[redacted]

Parameter Estimation for Stochastic Nonlinear Rational Models

Q.M. Zhu and S.A. Billings

Department of Automatic Control
and Systems Engineering

University of Sheffield

Mappin Street

Sheffield S1 4DU

U. K.

Research Report No. 435

September 1991

September 19, 1991

Parameter Estimation for Stochastic Nonlinear Rational Models

Q.M. Zhu, S.A. Billings

Department of Automatic Control and Systems Engineering,

University of Sheffield, Sheffield S1 4DU, UK

Abstract:
A general orthogonal parameter estimation algorithm is derived to estimate both the structure and the parameters for a wide range of stochastic nonlinear systems which can be described by a nonlinear rational model. Simulation studies are included to demonstrate the performance of the algorithm.

1 Introduction

Almost all applications of parameter estimation assume that the model structure is known a priori or that an appropriate structure will be obtained by a search over a limited model set. Thus if the system is known to be linear models are fitted over a range of time delays and model orders and the best fit is selected as the final model. When the system is nonlinear however the number of the combinations of models which have to be searched can rapidly become overwhelming and a more structured procedure is required. System identification then for nonlinear systems at least involves detecting the model structure and then estimating the unknown parameters. Apart from the obvious computational advantage that this offers the models obtained will tend to be concise, they will not be overparameterised and it may be possible to relate the terms in the model to specific components of the system thus providing additional insight.

When the system under investigation is only mildly nonlinear it can be represented by a polynomial NARMAX model and algorithms which solve both the structure detection and parameter estimation problems have been derived. These are based on an orthogonal estimator where the orthogonal property can be exploited to allow an optimal search for the model structure (Korenberg, Billings, Liu, and McIlroy 1988).

Polynomial models are fine for many applications but they are inadequate for severely nonlinear systems (Billings and Chen 1989) and recently the nonlinear rational model was introduced to overcome these problems. Sontag (1979) has studied the properties of the output affine model which can be considered as a subset of the rational model (Billings and Chen 1989 ,). Observability. realizability, and minimality were investigated and Sontag proved conditions under which these models are globally valid.

The extension of the rational model to nonlinear stochastic systems was introduced by Billings and Chen (1989) and Chen and Billings (1989). Parameterisation of systems using rational models should therefore offer substantial advantages compared to linear or polynomial expansions. However, the disadvantage is that rational models are nonlinear in the parameters. When the rational model contains nonlinear polynomial terms in the numerator and denominator attempts to multiply out and provide a linear in the parameter model fail because of the inherent noise bias which is induced even in the presence of just white noise. It is probably for these reasons that virtually no parameter estimation algorithms are available for nonlinear stochastic rational models. A prediction error estimation algorithm was derived by Billings and Chen (1989) but this was computationally demanding and therefore difficult to apply to real nonlinear systems.

In the present study a new orthogonal estimation algorithm which determines both the structure and provides estimates of the unknown parameters for nonlinear stochastic rational models is derived. The algorithm, called orthogonal rational model estimator (ORME), exploits the results from Billings and Zhu (1991) and Zhu and Billings (1991) to avoid the inherent bias problem. The new orthogonal formulation is shown to be easy to implement and provides an opportunity for the first time to fit concise nonlinear stochastic rational models from noise corrupted measurements. Simulation results are included to demonstrate the effectiveness of the new algorithm.

2 The rational model

Approximation studies can normally be divided into two aspects. The first concerns the approximation of functions whose computation is somewhat difficult, by simpler functions, such as polynomials or the ratio of two polynomials, which are easily evaluated with computers. This is called the function approximation problem, and many publications have appeared devoted to this approach (Garabedian 1965, Hayes 1970, Newman 1978). The approximating functions mainly take a static form due to the nature of the problem, for example a polynomial usually takes the finite power series form

$$y = 1 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n$$

and a rational function, the ratio of two polynomials, is of the form

$$y = \frac{1 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n}{1 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m}$$

The second problem concerns data fitting. In this case a simple function is usually constructed which will represent the data sequence as accurately as possible. This approach is appropriate to both static curve fitting and dynamic model estimation. System identification belongs to the latter case and linear model fitting based on differential or difference equation have been widely studied (Goodwin and Payne 1977, Ljung 1987).

Nonlinear approximation theory shows that rational functions may provide a more powerful approximation (i.e. a smaller number of parameters for a similar accuracy) than polynomial functions, such an approach is often abundantly justified, as shown by the work of Hastings (1955). Rational functions, are much more general and are capable of accurately representing certain types of singular, or near singular behaviour, and operating in infinite ranges. Rational functions have good extrapolation properties, they are easy to evaluate, and include the polynomial functions as special cases. Static rational functions have been successfully used to approximate some typical nonlinear functions such as e^x, e^{-x}, $\sqrt{x}$, and $|x|$ (Braess 1980) and dynamic rational functions have been primarily used in nonlinear system identification (Billings and Chen 1989, Billings and Zhu 1991, Zhu and Billings 1991).

An obvious example to show the concise expression obtained by using a rational function rather than a polynomial is given by considering the rational function

$$y = \frac{1}{1 + x}$$

which when expanded as a power series can be represented by the polynomial

$$y = 1 - x + x^2 - \cdots + x^{2n} - \cdots$$

There is, however, a need for some caution when attempting to use rational approximations. There is a possibility of degeneracy. The rational functions $\frac{a_1}{b_1}$ and $\frac{a_2}{b_2}$ are deemed equivalent if $a_1 b_2 = a_2 b_1$ and a rational function $\frac{a}{b}$ of degree (m, n), where m and $n \geq 1$ are the degrees of the numerator and denominator polynomials respectively, is called degenerate if $\frac{a}{b} = 0$ or $\frac{a}{b} \in (m-1, n-1)$ (Braess 1980). Although degeneracy does not affect uniqueness of the best approximation, it does spoil the continuity of the metric projection, which in turn implies an anomalous behaviour in many directions. For example such a degenerate function cannot be a best and uniformly convergent approximation, and most of the theories on rational approximation presented by Braess (1980) are conditioned with the non degenerate property. A rational function may also be less convenient for certain analytical manipulations, such as integration or differentiation. There is a bound on the derivatives of polynomials in terms of the functions but for rational functions the derivatives are unbounded (Feinerman and Newman 1974).

2.1 Stochastic rational representations

In a practical environment, some uncertain behaviours or stochastic phenomena are often encounted and model fitting based on a stochastic NARMAX model set will be necessary. Let t=1, 2, $\cdots$ be a time index. Let $a(t)$, and $b(t)$ be polymonials in $[y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t), \cdots, e(t-r)]$, where $u(t)$ and $y(t)$ represent the input and output at time t respectively, $e(t)$ is an unobservable independent noise sequence with zero mean and finite variance σ_e^2, the $r \geq 0$ is, an integer, the order of the polynomials. The stochastic NARMAX model set can be classified as (Sontag 1979, Chen and Billings 1989)

(1) Rational model

$$y(t) = \frac{a(y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))}{b(y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))} + e(t) \tag{2.1.1}$$

(2) Integral model

$$y^s(t) = \frac{a(y(t), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))}{b(u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))} + e(t) \tag{2.1.2}$$

where the degree of $y(t)$ in $a(t)$ is less than s

(3) Recursive model

$$y(t) = \frac{a(y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))}{b(u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))} + e(t) \tag{2.1.3}$$

(4) Output affine model

$$y(t) = \frac{\sum_{i=1}^{r} a_i(u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))y(t-i)}{a_0(u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))}$$

$$+ \frac{a_{r+1}(u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))}{a_0(u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))} + e(t) \qquad (2.1.4)$$

(5) Standard model (polynomial NARMAX model)

$$y(t) = a(y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r)) + e(t)$$

$$(2.1.5)$$

(6) Linear difference equation model (ARMAX model)

$$y(t) = \sum_{i=1}^{r}\alpha_i y(t-i) + \sum_{i=0}^{r}\beta_i u(t-i) + \sum_{i=1}^{r}\gamma_i e(t-i) + e(t) \qquad (2.1.6)$$

Inspection of the models shows that they are all types of NARMAX model (Chen and Billings 1989). In the above examples the rational model is the most general expression, in fact all the models are a subclass of the rational model. For example the polynomial NARMAX model in eqn (2.1.5) is a class of rational model with denominator $b(t) = 1$. All the results that follow which are derived for the rational model are therefore appropriate to all the models given above.

The NARMAX model is a very general model which includes variaties of discrete time linear and nonlinear difference equations as subsets. The NARMAX model can be easily and accurately used to describe a wide range of linear and nonlinear systems because it is concise and can give a high accuracy of approximation to a bounded system globally (Sontag 1979).

2.2 Model parametrization

A input output response map may be written as

$$y(t) = f(\mathbf{u}, \mathbf{y}, \mathbf{e}, \theta, t) \qquad (2.2.1)$$

where $f(.)$ is the input and output map, $\mathbf{u}$, $\mathbf{y}$, and $\mathbf{e}$ are input, output, and noise vectors, θ is the unknown parameter vector , and t is the time index. Two specific forms of this map are:

(i) Linear in the inputs, outputs, noise, and parameters

This gives

$$f(\lambda\mathbf{u}_1 + \mu\mathbf{u}_2, \mathbf{y}, \mathbf{e}, \boldsymbol{\theta}, t) = \lambda f(\mathbf{u}_1, \mathbf{y}, \mathbf{e}, \boldsymbol{\theta}, t) + \mu f(\mathbf{u}_2, \mathbf{y}, \mathbf{e}, \boldsymbol{\theta}, t)$$
$$f(\mathbf{u}, \lambda\mathbf{y}_1 + \mu\mathbf{y}_2, \mathbf{e}, \boldsymbol{\theta}, t) = \lambda f(\mathbf{u}, \mathbf{y}_1, \mathbf{e}, \boldsymbol{\theta}, t) + \mu f(\mathbf{u}, \mathbf{y}_2, \mathbf{e}, \boldsymbol{\theta}, t)$$
$$f(\mathbf{u}, \mathbf{y}, \lambda\mathbf{e}_1 + \mu\mathbf{e}_2, \boldsymbol{\theta}, t) = \lambda f(\mathbf{u}, \mathbf{y}, \mathbf{e}_1, \boldsymbol{\theta}, t) + \mu f(\mathbf{u}, \mathbf{y}, \mathbf{e}_2, \boldsymbol{\theta}, t)$$
$$f(\mathbf{u}, \mathbf{y}, \mathbf{e}, \lambda\boldsymbol{\theta}_1 + \mu\boldsymbol{\theta}_2, t) = \lambda f(\mathbf{u}, \mathbf{y}, \mathbf{e}, \boldsymbol{\theta}_1, t) + \mu f(\mathbf{u}, \mathbf{y}, \mathbf{e}, \boldsymbol{\theta}_2, t) \qquad (2.2.2)$$

which satisfy the superposition principle concerning the inputs, outputs, noise, and parameters. Where λ and μ are some real scalars.

(ii) Nonlinear in the inputs, outputs, noise, and parameters

This gives

$$f(\lambda\mathbf{u}_1 + \mu\mathbf{u}_2, \mathbf{y}, \boldsymbol{\theta}, t) \neq \lambda f(\mathbf{u}_1, \mathbf{y}, \boldsymbol{\theta}, t) + \mu f(\mathbf{u}_2, \mathbf{y}, \boldsymbol{\theta}, t)$$
$$f(\mathbf{u}, \lambda\mathbf{y}_1 + \mu\mathbf{y}_1, \boldsymbol{\theta}, t) \neq \lambda f(\mathbf{u}, \mathbf{y}_1, \boldsymbol{\theta}, t) + \mu f(\mathbf{u}, \mathbf{y}_2, \boldsymbol{\theta}, t)$$
$$f(\mathbf{u}, \mathbf{y}, \lambda\mathbf{e}_1 + \mu\mathbf{e}_2, \boldsymbol{\theta}, t) \neq \lambda f(\mathbf{u}, \mathbf{y}, \mathbf{e}_1, \boldsymbol{\theta}, t) + \mu f(\mathbf{u}, \mathbf{y}, \mathbf{e}_2, \boldsymbol{\theta}, t)$$
$$f(\mathbf{u}, \mathbf{y}, \lambda\boldsymbol{\theta}_1 + \mu\boldsymbol{\theta}_2, t) \neq \lambda f(\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_1, t) + \mu f(\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_2, t) \qquad (2.2.3)$$

which do not satisfy the superposition principle for any one of the inputs, outputs, noise, or parameters.

It is easy to show that all rational eqn (2.1.1), integral eqn (2.1.2), and recursive eqn (2.1.3) models are nonlinear in the parameters, input, output, and noise. The output affine model in eqn (2.1.4) is nonlinear in the parameters, input, and noise but linear in the output. The polynomial NARMAX model of eqn (2.1.5) is nonlinear in the input, output, and noise but linear in the parameters. The ARMAX model in eqn (2.1.6) is linear in the parameters, input, output, and noise.

From the classifications, the general NARMAX model includes all the combinations of linear and nonlinear in the parameters, input, output, and noise. Hence the parameterised NARMAX model provides a general basis for the modelling of systems.

The rational model is generally given in a parametric form. Define the numerator polynomial

$$a(t) = \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} \qquad (2.2.4)$$

and the denominator polynomial

$$b(t) = \sum_{j=1}^{den} p_{dj}(t)\theta_{dj} \tag{2.2.5}$$

where $p_{nj}(t)$, $p_{dj}(t)$ are referred to as terms consisting of $y(t), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r)$ and the total number of parameters is *num* + *den*.

The parametrized rational model may then be expressed as

$$y(t) = \frac{a(t)}{b(t)} + e(\iota) = \frac{\sum_{j=1}^{num} p_{nj}(t)\theta_{nj}}{\sum_{j=1}^{den} p_{dj}(t)\theta_{dj}} + e(t) \tag{2.2.6}$$

where $e(t)$ is an unobservable independent noise sequence with zero mean and finite variance σ_e^2 as defined in section 2.1.

2.3 Linear in the parameters expansion

Identification based on eqn (2.2.6) directly is very complex because the model is nonlinear in the parameters. Algorithms by Marquardt (1963) are available when the measurements are noise free but this is unrealistic. The only alternative is the prediction error algorithm of Billings and Chen (1989), which is computationally demanding.

An obvious solution would appear to be to multiply out to give a linear in the parameters model. Whilst this appears to simplify the problem it induces an inherent bias problem caused by the multiplication of lagged *u's*, *y's*, and *e's* in $p_{dj}(t)$ with $e(t)$ (Billings and Zhu 1991). This will be discussed in detail and a solution proposed in section 3.

Expanding eqn (2.2.6) therefore to give a linear in the parameters expression yields

$$\begin{aligned} Y(t) &= a(t) - y(t)\sum_{j=2}^{den} p_{dj}(t)\theta_{dj} + b(t)e(t) \\ &= \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} - \sum_{j=2}^{den} y(t)p_{dj}(t)\theta_{dj} + \zeta(t) \end{aligned} \tag{2.3.1}$$

where

$$\begin{aligned} Y(t) &= y(t)p_{d1}(t)|_{\theta_{d1}=1} \\ &= p_{d1}(t)\,\frac{a(t)}{b(t)} + p_{d1}(t)e(t) \end{aligned} \tag{2.3.2}$$

Alternatively devide all the right hand side terms by θ_{d1} and redifine symbols to give essentially $\theta_{d1} = 1$. Notice that

$$\begin{aligned}\zeta(t) &= b(t)e(t)\\ &= (\sum_{j=1}^{den} p_{dj}(t)\theta_{dj})e(t)\\ &= p_{d1}(t)e(t) + (\sum_{j=2}^{den} p_{dj}(t)\theta_{dj})e(t) \qquad (2.3.3)\end{aligned}$$

where

$$E[\zeta(t)] = E[b(t)]E[e(t)] = 0 \qquad (2.3.4)$$

providing $e(t)$ has been reduced to an uncorrelated sequence as defined in eqn (2.2.6).

Eqn (2.3.1) can alternatively be expressed as

$$\begin{aligned}Y(t) &= \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} - \sum_{j=2}^{den} y(t)p_{dj}(t)\theta_{dj} + b(t)e(t)\\ &= \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} - \sum_{j=2}^{den} \frac{a(t)}{b(t)} p_{dj}(t)\theta_{dj} + p_{d1}(t)e(t) \qquad (2.3.5)\end{aligned}$$

Although the term $\frac{a(t)}{b(t)} p_{dj}(t)$ in eqn (2.3.5) cannot be obtained directly the expression is very useful in the analysis of bias and the derivation of the new estimator.

Eqn (2.3.1) may be written in vector notation as

$$\begin{aligned}Y(t) &= \phi(t)\Theta + \zeta(t)\\ &= \hat{\phi}(t)\Theta + p_{d1}(t)e(t) \qquad (2.3.6)\end{aligned}$$

where

$$\begin{aligned}\phi(t) &= [\phi_n(t) \;\; \phi_d(t)]\\ &= [p_{n1}(t) \cdots p_{nnum}(t) \; -p_{d2}(t)y(t) \cdots -p_{dden}(t)y(t)]\\ &= [p_{n1}(t) \cdots p_{nnum}(t) \; -p_{d2}(t)(\frac{a(t)}{b(t)} + e(t)) \cdots -p_{dden}(t)(\frac{a(t)}{b(t)} + e(t))] \qquad (2.3.7)\end{aligned}$$

$$\begin{aligned}\Theta^T &= [\Theta_n \; \Theta_d]\\ &= [\theta_{n1} \cdots \theta_{nnum} \; \theta_{d2} \cdots \theta_{dden}] \qquad (2.3.8)\end{aligned}$$

and

$$\hat{\phi}(t) = [\phi_n(t) \;\; \hat{\phi}_d(t)]$$

$$= [p_{n1}(t) \cdots p_{nnum}(t) \; -p_{d2}(t)\frac{a(t)}{b(t)} \cdots -p_{dden}(t)\frac{a(t)}{b(t)}] \tag{2.3.9}$$

Notice that the matrix $\hat{\phi}(t)$ cannot be obtained directly because $\frac{a(t)}{b(t)}$ cannot be measured.

3 The bias problem

The bias problem which results from directly applying least squares type algorithms to the rational model identification exists even in the case of white noise corrupted data. It will be seen from following derivations that there is a noise element which includes $e(t)$ in the denomintor terms when the rational model is written as a linear in the parameters expression. Consider eqn (2.3.1) and take one of the denominator terms as an example on the right hand side

$$y(t)p_{dj}(t) = p_{dj}(t)\,\frac{a(t)}{b(t)} + p_{dj}(t)e(t) \tag{3.1}$$

where $y(t) = \frac{a(t)}{b(t)} + e(t)$ as defined in eqn(2.3.2), $p_{dj}(t)\,\frac{a(t)}{b(t)}$ represents the elements which are independent of $e(t)$ and $p_{dj}(t)e(t)$ represents the elements which involve the current noise term $e(t)$. We call this phenomenon an inherent error term which can not be removed from the regression terms or variables. The inherent errors will introduce bias in the parameter estimates associated with both numerator and denominator terms when linear least squares type algorithms are used.

A polynomial NARMAX model without denominator terms, ie $b(t) = 1$, is a linear in the parameters model but in this case there are no inherent error terms. Hence the estimates obtained from least squares algorithms are unbiased in case of the white noise for the polynomial NARMAX.

To show the bias problem, consider eqn (2.3.6). The least squares parameter estimate is computed as

$$\hat{\Theta} = [\Phi^T\Phi]^{-1}\,\Phi^T\vec{Y} \tag{3.2}$$

where

$$\Phi^T = [\phi^T(1) \cdots \phi^T(N)]$$

$$= \begin{bmatrix} \phi_n^T(1) & . & . & . & \phi_n^T(N) \\ \phi_d^T(1) & . & . & . & \phi_d^T(N) \end{bmatrix}$$

$$= \begin{bmatrix} p_{n1}(1) & . & p_{n1}(N) \\ . & . & . \\ . & . & . \\ . & . & . \\ p_{nnum}(1) & . & p_{nnum}(N) \\ -p_{d2}(1)(\frac{a(1)}{b(1)} + e(1)) & . & -p_{d2}(N)(\frac{a(N)}{b(N)} + e(N)) \\ . & . & . \\ . & . & . \\ . & . & . \\ -p_{dden}(1)(\frac{a(1)}{b(1)} + e(1)) & . & -p_{dden}(N)(\frac{a(N)}{b(N)} + e(N)) \end{bmatrix}$$

$$\vec{Y} = [Y(1) \cdots Y(N)]^T \tag{3.3}$$

It is clear from eqns (2.2.6), (2.3.1), and (2.3.6) that Φ may include lagged noise model terms where N is the data length. The normal matrix $\Phi^T\Phi$ and the correlation vector $\Phi^T\vec{Y}$ can be expressed as

$$\Phi^T\Phi = \begin{bmatrix} \sum_{t=1}^{N} \phi_n^T(t)\phi_n(t) & \sum_{t=1}^{N} \phi_n^T(t)\phi_d(t) \\ \sum_{t=1}^{N} \phi_d^T(t)\phi_n(t) & \sum_{t=1}^{N} \phi_d^T(t)\phi_d(t) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{t=1}^{N} \phi_n^T(t)\phi_n(t) & \sum_{t=1}^{N} \phi_n^T(t)\hat{\phi}_d(t) \\ \sum_{t=1}^{N} \hat{\phi}_d^T(t)\phi_n(t) & \sum_{t=1}^{N} \hat{\phi}_d^T(t)\hat{\phi}_d(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_e^2 \sum_{t=1}^{N} p_d^T(t)p_d(t) \end{bmatrix}$$

and

$$\Phi^T\vec{Y} = \begin{bmatrix} \sum_{t=1}^{N} \phi_n^T(t)p_{d1}(t)\frac{a(t)}{b(t)} \\ \sum_{t=1}^{N} \hat{\phi}_d^T(t)p_{d1}(t)\frac{a(t)}{b(t)} \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma_e^2 \sum_{t=1}^{N} p_d^T(t)p_{d1}(t) \end{bmatrix} \tag{3.4}$$

where

$$p_d(t) = [p_{d2}(t) \quad \cdots \quad p_{dden}(t)] \tag{3.5}$$

and $\hat{\phi}_d(t)$ is defined in eqn (2.3.9)

Rewritting eqn (3.4) gives

$$\Phi^T\Phi = [\Phi^T\Phi]_{(t-1)} + \sigma_e^2\,\Psi$$
$$\Phi^T\vec{Y} = [\Phi^T\vec{Y}]_{(t-1)} + \sigma_e^2\,\psi \qquad (3.6)$$

where the definition of terms follows directly and

$$\Psi = \begin{bmatrix} 0 & 0 \\ 0 & \sum_{t=1}^{N} p_d^T(t)p_d(t) \end{bmatrix} = \sum_{t=1}^{N}\rho^T(t)\rho(t) = \sum_{t=1}^{N}\Psi(t)$$

$$\psi = \begin{bmatrix} 0 \\ \sum_{t=1}^{N} p_d^T(t)p_{d1}(t) \end{bmatrix} = \sum_{t=1}^{N}\rho^T(t)p_{d1}(t) = \sum_{t=1}^{N}\psi(t) \qquad (3.7)$$

where

$$\rho(t) = [0\ p_d(t)] \qquad (3.8)$$

All terms involving $e(t)$ appear in $\sigma_e^2\,\Psi$ and $\sigma_e^2\,\psi$ which are called error terms and the subscript $(t-1)$ indicates that only lagged noise terms (eg $e(t-j)$ $j \geq 1$) are present.

Hence the estimate given in eqn (3.2) can be written as

$$\hat{\Theta} = [\Phi^T\Phi]^{-1}\,\Phi^T\vec{Y}$$
$$= [[\Phi^T\Phi]_{(t-1)} + \sigma_e^2\,\Psi]^{-1}\,[[\Phi^T\vec{Y}]_{(t-1)} + \sigma_e^2\,\psi] \qquad (3.9)$$

The two terms $\sigma_e^2\,\Psi$ and $\sigma_e^2\,\psi$ will cause bias even if the additive noise is white. Detailed derivations and simulation studies were presented in a previous publication (Billings and Zhu 1991). A new rational model estimator (RME) (Billings and Zhu 1991) based on an extended least squares formulation has been developed to remove the bias and a recursive implementation of this estimator (RRME) (Zhu and Billings 1991) has also been derived for on line applications.

Although the RME algorithm provides unbiased parameter estimates for the stochastic nonlinear rational model it does not solve the structure detection problem which involves a combinational explosion if all possible model structures are tested in a brute force manner. A natural solution to this problem is to develop an orthogonal algorithm for the rational model. The major strength of an orthogonal algorithm lies mainly in the fact that it should provide information regarding which terms in the model are significant. This is vital in the identification of both linear and especially nonlinear systems because the model structure of most real systems is rarely known a priori.

4 Orthogonal algorithm

The orthogonal algorithm developed by (Korenberg 1985, Korenberg, Billings, Liu, and McIlroy 1988) for the NARMAX model which is linear in the parameters can be applied to the rational model only when the data is totally noise free. This is clearly unrealistic and a new orthogonal formulation is required for nonlinear stochastic rational models. Transforming eqn (2.3.1) into an auxiliary model form

$$Y(t) = \sum_{j=1}^{num} w_{nj}(t)\hat{g}_{nj} + \sum_{j=2}^{den} w_{dj}(t)\hat{g}_{dj} + b(t)e(t) \tag{4.1}$$

where $w_{*j}(t)$ (* denotes either n or d) are constructed to be orthogonal over data record such that $\sum_{t=1}^{N} w_{*j}(t)\; w_{*i}(t) = 0, j \neq i$, and $\hat{g}_{*j}$ are the associated unknown parameters. This expression can now be used to derive a new orthogonal estimator for the nonlinear rational model.

4.1 Orthogonal transform

Consider the orthogonal equation in eqn (4.1), where the numerator term $w_{nj}(t)$ and denominator term $w_{dj}(t)$ are defined as

$$\begin{aligned} w_{nj}(t) &= p_{nj}(t) - \sum_{i=1}^{j-1} \hat{\alpha}_{nnij} w_{ni}(t) - \sum_{i=2}^{dl} \hat{\alpha}_{ndij} w_{di}(t) \\ w_{dj}(t) &= -\, p_{dj}(t)y(t) - \sum_{i=1}^{nl} \hat{\alpha}_{dnij} w_{ni}(t) - \sum_{i=2}^{j-1} \hat{\alpha}_{ddij} w_{di}(t) \end{aligned} \tag{4.1.1}$$

where the ordering of the numerator and denominator terms $w_{ni}(t)$ and $w_{di}(t)$ is arbitrary, dl is the number of denominator terms selected and nl is the number of numerator terms selected.

As discussed in section 3, there is an inherent error which is induced in the denomiantor term by the noise. From the orthogonal transform of the original equation given in eqn (2.3.1), this inherent error will propagate to numerator terms in the auxillary model eqn (4.1) because the orthogonally transformed terms are backwards dependent. This problem can be illustrated by considering the orthogonal transform of the first denominator term from eqn (4.1.1) and eqn (2.3.2) to give

$$\begin{aligned} w_{d2}(t) &= -\,p_{d2}(t)\;y(t) \\ &= -\,p_{d2}(t)\,\frac{a(t)}{b(t)} - p_{d2}(t)\;e(t) \end{aligned} \tag{4.1.2}$$

Let

$$\begin{aligned} ww_{d2}(t) &= -\,p_{d2}(t)\,\frac{a(t)}{b(t)} \\ e_{d2}(t) &= -\,p_{d2}(t) \end{aligned} \tag{4.1.3}$$

then eqn (4.1.2) may be rewritten as

$$w_{d2}(t) = ww_{d2}(t) + e_{d2}(t)\;e(t) \tag{4.1.4}$$

where $e_{d2}(t)e(t)$ indicates the existence of the inherent error in the orthogonal transform and represents all terms which include the factor $e(t)$, all other terms (which may include $e(t-j)$, $j > 0$ elements) are combined in $ww_{d2}(t)$. Therefore define

$$\begin{aligned} ww_{nj}(t) &= p_{nj}(t) - \sum_{i=1}^{j-1}\hat{\alpha}_{nnij}ww_{ni}(t) - \sum_{i=2}^{dl}\hat{\alpha}_{ndij}ww_{di}(t) \\ ww_{dj}(t) &= -\,p_{dj}(t)\frac{a(t)}{b(t)} - \sum_{i=1}^{nl}\hat{\alpha}_{dnij}ww_{ni}(t) - \sum_{i=2}^{j-1}\hat{\alpha}_{ddij}ww_{di}(t) \end{aligned} \tag{4.1.5}$$

and

$$\begin{aligned} e_{nj}(t) &= -\sum_{i=1}^{j-1}\hat{\alpha}_{nnij}e_{ni}(t) - \sum_{i=2}^{dl}\hat{\alpha}_{ndij}e_{di}(t) \\ e_{dj}(t) &= -\,p_{dj}(t) - \sum_{i=1}^{nl}\hat{\alpha}_{dnij}e_{ni}(t) - \sum_{i=2}^{j-1}\hat{\alpha}_{ddij}e_{di}(t) \end{aligned} \tag{4.1.6}$$

similarly where $e_{*j}(t)e(t)$ represents all terms which include the factor $e(t)$, all other terms (which may include $e(t-j)$, $j > 0$ elements) are combined in $ww_{*j}(t)$. Consequently as j is incremented in eqn (4.1.5) and eqn (4.1.6) $e_{ni}(t)$ and $e_{di}(t)$ are known for all $i<j$ and can be used to compute and remove the bias.

Hence the eqn (4.1.1) can be written as

$$\begin{aligned} w_{nj}(t) &= ww_{nj}(t) + e_{nj}(t)e(t) \\ w_{dj}(t) &= ww_{dj}(t) + e_{dj}(t)e(t) \end{aligned} \tag{4.1.7}$$

Alternatively eqn (4.1) can be written in the form of

$$Y(t) = \sum_{j=1}^{num} w_{nj}(t)\hat{g}_{nj} + \sum_{j=2}^{den} w_{dj}(t)\hat{g}_{dj} + b(t)e(t)$$

$$= \sum_{j=1}^{num} ww_{nj}(t)\hat{g}_{nj} + \sum_{j=2}^{den} ww_{dj}(t)\hat{g}_{dj} + \Big(\sum_{j=1}^{num} e_{nj}(t)\hat{g}_{nj} + \sum_{j=2}^{den} e_{dj}(t)\hat{g}_{dj} + b(t) \Big)e(t)$$

$$= \sum_{j=1}^{num} ww_{nj}(t)\hat{g}_{nj} + \sum_{j=2}^{den} ww_{dj}(t)\hat{g}_{dj} + p_{d1}(t)e(t) \tag{4.1.8}$$

With reference to the definitions given in eqn (4.1.7), it should be noticed that the terms $ww_{nj}(t)$ and $ww_{dj}(t)$ cannot be measured. These expressions will be used to derive some formulae for the orthogonal transform in this section and for the parameter estimation in sectoin 4.3 and some properties of the err test used for structure detection in section 4.4.

The coefficients $\hat{\alpha}_{nnij}$, $\hat{\alpha}_{ndij}$, $\hat{\alpha}_{dnij}$, and $\hat{\alpha}_{ddij}$ in eqn (4.1.1) can be computed directly.

Coefficient α_{nnij} is given by

$$\alpha_{nnij} = \frac{\overline{w_{ni}(t)\ p_{nj}(t)}}{\overline{w_{ni}^2(t)}}$$

$$= \frac{\overline{(ww_{ni}(t) + e_{ni}(t)e(t))\ p_{nj}(t)}}{\overline{(ww_{ni}(t) + e_{ni}(t)e(t))^2}} \tag{4.1.9}$$

where the overbar denotes time averaging, and by definition $e(t)$ is zero mean white noise sequence which is independent of $p_{nj}(t)$ and $e_{ni}(t)$ respectively such that

$$\alpha_{nnij} = \frac{\overline{ww_{ni}(t)\ p_{nj}(t)}}{\overline{ww_{ni}^2(t)} + \overline{e_{ni}^2(t)}\ \sigma_e^2} \tag{4.1.10}$$

An unbiased estimate of α_{nnij} can be obtained by substracting $\overline{e_{ni}^2(t)}\ \sigma_e^2$ in the denominator of α_{nnij}

$$\hat{\alpha}_{nnij} = \frac{\overline{ww_{ni}(t)\ p_{nj}(t)}}{\overline{ww_{ni}^2(t)} + \overline{e_{ni}^2(t)}\ \sigma_e^2 - \overline{e_{ni}^2(t)}\ \sigma_e^2}$$

$$= \frac{\overline{ww_{ni}(t)\ p_{nj}(t)}}{\overline{ww_{ni}^2(t)}} \tag{4.1.11}$$

Similarly the coefficient α_{ndij} is given by

$$\alpha_{ndij} = \frac{\overline{w_{di}(t)\ p_{nj}(t)}}{\overline{w_{di}^2(t)}}$$

$$= \frac{\overline{(ww_{di}(t) + e_{di}(t)e(t))\ p_{nj}(t)}}{\overline{(ww_{di}(t) + e_{di}(t)e(t))^2}} \tag{4.1.12}$$

Since $e(t)$ is a zero mean white noise sequence eqn (4.1.5) becomes

$$\alpha_{ndij} = \frac{\overline{ww_{di}(t)\ p_{nj}(t)}}{\overline{ww_{di}^2(t)} + \overline{e_{di}^2(t)}\ \sigma_e^2} \qquad (4.1.13)$$

An unbiased estimate of α_{ndij} can be obtained by substracting $\overline{e_{di}^2(t)}\ \sigma_e^2$ in the denominator of α_{ndij}

$$\hat{\alpha}_{ndij} = \frac{\overline{ww_{di}(t)\ p_{nj}(t)}}{\overline{ww_{di}^2(t)} + \overline{e_{di}^2(t)}\ \sigma_e^2 - \overline{e_{di}^2(t)}\ \sigma_e^2}$$

$$= \frac{\overline{ww_{di}(t)\ p_{nj}(t)}}{\overline{ww_{di}^2(t)}} \qquad (4.1.14)$$

The coefficient α_{dnij} is given by

$$\alpha_{dnij} = -\ \frac{\overline{w_{ni}(t)\ p_{dj}(t) y(t)}}{\overline{w_{ni}^2(t)}}$$

$$= -\ \frac{\overline{(\ ww_{ni}(t) + e_{ni}(t)e(t)\ )\ p_{dj}(t)(\frac{a(t)}{b(t)} + e(t))}}{\overline{(\ ww_{ni}(t) + e_{ni}(t)e(t)\ )^2}}$$

$$= -\ \frac{\overline{ww_{ni}(t)\ p_{dj}(t)\ \frac{a(t)}{b(t)}} + \overline{p_{dj}(t)\ e_{ni}(t)}\ \sigma_e^2}{\overline{ww_{ni}^2(t)} + \overline{e_{ni}^2(t)}\ \sigma_e^2} \qquad (4.1.15)$$

An unbiased estimate of α_{dnij} can be obtained by substracting $\overline{p_{dj}(t)\ e_{ni}(t)}\ \sigma_e^2$ and $\overline{e_{ni}^2(t)}\ \sigma_e^2$ in the numerator and denominator respectively of α_{dnij}

$$\hat{\alpha}_{dnij} = -\ \frac{\overline{ww_{ni}(t)\ p_{dj}(t)\ \frac{a(t)}{b(t)}} + \overline{p_{dj}(t)\ e_{ni}(t)}\ \sigma_e^2 - \overline{p_{dj}(t)\ e_{ni}(t)}\ \sigma_e^2}{\overline{ww_{ni}^2(t)} + \overline{e_{ni}^2(t)}\ \sigma_e^2 - \overline{e_{ni}^2(t)}\ \sigma_e^2}$$

$$= -\ \frac{\overline{ww_{ni}(t)\ p_{dj}(t)\ \frac{a(t)}{b(t)}}}{\overline{ww_{ni}^2(t)}} \qquad (4.1.16)$$

Similarly the coefficient α_{ddij} is given by

$$\alpha_{ddij} = -\ \frac{\overline{w_{di}(t)\ p_{dj}(t) y(t)}}{\overline{w_{di}^2(t)}}$$

$$= -\ \frac{\overline{(\ ww_{di}(t) + e_{di}(t)e(t)\ )\ p_{dj}(t)(\frac{a(t)}{b(t)} + e(t))}}{\overline{(\ ww_{di}(t) + e_{di}(t)e(t)\ )^2}}$$

$$= - \frac{\overline{ww_{di}(t)\, p_{dj}(t)\, \frac{a(t)}{b(t)}} + \overline{p_{dj}(t)\, e_{di}(t)}\, \sigma_e^2}{\overline{ww_{di}^2(t)} + \overline{e_{di}^2(t)}\, \sigma_e^2} \qquad (4.1.17)$$

An unbiased estimate of α_{ddij} can be obtained by substracting $\overline{p_{dj}(t)\, e_{di}(t)\, \sigma_e^2}$ and $\overline{e_{di}^2(t)}\, \sigma_e^2$ in the numerator and denominator respectively of α_{ddij}

$$\hat{\alpha}_{ddij} = - \frac{\overline{ww_{di}(t)\, p_{dj}(t)\, \frac{a(t)}{b(t)}} + \overline{p_{dj}(t)\, e_{di}(t)}\, \sigma_e^2 - \overline{p_{dj}(t)\, e_{di}(t)}\, \sigma_e^2}{\overline{ww_{di}^2(t)} + \overline{e_{di}^2(t)}\, \sigma_e^2 - \overline{e_{di}^2(t)}\, \sigma_e^2}$$

$$= - \frac{\overline{ww_{di}(t)\, p_{dj}(t)\, \frac{a(t)}{b(t)}}}{\overline{ww_{di}^2(t)}} \qquad (4.1.18)$$

4.2 Initial settings

Initial values are required in the orthogonal transform given in eqn (4.1.1). If the first term selected is a numerator term these values can be set as

$$w_{n1}(t) = p_{ni}(t) = ww_{n1}(t)$$

and

$$e_{n1}(t) = 0 \qquad (4.2.1)$$

Alternatively if the first term selected is a denominator term, then set

$$w_{d2}(t) = - p_{d2}(t)\, y(t)$$

$$= - \left(p_{d2}(t)\, \frac{a(t)}{b(t)} + p_{d2}(t)\, e(t) \right)$$

$$= ww_{d2}(t) + e_{d2}(t)\, e(t)$$

and

$$e_{d2}(t) = - p_{d2}(t) \qquad (4.2.2)$$

4.3 Parameter estimation

The unkown parameters $\hat{g}_{*j}$ associated with the orthogonal terms in the auxiliary eqaution eqn (4.1) must now be determined. From the orthogonality of $w_{*j}(t)$ and with refrence to eqn (2.4.2), eqn (4.1), and eqn (4.1.7), define

$$g_{nj} = \frac{\overline{w_{nj}(t)\, Y(t)}}{\overline{w_{nj}^2(t)}}$$

$$= \frac{\overline{(ww_{nj}(t) + e_{nj}(t)e(t))\ p_{d1}(t)(\frac{a(t)}{b(t)} + e(t))}}{\overline{(ww_{nj}(t) + e_{nj}(t)e(t))^2}}$$

$$= \frac{\overline{ww_{nj}(t)\ p_{d1}(t)\ \frac{a(t)}{b(t)}} + \overline{p_{d1}(t)\ e_{nj}(t)}\ \sigma_e^2}{\overline{ww_{nj}^2(t)} + \overline{e_{nj}^2(t)}\ \sigma_e^2}$$

and

$$g_{dj} = \frac{\overline{w_{dj}(t)\ Y(t)}}{\overline{w_{dj}^2(t)}}$$

$$= \frac{\overline{(ww_{dj}(t) + e_{dj}(t)e(t))\ p_{d1}(t)(\frac{a(t)}{b(t)} + e(t))}}{\overline{(ww_{dj}(t) + e_{dj}(t)e(t))^2}}$$

$$= \frac{\overline{ww_{dj}(t)\ p_{d1}(t)\ \frac{a(t)}{b(t)}} + \overline{p_{d1}(t)\ e_{dj}(t)}\ \sigma_e^2}{\overline{ww_{dj}^2(t)} + \overline{e_{dj}^2(t)}\ \sigma_e^2} \qquad (4.3.1)$$

Obviously the bias in the parameter estimates is caused by the noise $e(t)$ which is present in the terms $w_{nj}(t)$ and $w_{dj}(t)$. Unbiased estimates of g_{nj} and g_{dj} can be obtained by substracting the bias terms in g_{nj} and g_{dj} to yield

$$\hat{g}_{nj} = \frac{\overline{ww_{nj}(t)\ p_{d1}(t)\ \frac{a(t)}{b(t)}} + \overline{p_{d1}(t)\ e_{nj}(t)}\ \sigma_e^2 - \overline{p_{d1}(t)\ e_{nj}(t)}\ \sigma_e^2}{\overline{ww_{nj}^2(t)} + \overline{e_{nj}^2(t)}\ \sigma_e^2 - \overline{e_{nj}^2(t)}\ \sigma_e^2}$$

$$= \frac{\overline{ww_{nj}(t)\ p_{d1}(t)\ \frac{a(t)}{b(t)}}}{\overline{ww_{nj}^2(t)}}$$

and similarly

$$\hat{g}_{dj} = \frac{\overline{ww_{dj}(t)\ p_{d1}(t)\ \frac{a(t)}{b(t)}} + \overline{p_{d1}(t)\ e_{dj}(t)}\ \sigma_e^2 - \overline{p_{d1}(t)\ e_{dj}(t)}\ \sigma_e^2}{\overline{ww_{dj}^2(t)} + \overline{e_{dj}^2(t)}\ \sigma_e^2 - \overline{e_{dj}^2(t)}\ \sigma_e^2}$$

$$= \frac{\overline{ww_{dj}(t)\ p_{d1}(t)\ \frac{a(t)}{b(t)}}}{\overline{ww_{dj}^2(t)}} \qquad (4.3.2)$$

Notice that througtout the bias is dependent on σ_e^2 simply because of the assumption that $e(t)$ has been reduced to a zero mean white noise sequence by the action of fitting a noise model.

4.4 Error reduction ratio computation

Parameter estimation must be combined with structure detection if parsimonious nonlinear models are to be obtained. Detecting which terms should be included within the model and which should be discarded can be acheived by exploiting the properties of eqn (4.1) and by defining an error reduction ratio (err) test (Korenberg, Billings, Liu, and Mcllroy 1988). The err test which is a by-product of the orthogonal estimation algorithm must be rederived for the rational model because of the inherent error. Consider eqn (4.1)

$$Y(t) = \sum_{j=1}^{num} w_{nj}(t)\hat{g}_{nj} + \sum_{j=2}^{den} w_{dj}(t)\hat{g}_{dj} + b(t)e(t) \qquad (4.4.1)$$

Squaring eqn (4.4.1) and taking expected value gives

$$\overline{Y^2(t)} = \sum_{j=1}^{num} \hat{g}_{nj}^2 \overline{w_{nj}^2(t)} + \sum_{j=2}^{den} \hat{g}_{dj}^2 \overline{w_{dj}^2(t)}$$

$$+ 2\sum_{j=1}^{num} \hat{g}_{nj} \overline{w_{nj}(t)b(t)e(t)} + 2\sum_{j=2}^{den} \hat{g}_{dj} \overline{w_{dj}(t)b(t)e(t)} + \overline{b^2(t)}\, \sigma_e^2 \qquad (4.4.2)$$

From eqn (4.4.2) the first two terms on the right hand side are given by

$$\sum_{j=1}^{num} \hat{g}_{nj}^2 \overline{w_{nj}^2(t)} = \sum_{j=1}^{num} \hat{g}_{nj}^2 \overline{(ww_{nj}(t) + e_{nj}(t)e(t))^2}$$

$$= \sum_{j=1}^{num} (\hat{g}_{nj}^2 \overline{ww_{nj}^2(t)} + \overline{e_{nj}^2(t)}\, \sigma_e^2)$$

and

$$\sum_{j=2}^{den} \hat{g}_{dj}^2 \overline{w_{dj}^2(t)} = \sum_{j=2}^{den} \hat{g}_{dj}^2 \overline{(ww_{dj}(t) + e_{dj}(t)e(t))^2}$$

$$= \sum_{j=2}^{den} (\hat{g}_{dj}^2 \overline{ww_{dj}^2(t)} + \overline{e_{dj}^2(t)}\, \sigma_e^2) \qquad (4.4.3)$$

and the third and fourth terms on the right hand side are given by

$$\sum_{j=1}^{num} \hat{g}_{nj} \overline{w_{nj}(t)b(t)e(t)} = \sum_{j=1}^{num} \hat{g}_{nj} \overline{(ww_{nj}(t) + e_{nj}e(t))b(t)e(t)}$$

$$= \sum_{j=1}^{num} \hat{g}_{nj} \overline{e_{nj}(t)b(t)}\, \sigma_e^2$$

and

$$\sum_{j=2}^{den}\hat{g}_{dj}\ \overline{w_{dj}(t)b(t)e(t)} = \sum_{j=2}^{den}\hat{g}_{dj}\ \overline{(ww_{dj}(t) + e_{dj}e(t))b(t)e(t)}$$

$$= \sum_{j=2}^{den}\hat{g}_{dj}\ \overline{e_{dj}(t)b(t)}\ \sigma_e^2 \qquad (4.4.4)$$

Substituting eqn (4.4.3) and eqn (4.4.4) into eqn (4.4.2) and then multiplying by $\dfrac{1}{\overline{Y^2(t)}\ \overline{b^2(t)}}$ on both sides gives

$$\frac{1}{\overline{b^2(t)}} = \sum_{j=1}^{num}\frac{\hat{g}_{nj}^2\ \overline{ww_{nj}^2(t)} + \overline{e_{nj}^2(t)}\ \sigma_e^2 + 2\hat{g}_{nj}\ \overline{e_{nj}(t)b(t)}\ \sigma_e^2}{\overline{Y^2(t)}\ \overline{b^2(t)}}$$

$$+ \sum_{j=2}^{den}\frac{\hat{g}_{dj}^2\ \overline{ww_{dj}^2(t)} + \overline{e_{dj}^2(t)}\ \sigma_e^2 + 2\hat{g}_{dj}\ \overline{e_{dj}(t)b(t)}\ \sigma_e^2}{\overline{Y^2(t)}\ \overline{b^2(t)}} + \frac{\sigma_e^2}{\overline{Y^2(t)}} \qquad (4.4.5)$$

Define the estimated error reduction ratio as

$$err_{nj} = \frac{\hat{g}_{nj}^2\ \overline{ww_{nj}^2(t)} + \overline{e_{nj}^2(t)}\ \sigma_e^2 + 2\hat{g}_{nj}\ \overline{e_{nj}(t)b(t)}\ \sigma_e^2}{\overline{Y^2(t)}\ \overline{b^2(t)}}$$

$$err_{dj} = \frac{\hat{g}_{dj}^2\ \overline{ww_{dj}^2(t)} + \overline{e_{dj}^2(t)}\ \sigma_e^2 + 2\hat{g}_{dj}\ \overline{e_{dj}(t)b(t)}\ \sigma_e^2}{\overline{Y^2(t)}\ \overline{b^2(t)}} \qquad (4.4.6)$$

introduce

$$e\hat{r}r_{nj} = \frac{\hat{g}_{nj}^2\ \overline{ww_{nj}^2(t)}}{\overline{Y^2(t)}\ \overline{b^2(t)}}$$

$$e\hat{r}r_{dj} = \frac{\hat{g}_{dj}^2\ \overline{ww_{dj}^2(t)}}{\overline{Y^2(t)}\ \overline{b^2(t)}} \qquad (4.4.7)$$

as the err estimates that would arise if $e(t) = 0$, and

$$Bias\ [\ e\hat{r}r_{nj}\ ] = \frac{\overline{e_{nj}^2(t)}\ \sigma_e^2 + 2\hat{g}_{nj}\ \overline{e_{nj}(t)b(t)}\ \sigma_e^2}{\overline{Y^2(t)}\ \overline{b^2(t)}}$$

$$Bias\ [\ e\hat{r}r_{dj}\ ] = \frac{\overline{e_{dj}^2(t)}\ \sigma_e^2 + 2\hat{g}_{dj}\ \overline{e_{dj}(t)b(t)}\ \sigma_e^2}{\overline{Y^2(t)}\ \overline{b^2(t)}} \qquad (4.4.8)$$

as the biases which are induced in the err estimates for the realistic case of $e(t) \neq 0$.

An unbiased estimate of err for the rational model can therefore be estimated using

$$e\hat{r}r_{nj} = err_{nj} - Bias\ [\ e\hat{r}r_{nj}\ ]$$

$$e\hat{r}r_{dj} = err_{dj} - Bias\ [\ e\hat{r}r_{dj}\ ] \tag{4.4.9}$$

where err_{nj}, $Bias\ [\ e\hat{r}r_{nj}\ ]$, err_{dj}, and $Bias\ [\ e\hat{r}r_{dj}\ ]$ are obtained directly from the computations.

With reference to the definations in eqns (4.4.6), (4.4.7), and (4.4.8), eqn (4.4.5) can alternatively be written as

$$\frac{1}{\sigma_b^2} = \sum_{j=1}^{num} err_{nj} + \sum_{j=2}^{den} err_{dj} + \frac{\sigma_e^2}{\sigma_Y^2}$$

$$= \sum_{j=1}^{num} e\hat{r}r_{nj} + \sum_{j=2}^{den} e\hat{r}r_{dj} + \sum_{j=1}^{num} bias\ [\ e\hat{r}r_{nj}\ ] + \sum_{j=2}^{den} bias\ [\ e\hat{r}r_{dj}\ ] + \frac{\sigma_e^2}{\sigma_Y^2} \tag{4.4.10}$$

where

$$\sigma_b^2 = \overline{b^2(t)}$$
$$\sigma_Y^2 = \overline{Y^2(t)} \tag{4.4.11}$$

Enq (4.4.10) can be used as an information criterion for determining the number of terms to be included in the model, it therefore determines the model structure. The larger the value of err the more the ratio $\frac{\sigma_e^2}{\sigma_Y^2}$ will be reduced, and hence the corresponding term should be included in the model. Insignificant terms can be rejected by defining a cut off value of $1 - \sum e\hat{r}r_{*j}$ below which terms are deemed to be negligible. As a criterion err compromises the prediction accuracy and complexity of a final model.

There are several alternative term selection methods, most of these including a stepwise regression algorithm and a log determinant ratio have been studied respectively by Billings and Voon (1986) and Leontaritis and Billings (1987).

4.4 Implementation of the algorithm

Implementation of the algorithm requires a knowledge of the noise variance σ_e^2, this can be obtained by a recursive routine like

$$\hat{\Theta}(k) = Orth.\ Estimator\ (.,\ \hat{\sigma}_e^2(k-1))$$
$$\hat{\sigma}_e^2(k) = \frac{1}{N-md} \sum_{t=md+1}^{N} (y(t) - \frac{a(.,\ \hat{\Theta}(k))}{b(.,\ \hat{\Theta}(k))})^2 \tag{4.5.1}$$

where *Orth. Estimator* denotes the algorithm presented in this section, "." denotes the terms selected in the model, k is the iteration index, N is the data length and md is the maximum lag in the terms.

The general orthogonal rational model estimator (ORME) with arbitrary order selection for numerator and denominator terms can now be summarised as follows

$$w_{*j}(t) = \Delta_1(t) - \sum_{i=1}^{j-1} \hat{\alpha}_{**ij} w_{*i}(t)$$

$$e_{*j}(t) = \Delta_2(t) - \sum_{i=1}^{j-1} \hat{\alpha}_{**ij} e_{*i}(t)$$

$$\hat{\alpha}_{**ij} = \frac{\overline{w_{*i}(t)\Delta_1(t)} - \overline{e_{*i}(t)\Delta_2(t)}\hat{\sigma}_e^2}{\overline{w_{*i}^2(t)} - \overline{e_{*i}^2(t)}\hat{\sigma}_e^2}$$

$$\hat{g}_{*j} = \frac{\overline{w_{*j}(t)Y(t)} - \overline{e_{*j}(t)p_{d1}(t)}\hat{\sigma}_e^2}{\overline{w_{*j}^2(t)} - \overline{e_{*j}^2(t)}\hat{\sigma}_e^2}$$

$$\hat{b}(t) = 1 + \sum_{j=2}^{den} p_{dj}(t)\hat{\theta}_{dj}$$

$$e\hat{r}r_{*j} = \frac{\hat{g}_{*j}^2\overline{w_{*j}^2(t)} - \hat{g}_{*j}^2\overline{e_{*j}^2(t)}\ \hat{\sigma}_e^2 - 2\hat{g}_{*j}\overline{e_{*j}(t)\hat{b}(t)}\ \hat{\sigma}_e^2}{\overline{Y^2(t)}\ \overline{\hat{b}^2(t)}}$$

$$\hat{\sigma}_e^2 = \frac{1}{N-md}\sum_{t=md+1}^{N}(y(t) - \frac{a(.,\hat{\Theta}(k))}{b(.,\hat{\Theta}(k))})^2 \qquad (4.5.2)$$

where * denotes either n or d, k is the iteration index, and

$$\Delta_1(t) = \begin{cases} p_{nj}(t), & \text{for } \textit{the numerator} \\ -p_{dj}(t)y(t), & \text{for } \textit{the denominator} \end{cases}$$

$$\Delta_2(t) = \begin{cases} 0, & \text{for } \textit{the numerator} \\ -p_{dj}(t), & \text{for } \textit{the denominator} \end{cases} \qquad (4.5.3)$$

The parameters in the original model given in eqn (2.4.1) can then be computed from

$$\hat{\theta}_{num+den-1}(k) = \hat{g}_{num+den-1}(k)$$

$$\hat{\theta}_i(k) = \hat{g}_i(k) - \sum_{j=i+1}^{num+den-1} \hat{\alpha}_{ij}(k)\hat{\theta}_j(k), \ i = num+den-2, \cdots, 1 \qquad (4.5.4)$$

where $\hat{g}_i$ denotes either $\hat{g}_{ni}$ or $\hat{g}_{di}$ and similarly $\hat{\theta}_i$ for $\hat{\theta}_{ni}$ or $\hat{\theta}_{di}$, $\hat{\alpha}_i$ for $\hat{\alpha}_{**ij}$.

A comparison with the original algorithm derived for the polynomial NARMAX model, which was given in Korenberg (1985), Korenberg, Billings, Liu, and McIlroy (1988), and Chen, Billings, and Luo (1989), shows that the new ORME algorithm adds extra formulae for $e_{*j}(t)$, $\hat{\sigma}_e^2$, and $\hat{b}(t)$ which are used to remove the inherent bias.

Secondly all the averaging operations for $\hat{\alpha}_{**ij}$, $\hat{g}_{*j}$, and $e\hat{r}r_{*j}$ also include bias removal terms and thirdly an iteration framework for the estimation of the noise variance $\hat{\sigma}_e^2$ has to be adopted.

This algorithm will reduce to the original algorithm derived for polynomial NARMAX model if $b(t) = 1$.

The ORME algorithm may be programmed by the following steps:

(i) Set $\hat{\sigma}_e^2 = 0$, err cut off and max iteration.

(ii) Select terms and estimate associated parameters with the formulae of eqn (4.5.2).

(iii) Compute the noise sequence $e(t)$ and estimate the variance σ_e^2 with the selected model.

(iv) Go back to step (ii) and repeat until $\hat{\sigma}_e^2$ converges to a constant value or the max iteration is exceeded.

5 Simulation studies

Three simulation examples were chosen to illustrate the application of the ORME algorithm for term selection and parameter estimation of stochastic nonlinear rational models. In all three examples 1000 pairs of input and output data were used with the same input and noise signals. The input $u(t)$ was a zero mean uniform random sequence with amplitude eange ±1 (variance $\hat{\sigma}_u^2 = 0.33$) and the noise $e(t)$ was a zero mean Gaussian sequence with variance $\hat{\sigma}_e^2 = 0.01$.

The initial model specification consisted of 20 terms for the first two examples and this was used as the full model with *numerator degree* = *denominator degree* = 2 and *input lag* = *output lag* = *noise lag* = 1, and a full model consisting of 56 terms for the third example was specified with *numerator degree* = *denominator degree* = 2 and *input lag* = *output lag* = *noise lag* = 2. In each case the true model structure is represented by just three or four terms from the much larger model set. Both the structure, or terms to include in the model, and the unknown parameters are estimated using the ORME algorithm with no apriori information whatever.

Example S_1 consisted of the rational model

$$y(t) = \frac{a(t)}{b(t)} + e(t) = \frac{y^2(t-1) + u(t-1) + e(t-1)}{1 + y^2(t-1) + y(t-1)u(t-1)} + e(t) \tag{5.1}$$

The linear in the parameters expression for this model is

$$Y(t) = y^2(t-1) + u(t-1) + e(t-1) - y(t)y^2(t-1) - y(t)y(t-1)u(t-1) + b(t)e(t)$$

(5.2)

where

$$Y(t) = y(t) \tag{5.3}$$

The input and output data sequences for this example are shown in Fig. 1.1. For the detection of the model structure and the estimation of the parameter a cut off point of 0.02 was used for including terms in err and an initial value of $\hat{\sigma}_e^2 = 0.0$. A model with five selected terms was fitted and the parameter estimates after six iterations are listed in Table 1.1. The one step ahead predictions and residuals are illustrated in Fig. 1.2. The model validity tests are shown in Fig. 1.3. All these results indicate that the ORME algorithm produced a good unbiased S_1.

Example S_2 consisted of the output affine model

$$y(t) = \frac{a(t)}{b(t)} + e(t) = \frac{y(t-1) + u(t-1)e(t-1)}{1 + u^2(t-1)} + e(t) \tag{5.4}$$

The linear in the parameters expression for this model is

$$Y(t) = y(t-1) + u(t-1)e(t-1) - y(t)u^2(t-1) + b(t)e(t)$$

(5.5)

where

$$Y(t) = y(t) \tag{5.6}$$

The input and output data sequence for this example are shown in Fig. 2.1. For the detection of the model structure and the estimation of the parameters a cut off point of 0.34 was used for term selection with an initial value of $\hat{\sigma}_e^2 = 0.0$. A model with three selected terms was fitted and the parameter estimates after six iterations are given in Table 2.1. The one step ahead predictions and residuals are illustrated in Fig. 2.2. The model validation were all within the 95% confidence bands. Once again the ORME algorithm gave both a correct term selection and unbiased parameter estimation.

Example S_3 consisted of the rational model

$$y(t) = \frac{a(t)}{b(t)} + e(t) = \frac{y^2(t-1) + u(t-1)u(t-2) + y(t-1)e(t-2)}{1 + y^2(t-1) + u^2(t-2)} + e(t) \tag{5.7}$$

The linear in the parameters expression for this model is

$$Y(t) = y^2(t-1) + u(t-1)u(t-2) + y(t-1)e(t-2) - y(t)y^2(t-1) - y(t)u^2(t-2) + b(t)e(t)$$

(5.8)

where

$$Y(t) = y(t) \tag{5.9}$$

This is a more complicated model than the first two examples. The input and output data sequence for this example are shown in Fig. 3.1. For the detection of the model structure and the estimation of the parameters a cut off point of 0.05 was used for term selection with an initial value of $\hat{\sigma}_e^2 = 0.0$. A model with five selected terms was fitted and the parameter estimates after six iterations are given in Table 3.1. The one step ahead predictions and residuals are illustrated in Fig. 3.2. The model validation tests were all within the 95% confidence bands. Again the ORME algorithm gave both a correct term selection and unbiased parameter estimation for this higher order lag rational model identification.

Because the estimate of the variance of the noise tends to be overestiamted in the first few iterations in all the above examples σ_e^2 was weighted as 10, 50, 90, 99, and 100% for each iteration respectively.

6 Conclusions

A new orthogonal estimation algorithm has been derived for the identification of stochastic nonlinear rational models. The algorithm is computationally straightforward and provides, without any a priori information, a method of detecting the model structure or which terms should be included in the numerator and denominator together with unbiased estimates of the unknown parameters in the presence of linear and nonlinear noise corruption.

Acknowledgements

The authors gratefully acknowledge that this work is supported by SERC under grant GR/F2417.7. Finally we wish to thank Mr. H. Jamaluddin for his help running the part of the simulation program.

References

Billings, S.A. and W.S.F. Woon, "Least squares parameter estimation algorithms for nonlinear systems," Int. J. Systems Sci., vol. 19, 1984.

Billings, S.A. and S. Chen, "Identification of nonlinear rational systems using a predicton error estimation algorithm," Int. J. Systems Sci., vol. 20, 1989.

Billings, S.A. and Q.M. Zhu, "Rational model identification using an extended least squares algorithm," Int. J. Control, vol. 54, 1991.

Braess, D., Nonlinear approximation theory, Springer Verlag, 1986.

Chen, S. and S.A. Billings, "Represntations of nonlinear systems: the NARMAX model," Int. J. Control, vol. 48, 1989.

Chen, S., S.A. Billings, and W. Luo, "Orthogonal least squares methods and their application to nonlinear system identification," Int. J. Control, vol. 50, 1989.

Feinerman, R.P. and D.J. Newman, Polynomial approximations, Waverly press Inc., 1974.

Garabedian, H.L., Approximation of functions, Elsevier publishing company, 1965.

Goodwin, G. C. and R. L. Payne, Dynamic system identification: experiment design and data analysis, Academic Press, New York, 1977.

Handscomb(ed.), D.C., Methods of numerical approximation, Oxford, Pergamon, 1966.

Hastings, C., Aprroximations for digital computers, N.J. Princeton Univ. press, 1955.

Hayes, J.G., Numerical approximation to function and data, The Athlone press, 1970.

Korenberg, M., S.A. Billings, Y.P. Liu, and P.J. Mcllroy, "Orthogonal parameter estimation algorithm for nonlinear stochastic systems," Int. J. Control, vol. 48, 1988.

Korenberg, M.J., "Orthogonal identification of nonlinear difference equation models," Mid West Symp. on Circuits and Systems, Louisville, 1985.

Leontaritis, I.J. and S.A. Billings, "Model selection and validation methods for nonlinear systems," Int. J. Control, vol. 45, 1987.

Ljung, L., System identification--Theory for the user, Prentice Hall Englewood Cliffs , New Jersey, 1987.

Marquardt, D.W., "An algorithm for least squares estimation of nonlinear parameters," Journal of the society for industrial and applied mathematics, vol. 11, 1963.

Newman, D.J., Approximation with rational functions, American mathematical society, 1978.

Sontag, E.D., Polynomial response maps--Lecture notes in control and information sciences 13, Springer - Verlag, Berlin, 1979.

Zhu, Q.M. and S.A. Billings, "Recursive parameter estimation for nonlinear rational models," Journal of Systems Engineering (accepted for publication), 1991.

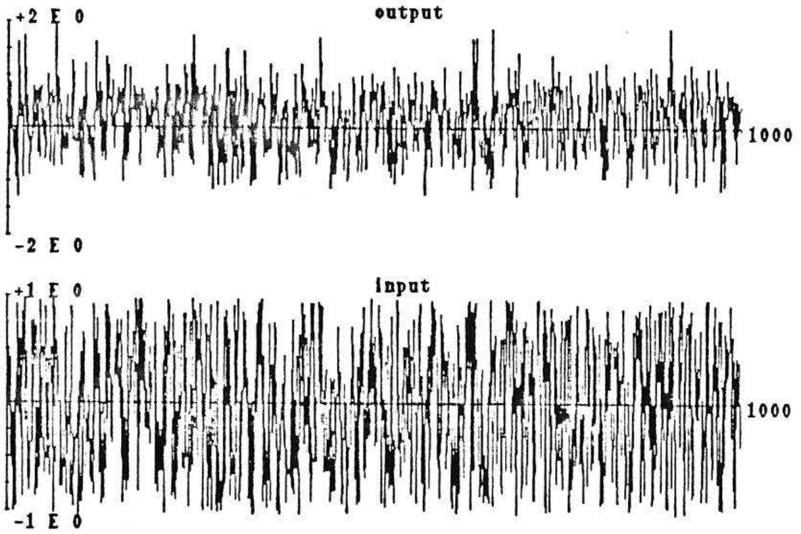


Figure 1.1 Input and output for example S_1

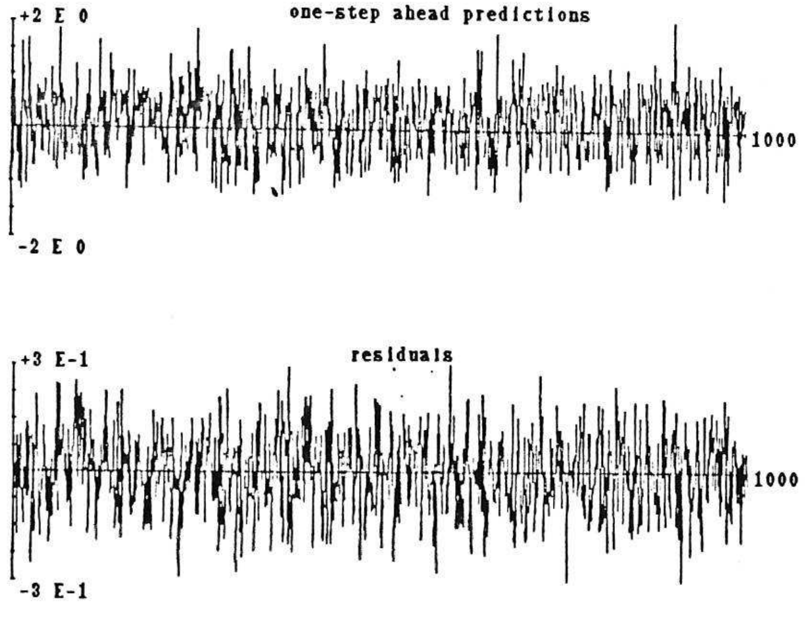


Figure 1.2 One step ahead predictions & residuals for example S_1

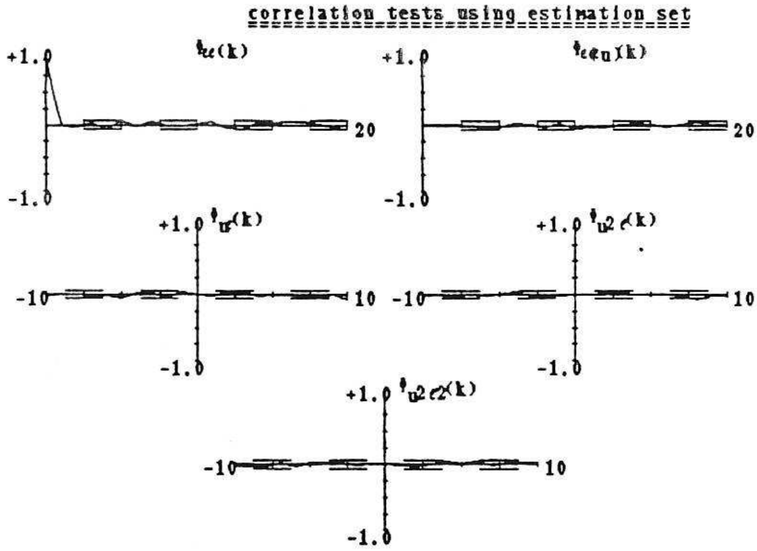


Figure 1.3 Model validation for example S_1

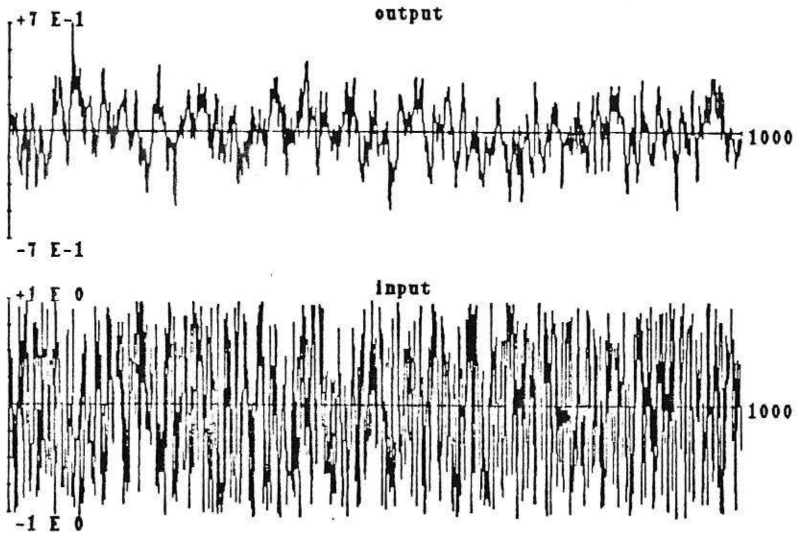


Figure 2.1 Input and output for example S_2

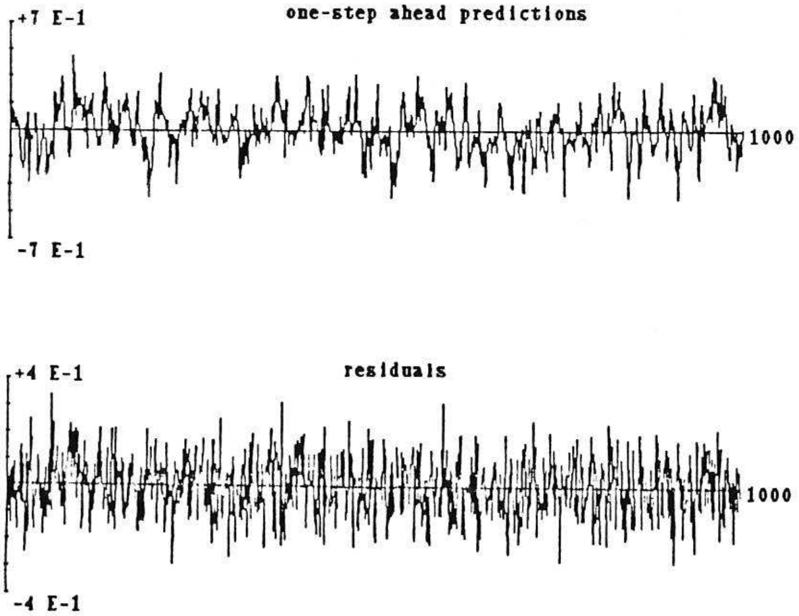


Figure 2.2 One step ahead predictions & residuals for example S_2

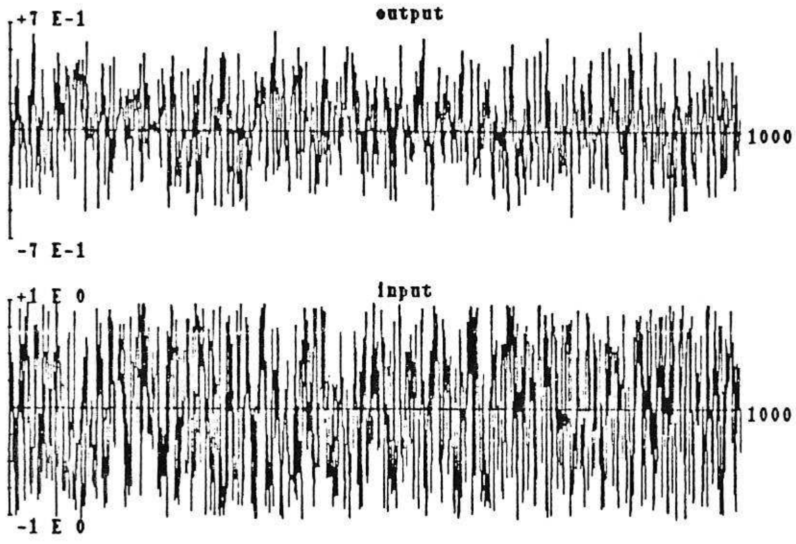


Figure 3.1 Input and output for example S_3

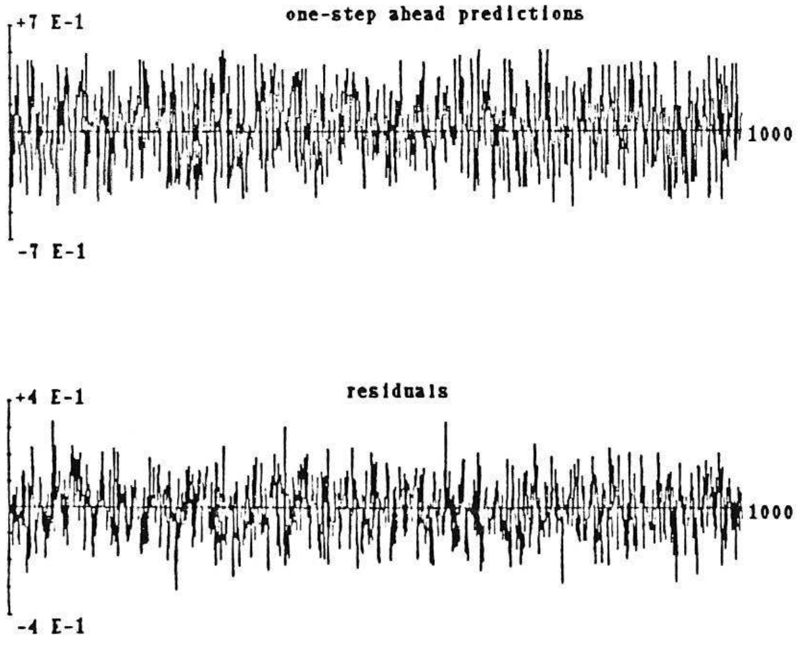


Figure 3.2 One step ahead predictions & residuals for example S_3

```
 degree of numerator=  2
 degree of denominator=  2
 order of output=   1
 order of input=   1
 order of noise=   1
 Data length= 1000, all used as estimation set
 term to regress upon:
( 11) y(t)
 sum of squares of this term over estim. set =   0.34262E+03
 variance of residuals over whole data set=  0.101849E-01
 accuracy for stopping regression:  0.020000000

           terms                                  estimates   e.r.r.s    st.de.s   o.s.

 numerator:
(  3) u(t- 1)**1=                               0.1011E+01 0.6658E+00 0.1038E-01(  1)
(  5) y(t- 1)**2=                               0.1021E+01 0.1664E+00 0.2109E-01(  2)
(  4) e(t- 1)**1=                               0.9743E+00 0.5709E-01 0.4591E-01(  5)
 denominator:
( 16) y(t- 1)**1*u(t- 1)**1*y(t)=              -0.1008E+01 0.5643E-01 0.2294E-01(  3)
( 15) y(t- 1)**2*y(t)=                         -0.1052E+01 0.8507E-01 0.3360E-01(  4)
```

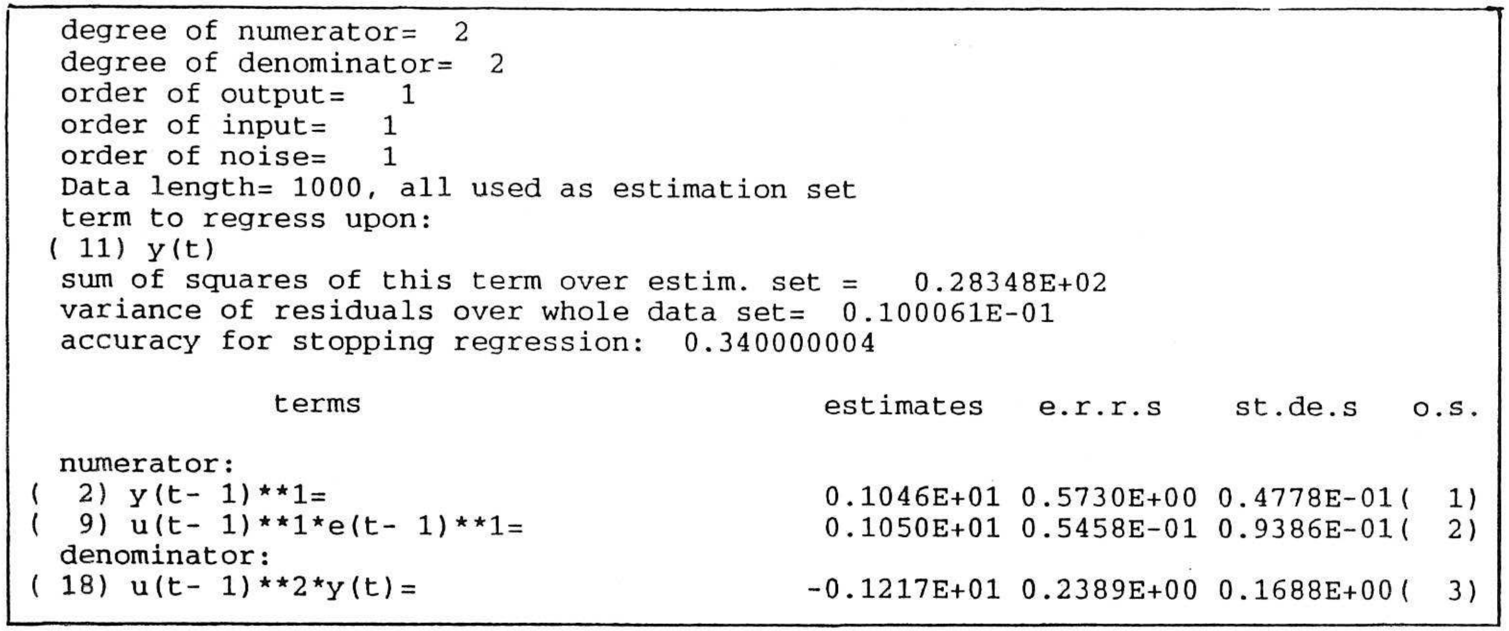
Table 1 Selected model for example S_1 using err criterion

```
 degree of numerator=  2
 degree of denominator=  2
 order of output=   2
 order of input=   2
 order of noise=   2
 Data length= 1000, all used as estimation set
 term to regress upon:
( 29) y(t)
 sum of squares of this term over estim. set =    0.59299E+02
 variance of residuals over whole data set=  0.988046E-02
 accuracy for stopping regression:  0.050000001

            terms                             estimates   e.r.r.s    st.de.s    o.s.

 numerator:
( 20) u(t- 1)**1*u(t- 2)**1=                 0.1080E+01 0.7340E+00 0.4918E-01(  1)
(  8) y(t- 1)**2=                            0.1158E+01 0.5348E-01 0.7729E-01(  2)
( 13) y(t- 1)**1*e(t- 2)**1=                 0.9606E+00 0.1962E-01 0.1968E+00(  5)
 denominator:
( 51) u(t- 2)**2*y(t)=                      -0.1213E+01 0.1330E+00 0.1299E+00(  3)
( 36) y(t- 1)**2*y(t)=                      -0.1335E+01 0.2798E-01 0.3261E+00(  4)
```

Table 3 Selected model for example S_3 using err criterion



```
  degree of numerator=  2
  degree of denominator=  2
  order of output=   1
  order of input=   1
  order of noise=   1
  Data length= 1000, all used as estimation set
  term to regress upon:
 ( 11) y(t)
  sum of squares of this term over estim. set =   0.28348E+02
  variance of residuals over whole data set=  0.100061E-01
  accuracy for stopping regression:  0.340000004

            terms                          estimates   e.r.r.s    st.de.s   o.s.

  numerator:
(  2) y(t- 1)**1=                         0.1046E+01 0.5730E+00 0.4778E-01(  1)
(  9) u(t- 1)**1*e(t- 1)**1=              0.1050E+01 0.5458E-01 0.9386E-01(  2)
  denominator:
( 18) u(t- 1)**2*y(t)=                   -0.1217E+01 0.2389E+00 0.1688E+00(  3)
```

Table 2 Selected model for example S_2 using err criterion