International Journal of Control

Identification of MIMO non-linear systems using a forward-regression orthogonal estimator

S. A. Billings [a]; S. Chen [a]; M. J. Korenberg [b]
[a] Department of Control Engineering, University of Sheffield, Sheffield, U.K
[b] Department of Electrical Engineering, Queen's University, Kingston, Canada

PLEASE SCROLL DOWN FOR ARTICLE

Identification of MIMO non-linear systems using a forward-regression orthogonal estimator

S. A. BILLINGS†, S. CHEN† and M. J. KORENBERG‡

An orthogonal least squares estimator, which was originally derived for single-input single-output systems, is extended to multi-input multi-output non-linear systems. The estimator can provide information about the structure, or which terms to include in the model, and final parameter estimates in a very simple and efficient manner. Multivariable non-linear model validity tests are also discussed and the results of applying the orthogonal estimator to both simulated and real data are included.

1. Introduction

Model structure determination is often vital for the identification of multivariable non-linear systems because simply increasing the orders and the non-linear degree of the model to achieve the desired accuracy will almost certainly result in an excessively complex model and numerical ill-conditioning. A very complex model is not only computationally expensive but also has limited practical value.

For many real single-input single-output (SISO) systems it has been shown that provided the significant terms in the model can be detected, models with about 10 terms are usually sufficient to capture the dynamics of highly non-linear processes (Billings 1986, Billings and Fadzil 1985, Billings et al. 1988 b, Billings et al. 1988 a). Several model structure selection methods have been derived for SISO non-linear systems (e.g. Billings and Voon 1986 b, Leontaritis and Billings 1987, Korenberg et al. 1988, Billings et al. 1988 c). Most of these methods can be extended to multi-input multi-output (MIMO) systems.

The orthogonal least squares estimator given by Korenberg et al. (1988) and Billings et al. (1988 c) has proved to be very efficient in determining the significant terms and providing final unbiased parameter estimates. The orthogonal property of the estimator results in a particularly simple estimation procedure. In the present study this orthogonal estimator is applied to the MIMO non-linear identification problem. An estimated model will in general be accepted only after model validity tests have confirmed that the fit is adequate, and simple correlation tests (Billings and Voon 1986 a) developed for validating SISO non-linear models and a chi-square statistical test (Bohlin 1978) are extended to multivariable non-linear model validation.

The work is organized as follows. Section 2 describes a MIMO non-linear system representation which will be used as the basis for the identification. Section 3 reviews the orthogonal least squares estimator and its forward-regression version. The application of the forward-regression orthogonal estimator to MIMO non-linear

---

models is given in § 4 and multivariable non-linear model validation methods are discussed in § 5. A simulation study is presented in § 6.

2. System representation

Under some mild assumptions a discrete-time multivariable non-linear stochastic control system with m outputs and r inputs can be described by the model (Leontaritis and Billings 1985)

$$y(t) = f(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u), e(t-1) ..., e(t-n_e)) + e(t) \qquad (1)$$

where

$$y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{bmatrix}, \quad u(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_r(t) \end{bmatrix}, \quad e(t) = \begin{bmatrix} e_1(t) \\ \vdots \\ e_m(t) \end{bmatrix} \qquad (2)$$

are the system output, input, and noise respectively; n_y, n_u and n_e are the maximum lags in the output, input, and noise; $\{e(t)\}$ is assumed to be a white sequence; and $f(\cdot)$ is some vector-valued non-linear function. The representation in (1) is referred to as a multistructural input–output innovation model by Leontaritis and Billings (1985) and, for the SISO case, it is known as the NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous inputs) model. Here the name NARMAX will be used to include both the SISO and MIMO cases.

Model (1) will be used as the basis for identification of MIMO non-linear systems in the present study. Expressing (1) in its component form gives, for the ith row

$$y_i(t) = f_i(y_1(t-1), ..., y_1(t-n_y), ..., y_m(t-1), ..., y_m(t-n_y), u_1(t-1), ...,$$

$$u_1(t-n_u) ..., u_r(t-1), ..., u_r(t-n_u), e_1(t-1), ..., e_1(t-n_e), ...,$$

$$e_m(t-1) ..., e_m(t-n_e)) + e_i(t), \quad i = 1, ..., m \qquad (3)$$

To increase the flexibility in the model structure selection the maximum lags for each output, input, and noise may be assigned to different values so that

$$y_i(t) = f_i(y_1(t-1), ..., y_1(t-n^i_{y_1}), ..., y_m(t-1), ..., y_m(t-n^i_{y_m}), u_1(t-1), ...,$$

$$u_1(t-n^i_{u_1}), ..., u_r(t-1), ..., u_r(t-n^i_{u_r}), e_1(t-1), ..., e_1(t-n^i_{e_1}), ...,$$

$$e_m(t-1), ..., e_m(t-n^i_{e_m})) + e_i(t), \quad i = 1, ..., m \qquad (4)$$

The non-linear form of $f_i(\cdot)$ is generally unknown. In order to use the model (4) for identification, a means of parametrization is required and a polynomial expansion of $f_i(\cdot)$ is a convenient but by no means the only choice. Expanding $f_i(\cdot)$ as a polynomial of degree L_i gives the representation

$$y_i(t) = \sum_{j=1}^{n_i} \theta_{ij} x_{ij}(t) + e_i(t), \quad i = 1, ..., m \qquad (5)$$

where

$$\left. \begin{array}{l} n_i = \sum_{j=0}^{L_i} n_{ij}, \quad n_{i0} = 1 \\[6mm] n_{ij} = \dfrac{n_{ij-1}\left[\sum_{k=1}^{m}(n^i_{y_k} + n^i_{e_k}) + \sum_{k=1}^{r} n^i_{u_k} + j - 1\right]}{j}, \quad j = 1, ..., L_i \end{array} \right\} \qquad (6)$$

and the $x_{ij}(t)$ are monomials of degrees 0 to L_i, each consisting of delayed outputs, inputs and/or noises (degree 0 corresponding to a constant term). Regrouping terms in (5) yields

$$y_i(t) = f_i^P(y_1(t-1), ..., y_1(t-n_{y_1}^i), ..., y_m(t-1), ..., y_m(t-n_{y_m}^i), u_1(t-1) ...,$$

$$u_1(t-n_{u_1}^i), ..., u_r(t-1), ..., u_r(t-n_{u_r}^i)) + f_i^n(y_1(t-1), ..., y_1(t-n_{y_1}^i), ...,$$

$$y_m(t-1), ..., y_m(t-n_{y_m}^i), u_1(t-1), ..., u_1(t-n_{u_1}^i), ..., u_r(t-1) ...,$$

$$u_r(t-n_{u_r}^i), e_1(t-1), ..., e_1(t-n_{e_1}^i), ..., e_m(t-1) ...,$$

$$e_m(t-n_{e_m}^i)) + e_i(t), \quad i = 1, ..., m \tag{7}$$

where $f_i^P(\ \cdot\ )$ includes all terms $\theta_{ij}x_{ij}(t)$ which do not contain noise elements. The remainder of the terms are included in $f_i^n(\ \cdot\ )$. Consequently $f_i^P(\ \cdot\ )$ will be referred to as the ith process model and $f_i^n(\ \cdot\ )$ as the ith noise model.

A full model of (5) can easily involve an excessive number of terms as indicated by (6). For example, let $m = r = 2$, all maximum lags be 2, and polynomial degrees $L_1 = L_2 = 2$, then the number of parameters is 182. If $m = r = 10$, all maximum lags are 12 and polynomial degrees are 4, the number of parameters will be in excess of the capacity of a 4-byte integer in computer store. Direct estimation based on the full model (5) will therefore result in an excessively complex model, and such an identification problem is almost certainly ill-conditioned. The determination of the structure or which terms to include in the model from the large number of candidate terms is therefore essential in MIMO non-linear identification. In reality each subsystem in (5) may involve only 10–20 significant terms. In the SISO case, an orthogonal least squares estimator (Korenberg *et al.* 1988, Billings *et al.* 1988 c) has been developed which can select significant terms very efficiently. This estimator can easily be extended to MIMO systems.

3. A forward-regression orthogonal estimator

This section reviews the orthogonal least squares estimator (Korenberg *et al.* 1988) and its forward-regression version (Billings *et al.* 1988 c). Consider the linear regression function

$$z(t) = \sum_{i=1}^{M} p_i(t)\theta_i + \xi(t), \quad t = 1, ..., N \tag{8}$$

where $z(t)$ is the dependent variable or the term to regress upon, the $p_i(t)$ are regressors or predictors, $\xi(t)$ is some modelling error, and θ_i are unknown parameters to be estimated. Equation (8) can be written as

$$\mathbf{Z} = \mathbf{P\Theta} + \Xi \tag{9}$$

with

$$\mathbf{Z} = \begin{bmatrix} z(1) \\ \vdots \\ z(N) \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} P^T(1) \\ \vdots \\ P^T(N) \end{bmatrix}, \quad \mathbf{\Theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_M \end{bmatrix}, \quad \Xi = \begin{bmatrix} \xi(1) \\ \vdots \\ \xi(N) \end{bmatrix} \tag{10}$$

and

$$P^T(t) = (p_1(t), ..., p_M(t)) \tag{11}$$

Assumptions for the regression equation (9) are now stated.

Assumption 1

$\xi(t)$ is a zero mean white sequence and is uncorrelated with $p_i(t)$, $i = 1, ..., M$.

Assumption 2

All stochastic processes involved are ergodic.

Assumption 3

$\mathbf{P}^T\mathbf{P}$ is positive definite.

Assumption 1 is to ensure unbiased estimates. Assumption 2 is reasonable and is introduced so that ensemble averages may be replaced by time averages over one realization. $\mathbf{P}^T\mathbf{P}$ is at least positive semidefinite. If $\mathbf{P}^T\mathbf{P}$ is singular or ill-conditioned there are many numerical methods to deal with it and Assumption 3 can be relaxed. For the time being, however, let us assume that Assumption 3 is satisfied.

Because $\mathbf{P}^T\mathbf{P}$ is symmetric positive definite it can be decomposed into

$$\mathbf{P}^T\mathbf{P} = \mathbf{A}^T\mathbf{D}\mathbf{A} \tag{12}$$

where $\mathbf{A}$ is an upper triangular matrix with unity diagonal elements

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1M} \\ & 1 & \alpha_{23} & \cdots & \alpha_{2M} \\ & & \ddots & & \vdots \\ & & & 1 & \alpha_{M-1M} \\ & & & & 1 \end{bmatrix} \tag{13}$$

and $\mathbf{D}$ is a diagonal matrix with positive diagonal elements. Now

$$\mathbf{Z} = \mathbf{P}(\mathbf{A}^{-1}\mathbf{A})\Theta + \Xi \tag{14}$$

or

$$\mathbf{Z} = \mathbf{W}\mathbf{g} + \Xi \tag{15}$$

where

$$\left. \begin{matrix} \mathbf{W} = \mathbf{PA}^{-1} \\ \mathbf{g} = \mathbf{A}\Theta \end{matrix} \right\} \tag{16}$$

The matrix

$$\mathbf{W} = \begin{bmatrix} w_1(1) & \cdots & w_M(1) \\ \vdots & & \vdots \\ w_1(N) & \cdots & w_M(N) \end{bmatrix} \tag{17}$$

is an orthogonal matrix because

$$\mathbf{W}^T\mathbf{W} = (\mathbf{PA}^{-1})^T(\mathbf{PA}^{-1}) = \mathbf{D} = \mathrm{Diag}\left[\sum_{t=1}^{N} w_1^2(t), ..., \sum_{t=1}^{N} w_M^2(t) \right] \tag{18}$$

Auxiliary regressors $w_i(t)$ can be obtained recursively from

$$\mathbf{W} = \mathbf{P} - \mathbf{W}(\mathbf{A} - \mathbf{I}) \tag{19}$$

that is

$$\left.\begin{aligned}
w_1(t) &= p_1(t) \\
w_k(t) &= p_k(t) - \sum_{i=1}^{k-1} \alpha_{ik} w_i(t), \quad k = 2, \ldots, M
\end{aligned}\right\} \tag{20}$$

The upper triangular matrix $\mathbf{A}$ satisfies

$$\mathbf{A} = \mathbf{D}^{-1} \mathbf{W}^T \mathbf{P} \tag{21}$$

therefore

$$\alpha_{ik} = \frac{\displaystyle\sum_{t=1}^{N} w_i(t) p_k(t)}{\displaystyle\sum_{t=1}^{N} w_i^2(t)}, \quad k = 2, \ldots, M \quad \text{and} \quad i < k \tag{22}$$

The auxiliary parameter vector

$$\mathbf{g} = \begin{bmatrix} g_1 \\ \vdots \\ g_M \end{bmatrix} \tag{23}$$

satisfies

$$\mathbf{g} = \mathbf{D}^{-1} \mathbf{W}^T \mathbf{Z} - \mathbf{D}^{-1} \mathbf{W}^T \Xi \tag{24}$$

so that the estimated $\mathbf{g}$ is given by

$$\hat{\mathbf{g}} = \mathbf{D}^{-1} \mathbf{W}^T \mathbf{Z} \tag{25}$$

or

$$\hat{g}_i = \frac{\displaystyle\sum_{t=1}^{N} w_i(t) z(t)}{\displaystyle\sum_{t=1}^{N} w_i^2(t)}, \quad i = 1, \ldots, M \tag{26}$$

The estimates of the original parameters can be computed from

$$\Theta = \hat{\mathbf{g}} - (\mathbf{A} - \mathbf{I}) \Theta \tag{27}$$

that is

$$\left.\begin{aligned}
\theta_M &= \hat{g}_M \\
\theta_i &= \hat{g}_i - \sum_{k=i+1}^{M} \alpha_{ik} \theta_k, \quad i = M-1, \ldots, 1
\end{aligned}\right\} \tag{28}$$

The whole procedure can be summarized as follows.

(a) *Step* 1. $\alpha_{11} = 1$, $w_1(t) = p_1(t)$, and $\hat{g}_1 = \dfrac{\displaystyle\sum_{t=1}^{N} w_1(t) z(t)}{\displaystyle\sum_{t=1}^{N} w_1^2(t)}$.

$$Step\ k. \quad k = 2, ..., M: \alpha_{ik} = \frac{\sum\limits_{t=1}^{N} w_i(t) p_k(t)}{\sum\limits_{t=1}^{N} w_i^2(t)}, \quad i = 1, ..., k-1, \quad \alpha_{kk} = 1$$

$$w_k(t) = p_k(t) - \sum_{i=1}^{k-1} \alpha_{ik} w_i(t), \quad \text{and} \quad \hat{g}_k = \frac{\sum\limits_{t=1}^{N} w_k(t) z(t)}{\sum\limits_{t=1}^{N} w_k^2(t)}.$$

(b) Compute $\hat{\theta}_i$ backward using (28).

The relationship between this orthogonal least squares estimator and other least squares algorithms has been discussed by Chen *et al.* (1989). Here some properties of the orthogonal estimator are given.

Property 1

The auxiliary regressors are orthogonal, that is,

$$\sum_{t=1}^{N} w_i(t) w_j(t) = 0, \quad i \neq j \tag{29}$$

This is obvious from (18).

Property 2

The estimate $\hat{\Theta}$ is equivalent to that obtained by solving the normal equation directly

$$(P^T P)\hat{\Theta} = P^T Z \tag{30}$$

That is, the estimate $\hat{\Theta}$ is unbiased and the covariance of the estimate is given by

$$COV\ [\hat{\Theta}] = \sigma_\xi^2 (P^T P)^{-1} \tag{31}$$

where

$$\sigma_\xi^2 = E[\xi^2(t)] \tag{32}$$

This is expected since the orthogonal algorithm is just another way of solving the normal equation. In fact,

$$E[\hat{\Theta}] = E[A^{-1}\hat{g}] = E[A^{-1}D^{-1}W^T Z] = E[A^{-1}D^{-1}W^T Wg] + E[A^{-1}D^{-1}W^T \Xi]$$

$$= \Theta + E[A^{-1}D^{-1}W^T \Xi]$$

Using Assumption 1 it can be shown that $E[A^{-1}D^{-1}W^T\Xi] = 0$, therefore, $E[\hat{\Theta}] = \Theta$. Also

$$COV\ [\hat{\Theta}] = COV\ [A^{-1}\hat{g}] = A^{-1}\ COV\ [\hat{g}]A^{-T}$$

but

$$COV\ [\hat{g}] = E[(\hat{g} - g)(\hat{g} - g)^T] = E[D^{-1}W^T \Xi\Xi^T WD^{-1}]$$

$$= D^{-1}W^T \sigma_\xi^2 IWD^{-1} = \sigma_\xi^2 D^{-1}$$

and

$$A^{-1}D^{-1}A^{-T} = (P^T P)^{-1}$$

Property 3

The proportion of the dependent variable variance explained by $w_i(t)$ is

$$\frac{g_i^2 \sum_{t=1}^{N} w_i^2(t)}{\sum_{t=1}^{N} z^2(t)} \tag{33}$$

Proof

$$\mathbf{Z^T Z = g^T W^T W g + \Xi^T \Xi + g^T W^T \Xi + \Xi^T W g}$$

Assumptions 1 and 2 ensure $\mathbf{W^T \Xi} = 0$, and hence

$$\sum_{t=1}^{N} z^2(t) = \sum_{i=1}^{M} \left[g_i^2 \sum_{t=1}^{N} w_i^2(t) \right] + \sum_{t=1}^{N} \zeta^2(t)$$

or

$$\frac{1}{N} \sum_{t=1}^{N} z^2(t) = \sum_{i=1}^{M} \left[g_i^2 \frac{1}{N} \sum_{t=1}^{N} w_i^2(t) \right] + \sigma_\zeta^2$$

Therefore the contribution to the dependent variable variance by the auxiliary regressor $w_i(t)$ is

$$g_i^2 \frac{1}{N} \sum_{t=1}^{N} w_i^2(t) \qquad \qquad \Box$$

Property 3 can be used as a criterion for model structure selection. Define the error reduction ratio due to the *i*th term as

$$[\text{err}]_i = \frac{\hat{g}_i^2 \sum_{t=1}^{N} w_i^2(t)}{\sum_{t=1}^{N} z^2(t)} \tag{34}$$

and give a threshold value ρ. Starting from $k = 1$, if $[\text{err}]_k < \rho$, $w_k(t)$ is deleted — and consequently $p_k(t)$. Notice that for $k > 1$ removing $w_k(t)$ does not influence the existing $w_i(t)$, $i < k$. Rename $p_{k+1}(t), ..., p_M(t)$ as $p_k(t), ..., p_{M-1}(t)$ and continue the procedure until all insignificant terms have been removed. Any numerical ill-conditioning can be avoided by simply deleting $w_k(t)$ if $\sum_{t=1}^{N} w_k^2(t)$ is less than a predetermined threshold.

Notice that $\sum_{t=1}^{N} w_k^2(t) = 0$ means that $p_k(t)$ is a linear combination of $p_1(t)$ to $p_{k-1}(t)$— see (20) and (22)—and $\mathbf{P^T P}$ is singular. It is thus seen that Assumption 3 was only included for the benefit of the derivation and is not really required. This is another advantage compared with a direct solution of the normal equation.

There is however one problem. The value of $[\text{err}]_i$ may depend upon the order in which the term $p_i(t)$ enters the equation. As a result, simply orthogonalizing the $p_i(t)$ into the orthogonal equation in the order in which they happened to be written down in (8) may produce the wrong information regarding their significance using the criterion (34). This difficulty can, however, be overcome by a forward-regression procedure proposed by Billings *et al.* (1988 c).

Instead of forward deleting terms, disregard the orders that the $p_i(t)$ enter (8) and forward add terms as follows.

In the first step, all the $p_i(t)$, $i = 1, ..., M$ are considered as possible candidates for $w_1(t)$. For $i = 1, ..., M$, calculate

$$w_1^{(i)}(t) = p_i(t), \quad \hat{g}_1^{(i)} = \frac{\sum_{t=1}^{N} w_1^{(i)}(t)z(t)}{\sum_{t=1}^{N} (w_1^{(i)}(t))^2}, \quad [\text{err}]_1^{(i)} = \frac{(\hat{g}_1^{(i)})^2 \sum_{t=1}^{N} (w_1^{(i)}(t))^2}{\sum_{t=1}^{N} z^2(t)}$$

Find the maximum of $[\text{err}]_1^{(i)}$, say, $[\text{err}]_1^{(j)} = \max \{[\text{err}]_1^{(i)}, 1 \leqslant i \leqslant M\}$. Then the first term $w_1(t) = w_1^{(j)}(t)$, $(p_j(t))$ is selected with $\hat{g}_1 = \hat{g}_1^{(j)}$ and $[\text{err}]_1 = [\text{err}]_1^{(j)}$.

In the second step, all the $p_i(t)$, $i = 1, ..., M$, $i \neq j$ are considered as possible candidates for $w_2(t)$. For $i = 1, ..., M$, $i \neq j$, calculate

$$w_2^{(i)}(t) = p_i(t) - \alpha_{12}^{(i)}w_1(t), \quad \hat{g}_2^{(i)} = \frac{\sum_{t=1}^{N} w_2^{(i)}(t)z(t)}{\sum_{t=1}^{N} (w_2^{(i)}(t))^2}, \quad [\text{err}]_2^{(i)} = \frac{(\hat{g}_2^{(i)})^2 \sum_{t=1}^{N} (w_2^{(i)}(t))^2}{\sum_{t=1}^{N} z^2(t)}$$

where

$$\alpha_{12}^{(i)} = \frac{\sum_{t=1}^{N} w_1(t)p_i(t)}{\sum_{t=1}^{N} w_1^2(t)}$$

Find the maximum of $[\text{err}]_2^{(i)}$, say, $[\text{err}]_2^{(k)} = \max \{[\text{err}]_2^{(i)}, 1 \leqslant i \leqslant M, i \neq j\}$. Then the second term $w_2(t) = w_2^{(k)}(t) = p_k(t) - \alpha_{12}w_1(t)$, $(p_k(t))$ is selected with $\alpha_{12} = \alpha_{12}^{(k)}$, $\hat{g}_2 = \hat{g}_2^{(k)}$ and $[\text{err}]_2 = [\text{err}]_2^{(k)}$.

The procedure is terminated at the M_sth step either when

$$1 - \sum_{i=1}^{M_s} [\text{err}]_i < a \text{ desired tolerance}, \quad M_s < M \tag{35}$$

or when $M_s = M$. Notice that the remainder $1 - \sum_{i=1}^{M_s} [\text{err}]_i$ is the proportion of the unexplained dependent variable variance and $100 \left(1 - \sum_{i=1}^{M_s} [\text{err}]_i\right)$ expresses this as a percentage. From the selected orthogonal equation

$$z(t) = \sum_{i=1}^{M_s} w_i(t)g_i + \xi(t) \tag{36}$$

it is straightforward to calculate the corresponding M_s parameters θ_i. Ill-conditioning can easily be avoided. If $\sum_{t=1}^{N} (w_k^{(i)}(t))^2$ is less than a predetermined threshold $p_i(t)$ will not be considered as a candidate for $w_k(t)$. Notice that the $[\text{err}]_i$ in (35) are not necessarily in descending order and it is possible that $[\text{err}]_i < [\text{err}]_j$, $i < j$.

Finally it should be emphasized that the forward-regression orthogonal procedure maximizes the increment to explained variance, not explained variance itself, and in general is a suboptimal method in the sense of maximizing explained variance.

4. Applications to identification of non-linear systems

This section considers the application of the forward-regression orthogonal estimator to the identification of MIMO non-linear systems which can be represented by the NARMAX model (7). Some possible ways of simplifying the identification procedure are also discussed.

4.1. *Application to NARX model*

A special case of the general MIMO NARMAX model (7) is

$$y_i(t) = f_i^p(y_1(t-1), ..., y_1(t-n_{y_1}^i), ..., y_m(t-1), ..., y_m(t-n_{y_m}^i), u_1(t-1) ...,$$

$$u_1(t-n_{n_1}^i), ..., u_r(t-1), ..., u_r(t-n_{u_r}^i)) + e_i(t), \quad i = 1, ..., m \tag{37}$$

This non-linear model will be referred to as the NARX (Nonlinear AutoRegressive with eXogenous inputs) model.

Each submodel in (37) is linear-in-the-parameters and is in the form of the regression function (8). Application of the forward-regression orthogonal estimator to this model is therefore straightforward. The identification of any subsystem is decoupled from the others, and for $i = 1$ to m, the estimator can easily be applied to determine the ith submodel structure and parameter values.

4.2. *Application to NARMAX model*

For the general NARMAX model (7), delayed noise terms $e(t-j)$ are included in the model and these have to be estimated using the prediction errors or residuals $\varepsilon(t-j)$. The orthogonal property of the estimator ensures that the selection of the process and noise model parameters can be decoupled. Significant terms in each row of the process model are selected initially. Which terms are included in these process models will not be affected by whatever noise models are produced later because of the orthogonal property. Initial residuals are then computed based on the process models and the structure or which terms to include in the noise models can then be selected. A revised residual sequence is calculated and improved noise models are determined. A few iterations are often enough to determine final noise models. The detailed procedure is as follows.

Step 1. For $i = 1$ to m, use the forward-regression orthogonal estimator to select terms in the process model $f_i^p(\cdot)$. The selection is terminated when

$$1 - \sum_{j=1}^{n_{pi}} [\text{err}]_j < \rho_{pi} \tag{38}$$

where n_{pi} is the number of significant terms selected for $f_i^p(\cdot)$ and ρ_{pi} is the tolerance for the ith row of the process model. The ith orthogonal process model is $\sum_{j=1}^{n_{pi}} w_{ij}(t)\hat{g}_{ij}$.

Step 2. Set $k = 0$ and calculate the initial residuals

$$\varepsilon_i^{(k)}(t) = y_i(t) - \sum_{j=1}^{n_{pi}} w_{ij}^{(k)}(t)\hat{g}_{ij}^{(k)}, \quad i = 1, ..., m \tag{39}$$

Step 3. Set $k = k + 1$. For $i = 1$ to m, use the forward-regression orthogonal estimator

to select terms in the noise model $f_i^n(\cdot)$. The selection is terminated when

$$1 - \sum_{j=1}^{n_{pi}+n_{ni}^{(k)}} [\text{err}]_j < \rho_{ni} \qquad (40)$$

where $n_{ni}^{(k)}$ is the number of significant terms selected for $f_i^n(\cdot)$ at the kth iteration and $\rho_{ni} < \rho_{pi}$.

Step 4. For $i = 1$ to m, calculate each submodel parameter $\theta_{ij}^{(k)}$, $j = 1, ..., n_{pi} + n_{ni}^{(k)}$ and compute $\varepsilon_i^{(k)}(t)$ recursively using

$$\varepsilon_i^{(k)}(t) = y_i(t) - \sum_{j=1}^{n_{pi}+n_{ni}^{(k)}} \theta_{ij}^{(k)} x_{ij}(t) \qquad (41)$$

Step 5. Test for convergence. If the convergence criterion is satisfied, stop. Otherwise go to Step 3.

Notice the difference in the computation of $\varepsilon^{(0)}(t)$ and $\varepsilon^{(k)}(t)$ for $k > 0$. If the latter were computed using

$$\varepsilon_i^{(k)}(t) = y_i(t) - \sum_{j=1}^{n_{pi}+n_{ni}^{(k)}} w_{ij}^{(k)}(t)\hat{g}_{ij}^{(k)}, \quad i = 1, ..., m \qquad (42)$$

then because some of the $w_{ij}^{(k)}(t)$ were calculated based on the old residuals, the fresh information obtained at the kth iteration would be wasted. Step 4 is obviously a better way to compute $\varepsilon^{(k)}(t)$, for $k > 0$. A possible convergence criterion would be

$$\|\varepsilon^{(k)}(t) - \varepsilon^{(k-1)}(t)\| < a \text{ tolerance} \qquad (43)$$

Numerous tests have shown that less than 10 iterations, typically 3–4 iterations, are sufficient for the algorithm to converge. In practice repeating Steps 3 and 4, 4–5 times is usually adequate. At each iteration, the selection of each noise model term is performed from all the original candidate terms for the noise model. Notice that no matter how many subsystems or how many terms are included in the model set the estimation remains very simple and consists of re-entering a routine of computing $\alpha_{jk}^{(i)}$, $w_k^{(i)}(t)$, $\hat{g}_k^{(i)}$ and $[\text{err}]_k^{(i)}$ for each possible term. This is a considerable advantage of this method and ensures that the coding of the algorithm is very straightforward and upwardly extendable.

4.3. *Some simplifying procedures*

Experience has shown that non-linear terms are often the combinations of linear terms in non-linear models. The forward-regression orthogonal estimator may therefore be used to identify the best linear model first, and then to extend the model set to include all non-linear combinations of these linear terms. Such a model set will be significantly smaller than a full non-linear model set and hence the saving in computational time will be substantial. There is, however, no guarantee that the non-linear model identified in this way will be as good as that selected from the full model set.

Alternatively, the noise model structure may be updated only once. Because $\varepsilon^{(0)}(t)$ is not generally white the model determined in such a simplified way may be slightly biased. A better way would be to select smaller ρ_{ni} at the first iteration to force more noise terms into the model than are actually needed, and then to use these terms as the noise model set selecting the noise models from within this set.

As shown in § 4.1, identifying NARX models is much simpler. Given a non-linear system with unknown structure, the best NARX model can be estimated and noise models can be added only when the model validity tests indicate that the NARX is inadequate.

5. Model validation

If the model set does not include all the significant terms of the true system, the final model selected from this set will not be a good representation of the system because some significant terms will be missing from the model. Model validation should indicate when such a deficient situation occurs.

Let $\hat{f}(\,\cdot\,)$ be an estimated model of the system $f(\,\cdot\,)$ and let the residuals $\varepsilon(t)$ be computed from

$$\varepsilon(t) = y(t) - \hat{f}(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u), \varepsilon(t-1), ..., \varepsilon(t-n_\varepsilon)) \quad (44)$$

If the model structure and parameter values are correct $\varepsilon(t)$ will be unpredictable from (uncorrelated with) all linear and non-linear combinations of past inputs and outputs. For SISO non-linear systems, this can be tested by means of the following correlation functions (Billings and Voon 1986 a, Billings and Chen 1988) under some mild assumptions:

$$\left.\begin{array}{ll} \Phi_{\varepsilon\varepsilon}(\tau) & \tau \neq 0 \\[4pt] \Phi_{u\varepsilon}(\tau) & \forall \tau \\[4pt] \Phi_{\varepsilon(\varepsilon u)}(\tau) & \tau \geqslant 0 \\[4pt] \Phi_{u'^2\varepsilon}(\tau) & \forall \tau \\[4pt] \Phi_{u'^2\varepsilon^2}(\tau) & \forall \tau \end{array}\right\} \quad (45)$$

where $u^{2'}(t) = u^2(t) - \overline{u^2(t)}$ and the bar indicates time averaging. If these correlation functions fall within the (95 %) confidence interval $\pm 1.96 N^{-1/2}$, the model is regarded as adequate. Notice that the linear covariance tests $\Phi_{\varepsilon\varepsilon}(\tau)$ and $\Phi_{u\varepsilon}(\tau)$ are certainly not sufficient. These correlation tests can be extended to validate MIMO non-linear models

$$\left.\begin{array}{llll} \Phi_{\varepsilon_i\varepsilon_j}(\tau) & \tau \neq 0 & i = 1, ..., m \quad \text{and} \quad j = i, ..., m \\[6pt] \Phi_{u_i\varepsilon_j}(\tau) & \forall \tau & i = 1, ..., r \quad \text{and} \quad j = 1, ..., m \\[6pt] \Phi_{\varepsilon_i(\varepsilon_j u_k)}(\tau) & \tau \geqslant 0 & i = 1, ..., m, \quad j = i, ..., m \quad \text{and} \quad k = 1, ..., r \\[6pt] \Phi_{(u_i u_j)'\varepsilon_k}(\tau) & \forall \tau & i = 1, ..., r, \quad j = i, ..., r \quad \text{and} \quad k = 1, ..., m \\[6pt] \Phi_{(u_i u_j)'(\varepsilon_k \varepsilon_l)}(\tau) & \forall \tau & i = 1, ..., r, \quad j = i, ..., r, \quad k = 1, ..., m \quad \text{and} \\[6pt] & & l = k, ..., m \end{array}\right\} \quad (46)$$

Alternatively, the chi-square statistical test introduced by Bohlin (1978) and adopted to the non-linear case by Leontaritis and Billings (1987) can be used to validate MIMO non-linear models. The problem of validating an estimated model is formulated as one of testing the null hypothesis H_0: that the input–output data were generated by the model. The procedure involves three steps. The first one is to find a statistic η which is a function of the available data and whose distribution is known if

H_0 is true. The second step is to define a domain D_α such that

$$\text{Prob}\{\eta \notin D_\alpha | H_0\} = \alpha \tag{47}$$

where α is called the significant level of the test. The third step is to reject H_0 if $\eta \notin D_\alpha$. The required statistic can be constructed as follows (Bohlin 1978). Define

$$y^t = \begin{bmatrix} y(1) \\ \vdots \\ y(t) \end{bmatrix}, \quad \varepsilon^t = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(t) \end{bmatrix}, \quad u^t = \begin{bmatrix} u(1) \\ \vdots \\ u(t) \end{bmatrix} \quad \text{and} \quad x^t = \begin{bmatrix} y^{t-1} \\ \varepsilon^{t-1} \\ u^{t-1} \end{bmatrix} \tag{48}$$

Let S be a square root of the covariance matrix of the residuals, that is,

$$\mathbf{S}^T\mathbf{S} = E[\varepsilon(t)\varepsilon^T(t)|x^t] \tag{49}$$

The normalized residuals are given by

$$\beta(t) = \mathbf{S}^{-T}\varepsilon(t) \tag{50}$$

Introduce a matrix-valued function of x^t as

$$\Omega(t) = \Omega(x^t) \tag{51}$$

where $\Omega(t)$ has dimension $s \times m$. It is assumed that $\Omega(t)$ satisfies the law of large numbers for the time average

$$\frac{1}{N}\sum_{t=1}^{N} \Omega(t)\Omega^T(t) = \Gamma^T\Gamma \tag{52}$$

and the central limit theorem holds for

$$\mu = \frac{1}{N}\sum_{t=1}^{N} \Omega(t)\beta(t) \tag{53}$$

Then the random vector μ is asymptotically normal with zero mean and covariance equal to $\Gamma^T\Gamma \, N^{-1}$ if H_0 is true (Bohlin 1978). The normalized random vector

$$\bar{\mu} = N^{-1/2}\Gamma^{-T}\mu \tag{54}$$

is thus asymptotically normal with zero mean and unit covariance if H_0 is true. The random variable

$$\eta = \bar{\mu}^T\bar{\mu} \tag{55}$$

is therefore asymptotically chi-square distributed with s degrees of freedom under the null hypothesis, and can be used as the desired statistic. For a specified level of significance α, the domain D_α where the null hypothesis is accepted can be chosen as

$$D_\alpha = \{\eta : \eta < \chi_s^2(\alpha)\} \tag{56}$$

where $\chi_s^2(\alpha)$ is the value of the chi-square variable with s degrees of freedom and given risk α.

Model validation using the above chi-square statistical test consists of the following steps: specifying a risk level α, choosing the matrix function $\Omega(t)$, computing the value of η and testing whether $\eta \in D_\alpha$. It is apparent that for different matrix functions $\Omega(t)$ the tests all have the same risk α of rejecting a model when it is actually valid. However, they do not have the same probability of rejecting a model when it is not valid. The optimal $\Omega(t)$ can be designed such that the test has asymptotic local

maximum power (Bohlin 1978). Such an optimal $\Omega(t)$ requires, however, knowledge of the derivatives of $\varepsilon(t)$ with respect to the parameters and is very expensive to compute for the MIMO non-linear model (7). A simple and convenient choice for $\Omega(t)$ is

$$\Omega(t) = \begin{bmatrix} \omega^T(t) \\ \omega^T(t-1) \\ \vdots \\ \omega^T(t-s+1) \end{bmatrix} \tag{57}$$

where

$$\omega^T(t) = (\omega_1(t), ..., \omega_m(t)) \tag{58}$$

and $\omega_i(t)$, $i = 1, ..., m$ are monomials of x^t. The disadvantage of this test is that several different types of $\omega^T(t)$ must be tried before any confidence that the model has been properly validated can be established. Whilst the correlation tests of (45) overcome this difficulty and work well for SISO systems the power of the tests appears to be diminished when applied in the multivariable case and this requires further study.

6. Simulation study

Example 1

This is a two-input two-output data set collected from a 50 MW turbo-alternator operating in parallel with an interconnected system having a capacity of approximately 5000 MW (Jenkins and Watts 1969; Appendix A11.3). The input $u_1(t)$ was the in-phase current deviation and $u_2(t)$ was the out-of-phase current deviation. The output $y_1(t)$ was the voltage deviation and $y_2(t)$ was the frequency deviation. The data set is illustrated in Fig. 1.

A linear model $(L_1 = L_2 = 1)$ with $n_{y_j}^i = n_{u_j}^i = 20$, $n_{e_j}^i = 0$, $i = 1, 2$ and $j = 1, 2$ was used initially to fit the data. The full model set consisted of 162 terms. The first 20 points of the data were used as the initial values and therefore only 80 points of the data were actually used in the identification. The forward-regression orthogonal estimator selected a model with 15 terms shown in Table 1. An idea of the efficiency of the algorithm can be judged by noting that it took less than 30 s on a Sun 3/50 workstation to produce the model given in Table 1. The response of the model is plotted in Fig. 2 where the one-step ahead prediction is defined as

$$\hat{y}(t) = \hat{f}(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u), \varepsilon(t-1), ..., \varepsilon(t-n_e)) \tag{59}$$

the (deterministic) model predicted output is given by

$$y_d(t) = \hat{f}(y_d(t-1), ..., y_d(t-n_y), u(t-1), ..., u(t-n_u), 0, ..., 0) \tag{60}$$

and the deterministic prediction error

$$\varepsilon_d(t) = y(t) - y_d(t) \tag{61}$$

The correlation tests and chi-square tests all confirm that the model is adequate. Figures 3 and 4 show the correlation tests and some of the chi-square tests used for the model validation respectively. Since the model validity tests show that the linear model is adequate the non-linear analysis was not pursued.

S. A. Billings et al.

output y₁

-1.7E 0

-2.5E 0 100

output y₂

5.2E 0

4.7E 0 100

input u₁

1.1E 0

1.5E -1 100

input u₂

1.6E 0

-1.0E 0 100
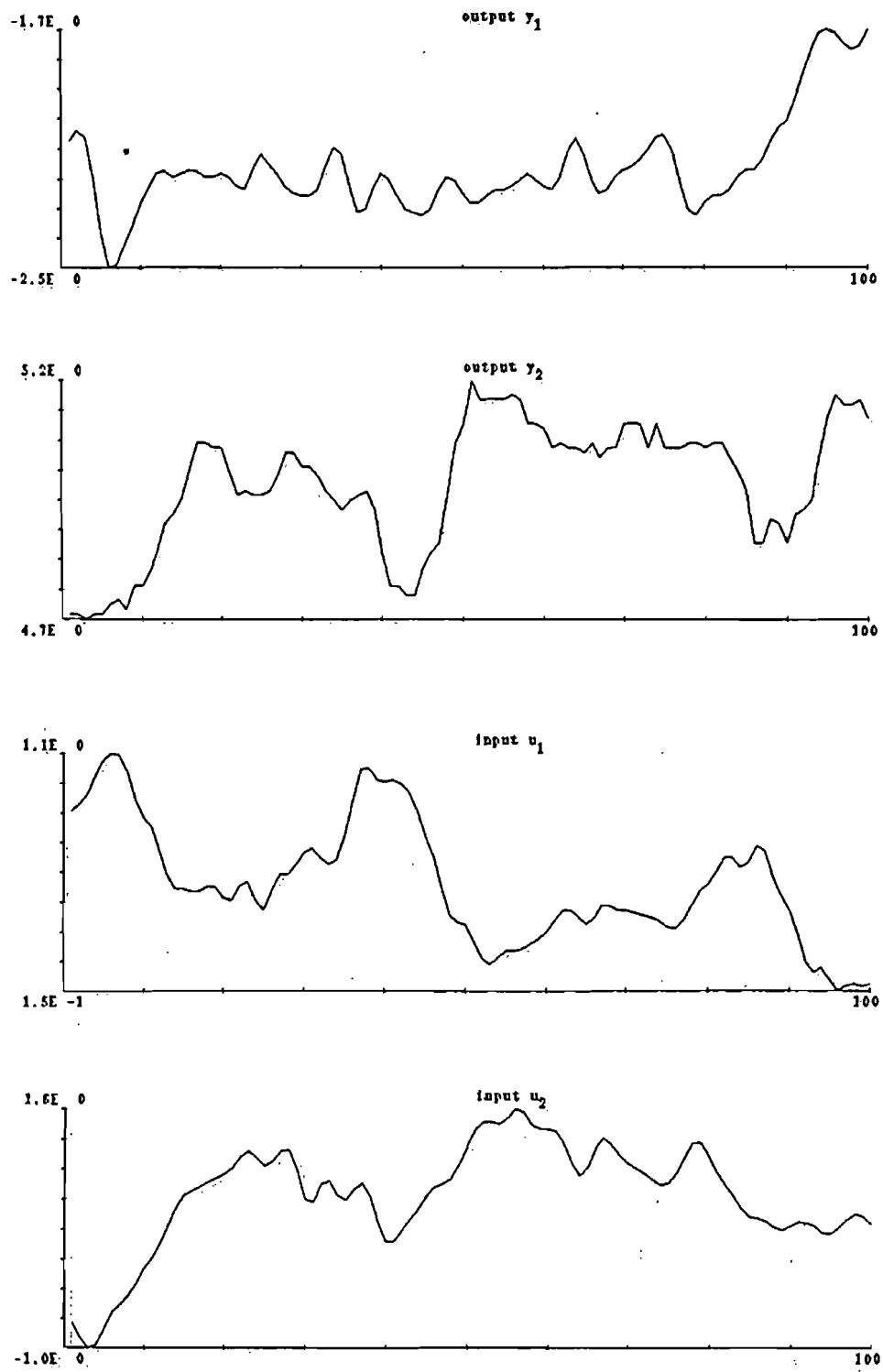
Figure 1.   Inputs and outputs of Example 1.

Subsystem	Terms	Estimates	$[\text{err}]_i$	Standard deviations
1	$y_1(t-1)$	0·22535E + 1	0·99965E + 0	0·10500E + 0
	$y_1(t-2)$	−0·24313E + 1	0·12983E − 3	0·24128E + 0
	$y_1(t-3)$	0·14672E + 1	0·11256E − 3	0·27117E + 0
	$y_1(t-4)$	−0·43287E + 0	0·12371E − 4	0·15491E + 0
	$y_1(t-6)$	0·83902E − 1	0·18386E − 5	0·47686E − 1
	$y_1(t-10)$	−0·13045E − 1	0·21199E − 5	0·33860E − 1
	$y_1(t-20)$	0·41910E − 1	0·17841E − 5	0·23195E − 1
	$u_2(t-2)$	−0·33549E − 1	0·12604E − 5	0·13320E − 1
	$u_1(t-3)$	−0·56494E − 1	0·56199E − 5	0·24734E − 1
2	$y_2(t-1)$	0·67664E + 0	0·99994E + 0	0·52896E − 1
	$u_1(t-8)$	0·21513E + 0	0·99848E − 5	0·21690E − 1
	$u_1(t-3)$	−0·25209E + 0	0·11360E − 4	0·24057E − 1
	$y_2(t-11)$	0·14139E + 0	0·15998E − 4	0·30151E − 1
	$u_2(t-4)$	0·34826E − 1	0·21750E − 5	0·94300E − 2
	constant	0·90807E + 0	0·16729E − 5	0·35033E + 0

$\rho_{p1} = 0\cdot00009$
$\rho_{p2} = 0\cdot00002$ } covariance of residuals = $\begin{bmatrix} 0\cdot41051E-3 & -0\cdot44232E-4 \\ -0\cdot44232E-4 & 0\cdot50953E-3 \end{bmatrix}$

Table 1. Estimated model of Example 1.

Subsystem	Terms	Estimates	$[\text{err}]_i$	Standard deviations
1	$u_1(t-2)$	0·10172E + 1	0·66343E + 0	0·17078E − 1
	$y_1(t-1)$	0·50962E + 0	0·26657E + 0	0·16527E − 1
	$u_1(t-1)$	0·22921E + 0	0·25808E − 1	0·16528E − 1
	$u_2(t-1)$	−0·16081E − 1	0·21991E − 3	0·16671E − 1
	$y_2(t-2)$	0·16677E − 1	0·17319E − 3	0·84644E − 2
	constant	−0·79677E − 1	0·69261E − 3	0·28167E − 1
	$y_2(t-1)$	0·15069E − 1	0·29243E − 3	0·91441E − 2
	$y_1(t-2)$	−0·10137E − 1	0·20163E − 3	0·15993E − 1
	$u_2(t-2)$	−0·73543E − 2	0·34263E − 5	0·18533E − 1
	$e_1(t-1)$	0·29245E + 0	0·29667E − 2	0·60357E − 1
	$e_1(t-2)$	0·18426E + 0	0·12234E − 2	0·59296E − 1
	$e_2(t-2)$	0·22001E − 1	0·45035E − 4	0·37683E − 1
	$e_2(t-1)$	−0·20432E − 1	0·37422E − 4	0·37757E − 1
2	$y_2(t-2)$	0·93526E + 0	0·84592E + 0	0·85669E − 2
	$u_2(t-1)$	0·10188E + 1	0·10754E + 0	0·25993E − 1
	$u_2(t-2)$	0·46579E + 0	0·23791E − 1	0·26037E − 1
	$e_2(t-1)$	0·17536E + 0	0·67034E − 3	0·57984E − 1
	$e_1(t-2)$	0·11953E + 0	0·12337E − 3	0·91431E − 1
	$e_2(t-2)$	0·17879E − 1	0·68212E − 5	0·58613E − 1
	$e_1(t-1)$	−0·17508E − 1	0·26867E − 5	0·91353E − 1

$\rho_{p1} = 0\cdot040$ $\rho_{n1} = 0\cdot010$
$\rho_{p2} = 0\cdot030$ $\rho_{n2} = 0\cdot009$ } covariance of residuals = $\begin{bmatrix} 0\cdot80913E-1 & -0\cdot38546E-2 \\ -0\cdot38546E-2 & 0\cdot20089E+0 \end{bmatrix}$

Table 2. Linear model for Example 2.

S. A. Billings et al.

-1.7E 0   measured output y₁

-2.5E 0   100

-1.7E 0   one-step ahead prediction of y₁

-2.5E 0   100

5.6E -2   residual e₁

                                                              100

-5.6E -2

2.1E -1   deterministic prediction error e_d₁

                                                              100

-2.1E -1

model predicted output of y_1
.... system output y_1

measured output y_2

one-step ahead prediction of y_2
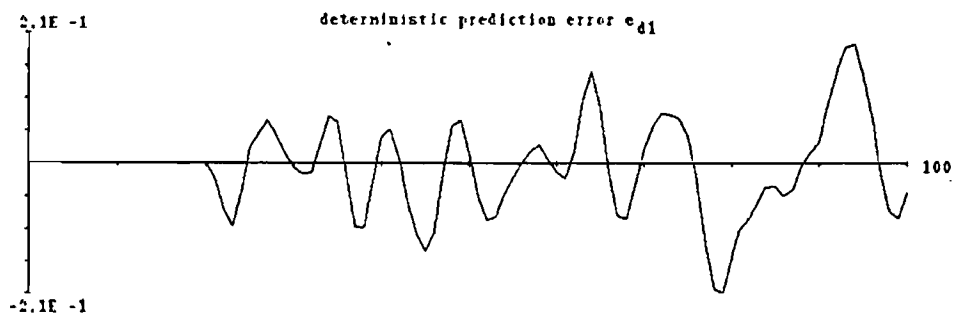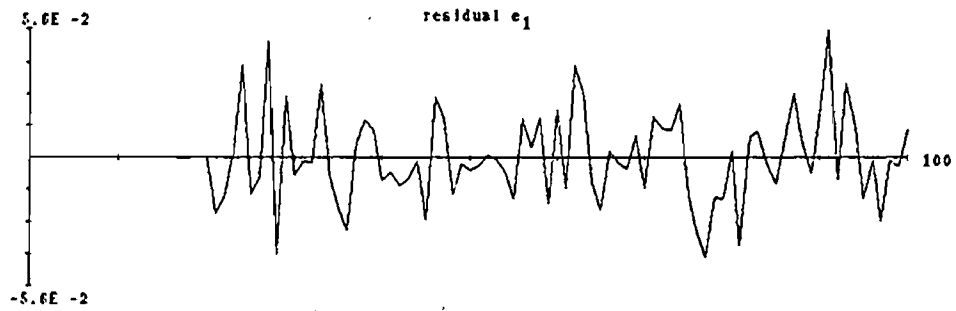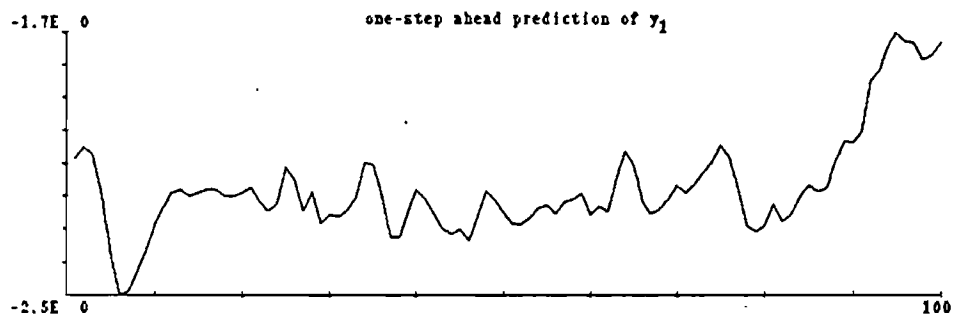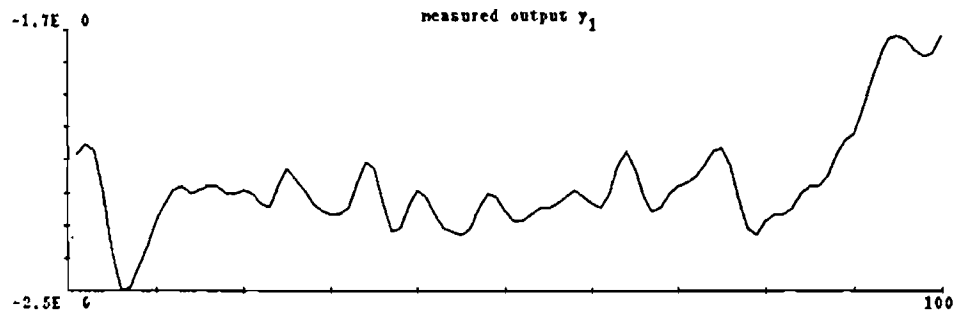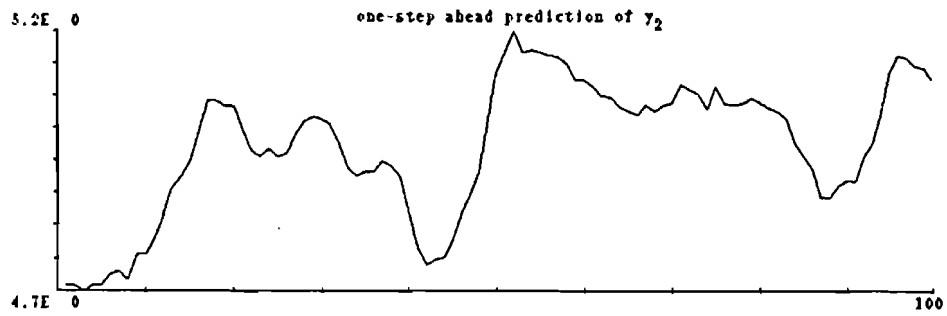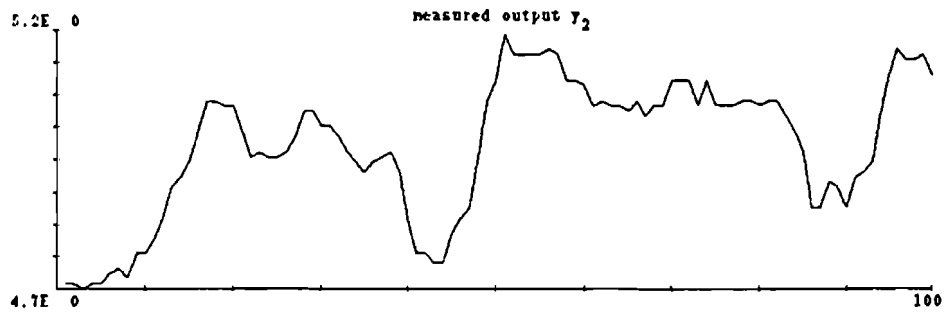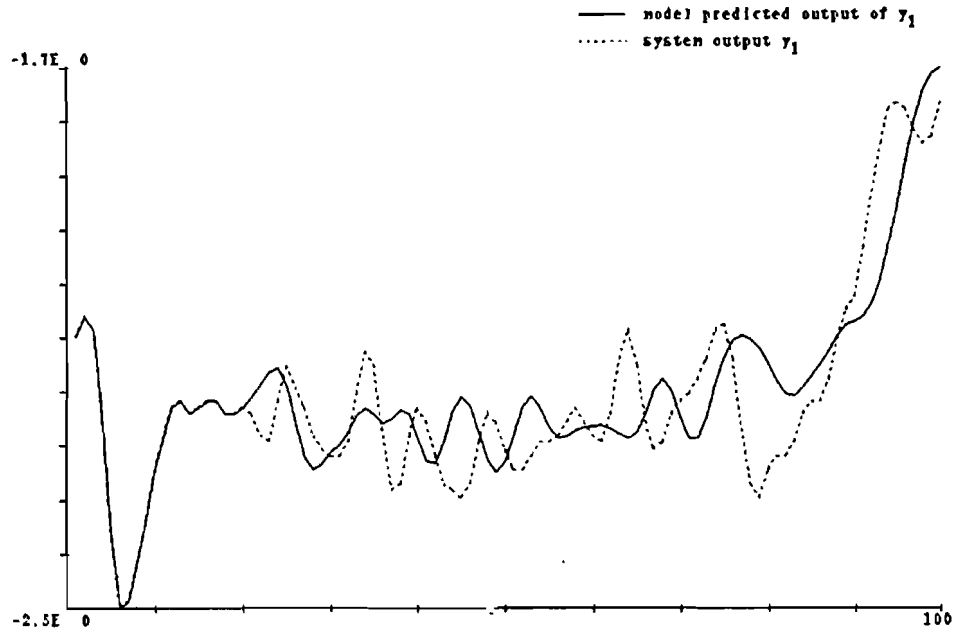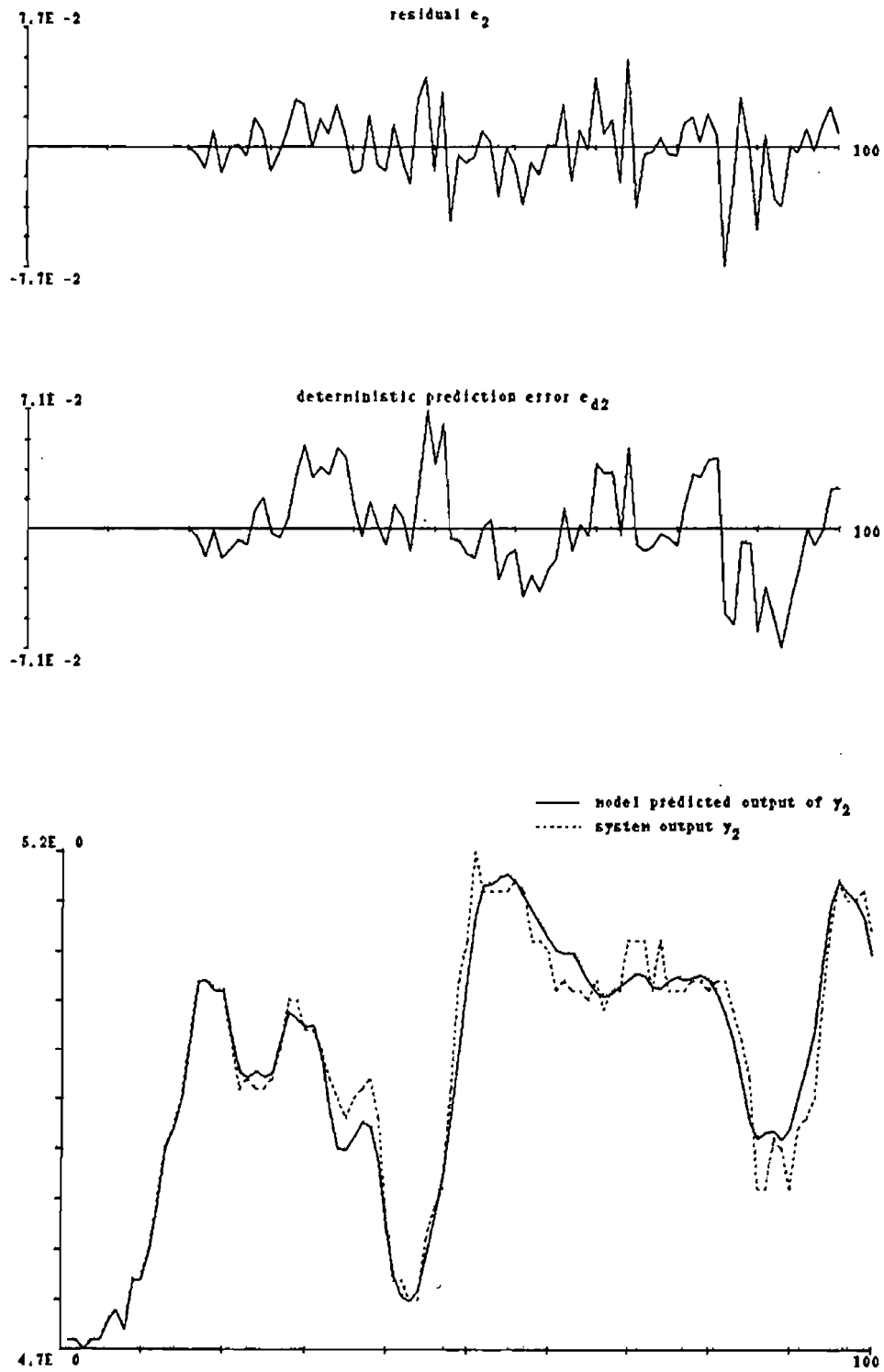
continued

Figure 2.   Model response of Example 1.
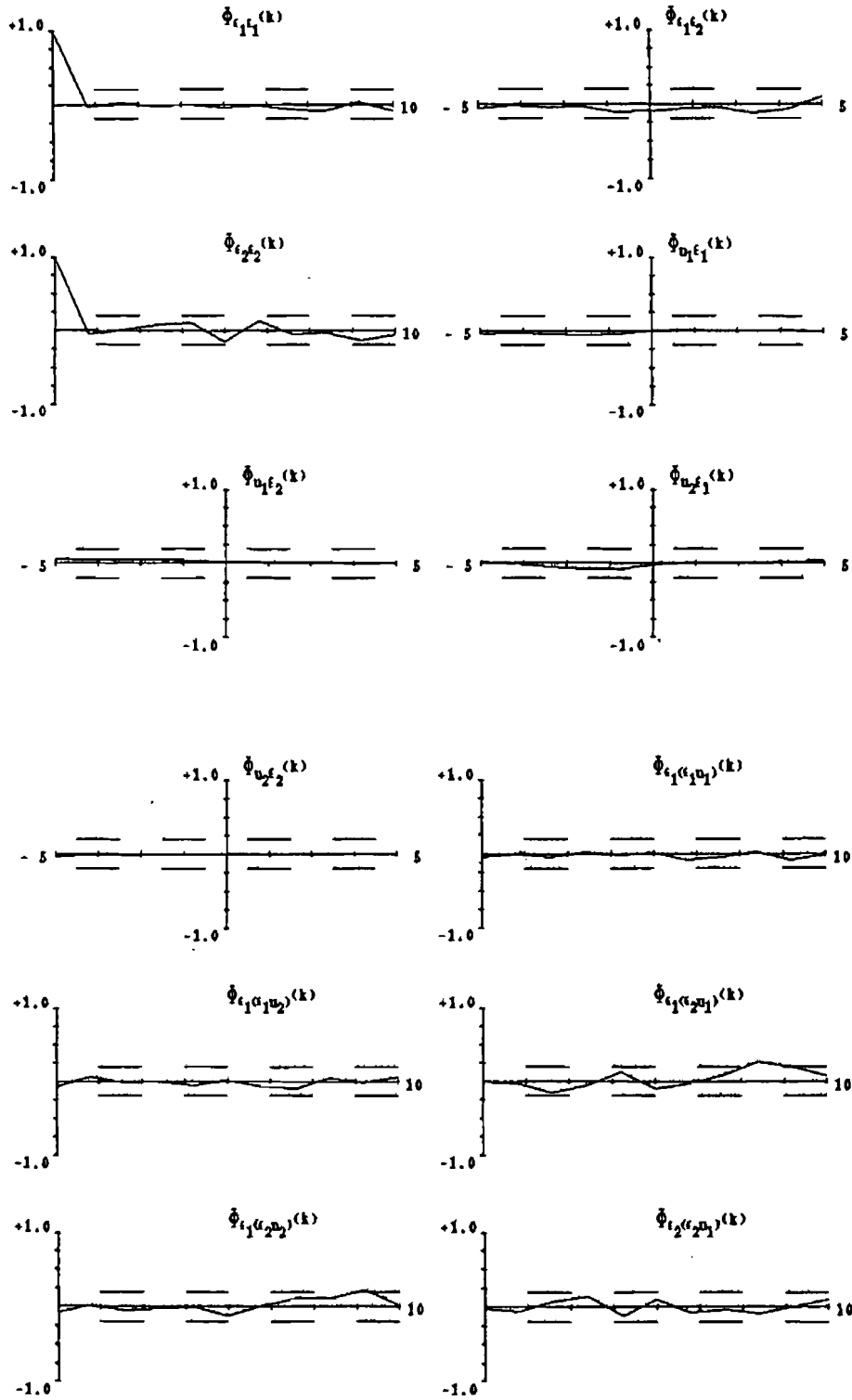
Figure 3.

Figure 3.
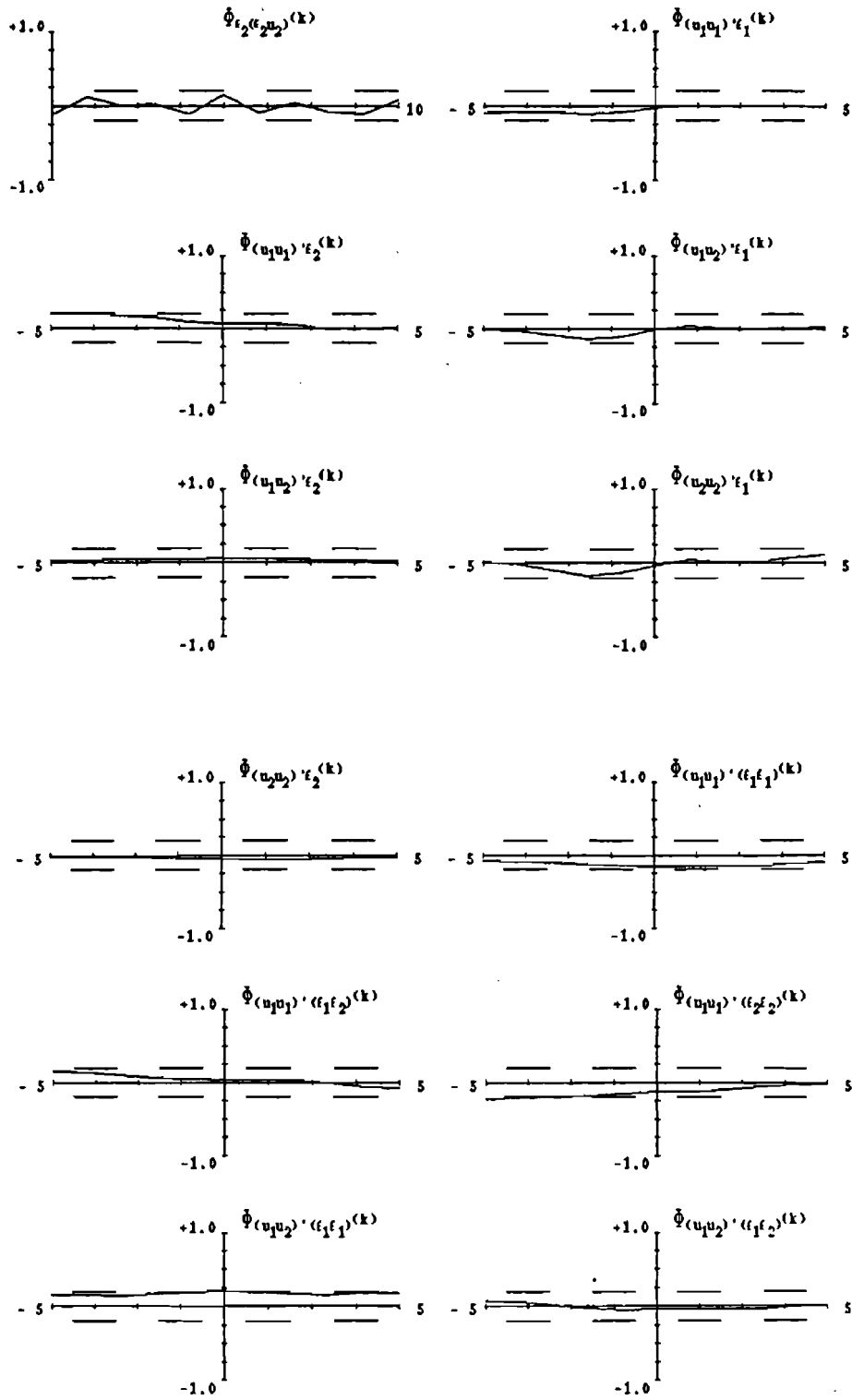
continued
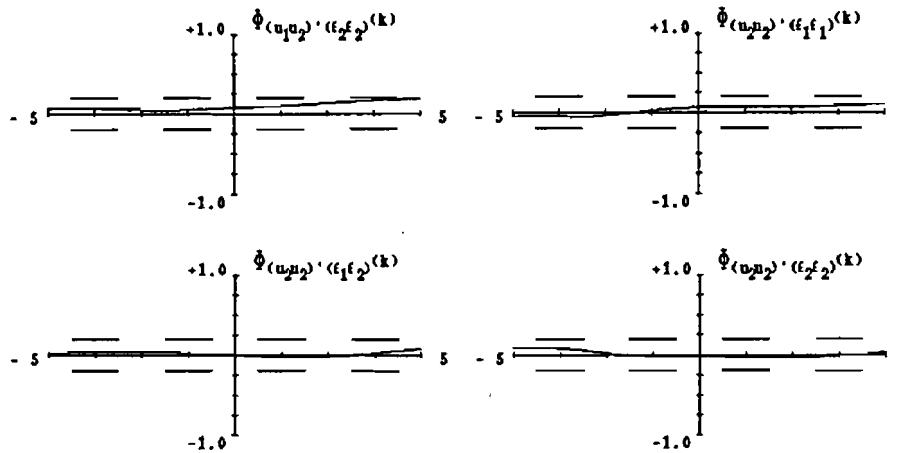
Figure 3.   Correlation tests of Example 1.

Example 2

This is a simulated two-input two-output system. The data was generated by

$$y_1(t) = 0{\cdot}5y_1(t-1) + u_1(t-2) + 0{\cdot}1y_2(t-1)u_1(t-1) + 0{\cdot}5e_1(t-1)$$

$$+ 0{\cdot}2y_1(t-2)e_1(t-2) + e_1(t)$$

$$y_2(t) = 0{\cdot}9y_2(t-2) + u_2(t-1) + 0{\cdot}2y_2(t-1)u_2(t-2) + 0{\cdot}5e_2(t-1)$$

$$+ 0{\cdot}1y_2(t-1)e_1(t-2) + e_2(t)$$

The system noise $e(t)$ was a gaussian white sequence with mean zero and covariance

$$\text{COV}\,[e(t)] = \begin{bmatrix} 0{\cdot}04 & 0{\cdot}0 \\ 0{\cdot}0 & 0{\cdot}04 \end{bmatrix}$$

$u_1(t)$ was a gaussian sequence with mean zero and variance $1{\cdot}0$, and $u_2(t)$ was an independent sequence of uniform distribution with mean zero and variance $1{\cdot}0$; $e(t)$ and $u(t)$ were independent. The inputs and outputs of the system are shown in Fig. 5.

First a linear model with $L_i = 1$, $n^i_{y_j} = n^i_{u_j} = n^i_{e_j} = 2$, $i = 1, 2$ and $j = 1, 2$ was used to fit the data. The model determined by the forward-regression orthogonal estimator is given in Table 2 where the number of iterations for updating the noise models was 4. The response of this linear model which are shown in Fig. 6 clearly indicate that the residuals are unacceptably large. Some of the chi-square tests are plotted in Fig. 7 and they correctly reveal that the model is deficient.

Next the polynomial degrees L_i, $i = 1, 2$ were increased to 2. Notice that even for such a low degree of non-linearity and the small maximum lags the full model set contained 182 terms and direct estimation based on this full model would indeed result in a very complicated model and possibly numerical ill-conditioning. This full model set, however, contained all terms of the true system and the forward-regression orthogonal estimator correctly selected the model structure as indicated in Table 3. The response of this non-linear model and some of the chi-square tests are illustrated in Figs 8 and 9 respectively.

S. A. Billings et al.

Subsystem	Terms	Estimates	$[\text{err}]_i$	Standard deviations
1	$u_1(t-2)$	0·99504E + 0	0·66343E + 0	0·11604E − 1
	$y_1(t-1)$	0·50723E + 0	0·26657E + 0	0·80900E − 2
	$y_2(t-1)u_1(t-1)$	0·99287E − 1	0·44573E − 1	0·37140E − 2
	$e_1(t-1)$	0·50587E + 0	0·42730E − 2	0·58465E − 1
	$y_1(t-2)e_1(t-2)$	0·24534E + 0	0·27833E − 2	0·36388E − 1
2	$y_2(t-2)$	0·90314E + 0	0·84592E + 0	0·39662E − 2
	$u_2(t-1)$	0·99220E + 0	0·10754E + 0	0·12114E − 1
	$y_2(t-1)u_2(t-2)$	0·19862E + 0	0·40604E − 1	0·40523E − 2
	$e_2(t-1)$	0·42562E + 0	0·86239E − 3	0·57947E − 1
	$y_2(t-1)e_1(t-2)$	0·82201E − 1	0·26118E − 3	0·20369E − 1

$\rho_{p1} = 0\text{·}030 \quad \rho_{n1} = 0\text{·}020$
$\rho_{p2} = 0\text{·}030 \quad \rho_{n2} = 0\text{·}005$
$\biggr\}$ covariance of residuals $= \begin{bmatrix} 0\text{·}38768E - 1 & 0\text{·}56083E - 3 \\ 0\text{·}56083E - 3 & 0\text{·}44041E - 1 \end{bmatrix}$

Table 3. Non-linear model for Example 2.

4 (a) $\omega_1(t)=y_1^2(t-1)$, $\omega_2(t)=y_2^2(t-1)$
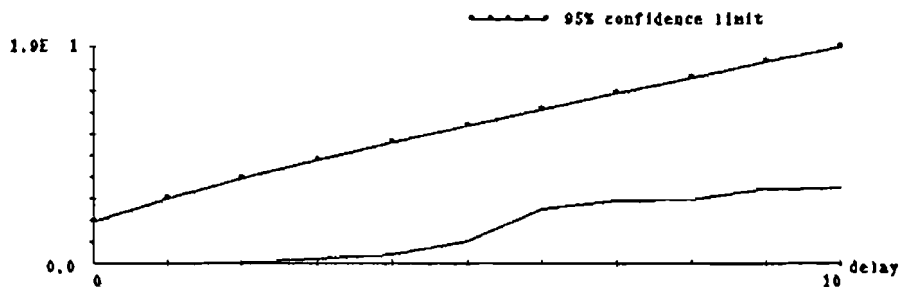
4 (b) $\omega_1(t)=e_1^2(t-1)$, $\omega_2(t)=e_2^2(t-1)$
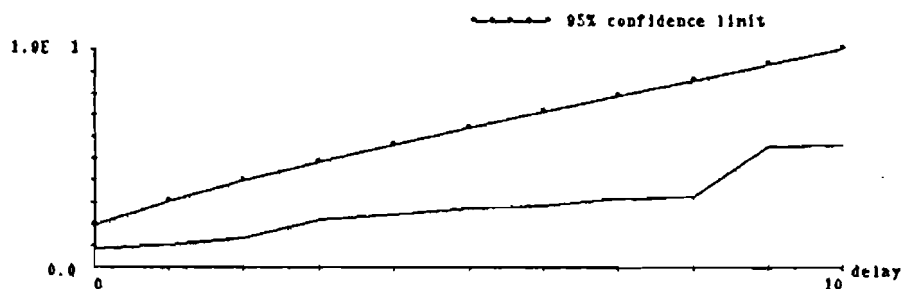
(c) $\omega_1(t)=u_1^2(t-9)$, $\omega_2(t)=u_2(t-9)y_1(t-9)$
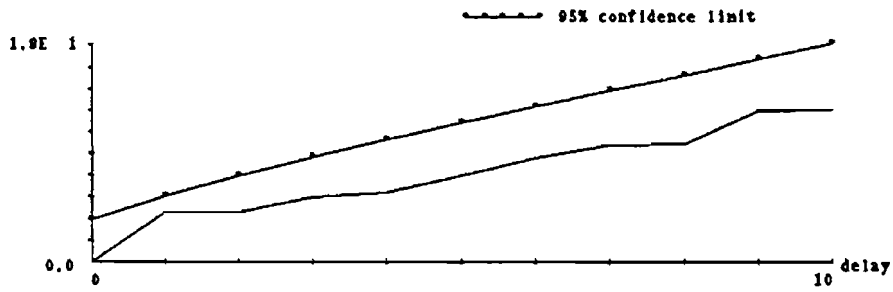
(d) $\omega_1(t)=u_1(t-9)y_2(t-9)$, $\omega_2(t)=u_2(t-1)y_1(t-1)$
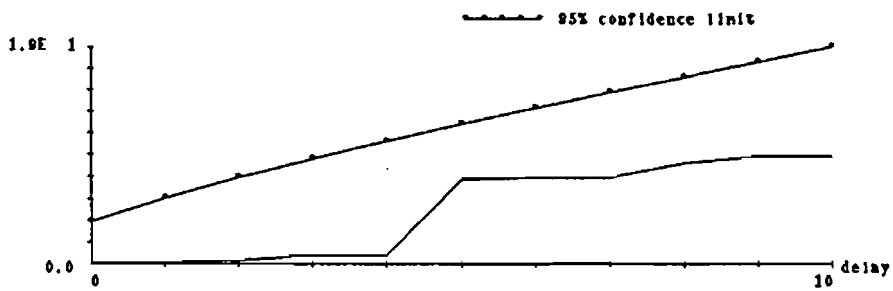
(e) $\omega_1(t)=e_1(t-1)y_2(t-7)$, $\omega_2(t)=0$
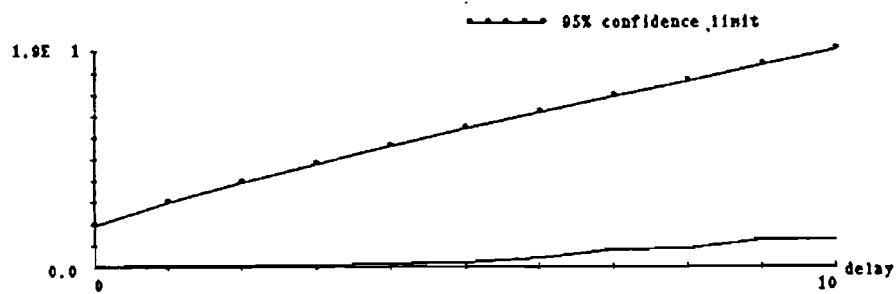
(f) $\omega_1(t)=0$, $\omega_2(t)=e_2(t-1)y_1(t-8)$
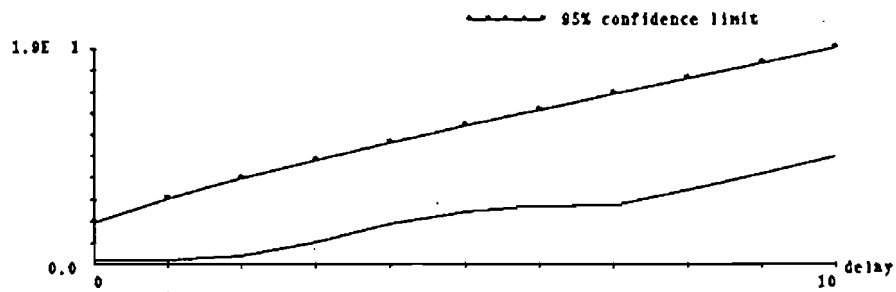
Figure 4. Chi-square tests of Example 1.

S. A. Billings et al.
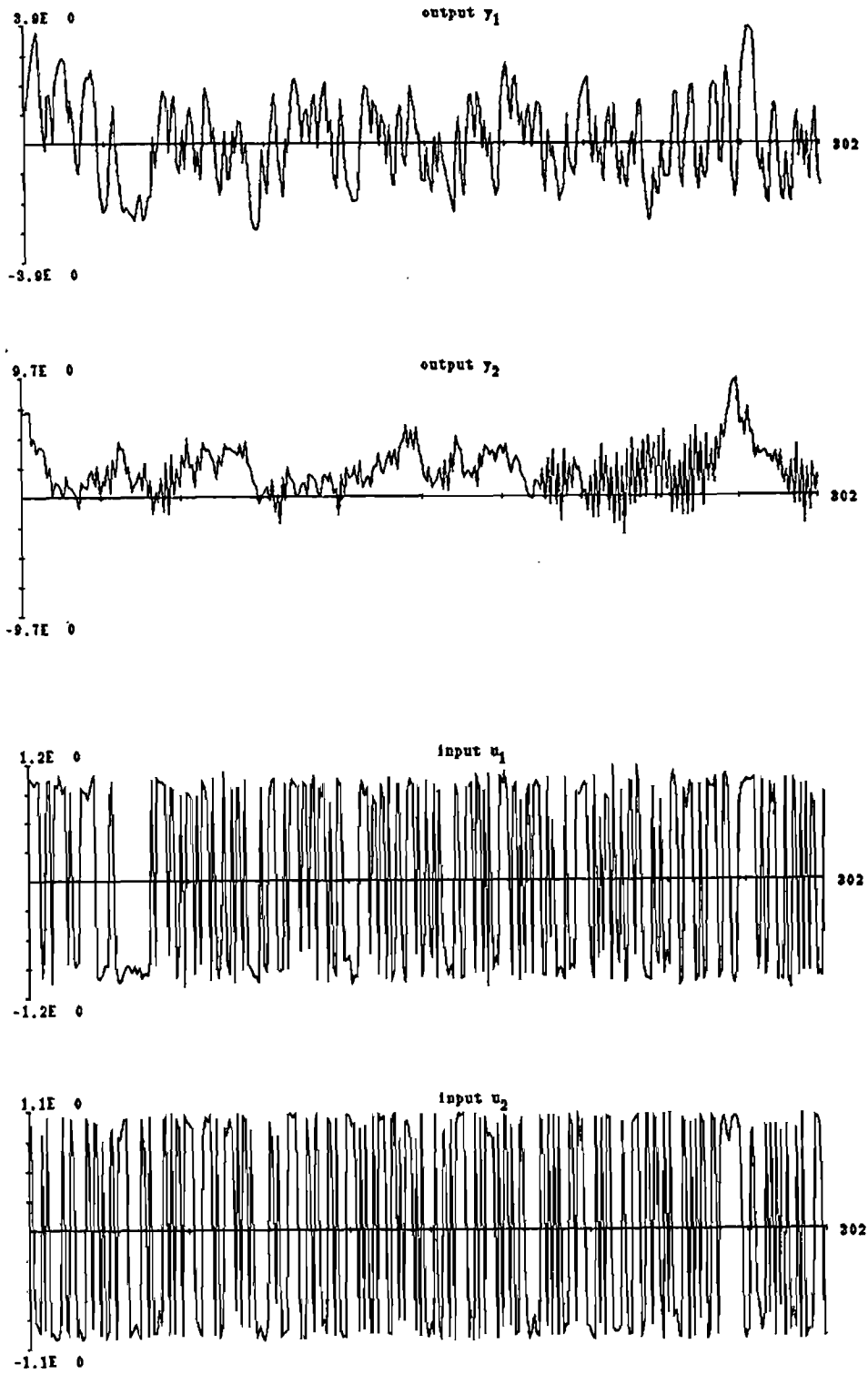
Figure 5.   Inputs and outputs of Example 2.

3.9E 0

measured output y_1

302

-3.9E 0

3.9E 0

one-step ahead prediction of y_1

302

-3.9E 0

8.2E -1

residual e_1

302

-8.2E -1

1.3E 0

deterministic prediction error e_{d1}

302

-1.3E 0

—— nodel predicted output of y_1

······ system output y_1

3.0E 0

302

-3.0E 0

9.7E 0   neasured output y_2

302

-9.7E 0

9.7E 0   one-step ahead prediction of y_2

302

-9.7E 0

continued

1.7E 0   residual e_2

302

-1.7E 0

5.0E 0   deterministic prediction error e_{d2}

302

-5.0E 0

——— model predicted output of y_2

······· system output y_2

0.7E 0

302

-0.7E 0

Figure 6. Linear model response of Example 2.

S. A. Billings et al.

(a) $\omega_1(t)=y_1(t-2)e_1(t-2)$, $\omega_2(t)=y_2(t-1)e_1(t-2)$
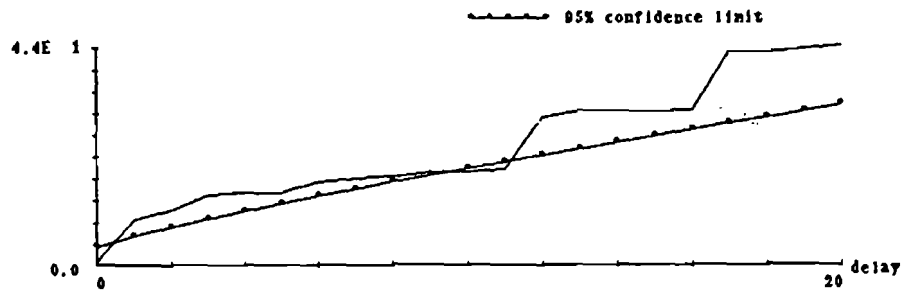
(b) $\omega_1(t)=y_2(t-1)u_1(t-1)$, $\omega_2(t)=0$
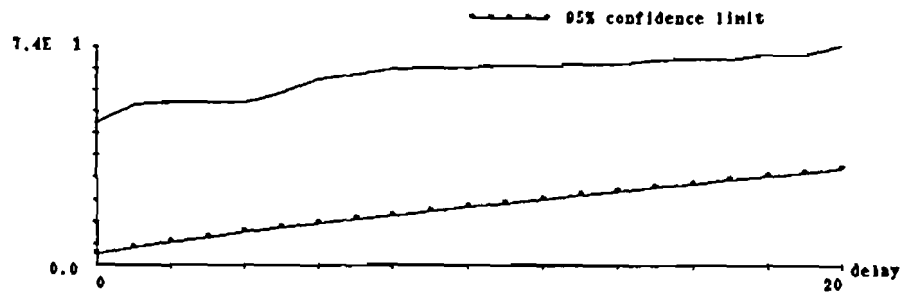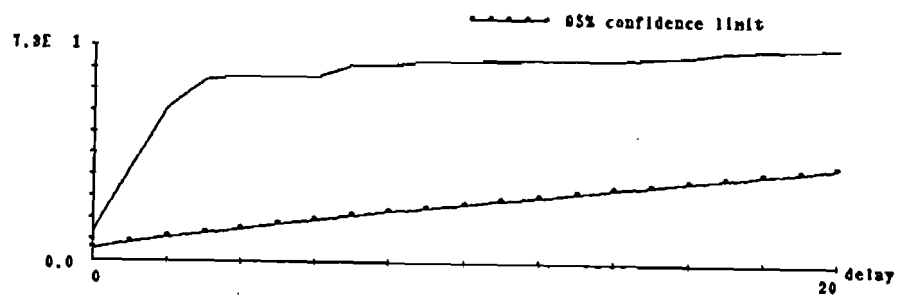
(c) $\omega_1(t)=y_1^2(t-1)$, $\omega_2(t)=y_2^2(t-1)$
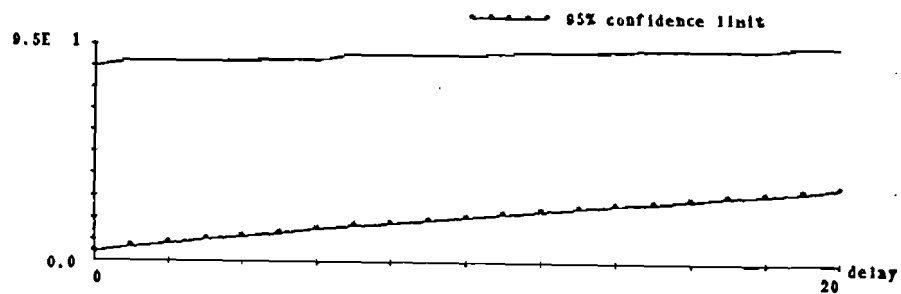
(d) $\omega_1(t)=0$, $\omega_2(t)=y_2(t-1)u_2(t-2)$

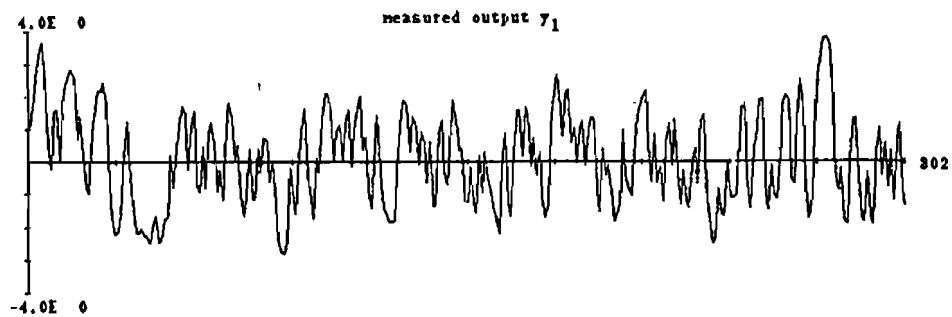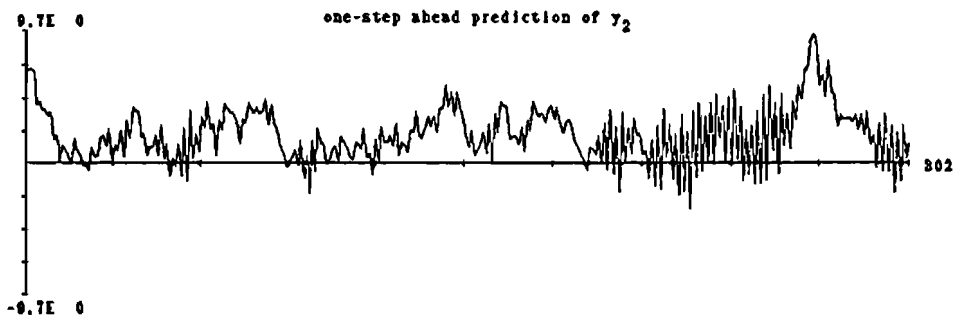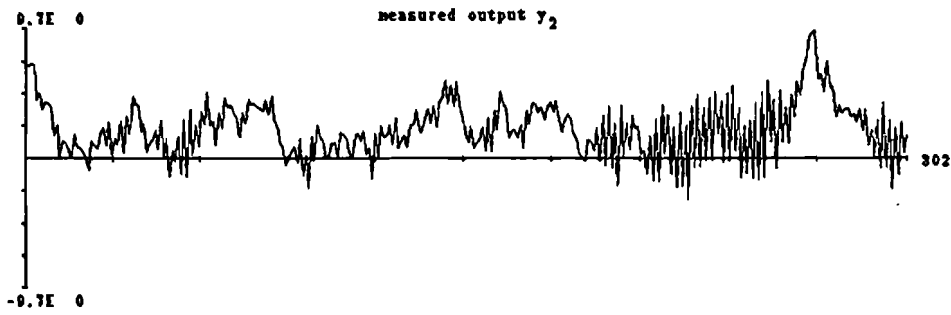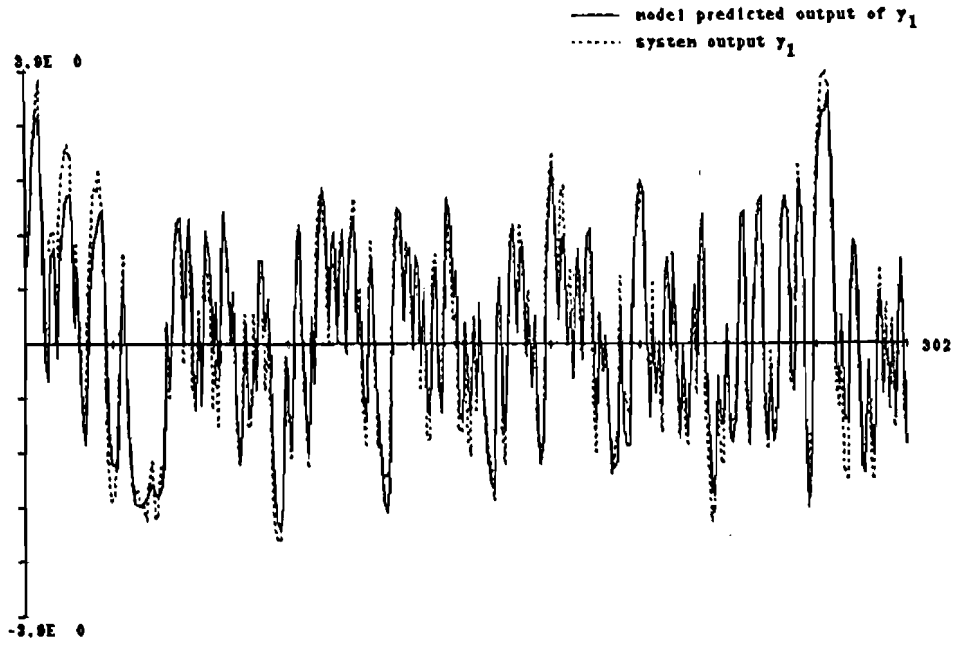Figure 7. Chi-square tests for linear model of Example 2.

4.0E 0   measured output y₁

302

-4.0E 0

4.0E 0   one-step ahead prediction of y₁

302

-4.0E 0

5.4E -1   residual e₁

302

-5.4E -1

8.7E -1   deterministic prediction error e_{d1}

302

-8.7E -1

S. A. Billings et al.

——— model predicted output of y_1

········ system output y_1

3.0E 0

302

-3.0E 0

0.7E 0   measured output y_2
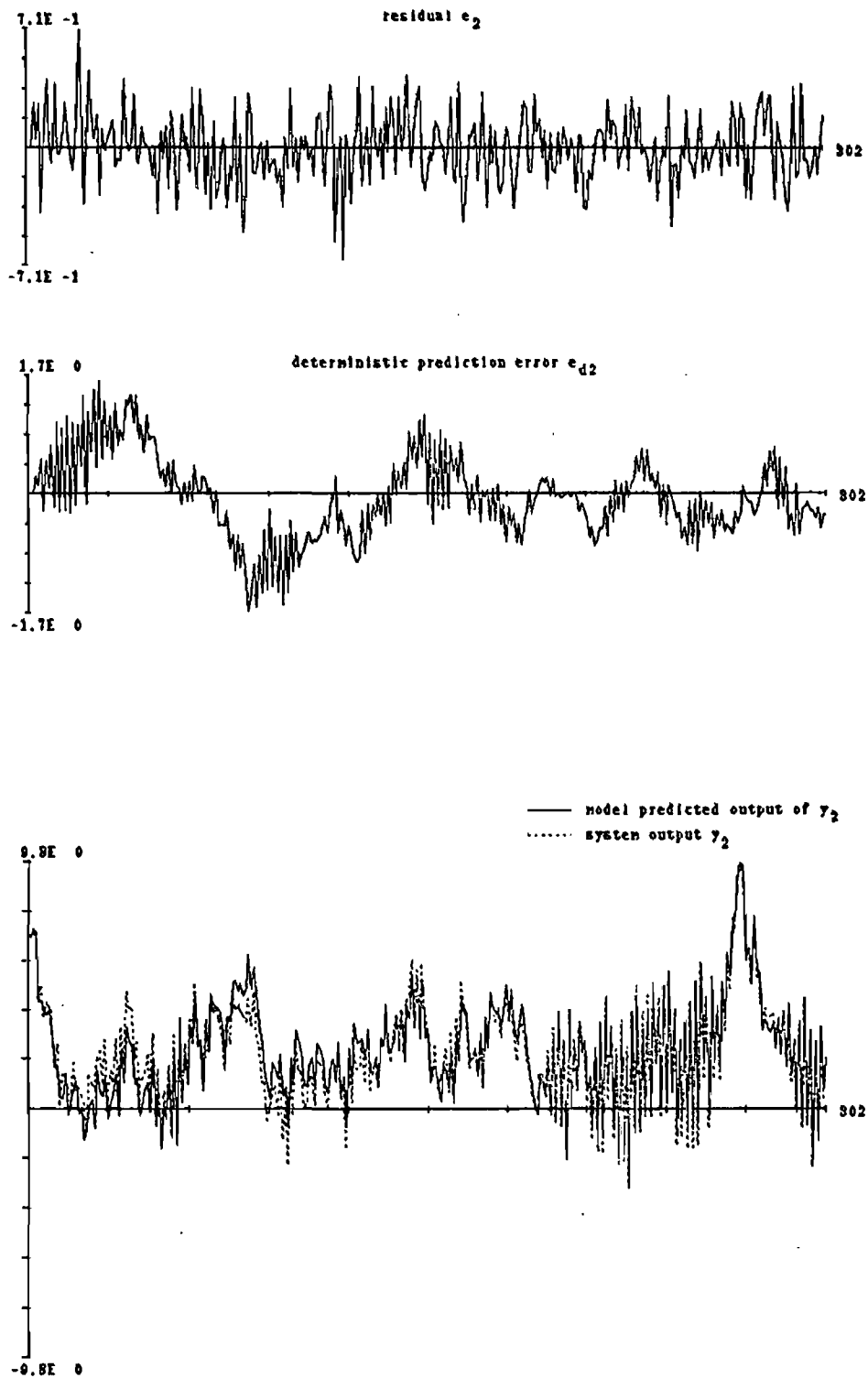
302

-0.7E 0

0.7E 0   one-step ahead prediction of y_2

302

-0.7E 0
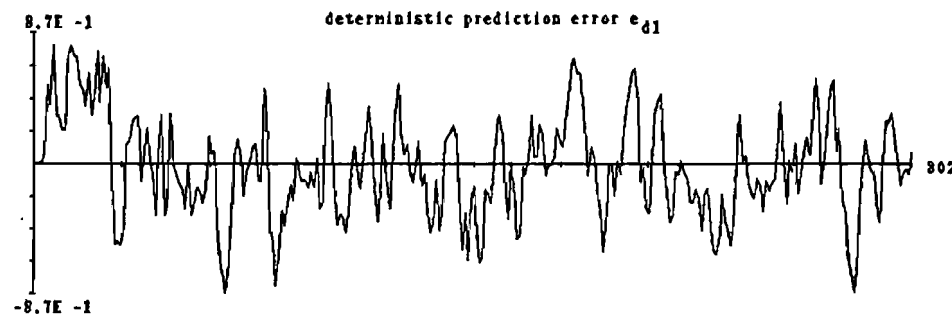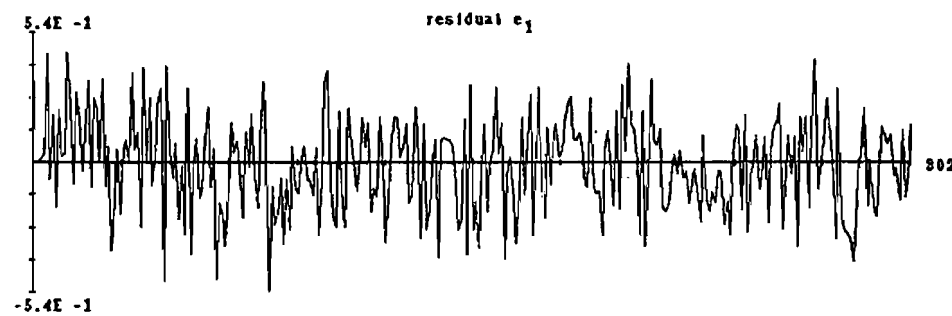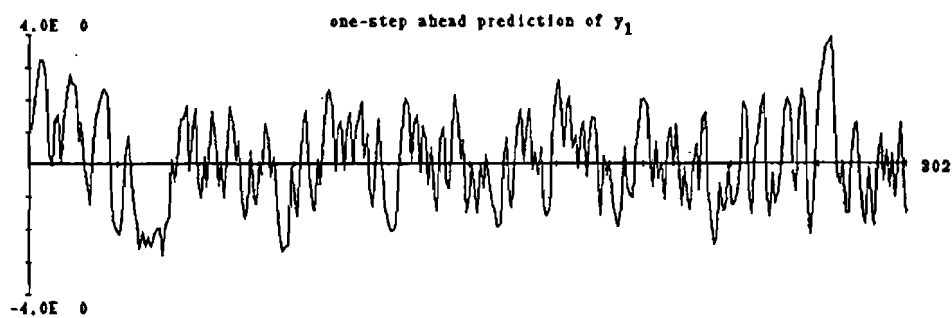
Figure 8.   Non-linear model response of Example 2.
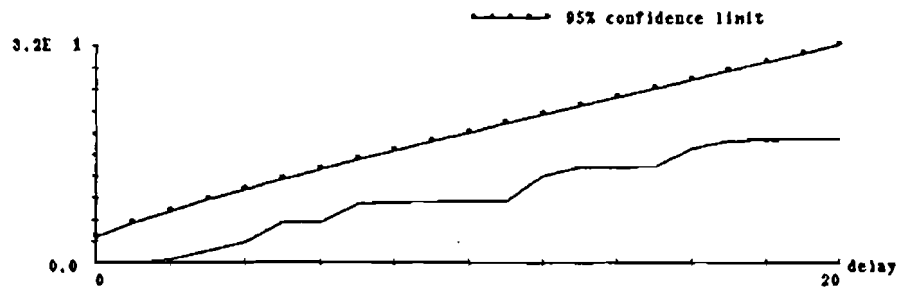
S. A. *Billings* et al.

(a) $\omega_1(t)=y_1(t-2)c_1(t-2)$, $\omega_2(t)=y_2(t-1)c_1(t-2)$

(b) $\omega_1(t)=y_2(t-1)u_1(t-1)$, $\omega_2(t)=0$

(c) $\omega_1(t)=y_1^2(t-1)$, $\omega_2(t)=y_2^2(t-1)$

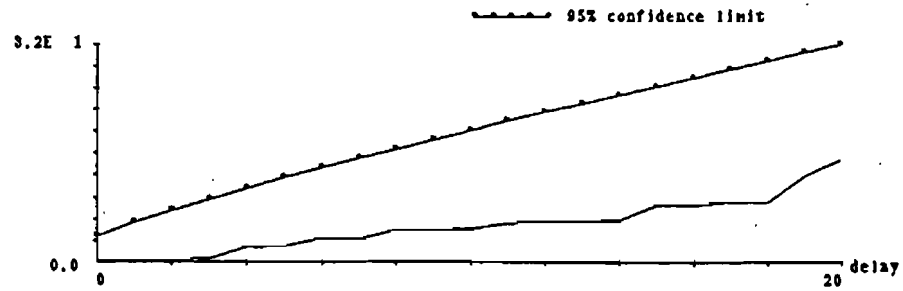(d) $\omega_1(t)=0$, $\omega_2(t)=y_2(t-1)u_2(t-2)$

Figure 9. Chi-square tests for non-linear model of Example 2.

7. Conclusions

A forward-regression orthogonal estimator has been derived for the identification of MIMO non-linear systems where model structure selection is often vital. It has been shown that this estimator efficiently combines structure determination with parameter estimation and, when coupled with model validity tests, is particularly powerful in identifying parsimonious models for structure-unknown systems. The application to both simulated and real data has been demonstrated.

REFERENCES

BILLINGS, S. A., 1986, Introduction to nonlinear systems analysis and identification. *Signal Processing for Control*, edited by K. Godfrey and P. Jones (New York: Springer-Verlag), pp. 261-294.

BILLINGS, S. A., and CHEN, S., 1988, Identification of non-linear rational systems using a prediction-error estimation algorithm. *International Journal of Systems Science*, **20**, 467-494.

BILLINGS, S. A., CHEN, S., and BACKHOUSE, R. J., 1988 a, The identification of linear and nonlinear models of a turbocharged automotive diesel engine. *Mechanical Systems and Signal Processing*, to be published.

BILLINGS, S. A., and FADZIL, M. B., 1985, The practical identification of systems with nonlinearities. *Proc. 7th IFAC Symp. on Identification and System Parameter Estimation*, York, U.K., pp. 155-160.

BILLINGS, S. A., FADZIL, M. B., SULLEY, J., and JOHNSON, P. M., 1988 b, Identification of a nonlinear difference equation model of an industrial diesel generator. *Mechanical Systems and Signal Processing*, **2**, 59-76.

BILLINGS, S. A., KORENBERG, M. J., and CHEN, S., 1988 c, Identification of non-linear output-affine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, **19**, 1559-1568.

BILLINGS, S. A., and VOON, W. S. F., 1986 a, Correlation based model validity tests for non-linear models. *International Journal of Control*, **44**, 235-244; 1986 b, A prediction-error and stepwise-regression estimation algorithm for non-linear systems. *Ibid.*, **44**, 803-822.

BOHLIN, T., 1978, Maximum-power validation of models without higher-order fitting. *Automatica*, **14**, 137-146.

CHEN, S., BILLINGS, S. A., and LUO, W., 1989, Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, to be published.

JENKINS, G. M., and WATTS, D. G., 1969, *Spectral Analysis and Its Applications* (San Francisco: Holden-Day).

KORENBERG, M. J., BILLINGS, S. A., LIU, Y. P., and McILROY, P. J., 1988, Orthogonal parameter estimation algorithm for non-linear stochastic systems. *International Journal of Control*, **48**, 193-210.

LEONTARITIS, I. J., and BILLINGS, S. A., 1985, Input-output parametric models for non-linear systems, Part I: deterministic non-linear systems; Part II: stochastic non-linear systems. *International Journal of Control*, **41**, 303-344; 1987, Model selection and validation methods for non-linear systems. *Ibid.*, **45**, 311-341.