Monograph:

Billings, S.A. and Zhu, Q.M. (1990) Rational Model Identification Using an Extended Least Squares Algorithm. Research Report. Acse Report 388 . Dept of Automatic Control and System Engineering. University of Sheffield

Rational Model Identification Using an Extended Least Squares Algorithm

S.A. Billings and Q.M. Zhu

Department of Control Engineering

University of Sheffield

Mappin Street

Sheffield S1 3JD

U. K.

Research Report No. 388

November 1990

October 16, 1990

Rational Model Identification
Using an Extended Least Squares Algorithm

S.A. Billings, Q.M. Zhu

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

Abstract:

A new least squares based parameter estimation algorithm is derived for nonlinear systems which can be represented by a rational model defined as the ratio of two polynomial expansions of past system inputs, outputs and noise. Simulation results are included to illustrate the performance of the new algorithm.

1 Introduction

Polynomial expansions are prevelant throughout nonlinear systems analysis. The classical Duffings and Van der Pol equations for example are differential equations that include polynomial terms and which are known to represent an enormous range of behaviours (Nayfeh and Mook 1979). Recent developments in the field include the introduction of the polynomial NARMAX model (Nonlinear AutoRegressive Moving Average model with eXogenous inputs) and the formulation of identification and controller design algorithms based on this description (Leontaritis and Billings 1985, Chen and Billings 1989, Haber and Unbehauen 1990, Sales and Billings 1990). Although the Weierstrass approximation theorem ensures that polynomial models will always play an important role in nonlinear analysis there are alternative model expansions which may provide a more concise description for some classes of nonlinear systems.

The rational model can be considered as the next natural progression after the linear and polynomial nonlinear models. The advantage of the rational model is the efficiency with which even severe nonlinearities can be described with just a few parameters. The NARMAX model which was derived for a general class of nonlinear systems is not restricted to polynomial systems and can be expanded as a rational model (Chen and Billings 1989). These results can be related to the models introduced by Sontag (1979) which when extended to the stochastic case provide a class of rational models which can be used as the basis for the development of parameter estimation algorithms.

The disadvantage of the rational model is that it is not naturally linear-in-the-parameters and consequently identification based on this description appears to be complex. Nonlinear least squares algorithms (Marquardt 1963) can be applied if the data is perfectly noise free but this is unrealistic in practice. Alternatively the prediction error algorithm developed by Billings and Chen (1989) can be used to determine the model structure and estimate the unknown parameters but this involves an iterative Gauss-Newton algorithm which can become computationally intense for large models.

The present study attempts to circumvent most of these problems by introducing a new rational model estimation (RME) algorithm to estimate the parameters in rational NARMAX models of known structure. It is shown that when the output is corrupted by measurement noise multiplying out the rational model to make it linear-in-the-parameters leads to biased estimates. Unlike polynomial models the bias remains even when the noise is white and it is this problem which has inhibited the development of least squares based estimators for rational systems. The new estimator is developed by carefully studying how the bias is induced and reformulating the well known extended least squares algorithm (Ljung and Soderstrom 1983) to provide an iterative solution which is unbiased even when the output is corrupted by unknown coloured noise. The new algorithm, which is called the rational model estimator (RME) maintains the advantages of linear least squares, it is easy to code and computationally cheap compared with the alternatives. Simulation studies are included to demonstrate the effectiveness of the new method.

2 The rational model

Classical approximation theory (Sontag 1979) shows that rational functions are very efficient descriptors of nonlinear effects. This suggests that an extension of these concepts to the nonlinear dynamic case may provide useful alternative models for nonlinear systems compared with polynomial system expansions.

2.1 Input-output description

Previous studies have shown that the nonlinear dynamic rational model can be derived as a particular expansion of the NARMAX model or via the results of Sontag (Chen and Billings 1988) to yield the stochastic rational model

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

$$(2.1)$$

where $u(k)$ and $y(k)$ represent the input and output at time k ($k = 1, 2, \cdots$) respectively, r is the order of the model, and $e(k)$ is an unobservable independent and identically distributed (iid) 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

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

and for the denominator

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

where $p_{nj}(k)$, $p_{dj}(k)$ are terms consisting of $y(k-1), \cdots, y(k-r), u(k-1), \cdots, u(k-r), e(k-1), \cdots, e(k-r)$ 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 $b(k)$ on both sides of eqn (2.1) and then moving all the terms except $y(k)p_{d1}(k)\theta_{d1}$ to the right hand side gives

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

where

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

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

$$\zeta(k) = b(k)e(k)$$

$$= (\sum_{j=1}^{den} p_{dj}(k)\theta_{dj})e(k)$$

$$= p_{d1}(k)e(k) + (\sum_{j=2}^{den} p_{dj}(k)\theta_{dj})e(k) \tag{2.6}$$

where

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

Providing $e(k)$ has been reduced to an uncorrelated sequence as defined in eqn (2.1). The disadvantage of this approach is now apparant all the denominator terms $y(k)p_{dj}(k)$ implicitly include a current noise term $e(k)$ which is highly correlated with $\zeta(k)$. This will introduce bias in the parameter estimation even when $e(k)$ is a zero mean white noise sequence. This problem is a direct consequence of the rational model description because inspection of eqn (2.4) shows that if a polynomial NARMAX model were used then $b(k) = 1$ and there would be no terms on the right hand side of eqn (2.4) involving $y(k)$.

The linear in the parameters expression of the rational model eqn (2.4) can alternatively be expressed as

$$Y(k) = \sum_{j=1}^{num} p_{nj}(k)\theta_{nj} - \sum_{j=2}^{den} y(k)p_{dj}(k)\theta_{dj} + b(k)e(k)$$

$$= \sum_{j=1}^{num} p_{nj}(k)\theta_{nj} - \sum_{j=2}^{den} \frac{a(k)}{b(k)} p_{dj}(k)\theta_{dj} + p_{d1}(k)e(k) \tag{2.8}$$

Although the term $\frac{a(k)}{b(k)}p_{dj}(k)$ in eqn (2.8) cannot be directly obtained the expression is very useful in the analysis of bias and the derivation of the new estimator.

Eqn (2.4) may be written in vector notation as

$$Y(k) = P(k)\Theta + \zeta(k)$$

$$= \hat{P}(k)\Theta + p_{d1}(k)e(k) \tag{2.9}$$

where

$$P(k) = [P_n(k) \;\; P_d(k)]$$

$$= [p_{n1}(k) \cdots p_{nnum}(k) \; -p_{d2}(k)y(k) \cdots -p_{dden}(k)y(k)]$$

$$= [p_{n1}(k) \cdots p_{nnum}(k) \; -p_{d2}(k)(\frac{a(k)}{b(k)} + e(k)) \cdots -p_{dden}(k)(\frac{a(k)}{b(k)} + e(k))] \tag{2.10}$$

$$\Theta^T = [\Theta_n \; \Theta_d]$$

$$= [\theta_{n1} \cdots \theta_{nnum}\ \theta_{d2} \cdots \theta_{dden}] \tag{2.11}$$

and

$$\hat{P}(k) = [P_n(k)\ \ \hat{P}_d(k)]$$

$$= [p_{n1}(k) \cdots p_{nnum}(k)\ -p_{d2}(k)\frac{a(k)}{b(k)} \cdots -p_{dden}(k)\frac{a(k)}{b(k)}] \tag{2.12}$$

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

3 Parameter estimation

The main reason for expanding the rational model to be linear in the parameters was to try and develop a least squares based parameter estimation routine. Whilst initially the noise problems induced by such an expression appear formidable a detailed analysis of the bias suggests a solution to this problem.

It will be assumed in the present study that the structure of the model is known a priori. This is unrealistic in general and will be solved in a later publication. For the current implementation therefore terms must be manually added or deleted from the model until a suitable model structure is determined in a manner analogous to the procedures used in linear parameter estimation.

3.1 Least squares estimation

Applying directly the well known least squares estimator yields

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

where

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

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

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

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

3.2 Bias analysis

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.

Probability limits refer to one particular way in which estimates may settle down as the number N of observations on which they are based is increased (Norton, 1986).

Consider a sequence of random variables $\xi(N)$ say. If $\xi(N)$ converges in probability to x, x is said to be the probability limit $Plim(\xi)$ of $\xi(N)$. Two useful properties of the probability limit (Wilks, 1962) are that for any continuous function $f(\xi)$

$$Plim\, f(\xi) = f(\, Plim\ \xi\ ) \tag{3.3}$$

and for two matrices A and B, both functions of the same random variables

$$Plim\ (A\ B) = Plim\ A\ \ Plim\ B \tag{3.4}$$

Returning to the analysis of the least squares estimate in eqn (3.1) and assuming that the probability limits

$$\begin{aligned} A &\equiv Plim\ [\ \frac{1}{N}\Phi^T\Phi\ ] \\ B &\equiv Plim\ [\ \frac{1}{N}\Phi^T\vec{Y}\ ] \end{aligned} \tag{3.5}$$

exist, taking the probability limit of eqn (3.1) yields

$$\begin{aligned} Plim[\hat{\Theta}] &= Plim[[\Phi^T\Phi]^{-1}\ \Phi^T\vec{Y}] \\ &= Plim[[\Phi^T\Phi]^{-1}\ ]\ Plim[\Phi^T\vec{Y}] \end{aligned} \tag{3.6}$$

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

$$\begin{aligned} Plim[\frac{1}{N}\Phi^T\Phi] &\approx \frac{1}{N}\Phi^T\Phi \\ Plim[\frac{1}{N}\Phi^T\vec{Y}] &\approx \frac{1}{N}\Phi^T\vec{Y} \end{aligned} \tag{3.7}$$

where

$$\Phi^T\Phi =$$

$$\begin{bmatrix}
\sum_{k=1}^{N} p_{n1}^2(k) & . & \sum_{k=1}^{N} p_{n1}(k)p_{nnum}(k) & . & -\sum_{k=1}^{N} p_{n1}(k)p_{d2}(k)\frac{a(k)}{b(k)} & . & -\sum_{k=1}^{N} p_{n1}(k)p_{dden}(k)\frac{a(k)}{b(k)} \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
\sum_{k=1}^{N} p_{nnum}(k)p_{n1}(k) & . & \sum_{k=1}^{N} p_{nnum}^2(k) & . & -\sum_{k=1}^{N} p_{nnum}(k)p_{d2}(k)\frac{a(k)}{b(k)} & . & -\sum_{k=1}^{N} p_{nnum}(k)p_{dden}(k)\frac{a(k)}{b(k)} \\
-\sum_{k=1}^{N} p_{d2}(k)p_{n1}(k)\frac{a(k)}{b(k)} & . & \sum_{k=1}^{N} p_{d2}(k)p_{nnum}(k)\frac{a(k)}{b(k)} & . & \sum_{k=1}^{N} p_{d2}^2(k)(\frac{a(k)}{b(k)})^2 & . & \sum_{k=1}^{N} p_{d2}(k)p_{dden}(k)(\frac{a(k)}{b(k)})^2 \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
-\sum_{k=1}^{N} p_{dden}(k)p_{n1}(k)\frac{a(k)}{b(k)} & . & \sum_{k=1}^{N} p_{dden}(k)p_{nnum}(k)\frac{a(k)}{b(k)} & . & \sum_{k=1}^{N} p_{dden}(k)p_{d2}(k)(\frac{a(k)}{b(k)})^2 & . & \sum_{k=1}^{N} p_{dden}^2(k)(\frac{a(k)}{b(k)})^2
\end{bmatrix}$$

$$+ \begin{bmatrix}
0 & . & 0 & . & 0 & . & 0 \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
0 & . & 0 & . & 0 & . & 0 \\
0 & . & 0 & . & \sigma_e^2 \sum_{k=1}^{N} p_{d2}^2(k) & . & \sigma_e^2 \sum_{k=1}^{N} p_{d2}(k)p_{dden}(k) \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
0 & . & 0 & . & \sigma_e^2 \sum_{k=1}^{N} p_{dden}(k)p_{d2}(k) & . & \sigma_e^2 \sum_{k=1}^{N} p_{dden}^2(k)
\end{bmatrix}$$

and

$$\Phi^T \vec{Y} =$$

$$\begin{bmatrix} \sum_{k=1}^{N} p_{n1}(k) p_{d1}(k) (\frac{a(k)}{b(k)} + e(k)) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{k=1}^{N} p_{nnum}(k) p_{d1}(k) (\frac{a(k)}{b(k)} + e(k)) \\ \sum_{k=1}^{N} p_{d2}(k) p_{d1}(k) (\frac{a(k)}{b(k)} + e(k))^2 \\ \cdot \\ \cdot \\ \cdot \\ \sum_{k=1}^{N} p_{dden}(k) p_{d1}(k) (\frac{a(k)}{b(k)} + e(k))^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{N} p_{n1}(k) p_{d1}(k) \frac{a(k)}{b(k)} \\ \cdot \\ \cdot \\ \cdot \\ \sum_{k=1}^{N} p_{nnum}(k) p_{d1}(k) \frac{a(k)}{b(k)} \\ -\sum_{k=1}^{N} p_{d2}(k) p_{d1}(k) (\frac{a(k)}{b(k)})^2 \\ \cdot \\ \cdot \\ \cdot \\ -\sum_{k=1}^{N} p_{dden}(k) p_{d1}(k) (\frac{a(k)}{b(k)})^2 \end{bmatrix} + \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ -\sigma_e^2 \sum_{k=1}^{N} p_{d2}(k) p_{d1}(k) \\ \cdot \\ \cdot \\ \cdot \\ -\sigma_e^2 \sum_{k=1}^{N} p_{dden}(k) p_{d1}(k) \end{bmatrix} \quad (3.8)$$

Rewritting eqn (3.8) gives

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

where the definition of terms follows directly and

$$\Psi = \begin{bmatrix} 0 & . & 0 & . & 0 & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & . & 0 & . & 0 & . & 0 \\ 0 & . & 0 & . & \sum_{k=1}^{N} p_{d2}^2(k) & . & \sum_{k=1}^{N} p_{d2}(k)p_{dden}(k) \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & . & 0 & . & \sum_{k=1}^{N} p_{dden}(k)p_{d2}(k) & . & \sum_{k=1}^{N} p_{dden}^2(k) \end{bmatrix}$$

$$\psi = \begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \\ -\sum_{k=1}^{N} p_{d2}(k)p_{d1}(k) \\ . \\ . \\ . \\ -\sum_{k=1}^{N} p_{dden}(k)p_{d1}(k) \end{bmatrix} \qquad (3.10)$$

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

Hence the estimate given in eqn (3.1) can be written as

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

$$= [[\Phi^T\Phi]_{(k-1)} + \sigma_e^2\, \Psi]^{-1}\, [[\Phi^T\vec{Y}]_{(k-1)} + \sigma_e^2\, \psi] \tag{3.11}$$

It is well known that the estimate of eqn (3.1) known as the extended least squares algorithm, for linear difference equation or polynomial NARMAX models yields asymptotically unbiased and consistent estimates providing a suitable noise model is estimated such that $e(k)$ is reduced to a white noise sequence. This is apparant from eqn (3.11) because $\sigma_e^2\, \Psi$ and $\sigma_e^2\, \psi$ are zero for both these models.

Inspection of eqn (3.11) however shows that the linear in the parameters expansion of the rational model induces two additional terms $\sigma_e^2\, \Psi$ and $\sigma_e^2\, \psi$. These terms exist even if the sequence $e(k)$ is white and will cause severe bias in the estimates. This problem arises because the denominator terms $y(k)p_{dj}(k)$ in eqn (2.4) implicitly contain $e(k)$. It is interesting to note that both the $\sigma_e^2\, \Psi$ and $\sigma_e^2\, \psi$ terms are affected and not just the latter as would be the case with coloured noise and a linear model say.

4 A new least squares estimator for the rational model

Extended least squares algorithms for linear difference equation models have been extensively studied for many years (Goodwin and Payne 1977, Norton 1986, and Ljung 1987). A common point in all the algorithms is that the correlated noise is reduced to a white noise sequence by incorporating a noise model. This results in an iterative procedure by which the process model and noise model parameters are alternatively estimated until the bias is reduced to zero. Ideally we would like to maintain the simplicity of this approach for the rational model. Inspection of eqn (3.11) reveals the terms which cause the bias and therefore indicates which terms we need to eliminate. Define the new estimator

$$\hat{\Theta} = [\Phi^T\Phi - \sigma_e^2\, \Psi]^{-1}\, [\Phi^T\vec{Y} - \sigma_e^2\, \psi]$$
$$= [[\Phi^T\Phi]_{(k-1)}]^{-1}\, [\Phi^T\vec{Y}]_{(k-1)} \tag{4.1}$$

It follows from eqn (3.11) and the analysis in section 3 that this new estimator called the rational model estimator (RME) will be unbiased assuming that $\Phi^T\Phi$, Ψ, $\Phi^T\vec{Y}$, and ψ can all be calculated when the noise sequence $e(k)$ is available.

In practice the noise sequence $e(k)$ and the noise variance σ_e^2 are 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. This suggests the following iterative algorithm.

(i) Use an Ordinary Least Squares (OLS) algorithm to compute $\hat{\Theta}$ as

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

This estimate provides initial parameter values for the subsequent computation.

(ii) Compute the noise sequence

$$e(k) = y(k) - \frac{a(\cdots, \hat{\Theta}(i-1))}{b(\cdots, \hat{\Theta}(i-1))}$$

and estimate the noise variance $\hat{\sigma}_e^2$ as

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

where i is the iteration index, N is the data length and md is the maximum lag in the terms.

(iii) Update the matrices $\Phi^T\Phi$ and Ψ, vectors $\Phi^T\vec{Y}$ and ψ using the noise sequence $e(k)$ from step (ii).

(iv) Compute the parameter estimate as

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

(v) Go back to step (ii) and repeat until the parameter estimates and $\hat{\sigma}_e^2$ converge to constant values.

Clearly the convergence of the RME algorithm depends on the convergence of the estimated noise sequence. It should be possible to study these ideas by extending the methods developed for the classical extended least squares algorithm (Ljung and Soderstrom 1983). This will be addressed in later publications. All the examples computed to date have converged in typically 5 to 10 iterations and whilst these results cannot be used to imply convergence in general they are very encouraging.

Another common stochastic rational model expression may be written as

$$y(k) = \frac{a(k)}{b(k)} + \frac{c(z^{-1})e(k)}{b(k)} \tag{4.3}$$

where $c(z^{-1})e(k)$ is the classical moving average noise model. This model is a subset of eqn (2.1) and it is easy to show that the REM algorithm for this model is given by replacing σ_e^2 by σ_e^2 / σ_b^2 in eqn (4.1), where $\sigma_b^2 = E[b^2(k)]$.

5 Simulation studies

Four simulated examples were selected to illustrate the application of the REM algorithm for parameter estimation in nonlinear rational models. In all the examples 500 pairs of input and output data were used.

Example S_1 consists of the model

$$y(k) = \frac{0.2y(k-1) + 0.1y(k-1)u(k-1) + u(k-1)}{1 + y^2(k-1) + y^2(k-2)} + \eta(k) \tag{5.1}$$

where

$$\eta(k) = \frac{0.8e(k-1) + e(k)}{1 + y^2(k-1) + y^2(k-2)} \tag{5.2}$$

The input $u(k)$ was a zero mean uniform random sequence with amplitude range ± 1 (variance $\sigma_u^2 = 0.33$), which was used in all of the four examples. The noise $e(k)$ was a zero mean Gaussian sequence with variance $\sigma_e^2 = 0.01$.

The linear in the parameter model for system S_1 is

$$Y(k) = 0.2y(k-1) + 0.1y(k-1)u(k-1) + u(k-1) - y(k)y^2(k-1) - y(k)y^2(k-2) + 0.8e(k-1) + e(k) \tag{5.3}$$

where

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

The input and output data sequences for this example are illustrated in Fig. 1.1. The one step ahead predictions and residuals computed using ELS and RME are illustrated in Fig. 1.2. The improvements using RME compared with ELS can be seen from the residual plots. The model validity plots for ELS and RME are illustrated in Fig. 1.3. Notice that $\Phi_{\varepsilon\varepsilon}(k)$ for ELS is well outside the confidence band for $k = 1, 2$ indicating that the estimated model is severely biased. All the tests for RME are within the 95% confidence bands showing that the bias has been removed. The parameter estimates obtained using RME are listed in table 1. The estimates at iteration 1 represent the classical ELS estimates. A comparison with the true parameter values shows that $\sigma_e^2\,\Psi$ and $\sigma_e^2\,\psi$ cause severe bias. The estimates converge quickly towards the true values and the estimated variance $\hat{\sigma}_e^2$ rapidly approaches the true variance $\sigma_e^2 = 0.01$.

Example S_2 consists of the same process model as example S_1

$$y(k) = \frac{0.2y(k-1) + 0.1y(k-1)u(k-1) + u(k-1)}{1 + y^2(k-1) + y^2(k-2)} + \eta(k) \tag{5.5}$$

but with a much more complex noise model

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

A zero mean Gaussian sequence with variance $\sigma_e^2 = 0.02$ was used to generate the noise $e(k)$.

The linear in the parameter model for S_2 is

$$Y(k) = 0.2y(k-1) + 0.1y(k-1)u(k-1) + u(k-1) - y(k)y^2(k-1) - y(k)y^2(k-2) + b(k)\eta(k) \tag{5.7}$$

where

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

The output data sequence for this example is illustrated in Fig. 2. The parameter estimates obtained using RME are listed in table 2. The estimates at iteration 1 represent the classical ELS estimates. The same conclusions can be made as those in example S_1.

During the iteration, the denominator polynomial $b(k)$ in eqn (5.8) was replaced by $\hat{b}(k)$, the latest estimate of $b(k)$, so that eqn (5.8) takes the form

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

Example S_3 consists of the model

$$y(k) = \frac{0.8y(k-1) + u(k-1)}{1 + u^2(k-1) + y(k-1)y(k-2)} + \eta(k) \tag{5.10}$$

where

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

A zero mean Gaussian sequence with variance $\sigma_e^2 = 0.01$ was used to generate the noise $e(k)$.

The linear in the parameter model for system S_3 is

$$Y(k) = 0.8y(k-1) + u(k-1) - y(k)u^2(k-1) - y(k)y(k-1)y(k-2) + e(k-1) + e(k) \tag{5.12}$$

where

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

The output data sequence for this example is illustrated in Fig. 3. The parameter estimates obtained using RME are listed in table 3. The estimates at iteration 1 represent the classical ELS estimates. The estimates after one step are severely biased but the RME algorithm yields excellent results after just a few iterations.

Example S_4 consists of the same process model as example S_3 but with a different noise model

$$y(k) = \frac{0.8y(k-1) + u(k-1)}{1 + u^2(k-1) + y(k-1)y(k-2)} + \eta(k) \tag{5.14}$$

where

$$\eta(k) = y(k-1)e(k-1) + e(k) \tag{5.15}$$

A zero mean Gaussian sequence with variance $\sigma_e^2 = 0.02$ was used to generate the noise $e(k)$.

The linear in the parameter model for S_4 is

$$Y(k) = 0.8y(k-1) + u(k-1) - y(k)u^2(k-1) - y(k)y(k-1)y(k-2) + b(k)\eta(k) \tag{5.16}$$

where

$$Y(k) = y(k)$$
$$b(k)\eta(k) = [1 + u^2(k-1) + y(k-1)y(k-2)]\ [y(k-1)e(k-1) + e(k)] \tag{5.17}$$

The output data sequence for this example is illustrated in Fig. 4. The parameter estimates obtained using RME are listed in table 4. The estimates at iteration 1 represent the classical ELS estimates. The same conclusions can be made as those in example S_3.

During the iteration, the denominator polynomial $b(k)$ in eqn (5.17) was replaced by $\hat{b}(k)$, the latest estimate of $b(k)$, hence eqn (5.17) is expressed as

$$\hat{b}(k)\eta(k) = \hat{b}(k)y(k-1)e(k-1) + \hat{b}(k)e(k) \tag{5.18}$$

The model validity tests for examples S_2, S_3, and S_4 have not been included because they show virtually the same effects as in Fig. 1.3.

6 Conclusions

The main difficulty of applying linear least squares algorithms to stochastic rational models is caused by the noise problem. When the models are expressed in a linear-in-the-parameters form noise terms are produced which lead to biased estimates using existing linear least squares algorithms even in for the case of additive white noise.

A new RME algorithm has been presented which overcomes these problems by removing the bias in both the normal matrix $\Phi^T\Phi$ and the correlation vector $\Phi^T\vec{Y}$ for the rational model. The algorithm is much simpler than nonlinear least squares procedures because it avoids complex minimization routines.

A theoretical analysis of the convergence of the algorithm is left as an open problem but all the simulation results suggest that convergence is fairly rapid and that the algorithm is well behaved.

Acknowledgements

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

References

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

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

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

Haber, R. and H. Unbehauen, "Structure identification of non-linear dynamic systems--A survey on input/output approaches," Automatica, vol. 26, 1990.

Leontaritis, I.J. and S.A. Billings, "Input-output parametric models for nonlinear systems," Int. J. Control, vol. 41, p. Part 1, Part 2, 1985.

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

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.

Nayfeh, A.H. and D.T. Mook, Nonlinear oscillations, Wiley, New York, 1979.

Norton, J. P., An introduction to identification, Academic Press Inc. (London) Ltd, 1986.

Sales, K.R. and S.A. Billings, "Self-tuning control of nonlinear ARMAX models," Int. J. Control, vol. 51, 1990.

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

Wilks, S., Mathematical statistics, Wiley, New York, 1962.

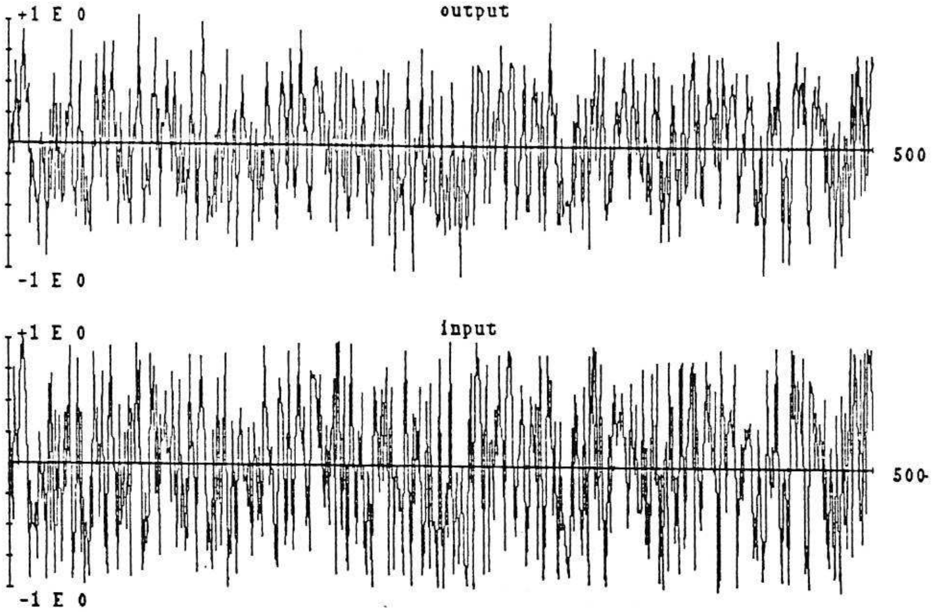
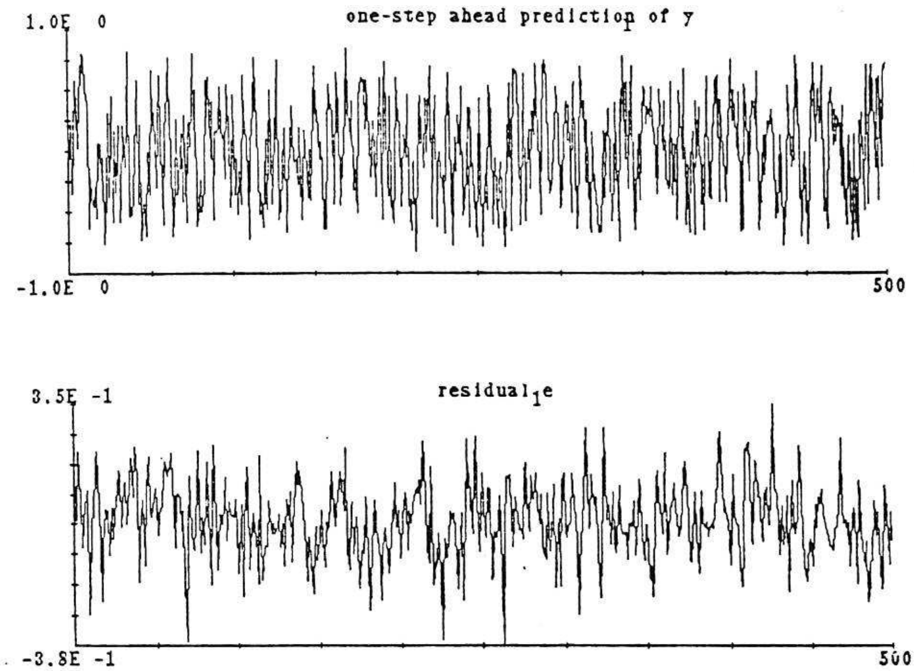
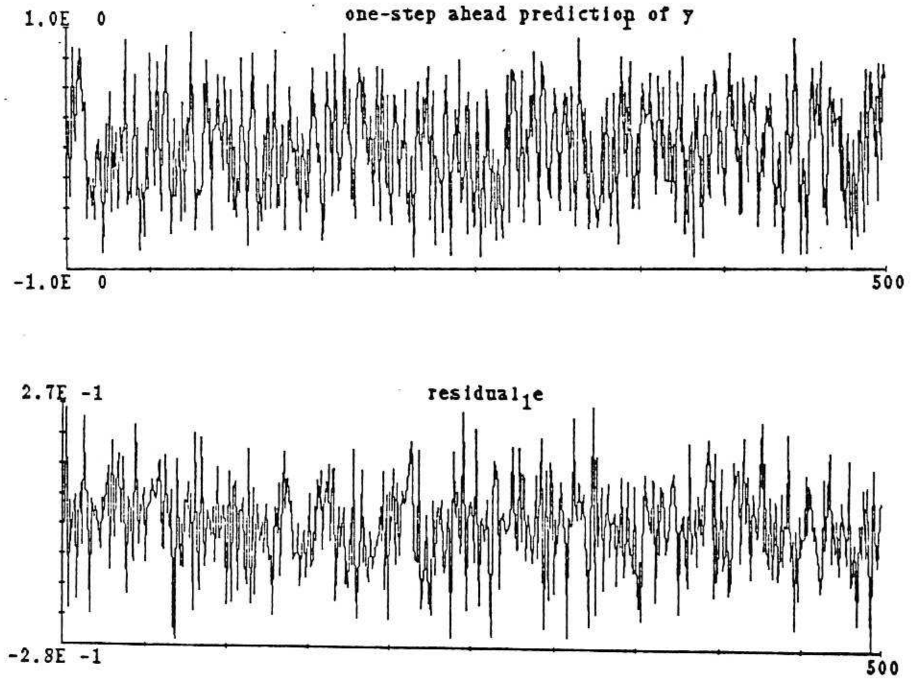


Figure 1.1 Input and output for example S_1



(a)



(b)

Figure 1.2 One step ahead predictions residuals for example S_1
(a) ELS and (b) RME

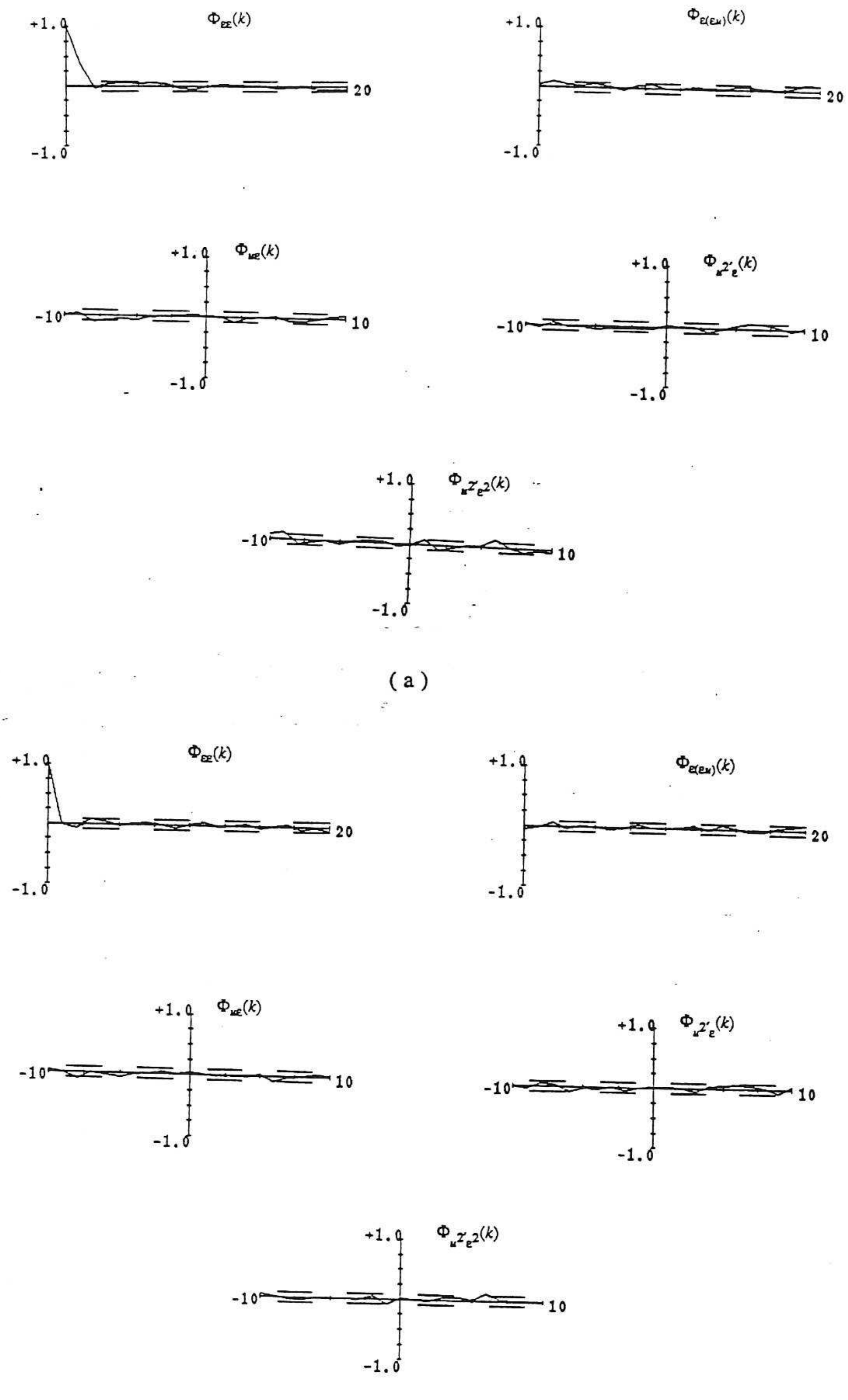


Figure 1.3 Model validition for example S_1 using correlation test (a) ELS and (b) RME

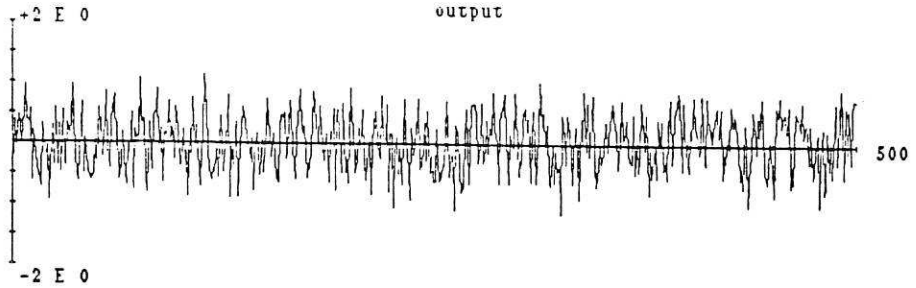


Figure 2 Output for example S_2

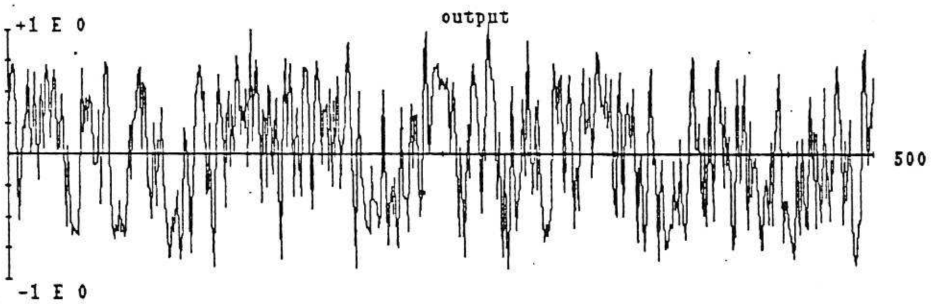


Figure 3 Output for example S_3

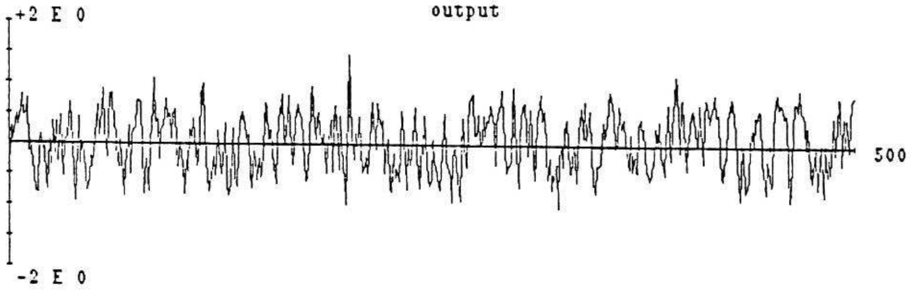


Figure 4 Output for example S_4

Itera-tion	Parameter Estimates						$\hat{\sigma}_e^2$
	y(k-1)	y(k-1)u(k-1)	u(k-1)	y(k)y(k-1)y(k-1)	y(k)y(k-2)y(k-2)	e(k-1)	
1	0.1732	0.0886	0.8790	-0.5072	-0.5370	0.0000	0.0127
2	0.1552	0.0914	0.8949	-0.6166	-0.5543	0.4191	0.0106
3	0.2208	0.1292	1.086	-1.3627	-1.2754	0.6020	0.0136
4	0.2189	0.1112	1.0493	-1.1851	-1.1869	0.5446	0.0119
5	0.2133	0.1119	1.0420	-1.1717	-1.1429	0.6150	0.0112
6	0.2105	0.1158	1.0347	-1.1471	-1.1159	0.6533	0.0107
7	0.2094	0.1168	1.0305	-1.1345	-1.1016	0.6707	0.0105
8	0.2087	0.1159	1.0294	-1.1310	-1.0932	0.6708	0.0106
9	0.2087	0.1152	1.0304	-1.1334	-1.0993	0.6771	0.0105
10	0.2085	0.1150	1.0291	-1.1292	-1.0944	0.6795	0.0104
true	0.2000	0.1000	1.0000	-1.0000	-1.0000	0.8000	0.0100

Table 1 Parameters estimated for example S_1

Itera-tion	Parameter Estimates						$\hat{\sigma}_e^2$
	y(k-1)	y(k-1)u(k-1)	u(k-1)	y(k)y(k-1)y(k-1)	y(k)y(k-2)y(k-2)	e(k-1)y(k-1)	
1	0.0675	0.0307	0.6467	0.4484	0.2084	0.0000	0.0370
2	0.0678	0.0323	0.6732	0.3320	0.1518	0.5465	0.0356
3	0.2972	0.3039	1.5889	-3.0272	-3.3478	2.1399	0.0286
4	0.2117	0.1857	1.0820	-1.1891	-1.4947	0.2072	0.0251
5	0.1775	0.1601	1.1237	-1.4571	-1.5520	0.8677	0.0221
6	0.2098	0.1551	1.0574	-1.2988	-1.1919	0.8429	0.0205
7	0.1907	0.1394	1.0268	-1.1753	-1.0331	0.9339	0.0198
8	0.1863	0.1361	0.9994	-1.0587	-0.9700	0.9125	0.0200
9	0.2039	0.1459	1.0443	-1.2541	-1.0921	1.1040	0.0189
10	0.2017	0.1392	0.9947	-1.0454	-1.0331	1.0220	0.0189
true	0.2000	0.1000	1.0000	-1.0000	-1.0000	1.0000	0.0200

Table 2 Parameters estimated for example S_2