Monograph:

Zhu, Q.M. and Billings, S.A. (1991) Recursive Parameter Estimation for Nonlinear Rational Models. Research Report. Acse Report 420 . Dept of Automatic Control and System Engineering. University of Sheffield

Recursive Parameter Estimation for Nonlinear Rational Models

Q.M. Zhu and S.A. Billings

Department of Control Engineering

University of Sheffield

Mappin Street

Sheffield S1 3JD

U. K.

Research Report No. 420

January 1991

January 17, 1991

Recursive Parameter Estimation for Nonlinear Rational Models

Q.M. Zhu, S.A. Billings

Department of Control Engineering, University of Sheffield, S1 3JD, UK

Abstract:

A new recursive parameter estimation algorithm is derived for a general class of stochastic nonlinear systems which can be represented by a rational model defined as the ratio of two polynomial expansions of past inputs, outputs and prediction error terms. Simulation results are included to illustrate the performance of the new algorithm.

1 Introduction

Recursive identification plays an important role in the identification of complex, possibly time varying systems, adaptive control and adaptive signal processing (Ljung and Soderstrom 1983). The recursive procedures are ideally suited to implementation on a digital computer, they are computationally elegant and provide alternative formulations of the conventional batch algorithms. Recursive parameter estimation provides a new dimension to system identification, providing the possibility of online data processing, prediction and adaptive control.

Most of the existing recursive parameter estimators were derived for linear systems described by the ARMAX (AutoRegressive Moving Average with eXogenous inputs) model or variants of it. There are several well known algorithms including recursive least squares (RLS), recursive instrumental variables (RIV), recursive extended least squares (RELS), and recursive maximum likelihood (RML). Ljung and Soderstrom (1983) showed how all these methods can be unified and provided an elegant convergence analysis. Recently some of these algorithms have been applied to nonlinear systems which can be described by the linear-in-the-parameter polynomial NARMAX (Nonlinear ARMAX) model (Billings and Voon 1984, Chen and Billings 1988, 1989a).

The nonlinear rational model (Sontag 1979, Chen and Billings 1989b) can be considered within the class of NARMAX models and provides a very concise structure which can be used to describe a wide range of nonlinear systems. The rational model is defined as the ratio of two polynomial expansions of past inputs, outputs and prediction errors and is therefore nonlinear in the unknown parameters. If the model is multiplied out to yield a linear-in-the-parameters description the presence of even additive white noise induces severe bias in the parameter estimates which can not be eliminated by conventional noise modelling procedures. This problem has constrained the development of parameter estimation algorithms for nonlinear rational models to methods which can accommodate nonlinear in the parameter models (Marquardt 1963, Billings and Chen 1989). Whilst these provide excellent results in many cases the necessity of numerical minimization techniques means that they can be rather complex and computationally demanding. In an attempt to avoid these difficulties Billings and Zhu (1990) have recently shown that the identification of rational models can be performed based on a linear in the parameter expansion if the induced bias terms are properly accommodated in the algorithm. The resulting rational model estimator (RME) provides unbiased estimates of the model parameters using essentially a modified least squares formulation.

In the present study a new recursive implementation of the RME algorithm is derived. It is shown that multiplying the rational model out to be linear-in-the-parameters will induce severe bias. An iterative solution is proposed to overcome this problem and it is shown how this can be formulated to provide a new recursive or on-line algorithm for stochastic nonlinear rational models. The new algorithm which will be called RRME (Recursive Rational Model Estimator) is illustrated using simulation studies.

2 The rational model

2.1 Input-output description

A input and output stochastic rational model (Chen and Billings 1988) is defined as the ratio of two polynomial expressions of past system inputs, outputs and noise, that is

$$y(t) = \frac{a(y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))}{b(y(t-1), \cdots, y(t-r), u(t-1), \cdots, u(t-r), e(t-1), \cdots, e(t-r))} + e(t) \tag{2.1}$$

where $u(t)$ and $y(t)$ represent the input and output at time t ($t = 1, 2, \cdots$) respectively, r is the order of the model, and $e(t)$ is an unobservable independent noise with zero mean and finite variance σ_e^2.

In order to use model (2.1) as a basis for identification, a means of parameterisation is required. Define for the numerator

$$a(t) = \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} \tag{2.2}$$

and for the denominator

$$b(t) = \sum_{j=1}^{den} p_{dj}(t)\theta_{dj} \tag{2.3}$$

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

2.2 A linear-in-the-parameters expression

Identification based on the model in eqn (2.1) is complex because the model is nonlinear in the parameters. A prediction error algorithm can however be formulated (Billings and Chen 1989) but this is computationaly expensive. An alternative approach is to multiply out eqn (2.1) so that the model becomes linear in the parameters. Thus multiplying $b(t)$ on both sides of eqn (2.1) and then moving all the terms except $y(t)p_{d1}(t)\theta_{d1}$ to the right hand side gives

$$\begin{aligned} Y(t) &= a(t) - y(t) \sum_{j=2}^{den} p_{dj}(t)\theta_{dj} + b(t)e(t) \\ &= \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} - \sum_{j=2}^{den} y(t)p_{dj}(t)\theta_{dj} + \zeta(t) \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} Y(t) &= y(t)p_{d1}(t)|_{\theta_{d1}=1} \\ &= p_{d1}(t)\,\frac{a(t)}{b(t)} + p_{d1}(t)e(t) \end{aligned} \tag{2.5}$$

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

$$\begin{aligned} \zeta(t) &= b(t)e(t) \\ &= (\sum_{j=1}^{den} p_{dj}(t)\theta_{dj})e(t) \\ &= p_{d1}(t)e(t) + (\sum_{j=2}^{den} p_{dj}(t)\theta_{dj})e(t) \end{aligned} \tag{2.6}$$

where providing $e(t)$ has been reduced to an uncorrelated sequence as defined in eqn (2.1)

$$E[\zeta(t)] = E[b(t)]E[e(t)] = 0 \tag{2.7}$$

Inspection of eqn (2.5) shows that all the denominator terms $y(t)p_{dj}(t)$ implicitly include a current noise term $e(t)$ which is highly correlated with $\zeta(t)$ and will introduce bias in the parameter estimates even if $e(t)$ is a zero mean white noise sequence. This problem has been induced by making the rational model linear in the parameters. If a polynomial NARMAX model were used in eqn (2.4) then $b(t) = 1$ and there would be no terms on the right hand side involving $y(t)$.

Eqn (2.4) can be expressed as

$$\begin{aligned} Y(t) &= \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} - \sum_{j=2}^{den} y(t)p_{dj}(t)\theta_{dj} + b(t)e(t) \\ &= \sum_{j=1}^{num} p_{nj}(t)\theta_{nj} - \sum_{j=2}^{den} \frac{a(t)}{b(t)} p_{dj}(t)\theta_{dj} + p_{d1}(t)e(t) \end{aligned} \tag{2.8}$$

In vector notation this becomes

$$\begin{aligned} Y(t) &= \phi(t)\Theta + \zeta(t) \\ &= \hat{\phi}(t)\Theta + p_{d1}(t)e(t) \end{aligned} \tag{2.9}$$

where

$$\begin{aligned} \phi(t) &= [\phi_n(t) \;\; \phi_d(t)] \\ &= [p_{n1}(t) \cdots p_{nnum}(t) \; -p_{d2}(t)y(t) \cdots -p_{dden}(t)y(t)] \\ &= [p_{n1}(t) \cdots p_{nnum}(t) \; -p_{d2}(t)(\frac{a(t)}{b(t)} + e(t)) \cdots -p_{dden}(t)(\frac{a(t)}{b(t)} + e(t))] \end{aligned} \tag{2.10}$$

$$\begin{aligned} \Theta^T &= [\Theta_n \; \Theta_d] \\ &= [\theta_{n1} \cdots \theta_{nnum} \; \theta_{d2} \cdots \theta_{dden}] \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} \hat{\phi}(t) &= [\phi_n(t) \;\; \hat{\phi}_d(t)] \\ &= [p_{n1}(t) \cdots p_{nnum}(t) \; -p_{d2}(t)\frac{a(t)}{b(t)} \cdots -p_{dden}(t)\frac{a(t)}{b(t)}] \end{aligned} \tag{2.12}$$

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

3 En bloc least squares parameter estimation

The least squares technique is well known in system identification because it provides a simple analytical solution to the parameter estimation problem, Billings and Zhu (1990) have recently proved that most of these properties can be retained even for rational model estimation providing the noise problem discussed above is suitably accommodated.

In the following sections, the error sources are analysed by evaluating the least squares bias and covariance and it is shown how these results can be used to produce an off line or en bloc linear-in-the-parameter rational model estimator (RME). A new on line version of the estimator called RRME, the recursive rational model estimator, is then derived. The structure of the model is assumed to be known as is the case for most recursive estimation algorithms.

3.1 Least squares estimation and error analysis

The well known least squares estimate is given by

$$\hat{\Theta} = [\Phi^T\Phi]^{-1}\,\Phi^T\vec{Y} \tag{3.1}$$

where

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

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

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

$$\vec{Y} = [Y(1) \cdots Y(N)]^T \tag{3.2}$$

It is clear from eqns (2.1), (2.4), and (2.9) that Φ may include lagged noise model terms and N is the data length.

The parameter estimates of eqn (3.1) will only be unbiased if $E[\hat{\Theta}] = \Theta$. It is convenient in the present analysis to study the bias associated with the least squares estimate using probability limit theory.

The probability limit of the estimated parameters is given by

$$\begin{aligned} Plim[\hat{\Theta}] &= Plim[[\Phi^T\Phi]^{-1}\ \Phi^T\vec{Y}] \\ &= Plim[[\Phi^T\Phi]^{-1}\ ]\ Plim[\Phi^T\vec{Y}] \end{aligned} \tag{3.3}$$

Assuming that both the input and output sequences are stationary it follows that for a sufficiently large data length N

$$\begin{aligned} Plim[\frac{1}{N}\Phi^T\Phi] &\approx \frac{1}{N}\Phi^T\Phi \\ Plim[\frac{1}{N}\Phi^T\vec{Y}] &\approx \frac{1}{N}\Phi^T\vec{Y} \end{aligned} \tag{3.4}$$

where

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

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

and

$$\Phi^T\vec{Y} = \begin{bmatrix} \sum_{t=1}^{N}\phi_n^T(t)p_{d1}(t)\frac{a(t)}{b(t)} \\ \sum_{t=1}^{N}\hat{\phi}_d^T(t)p_{d1}(t)\frac{a(t)}{b(t)} \end{bmatrix} + \begin{bmatrix} \sum_{t=1}^{N} 0 \sum_{t=1}^{N} \\ \sigma_e^2\sum_{t=1}^{N}p_d^T(t)p_{d1}(t) \end{bmatrix} \tag{3.5}$$

where

$$p_d(t) = [p_{d2}(t) \quad \cdots \quad p_{dden}(t)] \tag{3.6}$$

and $\hat{\phi}_d(t)$ is defined in eqn (2.12)

Rewritting eqn (3.6) gives

$$\Phi^T\Phi = [\Phi^T\Phi]_{(t-1)} + \sigma_e^2\,\Psi$$
$$\Phi^T\vec{Y} = [\Phi^T\vec{Y}]_{(t-1)} + \sigma_e^2\,\psi \tag{3.7}$$

where the definition of terms follows directly and

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

$$\psi = \begin{bmatrix} 0 \\ \sum_{t=1}^{N}p_d^T(t)p_{d1}(t) \end{bmatrix} = \sum_{t=1}^{N}\rho^T(t)p_{d1}(t) = \sum_{t=1}^{N}\psi(t) \tag{3.8}$$

$$\rho(t) = [0 \; p_d(t)] \tag{3.9}$$

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.

The estimate given in eqn (3.1) can therefore be written as

$$\hat{\Theta} = [\Phi^T\Phi]^{-1}\,\Phi^T\vec{Y}$$
$$= [[\Phi^T\Phi]_{(t-1)} + \sigma_e^2\,\Psi]^{-1}\,[[\Phi^T\vec{Y}]_{(t-1)} + \sigma_e^2\,\psi] \tag{3.10}$$

Even for the case of additive white noise therefore the two terms $\sigma_e^2\,\Psi$ and $\sigma_e^2\,\psi$ will cause bias. This is an inherent problem associated with the rational model expanded to be linear in the parameters because the denominator terms $y(t)p_{dj}(t)$ in eqn (2.4) implicitly contain $e(t)$ and this causes an σ_e^2 in the least squares estimates. This problem does not arise for the ARMAX and polynomial NARMAX models and hence unbiased estimates can be directly obtained.

Probability limit theory can also be used to formulate the asymptotic covariance matrix of the estimated parameters as follows

$$\begin{aligned} Cov\hat{\Theta} = \\ Plim\{[[\Phi^T\Phi]^{-1}\,\Phi^T\vec{Y} - \Theta]\;[[\Phi^T\Phi]^{-1}\,\Phi^T\vec{Y} - \Theta]^T\} \\ = Plim\{[[\Phi^T\Phi]^{-1}\,\Phi^T(\Phi\Theta + \vec{\zeta}) - \Theta]\;[[\Phi^T\Phi]^{-1}\,\Phi^T(\Phi\Theta + \vec{\zeta}) - \Theta]^T\} \\ = Plim\{[[\Phi^T\Phi]^{-1}\,\Phi^T\vec{\zeta}][[\Phi^T\Phi]^{-1}\,\Phi^T\vec{\zeta}]^T\} \end{aligned} \tag{3.11}$$

where with reference to eqn (2.6) and eqn (2.9)

$$\begin{aligned} \vec{\zeta} &= [\zeta(1) \;\cdots\; \zeta(N)]^T \\ &= [b(1)e(1) \;\cdots\; b(N)e(N)]^T \end{aligned} \tag{3.12}$$

Assuming that both the input and output sequences are stationary it follows that for a sufficiently large data length N

$$\begin{aligned} Cov\hat{\Theta} = \\ = Plim\{[[\Phi^T\Phi]^{-1}\,\Phi^T\vec{\zeta}][[\Phi^T\Phi]^{-1}\,\Phi^T\vec{\zeta}]^T\} \\ = Plim[\vec{\zeta}\vec{\zeta}^T]\; Plim[\Phi^T\Phi]^{-1} \\ = \sigma_e^2\sigma_b^2[\Phi^T\Phi]^{-1} \end{aligned} \tag{3.13}$$

where

$$\sigma_b^2 = \frac{1}{N}\sum_{t=1}^{N}[b(t)]^2 \tag{3.14}$$

which is the mean squared value of the denominator.

3.2 A least squares estimator for the rational model

Eqn (3.10) indicates that an unbiased parameter estimator for the rational model could be obtained if the error terms were removed. Define the new estimator

$$\begin{aligned} \hat{\Theta} &= [\Phi^T\Phi - \sigma_e^2\,\Psi]^{-1}\,[\Phi^T\vec{Y} - \sigma_e^2\,\psi] \\ &= [[\Phi^T\Phi]_{(t-1)}]^{-1}\,[\Phi^T\vec{Y}]_{(t-1)} \end{aligned} \tag{3.15}$$

as the rational model estimator (RME) providing that $\Phi^T\Phi$, Ψ, $\Phi^T\vec{Y}$, and ψ can all be calculated when the noise sequence $e(t)$ is available.

In practice the noise sequence $e(t)$ and the noise variance σ_e^2 will be unknown. However, a predicted noise sequence and an estimated noise variance σ_e^2 can be obtained using a simple extension of the traditional extended least squares algorithm. Billings and Zhu (1990) have implemented this idea to formulate the RME algorithm.

The covariance matrix of the estimates using RME can be derived with reference to eqn (2.9) and eqn (3.13) for the special case when $\sigma_e^2 < 1$ to yield

$$Cov\hat{\Theta} \approx \sigma_e^2\sigma_b^2[[\Phi^T\Phi]_{(t-1)}]^{-1} \tag{3.16}$$

this shows that the effect of removing the bias is to increase the covariance of the estimates because $[\Phi^T\Phi]_{(t-1)} = \Phi^T\Phi - \sigma_e^2\,\Psi$. A detailed derivation of the covariance matrix of the estimates is given in appendix A.

For the ARMAX and NARMAX models, eqn (3.16) becomes

$$Cov\hat{\Theta} = \sigma_e^2[\Phi^T\Phi]^{-1} \tag{3.17}$$

because $\sigma_b^2 = 1$ and $[[\Phi^T\Phi]_{(t-1)}]^{-1} = [\Phi^T\Phi]^{-1}$.

4 Online least squares estimator for the rational model

A unified recursive least squares (RLS) based algorithm (Soderstrom, Ljung and Gustavsson 1978) may be written as

$$\begin{aligned}
\hat{\Theta}(t) &= \hat{\Theta}(t-1) + K(t)\varepsilon(t) \\
K(t) &= P(t)\phi^T(t) \\
P(t) &= (P(t-1) - \frac{P(t-1)\phi^T(t)\phi(t)P(t-1)}{\lambda(t) + \phi(t)P(t-1)\phi^T(t)})/\lambda(t) \\
\lambda(t) &= \lambda_0\lambda(t-1) + (1 - \lambda_0) \\
\varepsilon(t) &= Y(t) - \phi(t)\hat{\Theta}(t-1)
\end{aligned} \tag{4.1}$$

where $\lambda(t)$ is a variable forgetting factor and $\phi(t)$ is defined as in eqn (2.10).

Eliminating $K(t)$ yields

$$\begin{aligned}
\hat{\Theta}(t) &= \hat{\Theta}(t-1) + P(t)\phi^T(t)[Y(t) - \phi(t)\hat{\Theta}(t-1)] \\
P(t) &= (P(t-1) - \frac{P(t-1)\phi^T(t)\phi(t)P(t-1)}{\lambda(t) + \phi(t)P(t-1)\phi^T(t)})/\lambda(t) \\
\lambda(t) &= \lambda_0\lambda(t-1) + (1 - \lambda_0)
\end{aligned} \tag{4.2}$$

such that the algorithm consists of three updating formulae, the parameter vector $\hat{\Theta}(t)$, covariance matrix $P(t)$, and forgetting factor $\lambda(t)$.

To obtain the unbiased estimates for the rational model, a modified version of the recursive algorithm which implements the ideas from RME will be necessary

From eqn (3.15)

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

$$= P(t)\,f(t) \tag{4.3}$$

Updating expressions for $P(t)$ and $f(t)$ can be derived as

$$[P(t)]^{-1} = \Phi^T\Phi - \sigma_e^2\,\Psi$$

$$= \sum_{i=1}^{t-1}[\phi^T(t)\phi(t) - \sigma_e^2\rho(t)^T\rho(t)] + \phi^T(t)\phi(t) - \sigma_e^2\rho^T(t)\rho(t)]$$

$$= [P(t-1)]^{-1} + \phi^T(t)\phi(t) - \sigma_e^2\rho^T(t)\rho(t)$$

and

$$f(t) = \Phi^T\vec{Y} - \sigma_e^2\,\psi$$

$$= \sum_{i=1}^{t-1}[\phi^T(t)Y(t) - \sigma_e^2\rho(t)^T p_{d1}(t)] + \phi^T(t)Y(t) - \sigma_e^2\rho^T(t)p_{d1}(t)$$

$$= f(t-1) + \phi^T(t)Y(t) - \sigma_e^2\rho^T(t)p_{d1}(t) \tag{4.4}$$

From eqn (4.3)

$$f(t) = [P(t)]^{-1}\,\hat{\Theta}(t)$$

and hence

$$f(t-1) = [P(t-1)]^{-1}\,\hat{\Theta}(t-1) \tag{4.5}$$

Eqn (4.3) can now be written in the recursive form

$$\hat{\Theta}(t) = [P(t)]\,f(t)$$

$$= [P(t)]\,\{f(t-1) + \phi^T(t)Y(t) - \sigma_e^2\rho^T(t)p_{d1}(t)\}$$

$$= [P(t)]\,\{[P(t-1)]^{-1}\,\hat{\Theta}(t-1) + \phi^T(t)Y(t) - \sigma_e^2\rho^T(t)p_{d1}(t)\}$$

$$= [P(t)]\,\{([P(t)]^{-1} - \phi^T(t)\phi(t) + \sigma_e^2\rho^T(t)\rho(t))\hat{\Theta}(t-1)$$

$$+ \phi^T(t)Y(t) - \sigma_e^2\rho^T(t)p_{d1}(t)\}$$

$$= \hat{\Theta}(t-1) + [P(t)]\,\{\phi^T(t)Y(t) - \sigma_e^2\rho^T(t)p_{d1}(t)$$

$$- [\phi^T(t)\phi(t) - \sigma_e^2\rho^T(t)\rho(t)]\hat{\Theta}(t-1)\} \tag{4.6}$$

which will be used to update the parameter estimates.

A recursion for the covariance matrix $P(t)$ can be derived from eqn (4.4)

$$[P(t)] = [[P(t-1)]^{-1} + \phi^T(t)\phi(t) - \sigma_e^2\rho^T(t)\rho(t)]^{-1} \tag{4.7}$$

By the well known matrix inversion lemma

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(t + CA^{-1}B)^{-1}CA^{-1} \tag{4.8}$$

let

$$A = [P(t-1)]^{-1} + \phi^T(t)\phi(t)$$

$$B = -\sigma_e^2\rho^T(t)$$

$$C = \rho(t) \tag{4.9}$$

then eqn (4.7) becomes

$$[P(t)] = A^{-1} + \frac{\sigma_e^2 A^{-1}\rho^T(t)\rho(t)A^{-1}}{(1 - \sigma_e^2\rho(t)A^{-1}\rho^T(t))^{-1}} \tag{4.10}$$

Repeating the matrix inversion lemma for A^{-1}

$$A^{-1} = [[P(t-1)]^{-1} + \phi^T(t)\phi(t)]^{-1}$$

$$= P(t-1) - \frac{P(t-1)\phi^T(t)\phi(t)P(t-1)}{1 + \phi(t)P(t-1)\phi^T(t)} \tag{4.11}$$

Finally the RRME (Recursive Rational Model Estimator) algorithm can be summarised as

$$\hat{\Theta}(t) = \hat{\Theta}(t-1) + P(t)\{[\phi^T(t)Y(t)]_{(t-1)} - [\phi^T(t)\phi(t)]_{(t-1)}\hat{\Theta}(t-1)\}$$

$$P(t) = A^{-1} + \frac{\sigma_e^2 A^{-1}\rho^T(t)\rho(t)A^{-1}}{(1 - \sigma_e^2\rho(t)A^{-1}\rho^T(t))^{-1}}$$

$$A^{-1} = (P(t-1) - \frac{P(t-1)\phi^T(t)\phi(t)P(t-1)}{\lambda(t) + \phi(t)P(t-1)\phi^T(t)})/\lambda(t)$$

$$\lambda(t) = \lambda_0\lambda(t-1) + (1 - \lambda_0)$$

$$[\phi^T(t)Y(t)]_{(t-1)} = \phi^T(t)Y(t) - \hat{\sigma}_e^2(t-1)\psi(t)$$

$$[\phi^T(t)\phi(t)]_{(t-1)} = \phi^T(t)\phi(t) - \hat{\sigma}_e^2(t-1)\Psi(t)$$

$$\hat{e}(t) = (Y(t) - \phi(t)\hat{\Theta}(t-1))/b(t)$$

$$\hat{\sigma}_e^2(t) = \frac{(t-1)}{t}\hat{\sigma}_e^2(t-1) + \frac{((Y(t) - \phi(t)\hat{\Theta}(t-1))/b(t))^2}{t} \tag{4.12}$$

Implementation of the algorithm consists of the following steps:

(i) Set initial values $\hat{\Theta}(0)$, $P(0)$, $\lambda(0)$, λ_0, $e(0)$, $\hat{\sigma}_e^2(0)$, and t=0.

(ii) Set t=1+t and form the vectors $\phi(t)$ and $\rho(t)$

(iii) Compute $[\phi^T(t)Y(t)]_{(t-1)}$, $[\phi^T(t)\phi(t)]_{(t-1)}$, and $P(t)$.

(iv) Estimate the parameter vector $\hat{\Theta}(t)$.

(v) Predict the noise $\hat{e}(t)$ and compute the noise variance $\hat{\sigma}_e^2(t)$.

(vi) if $t<N$ go back to step (ii), otherwise end with $t=N$.

The RRME algorithm depends on the following assumptions which arise from eqn (3.5) and eqn (3.8)

$$plim[\phi^T(t)Y(t)]_{(t-1)} \approx [\Phi^T\vec{Y} - \sigma_e^2\,\psi] = [\Phi^T\vec{Y}]_{(t-1)}$$
$$plim[\phi^T(t)\phi(t)]_{(t-1)} \approx [\Phi^T\Phi - \sigma_e^2\,\Psi] = [\Phi^T\Phi]_{(t-1)} \tag{4.13}$$

For the noise free case, the RRME algorithm reduces to the classical recursive least squares algorithm, this is easily proved by setting $\hat{\sigma}_e^2(t-1) = 0$ to give eqn (4.2).

For a recursive algorithm, initial values for the parameter estimates $\hat{\Theta}(t)$, covariance matrix $P(t)$, and forgetting factor $\lambda(t)$ must be set before starting the algorithm. The choice of the initial values has been studied extensively by Ljung and Soderstrom (1983), a large value of $P(0)$ makes the value of $\hat{\Theta}(0)$ only marginally important, the parameter estimates will change quickly in the transient phase (for small values of t). A small value of $P(0)$ will give only small corrections of $\hat{\Theta}(t)$ since the gain for the error correction will be small for all t, convergence will then be slow unless $\hat{\Theta}(0)$ is close to the convergent limit.

Assuming no a priori information about the initial values, we set $P(0)=10000*I$, where I is an identity matrix, $\hat{\Theta}(0)=0$, predicted noise sequence $\hat{e}(0)=E[e(t)]=0$, $\lambda(0)=1$, $\lambda_0=1$, and $\hat{\sigma}_e^2(0)=0$.

5 Simulation studies

Three simulated examples were selected to illustrate the application of the on line RREM algorithm for parameter estimation of nonlinear stochastic rational models.A zero mean uniform random sequence with amplitude range ±1 (variance $\sigma_u^2 = 0.33$) was used as input and a zero mean Gaussian noise sequence $e(t)$ with variance $\sigma_e^2 = 0.01$ was used in all three examples. 500 data pairs were used in each case.

Example S_1 consisted of the model

$$y(t) = \frac{0.8y(t-1) + u(t-1)}{1 + u^2(t-1) + y(t-1)y(t-2)} + \eta(t) \tag{5.1}$$

where

$$\eta(t) = \frac{0.5e(t-1) + e(t)}{1 + u^2(t-1) + y(t-1)y(t-2)} \tag{5.2}$$

Expanding S_1 to give a linear in the parameter model yields

$$Y(t) = 0.8y(t-1) + u(t-1) - y(t)u^2(t-1) - y(t)y(t-1)y(t-2) + 0.5e(t-1) + e(t) \tag{5.3}$$

where

$$Y(t) = y(t) \tag{5.4}$$

The input and output data sequences for this example are illustrated in Fig. 1.1. The evolution of the parameter estimates are illustrated in Fig. 1.2 (a) to (e). The lines parallel to the time axis in plots (a) to (e) indicate the corresponding true parameters. Inspection of Fig. 1.2 shows that the estimates associated with the terms from the process and noise model converge rapidly after about 250 data samples. The estimated noise variance $\hat{\sigma}_e^2$ illustrated in Fig. 1.2 (f) which is used to check the performance of the algorithm shows a steady convergent characteristic.

The estimates obtained using the RLS and RRME algorithms at t=500 are given in Table 1 and these results show the significant improvement of the RRME algorithm for rational model parameter estimation.

Example S_2 consists of the same process model as example S_1

$$y(t) = \frac{0.8y(t-1) + u(t-1)}{1 + u^2(t-1) + y(t-1)y(t-2)} + \eta(t) \tag{5.5}$$

but with a much more complex noise model

$$\eta(t) = 0.5y(t-1)e(t-1) + e(t) \tag{5.6}$$

The linear in the parameter expansion for model S_2 is

$$Y(t) = 0.8y(t-1) + u(t-1) - y(t)u^2(t-1) - y(t)y(t-1)y(t-2) + b(t)\eta(t) \tag{5.7}$$

where

$$Y(t) = y(t)$$

$$b(t)\eta(t) = [1 + u^2(t-1) + y(t-1)y(t-2)]\ [0.5y(t-1)e(t-1) + e(t)] \tag{5.8}$$

The input and output data sequences for this example are illustrated in Fig. 2.1. The evolution of the parameter estimates are illustrated in Fig. 2.2 (a) to (e). Again the parallel lines in each plot indicate the corresponding true parameter values. The estimated noise variance $\hat{\sigma}_e^2$ is shown in Fig. 2.2 (f).

During the iteration, the denominator polynomial $b(t)$ in eqn (5.8) was replaced by $\hat{b}(t)$, the estimate of $b(t)$, hence eqn (5.8) is expressed as

$$\hat{b}(t)\eta(t) = 0.5\hat{b}(t)y(t-1)e(t-1) + \hat{b}(t)e(t) \tag{5.9}$$

The estimates obtained using the RLS and RRME algorithms at t=500 are given in Table 2 and clearly show how RRME provides unbiased estimates.

Example S_3 was modelled as

$$y(t) = \frac{0.5y(t-1)u(t-2) + u(t-1)}{1 + u^2(t-1) + y^2(t-1)} + \eta(t) \tag{5.10}$$

where

$$\eta(t) = \frac{0.7e(t-1) + e(t)}{1 + u^2(t-1) + y^2(t-1)} \tag{5.11}$$

Multiplying the rational model out to be a linear in the parametes gives

$$Y(t) = 0.5y(t-1)u(t-2) + u(t-1) - y(t)u^2(t-1) - y(t)y^2(t-1) + 0.7e(t-1) + e(t) \tag{5.12}$$

where

$$Y(t) = y(t) \tag{5.13}$$

The input and output data sequences for this example are illustrated in Fig. 3.1. The evolution of the parameter estimates are illustrated in Fig. 3.2 (a) to (e), where the parallel line in each plot represents the corresponding true parameter value. The evolution of the estimated noise variance is shown in Fig. 3.2 (f). Once again the fast convergence of the estimated parameters and noise variance is demonstrated in this example.

6 Conclusions

A new recursive least squares type parameter estimation routine has been derived for nonlinear stochastic rational models. A major advantage of the new algorithm is that unbiased parameter estimates are obtained from a linear-in-the-parameter model formulation. Simulation results suggest that the algorithm has good convergence properties but a thorough theoretical analysis of convergence is left for a later publication.

Acknowledgements

The authors gratefully acknowledge that this work is supported by SERC under grant GR/F2417.7.

7 Appendix A

The covariance matrix of the estimates is defined as

$$Cov\ \hat{\Theta} = plim[(\ \hat{\Theta} - \Theta\ )\ (\ \hat{\Theta} - \Theta\ )^T\ ] \tag{A.1}$$

Consider

$$\begin{aligned} \hat{\Theta} - \Theta &= [[\Phi^T\Phi]_{(t-1)}]^{-1}\ [\Phi^T\vec{Y}]_{(t-1)} - \Theta \\ &= [[\Phi^T\Phi]_{(t-1)}]^{-1}\ [\Phi^T(\Phi\Theta + \vec{\zeta}) - \sigma_e^2\psi] - \Theta \\ &= [[\Phi^T\Phi]_{(t-1)}]^{-1}\ [([\Phi^T\Phi]_{(t-1)} + \sigma_e^2\Psi)\Theta + \Phi^T\vec{\zeta} - \sigma_e^2\psi] - \Theta \\ &= \Theta + [[\Phi^T\Phi]_{(t-1)}]^{-1}\ [\sigma_e^2\Psi\Theta + \Phi^T\vec{\zeta} - \sigma_e^2\psi] - \Theta \\ &= [[\Phi^T\Phi]_{(t-1)}]^{-1}\ [\sigma_e^2\Psi\Theta + \Phi^T\vec{\zeta} - \sigma_e^2\psi] \end{aligned} \tag{A.2}$$

Define

$$\Delta = \sigma_e^2\Psi\Theta + \Phi^T\vec{\zeta} - \sigma_e^2\psi = \Phi^T\vec{\zeta} + \sigma_e^2\delta \tag{A.3}$$

where

$$\delta = \Psi\Theta - \psi \tag{A.4}$$

Now the asymptotic covariance matrix can be written as

$$\begin{aligned} Cov\ \hat{\Theta} &= plim[(\ \hat{\Theta} - \Theta\ )\ (\ \hat{\Theta} - \Theta\ )^T\ ] \\ &= plim\{\ [[\Phi^T\Phi]_{(t-1)}]^{-1}\ \Delta\Delta^T\ [[\Phi^T\Phi]_{(t-1)}]^{-1}\ \} \\ &= \sigma_e^2\sigma_b^2[[\Phi^T\Phi]_{(t-1)}]^{-1} + [[\Phi^T\Phi]_{(t-1)}]^{-1}\ [\sigma_e^2\Psi\vec{\zeta}\vec{\zeta}^T \\ &\quad + 2\sigma_e^2\Phi^T\vec{\zeta}\delta^T + \sigma_e^4\delta\delta^T]\ [[\Phi^T\Phi]_{(t-1)}]^{-1} \end{aligned} \tag{A.5}$$

where

$$\Delta\Delta^T = \Phi^T\vec{\Phi}\vec{\zeta}\zeta^T + 2\sigma_e^2\Phi^T\vec{\zeta}\delta^T + \sigma_e^4\delta\delta^T$$

$$= [[\Phi^T\Phi]_{(t-1)} + \sigma_e^2\Psi]\vec{\zeta}\vec{\zeta}^T + 2\sigma_e^2\Phi^T\vec{\zeta}\delta^T + \sigma_e^4\delta\delta^T$$

$$= [\Phi^T\Phi]_{(t-1)}\vec{\zeta}\vec{\zeta}^T + \sigma_e^2\Psi\vec{\zeta}\vec{\zeta}^T + 2\sigma_e^2\Phi^T\vec{\zeta}\delta^T + \sigma_e^4\delta\delta^T \tag{A.6}$$

Inspection of eqn (A.5) shows that the three terms $\sigma_e^2\Psi\vec{\zeta}\vec{\zeta}^T$, $\sigma_e^2\Phi^T\vec{\zeta}\delta^T$, and $\sigma_e^4\delta\delta^T$ all include a multiplier σ_e^4. It is reasonable to assume therefore that the effect of these terms on $Cov\ \hat{\Theta}$ can be ignored when $\sigma_e^2<1$, such that eqn (A.5) reduces to the approximate expression

$$Cov\hat{\Theta} \approx \sigma_e^2\sigma_b^2[[\Phi^T\Phi]_{(t-1)}]^{-1}|_{\sigma_e^2<1} \tag{A.7}$$

References

Billings, S.A. and W.S.F. Voon, "Least squares parameter estimation algorithms for nonlinear systems," Int. J. Control, vol. 15, 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 (accepted for publication).

Chen, S. and S.A. Billings, "Recursive maximum likelihood identification of a nonlinear output affine model," Int. J. Control, vol. 48, 1988.

Chen, S. and S.A. Billings, "Recursive prediction error parameter estimator for nonlinear models," Int. J. Control, vol. 49, 1989a.

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

Ljung, L. and T. Soderstrom, Theory and practice of recursive identification, MIT Press, Cambridge, MA, 1983. 629.831(2)

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

Soderstrom, T., L. Ljung, and I. Gustavsson, "A theoretical analysis of recursive identification methods," Automatica, vol. 14, 1978.

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

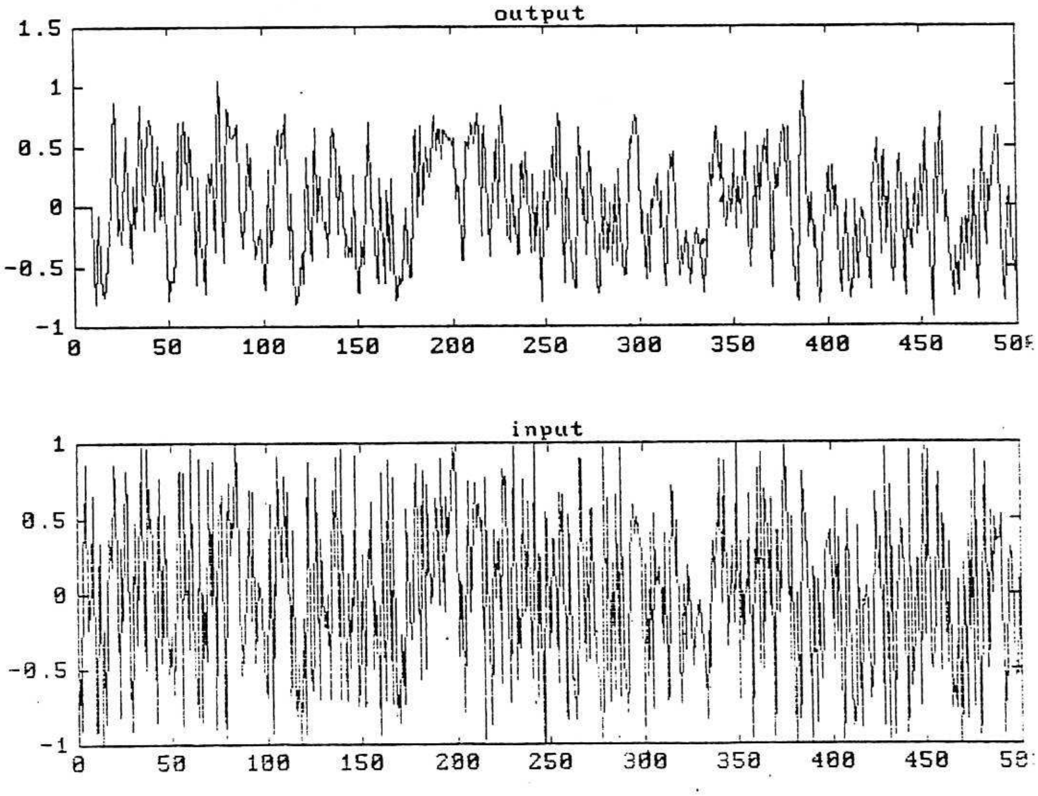
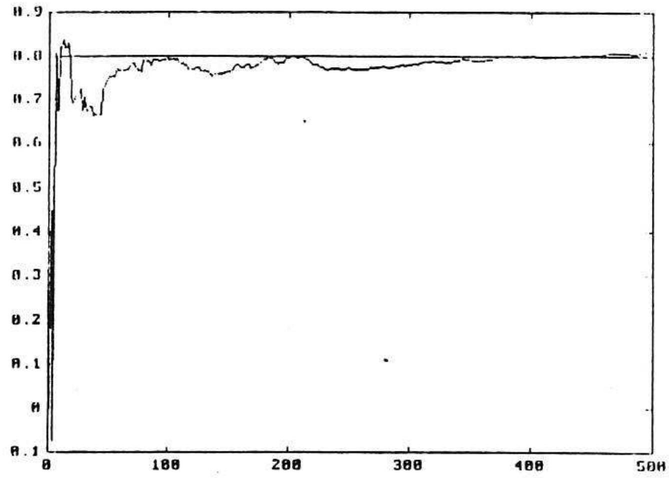
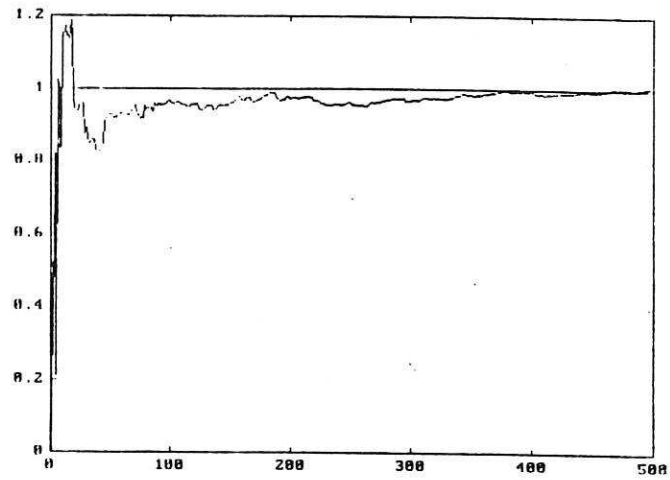


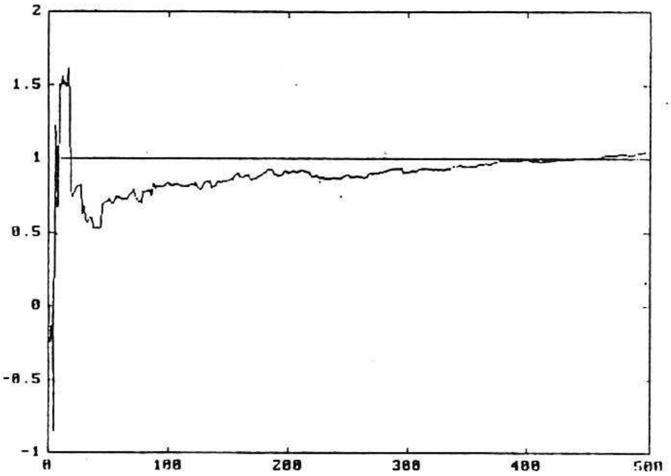
Figure 1.1 Input and output sequences for example S_1



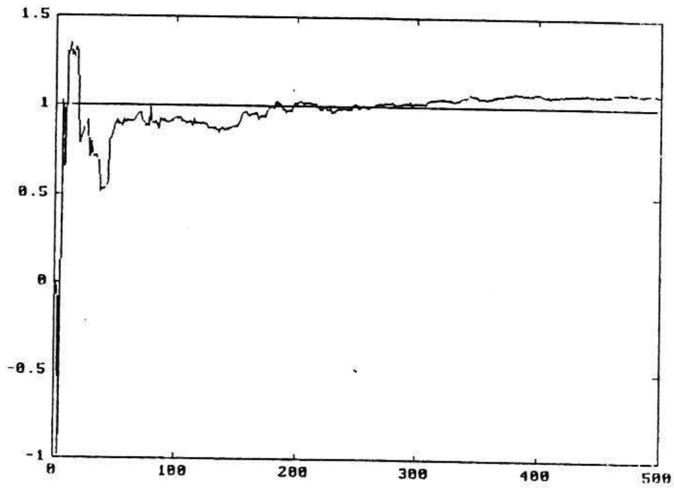
(a) Parameter associated with y(t-1)



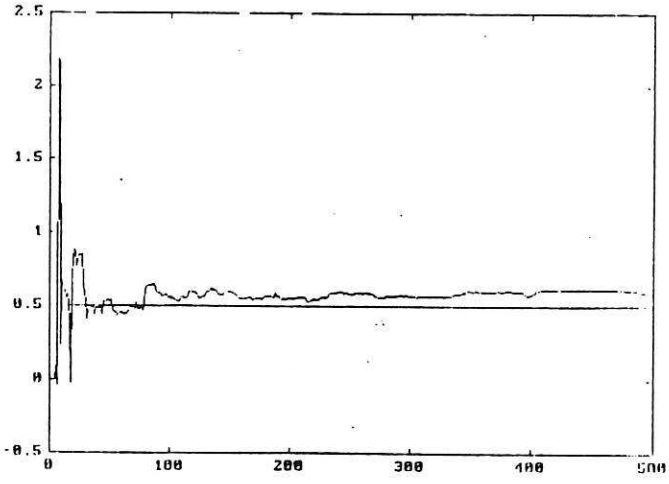
(b) Parameter associated with u(t-1)



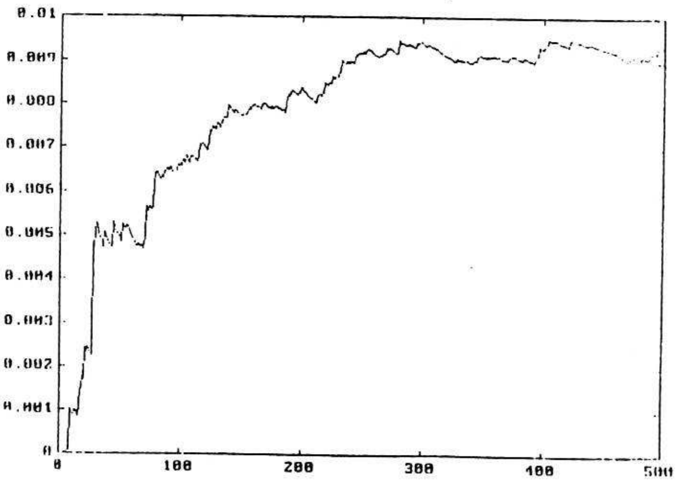
(c) Parameter associated with $y(t)U^2(t-1)$



(d) Parameter associated with y(t)y(t-1)y(t-2)



(e) Parameter associated with e(t-1)



(f) Estimated noise variance

Figure 1.2 Estimates (a)-(e) and estimated noise variance (f) for example S_1

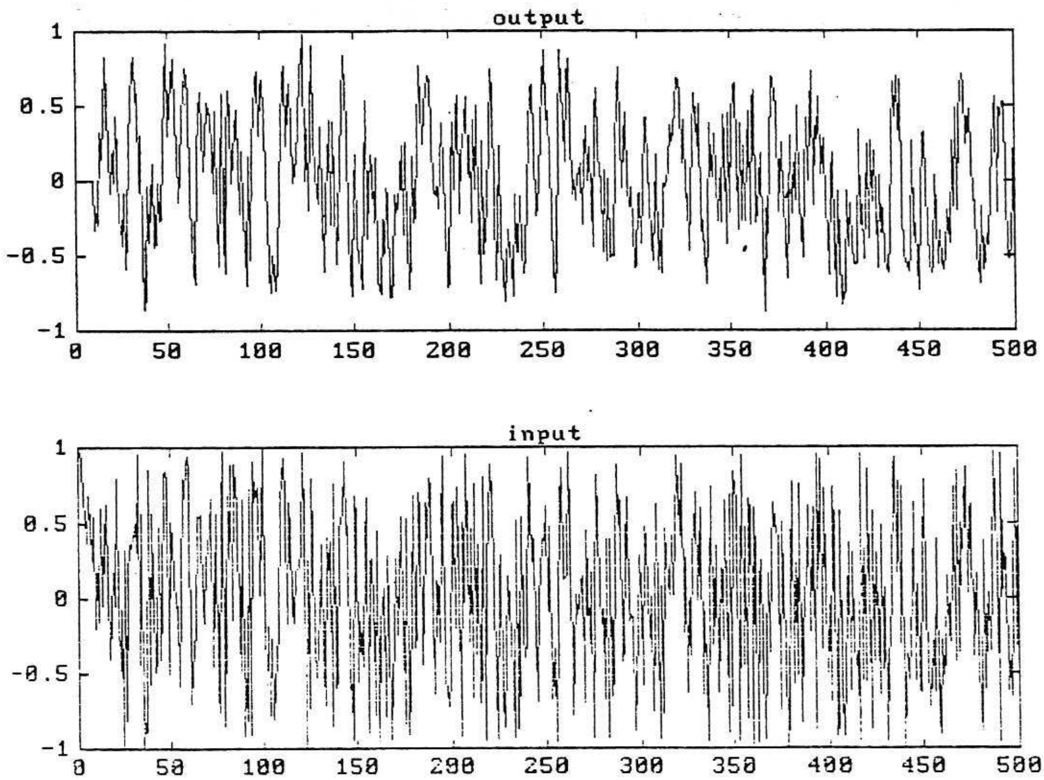
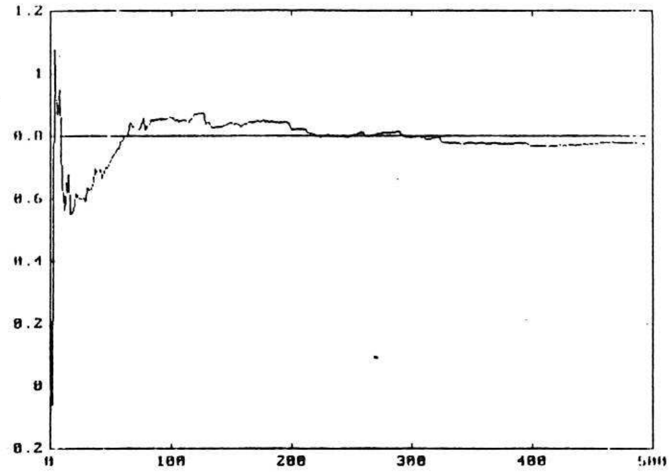
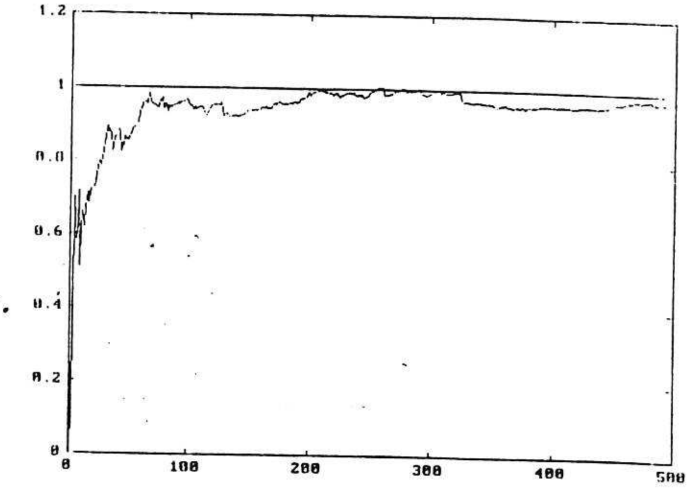


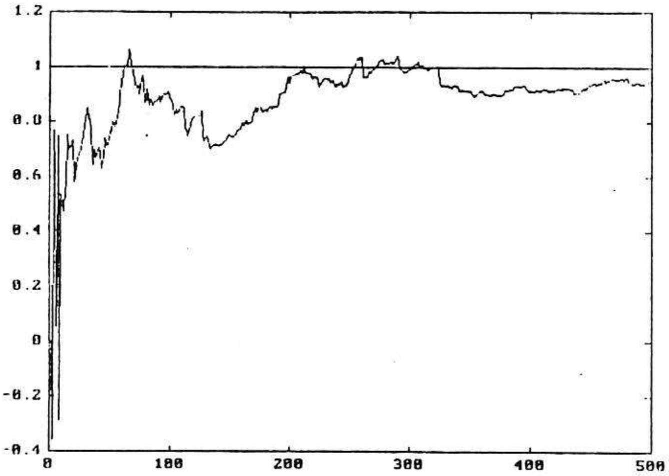
Figure 2.1 Input and output sequences for example S_2



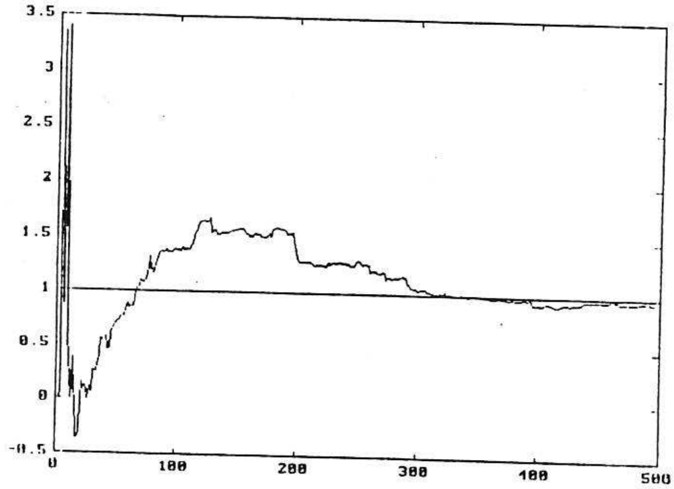
(a) Parameter associated with y(t-1)



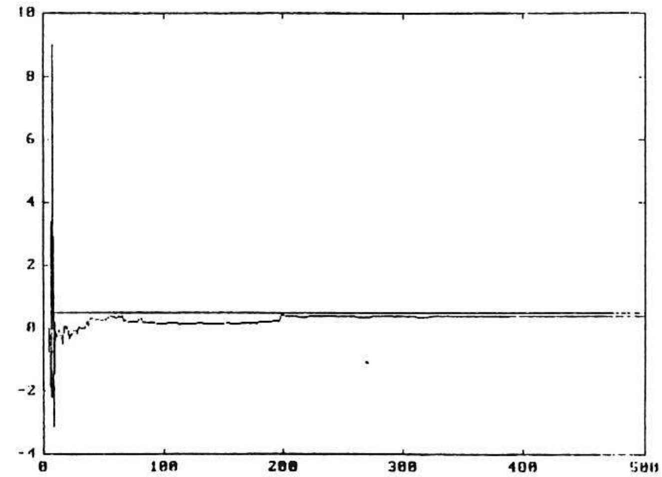
(b) Parameter associated with u(t-1)



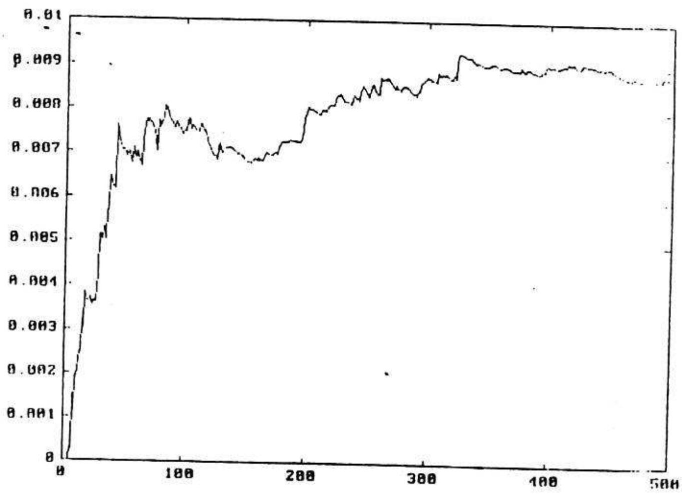
(c) Parameter associated with $y(t)U^2(t-1)$



(d) Parameter associated with y(t)y(t-1)y(t-2)



(e) Parameter associated with y(t-1)e(t-1)



(f) Estimated noise variance

Figure 2.2 Estimates (a)-(e) and estimated noise variance (f) for example S_2

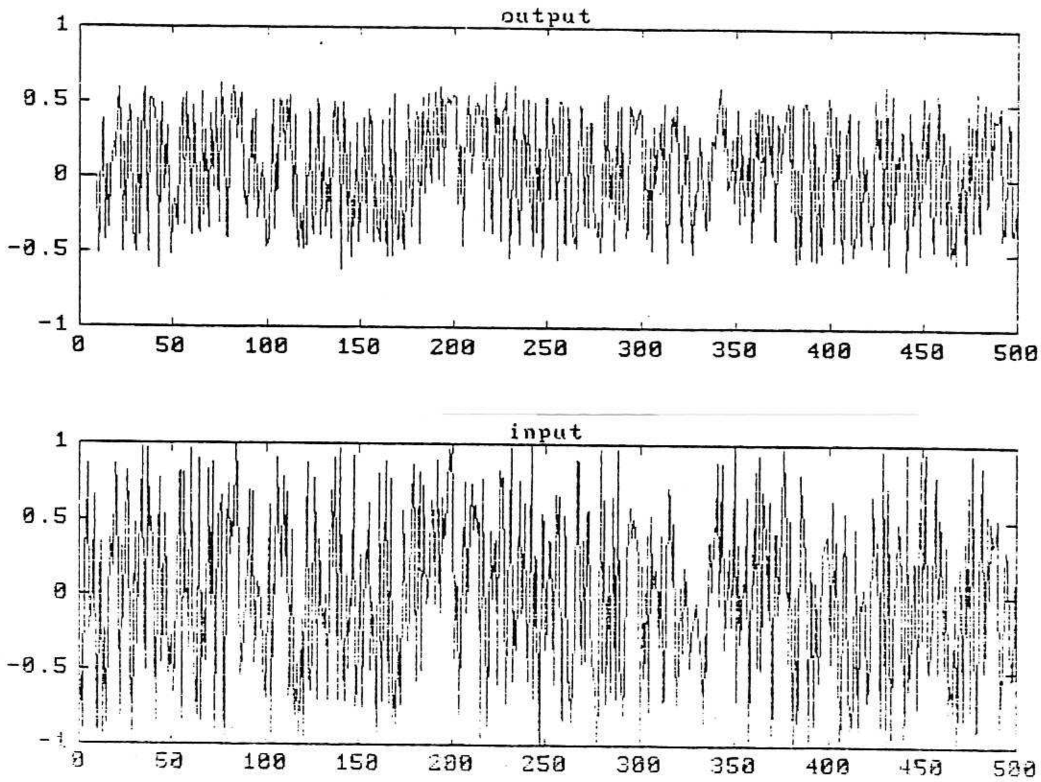
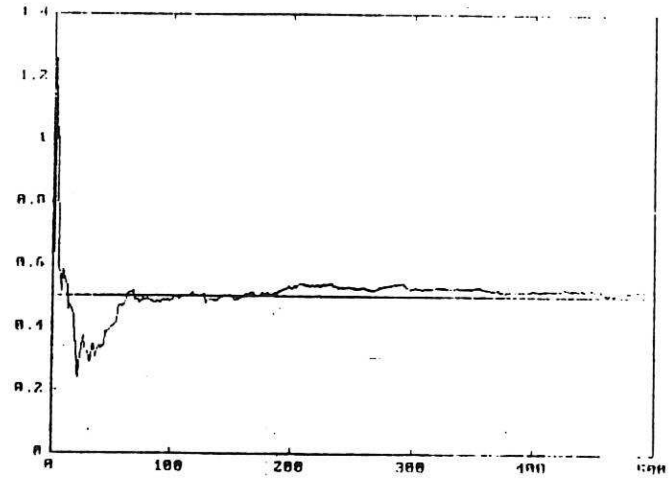
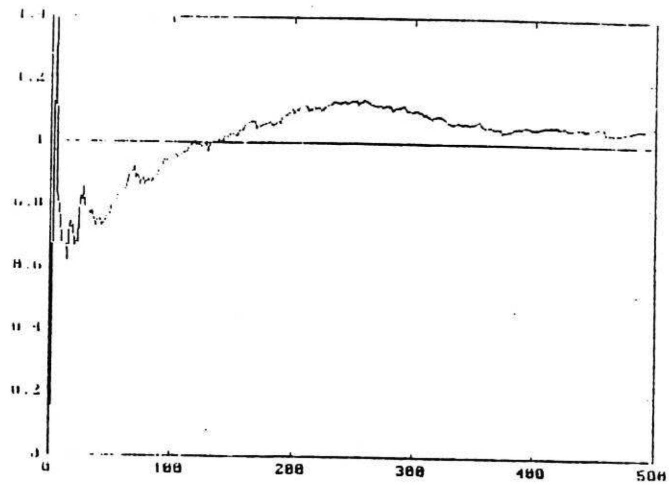


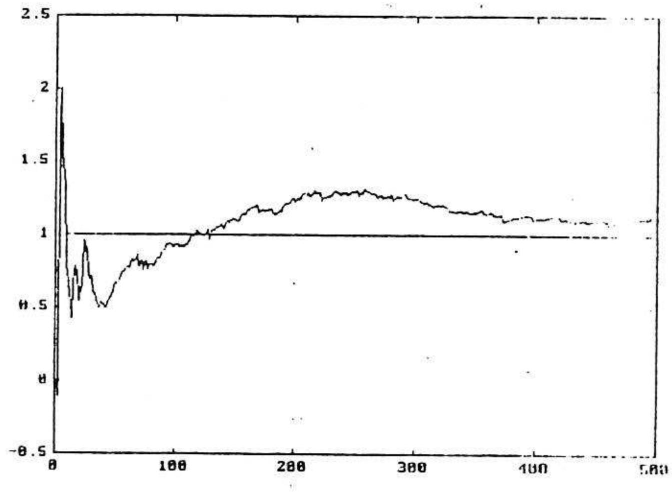
Figure 3.1 Input and output sequences for example S_3



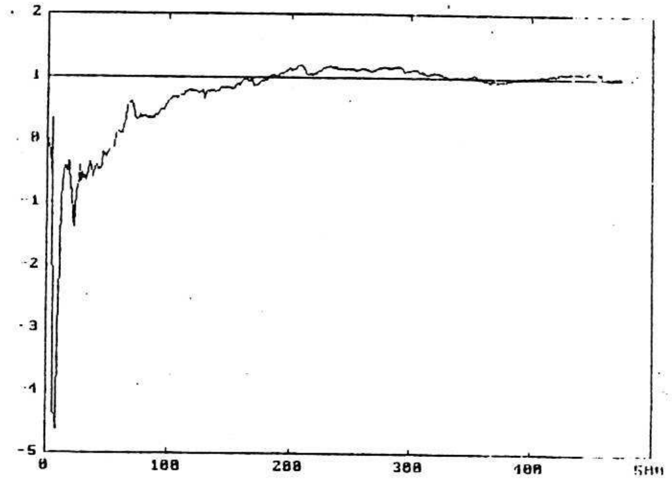
(a) Parameter associated with $y(t-1)U(t-2)$



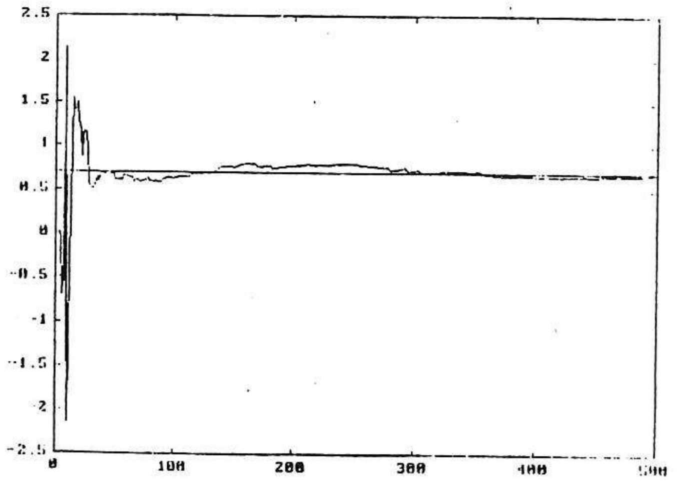
(b) Parameter associated with $u(t-1)$



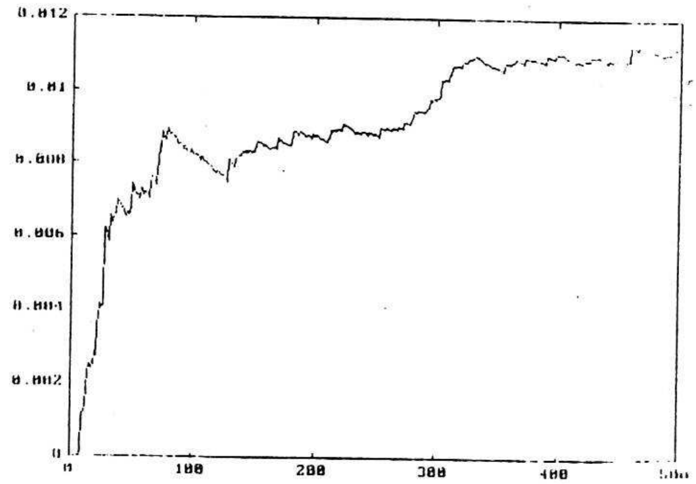
(c) Parameter associated with $U^2(t-1)y(t)$



(d) Parameter associated with $y^2(t-1)y(t)$



(e) Parameter associated with $e(t-1)$



(f) Estimated noise variance

Figure 3.2 Estimates (a)-(e) and estimated noise variance (f) for example S_3

Terms	Parameter Estimates		
	$\hat{\Theta}_{RRME}(500)$	$\hat{\Theta}_{RLS}(500)$	Θ_{true}
$y(t-1)$	0.8094	0.7045	0.8
$u(t-1)$	1.0025	0.8420	1.0
$y(t)u^2(t-1)$	-1.0373	-0.6262	-1.0
$y(t)y(t-1)y(t-2)$	-1.0709	-0.6990	-1.0
$e(t-1)$	0.5886	0.4858	0.5

Table 1 Parameters estimated for example S_1

Terms	Parameter Estimates		
	$\hat{\Theta}_{RRME}(500)$	$\hat{\Theta}_{RLS}(500)$	Θ_{true}
$y(t-1)$	0.7792	0.6502	0.8
$u(t-1)$	0.9788	0.7674	1.0
$y(t)u^2(t-1)$	-0.9489	-0.4037	-1.0
$y(t)y(t-1)y(t-2)$	-0.9711	-0.4916	-1.0
$y(t-1)e(t-1)$	0.3905	0.3082	0.5

Table 2 Parameters estimated for example S_2

Terms	Parameter Estimates		
	$\hat{\Theta}_{RRME}(500)$	$\hat{\Theta}_{RLS}(500)$	Θ_{true}
$y(t-1)u(t-2)$	0.5272	0.3694	0.5
$u(t-1)$	1.0604	0.7235	1.0
$y(t)u^2(t-1)$	-1.1374	-0.4281	-1.0
$y(t)y^2(t-1)$	-1.0953	0.2552	-1.0
$e(t-1)$	0.6934	0.5405	0.7

Table 3 Parameters estimated for example S_3