Copyright © 1966, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

IDENTIFICATION OF LINEAR DISCRETE TIME SYSTEMS USING THE INSTRUMENTAL VARIABLE METHOD

by

K. Y. Wong and E. Polak

Memorandum No. ERL-M187
9 December 1966

ELECTRONICS RESEARCH LABORATORY
College of Engineering
University of California, Berkeley
94720

Manuscript submitted: 23 September 1966.

The research reported herein was supported wholly by the National Aeronautics and Space Administration under Grant NsG-354, Supplement 2 and 3.

ABSTRACT

This paper explores the possibility of using the instrumental variable method to estimate the parameters of linear time invariant discrete time system. The existence of optimal estimates is established, methods for their approximate computation are given and an on-line identification scheme based on recursive computation is proposed. Experimental results are included.

INTRODUCTION

The problem of identification has received a great deal of attention in the literature of control theory. However, it is not peculiar to control and is, in fact, a problem common to many branches of science and engineering. This paper explores the potential of the "Instrument Variable Method" first introduced in economics by Reiersol [1] in 1941, as a tool for identifying the parameters of a dynamical system described by a linear difference equation. It will be assumed that the system input, over which one is given no control, and the additive noise corrupted output can be observed.

The instrumental variable method has two salient features. The first is that given a system of linear equations involving the unknown parameters, observations, and noise terms, with the number of equations N greater than the number of unknown parameters p, one reduces this system to an invertible square array of p equations. The second feature is that this is done in such a way that the contribution of the noise terms to the square array goes to zero in probability as N, the dimension of the original array, goes to infinity, i.e. the resulting estimate of the unknown parameters is consistent.

In the context of system identification, one can raise a series of questions such as some of the following. "Since the reduction to a square array can be done in more than one way, are there optimal ways of doing so?" "Are the optimal estimates computable?" "How can one simplify some of the resulting, very difficult computations, such as inversion of large matrices?" "Can this method be adapted for iterative on-line identification?"

All of these questions are answered here. And from the computations described in Section 8, it is clear that there is good experimental evidence supporting the authors' conviction that the instrumental variable method can be a most valuable tool for identifying linear discrete time systems.

1. STATEMENT OF THE PROBLEM

Suppose we are given a dynamical system described by a scalar difference equation of order (p-1) of the form

$$\sum_{j=1}^{p} a_j y_{i-p+j} = \sum_{j=1}^{p} b_j u_{i-p+j}, \quad i=1,2,\ldots \tag{1}$$

where y_i is the system output and u_i is the system input at time $i=1,2,\ldots$ We assume that Eq. (1) is normalized so that $b_p=1$.

We are required to identify the system parameters $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots b_{p-1}$ on the basis of the following $N+p-1$ observations, where $N \geq p$. For $i=1,2,\ldots, N+p-1$, we are allowed to observe the exact system inputs u_i and the noise-corrupted system output x_i defined by

$$x_i = y_i + d_i, \quad i=1,2,\ldots \tag{2}$$

where $d_i \epsilon\ d \triangleq \{d_i\}_{i=1}^{\infty}$, a sample from a zero mean, stationary noise process whose convariance function $r_d(k)$ tends to zero at a rate faster than 1/k, as $k \to \infty$. For $i=1, 2,\ldots, N+p-1$, the inputs u_i are elements of a sequence $u \triangleq \{u_i\}_{i=1}^{\infty}$ which is either deterministic or else a sample of a stationary random process. Furthermore, we assume that the noise process and the input process are statistically independent.

In keeping with the notation for the input and the noise, we define the sequences $y \triangleq \{y_i\}_{i=1}^{\infty}$, whose elements are the system outputs corresponding to the input sequence u, and $x=y+d=\{x_i\}_{i=1}^{\infty}$.

We shall always assume that the eigenvalues of the system (1), i.e., the roots of the equation

$$\sum_{j=1}^{p} a_j \lambda^{j-1} = 0,$$

lie inside the unit circle.

Our discussion becomes considerably simplified if we restrict ourselves to the particular case of (1) where $b_1=b_2=\ldots=b_{p-1}=0$, i.e., to the system

$$a_p y_i + a_{p-1} y_{i-1} + \ldots + a_1 y_{i-p+1} = u_i, \qquad i=1,2,3\ldots \tag{3}$$

Most of our results for (3) are readily applied to the system (1) by rewriting (1) into the form

$$a_p y_i + a_{p-1} y_{i-1} + \ldots + a_1 y_{i-p+1} - b_{p-1} u_{i-1} \ldots - b_1 u_{i-p+1} = u_i, \tag{4}$$

and then making appropriate substitutions for the matrix elements which appear in our computations. We shall always indicate when a particular result is valid for (3) only.

2. THE INSTRUMENTAL VARIABLE ESTIMATE

We now introduce functions $V_N(\cdot)$, $N=p, p+1, \ldots,$ which map scalar sequences, such as $x=\{x_i\}_{i=1}^{\infty}$, into $N\times p$ matrices, as follows:

$$V_N(x) = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \\ x_2 & x_3 & \cdots & x_{p+1} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ x_N & x_{N+1} & \cdots & x_{N+p-1} \end{bmatrix} \tag{5}$$

Let $V^j_N(\cdot)$ denote the jth column of the matrix $V_N(\cdot)$, with $j=1,2,\ldots,p$, and let $a=(a_1, a_2, \ldots, a_p)$. With i taking the values of $1,2,\ldots,N+p-1$, $N\geq p$, we obtain from (3) (N+p-1) equations which can be written in matrix form as follows:

$$V_N(y)a = V^p_N(u). \tag{6}$$

Since $x=y+d$, we obtain from (6)

$$V_N(x)a=V_N^p(u)+V_N(d)a\,. \tag{7}$$

Let Z_N be any $N\times p$ matrix such that $Z_N^T\, V_N(x)$ is nonsingular, then from (7),

$$a=\left(Z_N^T\, V_N(x)\right)^{-1} Z_N^T(u)+\left(Z_N^T\, V_N(x)\right)^{-1} Z_N\, V_N(d)a\,. \tag{8}$$

<u>Definition:</u> <u>Given an input sequence</u> u for (3), we shall call a sequence of $N\times p$ matrices $\{Z_N\}_{N=p}^{\infty}$ <u>an instrumental matrix sequence</u> if

$$\begin{aligned}
&\text{(i)} \qquad p\lim_{N\to\infty} \frac{1}{N}\, Z_N^T\, V_N(d)=0\\
&\text{(ii)} \qquad p\lim_{N\to\infty} \frac{1}{N}\, Z_N^T\, V_N(y) \text{ is nonsingular.}
\end{aligned} \tag{9}$$

<u>Remark:</u> When the elements of the matrices $Z_N^T\, V_N(y)$ are deterministic, the limit in (ii) is taken in the ordinary sense of convergence.

<u>Definition:</u> We define $\hat{a}_N$, the <u>instrumental variable estimate</u> of the unknown parameter a, computed on the basis of $N+p-1$ observations, to be

$$\hat{a}_N=\left(Z_N^T\, V_N(x)\right)^{-1} Z_N^T\, V_N^p(u)\,, \qquad N=p,\ p+1,\ p+2,\ldots \tag{10}$$

<u>Theorem 1:</u> For $N=p,\ p+1,\ldots,$ the instrumental variable estimate $\hat{a}_N$ is a consistent estimate of the unknown parameter a.

<u>Proof:</u> By definition, $\hat{a}_N$ is a consistent estimate of a if

$$p\lim_{N\to\infty} \hat{a}_N=a\,. \tag{11}$$

$$p\lim_{N\to\infty} \hat{a}_N=a-p\lim_{N\to\infty}\left(Z_N^T\, V_N(x)\right)^{-1} Z_N^T\, V_N(d)\, a\,. \tag{12}$$

Applying Slutsky's theorem [2], we get

$$\underset{N\to\infty}{p\lim}\ \hat{a}_N = a - \underset{N\to\infty}{p\lim}\left(\frac{1}{N} Z_N^T V_N(y) + \frac{1}{N} Z_N^T V_N(d)\right)^{-1} \times$$

$$\underset{N\to\infty}{p\lim}\ \frac{1}{N} Z_N^T V_N(d) a = a, \tag{13}$$

since by assumption $\underset{N\to\infty}{p\lim}\ \frac{1}{N} Z_N^T V_N(d) = 0$ and $\underset{N\to\infty}{p\lim}\ \frac{1}{N} Z_N^T V_N(y)$ is a nonsingular matrix.

<u>Remark</u>: Since $\hat{a}_N$ is a consistent estimate of a, it is clear that it is asymptotically unbiased.* However, for any finite value of N, $\hat{a}_N$ may be considerably biased. Consequently, after considering in the next section the question of existence of instrument matrix sequence, we shall proceed in section 4 to determine whether it is possible to choose an <u>optimal</u> instrumental matrix sequence.

3. PROPERTIES OF INSTRUMENTAL MATRIX SEQUENCES AND CORRESPONDING ESTIMATES

By definition, an instrumental matrix sequence $\{Z_N\}_{N=p}^{\infty}$ must satisfy conditions (9) (i) and (ii). Condition (i) depends on the noise process. Condition (ii) depends on the input sequence u which so far has not been particularly restricted. We therefore have to determine whether instrumental matrix sequences exist for arbitrary inputs, since this determines our ability to identify the system (3).

<u>Theorem 2</u>: Let $\{Z_N\}_{N=p}^{\infty}$ be any sequence of $N\times p$ matrices whose elements are uniformly bounded and are statistically independent of the elements of the matrices $V_N(d)$,

*The least squares estimate of a, which is obtained by putting $Z_N = V_N(x)$ in (10) can be shown to be non-consistent and therefore asymptotically inferior to an instrumental variable estimate.

then $p \lim_{N\to\infty} \frac{1}{N} Z_N^T V_N(d) = 0$.

<u>Proof</u>: Let $G_N = Z_N^T V_N(d)$. We shall denote the $\underline{i}\ \underline{j}$th element of the matrix G_N by g_{ij}^N and the $\underline{i}\ \underline{j}$th element of the matrix Z_N by z_{ij}^N.

Hence,

$$g_{ij}^N = \sum_{k=1}^{N} z_{ki}^N d_{j+k-1}, \qquad i, j = 1, 2, \ldots, p. \tag{14}$$

Since the elements of z_{ij}^N are bounded and are statistically independent of the elements of the sequence d, and since $Ed_i = 0$ for $i = 1, 2, \ldots,$

we have

$$E \frac{1}{N} g_{ij}^N = 0 \quad \text{for } i, j = 1, 2, \ldots, p \quad \text{and} \quad N = p, p+1, \ldots \tag{15}$$

Now, the variance of $\frac{1}{N} g_{ij}^N$ is given by

$$E\left(\frac{1}{N} g_{ij}^N\right)^2 = \frac{1}{N^2} E \left(\sum_{k=1}^{N} z_{ki}^N d_{j+k-1}\right)^2. \tag{16}$$

Since by assumption the z_{ij}^N are bounded, let z_m be such that $|z_{ij}^N| \leq z_m$ for $i, j = 1, 2, \ldots, p$ and $N = p, p+1, \ldots$ Then (16) yields

$$E\left(\frac{1}{N} g_{ij}^N\right)^2 \leq \frac{z_m^2}{N^2} \sum_{k=1}^{N} \sum_{m=1}^{N} \left| Ed_{j+k-1} d_{j+m-1} \right| = \frac{z_m^2}{N^2} \sum_{k=1}^{N} \sum_{m=1}^{N} \left| r_d(k-m) \right|$$

$$= \frac{z_n^2}{N} \left[r_d(0) + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \left| r_d(k) \right| \right] \tag{17}$$

where $r_d(j)$ is the correlation function of the noise process, which was assumed to satisfy $\lim_{N\to\infty} Nr_d(N)=0$. Hence from (17),

$$\lim_{N\to\infty} E\frac{1}{N}g_{ij}^{N^2} = 0 \tag{18}$$

Invoking Tchebychev's inequality, we deduce from (18) that

$$p\lim_{N\to\infty} g_{ij}^{N} = 0$$

<u>Theorem 3:</u>* Suppose that the input sequence $u=\{u_i\}_{i=1}^{\infty}$ is deterministic and bounded, then <u>a ncessary condition</u> for the existence of an instrumental matrix sequence $\{Z_N\}_{N=p}^{\infty}$ is that the $p\times p$ matrix $\lim_{N\to\infty}\frac{1}{N}V_N^T(u)\,V_N(u)$ be nonsingular.

Proof: Let $S_N(\cdot)$ be a map which takes vectors in R^k, $k=1,2,\ldots,$ into $N\times N$ matrices, defined as follows. For $a=(a_1,a_2,\ldots,a_p)$,

$$S_N(a)=\begin{bmatrix} a_p & 0 & 0 & \cdots & 0 & 0 \\ a_{p-1} & a_p & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ a_2 & a_3 & \cdots & a_p & 0 & 0 \\ a_1 & a_2 & \cdots & a_{p-1} & & \\ 0 & a_1 & a_2 & \cdots & a_p & 0 \\ 0 & 0 & a_1 & a_2 & a_{p-1} & a_p \end{bmatrix} \tag{19}$$

*This theorem is only valid for systems of the form in (3).

Then, the input-output relation (3) yields

$$S_N(a)\ [V_N(y)+Y] = V_N(u), \qquad N=1,2,\ldots \tag{20}$$

where Y is a $N\times p$ matrix whose last $N-p+1$ rows all are zero. Since by assumption the system (3) is strictly stable, which implies that ℓ_2 input sequences results in ℓ_2 output sequences, we conclude that $\|S_N(a)^{-1}\|$ (The Enclidean norm of $S_N(a)^{-1}$) is bounded, i.e., there exists a constant $M < \infty$ such that

$$\text{for all } N=1,2,\ldots, \quad \|S_N(a)^{-1}\| \le M.$$

Now, suppose that $\lim\limits_{N\to\infty} \frac{1}{N} V_N^T(u)\, V_N(u)$ is a singular matrix. Then there exists a vector $\xi \in R^p$ such that

$$\lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(u)\ \xi \right\| = 0 \tag{21}$$

But from (20), for $N\to\infty$,

$$\lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(y)\ \xi \right\| = \lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} S_N(a)^{-1} V_N(u)\ \xi \right\|$$

$$\le M \lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(u)\ \xi \right\| = 0\ . \tag{22}$$

Hence $\lim\limits_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(y)\ \xi \right\| = 0$ and, since by assumption $|z_{ij}^N| \le z_m$, for $i=1,2,\ldots,N$, $j=1,2,\ldots,p$, we conclude that $\lim\limits_{N\to\infty} \frac{1}{N} Z_N^T V_N(y)\ \xi = 0$, i.e., the matrix $\lim\limits_{N\to\infty} \frac{1}{N} Z_N^T V_N(y)$ is singular. Q.E.D.

<u>Example:</u> We now give an example of an instrumental matrix sequence $\{Z_N\}_{N=p}^{\infty}$ which satisfies all our assumptions. Suppose that the input sequence u is periodic with period p and that the matrix $V_p(u)$ is nonsingular. (This last condition is very easily satisfied). Let y' be the decaying part and let y'' be the steady state part of the output sequence y,

i.e., $y=y'+y''$. Then y'' is also periodic with period p and it is easily shown that the matrix $V_p(y'')$ is also nonsingular. We claim that $\{V_N(u)\}_p^\infty$ is an instrumental matrix sequence. It is clear from Theorem 2 that $p \lim_{N\to\infty} \frac{1}{N} V_N^T(u)\ V_N(d)=0$. Hence we only need to show that $\lim_{N\to\infty} \frac{1}{N} V_N^T(u)\ V_N(y) = \lim_{N\to\infty} \frac{1}{N} V_N^T(u)\ V_N(y'')$ is nonsingular. For $i=1, 2, 3, \ldots$, let $u(i) = (u_i, u_{i+1}, \ldots u_{i+p-1})$, and $y''(i)=(y''_i, y''_{i+1}, \ldots, y''_{i+p-1})$, obviously, $u(i)=u(i+p)$, $y''(i)=y''(i+p)$. Then, with $N=mp+n$,

$$\frac{1}{N} V_N^T(u)\ V_N(y'') = \frac{1}{mp+n} \sum_{i=1}^{mp+n} u(i) >< y''(i)$$

$$= \frac{m}{mp+n}\left(\sum_{i=1}^{p} u(i) >< y''(i) + \sum_{i=1}^{n} u(i) >< y''(i) \right)^{*} \tag{23}$$

Consequently,

$$\lim_{N\to\infty} V_N^T(u)\ V_N(y'') = \frac{1}{p} \sum_{i=1}^{p} u(i) >< y''(i) = \frac{1}{p} V_p^T(u)\ V_p(y''),$$

which is obviously nonsingular, Hence, $\{V_N(u)\}_{N=p}^{\infty}$ is an instrumental matrix sequence.

We shall now show that under mild assumptions the sequences of estimates $\{\hat{a}_N\}$ converges to the parameter vector a at a rate proportional to $1/N$.

Theorem 4: Suppose that the input sequence $u=(u_1, u_2, \ldots)$ is deterministic and bounded, and that $\{Z_N\}_{N=p}^{\infty}$ is a deterministic sequence of instrumental matrices, with bounded components, i.e., $|z_{ij}^N| \le z_m$ for $i=1, 2, \ldots, N$, $j=1, 2, \ldots, p$, $N=., 2, \ldots$ Also, suppose

*Given $\alpha=(\alpha_1, \alpha_2, \ldots, \alpha_s)$, $\beta=(\beta_1, \beta_2, \ldots \beta_s)$ we denote by $\alpha><\beta$ the $s\times s$ matrix whose ij^{th} element is $\alpha_i\beta_j$.

that the disturbance sequence $d=(d_1, d_2, \ldots)$ is bounded with probability 1. Then, $E\|\hat{a}_N - a\|^2$ is proportional to $1/N$ for $N\to\infty$.

<u>Proof:</u> Let $\Omega_N = E\left(V_N(d)\, a >< V_N(d)\, a\right)$ and let $C_N = E\left((\hat{a}_N - a) >< (\hat{a}_N - a)\right)$. Then from (8) and (10)

$$C_N = E\left[\left(Z_N^T V_N(x)\right)^{-1} Z_N V_N(d)\, a >< \left(Z_N^T V_N(x)\right)^{-1} Z_N V_N(d)\, a\right] \tag{24}$$

Since by assumption (i) the input sequence u, and consequently also the output sequence y, is bounded, (ii) the elements z_{ij}^N of the matrices Z_N are bounded and, (iii) the noise sequence d is bounded with probability 1, it can be shown that [3]

$$\lim_{N\to\infty} C_N = \lim_{N\to\infty} \frac{1}{N}\left(\frac{1}{N} Z_N^T V_N(y)\right)^{-1}\left(\frac{1}{N} Z_N^T \Omega_N Z_N\right)\left(\frac{1}{N} V_N(y)^T Z_N\right)^{-1} \tag{25}$$

Observe that, for <u>finite</u> N, $C_N \neq \left(Z_N^T V_N(y)\right)^{-1}\left(Z_N^T \Omega_N Z_N\right)\left(V_N(y)^T Z_N\right)^{-1}$ because $V_N(x) = V_N(y) + V_N(d)$.

Now, $\lim_{N\to\infty}\left(\frac{1}{N} Z_N^T V_N(y)\right)$ is by assumption a finite nonsingular matrix, hence, its inverse is also finite. The <u>i jth</u> element of Ω_N is $r_e(i-j) \triangleq E e_i e_j$, the correlation function of the sequence $e \triangleq V_\infty(d)a$, which is easily shown to satisfy $k\, r_e(k) \to 0$ as $k\to\infty$, since $k\, r_d(k) \to 0$ as $k\to\infty$ by assumption. Consequently, $|\Omega_N|$, the sup norm of Ω_N, for each $N=p, p+1\ldots$, is bounded by $\sum_{k=0}^{\infty} |r_e(k)|$, which is finite. Hence, since $|z_{ij}^N| \le z_m < \infty$, we conclude that $\lim_{N\to\infty}\left(\frac{1}{N} Z_N \Omega_N Z_N^T\right)$ is a finite element matrix. Now, for $j = 1, 2, \ldots, p$, let $f_j = (0, 0, \ldots, 1, 0 \ldots 0) \in R^p$ be a unit vector with a 1 in the <u>j</u>th position. Then,

$$\lim_{N\to\infty} \sum_{j=1}^{p} \langle f_j, C_N f_j \rangle = \lim_{N\to\infty} E\|\hat{a}_N - a\|^2, \tag{26}$$

and from (25), $E\|\hat{a}_N - a\|^2$ is proportional to $1/N$ for $N\to\infty$.

5. OPTIMAL INSTRUMENTAL MATRIX SEQUENCES

Our entire discussion of optimal instrumental matrix sequences will be based upon a single matric inequality, which we shall now establish.

Lemma 1. If a $n \times n$ matrix B is idempotent, i.e., symmetric and such that $B^2 = B$, then all its eigenvalues are either zero or one.

Proof: Let λ be any eigenvalue of B and ξ a corresponding eigenvector. Then $B(B\xi) = \lambda^2 \xi$. But $B(B\xi) = B\,\xi = \lambda\,\xi$ and hence $\lambda = \lambda^2$, i.e., $\lambda = 0$ or 1.

Lemma 2. If the $n \times n$ matrix B is idempotent, then

$$< \eta, B\eta > \quad \leq < \eta, \eta > \text{ for every nonzero } \eta \in R^n . \tag{27}$$

Proof: Since B is idempotent, $B = H^{-1} DH$, where D is a diagonal matrix with elements on the diagonal 0 or 1 and H is an $n \times n$ matrix such that $H^{-1} = H^T$. Hence,

$$< \eta, B\eta > = < H\eta, DH\eta > \leq < H\eta, H\eta > = < \eta, \eta > . \tag{28}$$

Lemma 3: Let A, B be symmetric $n \times n$ positive definite (> 0) matrices such that $A-B$ is positive semidefinite (≥ 0), then the matrix $(B^{-1} - A^{-1})$ is positive semidefinite (≥ 0).

Proof: Since A, B are positive definite symmetric matrices, there exists an $n \times n$ matrix Q, such that $Q^T = Q^{-1}$ and such that $Q^T AQ = I$, $Q^T BQ = D > 0$, a diagonal matrix (see Friedman p. 109, [4]). Since $A-B \geq 0$, $Q^T (A-B) Q = I - D \geq 0$, and $D^{-1/2} (I-D) D^{-1/2} = D^{-1} - I \geq 0$. But $Q^T (B^{-1} - A^{-1}) Q = D^{-1} - I$ and hence $B^{-1} - A^{-1} \geq 0$.

Theorem 5: Let Z_N, V_N be any two $N \times p$, $N \geq p$, matrices and let Q_N be any symmetric, positive definite (> 0) $N \times N$ matrix. If $(V_N^T Z_N)^{-1}$, $(V_N^T Q_N^{-1} V_N)^{-1}$ and $(Z_N^T Q_N Z_N)^{-1}$ exist, then the $p \times p$ matrix $(Z_N^T V_N)^{-1} Z_N^T Q_N Z_N (V_N^T Z_N)^{-1} - (V_N^T Q_N^{-1} V_N)^{-1}$ is positive semidefinite (≥ 0).

Proof: From Lemma 3, $(Z_N^T V_N)^{-1} Z_N^T Q_N Z_N (V_N^T Z_N)^{-1} - (V_N^T Q_N^{-1} V_N)^{-1} \geq 0$ if

$V_N^T Q_N^{-1} V_N - (V_N^T Z_N)(Z_N^T Q_N Z_N)^{-1}(Z_N^T V_N) \geq 0$. Now, since $(V_N^T Z_N)^{-1}$

exists by assumption, V_N must be a matrix of rank p, and hence, the theorem is true if $Q_N^{-1} - Z_N(Z_N^T Q_N Z_N)^{-1} Z_N^T \geq 0$, i.e., if $I - Q_N^{1/2} Z_N (Z_N^T Q_N Z_N)^{-1} Z_N^T Q_N^{1/2} \geq 0$. But the matrix $Q_N^{1/2} Z_N (Z_N^T Q_N Z_N)^{-1} Z_N^T Q_N^{1/2}$ is idempotent and hence, it follows from Lemma 2 that Theorem 5 is true. From (8) and (10), the instrumental variable estimate $\hat{a}_N$ (when it exists) can be expressed in the form

$$\hat{a}_N = a - (Z_N^T V_N(x))^{-1} Z_N^T V_N(d)\, a. \tag{29}$$

Let $e = V_\infty(d)\, a = (e_1, e_2, \ldots)$, then $V_N^1(e) = V_N(d)\, a$ and, since $p \lim_{N\to\infty} \frac{1}{N} Z_N^T V_N(d) = 0$, for N large, (29) approximately reduces to

$$\hat{a}_N = a - (Z_N^T V_N(y))^{-1} Z_N^T V_N^1(e)\ . \tag{30}$$

When our knowledge of the statistical properties of the noise process is inadequate, say we do not know its covariance function $r_d(\cdot)$, we can still optimize our estaimte $\hat{a}_N$, given by (30), on a minimax basis, as shown below.

Theorem 6: Suppose that the matrix $\lim_{N\to\infty} \frac{1}{N} V_N^T(u) V_N(u)$ is nonsingular and that $\| V_N^1(e) \| \leq 1$ with probability 1. Let Z be the class of instrumental matrices which can be used for identifying (3) with the input sequence u and let $\hat{a}_N(Z_N)$ denote the estimate resulting from the use of some instrumental matrix $Z_N \in Z$. Then the sequence $\{ V_N(y) \}$ $\{ V_N(y) \}_p^\infty$ is an instrumental matrix sequence and, for N large,

$$\| \hat{a}_N(V_N(y)) - a \| = \min_{Z_N \in Z} \ \max_{\| V_N^1(e) \| = 1} \| \hat{a}_N(Z_N) - a \|. \tag{31}$$

<u>Proof</u>: Since by assumption the input sequence u is bounded and statistically independent of the noise sequence d, and since the system (3) is strictly stable, the output sequence y is bounded and statistically independent of d. Hence from Theorem 2,

$$p \lim_{N\to\infty} \frac{1}{N} V_N^T(y)\, V_N(d) = 0.$$

Now, since $\lim_{N\to\infty} \frac{1}{N} V_N^T(u)\, V_N(u)$ is nonsingular, and (see (19)) since for N=p, p+1,... it is proved in Appendix I that $0 < \|S_N(a)\| < (\alpha p)^{1/2}$, where $\alpha = \sum_{i=1}^{p} a_i^2$, we have for every $\xi \in R^p$, $\xi \neq 0$,

$$\lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} S_N(a)\, V_N(y)\, \xi \right\| = \lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(u)\, \xi \right\| > 0 \tag{32}$$

Hence, since the elements of $V_N(y)$ are uniformly bounded in magnitude,

$$\infty > \lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(y)\, \xi \right\| \geq \frac{1}{(\alpha p)^{1/2}} \lim_{N\to\infty} \left\| \frac{1}{\sqrt{N}} V_N(u)\, \xi \right\| > 0, \tag{33}$$

and $\lim_{N\to\infty} \frac{1}{N} V_N^T(y)\, V_N(y)$ is a nonsingular matrix. Hence $\{V_N(y)\}_{N=p}^{\infty}$ is a sequence of instrumental matrices.

Now, to prove (31) we only need to show that $\|(Z_N^T Y_N)^{-1} Z_N^T\|$ assumes a minimum for $Z_N = Y_N$. Let B be any $N\times p$ matrix. Then $\|B\|^2 = \max_{\|\eta\|=1} <\eta, B^T B \eta> = \lambda_{max}$, is the maximum eigenvalue of the matrix $B^T B$. But (see Gantmacher, pp. 45-46, [5]) $\det(\lambda I - B^T B) = \lambda^{N-p} \det(\lambda I - BB^T)$ and hence $\lambda_{max} = \lambda^*_{max}$, where λ^*_{max} is the largest eigenvalue of BB^T. Thus, $\|(Z_N^T V_N)^{-1} Z_N^T\|^2 = \max_{\|\xi\|=1} <\xi, (Z_N^T V_N(y))^{-1}$ $Z_N^T Z_N (V_N^T(y) Z_N)\, \xi>$.

But, from Theorem 5, with Q=I, the identity matrix,

$$(Z_N^T V_N(y))^{-1} Z_N^T Z_N (V_N(y) Z_N^T)^{-1} - (V_N^T(y)\, V_N(y))^{-1} \geq 0 \tag{34}$$

and hence $\| (Z_N^T V_N(y))^{-1} Z_N^T \|$ is minimized by choosing $Z_N = V_N(y)$.

When we do know the covariance function $r_d(\cdot)$ of the noise process, we can take it into account in optimizing our estimate $\hat{a}_N$, to obtain an estimate, which, for N large, can be expected to be better than the one we derived by optimizing on a minimax basis.

Theorem 7: Suppose that the input sequence $u=(u_1, u_2, \ldots)$ is deterministic and bounded, and that the matrix $\lim_{N\to\infty} \frac{1}{N} V_N(u) V_N(u)$ is nonsingular. Also suppose that the disturbance sequence $d=(d_1, d_2, \ldots)$ is bounded with probability 1 and that

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{-\infty} r_e(k) e^{ik\omega} \geq f_{min} > 0$$

where

$$r_e(k) = E\, e_j e_{j+k}, \text{ with } V_\infty(d)a \triangleq e = (e_1, e_2, \ldots).$$

Let $\{Z_N\}_{N=p}^{\infty}$ be any sequence of instrumental matrices, and let $\{\hat{a}_N(Z_N)\}_{N=p}^{\infty}$ denote the corresponding sequence of estimate of the parameter a.† For N=p, p+1, ... let $Z_N^* = \Omega_N^{-1} V_N(y)$, where the N×N matrix $\Omega_N = V_N^1(e) >< V_N^1(e) \triangleq E\, V_N(d)\, a >< V_N(d)\, a$, then $\{Z_N^*\}_{N=p}^{\infty}$ is an instrumental matrix sequence and, for N large,

$$E \| \hat{a}_N(Z_N^*) - a \|^2 \leq E \| \hat{a}_N(Z_N) - a \|^2 \tag{35}$$

for all admissible instrumental matrix sequences $\{Z_N\}_p^{\infty}$.

Proof: We begin by showing that

$$\lim_{N\to\infty} \frac{1}{N} Z_N^{*T} V_N(y) = \lim_{N\to\infty} \frac{1}{N} V_N^T(y)\, \Omega_N^{-1} V_N(y)$$

exists. The ijth element of the symmetric N×N matrix Ω_N is seen to be $r_e(i-j)$

† It is understood that not all $\hat{a}_N$ may exist.

$= r_e(j-i)$, and $|\Omega_N|$, the sup norm of Ω_N, obviously satisfies

$$|\Omega_N| \leq \sum_{k=-\infty}^{+\infty} |r_e(k)| = M < \infty.$$

Thus, for all N, the sup norm of Ω_N is finite. Now, Ω_N is positive definite since for all $\xi \in R^N$ with $\|\xi\| = 1$,

$$< \xi , \Omega_N \xi > = \sum_{i=1}^{N} \sum_{j=1}^{N} \xi_i r_e (i-j) = \int_{-\pi}^{+\pi} f(\omega) \; | \sum_{k=1}^{N} \xi_k e^{-ik\omega} |^2 d\omega \quad (36)$$

$$\geq f_{min} \int_{-\pi}^{+\pi} | \sum_{k=1}^{N} \xi_k e^{-ik\omega} |^2 d\omega = 2\pi f_{min} > 0 .$$

It is obvious that λ_N^*, the smallest eigenvalue of Ω_N satisfies $\lambda_N^* \geq 2\pi f_{min}$. Hence Ω_N^{-1} exists for all N and, for all $\eta \in R^N$, with $\|\eta\| = 1$, we have

$$\alpha \leq < \eta , \Omega_N^{-1} \eta > \leq \beta$$

where α, β are finite positive constants. But for any $\xi \neq 0$ in R^p, $0 < \lim_{N\to\infty} \frac{1}{\sqrt{N}} \|V_N(y) \xi\| < \infty$, since $\lim_{N\to\infty} V_N^T(u) V_N(u)$ is nonsingular, and the sequence u is bounded [see Theorem 6] and hence, for all $\xi \neq 0$ in R^p,

$$0 < \lim_{N\to\infty} \frac{1}{N} < \xi , V_N^T(y) \Omega_N^{-1} V_N(y) \xi > < \infty$$

We conclude that $\lim_{N\to\infty} \frac{1}{N} V_N^T(y) \Omega_N^{-1} V_N(y)$ is a finite $p \times p$ positive definite matrix and hence nonsingular. Thus the matrices $\{Z_N^*\}_p^\infty$ form an instrumental matrix sequence.

It will be recalled from the discussion in Theorem 4 that to prove (35) we only need to show that the difference of asymptotic covariance matrices

$$(Z_N^T V_N(y))^{-1} Z_N^T \Omega_N Z_N (V_N^T(y) Z_N)^{-1} - (Z_N^{*T} V_N(y))^{-1} Z_N^{*T} \Omega_N Z_N^* (V_N(y) Z_N^{*T})^{-1}$$

$$= (Z_N^T V_N(y))^{-1} Z_N^T \Omega_N Z_N (V_N^T(y) Z_N)^{-1} - V_N^T(y) \Omega_N^{-1} V_N(y) \quad (37)$$

is always positive semidefinite. But this follows immediately from Theorem 5, which completes our proof. We shall call the matrix sequence $\{Z_N^*\}_p^\infty$, with $Z_N^* = \Omega_N^{-1} V_N(y)$, the <u>asymptotically optimal instrumental matrix sequence</u>.

In practice we cannot compute $\{Z_N^*\}_{N=p}^\infty$, because, on the one hand, we cannot observe $V_N(y)$, and, on the other, we need to know the unknown parameters a as well as the covariance function $r_d(k)$ of the noise to compute Ω_N. However, we can estimate the unknown parameter a by first choosing a reasonable instrumental variable matrix Z_N, and then estimate $V_N(y)$ and Ω_N^{-1}. Experimental results which will be given in Section 8, supports this method of approach.

6. APPROXIMATE COMPUTATION OF ASYMPTOTICALLY OPTIMAL INSTRUMENTAL MATRICES

To compute the asymptotically optimal matrix $Z_N^* = \Omega_N^{-1} V_N(y)$, it is necessary to invert the $N\times N$ matrix Ω_N, which becomes prohibitively difficult when N is large. We shall now sketch out a method for computing Z_N^* approximately by solving a set of difference equations, when the disturbance sequence is filtered white noise, i.e., it is the solution of the difference equation

$$\sum_{i=1}^{r} \alpha_i d_{k-r+i} = \sum_{i=1}^{s} \beta_i \omega_{k-s+i}, \qquad k=1, 2, \ldots \tag{38}$$

where $\beta_s \neq 0$, $s \leq r$, and the difference equation (38) is assumed to be input-output stable. The sequence $\omega = \{\omega_k\}$ is a white noise sequence, i.e., $E\omega_k = 0$ and $E\omega_k \omega_j = \delta_{kj}\sigma^2$. Without loss of generality, we can assume that $\sigma^2 = 1$.

Recalling that $V_N^1(d)^T = (d_1, d_2, \ldots, d_N)$, we find that if we neglect the zero input response, (38) yields

$$S_{N+p-1}(\alpha)\, V^1_{N+p-1}(d) = S_{N+p-1}(\beta)\, V^1_{N+p-1}(\omega)\ , \tag{39}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_s)$ and $S_N(\cdot)$ was defined in (19). Now, let A_N be an $N\times (N+p-1)$ matrix whose coefficients are the parameters of the system (3), defined as follows:

$$A_N = \begin{bmatrix} a_1 & a_2 & \cdots & a_p & 0 & 0 \\ 0 & a_1 & \cdots & \cdots & a_p & 0 \\ 0 & 0 & a_1 & \cdots & \cdots & a_p \end{bmatrix} \tag{40}$$

We recall that $V^1_N(e) = V_N(d)\, a$, by definition. This can obviously also be written as

$$V^1_N(e) = A_N\, V^1_{N+p-1}(d)\,, \tag{41}$$

which becomes, after substitution from (39),

$$V^1_N(e) = A_N\, S_{N+p-1}(\alpha)^{-1}\, S_{N+p-1}(\beta)\, V^1_{N+p-1}(\omega)\,, \tag{42}$$

and hence,

$$\begin{aligned} \Omega_N &= E\, V^1_N(e) >< V^1_N(e) \\ &= \left(A_N\left(S_{N+p-1}(\alpha)\right)^{-1} S_{N+p-1}(\beta)\right)\left(A_N\left(S_{N+p-1}(\alpha)\right)^{-1} S_{N+p-1}(\beta)\right)^T , \end{aligned} \tag{43}$$

since $E\, V^1_N(\omega) >< V^1_N(\omega) = I_N$, the $N\times N$ identity matrix.

Now, since $Z^*_N = \Omega_N^{-1}\, V_N(y)$, we have

$$\Omega_N\, Z^*_N = V_N(y)\,. \tag{44}$$

To show how Z^*_N can be computed approximately, consider the simplest case, $S_{N+p-1}(\alpha) = S_{N+p-1}(\beta) = I$, the identity matrix. Then, from (43) and (44),

$$A_N\, A^T_N\, Z^*_N = V_N(y)\,. \tag{45}$$

Now, with $z = (z_1, z_2, \ldots z_N)$, consider the equation

$$A_N A_N^T z = V_N^1(y). \tag{46}$$

Let $v = (v_1, v_2, \ldots v_{N+p-1})$ be given by $v = A_N^T z$.

Then, for $i=1, 2, \ldots, N+p-1$, the sequence v can be obtained as a solution to the difference equation

$$a_1 v_i + a_2 v_{i+1} + \ldots + a_p v_{i+p-1} = y_i, \quad i=1, 2, \ldots, N. \tag{47}$$

Let us denote by $Z_N^{*1} = (z_1^{*1}, z_2^{*1}, \ldots, z_N^{*1})$ the first column of Z_N^* and let $V_N^{*1} = A_N^T Z_N^{*1}$. We can compute $V^{*1} = (V_1^{*1}, V_2^{*1}, \ldots, V_{N+p-1}^{*1})$ from (47) by using the initial conditions $v_i = v_i^{*1}$, for $i=1, 2, \ldots, p-1$. If we know these, we can then compute Z_N^{*1} by solving the array $A_N^T Z_N^{*1} = V^{*1}$, from the bottom up. Now suppose that we solve (47) with arbitrary bounded initial conditions and input sequence $(y_1, y_2, \ldots, y_N) = V_N^1(y)$. Then, since (47) is stable by assumption, we find $|v_i - v_i^{*1}| \to 0$ as $i \to \infty$. If we now solve the array $A_N^T z = v$, with $z = (z_1, z_2, \ldots, z_{N+p-1})$ and $v = (v_1, \ldots, v_{N+p-1})$, compared as above, we find that $|z_i - z_i^{*1}| \to 0$ as i(and N) $\to \infty$. Consequently, if we denote by $\tilde{Z}_N$ the $N \times p$ matrix whose ith column corresponds to the solution of (47) from zero initial conditions and input sequence $V_N^i(y)$, we find that $\tilde{Z}_N \to Z_N^*$, in the sense that as $N \to \infty$,

$$\frac{1}{N}\left(\tilde{Z}_N^T V_N(y)\right)^{-1} \to \frac{1}{N} Z_N^{*T} V_N(y)^{-1}, \text{ and } \frac{1}{N} \tilde{Z}_N^T V_N^p(u) \to \frac{1}{N} Z_N^{*T} V_N^p(u),$$

and hence for $N \to \infty$

$$\left[\tilde{Z}_N^T V_N(y)\right]^{-1} \tilde{Z}_N^T V_N^p(u) \to \left[Z_N^{*T} V_N(y)\right]^{-1} Z_N^{*T} V_N^p(u), \tag{48}$$

which show that $\hat{a}_N(\tilde{Z}_N) \to \hat{a}_N(\tilde{Z}_N^*)$.

When the matrices $S_{N+p-1}(\alpha)$, $S_{N+p-1}(\beta)$ are not identity matrices, Z_N^* can be computed approximately in a similar way, by solving more difference equations.

7. ON-LINE ESTIMATION OF THE PARAMETER a

We shall now outline a method for computing on-line the minimax estimate of a, $\hat{a}_N(V_N(y))$, N=p, p+1, ..., derived in Section 5.

For k=1, 2, ..., let $M_k = Z_k^T V_k(x)$, where $\{Z_k\}_{k=p}^{\infty}$ is any instrumental matrix sequence, such that $Z_k = V_k(z)$, (see definition of $V_k(\cdot)$ in (5)) for some scaled sequence $z = \{z_1, z_2, \dots\}$, and where $x = (x_1, x_2, \dots)$ is the noise corrupted output sequence. For i=1, 2, ..., k, let z(i), be the ith row of Z_k and let x(i) be the ith row of $V_k(x)$, then

$$M_{k+1} = M_k + z(k+1) >< x(k+1), \qquad k=1, 2, \dots, \tag{49}$$

Now, $M_{k+1}^{-1} M_{K+1} = I$, the identity matrix, and hence

$$M_{k+1}^{-1}\left(M_k + z(k+1) >< x(k+1)\right) = I, \tag{50}$$

which may be rewritten as

$$M_{k+1}^{-1} = M_k^{-1}\left(I + z(k+1) >< x(k+1)\, M_k^{-1}\right)^{-1} \tag{51}$$

Making use of a well-known formula for the inversion of an identity plus a dyad (see Friedman p. 31 [4]), we get from (51)

$$M_{k+1}^{-1} = M_k^{-1}\left[\left(< x(k+1),\ M_k^{-1} z(k+1) > +1\right)^{-1} z(k+1) >< x(k+1)\right] M_k^{-1}\ . \tag{52}$$

Simplifying out, we obtain

$$M_{k+1}^{-1} = M_k^{-1} - \frac{1}{\alpha_{k+1}}\ M_k^{-1} z(k+1) >< x(k+1)\, M_k^{-1}\ , \tag{53}$$

where $\alpha_{k+1} \triangleq 1 + < x(k+1),\ M_k^{-1} z(k+1) >$

Similarly, we can develop an iterative formula for computing the estimate $\hat{a}_k$. Thus,

$$\hat{a}_{k+1} = \hat{a}_k - \frac{1}{\alpha_{k+1}} M_k^{-1} z(k+1)^T \left(< x(k+1), \hat{a}_k > - u_{k+p} \right) \tag{54}$$

Of course, given that M_k^{-1} exists, (53) and (54) are only meaningful when $\alpha_{k+1} \neq 0$.

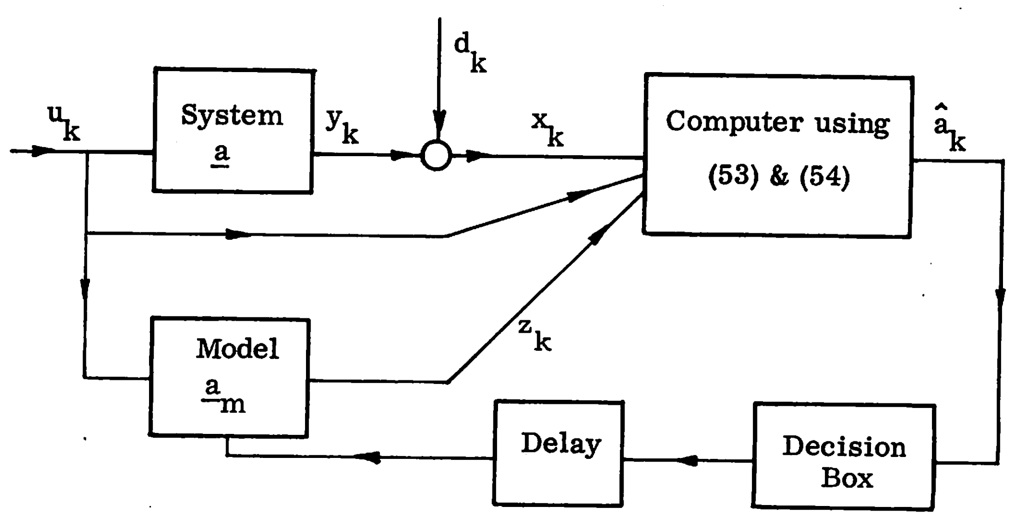
In practice, if $\alpha_{k+1} = 0$, we do not use the vectors $z(k+1)$ and $x(k+1)$ but wait until the first time j, $j=2, 3, \ldots,$ where α_{k+j}, becomes different from zero at which time we compute $\hat{a}'_{k+j}$ from the formula

$$\hat{a}'_{k+j} = \hat{a}_k - \frac{1}{\alpha_{k+j}} M^{-1} z(k+j)^T \left(< x(k+j), \hat{a}_k > - u_{k+j-1+p} \right) \tag{55}$$

This is equivalent to putting $z(k+1) = 0$ whenever $\alpha_{k+1} = 0$.

The main, and considerable, advantage offered by (53) and (54) is that these formulas eliminate the need for storing large numbers of past observations, as well as simplifying the computation of M_k^{-1}.

Although, the theoretical questions of the stability of the computation scheme proposed below still remains unanswered, the authors have found that in practice it works very well. Thus, the authors suggest that the minimax estimate $\hat{a}_N(V_N(y))$, can be computed iteratively, on-line as follows, assuming that d_k is statistically independing of d_{k+q} for some finite q. Consider the block diagram in Figure 1.



In order to estimate the unknown system parameters a in real time, we must be able to generate the elements of the instrumental matrices Z_N in real time. Let $Z_N = V_N(z)$, where the sequence z is the output from the model, i.e.,

$$\sum_{i=1}^{p} a_{mi}(k)\, z_{k-p+i} = u_k, \qquad k=1, 2, \ldots, . \tag{56}$$

For $k=1$, the vector $a_m(k)$ is set equal to our best guess of the parameter a and it is then modified with k as follows. The computer calculates $\hat{a}_k$, using the recursion formulas (53) and (54), and the Decision Box checks whether the roots λ_i of

$$\sum_{i=1}^{p} \hat{a}_i\,(1+(p+q-1)j)\;\lambda^{i-1} = 0, \qquad \text{for } j=0, 1, 2, \ldots \tag{57}$$

are inside the unit circle, using the Jury table stability criterion [6]. If all of the roots of (57) are inside the unit circle, then for $j=0, 1, 2, \ldots,$ $a_m(1+(p+q-1)(j+1))$ is set equal to $\hat{a}\,(1+(p+q-1)\,j)$. This delay ensures that $\operatorname*{p\,lim}_{N\to\infty} \frac{1}{N} V_N^T(z)\, V_N(d)\, a = 0$. Computational experience shows that with the input sequence u chosen to be a sample from a stationary random process the matrix $\frac{1}{N}\left(V_N(z)^T\, V_N(y)\right)^{-1}$ exists at least for N up to 500. Experimental results using the real time estimation scheme are given in Section 8.

8. EXPERIMENTAL RESULTS

Although many experiments were performed, we report here one set only, which would be adequate to illustrate the methods proposed.

(i) The real time estimation scheme shown in Figure 1, was simulated on a digital computer and the results of this simulation are given in Figure 2. The system difference equation, corresponding to (3), was

$$-0.269y(k-4) - 0.832y(k-3) - 0.22y(k-2) + 1.3y(k-1) + y(k) = u(k), \qquad k=1, 2, \ldots, (N+p-1). \tag{58}$$

The disturbance process $\{d(k)\}$ used is given by

$$d(k) = \omega(k) \times (\text{NOISE FACTOR}) \tag{59}$$

where $\{\omega(k)\}$ is a zero mean white noise process with unit variance and is independent of the input process $\{u(k)\}$. The input sequence is obtained by passing a white noise sequence through a stable linear filter with z transfer function $\frac{1}{z+0.6}$. The value of delay shown in Figure 1 was chosen to be five.

(ii) For comparison, the system parameters in (58) were also estimated, under the same testing conditions as in (i), by the recursive least squares method, i.e., by up-dating

$$\hat{a}_{L.S.}(k) = V_k(x)^T \; V_k(x)^{-1} \; V_k^T(x) \; V_k^p(a)$$

at each sampling instant. The results of this experiment are given in Figure 3, which clearly shown that $\hat{a}_{L.S.}$ is biased.

(iii) To estimate $Z_N^* = \Omega_N^{-1} V_N(y)$ for the system in (58) with d_k defined by (59), we proceeded as follows. For a given value of N, e.g., N=100, 200, ..., 500, we used the on-line estimation scheme, described under (i), to find an initial estimate of a, $\hat{a}^1(N)$, and we then generated the sequence $\hat{y}$, estimate of y, using the difference equation

$$\sum_{i=1}^{p} \hat{a}_i^1(N)\, \hat{y}(i-p+1) = u_i, \qquad i=1, 2, \ldots, (N+p-1)$$

with initial conditions equal to zero. Then, $\hat{Z}_N^* = \hat{\Omega}_N^{-1} V_N(\hat{y})$, the estimate of Z_N^*, was computed by the method outlined in Section 6 with $\hat{\Omega}_N = \hat{A}_N \hat{A}_N^T$, where $\hat{A}_N$ is the matrix in (40) with the components of $\hat{a}^1(N)$ taking the place of the components of a. A new estimate $\hat{a}^2$ was obtained using the formula

$$\hat{a}^2(N) = (\hat{Z}_N^{*T} \hat{Y}_N)^{-1} \hat{Z}_N^{*T} V^p(u) .$$

The final estimate $\hat{a}^3(N)$ is obtained by using $\hat{a}^2(N)$ to estimate Z_N^* and $V_N(y)$ again as before. Observe from Figure 4 that the estimates $\hat{a}^2$ (200) and $\hat{a}^2$ (300) are much better than $\hat{a}^1$ (200) and $\hat{a}^1$ (300) respectively.

CONCLUSION

The purpose of this paper was to explore the instrumental variable method as a tool for estimating the parameters of linear discrete time systems. It was shown that it always yields consistent estimates, in contrast to the least squares method, and that these estimates can be optimized either in the minimax sense or else in the least variance sense. Although the optimal estimates cannot be computed directly, it can be approximated quite closly. It was shown that the instrumental variable method can be adapted to yield a good on-line estimation scheme, which was checked out experimentally with satisfactory results.

Although not discussed in this paper, it was shown by one of the authors [3] that the problem of estimating the parameters of a linear differential system can also be tackled by means of the instrumental variable method to yield consistent parameter estimates. Thus, the instrumental variable method has broad applicability to the parameter estimation problem of dynamical systems and can find frequent utilization.

References

1. Reiersol, O., "Confluence Analysis by Means of Lag Moments and Other Methods Confluence Analysis", Econometrica, Vol. 9, No. 1, 1941, pp. 1-23.
2. Wilks, S., "Mathematical Statistics", Wiley, 1962, pp. 102-106.
3. Wong, K. Y., "Estimation of Parameters of Linear Systems Using The Instrumental Variable Method", PhD Thesis, June, 1966, University of California, Berkeley.
4. Friedman, B., Principles and Techniques of Applied Mathematics, John Wiley, 1954.
5. Gantmacher, F. R., Theory of Matrices, Chelsea Publishing Company, 1959.
6. Jury, E. I., Theory and Application of the z-Transform Method, Wiley, 1964, pp. 79-141.

Appendix I

Let $\xi \in R^N$ be such that $\|\xi\| = 1$, and let $\eta = S_N(a)\,\xi$, where $S_N(a)$ was defined in (19). Then,

$$\eta_j = \sum_{i=1}^{q} a_{p-i+1}\,\xi_{-i+j+1}, \quad j = 1, 2, \ldots, N$$

where $q = j$ if $j \le p$ and $q = p$ if $j > p$. Hence,

$$\eta_j^2 \le \alpha \sum_{i=1}^{p} \xi^2_{-i+j+1}, \quad j = 1, 2, \ldots, N$$

where $\alpha = \sum_{i=1}^{p} a_i^2$. Consequently,

$$\sum_{j=1}^{N} \eta_j^2 \le \alpha p \sum_{j=1}^{N} \xi_j^2 = \alpha p$$

and $\|S_N(a)\| \le (\alpha p)^{1/2}$.

NOISE = WHITE

NOISE FACTOR = 1.0

DELAY = 5

N	A1=-0.269	A2=-0.832	A3=-0.22	A4=+1.3	A5=+1.0	y_N	z_N
1	0.0240	-0.2020	-0.4500	0.6000	1.0000	-1.8955	0.7415
25	-0.5978	0.3613	1.7548	1.5535	-1.1170	1.1812	0.3433
50	-1.0651	-1.0364	0.5117	1.7497	0.9425	-4.8965	-0.5003
75	-0.4314	-0.9219	-0.0747	1.4609	1.0349	3.8628	1.6808
100	-0.3917	-0.9139	-0.1907	1.4196	1.1353	7.5516	8.8305
125	-0.3980	-0.9265	-0.0777	1.4495	1.0346	-1.1199	-2.7554
150	-0.4453	-0.9390	-0.0232	1.4410	1.0039	-2.4293	-0.1622
175	-0.3537	-0.9768	-0.1951	1.4599	1.0735	-3.0545	-1.8814
200	-0.3142	-0.9171	-0.2087	1.3555	1.0032	0.1055	0.3110
225	-0.3036	-0.8507	-0.1818	1.3049	0.9807	7.3331	8.1422
250	-0.2932	-0.8127	-0.1445	1.2915	0.9534	-3.6704	-3.9958
275	-0.2919	-0.8016	-0.1361	1.2864	0.9494	-4.6100	-4.9833
300	-0.2947	-0.8073	-0.1427	1.2814	0.9474	-3.3804	-3.4547
325	-0.2951	-0.8128	-0.1498	1.2778	0.9463	-3.0705	-3.6767
350	-0.3189	-0.8490	-0.1361	1.3182	0.9607	-1.8811	-1.5930
375	-0.3122	-0.8472	-0.1568	1.3150	0.9726	-6.6169	-7.0380
400	-0.3273	-0.8651	-0.1472	1.3291	0.9745	0.9522	0.9575
425	-0.3275	-0.8663	-0.1379	1.3426	0.9801	8.9385	9.7884
450	-0.3196	-0.8453	-0.1312	1.3213	0.9626	-4.4318	-4.7762
475	-0.3183	-0.8081	-0.0973	1.2851	0.9278	5.9764	6.8158
500	-0.3040	-0.8017	-0.1102	1.2899	0.9390	-5.8468	-6.5071

ON LINE INSTRUMENTAL VARIABLE ESTIMATES

Figure 2

NOISE = WHITE

NOISE FACTOR = 1.0

N	A1=-0.269	A2=-0.832	A3=-0.22	A4=+1.3	A5=+1.0	y_N	x_N
1	0.0240	-0.2020	-0.4500	0.6000	1.0000	-1.8955	-0.2320
25	0.1401	0.2042	0.2611	0.5563	0.3045	1.1812	1.9009
50	0.0595	-0.0023	0.0312	0.4543	0.3823	-4.8965	-3.8458
75	0.0348	-0.0734	-0.0167	0.4740	0.4099	3.8628	4.6301
100	0.0478	-0.0714	-0.0430	0.4397	0.4033	7.5516	7.6129
125	0.0602	-0.0217	-0.0185	0.4105	0.3721	-1.1199	-0.4874
150	0.0360	-0.0443	-0.0061	0.4106	0.3619	-2.4393	-4.0617
175	0.0500	-0.0336	-0.0315	0.3860	0.3599	-3.0545	-1.9085
200	0.0472	-0.0441	-0.0418	0.3695	0.3444	0.1055	-0.8602
225	0.0474	-0.0190	-0.0167	0.3587	0.3329	7.3331	6.6169
250	0.0397	-0.0220	-0.0004	0.3743	0.3355	-3.6704	-4.3196
275	0.0383	-0.0266	-0.0054	0.3664	0.3284	-4.6100	-3.5970
300	0.0232	-0.0427	-0.0006	0.3859	0.3411	-3.3804	-3.4515
325	0.0170	-0.0456	-0.0028	0.3831	0.3382	-3.0705	-2.8094
350	0.0120	-0.0473	0.0136	0.3982	0.3463	-1.8811	-2.9019
375	0.0060	-0.0492	0.0117	0.3962	0.3495	-6.6169	-4.9757
400	-0.0009	-0.0572	0.0129	0.3994	0.3505	0.9522	1.2164
425	-0.0028	-0.0585	0.0172	0.4039	0.3536	8.9385	9.5625
450	-0.0021	-0.0510	0.0230	0.4068	0.3556	-4.4318	-3.8517
475	-0.0149	-0.0626	0.0318	0.4145	0.3567	5.9764	4.2159
500	-0.0135	-0.0601	0.0308	0.4162	0.3599	-5.8638	-6.3396

<u>LEAST SQUARES ESTIMATES</u>

Figure 3

NOISE = WHITE, NOISE FACTOR = 1.0

N		A1=-0.269	A2=-0.832	A3=-0.220	A4=+1.300	A5=+1.000
100	$\hat{a}^1(100)$	-0.2977	-0.5743	0.5274	1.9763	1.2167
	$\hat{a}^2(100)$	-0.1081	-0.5052	-0.1756	1.0529	0.8743
	$\hat{a}^3(100)$	-0.1707	-0.6681	-0.2886	1.0659	0.9002
200	$\hat{a}^1(200)$	-0.4689	-1.0323	0.0874	1.8001	1.1876
	$\hat{a}^2(200)$	-0.2359	-0.7805	-0.2750	1.1687	0.9420
	$\hat{a}^3(200)$	-0.2520	-0.8090	-0.2666	1.2084	0.9607
300	$\hat{a}^1(300)$	-0.2970	-0.7807	-0.0110	1.4707	1.0379
	$\hat{a}^2(300)$	-0.2859	-0.8741	-0.2575	1.2848	0.9970
	$\hat{a}^3(300)$	-0.2898	-0.8814	-0.2559	1.2919	1.0022
400	$\hat{a}^1(400)$	-0.3299	-0.8431	-0.0334	1.4940	1.0534
	$\hat{a}^2(400)$	-0.2851	-0.8883	-0.2909	1.2658	0.9965
	$\hat{a}^3(400)$	-0.2885	-0.8881	-0.2700	1.2947	1.0081
500	$\hat{a}^1(500)$	-0.2532	-0.7166	-0.0399	1.3710	0.9883
	$\hat{a}^2(500)$	-0.2802	-0.8711	-0.2743	1.2647	0.9912
	$\hat{a}^3(500)$	-0.2806	-0.8685	-0.2637	1.2766	0.9956

NOISE = WHITE, NOISE FACTOR = 2.0

		A1=-0.269	A2=-0.832	A3=-0.220	A4=+1.300	A5=+1.000
100	$\hat{a}^1(100)$	-2.3305	-4.9473	-0.8722	5.1349	3.4944
	$\hat{a}^2(100)$	0.1073	-0.0109	-0.0146	0.7675	0.7059
	$\hat{a}^3(100)$	-0.1165	-0.5627	-0.2950	0.9619	0.8538
200	$\hat{a}^1(200)$	-1.5718	-3.1889	-0.4289	3.4075	2.2777
	$\hat{a}^2(200)$	-0.1342	-0.5109	-0.0825	1.1522	0.9009
	$\hat{a}^3(200)$	-0.2488	-0.8262	-0.3509	1.1078	0.9239
300	$\hat{a}^1(300)$	-0.5069	-1.2048	-0.1116	1.8315	1.2864
	$\hat{a}^2(300)$	-0.2859	-0.8815	-0.2891	1.2439	0.9800
	$\hat{a}^3(300)$	-0.2992	-0.9067	-0.2837	1.2790	0.9984
400	$\hat{a}^1(400)$	-0.5568	-1.2687	-0.0941	1.8717	1.2926
	$\hat{a}^2(400)$	-0.2754	-0.8785	-0.3137	1.2306	0.9843
	$\hat{a}^3(400)$	-0.3067	-0.9422	-0.3188	1.2913	1.0178
500	$\hat{a}^1(500)$	-0.2824	-0.7691	-0.0133	1.4779	1.0457
	$\hat{a}^2(500)$	-0.2799	-0.8928	-0.3417	1.1954	0.9678
	$\hat{a}^3(500)$	-0.2848	-0.8869	-0.2947	1.2539	0.9900

Figure 4