15th Triennial World Congress, Barcelona, Spain

ON LEAST-SQUARES IDENTIFICATION OF ARMAX MODELS

Wei Xing Zheng [*,1]

[*] *School of Quantitative Methods and Mathematical Sciences*
University of Western Sydney
Penrith South DC NSW 1797, Australia

Abstract: In this paper the problem of least-squares (LS) identification of ARMAX models is investigated from a new point of view. An efficient scheme for estimating the noise-induced bias in the LS parameter is introduced by exploiting the unique structure of the ARMAX model and utilizing extra delayed outputs. Then a new type of LS based method is developed in combination with the bias correction technique. The proposed method makes no use of a prefilter and deals directly with the underlying ARMAX model. The important characteristics of the proposed method includes desired computational efficiency and superior estimation accuracy. The behavior of the proposed LS based method is also substantiated using simulation data while in comparison with other identification methods.

Keywords: System identification, Parameter estimation, Least-squares method, ARMAX models, Unbiased estimators.

1. INTRODUCTION

Systems identification has been an active research topic for more than two decades in such areas as control, signal processing and communications (Ljung, 1987; Söderström and Stoica, 1989). Parameter estimation via the least-squares (LS) criterion has become perhaps the most studied and implemented means of model identification for stochastic systems. The popularity of the standard LS method are usually attributed to the simple concept and the easy implementation. The major drawback, however, is that the LS parameter estimates are unbiased only in the special (and probably less practically meaningful) situation when the equation error of the underlying system model is white noise. As a result, many variants of the standard LS method have been developed to consistently identify system models frequently encountered in real-world applications. Among them, there are the prediction error (PE) methods (Ljung, 1987), the instrumental variable (IV) methods (Söderström and Stoica, 1989), and the bias correction methods (James, *et al.*, 1972; Sagara and Wada, 1977; Stoica and Söderström, 1982), to just mention a few. The PE methods are perhaps the most accurate and statistically efficient parametric algorithms, but they are involved with modelling the process noise and are highly computationally demanding. On the other hand, the IV methods can be implemented at a reduced numerical cost, with their consistency being irrespective of the noise dynamics. However, the choice of instruments may affect ultimate identification results substantially in certain situations, whereas simple and efficient methods for selecting appropriate instruments to attain some optimal properties are expected to be developed yet for the IV methods.

The bias correction methods are motivated by the intuition that once the noise-induced bias

[1] This work was supported in part by a Research Grant from the Australian Research Council and in part by a Research Grant from the University of Western Sydney, Australia.

in the LS parameter estimate is obtainable, an unbiased parameter estimate will result straightforwardly from eliminating the bias from the LS estimate. So how to estimate the noise-induced bias is the kernel of the bias correction methods. The prefiltering based bias-eliminated least-squares (PBELS) method (Feng and Zheng, 1991) is featured by the use of a prefilter to preprocess the system input to enable estimation of the noise-induced bias. It is equipped with the ability to cope with a wide range of noise models. The bias-compensating least-squares (BCLS) method (Zhang, *et al.*, 1997) makes use of a composite signal to estimate the noise-induced bias. The direct bias-eliminated least-squares (DBELS) method (Zheng, 1998) is aimed at improving the computational efficiency of the identification algorithm and maintaining desirable estimation accuracy but without employing a prefilter. The BELS based methods are also shown to be closely related to the weighted IV methods (Stoica, *et al.*, 1995; Zhang, *et al.*, 1997). In particular, the DBELS method is identical to the simple instrumental variable (SIV) method which uses delayed inputs as instruments (Söderström, *et al.*, 1999).

This paper is concerned with unbiased identification of autoregressive moving average models with an exogenous signal (termed as ARMAX model). ARMAX models are one type of dynamic models commonly used in representation of stochastic linear systems found in practice. The unique structure of ARMAX models presents an interesting and yet important feature. That is, the model noise (or equation error) is indeed colored but is finitely auto-correlated. This in turn implies that the system output can be made orthogonal to some subspace spanned by stochastic disturbances. It is shown in this paper that the mentioned important feature of ARMAX models may be manoeuvred to construct an efficient scheme of using extra delayed output values to estimate the noise-induced bias. This innovative estimation scheme is illustrated in the following. Firstly, an LS parameter estimate of the underlying ARMAX model is evaluated which is known to be biased. Secondly, an auxiliary linear regression model, which is equivalent to the underlying ARMAX model, is introduced. An estimate of the noise covariance vector, which specifies the source of the noise-induced bias in the LS estimate, is then derived from this auxiliary model in an innovative way of making use of extra delayed outputs. Thirdly, the unbiased parameter estimate is obtained via the bias-correction principle. The proposed method retains the merits of the previous BELS based methods. More importantly, it can achieve almost the same estimation accuracy as the PBELS method with a significantly reduced computational load; and it performs better than the DBELS method in terms of much low variance, particularly in the presence of high noise. This new BELS based method is also linked to the IV methods in certain way, and this connection largely assists the comprehension and investigation of the proposed method.

2. DESCRIPTION OF ARMAX MODELS

Let q^{-1} be the unit backward shift operator, and define the three polynomials in q^{-1} as

$$A(q^{-1}) = 1 - a_1 q^{-1} - \dots - a_n q^{-n} \quad (1a)$$

$$B(q^{-1}) = b_1 q^{-1} + \dots + b_m q^{-m} \quad (1b)$$

$$C(q^{-1}) = c_0 + c_1 q^{-1} + \dots + c_n q^{-n}, \quad c_0 = 1 \quad (1c)$$

Then the output $y(t)$ of a linear time-invariant discrete-time system in relation to its input $u(t)$ can be described by the following ARMAX model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})w(t) \quad (2)$$

where

$$e(t) = C(q^{-1})w(t) = \sum_{j=0}^{n} c_j w(t-j) \quad (3)$$

denotes the stochastic disturbance acting on the system, and $w(t)$ stands for the source of the disturbance.

Introduce the system parameter vector $\boldsymbol{\theta}$ and the data regression vector $\boldsymbol{\phi}_t$ respectively as

$$\boldsymbol{\theta}^\top = [\mathbf{a}^\top;\ \mathbf{b}^\top] = [a_1 \dots a_n;\ b_1 \dots b_m] \quad (4)$$

$$\boldsymbol{\phi}_t^\top = [\mathbf{y}_t^\top;\ \mathbf{u}_t^\top] = [y(t-1) \dots y(t-n);\ u(t-1) \dots u(t-m)] \quad (5)$$

The ARMAX model (2) can be further put in the following linear regression form

$$y(t) = \boldsymbol{\phi}_t^\top \boldsymbol{\theta} + e(t) \quad (6)$$

Several standard assumptions are made as follows:

(A1) The polynomial $A(q^{-1})$ is a Hurwitz polynomial, *i.e.* $A(.)$ has all zeros strictly outside the unit disc.
(A2) The input $u(t)$ is stationary and persistently exciting of a sufficient order.
(A3) The disturbance source $w(t)$ is white noise, and statistically uncorrelated with $u(t)$.
(A4) The models (n, m) are known.

By assumption (A3), the input $u(t)$ and the stochastic disturbance $e(t)$ are orthogonal to each other, which implies that the system under consideration operates in open loop. Moreover, another important feature of the ARMAX model (2) is that the output $y(t)$ is orthogonal to the subspace spanned by the stochastic disturbances $e(t+n+1), e(t+n+2), \dots$. This property is stated in Theorem 1.

Theorem 1. Define the cross-covariance function $r_{ye}(-k) = E[y(t)e(t+k)]$. Then

$$r_{ye}(-k) = 0, \quad k = n+1, n+2, ... \tag{7}$$

Proof: First, let $r_{yw}(-k) = E[y(t)w(t+k)]$. The assumption that $w(t)$ is a white noise sequence immediately leads to

$$r_{yw}(-k) = 0, \quad k = 1, 2, 3, ... \tag{8}$$

Equation (8) bears a clear physical interpretation in that the current plant output signal $y(t)$ is statistically uncorrelated with any future disturbance source $w(k)$, where $k > t$. Using (3), we have

$$r_{ye}(-k) = E[y(t)\sum_{j=0}^{n} c_j w(t-j+k)]$$
$$= \sum_{j=0}^{n} c_j E[y(t)w(t-j+k)] = \sum_{j=0}^{n} c_j r_{yw}(j-k). \tag{9}$$

By (8), it is easy to see that

$$r_{yw}(j-k) = 0 \tag{10}$$

for $j = 0, 1, ..., n$ and $k = n+1, n+2, ...$ Thus, (7) follows easily from (9) and (10). ∎

3. LEAST-SQUARES ESTIMATOR

Systems identification is aimed at estimating the system parameter vector $\boldsymbol{\theta}$ from the sampled input-output data $\{u(t), y(t)\}$. In particular, it is desired to obtain an unbiased estimate of $\boldsymbol{\theta}$. However, the popular LS method produces biased parameter estimates for ARMAX models, as will be illustrated below.

Consider the LS criterion $J(\boldsymbol{\theta}) = E[e(t)^2]$. As shown in Davis and Vinter (1985), minimization of $J(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ gives rise to the LS estimate $\boldsymbol{\theta}_{LS}$ as

$$\boldsymbol{\theta}_{LS} = \mathbf{R}_{\phi\phi}^{-1}\mathbf{R}_{\phi y} \tag{11}$$

where $\mathbf{R}_{\phi\phi} = E[\boldsymbol{\phi}_t\boldsymbol{\phi}_t^\top]$ and $\mathbf{R}_{\phi y} = E[\boldsymbol{\phi}_t y(t)]$.

Following assumptions (A1)-(A4), it can be shown that $\boldsymbol{\theta}_{LS}$ has the asymptotic expression

$$\boldsymbol{\theta}_{LS} = \boldsymbol{\theta} + \mathbf{R}_{\phi\phi}^{-1}\mathbf{R}_{\phi e} = \boldsymbol{\theta} + \mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye} \tag{12}$$

The second equality in (12) is due to

$$\mathbf{R}_{\phi e} = E[\boldsymbol{\phi}_t e(t)] = \begin{bmatrix}\mathbf{R}_{ye}\\ \mathbf{R}_{ue}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_n\\ \mathbf{0}\end{bmatrix}\mathbf{R}_{ye} = \mathbf{D}\mathbf{R}_{ye} \tag{13}$$

$$\mathbf{R}_{ye}^\top = [r_{ye}(-1) \; ... \; r_{ye}(-n)] \tag{14}$$

where $\mathbf{D}^\top = [\mathbf{I}_n \;\; \mathbf{0}] \in \mathbb{R}^{n\times(n+m)}$, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. Since the $n \times 1$ noise covariance vector $\mathbf{R}_{ye}$ is not a zero vector, (12) indicates that the LS estimate $\boldsymbol{\theta}_{LS}$ is bound to be biased, with $\mathbf{R}_{ye}$ specifying the source of the noise-induced bias $\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye}$. Thus, in general the standard LS method is not workable for unbiased identification of ARMAX models.

4. NEW BIAS CORRECTION SCHEME

In this section, we will develop a new bias correction scheme so as to arrive at an unbiased parameter estimate of $\boldsymbol{\theta}$. By means of the well-known bias correction principle (James, *et al.*, 1972), a bias-eliminated least-squares (BELS) estimate of the system parameter vector $\boldsymbol{\theta}$ may be acquired from (12) as

$$\boldsymbol{\theta}_{BELS} = \boldsymbol{\theta}_{LS} - \mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye} \tag{15}$$

on the condition that an estimate of the noise covariance vector $\mathbf{R}_{ye}$ is attainable in some manner. In order to make the bias correction scheme (15) practically implementable, our focus is now on estimation of $\mathbf{R}_{ye}$.

For this purpose, we introduce a $(2n+m) \times 1$ auxiliary parameter vector $\boldsymbol{\Theta}$ and a $(2n+m) \times 1$ auxiliary regression vector $\boldsymbol{\Phi}_t$ as follows:

$$\boldsymbol{\Theta} = \begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\alpha}\end{bmatrix}, \quad \boldsymbol{\Phi}_t = \begin{bmatrix}\boldsymbol{\phi}_t\\ \boldsymbol{\rho}_t\end{bmatrix} \tag{16}$$

where $\boldsymbol{\rho}_t^\top = [y(t-n-1) \; ... \; y(t-2n)]$ and $\boldsymbol{\alpha}^\top = [0 \; ... \; 0] \in \mathbb{R}^n$. With this, it is easy to verify that $\boldsymbol{\Phi}_t^\top\boldsymbol{\Theta} = \boldsymbol{\phi}_t^\top\boldsymbol{\theta}$. Thus the underlying linear regression model (6) can be equivalently represented by an auxiliary linear regression model

$$y(t) = \boldsymbol{\Phi}_t^\top\boldsymbol{\Theta} + e(t) \tag{17}$$

By a similar procedure to (11), the LS estimate of the auxiliary parameter vector $\boldsymbol{\Theta}$ is found to be

$$\boldsymbol{\Theta}_{LS} = \mathbf{R}_{\Phi\Phi}^{-1}\mathbf{R}_{\Phi y} \tag{18}$$

where $\mathbf{R}_{\Phi\Phi} = E[\boldsymbol{\Phi}_t\boldsymbol{\Phi}_t^\top]$ and $\mathbf{R}_{\Phi y} = E[\boldsymbol{\Phi}_t y(t)]$. Moreover, using assumptions (A1)-(A4), $\boldsymbol{\Theta}_{LS}$ is expressible asymptotically as

$$\boldsymbol{\Theta}_{LS} = \boldsymbol{\Theta} + \mathbf{R}_{\Phi\Phi}^{-1}\mathbf{R}_{\Phi e} \tag{19}$$

where

$$\mathbf{R}_{\Phi e} = E[\boldsymbol{\Phi}_t e(t)] = \begin{bmatrix}\mathbf{R}_{\phi e}\\ \mathbf{R}_{\rho e}\end{bmatrix} = \begin{bmatrix}\mathbf{D}\mathbf{R}_{ye}\\ \mathbf{R}_{\rho e}\end{bmatrix} \tag{20}$$

$$\mathbf{R}_{\rho e}^\top = [r_{ye}(-n-1) \; ... \; r_{ye}(-2n)] \tag{21}$$

Note that (19) is analogous to (12), while it follows from (7) that every element of $\mathbf{R}_{\rho e}$ is zero, yielding $\mathbf{R}_{\rho e} = \mathbf{0}$. So the $(2n+m) \times 1$ noise covariance vector $\mathbf{R}_{\Phi e}$ may be simply expressed as

$$\mathbf{R}_{\Phi e} = \begin{bmatrix}\mathbf{D}\mathbf{R}_{ye}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_n\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}\mathbf{R}_{ye} = \boldsymbol{\mathcal{D}}\mathbf{R}_{ye} \tag{22}$$

where $\boldsymbol{\mathcal{D}}^\top = [\mathbf{I}_n \;\; \mathbf{0}] \in \mathbb{R}^{n\times(2n+m)}$. It is seen from (19) and (22) that, similarly to (12), the same $n \times 1$ noise covariance vector $\mathbf{R}_{ye}$ also specifies the source of the noise-induced bias in $\boldsymbol{\Theta}_{LS}$.

As mentioned above, our aim is to acquire an estimate of the noise covariance vector $\mathbf{R}_{ye}$. The following theorem shows that this can now be achieved by finding the LS estimate of the intermediate parameter vector $\boldsymbol{\alpha}$ out of $\boldsymbol{\Theta}_{LS}$.

Theorem 2. Let $\boldsymbol{R}_{\phi\rho}=E[\boldsymbol{\phi}_t\boldsymbol{\rho}_t^\top]$, $\boldsymbol{R}_{\rho y}=E[\boldsymbol{\rho}_t y(t)]$. We have the following matrix-vector equation with regard to the noise covariance vector $\mathbf{R}_{ye}$:

$$\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye}=\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta}_{LS}-\mathbf{R}_{\rho y} \qquad (23)$$

Proof: By virtue of $\mathbf{R}_{\phi\rho}$ and $\mathbf{R}_{\rho y}$, the covariance matrix $\mathbf{R}_{\Phi\Phi}$ and the covariance vector $\mathbf{R}_{\Phi y}$ are expressible as

$$\mathbf{R}_{\Phi\Phi}=\begin{bmatrix}\mathbf{R}_{\phi\phi} & \mathbf{R}_{\phi\rho}\\ \mathbf{R}_{\phi\rho}^\top & \mathbf{R}_{\rho\rho}\end{bmatrix},\quad \mathbf{R}_{\Phi y}=\begin{bmatrix}\mathbf{R}_{\phi y}\\ \mathbf{R}_{\rho y}\end{bmatrix} \qquad (24)$$

where $\mathbf{R}_{\rho\rho}=E[\boldsymbol{\rho}_t\boldsymbol{\rho}_t^\top]$. Applying the matrix inversion formula (see *e.g.* Davis and Vinter, 1994) to $\mathbf{R}_{\Phi\Phi}$ gives rise to

$$\mathbf{R}_{\Phi\Phi}^{-1}=\begin{bmatrix}\mathbf{R}_{11} & -\mathbf{R}_{\phi\phi}^{-1}\mathbf{R}_{\phi\rho}\boldsymbol{\Delta}^{-1}\\ -\boldsymbol{\Delta}^{-1}\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1} & \boldsymbol{\Delta}^{-1}\end{bmatrix} \qquad (25)$$

where $\mathbf{R}_{11}=\mathbf{R}_{\phi\phi}^{-1}+\mathbf{R}_{\phi\phi}^{-1}\mathbf{R}_{\phi\rho}\boldsymbol{\Delta}^{-1}\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}$ and $\boldsymbol{\Delta}=\mathbf{R}_{\rho\rho}-\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{R}_{\phi\rho}$. Let the last n elements in $\boldsymbol{\Theta}_{LS}$ be denoted by $\boldsymbol{\alpha}_{LS}$. Using (24) and (25) in (18), we get

$$\boldsymbol{\alpha}_{LS}=-\boldsymbol{\Delta}^{-1}\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{R}_{\phi y}+\boldsymbol{\Delta}^{-1}\mathbf{R}_{\rho y} \qquad (26)$$

Combining (19), (22) and (25) together and noticing that $\boldsymbol{\alpha}=\mathbf{0}$, we have

$$\boldsymbol{\alpha}_{LS}=-\boldsymbol{\Delta}^{-1}\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye} \qquad (27)$$

Finally, (23) follows immediately from equating (26) and (27) with $\boldsymbol{\alpha}_{LS}$ and premultiplying the resulting equation with $\boldsymbol{\Delta}$. ■

By Theorem 2, an unbiased estimate of the noise covariance vector $\mathbf{R}_{ye}$ may be obtained from (23) as

$$\hat{\mathbf{R}}_{ye}=(\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D})^{-1}(\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta}_{LS}-\mathbf{R}_{\rho y}) \qquad (28)$$

Hence, for unbiased identification of ARMAX models, the following new bias correction scheme can be implemented, which is called the BELSX method for short.

The BELSX Method

Step 1. From the sampled input-output data $\{u(t),y(t),1\le t\le N\}$, calculate the estimates of the covariance matrices and vectors:

$$\hat{\mathbf{R}}_{\phi\phi}(N)=\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\phi}_t\boldsymbol{\phi}_t^\top,\ \ \hat{\mathbf{R}}_{\phi\rho}(N)=\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\phi}_t\boldsymbol{\rho}_t^\top \qquad (29)$$

$$\hat{\mathbf{R}}_{\phi y}(N)=\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\phi}_t y(t),\ \ \hat{\mathbf{R}}_{\rho y}(N)=\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\rho}_t y(t) \qquad (30)$$

Step 2. Evaluate the LS estimate of the system parameter vector $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{LS}(N)=\hat{\mathbf{R}}_{\phi\phi}^{-1}(N)\hat{\mathbf{R}}_{\phi y}(N) \qquad (31)$$

Step 3. Compute the estimate of the noise covariance $\mathbf{R}_{ye}$:

$$\hat{\mathbf{R}}_{ye}(N)=(\hat{\mathbf{R}}_{\phi\rho}^\top(N)\hat{\mathbf{R}}_{\phi\phi}^{-1}(N)\mathbf{D})^{-1}(\hat{\mathbf{R}}_{\phi\rho}^\top(N)\hat{\boldsymbol{\theta}}_{LS}(N)-\hat{\mathbf{R}}_{\rho y}(N)) \qquad (32)$$

Step 4. Calculate the BELS estimate of the system parameter vector $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{BELS}(N)=\hat{\boldsymbol{\theta}}_{LS}(N)-\hat{\mathbf{R}}_{\phi\phi}^{-1}\mathbf{D}\hat{\mathbf{R}}_{ye}(N) \qquad (33)$$

5. FURTHER ANALYSIS

First of all, the estimation consistency of the proposed BELSX method is studied.

Theorem 3. Consider the BELSX method applied to the ARMAX model (2) and subject to assumptions (A1)-(A4). We may conclude parameter consistent convergence for $\hat{\boldsymbol{\theta}}_{BELS}(N)$, namely

$$\lim_{N\to\infty}\hat{\boldsymbol{\theta}}_{BELS}(N)=\boldsymbol{\theta}\quad \text{w.p.1} \qquad (34)$$

Proof: Premultiplying (6) with $\boldsymbol{\rho}_t$ and taking the mathematical expectation gives

$$\mathbf{R}_{\rho y}=\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta}+\mathbf{R}_{\rho e} \qquad (35)$$

Since by Theorem 1 $\mathbf{R}_{\rho e}=\mathbf{0}$, (35) further reduces to

$$\mathbf{R}_{\rho y}=\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta} \qquad (36)$$

Letting $N\to\infty$ in (32) and noticing that the sample covariance estimates given by (29) and (30) converge to its respective true covariances, we have

$$\begin{aligned}\lim_{N\to\infty}\hat{\mathbf{R}}_{ye}(N)&=(\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D})^{-1}(\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta}_{LS}-\mathbf{R}_{\rho y})\\ &=(\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D})^{-1}(\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta}_{LS}-\mathbf{R}_{\phi\rho}^\top\boldsymbol{\theta})\\ &=(\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D})^{-1}\mathbf{R}_{\phi\rho}^\top(\boldsymbol{\theta}_{LS}-\boldsymbol{\theta})\\ &=(\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D})^{-1}\mathbf{R}_{\phi\rho}^\top\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye}\\ &=\mathbf{R}_{ye}\qquad \text{w.p.1}\end{aligned} \qquad (37)$$

where (12) and (36) are utilized in deriving (37). Finally, letting $N\to\infty$ in (33) and using (12), we obtain

$$\begin{aligned}\lim_{N\to\infty}\hat{\boldsymbol{\theta}}_{BELS}(N)&=\boldsymbol{\theta}_{LS}-\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye}\\ &=(\boldsymbol{\theta}+\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye})-\mathbf{R}_{\phi\phi}^{-1}\mathbf{D}\mathbf{R}_{ye}\\ &=\boldsymbol{\theta}\qquad \text{w.p.1}\end{aligned} \qquad (38)$$

which completes the proof. ■

The proposed BELSX method may also be implemented recursively in on-line identification of ARMAX models. This is because the LS estimate $\hat{\boldsymbol{\theta}}_{LS}(N)$ can be readily calculated in a recursive manner (Davis and Vinter, 1985).

In comparison with other identification methods, the BELSX method retains the desirable features of the existing BELS methods (Feng and Zheng, 1991; Zheng, 1998), such as good estimation accuracy, implementation convenience with low computational complexity, desired robustness against noise, and so on.

Similarly to the DBELS method, the proposed BELSX method is superior to the PBELS method computationally due to its attractive algorithmic structure. Unlike the PBELS method, the BELSX method does not involve design of a prefilter or prefiltering of the sampled input-output data in its implementation. Instead, the sampled data are utilized directly in parameter estimation by the BELSX method. Moreover, no parameter extraction is needed for the BELSX method as the original system parameter vector $\boldsymbol{\theta}$ is estimated in a direct manner. These algorithmic advantages make the BELSX method very appealing in implementation in terms of computational load involved.

It is interesting to note that like the DBELS method, the proposed BELSX method can also be interpreted as a kind of the IV methods. Hence, it will inherit attractive properties of the IV methods as described in Söderström and Stoica (1989). In particular, an unbiased estimate with the minimum asymptotic covariance matrix can be attained with the proposed method. The relation of the BELSX method to the IV methods is studied in the following.

Theorem 4. Introduce $\boldsymbol{\zeta}_t^\top = [\boldsymbol{\rho}_t^\top;\ \mathbf{u}_t^\top]$. Then
(i) $\text{rank}(\mathbf{R}_{\zeta\phi}) = n + m$, where $\mathbf{R}_{\zeta\phi} = E[\boldsymbol{\zeta}_t \boldsymbol{\phi}_t^\top]$.
(ii) $\mathbf{R}_{\zeta e} = E[\boldsymbol{\zeta}_t e(t)] = \mathbf{0}$.

Proof: The proof is rather straightforward. First, $\text{rank}(\mathbf{R}_{\zeta\phi}) = n + m$ is assured by assumptions (A1)-(A2). Second, $\boldsymbol{R}_{\zeta e} = \mathbf{0}$ follows easily from assumption (A3) and Theorem 1. ■

By Theorem 4, $\boldsymbol{\zeta}_t$ is a qualified instrumental variable (Söderström and Stoica, 1989). So an IV estimate of the system parameter vector $\boldsymbol{\theta}$ may be obtained by

$$\hat{\boldsymbol{\theta}}_{IV}(N) = \hat{\mathbf{R}}_{\zeta\phi}^{-1}(N)\hat{\mathbf{R}}_{\zeta y}(N) \tag{39}$$

where $\hat{\mathbf{R}}_{\zeta\phi}(N) = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\zeta}_t \boldsymbol{\phi}_t^\top$ and $\hat{\mathbf{R}}_{\zeta y}(N) = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\zeta}_t y(t)$.

Theorem 5. The BELS parameter estimate $\hat{\boldsymbol{\theta}}_{BELS}(N)$ given by (33) is equal to the IV parameter estimate $\hat{\boldsymbol{\theta}}_{IV}(N)$ given by (39), namely

$$\hat{\boldsymbol{\theta}}_{BELS}(N) = \hat{\boldsymbol{\theta}}_{IV}(N) \tag{40}$$

Proof: Let

$$\hat{\mathbf{R}}_{\phi\rho}^\top(N)\hat{\mathbf{R}}_{\phi\phi}^{-1}(N) = \mathbf{F} = [\mathbf{F}_1 \quad \mathbf{F}_2] \tag{41}$$

where $\mathbf{F}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{F}_2 \in \mathbb{R}^{n\times m}$. Using (31) and (41), (32) may be rewritten as

$$\hat{\mathbf{R}}_{ye}(N) = \mathbf{F}_1^{-1}(\mathbf{F}\hat{\mathbf{R}}_{\phi y}(N) - \hat{\mathbf{R}}_{\rho y}(N)) \tag{42}$$

where it is assumed that $\mathbf{F}_1$ is of full rank. Substitution of (31) and (42) into (33) yields

$$\hat{\boldsymbol{\theta}}_{BELS}(N) = \hat{\mathbf{R}}_{\phi\phi}^{-1}(N)\mathbf{M}\begin{bmatrix}\hat{\mathbf{R}}_{\phi y}(N)\\ \hat{\mathbf{R}}_{\rho y}(N)\end{bmatrix} \tag{43}$$

where $\mathbf{M} = [\mathbf{I}_{n+m} - \mathbf{D}\mathbf{F}_1^{-1}\mathbf{F} \quad \mathbf{D}\mathbf{F}_1^{-1}]$. But

$$\mathbf{D}\mathbf{F}_1^{-1}\mathbf{F} = \begin{bmatrix}\mathbf{I}_n\\ \mathbf{0}\end{bmatrix}\mathbf{F}_1^{-1}[\mathbf{F}_1 \quad \mathbf{F}_2] = \begin{bmatrix}\mathbf{I}_n & \mathbf{F}_1^{-1}\mathbf{F}_2\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{44}$$

$$\mathbf{D}\mathbf{F}_1^{-1} = \begin{bmatrix}\mathbf{I}_n\\ \mathbf{0}\end{bmatrix}\mathbf{F}_1^{-1} = \begin{bmatrix}\mathbf{F}_1^{-1}\\ \mathbf{0}\end{bmatrix} \tag{45}$$

So $\mathbf{M}$ is expressible as

$$\mathbf{M} = \begin{bmatrix}\mathbf{F}_1 & \mathbf{F}_2\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{I}_n\\ \mathbf{0} & \mathbf{I}_m & \mathbf{0}\end{bmatrix} \tag{46}$$

Let $\mathbf{H} = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{I}_n\\ \mathbf{0} & \mathbf{I}_m & \mathbf{0}\end{bmatrix}$. Since

$$\begin{bmatrix}\mathbf{F}_1 & \mathbf{F}_2\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{I}_n\\ \mathbf{0} & \mathbf{I}_m & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\\ \mathbf{F}_1 & \mathbf{F}_2\end{bmatrix} = \mathbf{H}\begin{bmatrix}\mathbf{I}_{n+m}\\ \mathbf{F}\end{bmatrix} \tag{47}$$

$\mathbf{M}$ may be further written as

$$\mathbf{M} = \left(\mathbf{H}\begin{bmatrix}\mathbf{I}_{n+m}\\ \mathbf{F}\end{bmatrix}\right)^{-1}\mathbf{H} \tag{48}$$

It is easy to verify that

$$\mathbf{H}\begin{bmatrix}\hat{\mathbf{R}}_{\phi y}(N)\\ \hat{\mathbf{R}}_{\rho y}(N)\end{bmatrix} = \hat{\mathbf{R}}_{\zeta y}(N) \tag{49}$$

$$\mathbf{H}\begin{bmatrix}\mathbf{I}_{n+m}\\ \mathbf{F}\end{bmatrix}\hat{\mathbf{R}}_{\phi\phi}(N) = \mathbf{H}\begin{bmatrix}\hat{\mathbf{R}}_{\phi\phi}(N)\\ \hat{\mathbf{R}}_{\phi\rho}^\top(N)\end{bmatrix} = \hat{\mathbf{R}}_{\zeta\phi}(N) \tag{50}$$

Substituting (48), (49) and (50) into (43) immediately gives rise to

$$\hat{\boldsymbol{\theta}}_{BELS}(N) = \hat{\mathbf{R}}_{\zeta\phi}^{-1}(N)\hat{\mathbf{R}}_{\zeta y}(N) = \hat{\boldsymbol{\theta}}_{IV}(N) \tag{51}$$

Thus, the proof is complete. ■

6. SIMULATION ILLUSTRATIONS

Stochastic simulation results are now presented to validate the theoretical analysis made about the proposed BELSX method. Consider a second-order ARMAX model described by (2) and with

$$A(q^{-1}) = 1 - 1.5q^{-1} + 0.7q^{-2} \tag{52a}$$

$$B(q^{-1}) = 1.0q^{-1} + 0.5q^{-2} \tag{52b}$$

$$C(q^{-1}) = 1 - 1.0q^{-1} + 0.2q^{-2} \tag{52c}$$

The input $u(t)$ is taken as white noise with unit variance. The system has a signal-to-noise ratio (SNR) of around 2dB at its output, which implies the presence of very high noise.

The six methods, namely, the standard LS method, the PE (ARMAX) method, the optimal IV (OIV) method, the standard IV (SIV) method, the PBELS method and the BELSX method, are employed to identify this ARMAX model from $N = 2000$ sampled input-output data. In particular, delayed inputs are used as instruments with the SIV method; the MATLAB codes `armax` and `iv4` are used to implement the PE (ARMAX) method and the OIV method, respectively; and a prefilter designed as $F(q^{-1}) = 1/(1 - 1.7q^{-1} + 0.72q^{-2})$ is used with the PBELS method. A total of $M = 1000$ stochastic trials have been conducted to compute the arithmetic means and standard deviations of the various estimates, which are displayed in Table 1. Further, the overall behavior of these methods is measured in terms of the

Table 1. Comparative Identification Results
(N=2000, SNR $\approx$ 2dB, 1000 stochastic trials, NFPT = Number of `flops` per trial)

method	a_1	a_2	b_1	b_2	Gain	RE	RMSE	NFPT
LS	0.8242 $\pm$0.0252	$-$0.1370 $\pm$0.0222	0.9997 $\pm$0.0745	1.1758 $\pm$0.0693	6.9594 $\pm$0.3650	55.53%	55.78%	56300
PE(ARMAX)	1.4990 $\pm$0.0167	$-$0.6995 $\pm$0.0142	0.9990 $\pm$0.0646	0.5039 $\pm$0.0790	7.4989 $\pm$0.2137	0.21%	5.23%	3251138
OIV	1.4973 $\pm$0.0211	$-$0.6979 $\pm$0.0179	0.9980 $\pm$0.0656	0.5082 $\pm$0.0857	7.5103 $\pm$0.2316	0.45%	5.59%	720205
SIV	1.5163 $\pm$0.1549	$-$0.7245 $\pm$0.2178	0.9979 $\pm$0.0925	0.4835 $\pm$0.2130	8.3956 $\pm$6.6335	1.69%	17.81%	88311
PBELS	1.5002 $\pm$0.0333	$-$0.7001 $\pm$0.0278	0.9971 $\pm$0.0891	0.4997 $\pm$0.1087	7.4936 $\pm$0.3630	0.14%	7.36%	121706
BELSX	1.5015 $\pm$0.0814	$-$0.7011 $\pm$0.0472	0.9974 $\pm$0.0894	0.4989 $\pm$0.1182	7.6963 $\pm$1.3018	0.17%	8.78%	96455
true value	1.5000	$-$0.7000	1.0000	0.5000	7.5000			

MATLAB code `flops`, the relative error (RE) and the normalized root mean squared error (RMSE). The RE and RMSE are defined respectively as

$$\text{RE}=\frac{\|\mathbf{m}(\hat{\boldsymbol{\theta}})-\boldsymbol{\theta}\|}{\|\boldsymbol{\theta}\|},\quad \text{RMSE}=\sqrt{\frac{1}{M}\sum_{k=1}^{M}\frac{\|\hat{\boldsymbol{\theta}}_k-\boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^2}}$$

where $\mathbf{m}(\hat{\boldsymbol{\theta}}) = \frac{1}{M}\sum_{k=1}^{M}\hat{\boldsymbol{\theta}}_k$, and $\hat{\boldsymbol{\theta}}_k$ denotes a parameter estimator in the kth stochastic trial over a total of M trials.

It is observed from Table 1 that in spite of their favorable estimation accuracy, the PE (ARMAX) method and the OIV method involve the highest numerical cost among the six methods. On the contrary, the two BELS based methods both demonstrate the superior performance, including their low computational complexity and estimation unbiasedness for ARMAX models. In agreement with the preceding analysis, while achieving the almost same estimation accuracy, the proposed BELSX method exhibits obvious computational advantages over the PBELS method because of no use of a prefilter. It is interesting to note that with considerably smaller values of the RE and the RMSE, the BELSX method is much more accurate than the SIV method although the latter requires fewer computations (about 8% less) than the former.

7. CONCLUSIONS

The novelty of the present work in comparison with the previous work is in that it has shown how the noise-induced bias can be estimated by taking advantage of the ARMAX structure and making use of extra delayed outputs. On the basis of this, the BELSX method has been established to perform unbiased identification of ARMAX models. Computer simulations have indicated that if one wants to achieve a high estimation accuracy at a moderate numerical cost, the proposed BELSX method will be a much preferable algorithm for use because its accuracy is very close to that of the PE (ARMAX) method while its computational load is only slightly more than that of the SIV method. Finally, we note that the invertibility of the matrix $\hat{\mathbf{R}}_{\phi\rho}^{\top}(N)\hat{\mathbf{R}}_{\phi\phi}^{-1}(N)\mathbf{D}$ in (32) is yet to be proven. Another interesting issue is to analyze the impact of the bias on the results stemmed from the Monte-Carlo experimentations via the frequency study. These are the topics of further investigations.

REFERENCES

Davis, M. H. A. and R. B. Vinter (1985). *Stochastic Modelling and Control.* Chapman and Hall.

Feng, C. B. and W. X. Zheng (1991). Robust identification of stochastic linear systems with correlated noise. *IEE Proc.-Control Theory and Applications*, **138**, 484-492.

James, P. N., P. Souter and D. C. Dixon (1972). Suboptimal estimation of the parameters of discrete systems in the presence of correlated noise. *Electronics Letters*, **8**, 411-412.

Ljung, L. (1987). *System Identification: Theory for the User.* Prentice-Hall.

Sagara, S. and K. Wada (1977). On-line modified least-squares parameters estimation of linear discrete dynamic systems. *Int. J. Control*, **25**, 329-343.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice-Hall.

Söderström, T., W. X. Zheng and P. Stoica (1999). Comments on 'On a least-squares based algorithm for identification of stochastic linear systems'. *IEEE Trans. Signal Processing*, **47**, 1395-1396.

Stoica, P. and T. Söderström (1982). Bias correction in least-squares identification. *Int. J. Control*, **35**, 449-457.

Stoica, P., T. Söderström and V. Simonyte (1995). Study of a bias-free least squares parameter estimator. *IEE Proc.-Control Theory and Applications*, **142**, 1-6.

Zhang, Y., T. T. Lie and C. B. Soh (1997). Consistent parameter estimation of systems disturbed by correlated noise. *IEE Proc.-Control Theory and Applications*, **144**, 40-44.

Zheng, W. X. (1998). On a least-squares based algorithm for identification of stochastic linear systems. *IEEE Trans. Signal Processing*, **46**, 1631-1638.