot\mathrm{col}[\mathbf{B}(j,\theta_2)] - \sum_{j=1}^{na}\mathbf{Y}_F[t-j]\cdot\mathrm{col}[\mathbf{A}(j,\theta_2)] + \epsilon'[t/\theta_2] \Rightarrow$$

$$\Rightarrow \mathbf{y}_F[t] = \mathbf{U}_{1F}[t]\cdot\theta_2 + \epsilon'[t/\theta_2] \tag{25}$$

in which the matrix $\mathbf{U}_{1F} \in \Re^{s\times(s^2\cdot na+m\cdot s\cdot nb)}$ is defined as:

$$\mathbf{U}_{1F}[t] \triangleq \left[-\mathbf{Y}_F[t-1] \;\; -\mathbf{Y}_F[t-2]\cdots-\mathbf{Y}_F[t-na] \;\vdots \mathbf{X}_F[t-1]\cdots\mathbf{X}_F[t-nb]\right] \tag{26}$$

Since the model structure M_2 is linear in $\boldsymbol{\theta}_2$, minimization of the PE criterion:

$$J \triangleq trace\Big(\frac{1}{N}\sum_{t=1}^{N}\epsilon'[t/\boldsymbol{\theta}_2]\cdot\epsilon'^T[t/\boldsymbol{\theta}_2]\Big) \tag{27}$$

leads to the following LLS estimator:

$$\hat{\theta}_2 = \left[\sum_{t=1}^{N}\mathbf{U}_{1F}^T[t]\cdot\mathbf{U}_{1F}[t]\right]^{-1}\cdot\left[\sum_{t=1}^{N}\mathbf{U}_{1F}^T[t]\cdot\mathbf{y}_F[t]\right] \tag{28}$$

In implementing, however, this estimator, the filtering operations that define the vector and matrix sequences $\{\mathbf{y}_F[t]\}_{t=1}^N$, $\{\mathbf{Y}_F[t]\}_{t=1}^N$, and $\{\mathbf{X}_F[t]\}_{t=1}^N$, have to be performed. Towards this end consider the matrix $\mathbf{Y}_F[t]$ first:

$$\begin{aligned}\mathbf{Y}_F[t] &\triangleq \mathbf{y}^T[t]\otimes\hat{\mathbf{C}}^{-1}(B) = \left[\,y_1[t]\cdot\hat{\mathbf{C}}^{-1}(B)\vdots y_2[t]\cdot\hat{\mathbf{C}}^{-1}(B)\vdots\cdots\vdots y_s[t]\cdot\hat{\mathbf{C}}^{-1}(B)\,\right]\\ &\triangleq \left[\;\mathbf{Y}_1[t]\;\vdots\;\mathbf{Y}_2[t]\;\vdots\cdots\vdots\;\mathbf{Y}_s[t]\;\right]\end{aligned} \tag{29}$$

In this expression the i-th block element $\mathbf{Y}_i[t]$ of $\mathbf{Y}_F[t]$ is defined as:

$$\mathbf{Y}_i[t] \triangleq y_i[t]\cdot\hat{\mathbf{C}}^{-1}(B) \qquad [s\times s] \tag{30}$$

from which the following recursive equation may be obtained:

$$\sum_{j=0}^{nc}\mathbf{Y}_i[t-j]\cdot\hat{\mathbf{C}}(j) = y_i[t]\cdot\mathbf{I}_s \tag{31}$$

with $\hat{\mathbf{C}}(0)\triangleq\mathbf{I}_s$. Based on this, and by assuming $\mathbf{Y}_i[t]\equiv\mathbf{0}$ for $t<0$, the sequence $\{\mathbf{Y}_i[t]\}$ may be obtained. $\{\mathbf{y}_F[t]\}_{t=1}^N$ thus becomes also available, whereas the sequence $\{\mathbf{X}_F[t]\}_{t=1}^N$ is similarly obtained by assuming that $\mathbf{X}_F[t]\equiv\mathbf{0}$ for $t<0$. □

Stage 4: Estimation of the MA matrices and the innovations covariance.

After the AR and X matrices have been obtained, the MA matrix estimates are updated by using an estimator expression that is derived from (15) for $i=1,\cdots,nc$, after replacing the theoretical impulse response samples by their corresponding estimates:

$$\sum_{j=0}^{i}\hat{\mathbf{C}}(j)\cdot\hat{\mathbf{H}}_y(i-j) = \hat{\mathbf{A}}(i) \qquad i=0,1,2,\cdots,nc \tag{32}$$

in which $\hat{\mathbf{A}}(i)\triangleq\mathbf{0}$ for $i>na$.

After the AR, MA, and X matrices have been all obtained, the prediction error sequence $\{\mathbf{e}[t/\hat{\boldsymbol{\theta}}_1]\}$ with:

$$\boldsymbol{\theta}_1 \triangleq \text{col}\Big[\, \mathbf{A}(1)\, \mathbf{A}(2) \cdots \mathbf{A}(na)\, \mathbf{B}(1) \cdots \mathbf{B}(nb)\, \mathbf{C}(1) \ldots \mathbf{C}(nc) \,\Big] \tag{33}$$

is estimated by using the model expression (5) in which the estimates $\hat{\mathbf{A}}(B)$, $\hat{\mathbf{B}}(B)$ and $\hat{\mathbf{C}}(B)$ have replaced $\mathbf{A}(B,\boldsymbol{\theta}), \mathbf{B}(B,\boldsymbol{\theta})$, and $\mathbf{C}(B,\boldsymbol{\theta})$, respectively, and the innovations covariance is finally estimated as:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{e}[t/\hat{\boldsymbol{\theta}}_1] \cdot \mathbf{e}^T[t/\hat{\boldsymbol{\theta}}_1] \tag{34}$$

□

Remarks:

(a). The order p of the ARX model of Eq.(10) needs to be selected such that:

$$r \;\leq\; p \;<<\; N$$

An appropriate value of p may be selected based on statistical order selection criteria, such as the BIC. As it has been however observed, the specific value of p is not of critical importance as long as a value that is sufficiently higher than the model order is used (also see Section 3 of Part II).

(b). Stages 3 and 4 of the estimation algorithm may be iterated until convergence in the parameter vector $\boldsymbol{\theta}_1$ is achieved. Although the improvement thus obtained is typically not very significant, this iterative procedure may be used as a means of refining the ARMAX process parameter estimates (also see Section 3 of Part II).

4. THE GUARANTEED-STABILITY FORM

A very significant problem that is inherent in most ARMAX parameter estimation methods is that of algorithmic instability. Algorithmic instability may occur in connection with all types of processes, but it almost always occurs in cases where ARMAX processes with MA roots close to the unit circle are to be modeled, and this inevitably prevents the completion of the estimation procedure. Although, in principle, the estimator stability can be monitored and (typically ad hoc) stabilization procedures applied, the implementation of "continuous" stability monitoring (for instance in every iteration of the Maximum Likelihood method) and stabilization procedures is cumbersome, and very significantly increases the method's computational complexity. In the multivariate case, in particular, the estimation methods are known to be much more prone to instability occurence problems, while stability monitoring and stabilization procedures become much more complicated and computationally expensive.

As a consequence, estimation methods characterized by mathematically guaranteed stability, that completely overcome the problem and eliminate the need for stability monitoring and stabilization, are highly desirable. In the present section this problem is addressed, and a guaranteed-stability form of the proposed linear multi-stage method introduced.

Evidently, within the context of the proposed linear multi-stage method, instabilities may occur only in cases where the initial MA polynomial matrix estimate obtained in Stage 2 is characterized by roots inside or on the unit circle, so that the recursive filtering operations (22)-(24) become unstable. In order to overcome this difficulty Stage 2 of the algorithm needs to be appropriately modified, and an estimator that is bound to yield strictly minimum phase (i.e. Hurwitz) initial MA polynomial matrix estimates constructed.

Towards this end consider the following expression that may be derived from (7) by using the causality property of the multivariate ARMAX process S:

$$\mathbf{H}_y^\circ(i) + \sum_{j=1}^{nc} \mathbf{C}^\circ(j) \cdot \mathbf{H}_y^\circ(i-j) = \mathbf{0} \qquad i \notin [0, \max(na, nc)] \tag{35}$$

in which $\mathbf{C}^\circ(0) \triangleq \mathbf{I}_s$. By approximating the impulse response samples $\mathbf{H}_y^\circ(i)$ by zero outside the interval $[\,max(na,nb), p]$, that is:

<u>*Approximation 1*</u> : $\mathbf{H}_y^\circ(i) \approx \mathbf{0}$ for $i < l \triangleq \max(na, nc) + 1$ and $i > p$

or, alternatively, by approximating $\mathbf{H}_y^\circ(i)$ by zero for $i > p$ and neglecting the values of the AR matrices, that is:

<u>*Approximation 2*</u> : $\mathbf{H}_y^\circ(i) \approx \mathbf{0}$ for $i > p$ and $\mathbf{A}^\circ(j) \approx \mathbf{0}$ for $j = 0, 1, \cdots, na$.

the range of validity of (35) can be extended to all values of i. Furthermore, by premultiplying the transpose of (35) by $\mathbf{H}_y^\circ(i-k)$ and taking summation over $-\infty < i < \infty$, the following expression is obtained:

$$\mathbf{R}_H^\circ(k, 0) + \sum_{j=1}^{nc} \mathbf{R}_H^\circ(k, j) \cdot \mathbf{C}^{\circ T}(j) \approx \mathbf{0} \tag{36}$$

in which $\mathbf{R}_H^\circ(k,j)$ represents the theoretical autocorrelation of the impulse response sequence $\{\mathbf{H}_y^\circ(i)\}$ defined as:

$$\mathbf{R}_H^\circ(k, j) \triangleq \sum_{i=-\infty}^{\infty} \mathbf{H}_y^\circ(i-k) \cdot \mathbf{H}_y^{\circ T}(i-j) \tag{37}$$

In light of Approximation 1 or 2 and the definition of $\mathbf{R}_H^\circ(k,j)$, the error involved in (36) is a decreasing function of $p - max(na, nc)$, and the approximation reasonable for a relatively large value of p. By replacing the theoretical autocorrelation $\mathbf{R}_H^\circ(k,j)$ by its biased estimate $\mathbf{R}_H(k,j)$ defined as follows for each type of approximation:

Approximation 1:

$$\mathbf{R}_H(k,j) = \begin{cases} \sum_{i=l}^{p-(k-j)} \hat{\mathbf{H}}_y(i)\cdot\hat{\mathbf{H}}_y^T(i+k-j) \triangleq \mathbf{R}_H(k-j) & (k \geq j) \\ \sum_{i=l}^{p-(j-k)} \hat{\mathbf{H}}_y(i+j-k)\cdot\hat{\mathbf{H}}_y^T(i) \triangleq \mathbf{R}_H(k-j) & (k \leq j) \end{cases} \tag{38}$$

Approximation 2:

$$\mathbf{R}_H(k,j) = \begin{cases} \sum_{i=0}^{p-(k-j)} \hat{\mathbf{H}}_y(i)\cdot\hat{\mathbf{H}}_y^T(i+k-j) \triangleq \mathbf{R}_H(k-j) & (k \geq j) \\ \sum_{i=0}^{p-(j-k)} \hat{\mathbf{H}}_y(i+j-k)\cdot\hat{\mathbf{H}}_y^T(i) \triangleq \mathbf{R}_H(k-j) & (k \leq j) \end{cases} \tag{39}$$

where $\{\hat{\mathbf{H}}_y(i)\}$ represents the impulse response estimates obtained in Stage 1 of the algorithm, and collecting the nc expressions (corresponding to $k = 1, 2, \cdots, nc$) from (36), the following estimator expression may be formulated for the MA matrices:

$$\begin{bmatrix} \mathbf{R}_H(0) & \mathbf{R}_H(-1) & \cdots & \mathbf{R}_H(1-nc) \\ \mathbf{R}_H(1) & \mathbf{R}_H(0) & \cdots & \mathbf{R}_H(2-nc) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_H(nc-1) & \mathbf{R}_H(nc-2) & \cdots & \mathbf{R}_H(0) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{C}}^T(1) \\ \hat{\mathbf{C}}^T(2) \\ \vdots \\ \hat{\mathbf{C}}^T(nc) \end{bmatrix} = - \begin{bmatrix} \mathbf{R}_H(1) \\ \mathbf{R}_H(2) \\ \vdots \\ \mathbf{R}_H(nc) \end{bmatrix} \tag{40}$$

in which the leftmost matrix is block Toeplitz and Hermitian [since $\mathbf{R}_H(-k) = \mathbf{R}_H^T(k)$]. This estimator expression is characterized by a special structure that closely resembles that of the Yule-Walker equations associated with a multivariate AR model (Kay, 1988), and therefore yields a positive definite, and hence invertible, block system of equations, and an estimated polynomial matrix that is strictly minimum phase. As a consequence the MA polynomial matrix $\hat{\mathbf{C}}^T(B)$ thus estimated will always be strictly minimum phase, and since $\det|\hat{\mathbf{C}}| = \det|\hat{\mathbf{C}}^T|$, the initial estimate $\hat{\mathbf{C}}(B)$ of the MA polynomial matrix obtained from (40) will be strictly minimum phase in both approximation cases.

This form of the estimator completely overcomes the algorithmic instability problem, and hence eliminates the need for stability checks and subsequent stabilization procedures, while making the proposed method capable of estimating all types of multivariate ARMAX processes, including those characterized by MA roots close to the unit circle.

5. CONCLUSION

In this paper the main and guaranteed-stability forms of a novel linear multi-stage method for the estimation of parameters in multivariate ARMAX processes represented in the fully-parametrized pseudo-canonical form were introduced. The main form of the proposed method is based on the replacement of the original and nonquadratic Prediction Error (PE) problem by an appropriate sequence of simpler quadratic PE and/or linear subproblems, for which uniquely-determined closed-form solutions may be obtained. This is achieved by using basic ARMAX process properties, a truncated ARX representation, and appropriate filtering operations. The method is thus based on exclusively linear techniques, and by offering low computational complexity and uniquely-determined estimates, circumventing local extrema/wrong convergence problems, and requiring no initial guess parameter values, it overcomes some of the major drawbacks that render multivariate identification unrealistic in many practical applications. In addition, a guaranteed-stability form of the method that overcomes the severe instability problems of alternative approaches and allows for the identification of all types of ARMAX processes, including those characterized by MA roots close to the unit circle, was developed.

In the second part of this paper (Fassois and Lee, 1990b) the very important problem of unknown model structure will be addressed, an effective structure and parameter estimation procedure that overcomes the enormous computational requirements of alternative approaches developed, and the performance characteristics of the linear multi-stage method examined via numerical simulations.

REFERENCES

Akaike, H., 1974, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, Vol. AC-19, pp. 716-723.

Akaike, H., 1977, "On Entropy Maximization Principle," in *Applications of Statistics*, edited by Krishnaiah, P.R., North-Holland Publishing Company, Amsterdam, pp. 27-41.

Brockwell, P.J., and Davis, R.A., 1987, *Time Series: Theory and Methods*, Springer-Verlag.

Caines, P.E., and Rissanen, J., 1974, "Maximum Likelihood Estimation of Parameters in Multivariate Gaussian Stochastic Processes", *IEEE Transactions on Information Theory*, Vol. IT-20, pp. 102-104.

Deistler, M., 1983, "The Properties of the Parameterization of ARMAX Systems and their Relevance for Structural Estimation and Dynamic Specifiction," *Econometrica*, Vol. 51, No. 4, pp. 1187-1207.

Durbin, J., 1960, "Fitting of Time Series Models," *Rev. Inst. Stat.*, Vol. 28, pp. 233-244.

El-Sherief, H., 1982, "Recent Advances in Modeling and Identification of Stochastic Multivariable Systems and Their Applications," *6th IFAC Symposium on Identification and System Parameter Estimation*, Washington D.C., pp. 367-372.

Fassois, S. D., and Lee, J. E., 1990a, "Suboptimum Maximum Likelihood Identification of ARMAX Processes," *ASME Journal of Dynamic Systems, Measurement, and Control*, in press. Also ASME paper 88-WA/DSC-32.

Fassois, S.D., and Lee, J.E., 1990b, "A Linear Multi-Stage Method for Multivariate ARMAX Process Identification - Part II: Effective Structure/Parameter Estimation and Performance Evaluation", submitted to the *ASME Journal of Dynamic Systems, Measurement, and Control.*

Gevers, M., and Wertz, V., 1987, "Techniques for the Selection of Identifiable Parametrization for Multivariable Linear Systems," in *Control and Dynamic Systems: Advances in Theory and Applications*, edited by Leondes, C.T., Vol. 26, pp. 35-86.

Goodwin, G. C., Payne, R. L., and Kabaila, P., 1978, "On Canonical Forms and Algorithms for Multivariable Time Series Analysis," *IFAC Symposium on Identification and System Parameter Estimation*, North-Holland Publishing Company, pp. 1965-1973.

Hannan, E. J., 1976, "The Identification and Parametrization of ARMAX and State Space Forms," *Econometrica*, Vol. 44, pp. 713-723.

Hannan, E. J., and Kavalieris, L., 1984, "Multivariable Linear Time Series Models," *Advances in Applied Probability*, Vol. 16, pp. 492-561.

Jakeman, A., and Young, P., 1979, "Refined Instrumental Variable Methods of Recursive Time-Series Analysis - Part II: Multivariable Systems," *International Journal of Control*, Vol. 29, No. 4, pp. 621-644.

Jones, R.H., 1985, "Fitting Multivariable Models to Unequally Spaced Data", in *Time Series*

Analysis of Irregularly Observed Data, edited by Perzen, E., Springer-Verlag Notes in Statistics, Vol. 25, pp. 158-188.

Kalman, R.E., 1974, "Algebraic Geometric Description of the Class of Linear Systems of Constant Dimension", *8th Annual Princeton Conference on Information Sciences and Systems.*

Kashyap, R. L., and Nasburg, R. E., 1974, "Parameter Estimation in Multivariate Stochastic Difference Equations," *IEEE Transactions in Automatic Control*, Vol. AC-19, pp. 784-797.

Kashyap, R. L., and Rao, A. R., 1976, *Dynamic Stochastic Models from Empirical Data*, Academic Press.

Kay, S. M., 1988, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall.

Keviczky, L. K., and Bányász, Cs. M, 1978, "Some New Results on Multiple Input Multiple Output Identification Methods," *IFAC Symposium on Identification and System Parameter Estimation*, North-Holland Publishing Company, pp. 1989-2001.

Keviczky, L.K., Bokor, J., and Sandor, V., 1987, "Strong Consistency of ML Estimators Using Partial Fraction and Elementary Subsystem Representation of Multivariable Systems", *IEEE Transactions on Automatic Control*, Vol. AC-32, pp. 867-876.

Ljung, L., 1987, *System Identification: Theory for the User*, Prentice-Hall.

Magnus, J.R., and Neudecker, H., 1988, *Matrix Differential Calculus*, John Wiley.

Pandit, S. M., and Wu, S. M., 1983, *Time Series and System Analysis with Applications*, John-Wiley.

Priestley, M.B., 1981, *Spectral Analysis and Time Series*, Academic Press.

Stoica, P., and Söderström, T., 1982, "Identification of Multivariable Systems Using Instrumental Variable Methods," *6th IFAC Symposium on Identification and System Parameter Estimation*, Washington D.C., pp. 1073-1078.

Stoica, P., and Söderström, T., 1983, "Optimal Instrumental Variable Methods for Identification of Multivariable Linear Systems," *Automatica*, Vol. 19, No. 4, pp. 425-429.

Tse, E., and Weinert, H.L., 1975, "Structure Determination and Parameter Identification for Multivariable Stochastic Linear Systems", *IEEE Transactions on Automatic Control*, Vol. AC-20, pp. 603-613.

LIST OF SYMBOLS

$\mathbf{A}(B)$	$[\,s \times s\,]$	autoregressive (AR) polynomial matrix
$\mathbf{B}(B)$	$[\,s \times m\,]$	exogenous (X) polynomial matrix
$\mathbf{C}(B)$	$[\,s \times s\,]$	moving average (MA) polynomial matrix
$\mathbf{e}[t/\theta]$	$[\,s \times 1\,]$	one-step-ahead prediction error of the output
$\mathbf{h}'_j$	$[\,(s+m) \times 1\,]$	1-st column of $[-\mathbf{H}_y(j)\,\mathbf{H}_x(j)]^T$
$\mathbf{h}_1(j)$	$[\,p(s+m) \times 1\,]$	j-th column of the matrix $\mathcal{H}_1^T$
$\mathbf{h}_2(j)$	$[\,p(s+m) \times 1\,]$	j-th column of the matrix $\mathcal{H}_2^T$
$\boldsymbol{h}_1$	$[\,sp(s+m) \times 1\,]$	column vector containing in order the columns of $\mathcal{H}_1^T$
$\boldsymbol{h}_2$	$[\,sp(s+m) \times 1\,]$	column vector containing in order the columns of $\mathcal{H}_2^T$
$\mathbf{H}_x(j)$	$[\,s \times m\,]$	j-th sample of an impulse response function
$\mathbf{H}_y(j)$	$[\,s \times s\,]$	j-th sample of an impulse response function
$\mathcal{H}_1$	$[\,s \times p(s+m)\,]$	matrix containing the $\mathbf{H}_x(j)$ and $\mathbf{H}_y(j)$ sequences
$\mathcal{H}_2$	$[\,s \times p(s+m)\,]$	matrix containing the $\mathbf{H}_x(j)$ and $\mathbf{H}_y(j)$ sequences
$r_{ij}(k)$		i,j-th element of the covariance matrix $\mathbf{R}_{zz}(k)$
$\tilde{\mathbf{r}}_{ij}(k)$	$[\,s \times s\,]$	filtered version of $r_{ij}(k)$
$\mathbf{r}_{zy_i}(k)$	$[\,(s+m) \times 1\,]$	cross covariance between $\mathbf{z}[t]$ and $y_i[t]$
$\tilde{\mathbf{r}}_{zy}(k)$	$[\,s(s+m) \times 1\,]$	cross covariance between $\mathbf{Z}_F[t]$ and $\mathbf{y}_F[t]$
$\mathbf{R}_H(k)$	$[\,s \times s\,]$	autocovariance of the impulse response function $\{\mathbf{H}_y(j)\}$
$\mathbf{R}_{zy}(k)$	$[\,(s+m) \times s\,]$	cross covariance between $\mathbf{z}[t]$ and $\mathbf{y}[t]$
$\mathbf{R}_{zz}(k)$	$[\,(s+m) \times (s+m)\,]$	autocovariance of $\mathbf{z}[t]$
$\tilde{\mathbf{R}}_{zy}(k)$	$[\,s(s+m) \times s^2\,]$	filtered cross covariance between $\mathbf{z}[t]$ and $\mathbf{y}[t]$
$\tilde{\mathbf{R}}_{zz}(k)$	$[\,s(s+m) \times s(s+m)\,]$	filtered autocovariance of $\mathbf{z}[t]$
$\mathbf{u}_1[t]$	$[\,p(s+m) \times 1\,]$	vector containing samples of the input/output sequences
$\mathbf{u}_2[t]$	$[\,p(s+m) \times 1\,]$	vector containing samples of the input/output sequences
$\mathbf{U}_{1F}[t]$	$[\,s \times s(s \cdot na + m \cdot nb)\,]$	matrix containing samples of the filtered input/output sequence
$\mathbf{U}_{2F}[t]$	$[\,s \times sn(s+m)\,]$	matrix containing samples of the filtered input/output sequence
$\mathbf{v}_j$	$[\,s(s+m) \times 1\,]$	column vector containing in order the columns of $[-\mathbf{A}(j)\,\mathbf{B}(j)]$
$\mathbf{w}[t]$	$[\,s \times 1\,]$	innovations sequence
$\mathbf{x}[t]$	$[\,m \times 1\,]$	input sequence
$\mathbf{X}_F[t]$	$[\,s \times m \cdot s\,]$	matrix containing samples of the filtered
$\mathbf{y}[t]$	$[\,s \times 1\,]$	noise-corrupted output sequence
$\mathbf{y}_F[t]$	$[\,s \times 1\,]$	filtered output sequence
$\mathbf{Y}_F[t]$	$[\,s \times s^2\,]$	filtered output matrix sequence
$\mathbf{z}[t]$	$[\,(s+m) \times 1\,]$	joint input/output vector
$\mathbf{Z}_F[t]$	$[\,s \times s(s+m)\,]$	filtered version of $\mathbf{z}[t]$

$\boldsymbol{\theta}$		set of ARMAX process parameters
$\boldsymbol{\theta}_1$	$[s(s \cdot na + m \cdot nb + s \cdot nc) \times 1]$	AR, MA, and X parameter vector
$\boldsymbol{\theta}_2$	$[s(s \cdot na + m \cdot nb) \times 1]$	AR and X parameter vector
$\boldsymbol{\theta}_3$	$[sn(s + m) \times 1]$	AR and X parameter vector
$\boldsymbol{\Sigma}$	$[\, s \times s \,]$	innovations covariance