**Monograph:**
Mao, K.Z., Billings, S.A. and Zhu, Q.M. (1996) A Regularised Least Squares Algorithm for Nonlinear Rational Model Identification. Research Report. ACSE Research Report 609 . Department of Automatic Control and Systems Engineering

# A Regularized Least Squares Algorithm for Nonlinear Rational Model Identification

K.Z.Mao S.A.Billings
Department of Automatic Control and Systems Engineering
University of Sheffield, UK

Q.M.Zhu
Department of Mechanical and Manufacturing Engineering
University of Brighton, UK

Research Report No: 609

January 31, 1996

**Abstract**

A new regularized least squares estimation algorithm is derived for the estimation of nonlinear dynamic rational models. Theoretical analysis and numerical simulations demonstrate that the new algorithm provides improved estimates compared with previously developed rational model estimators.

# 1 Introduction

The NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous input) model provides a general input-output description for a wide class of nonlinear systems and can be expanded to provide polynomial, rational and neural network model forms. The nonlinear polynomial NARMAX model has been extensively studied (Leontaritis and Billings 1985a, 1985b, Chen and Billings 1989, Haber and Unbehauen 1990, Sales and Billings 1990) and widely applied (Billings *et al* 1988, Hernandez and Arkun 1993, Proll and Karim 1994, Sriniwas *et al* 1995). Recently, the nonlinear rational model formulation, defined as the ratio of two polynomial expansions, has been introduced. Rational functions provide superior approximation properties compared to polynomials and are widely used in static function approximation (Ratkowsky 1987, Ford *et al* 1989, Ponton 1993). These properties carry over to the dynamic rational NARMAX model which can be considered as a general form with polynomial and neural network expansions as special cases (Billings and Chen 1989, Billings and Zhu 1991, Zhu and Billings 1993, Billings and Zhu 1994a).

Despite the attractive advantages the nonlinear rational model is much more difficult to identify than the polynomial model because rational models are inherently nonlinear in the parameters. One of the first estimators for dynamic nonlinear rational models was the prediction error (PE) algorithm proposed by Billings and Chen (1989). Because the prediction error (PE) algorithm is based on nonlinear optimization, it is computaionally expensive and the estimates may not correspond to the global minimum. To avoid this problem Billings and Zhu (1991) developed the Rational Model Estimator (RME) by multiplying out the rational model to form a linear-in-the-parameters expression. The main contribution of the RME algorithm was to show how the severe noise problems which are induced by multiplying out the model can be avoided by reformulating the least squares solution. The noise problems arise as a direct consequence of making the rational model linear in the parameters and, unlike linear polynomial or neural network models, this induces severe bias in the rational model estimates even for low levels of white noise. Several algorithms have been developed based on this result, Zhu and Billings (1991), Zhu and Billings (1993), Billings and Zhu (1994a), and the RME has been shown to provide consistent estimates for nonlinear rational models.

Often the first requirement when designing an estimator is to ensure that the estimates are consistent. Under this constraint, however, the estimator can not guarantee that the minimum mean squared error (MSE) or covariance will be achieved. Both of these measure estimator precison and are associated with prediction performance, so that strengthing the consistency property might lead to an inflation in mean squared error (MSE). Consequently, the estimates obtained from fitting the model to particular samples might be far from the true values. Regularization can be used to prevent mean squared error (MSE) inflation (Hoerl and Kennard 1970, Barron and Xiao 1991, Bishop 1991, Orr 1995) by trading off the MSE reduction at the expense of the consistency property of the algorithms.

In the present study, the mean squared error of the RME (Billings and Zhu, 1991) estimate is initially investigated. This reveals that the MSE inflation problem exists in RME and it is shown how this can lead to a deterioration in prediction performance. To overcome this problem, a new regularized rational model identification algorithm is derived where the requirements of consistent estimates and minimum mean squared error can be balanced. Theoretical analysis and numerical simulations demonstrate that the new algorithm provides improved estimates.

This paper is organized as follows, in Section 2 the RME estimator is reviewed and the reason why the prediction performance may deteriorate is analysized from the MSE inflation point of view. A new regularized RME is derived in Section 3.1, and the orthogonal regularized RME is developped in Section 3.2. Section 4 provides some numerical examples to demonstrate the performance of the new algorithm.

# 2 Rational Model Estimator (RME) and associated statistical properties

## 2.1 Brief review of RME

Consider the dynamical nonlinear rational model given by:

$$y(k) = \frac{F_1[y(k-1)\dots y(k-n_{ny}), u(k-1)\dots u(k-n_{nu}), e(k-1)\dots e(k-n_{ne})]}{F_2[y(k-1)\dots y(k-n_{dy}), u(k-1)\dots u(k-n_{du}), e(k-1)\dots e(k-n_{de})]} + e(k) \tag{1}$$

where $u(k)$ and $y(k)$ denote the input and output at time $k(k=0,1,\dots)$ respectively, $\{e(k)\}$ is an unobservable independent noise sequence with zero mean and finite variance $\sigma^2$ . $F_1(\bullet)$ and $F_2(\bullet)$ are nonlinear polynomials, $n_\bullet$ denotes order.

Equation (1) can also be written to the following regressor-parameter form:

$$y(k) = \frac{\sum_{j=1}^{nnum} \varphi_{nj}(k)\theta_{nj}}{\sum_{j=1}^{nden} \varphi_{dj}(k)\theta_{dj}} + e(k) \tag{2}$$

where $\varphi_\bullet(k)$ and $\theta_\bullet$ denote regressor and parameter respectively, $n_{num} = n_{ny} + n_{nu} + n_{ne}$, $n_{den} = n_{dy} + n_{du} + n_{de}$.

Expresson (2) is nonlinear in the parameters and conventional least squares (LS) estimators can not be applied directly to obtain unbiased parameter estimates. One of the first estimation algorithms for nonlinear rational models was the prediction error (PE) algorithm proposed by Billings and Chen (1989). But this is based on a nonlinear optimization method which can be computationally expensive and can lead to local minima problems.

An alternative is to multiply out the rational model into a linear-in-the -parameters form (Billings and Zhu 1991) to yield

$$\begin{aligned}
Y(k) &= F_1(k) - y(k)[F_2(k) - \varphi_{d1}(k)\theta_{d1}] + F_2(k)e(k) \\
&= \sum_{j=1}^{nnum} \varphi_{nj}(k)\theta_{nj} - \sum_{j=2}^{nden} \varphi_{dj}(k)y(k)\theta_{dj} + \xi(k) \\
&= \phi(k)\Theta + \xi(k) \\
&= [\phi_{free}(k) + \phi_{noise}(k)]\Theta + \xi(k) \\
&= \phi_{free}(k)\Theta + \xi_1(k)
\end{aligned} \tag{3}$$

where:

$$Y(k) = y(k)\varphi_{d1}(k)\theta_{d1} \mid_{\theta_{d1}=1}$$

$$\xi(k) = F_2(k)e(k)$$

$$\xi_1(k) = \varphi_{d1}(k)e(k)$$

$$\overline{y}(k) = y(k) - e(k) = \frac{F_1(\bullet)}{F_2(\bullet)}$$

$$\Theta = [\theta_{n1} \ldots \theta_{nnnum} \;\; \theta_{d2} \ldots \theta_{dnden}]^T$$

$$\phi(k) = [\varphi_{n1}(k) \ldots \varphi_{nnnum}(k), \;\; -\varphi_{d2}(k)y(k) \ldots - \varphi_{dnden}(k)y(k)]$$

$$\phi_{free}(k) = [\varphi_{n1}(k) \ldots \varphi_{nnnum}(k), \;\; -\varphi_{d2}(k)\overline{y}(k) \ldots - \varphi_{dnden}(k)\overline{y}(k)]$$

$$\phi_{noise}(k) = [0 \ldots 0, \;\; -\varphi_{d2}(k)e(k) \ldots - \varphi_{dnden}(k)e(k)]$$

Define

$$Y = [Y(1), Y(2) \ldots Y(N)]^T$$

$$\Phi = [\phi^T(1), \phi^T(2) \ldots \phi^T(N)]^T$$

$$\Phi_{free} = [\phi^T_{free}(1), \phi^T_{free}(2) \ldots \phi^T_{free}(N)]^T$$

$$\Phi_{noise} = [\phi^T_{noise}(1), \phi^T_{noise}(2) \ldots \phi^T_{noise}(N)]^T$$

$$\Xi = [\xi(1), \xi(2) \ldots \xi(N)]^T$$

$$\Xi_1 = [\xi_1(1), \xi_1(2) \ldots \xi_1(N)]^T$$

to give the following vector equation

$$\begin{aligned} Y &= \Phi\Theta + \Xi = (\Phi_{free} + \Phi_{noise})\Theta + \Xi & (4) \\ &= \Phi_{free}\Theta + \Xi_1 & (5) \end{aligned}$$

If ordinary least squares (LS) is applied the estimate would be of the form

$$\hat{\Theta}_{LS} = (\Phi^T\Phi)^{-1}\Phi^T Y \tag{6}$$

However, the probability limit (denoted by $Plim$) of this estimate

$$\begin{aligned} Plim(\hat{\Theta}_{LS}) &= Plim[(\Phi^T\Phi)^{-1}\Phi^T Y] \\ &= Plim[(\Phi^T\Phi)^{-1}\Phi^T(\Phi\Theta + \Xi)] \\ &= \Theta + Plim(\frac{\Phi^T\Phi}{N})^{-1} \bullet Plim(\frac{\Phi^T\Xi}{N}) \end{aligned}$$

indicates that the ordinary LS estimate is not consistent because of the correlation between the noise vector $\Xi$ and $\Phi_{noise}$ included in the regressor vector $\Phi$, eqn(4).

To overcome the problem, Billings and Zhu (1991) proposed a modified least squares algorithm tailored specially for the rational model and called RME by taking the correlation between $\Xi$ and $\Phi$, and that between $\Phi$ and $\Phi$ into consideration to yield

$$\hat{\Theta}_{RME} = (\Phi^T\Phi - N\psi_1)^{-1}(\Phi^T Y - N\psi_2) \tag{7}$$

where $E(\bullet)$ denotes the expection of $(\bullet)$, and

$$\psi_1 = E[\phi^T_{noise}(k)\phi_{noise}(k)] \qquad \psi_2 = E[\phi^T_{noise}(k)\xi_1(k)]$$

## 2.2 Mean squared error (MSE) inflation and RME

When an estimator and the quality of a model is evaluated it is often necessary to examine both the mean and the mean squared error (MSE) or covariance of the estimate. The mean is related to consistency and the mean squared error (MSE) measures both the bias and variance of the estimate.

It has been demonstrated (Billings and. Zhu 1991, Zhu and Billings 1993, Billings and Zhu 1994a) that the Rational Model Estimator (RME) provides consistent estimates. Consider therefore an analysis of the mean squared error (MSE) of $\hat{\Theta}_{RME}$:

$$\begin{aligned} MSE(\hat{\Theta}_{RME}) &= E[(\hat{\Theta}_{RME} - \Theta)^T(\hat{\Theta}_{RME} - \Theta)] \\ &= Tr[cov(\hat{\Theta}_{RME})] \end{aligned}$$

where $cov(\bullet), E(\bullet)$ and $Tr(\bullet)$ denote the covariance, the expection and the trace of $(\bullet)$ respectively.

$$\begin{aligned} cov(\hat{\Theta}) &= E[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] \\ &= E\left\{[(\Phi^T\Phi - N\psi_1)^{-1}(\Phi^TY - N\psi_2) - \Theta][(\Phi^T\Phi - N\psi_1)^{-1}(\Phi^TY - N\psi_2) - \Theta]^T\right\} \\ &= E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\ &\quad +E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta\Xi_1^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\ &\quad -E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta\Theta^T] \\ &\quad -E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta N\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] \\ &\quad +E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Xi_1\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\ &\quad +E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}]\sigma_{\xi_1}^2 \\ &\quad -E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Xi_1N\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] \\ &\quad -E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Xi_1\Theta^T] \\ &\quad -E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\ &\quad +E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2\Theta^T] \end{aligned}$$

$$
\begin{aligned}
&-E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2\Xi_1^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&+E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2 N\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] \\
&-E[\Theta\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&-E[\Theta\Xi_1^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] + E[\Theta\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] + \Theta\Theta^T
\end{aligned}
$$

where $\sigma_{\xi_1}^2$ denotes the variance of noise sequence $\{\xi_1(t)\}$ in eqn (3).

The probability limit of the mean squared error of $\hat{\Theta}_{RME}$ is then obtained (see Appendix A for details):

$$
\begin{aligned}
Plim[MSE(\hat{\Theta}_{RME})] &= Tr\left\{[Plim(\frac{\Phi^T\Phi}{N}) - \psi_1]^{-1}[\psi_1\frac{\sigma_{\xi_1}^2}{N} - \psi_2\psi_2^T][Plim(\frac{\Phi^T\Phi}{N}) - \psi_1]^{-1}\right\} \\
&+Tr\left\{[Plim(\frac{\Phi^T\Phi}{N}) - \psi_1]^{-1}\frac{\sigma_{\xi_1}^2}{N}\right\} \qquad (8)
\end{aligned}
$$

Note that the mean square error (MSE) of the ordinary LS estimate eqn (5) is

$$
MSE(\hat{\Theta}_{LS}) = Tr\left\{[Plim(\frac{\Phi^T\Phi}{N})]^{-1}\frac{\sigma^2}{N}\right\} \qquad (9)
$$

Comparing (8) and (9), shows that the MSE of $\hat{\Theta}_{RME}$ will be inflated if

- $\sigma_{\xi_1}^2 > \sigma^2$
- $Tr(\psi_1\frac{\sigma_{\xi_1}^2}{N}) > Tr(\psi_2\psi_2^T)$
- $Tr\left\{[Plim(\frac{\Phi^T\Phi}{N}) - \psi_1]^{-1}\right\} > Tr\left\{Plim[(\frac{\Phi^T\Phi}{N})^{-1}]\right\}$

The first and the second conditions are system and signal dependent, but the third is definitely true because the matrix $\psi_1$ is positive definite.

The investigation of the mean squared error (MSE) above therefore shows that while RME provides consistent estimates, this may be accompanied by MSE inflation as shown in Fig.1. A direct result of mean squared error (MSE) inflation is that the estimates obtained from fitting the model to a particular set of data may differ from the true values and result in the predictions differing from the observations. Notice that the ordinary least squares (LS) estimate will be biased and inconsistent due to the inherent noise problems discussed in Section 2.1. Conversely, MSE inflation seems to be a drawback of RME and the factor that can lead to a deterioration in predictive performance.

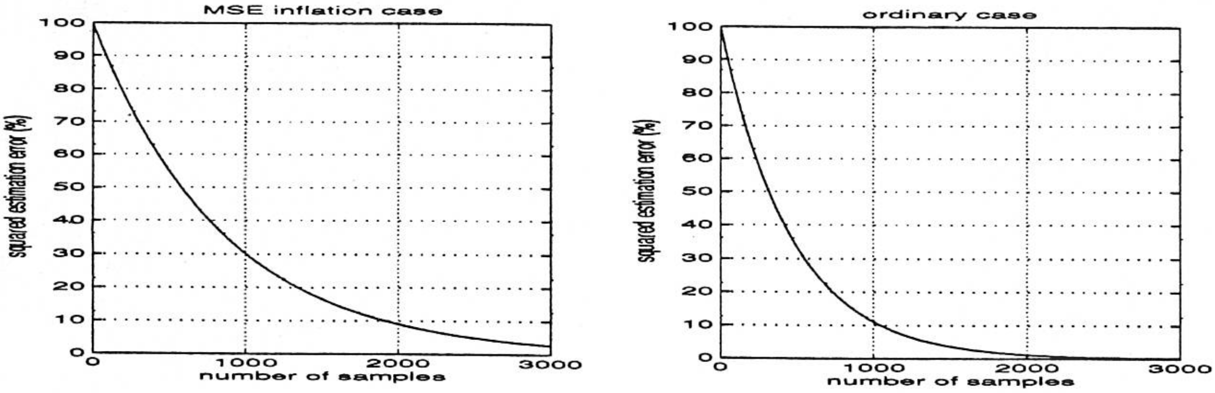


Figure 1: MSE comparison of inflation case eqn (8) and ordinary case eqn (9)

# 3 Regularized Rational Model Estimator (RRME)

## 3.1 The basic RRME

In Section 2, the deterioration in prediction performance from a MSE inflation point of view was analysed. The consistency property of RME conflicts to some extent with the prediction performance. To balance both requirements a compromise between these two factors can be made by introducing

$$
\begin{aligned}
J_1 &= \sum_{k=1}^{N} \left\{ [Y(k) - \phi(k)\hat{\Theta}]^2 + \hat{\Theta}^T \lambda \hat{\Theta} \right\} \\
&= [Y - \Phi\hat{\Theta}]^T [Y - \Phi\hat{\Theta}] + N\hat{\Theta}^T \lambda \hat{\Theta} \qquad (10)
\end{aligned}
$$

where $\lambda = Diag\{\lambda_i, \quad i = 1, 2...\}$ and $\lambda_i \geq 0$ is the regularized parameter. Optimizing $J_1$ with respect to $\hat{\Theta}$, yields

$$
\hat{\Theta} = (\Phi^T\Phi + N\lambda)^{-1}\Phi^T Y \qquad (11)
$$

Taking the correlation between $\Phi$ and $\Xi$ and that between $\Phi$ and $\Phi$ into consideration (Billings and Zhu 1991), yields the following Regularized Rational Model Estimator (RRME)

$$
\hat{\Theta}_{RRME} = [\Phi^T\Phi + N\lambda - N\psi_1]^{-1}[\Phi^T Y - N\psi_2] \qquad (12)
$$

The probability limit of $\hat{\Theta}_{RRME}$ is given as follows

$$
\begin{aligned}
Plim(\hat{\Theta}_{RRME}) &= [Plim(\frac{\Phi^T\Phi}{N}) + \lambda - \psi_1]^{-1} Plim[\frac{\Phi^T(\Phi_{free} + \Xi_1)}{N} - \psi_2] \\
&= [Plim(\frac{\Phi_{free}^T\Phi_{free}}{N}) + \lambda]^{-1} \bullet Plim(\frac{\Phi_{free}^T\Phi_{free}}{N})\Theta \\
&= \Theta - \lambda[Plim(\frac{\Phi_{free}^T\Phi_{free}}{N}) + \lambda]^{-1}\Theta \\
&= \Theta + Bias(\hat{\Theta}_{RRME}) \qquad (13)
\end{aligned}
$$

Obviously, the estimate via regularization is not consistent. If $\lambda = 0$, $Bias(\hat{\Theta}_{RRME}) = 0$ and the regularized rational model estimator (RRME) is reduced to RME.

## 3.2 Orthogonal RRME (ORRME) and regularization parameter selection

Although the analysis in previous sections shows how to obtain a regularized rational model estimate, it is also important to consider how to determine the structure or which terms to include in model eqn (1). Indeed the structure detection problem is much more challenging than parameter estimation and the following orthogonal transformation is introduced to address the problem

$$W = \Phi T^{-1} \tag{14}$$

where T is the orthogonal transformation matrix

$$T = \begin{bmatrix} 1 & T_{12} & \dots & T_{1p} \\ 0 & 1 & \dots & T_{2p-1} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$W = W_{free} + W_{noise} = \begin{bmatrix} w_{1free}(1) & \dots & w_{pfree}(1) \\ \vdots & \dots & \vdots \\ w_{1free}(N) & \dots & w_{pfree}(N) \end{bmatrix} + \begin{bmatrix} w_{1noise}(1) & \dots & w_{pnoise}(1) \\ \vdots & \dots & \vdots \\ w_{1noise}(N) & \dots & w_{pnoise}(N) \end{bmatrix}$$

$$W^T W / N = Diag\{d_i, \quad i = 1, 2 \dots p\}$$

$$p = n_{num} + n_{den} - 1$$

The orthogonal transformation matrix T can be computed using a Gram-Schmidt scheme, full details of which can be found in Billings and Zhu (1994). The regressors $w_{ifree}(k)$ and $w_{inoise}(k)$ respectively correspond to $ww_i(k)$ and $e_i(k)$ defined in the aforementioned paper.

Applying the orthogonal transformation to (4), the regression equation is converted to:

$$\begin{aligned} Y &= WG + \Xi = (W_{free} + W_{noise})G + \Xi \\ &= W_{free}G + \Xi_1 \end{aligned} \tag{15}$$

where $G = T\Theta$ is the new parameter vector.

Analogous to the basic RRME, the Orthogonal Regularized Rational Model Estimator (ORRME) is given by:

$$\hat{G} = [W^T W + N\lambda - N\Psi_1]^{-1}[W^T Y - N\Psi_2] \tag{16}$$

where

$$w_{noise}(k) = [w_{1noise}(k) \dots w_{pnoise}(k)]$$

$$\Psi_1 = E[w_{noise}(k)^T w_{noise}(k)] = Diag\,\{\Psi_{1i}, \quad i = 1, 2 \dots p\}$$

$$\Psi_2 = E[w_{noise}^T(k)\xi_1(k)] = [\Psi_{21} \dots \Psi_{2p}]^T$$

The $i-th$ element of $\hat{G}$ is:

$$\hat{g}_i = \frac{w_i Y/N - \Psi_{2i}}{d_i + \lambda_i - \Psi_{1i}} \tag{17}$$

where $w_i = [w_i(1) \dots w_i(N)] = [w_{ifree}(1) + w_{inoise}(1) \dots w_{ifree}(N) + w_{inoise}(N)]$.

The mean and the mean squared error of $\hat{g}_i$ can now be investigated using

$$\begin{aligned} \hat{g}_i &= \frac{w_i Y/N - \Psi_{2i}}{d_i + \lambda_i - \Psi_{1i}} \\ &= \frac{w_i(W_{free}G + \Xi_1)/N - \Psi_{2i}}{d_i + \lambda_i - \Psi_{1i}} \\ &= \frac{(w_{ifree} + w_{inoise})(W_{ifree}G + \Xi_1)/N - \Psi_{2i}}{d_i + \lambda_i - \Psi_{1i}} \end{aligned} \tag{18}$$

The probability limit of the estimate is

$$\begin{aligned}
\overline{g_i} = Plim(\hat{g}_i) &= \frac{Plim(w_{ifree}w_{ifree}^T/N)g_i + Plim(w_i\Xi_1/N) - \Psi_{2i}}{Plim(d_i) + \lambda_i - \Psi_{1i}} \\
&= \frac{d_i' g_i}{d_i' + \lambda_i} \\
&= g_i + Bias[\hat{g}_i] \qquad (19)
\end{aligned}$$

where

$$d_i = \frac{w_i w_i^T}{N}$$

$$d_i' = Plim(\frac{w_{ifree}w_{ifree}^T}{N}) = Plim(d_i) - \Psi_{1i}$$

$$Bias[\hat{g}_i] = -\frac{\lambda_i g_i}{d_i' + \lambda_i}$$

The probability limit of the covariance is given as follow

$$\begin{aligned}
\overline{cov(\hat{g}_i)} &= Plim[E(\hat{g}_i - g_i)^2] \\
&= Plim[E(\hat{g}_i - \overline{g_i} + \overline{g_i} - g_i)^2] \\
&= Plim[E(\hat{g}_i - \overline{g_i})^2] + Plim[E(\overline{g_i} - g_i)^2] \qquad (20)
\end{aligned}$$

where the first part of the right side of (20) is equal to (see Appendix B for details)

$$\begin{aligned}
Plim[E(\hat{g}_i - \overline{g_i})^2] &= Plim\left\{E\left[\frac{w_iY/N - \Psi_{2i}}{d_i + \lambda_i - \Psi_{1i}} - g_i + \frac{\lambda_i g_i}{d_i' + \lambda_i}\right]^2\right\} \\
&= \frac{Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T\Psi_{2i}}{(d_i' + \lambda_i)^2} \qquad (21)
\end{aligned}$$

The second part of the right side of (20) is equal to

$$Plim[E(\overline{g_i} - g_i)^2] = \frac{\lambda_i^2 g_i^2}{(\lambda_i + d_i')^2} \qquad (22)$$

Therefore the mean squared error (MSE) of the regularized estimate is

$$MSE(\hat{G}_{ORRME}) = \sum_{i=1}^{p} \frac{Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T \Psi_{2i}}{(d_i' + \lambda_i)^2} + \frac{\lambda_i^2 g_i^2}{(\lambda_i + d_i')^2} \tag{23}$$

The MSE of nonregularized estimator can be obtained by substituting $\lambda_i = 0$ into (23) to yield

$$MSE(\hat{G}_{ORME}) = \sum_{i=1}^{p} \frac{Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T \Psi_{2i}}{d_i'^2} \tag{24}$$

The results eqn (23) and eqn (24) are obtained under the assumption that sample number $N$ goes to infinity. For a large finite $N$, the results are approximately true. A reduction of mean squared error (MSE) is the aim of employing regularization and the regularized parameter $\lambda_i$ should guarantee

$$\frac{Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T \Psi_{2i}}{(d_i' + \lambda_i)^2} + \frac{\lambda_i^2 g_i^2}{(\lambda_i + d_i')^2} \leq \frac{Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T \Psi_{2i}}{d_i'^2} \tag{25}$$

Solving the inequality, provides a range for the regularized parameter

$$0 < \lambda_i \leq \frac{[2Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T \Psi_{2i}]d_i'}{g_i^2 d_i'^2 + \Psi_{2i}^T \Psi_{2i} - Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N} \tag{26}$$

which will result in an improvement in the MSE.

Expression (25) provides no practical guidance because the upper bound contains the unknown true parameters. Comparing (18) and (19), yields

$$g_i d_i' = Plim(w_i Y/N) - \Psi_{2i} \tag{27}$$

Substituting (27) into (26), the range of $\lambda_i$ is then obtained

$$0 < \lambda_i \leq \frac{[2Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T\Psi_{2i}][Plim(w_i w_i^T/N) - \Psi_{1i}]}{[Plim(w_i Y/N) - \Psi_{2i}]^2 + \Psi_{2i}^T\Psi_{2i} - Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N} \tag{28}$$

In practice the upper bound of $\lambda_i$ can be computed by replacing $Plim(\bullet/N)$ with $(\bullet)/N$ in (28) when N is sufficiently large.

After obtaining the orthogonal parameter estimates $\hat{g}_i$, the parameters in the original model (2) can be computed from

$$\hat{\theta}_p = \hat{g}_p$$

$$\hat{\theta}_i = \hat{g}_i - \sum_{j=i+1}^{p} T_{ij}\hat{\theta}_j, \qquad i = p-1\ldots 2, 1$$

**Remark 1**

When solving inequality (25) the following assumption has been made

$$(g_i d_i')^2 > Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T\Psi_{2i}$$

which is equivalent to

$$g_i^2 > \frac{Plim(w_i w_i^T/N)\sigma_{\xi_1}^2/N - \Psi_{2i}^T\Psi_{2i}}{d_i'^2}$$

The physical explanation of the above inequality is that the size of the parameter to be estimated is larger than that of the estimation error (in terms of eqn (24)), which is almost always true. Therefore the assumption is reasonable.

**Remark 2**

The regularization approach was introduced for the purpose of improving and enhancing RME estimation. In fact, regularization can also overcome overfitting when the raw data is very noisy, and alleviate ill-conditioning which often occurs when the elements of the parameter vector are of different orders of magnitude.

# 4 Numerical simulations

To test the performance of the proposed regularization algorithm, two examples will be used as a demonstration. In both examples, the input $u(k)$ was a uniformly distributed random sequence with zero mean and amplitude $\pm 1$, the noise $e(k)$ was a normally distrubuted disturbance sequence with zero mean and variance 0.01. In each example, 600 pairs of input/output data were produced, where 500 pairs were used for fitting the model, and the remaining 100 pairs were used to test the predictive performance of the estimator.

**Example 1**

Consider the nonlinear stochastic model:

$$y(k) = \frac{y(k-1) + u(k-1)e(k-1)}{1 + y^2(k-1)} + e(k) \tag{29}$$

The equivalent linear-in-the-parameter expression is given as follows

$$Y(k) = y(k-1) + u(k-1)e(k-1) - y^2(k-1)y(k) + \xi(k) \tag{30}$$

where

$$Y(k) = y(k)$$

$$\xi(k) = [1 + y^2(k-1)]e(k)$$

Applying the regularization algorithm RRME and non-regularized RME (Billings and Zhu 1991) to the unknown model, the parameter estimates obtained are listed in Table 1. The true output and one-step ahead predicted output for the RRME algorithm is shown in Fig.2, the model validity tests (Billings and Zhu, 1994b) are shown in Fig.3

Comparing the results obtained from the two algorithms, shows that the new regularization algorithm proposed in this paper provides more accurate estimates and than the non-regularized RME.

**Example 2**

Consider a more complex dynamical stochastic nonlinear rational model:

$$y(k) = \frac{y(k-1) + u(k-1)u(k-2) + y(k-2)e(k-2)}{1 + y^2(k-1) + u^2(k-2)} + e(k) \tag{31}$$

The linear-in-the-parameter expression of the rational model is given as follows

$$\begin{aligned} Y(k) &= y(k-1) + u(k-1)u(k-2) + y(k-2)e(k-2) \\ &\quad -y^2(k-1)y(k) - u^2(k-2)y(k) + \xi(k) \end{aligned} \tag{32}$$

where

$$Y(k) = y(k)$$

$$\xi(k) = [1 + y^2(k-1) + u^2(k-2)]e(k)$$

Applying the regularization algorithm RRME and the non-regularized RME respectively to the unknown model, the estimates are listed in Table 2. The true output and the one-step ahead predicted output for the RRME algorithm is shown in Fig.4, the model validity tests (Billings and Zhu, 1994b) are shown in Fig.5

A comparison of the results again shows that the regularized algorithm provides improved results compared to the nonregularized scheme.

# 5 Conclusions

A new regularized least squares estimator has been derived for the dynamic rational model. A detailed analysis of the property of the new estimator showed how regularization can be used to balance both mean squared error performance and consistency of the estimates. Simulation results confirmed the advantages offered by the new algorithm.

# 6 Acknowledgement

SAB gratefully acknowledges that part of this work was supported by EPSRC.

# References

[1] Barron,A.R. and X.Xiao, 1991, Discussion of multivariate adaptive regression splines, *Annuals of Statistics*, 19, 67-82.

[2] Billings, S.A., Fadzil, M.B., Sulley, J.L. and P.M.Johnson, 1988, Identification of a nonlinear difference equation model of an industrial diesel generator, *Mechanical Systems and Signal Processing* , 2, 59-76.

[3] Billings, S.A. and S.Chen, 1989, Identification of nonlinear rational systems using a prediction error estimation algorithm, *International Journal of System Science*, 20, 467-494.

[4] Billings, S.A. and Q.M.Zhu, 1991, Rational model identification using an extended least square algorithm, *International Journal of Control* , 54, 529-564.

[5] Billings, S.A. and Q.M.Zhu, 1994a, A structure detection algorithm for nonlinear dynamic rational models, *International Journal of Control* , 59, 1439-1463.

[6] Billings, S.A. and Q.M.Zhu, 1994b, Nonlinear model validation using correlation tests, *International Journal of Control* , 60, 1107-1120.

[7] Bishop,C., 1991, Improving the generalization properties of radial basis function neural networks, *Neural Computation*, 3, 579-588.

[8] Chen, S. and S.A.Billings, 1989, Representations of nonlinear systems:the NARMAX model, *International Journal of Control* , 48, 1013-1032.

[9] Ford,I., Titterington, D.M. and C.P.kitsos, 1989, Recent advance in nonlinear experimental design, *Technometrics* , 31, 49-60.

[10] Haber,R. and H.Unbehauen, 1990, Structure identification of non-linear dynamic system-a survey on input/output approaches, *Automatica*, 26, 651-677.

[11] Hernandez, E. and Y.Arkun, 1993, Control of nonlinear system using polynomial ARMA models, *AIChe Journal* , 39, 446-462.

[12] Hoerl, A. and R.Kennard, 1970, Ridge regression:application to nonorthogonal problems, *Technometrics*, 12, 55-67.

[13] Leontaritis,I.J. and S.A. Billings, 1985a, Input-output parametric models for nonlinear systems, Part One, *International Journal of Control*, 41, 303-328.

[14] Leontaritis,I.J. and S.A. Billings, 1985b, Input-output parametric models for nonlinear systems, Part Two, *International Journal of Control*, 41, 329-344.

[15] Norton, J.P., 1986, *An introduction to identification* (London:Academic Press).

[16] Orr, M.J.L., 1995, Regularization in the selection of radial basis function center, *Neural Computation* , 7, 606-623.

[17] Ponton, J.W., 1993, The Use of multivariable rational functions for nonlinear data presentation and classification, *Computers and Chemical Engineering*, 17, 1047-1052.

[18] Proll,T. and M.N. Karim, 1994, Model-predictive PH control using real-time NARX approach, *AIChe Journal*, 40, 269-184.

[19] Ratkowsky, D.A., 1987, The statistical properties and utility of rational functions models, *The Canadian Journal of Chemical Engineering*, 65, 845-851.

[20] Sales, K.R. and S.A. Billings, 1990, Self-tuning control of non-linear ARMAX models,. *International Journal of Control*, 51, 753-769.

[21] Sriniwas,G.R., Arkun,Y., Chien,I. and B.A. Ogunnaike, 1995, Nonlinear identification and control of a high-purity distillation collumn: a case study, *Journal of Process Control*, 5, 149-162.

[22] Zhu, Q.M., and S.A.Billings, 1991, Recursive parameter estimation for nonlinear rational models, *Journal of System Engineering*, 1, 63-76.

[23] Zhu, Q.M. and S.A.Billings, 1993, Parameter estimation for stochastic nonlinear rational models, *International Journal of Control* , 57, 309-333.

# Appendix A: Properties of $\hat{\Theta}_{RME}$

The RME estimate is given by eqn (6) as

$$\begin{aligned}
\hat{\Theta}_{RME} &= (\Phi^T\Phi - N\Psi_1)^{-1}(\Phi^T Y - N\Psi_2) \\
&= (\Phi^T\Phi - N\Psi_1)^{-1}[\Phi^T(\Phi_{free}\Theta + \Xi_1) - N\Psi_2] \\
&= (\Phi^T\Phi - N\Psi_1)^{-1}\Phi^T\Phi_{free}\Theta + (\Phi^T\Phi - N\Psi_1)^{-1}\Phi^T\Xi_1 - (\Phi^T\Phi - N\Psi_1)^{-1}N\Psi_2
\end{aligned}$$

The covariance of the estimate is given by:

$$\begin{aligned}
cov(\hat{\Theta}) &= E[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] \\
&= E\left\{[(\Phi^T\Phi - N\psi_1)^{-1}(\Phi^T Y - N\psi_2) - \Theta][(\Phi^T\Phi - N\psi_1)^{-1}(\Phi^T Y - N\psi_2) - \Theta]^T\right\} \\
&= E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad + E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta\Xi_1^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad - E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta\Theta^T] \\
&\quad - E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi_{free}\Theta N\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad + E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Xi_1\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad + E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}]\sigma_{\xi_1}^2 \\
&\quad - E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Xi_1 N\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad - E[(\Phi^T\Phi - N\psi_1)^{-1}\Phi^T\Xi_1\Theta^T] \\
&\quad - E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] + E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2\Theta^T] \\
&\quad - E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2\Xi_1^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad + E[(\Phi^T\Phi - N\psi_1)^{-1}N\psi_2 N\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] - E[\Theta\Theta^T\Phi_{free}^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] \\
&\quad - E[\Theta\Xi_1^T\Phi(\Phi^T\Phi - N\psi_1)^{-1}] + E[\Theta\psi_2^T(\Phi^T\Phi - N\psi_1)^{-1}] + \Theta\Theta^T
\end{aligned}$$

where $\sigma_{\xi_1}^2$ denotes the variance of noise sequence $\{\xi_1(t)\}$.

Therefore the Probability limit of the covariance is:

$$
\begin{aligned}
Plim[cov(\hat{\Theta})] &= [Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Phi_{free}}{N})\Theta\Theta^T Plim(\frac{\Phi^T_{free}\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&+[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Phi_{free}}{N})\Theta Plim(\frac{\Xi_1^T\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&-[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Phi_{free}}{N})\Theta\psi_2^T[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&-[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Phi_{free}}{N})\Theta\Theta^T \\
&+[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Xi_1}{N})\Theta^T Plim(\frac{\Phi^T_{free}\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&+[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}\frac{\sigma^2_{\xi_1}}{N} \\
&-[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Xi_1}{N}\psi_2^T[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&-[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}Plim(\frac{\Phi^T\Xi_1}{N})\Theta^T \\
&-[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}\psi_2\Theta^T Plim(\frac{\Phi^T_{free}\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&-[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}\psi_2 Plim(\frac{\Xi_1^T\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&+[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}\psi_2\psi_2^T[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}+[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}\psi_2\Theta^T \\
&-\Theta\Theta^T Plim(\frac{\Phi^T_{free}\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}-\Theta Plim(\frac{\Xi_1^T\Phi}{N})[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1} \\
&+\Theta\psi_2^T[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}+\Theta\Theta^T \\
&= \Theta\Theta^T+\Theta\psi_2^T[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}-\Theta\psi_2^T[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1} \\
&-\Theta\Theta^T+[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\Theta^T \\
&+[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}Plim(\frac{\Phi^T\Phi}{N})[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\frac{\sigma^2_{\xi_1}}{N} \\
&-[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\psi_2^T[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}-[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\Theta^T \\
&-[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\Theta^T-[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\psi_2^T[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1} \\
&+[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\psi_2^T[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}-\Theta\Theta^T \\
&+[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}\psi_2\Theta^T-\Theta\psi_2^T[Plim(\frac{\Phi^T_{free}\Phi_{free}}{N})]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
& +\Theta\psi_2^T[Plim(\frac{\Phi_{free}^T\Phi_{free}}{N})]^{-1}+\Theta\Theta^T \\
= \; & [Plim(\frac{\Phi_{free}^T\Phi_{free}}{N})]^{-1}Plim(\frac{\Phi^T\Phi}{N})[Plim(\frac{\Phi_{free}^T\Phi_{free}}{N})]^{-1}\frac{\sigma_{\xi_1}^2}{N} \\
& -[Plim(\frac{\Phi_{free}^T\Phi_{free}}{N})]^{-1}\psi_2\psi_2^T[Plim(\frac{\Phi_{free}^T\Phi_{free}}{N})]^{-1} \\
= \; & [Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}[\psi_1\frac{\sigma_{\xi_1}^2}{N}-\psi_2\psi_2^T][Plim(\frac{\Phi^T\Phi)}{N}-\psi_1]^{-1} \\
& +[Plim(\frac{\Phi^T\Phi}{N})-\psi_1]^{-1}\frac{\sigma_{\xi_1}^2}{N}
\end{aligned}
\tag{1}
$$

# Appendix B: Properties of Eqn (21)

From eqn (20)

$$
\begin{aligned}
Plim[E(\hat{g}_i-\overline{g_i})^2] &= Plim\left\{E\left(\frac{w_iY/N-\Psi_{2i}}{d_i+\lambda_i-\Psi_{1i}}-g_i+\frac{\lambda_i g_i}{d_i'+\lambda_i}\right)^2\right\} \\
&= Plim\left\{E\left(\frac{w_iY/N-\Psi_{2i}}{d_i+\lambda_i-\Psi_{1i}}\right)^2\right\}+\left(\frac{d_i' g_i}{d_i'+\lambda_i}\right)^2 \\
&\quad -2Plim\left\{E\left(\frac{w_iY/N-\Psi_{2i}}{d_i+\lambda_i-\Psi_{1i}}\right)\frac{d_i' g_i}{d_i'+\lambda_i}\right\} \\
&= \frac{Plim(w_iw_i^T/N)\sigma_{\xi_1}^2/N-\Psi_{2i}\Psi_{2i}^T+d_i'^2g_i^2}{(d_i'+\lambda_i)^2} \\
&\quad +\left(\frac{d_i' g_i}{d_i'+\lambda_i}\right)^2-2\left(\frac{d_i' g_i}{d_i'+\lambda_i}\right)^2 \\
&= \frac{Plim(w_iw_i^T/N)\sigma_{\xi_1}^2/N-\Psi_{2i}\Psi_{2i}^T}{(d_i'+\lambda_i)^2}
\end{aligned}
\tag{2}
$$

| numerator polynomial | RRME estimate | RME estimate | true value |
|---|---|---|---|
| $y(k-1)$ | 1.0205 | 1.0338 | 1 |
| $u(k-1)e(k-1)$ | 0.9019 | 0.9100 | 1 |
| denominator polynomial | RRME estimate | RME estimate | true value |
| $y^2(k-1)$ | 1.0659 | 1.1843 | 1 |

Table 1: Parameter estimates for Example 1

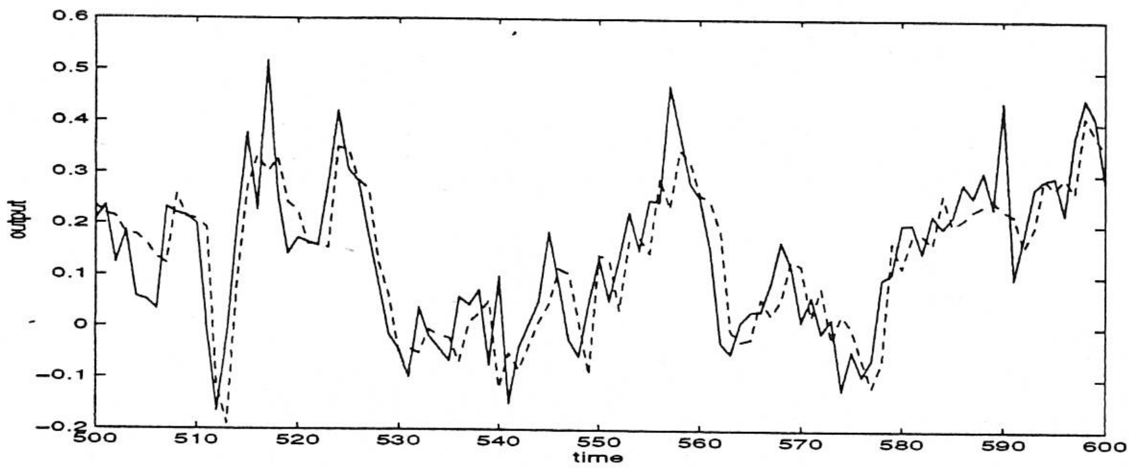


Figure 2: The output and one-step ahead prediction of Example 1 (RRME)
(dotted line –predicted output solid line–true output)

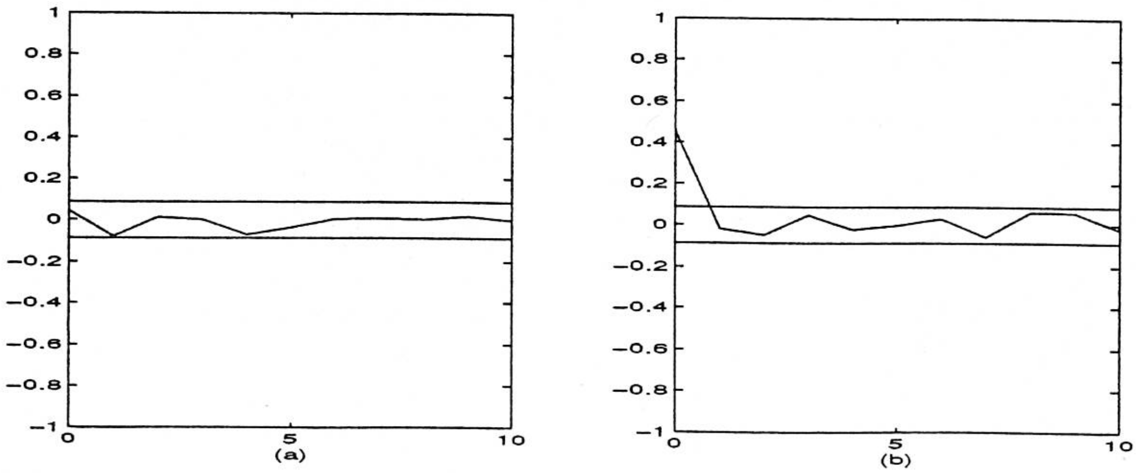


Figure 3: The model validity tests for Example 1
( a $-\Phi_{(y\xi)u^2}$ b$-\Phi_{(y\xi)\xi^2}$)

| numerator polynomial | RRME estimate | RME estimate | true value |
|---|---|---|---|
| $y(k-1)$ | 1.0248 | 1.0585 | 1 |
| $u(k-1)u(k-2)$ | 0.9706 | 0.9852 | 1 |
| $y(k-2)e(k-2)$ | 0.9010 | 1.1222 | 1 |
| denominator polynomial | RRME estimate | RME estimate | true value |
| $y^2(k-1)$ | 0.9965 | 1.2556 | 1 |
| $u^2(k-2)$ | 0.9903 | 0.9913 | 1 |

Table 2: Parameter estimates for Example 2

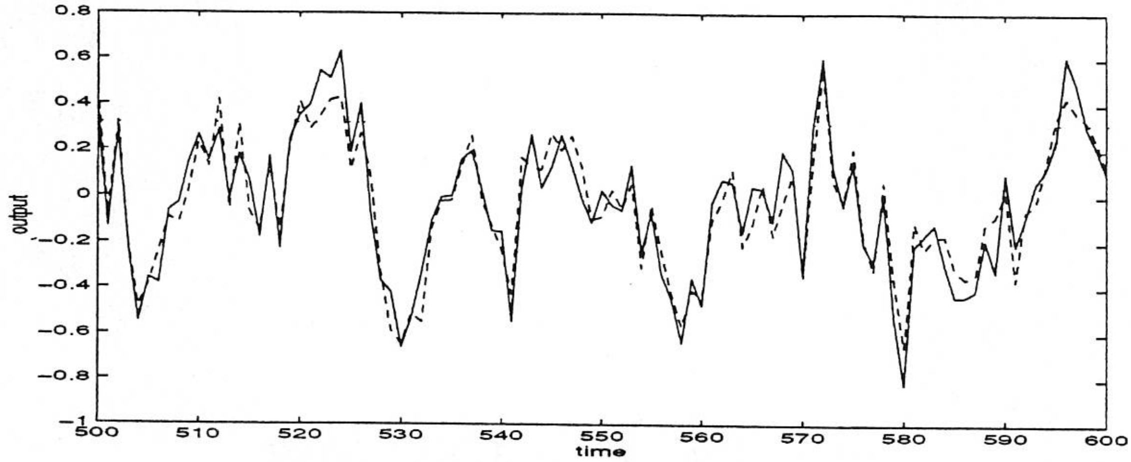


Figure 4: The output and one-step ahead prediction of Example 2 (RRME)
(dotted line –predicted output solid line–true output)

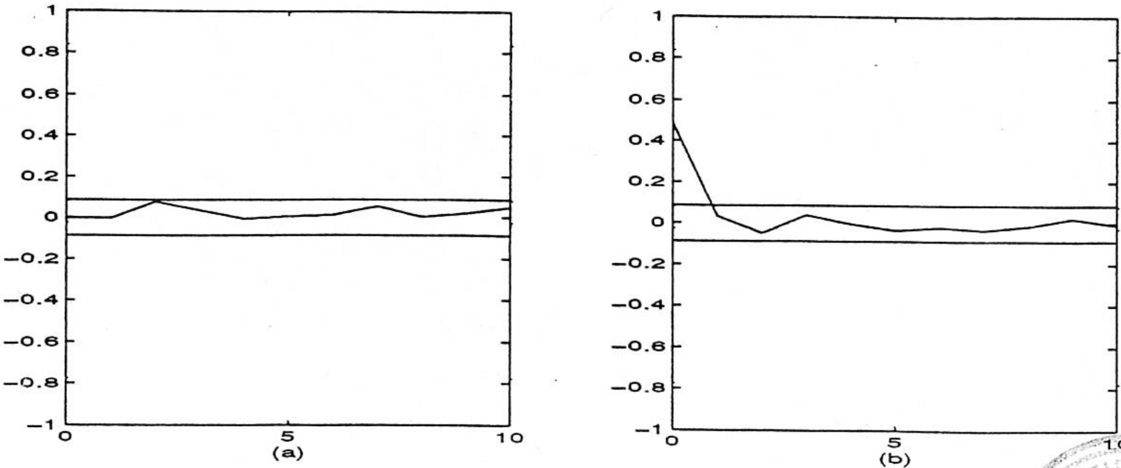


Figure 5: The model validity tests for Example 2
( a $-\Phi_{(y\xi)u^2}$ b$-\Phi_{(y\xi)\xi^2}$ )