AUTOREGRESSIVE MODEL FITTING FOR CONTROL

HIROTUGU AKAIKE

(Received Dec. 25, 1970)

Summary

The use of a multidimensional extension of the minimum final prediction error (FPE) criterion which was originally developed for the decision of the order of one-dimensional autoregressive process [1] is discussed from the standpoint of controller design. It is shown by numerical examples that the criterion will also be useful for the decision of inclusion or exclusion of a variable into the model. Practical utility of the procedure was verified in the real controller design process of cement rotary kilns.

Introduction

In a practical situation of the autoregressive model fitting the order of the model is not generally known. The order may not be finite and rather will only be an artificial variable for the purpose of developping an approximation to the real world. Thus the decision of the order forms a crucial point in the autoregressive model fitting and any fitting procedure which lacks a description of this point may not be considered to be very efficient for practical applications. The situation is the same for any fitting procedure of finite parameter models and remains at present as a most challenging subject of study which deeply concerns with the practical utility of the whole statistical theories.

The natural approach to this type of problem will be to estimate the possible risk of fitting each model and adopt the one which gives the minimum of the estimates. The main problem in this approach is the choice of the risk function. In recent papers by the present author [1], [2] it has been demonstrated that the use of FPE (final prediction error), which is the mean square one-step prediction error when the set of the fitted coefficients is applied to another independent realization of the process, produces quite reasonable results. This fact is also confirmed by the results of applications of the procedure to many practical data.

Although, from the standpoint of evaluation of the linear transformation of the process under quadratic error criterion, the ultimate purpose of the model fitting may always be considered to be the identification of the spectral characteristics of the process the controller design of the process poses various interesting problems for the model fitting. As will be described briefly in the next section, the autoregressive representation of a multivariate stationary process can directly serve as a starting point of the controller design.

In a real situation of controlling a complex and noisy system another very important problem is the decision on the inclusion or exclusion of a variable into the model. Thus, besides the decision of the order, a procedure must be developed for the decision of the dimensionality of the multivariate vector process.

Extensions of the definition of FPE to multidimensional case have been proposed in [3] and the use of the estimate of generalized variance of the one step prediction error is suggested. Taking into account that this is the only quantity which appears in the maximum likelihood in Gaussian case it seems that this choice is a reasonable one for the purpose of the determination of the spectral characteristics of the process. In the case of controller design we are only interested in the predictability by the model of the system output, since the system input will eventually be under our complete control. This consideration suggests the replacement of the generalized variance by that of the system output variables. This is the procedure to be proposed in the present paper and this definition of the criterion naturally suggests an extension of its use for the purpose of inclusion and exclusion of the input variables. Once the extension of the decision procedure to the selection of input variables is admitted the procedure can also be utilized for the decision of the inclusion and exclusion of the output variables. Naturally the procedure of selecting variables excludes a complete theoretical analysis at present and it must be used with a sound scientific reasoning. The utility of the procedure is illustrated by using real and artificial data.

1. Use of autoregressive models for controller design

As a preliminary for the discussions in the following sections the use of the multivariate autoregressive model for the controller design under the quadratic criterion is briefly described in this section. We assume that the r-dimensional vector of the system output variables and the l-dimensional vector of the system input variables at time n are represented by x_n and y_n, respectively. We will call the components of x_n controlled variables and those of y_n manipulated variables. The $(r+l)$-dimensional vector X_n is defined by

$$X_n=\begin{array}{cc} \left[\begin{array}{c} x_n \\ y_n \end{array}\right] & \begin{array}{c} \updownarrow r \\ \updownarrow l \end{array} \end{array} . \tag{1.1}$$

For the sake of simplicity it is assumed that x_n has a zero mean vector. We assume that X_n admits the following autoregressive representation:

$$X_n=\sum_{m=1}^{M} A_m X_{n-m}+U_n , \tag{1.2}$$

where A_m is an $(r+l)\times(r+l)$ matrix and U_n is a random $(r+l)\times 1$ vector satisfying the relations

$$\begin{aligned} &EU_n=0 \qquad \text{(zero vector)} \\ &EU_nX'_{n-m}=0 \text{ (zero matrix)} \qquad \text{for } m\geqq 1 \\ &EU_nU'_m=\delta_{nm}S , \end{aligned} \tag{1.3}$$

where $\delta_{nm}=1$ $(n=m)$, $=0$ $(n\neq m)$ and S is a positive definite $(r+l)\times(r+l)$ matrix and the symbol $'$ denotes the transpose. From the representation (1.2) we get the following representation of controlled variables, which will be used for the controller design:

$$x_n=\sum_{m=1}^{M} a_m x_{n-m}+\sum_{m=1}^{M} b_m y_{n-m}+w_n , \tag{1.4}$$

where a_m, b_m and w_n are given by the relations

$$\begin{aligned} A_m&=\begin{array}{cc} & \begin{array}{cc} \leftarrow r\rightarrow & \leftarrow l\rightarrow \end{array} \\ \begin{array}{c} \updownarrow r \\ \updownarrow l \end{array} & \left[\begin{array}{cc} a_m & b_m \\ * & * \end{array}\right] \end{array} \\ U_n&=\begin{array}{cc} & \leftarrow 1\rightarrow \\ \begin{array}{c} \updownarrow r \\ \updownarrow l \end{array} & \left[\begin{array}{c} w_n \\ * \end{array}\right] \end{array} \end{aligned} \tag{1.5}$$

where $*$ denotes the irrelevant quantities for the present representation.

For the purpose of controller design, (1.4) is transformed into the following state space representation [10]:

$$Z_n=\Phi Z_{n-1}+\Gamma Y_{n-1}+W_n , \tag{1.6}$$

where

$$Z_n=\begin{array}{cc} & \leftarrow 1\rightarrow \\ \begin{array}{c} \updownarrow r \\ \updownarrow r \\ \\ \\ \updownarrow r \end{array} & \left[\begin{array}{c} z_n^{(1)} \\ z_n^{(2)} \\ \vdots \\ \\ z_n^{(M)} \end{array}\right] \end{array}$$

$$\Phi = \begin{array}{c} \\ r \\ r \\ \\ r \\ r \end{array} \begin{array}{c} \begin{array}{ccccc} \leftarrow r \rightarrow & \leftarrow r \rightarrow & \leftarrow r \rightarrow & & \leftarrow r \rightarrow \end{array} \\ \begin{bmatrix} a_1 & I & 0 & \cdots & 0 \\ a_2 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{M-1} & 0 & 0 & \cdots & I \\ a_M & 0 & 0 & \cdots & 0 \end{bmatrix} \end{array}$$

(1.7)
$$\Gamma = \begin{array}{c} r \\ r \\ \\ r \end{array} \begin{array}{c} \leftarrow l \rightarrow \\ \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} \end{array}$$

$$Y_{n-1} = \begin{array}{c} \leftarrow 1 \rightarrow \\ l \;\; y_{n-1} \end{array}$$

$$W_n = \begin{array}{c} r \\ r \\ \\ r \end{array} \begin{array}{c} \leftarrow 1 \rightarrow \\ \begin{bmatrix} w_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{array}$$

and

$$z_n^{(1)} = \begin{array}{c} \leftarrow 1 \rightarrow \\ r \;\; x_n \end{array} .$$

A simple controller design under quadratic performance criterion proceeds as follows: We assume that the performance of the control is evaluated by the quantity

$$J_T = E\left\{\sum_{n=1}^{T} (Z_n' Q Z_n + Y_{n-1}' R Y_{n-1})\right\} , \tag{1.8}$$

where Q and R are positive definite matrices of $Mr \times Mr$ and $l \times l$ dimensions and T is a properly chosen large integer and the control input Y_n is chosen so as to make J_T minimum. Obviously J_T admits the representation

$$\begin{aligned} J_T = EW_T' Q W_T + E\{&Z_{T-1}'(\Phi' Q \Phi') Z_{T-1} + Y_{T-1}' \Gamma' Q \Phi Z_{T-1} \\ &+ Z_{T-1}' \Phi' Q \Gamma Y_{T-1} + Y_{T-1}'(\Gamma' Q \Gamma + R) Y_{T-1}\} + J_{T-1} , \end{aligned} \tag{1.9}$$

and it can be seen that the optimal control Y_{T-1} is given by

$$Y_{T-1} = G_{T-1} Z_{T-1} , \tag{1.10}$$

where

$$G_{T-1} = -(\Gamma' Q \Gamma + R)^{-1} \Gamma' Q \Phi .$$

By inserting the result of (1.10) into (1.9) and successively applying Bellman's optimality principle we get

$$Y_n = G_n Z_n \,, \tag{1.11}$$

where $G_n = -(\Gamma' P_{T-n} \Gamma + R)^{-1} \Gamma' P_{T-n} \Phi$ and P_{T-n} is given by the recursive relation

$$\begin{aligned} &P_1 = Q \\ &\left.\begin{aligned} &M_i = P_{i-1} - P_{i-1}\Gamma(\Gamma' P_{i-1} \Gamma + R)^{-1} \Gamma' P_{i-1} \\ &P_i = \Phi' M_i \Phi \end{aligned}\right\} \quad i = 2, 3, \cdots, T\,. \end{aligned} \tag{1.12}$$

When T is sufficiently large G_1 will show little change for the further increase of T and for the control of the stationary process this G_1 is used as the fixed controller gain G and the control is realized by

$$Y_n = G Z_n \,. \tag{1.13}$$

Taking into account the fact that under fairly general condition the stationary process admits an autoregressive representation (1.2) [6], the discussion of this section can be considered to have given a theoretical justification for the controller design based on the model of (1.4) to be a generally useful procedure.

2. Fitting autoregressive models with FPE

In the following discussions we shall adopt the convention to denote by $X(i)$ the ith element of a vector X and by $A(i, j)$ the (i, j)th element of a matrix A.

We assume the model (1.2) of autoregression and consider the case where U_n's are independently identically distributed and X_n is stationary with finite second order moments. Under the assumption of finiteness of all order moments of U_n it was shown by Mann and Wald [5] that distribution of the least squares estimate $\hat{A}_m$ of A_m based on a set of observations $\{X_n;\ n=-M+1, -M+2, \cdots, N\}$ tends to be Gaussian, i.e., the distribution of $\sqrt{N}(\hat{A}_m(i, j) - A_m(i, j))$ $(i, j = 1, 2, \cdots, k,\ m = 1, 2 \cdots, M)$, where $k = r + l$, tends to be Gaussian with a zero mean vector and the covariance corresponding to $EN(\hat{A}_{m_1}(i_1, j_1) - A_{m_1}(i_1, j_1))(\hat{A}_{m_2}(i_2, j_2) - A_{m_2}(i_2, j_2))$ equal to $S(i_1, i_2) R_{xx}^{-1}(m_1, j_1; m_2, j_2)$, where S is the variance matrix of innovations as given by (1.3) and $R_{xx}^{-1}(m_1, j_1; m_2, j_2)$ is the $((m_1-1)\cdot k + j_1, (m_2-1)k + j_2)$th element of the inverse of the matrix R_{xx} which is an $Mk \times Mk$ matrix with the $((m_1-1)k + j_1, (m_2-1)k + j_2)$th element $R_{xx}(m_1, j_1; m_2, j_2)$ equal to $EX_{n-m_1}(j_1) X_{n-m_2}(j_2)$*. Hereafter we denote the ex-

* cf. Proposition in the Appendix.

pectation of a random variable assuming this limiting Gaussian distribution to be exact by E_∞, so that we have

$$E_\infty\{\sqrt{N}(\hat{A}_m(i,j)-A_m(i,j))\}=0$$

and

$$E_\infty\{N(\hat{A}_{m_1}(i_1,j_1)-A_{m_1}(i_1,j_1))(\hat{A}_{m_2}(i_2,j_2)-A_{m_2}(i_2,j_2))\} \\ =S(i_1,i_2)R_{xx}^{-1}(m_1,j_1;m_2,j_2)\ . \tag{2.1}$$

The one-step prediction error when $\hat{A}_m$ is applied to another independent realization of X_n is given by

$$D_n=\sum_{m=1}^{M}(A_m-\hat{A}_m)X_{n-m}+U_n\ . \tag{2.2}$$

We take the expectation E_x of D_nD_n' with respect to the realization X_n and get

$$E_xD_nD_n'=S+\sum_{m=1}^{M}\sum_{l=1}^{M}\Delta A_mEX_{n-m}X_{n-l}'(\Delta A_l)'\ , \tag{2.3}$$

where $\Delta A_m=\hat{A}_m-A_m$. We then consider the expectation of the last term with respect to this limiting distribution. We have

$$E_\infty\left\{N\sum_{m=1}^{M}\sum_{l=1}^{M}\Delta A_mEX_{n-m}X_{n-l}'(\Delta A_l)'\right\}(i,h) \\ =\sum_{m=1}^{M}\sum_{l=1}^{M}\sum_{j=1}^{k}\sum_{g=1}^{k}E_\infty\{N\Delta A_m(i,j)\Delta A_l(h,g)\}R_{xx}(m,j;l,g)\ ,$$

and by using (2.1)

$$=S(i,h)\sum_{m=1}^{M}\sum_{l=1}^{M}\sum_{j=1}^{k}\sum_{g=1}^{k}R_{xx}^{-1}(m,j;l,g)R_{xx}(m,j;l,g)$$

$$=MkS(i,h)\ . \tag{2.4}$$

Thus we have

$$E_\infty E_xD_ND_N'=\left(1+\frac{Mk}{N}\right)S\ , \tag{2.5}$$

and

$$\|E_\infty E_xD_ND_N'\|=\left(1+\frac{Mk}{N}\right)^k\|S\|\ , \tag{2.6}$$

where $\|A\|$ denotes the determinant of a matrix A. We shall call this last quantity given by (2.6) MFPE (multiple final prediction error) which coincides with our definition of FPE when $k=1$.

The meaning of MFPE is intuitively clear but it can not enjoy the unique status of FPE in one-dimensional case as an index of of the prediction error variance. For the adoption of MFPE as our criterion of the autoregressive model fitting we need a further justification. Since we are interested in the discrimination of the fitted models it will be reasonable to take into account the fact that when the process X_n is Gaussian the only sample function entering into the maximized asymptotic likelihood is $\|\hat{S}_M\|$ [8], where

$$\hat{S}_M=\frac{1}{N}\sum_{n=1}^{N}\left(X_n-\sum_{m=1}^{M}\hat{A}_m X_{n-m}\right)\left(X_n-\sum_{l=1}^{M}\hat{A}_l X_{n-l}\right)' . \tag{2.7}$$

Since $\hat{A}_m$ is obtained by minimizing the trace of $\hat{S}_M$ in (2.7), $\Delta A_m=\hat{A}_m-A_m$ is minimizing the trace of

$$\hat{S}_M=\frac{1}{N}\sum_{n=1}^{N}\left(U_n-\sum_{m=1}^{M}\Delta A_m X_{n-m}\right)\left(U_n-\sum_{l=1}^{M}\Delta A_l X_{n-l}\right)' \tag{2.8}$$

and it must hold that

$$\frac{1}{N}\sum_{n=1}^{N}\left(U_n-\sum_{m=1}^{M}\Delta A_m X_{n-m}\right)X'_{n-l}=0 . \tag{2.9}$$

From this relation we can get

$$\hat{S}_M=\frac{1}{N}\sum_{n=1}^{N}U_nU'_n-\sum_{l=1}^{M}\sum_{m=1}^{M}\Delta A_m\frac{1}{N}\sum_{n=1}^{N}X_{n-m}X'_{n-l}(\Delta A_l)' . \tag{2.10}$$

We have

$$E\left(\frac{1}{N}\sum_{n=1}^{N}U_nU'_n\right)=S , \tag{2.11}$$

and similarly as in (2.5)

$$E_\infty\left\{N\sum_{l=1}^{M}\sum_{m=1}^{M}\Delta A_m\frac{1}{N}\sum_{n=1}^{N}X_{n-m}X'_{n-l}(\Delta A_l)'\right\}=MkS . \tag{2.12}$$

where as was defined before E_∞ denotes the expectation of the limit distribution of the quantity within the braces as N tends to infinity. This observation suggests that for the present discussion where the quantity of the order of $1/N$ is playing crucial role it will be reasonable to adopt $(1-Mk/N)^{-1}\hat{S}_M$ as our estimate of $\hat{S}_M$ and accordingly to adopt

$$\left(1-\frac{Mk}{N}\right)^{-k}\|\hat{S}_M\| \tag{2.13}$$

as our estimate of $\|S\|$.

Based on these observations we propose the following procedure for autoregressive model fitting: We fit models with order M $(0 \leqq M \leqq L)$ by using the least square method and adopt the one which gives the minimum of

$$\text{MFPE}(M) = \left(1+\frac{Mk}{N}\right)^k \left(1-\frac{Mk}{N}\right)^{-k} \|\hat{S}_M\| \tag{2.14}$$

as our final choice.

3. System identification for control

The procedure described in the preceding section may suffer some lack of efficiency by assuming one and the same value of M for all the components, though this allows us to fully enjoy the efficient computing procedure developed by Whittle [9]. From the stand point of maximizing the likelihood function in Gaussian case it will be reasonable to focus our attension to the behavior of the estimate of the prediction error variance of any subset of the component variables of X_n when the corresponding components of U_n are independent from the rest. As was discussed in Section 1, in the case of controller design we need only the estimates of the rows of A_m which are giving the system outputs or the controlled variables. In many practical situations, if only sufficient number of variables are taken into account, it is quite possible that the disturbances originating in the controller or the manipulated variables are independent of those originating in the system, in the sense that the set of the components of U_n corresponding to the controlled variables are independent of those corresponding to the manipulated variables. In this case it will be most practical and useful to concentrate our attention to the subset of the controlled variables and replace the definition of MFPE(M) by

$$\text{FPEC}(M) = \left(1+\frac{Mk}{N}\right)^r \left(1-\frac{Mk}{N}\right)^{-r} \|\hat{S}_{r,M}\|, \tag{3.1}$$

where $\hat{S}_{r,M}$ is the $r \times r$ submatrix of $\hat{S}_M$ with the rows and columns corresponding to the controlled varibales. FPEC stands for final prediction error of the controlled variables. By the formulation of Section 1, $\hat{S}_{r,M}$ is the $r \times r$ submatrix in the upper left hand corner of $\hat{S}_M$.

By introducing FPEC(M) our model fitting procedure for the controller design will be realized by replacing MFPE(M) in the case of general autoregressive model fitting by FPEC(M) and by adopting $\hat{a}_m$, $\hat{b}_m$ which are obtained from $\hat{A}_m$ and the definition (1.5) as the coefficients of the model and $\hat{S}_{r,M}$, or $(1-Mr/N)^{-1}\hat{S}_{r,M}$, as an estimate of the variance matrix of the innovations w_n within the system. This procedure will guard us against adopting inadequately large or small values of M for the system due to the effect of the structure of manipulated variables.

It should be mentioned here that by following the line of the discussion of Section 5 of [1] it is possible to show that for $K \leqq M_1 \leqq M_2 \leqq L$, where K is the order of the autoregressive process, i.e., $A_K \neq 0$ and $A_m = 0$ for $m > K$, $N \log_e (\|\hat{S}_{M_1}\| / \|\hat{S}_{M_2}\|)$ is asymptotically distributed as the sum $\sum_{j=M_1+1}^{M_2} \chi_j^2$ of mutually independently distributed chi-square variables χ_j^2 each with k^2 degrees of freedom. Analogously $N \log_e (\|\hat{S}_{r,M_1}\| / \|\hat{S}_{r,M_2}\|)$ is asymptotically distributed as the sum $\sum_{j=M_1+1}^{M_2} \chi_j'^2$ of mutually independently distributed chi-square variables $\chi_j'^2$ each with kr degrees of freedom for $K_r \leqq M_1 < M_2 \leqq L$, where K_r is such that $a_m = 0$ for $m > K_r$ and $\neq 0$ for $m = K_r$. A proof of this fact will be given in Appendix. Statistical properties of MFPE (M) and FPEC (M) can be deduced from these results as in the one dimensional case. It should be noted that as k or r tends to be large the probabilities of adopting higher values of M_0 tend to be small.

The present minimum FPEC procedure will be a reasonable one in practical applications if only the dependence between the subsets of innovations is not quite significant and this will be the case if there exist lags in the responses of the controlled variables to the variations of the manipulated variables and also there does not exist any hidden variable which is influencing on both the controlled and manipulated variables simultaneously. This consideration suggests the utility of evaluation of independence between the innovations of controlled and manipulated variables, since this will give some indication of possible inadequacy of sampling interval for time sampled observations of a continuous process and/or of possible existence of some hidden influencing variables. For this purpose we propose the use of the statistic

$$\lambda = \frac{\|\hat{S}_M\|}{\|\hat{S}_{r,M}\| \, \|\hat{S}_{l,M}\|}, \tag{3.2}$$

where $\hat{S}_{l,M}$ denotes the $l \times l$ matrix in the lower right hand corner of $\hat{S}_M$. When it is desired to test the significance of λ being smaller than 1, the result by Whittle [8] can be used, which tells that asymptotically

$$\xi = -N \log_e \lambda \tag{3.2}$$

is distributed as a chi-square variable with $r \times l$ degrees of freedom.

Here we consider the problem of inclusion and exclusion of variables.

Given a set of controlled variables we adopt the set of manipulated variables which gives the minimum value of FPEC(M) of (3.1) for the modelling by the present data. Thus even if there is a record of another manipulated variable we do not think it profitable to include it into the present model if its inclusion increase the minimum of FPEC. Exclusion of a manipulated variable can be treated analogously. As to the inclusion and exclusion of a controlled variables we have only to place it in the set of manipulated variables and apply the above stated procedure.

It must be stressed here that this kind of simple and general decision may be allowed and useful only when there does not exists any definite structural information of the process under observation. Any structural information available of the process should be paid carefull attention at the time of modelling. FPEC will give an indication of relative merit of each model with respect to the given set of observation data.

4. A practical computation procedure

In practical applications when a set of data $\{X_n;\ n=1, 2, \cdots, N\}$ is given the sample means of each variables are first deleted and the sample covariance functions are computed by the formulae:

$$(4.1)\quad \begin{aligned} &C(X_{n-m}(j), X_{n-l}(h)) \\ &= \frac{1}{N} \sum_{n=1}^{N-m+l} (X_n(j)-\overline{X(j)})(X_{n-l+m}(h)-\overline{X(h)}) \qquad \text{for } m \geqq l \\ &= \frac{1}{N} \sum_{n=1}^{N-l+m} (X_{n-m+l}(j)-\overline{X(j)})(X_n(h)-\overline{X(h)}) \qquad \text{for } m \geqq l\,, \end{aligned}$$

where

$$\overline{X(j)} = \frac{1}{N} \sum_{n=1}^{N} X_n(j)\,.$$

Instead of solving for the least squares estimate $\hat{A}_m$ described in Section 2 we fit the autoregressive model with the second order moments equal to those given by (4.1) up to the order M. This gives the coefficients of autoregression A_m^M by the efficient recursive computational procedure based on a formula given by Whittle [4], [9]. We define $k \times k$ $(k=r+l)$ matrix C_m, for $m=0, 1, \cdots, L$, by

$$(4.2)\qquad C_m(i, j) = C(X_n(i), X_{n-m}(j))\,.$$

The matrices A_m^M $(m=1, 2, \cdots, M,\ M=1, 2, \cdots, L)$ are obtained recursively by the following relation with initials $A_l^0 = B_l^0 = 0$ (zero matrix),

$d_0=f_0=C_0$ and $e_0=C_1$:

$$
(4.3)\qquad
\begin{aligned}
A_l^{M+1}&=\begin{cases} A_l^M-D^M B_{M+1-l}^M & l=1,2,\cdots,M \\ D^M & l=M+1 \end{cases}\\
B_l^{M+1}&=\begin{cases} B_l^M-E^M A_{M+1-l}^M & l=1,2,\cdots,M \\ E^M & l=M+1\,, \end{cases}
\end{aligned}
$$

where

$$
(4.4)\qquad
\begin{aligned}
D^M&=e_M f_M^{-1}\\
E^M&=e_M' d_M^{-1}
\end{aligned}
$$

and

$$
(4.5)\qquad
\begin{aligned}
d_M&=C_0-\sum_{l=1}^{M} A_l^M C_l'\\
e_M&=C_{M+1}-\sum_{l=1}^{M} A_l^M C_{M+1-l}\\
f_M&=C_0-\sum_{l=1}^{M} B_l^M C_l\,,
\end{aligned}
$$

where $'$ denotes transpose. We replace $\hat{S}_M$ of the preceding section by d_M and taking into account of the effect of adjusting for the mean we modify the definitions of MFPE and FPEC into

$$\text{MFPE}\,(M)=\left(1+\frac{Mk+1}{N}\right)^k\left(1-\frac{Mk+1}{N}\right)^{-k}\|d_M\|$$

and

$$\text{FPEC}\,(M)=\left(1+\frac{Mk+1}{N}\right)^r\left(1-\frac{Mk+1}{N}\right)^{-r}\|d_{r,M}\|\,,$$

respectively, where $d_{r,M}$ is the $r\times r$ submatrix in the upper left hand corner of d_M.

Taking into account the required degrees of freedom of the related statistics, it seems reasonable to keep L within the limit of $N/(10k)$ or $N/(5k)$.

5. Numerical examples

The decision procedure described in the preceding section was applied to the real data of cement rotary kilns of Chichibu Cement Company, Kumagaya, Japan, by Dr. T. Nakagawa and other members of the company. It was found that the procedure is very useful for the design of actual controllers and is eliminating the tedious trial and error process of identification of the basic model. Technical details of the recent results will be published shortly. We shall here content ourselves by presenting some of the results obtained by applying the procedure to a set of old data of which analysis was reported in [7]. The data were obtained while the kiln was under human control. By the result of the former analysis three variables were selected as controlled variables, which will be represented by $x_n(1)$, $x_n(2)$ and $x_n(3)$, and four variables $y_n(1)$, $y_n(2)$, $y_n(3)$ and $y_n(4)$ were identified as manipulated variables. For various combinations of variables the values of the minimum FPEC (M) are given in the Table 1 along with the values M_0 of M which gave the

Table 1. Application to kiln data

Controlled variable	Manipulated variable	M_0	FPEC $(M_0)\times 10^{-4}$
$x_n(1)$, $x_n(2)$, $x_n(3)$		6	2.22927
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(1)$	6	2.23594
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(2)$	7	2.27954
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$	6	2.15254
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(4)$	6	2.17707
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$, $y_n(1)$	6	2.10345
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$, $y_n(2)$	6	2.20113
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$, $y_n(4)$	6	2.08433
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$, $y_n(4)$, $y_n(1)$	6	2.05687
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$, $y_n(4)$, $y_n(2)$	6	2.13427
$x_n(1)$, $x_n(2)$, $x_n(3)$	$y_n(3)$, $y_n(4)$, $y_n(1)$, $y_n(2)$	6	2.09615

minima. The length of observation was $N=511$ and L was set equal to 15. The results in Table 1 show that the variables $y_n(3)$, $y_n(4)$, $y_n(1)$ and $y_n(2)$ are contributing to reduce the FPEC (M_0) in this order and that the inclusion of $y_n(2)$ into the model may not be profitable. Later analysis has shown that this result may be due to the very noisy behavior of $y_n(2)$ at the time of the experiment and the data provided by the later experiments under more natural running conditions have proven the inclusion of $y_n(2)$ necessary. The result is also considered to be due to the fact that the effect of manipulating $y_n(2)$ tends to be significantly non-linear when the system is too noisy or abnormal.

The present use of FPEC may find application in many other situations, such as the case of cross-spectrum estimation by autoregressive model fitting. In this case if uncorrelated observations are included into the fitting procedure this will introduce some bias into the values of M_0, the value of M which gives the minimum of FPEC (M) $(M=0, 1, \cdots, L)$. In the case of the example of Table 1 the value of M_0 which gave the minimum of MFPE for the total seven variables was equal to 4 contrary to 6 giving the minimum of FPEC. Thus a reasonable procedure may be to watch the behaviour or FPEC (M_0) at each inclusion of new variables into the analysis and divide the variables into small number of groups and adopt a different value of M_0 for each group. To show the feasibility of this type of procedure a numerical experiment was performed on the variables $x_n(i)$ $(i=1, 2, 3, 4)$, where $x_n(1)$ and $x_n(2)$ were given by the relations

$$
(5.1)\quad \begin{aligned} x_n(1) &= 0.5x_{n-1}(1)+0.4x_{n-1}(2)+v_n(1) \\ x_n(2) &= -0.6x_{n-1}(1)+0.7x_{n-1}(2)+0.3x_{n-2}(1)+0.2x_{n-2}(2)+v_n(2) \end{aligned}
$$

and

$$
(5.2)\quad \begin{aligned} v_n(1) &= u_n(1)+0.4u_n(2) \\ v_n(2) &= u_n(2)+0.4u_n(1)\ , \end{aligned}
$$

where $u_n(1)$ and $u_n(2)$ are the sequences of mutually independent random numbers both uniformly distributed over $[-0.5, 0.5]$, and $x_n(3)$ and $x_n(4)$ were realizations of physical processes which were mutually independent and also independent from $x_n(1)$ and $x_n(2)$. The results of application of our procedure are given in Table 2. For the sake of simplicity we denoted the variables for which we have evaluated FPEC (M) as controlled variables and other variables taken into consideration as manipulated variables. FPEC (M) is equal to MFPE (M) when the manipulated variables are absent. The values $L=15$ and $N=500$ were adopted. The

Table 2. Application to artificial data

Controlled variable	Manipulated variable	M_0	FPEC (M_0)	ξ (d.f.)
$x_n(1)$		3	0.558057	
$x_n(1)$	$x_n(2)$	1	0.094309	2.3×10^2 (1)
$x_n(1)$, $x_n(2)$		2	0.004971	
$x_n(1)$, $x_n(2)$	$x_n(3)$	2	0.005031	
$x_n(1)$, $x_n(2)$	$x_n(3)$, $x_n(4)$	2	0.005079	2.3 (4)
$x_n(1)$, $x_n(2)$, $x_n(3)$, $x_n(3)$		4	0.000672	
$x_n(3)$, $x_n(4)$		13	0.003382	

results show that the inclusion of $x_n(2)$ into the analysis along with $x_n(1)$ is profitable but the inclusion of $x_n(3)$ and $x_n(4)$ is unprofitable. Also the inclusion of $x_n(3)$ or $x_n(3)$ and $x_n(4)$ causes an unnecessary increase of M for $x_1(n)$ and $x_2(n)$ and decrease for $x_3(n)$ and $x_4(n)$. Though the almost perfect identifications of orders of the model is partly by coincidence the results give a very clear illustration of the possible use of the procedure discussed in the preceding paragraph. Two of the values of the chi-square variable ξ defined by (3.2) are also illustrated in Table 2 with corresponding degrees of freedom within the parentheses. The values very clearly show the dependence and independence of innovations in the two cases and suggest the utility of this statistic.

6. Concluding remarks

The minimum FPEC procedure described in this paper will be valid even if the manipulate variables $y_n(i)$ contain deterministic components if only it is assured that the sample variance and covariance matrices $C(X_{n-m_1}(j_1), X_{n-m_2}(j_2))$ tend to some definite constant matrices which can replace $EX_{n-m_1}(j_1)X_{n-m_2}(j_2)$ in the definition of R_{xx} of Section 2, which is assumed to be non-singular for any choice of M. In this way the procedure may find an application in the field of econometric model building. It must be born in mind that generally a manipulated variable under completely noise-free linear feedback control is not allowed in the present identification procedure. If it is considered to be desirable to adopt different values of M for each controlled variables we can apply the procedure by assuming each controlled variable in turn as the only output of the system and assuming all the variables except this one as manipulated variables in the definition of FPEC. The use of this procedure has been discussed in [3]. This is equivalent to ignoring the effect of possible correlations within the components of innovations. There are obvious modifications of the procedure to allow finer specifications of the system equations, but except for some special situations the procedure described in the present paper will serve as the most practical and generally useful one for the identification of a system for the purpose of controller design.

Acknowledgements

The result reported in this paper is a direct consequence of the cooperative research of cement rotary kilns by Dr. T. Nakagawa and the present author. Necessary programmings for the computations were performed by Miss E. Arahata.

Appendix

A derivation of the asymptotic distribution of statistics

$$\kappa_M = -N \log_e \| W_r^{-1} \hat{S}_{r,M} \| , \tag{A.1}$$

where

$$W_r = \frac{1}{N} \sum_{n=1}^{N} w_n w_n' \tag{A.2}$$

and w_n is given by (1.5), is presented here. The distribution of these statistics forms a theoretical background of the minimum FPEC procedure proposed in the text. We shall hereafter denote the operation $\frac{1}{N}\sum_{n=1}^{N}$ by $\overline{}$. We sometimes use the notation

$$c_m = \updownarrow r \, [\overset{\leftarrow r \rightarrow}{a_m}, \ \overset{\leftarrow l \rightarrow}{b_m}] \tag{A.3}$$

and also recall the definition $X_n' = [x_n' y_n']$.

We have

$$\begin{aligned} \hat{S}_{r,M} &= \overline{x_n x_n'} - \sum_{m=1}^{M} \hat{c}_m \overline{X_{n-m} x_n'} \\ &= H(M) - N^{-1} Q(M) , \end{aligned} \tag{A.4}$$

where

$$H(M) = \overline{x_n x_n'} - \sum_{m=1}^{M} (c_m \overline{X_{n-m} x_n'} + x_n \overline{X_{n-m}' c_m'}) + \sum_{m=1}^{M} \sum_{l=1}^{M} c_m \overline{X_{n-m} X_{n-l}'} c_l' \tag{A.5}$$

and

$$Q(M) = N \sum_{m=1}^{M} \sum_{l=1}^{M} \Delta c_m \overline{X_{n-m} X_{n-l}'} (\Delta c_l)' , \tag{A.6}$$

where $\Delta c_m = \hat{c}_m - c_m$ and $\hat{c}_m$ is the least squares estimate of c_m. We assume $c_m = 0$ for $m \geqq K_r$ and get

$$\begin{aligned} H(M) &= \overline{w_n w_n'} \\ &= W_r \qquad \text{for } M \geqq K_r . \end{aligned} \tag{A.7}$$

We also have for $M \geqq K_r$

$$[\Delta c_1, \Delta c_2, \cdots, \Delta c_M] = [\overline{w_n X_{n-1}'}, \overline{w_n X_{n-2}'}, \cdots, \overline{w_n X_{n-M}'}] C_{xx}^{-1} , \tag{A.8}$$

where C_{xx} is an $Mk \times Mk$ matrix with

$$C_{xx}((l-1)k+i, (m-1)k+j) = \overline{X_{n-l}(i) X_{n-m}(j)} . \tag{A.9}$$

For this case $Q(M)$ can be represented in the form

$$Q(M)=N[\overline{wX'}]C_{xx}^{-1}[\overline{wX'}]' , \tag{A.10}$$

where

$$X'=[X'_{n-1}, X'_{n-2}, \cdots, X'_{n-M}]$$

(A.11) and

$$[\overline{wX'}]=[\overline{w_nX'_{n-1}}, \cdots, \overline{w_nX'_{n-M}}] .$$

The whole discussion will depend on the following

PROPOSITION (Mann and Wald [5]). The limit distribution of $\sqrt{N}[\overline{wX'}]\cdot(i, (l-1)k+j)$ $(i=1, 2, \cdots, r;\ l=1, 2, \cdots, M;\ j=1, 2, \cdots, k)$ is Gaussian with a zero mean vector and with the variance covariance matrix given by the relation

$$E_\infty N([\overline{wX'}](i, \cdot)'([\overline{wX'}](j, \cdot))=S(i, j)R_{xx} , \tag{A.12}$$

where $[\ ](i, \cdot)$ denotes the ith row of the matrix $[\ ]$ and $R_{xx}=EXX'$.

Since we have for $M\geqq K_r$

$$\|W_r^{-1}\hat{S}_{r,M}\|=\|I-N^{-1}W_r^{-1/2}Q(M)W_r^{-1/2}\| , \tag{A.13}$$

we have

$$-N \log_e \|W_r^{-1}\hat{S}_{r,M}\|=\text{Trace}\,(W_r^{-1/2}Q(M)W_r^{-1/2})+o_p(1) , \tag{A.14}$$

where $o_p(1)$ denotes a term which is stochastically vanishing as N tends to infinity. Thus we have only to concern ourselves with the limiting distribution of Trace $(W_r^{-1/2}Q(M)W_r^{-1/2})$. The limit distribution of this quantity is identical if we replace C_{xx} in the definition of $Q(M)$ by its theoretical value R_{xx} and W_r by S_r which is the $r\times r$ matrix at the upper left corner of S. We know that there is a special orthogonalization procedure of $X_{n-1}, X_{n-2}, \cdots, X_{n-M}$, which is implicitly described in Section 4 and, by using the same notation for the corresponding quantities, is realized in the form

$$Z_{n-l}=X_{n-l}-\sum_{m=1}^{l-1} B_m^{l-1}X_{n-l+m} \qquad l=1, 2, \cdots, M . \tag{A.15}$$

We further orthonormalize the components within Z_{n-l} by a transformation

$$V_l=O_lZ_{n-l} , \tag{A.16}$$

where O_l is a $k\times k$ matrix. Combining these transformations we get a transformation T_M

(A.17) $$V = T_M X,$$

where $V' = [V_1', V_2', \cdots, V_M']$ and T_M is an $M \times M$ block matrix of $k \times k$ matrices, of which superdiagonal triangular part is filled with $k \times k$ zero matrices. From the structure of T_M we have

(A.18) $$\begin{aligned} EVV' &= I \\ &= T_M R_{xx} T_M', \end{aligned}$$

where I denotes an $Mk \times Mk$ identity matrix, and $Q(M)$ can be replaced by

(A.19) $$Q(M) = N[\overline{wV'}][\overline{wV'}]',$$

where it holds that

(A.20) $$E_\infty N([\overline{wV'}])'(i,\cdot)[\overline{wV'}](j,\cdot) = S_r(i,j)I.$$

This shows that in the limit distribution the columns of $\sqrt{N}[\overline{wV'}]$ are distributed independently with covariance matrices equal to S_r and consequently the components of the $r \times Mk$ matrix $\sqrt{N} S_r^{-1/2}[\overline{wV'}]$ are asymptotically distributed mutually independently as Gaussian with zero means and unit variances. Since W_r converges to S_r with probability one as N tends to infinity we can see that the limit distribution of Trace $(W_r^{-1/2} \cdot Q(M) W_r^{-1/2})$ is identical to that of

(A.21) $$\sum_{l=1}^{M} \sum_{j=1}^{k} \sum_{i=1}^{r} \{\sqrt{N}(S_r^{-1/2}[\overline{wV'}])(i,(l-1)k+j)\}^2.$$

It is obvious that T_M is identical to the $Mk \times Mk$ matrix at the left upper corner of T_L for $M \leqq L$ and since $V = T_M X$ we can see that (A.21) is identical to the partial sum

(A.22) $$\sum_{l=1}^{M} \chi_{kr}^2(l) \qquad \text{for } M \leqq L,$$

where

$$\chi_{kr}^2(l) = \sum_{j=1}^{k} \sum_{i=1}^{r} \{\sqrt{N}(S_r^{-1/2}[\overline{wV'}])(i,(l-1)k+j)\}^2$$

$(l = 1, 2, \cdots, L)$ are defined by putting $M = L$ in (A.21) and are asymptotically mutually independently distributed as chi-square variables with d.f. kr. This gives the asymptotic property of the statistic κ_M defined by (A.1) and completes the proof of the statement made at Section 3 about the asymptotic distribution of $N \log_e(\|\hat{S}_{r,M_1}\|/\|\hat{S}_{r,M_2}\|)$ which is equal to $\kappa_{M_2} - \kappa_{M_1}$ $(k_r \leqq M_1 \leqq M_2 \leqq L)$.

Since we have

$$\text{(A.23)} \quad N\left(\frac{\text{FPEC}\,(M_2)}{\text{FPEC}\,(M_1)}-1\right)=2(M_2-M_1)kr-N\log_e\frac{\|\hat{S}_{r,M_1}\|}{\|\hat{S}_{r,M_2}\|}+o_p(1)\ ,$$

we can see from the above result that for $M_2 \geqq M_1 \geqq K_r$ $N(\text{FPEC}\,(M_2)/\text{FPEC}\,(M_1)-1)$ will asymptotically behave as a realization of a random walk with variance of each step equal to $2kr$ and with upward drift of amount kr. Thus when kr tends to be significant compared with N the present minimum FPEC procedure may show the tendency to underestimate the order of the process, ignoring the effect of the bias of order kr/N.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1970). Statistical predictor identification, *Ann. Inst. Statist. Math.*, **22**, 203-217.

[2] Akaike, H. (1970). On a semi-automatic power spectrum estimation procedure, *Proc. 3rd Hawaii International Conference on System Sciences*, 974-977.

[3] Akaike, H. (1970). On a decision procedure for system identification, *Preprints, IFAC Kyoto Symposium on System Engineering Approach to Computer Control*, 485-490.

[4] Jones, R. H. (1964). Prediction of multivariate time series, *J. Appl. Meteor.*, **3**, 285-289.

[5] Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations, *Econometrica*, **11**, 173-220.

[6] Masani, P. (1966). Recent trends in multivariate prediction theory, *Multivariate Analysis*, (P. R. Krishnaiah ed.), Academic Press, New York, 351-382.

[7] Otomo, T., Nakagawa, T. and Akaike, H. (1969). Implementation of computer control of a cement rotary kiln through data analysis, *Preprints, IFAC 4th Congress, Warszawa.*

[8] Whittle, P. (1953). The analysis of multiple stationary time series, *J. R. Statist. Soc.*, **B 15**, 125-139.

[9] Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, *Biometrika*, **50**, 129-134.

[10] Wong, K. Y., Wiig, K. M. and Allbritton, E. J. (1968). Computer control of the Clarksville cement plant by state space design method, *IEEE Cement Industry Technical Conference, St. Louis, U.S.A., May* 20-24.

CORRECTION TO

"AUTOREGRESSIVE MODEL FITTING FOR CONTROL"

H. Akaike

This Annals, 23 (1971), 163–180. On page 167, line 8, $P_t = \Phi' M_t \Phi$ should be read as $P_t = \Phi' M_t \Phi + Q$.