Unbiased Identification of Stochastic Linear Systems from Noisy Input and Output Measurements

Wei Xing Zheng
School of Science, University of Western Sydney, Nepean
Kingswood, Sydney, NSW 2747, Australia
w.zheng@uws.edu.au

Abstract

This paper is concerned with identification of stochastic linear systems from noisy input and output measurements. A modified scheme that employs extra delayed noisy measurements is derived to estimate the variances of white input and output noises. These estimated noise variances are then applied for removal of the bias from a least-squares parameter estimate via an iterative procedure to achieve estimation consistency. The new identification algorithm incorporated with this modified estimation scheme for the noise variances demonstrates greatly improved performances. Compared with the previously developed method, the new identification algorithm can converge at a much faster rate and produce much more accurate parameter estimates at only a slightly increased numerical cost. The theoretical predictions are confirmed through Monte-Carlo stochastic simulation studies.

1 Introduction

Consider the following linear time-invariant system:

$$x(t) = G(q^{-1})r(t) \quad (1)$$
$$u(t) = r(t) + v(t) \quad (2)$$
$$y(t) = x(t) + w(t) \quad (3)$$

where

$r(t)$	true input signal
$x(t)$	true output signal
$v(t)$	additive white noise with variance σ_v^2
$w(t)$	additive white noise with variance σ_w^2
$u(t)$	measured noisy input
$y(t)$	measured noisy output
$G(q^{-1})$	system transfer function defined by

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{b_1 q^{-1} + \dots + b_m q^{-m}}{1 - a_1 q^{-1} - \dots - a_n q^{-n}}. \quad (4)$$

The model given by (1)–(4) takes into account the output noise $w(t)$ as well as the input noise $v(t)$ to describe the underlying dynamic system, and is known as the dynamic errors-in-variables (DEV) model. It can be recast in linear regression form:

$$y(t) = \boldsymbol{\psi}_t^\top \boldsymbol{\theta} + \epsilon(t) \quad (5)$$

where

$$\boldsymbol{\theta}^\top = [\boldsymbol{a}^\top;\ \boldsymbol{b}^\top] = [a_1\ \dots\ a_n;\ b_1\ \dots\ b_m]. \quad (6)$$

$$\boldsymbol{\psi}_t^\top = [\boldsymbol{y}_t^\top; \boldsymbol{u}_t^\top] = [y(t-1) \dots y(t-n); u(t-1) \dots u(t-m)]. \quad (7)$$

The equation error $\epsilon(t)$ in (5) is given by

$$\epsilon(t) = w(t) - \boldsymbol{\epsilon}_t^\top \boldsymbol{\theta} \quad (8)$$

where

$$\boldsymbol{\epsilon}_t^\top = [\boldsymbol{w}_t^\top; \mathbf{v}_t^\top] = [w(t-1) \dots w(t-n); v(t-1) \dots v(t-m)]. \quad (9)$$

This DEV model has a wide range of important applications in control systems, signal processing, communications, and econometrics [5], [9], [7].

DEV model identification represents a special type of system identification problem. Anderson, Deistler and the others have made a thorough investigation of identifiability and structure theory of linear DEV models [1], [3], [8]. In estimating the system parameter vector $\boldsymbol{\theta}$ from noisy input-output measurements $\{u(t), y(t)\}$, the well-known least-squares (LS) method (see e.g. [2]) is defined by minimizing the mean squared error

$$J(\boldsymbol{\theta}) = E[\epsilon(t)^2] \quad (10)$$

which gives rise to

$$\boldsymbol{\theta}_{LS} = \boldsymbol{R}_{\psi\psi}^{-1} \boldsymbol{R}_{\psi y} \quad (11)$$

where

$$\boldsymbol{R}_{\psi\psi} = E[\boldsymbol{\psi}_t \boldsymbol{\psi}_t^\top], \quad \boldsymbol{R}_{\psi y} = E[\boldsymbol{\psi}_t y(t)]. \quad (12)$$

However, the equation error $\epsilon(t)$ is no longer white noise even if both $v(t)$ and $w(t)$ are now white noises. This implies that there is no possibility for $\boldsymbol{\theta}_{LS}$ to be a consistent estimator of $\boldsymbol{\theta}$. In fact, as shown in [12], the following asymptotic expression holds for $\boldsymbol{\theta}_{LS}$:

$$\boldsymbol{\theta}_{LS} = \boldsymbol{\theta} + \Delta\boldsymbol{\theta} \quad (13)$$

where $\Delta\boldsymbol{\theta}$ is the noise-induced bias given by

$$\Delta\boldsymbol{\theta} = -\boldsymbol{R}_{\psi\psi}^{-1} \boldsymbol{D} \boldsymbol{\theta} \quad (14)$$

with

$$\boldsymbol{D} = \text{diag}\{\sigma_w^2 \boldsymbol{I}_n;\ \sigma_v^2 \boldsymbol{I}_m\} \quad (15)$$

and $\boldsymbol{I}_n$ being an $n \times n$ identity matrix.

Considerable efforts have been made to implement unbiased parameter estimation of linear DEV models, leading to a number of parametric algorithms. These include the Koopmans-Levin (KL) method [4], the joint-output (JO) method [10], the logarithmic least-squares frequency-domain (LLS-FD) method [6], the combined instrumental variable and weighted subspace fitting (IV-WSF) method [11], and the bias-eliminated least-squares (BELS) methods [12], [13]. In particular, the BELS methods are developed by the way of directly estimating the two noise variances σ_v^2 and σ_w^2, which, as shown in (14), decide the noise-induced bias $\Delta\boldsymbol{\theta}$. Denoting estimates of the noise variances by $\hat{\sigma}_v^2(k)$ and $\hat{\sigma}_w^2(k)$ and forming the diagonal matrix

$$\hat{\boldsymbol{D}}(k) = \mathrm{diag}\{\hat{\sigma}_w^2(k)\boldsymbol{I}_n;\ \hat{\sigma}_v^2(k)\boldsymbol{I}_m\} \tag{16}$$

an unbiased estimate of the system parameter vector $\boldsymbol{\theta}$ is then obtained iteratively via the bias correction procedure:

$$\hat{\boldsymbol{\theta}}_{BELS}(k) = \boldsymbol{\theta}_{LS} + \boldsymbol{R}_{\psi\psi}^{-1}\hat{\boldsymbol{D}}(k)\hat{\boldsymbol{\theta}}_{BELS}(k-1) \tag{17}$$

where k stands for the iteration step. As a linear regression based algorithm, the BELS methods involve low computational complexity and are well suited for real-time operation. However, their limitation is perhaps that they treat linear DEV models with white input and white output noises.

This paper is aimed at making further performance improvements on the BELS based methods for identification of the linear DEV model given by (1)–(4). The key ingredient in the BELS based methods is an estimation scheme for the two noise variances σ_v^2 and σ_w^2. This estimation scheme is utilized alternately with the bias correction procedure (17) to arrive at the BELS parameter estimate in an iterative manner. It goes without saying that performances of the BELS based methods depend critically upon how these noise variances are estimated. The estimation scheme of the BELS1 method presented in [13] consists of two linear equations with respect to σ_v^2 and σ_w^2. One equation is derived from asymptotic analysis of the mean LS error, whereas the other is obtained by the way of increasing the denominator of the transfer function $G(q^{-1})$ by one dimension, which is equivalent to using one extra delayed noisy output measurement. However, it has been observed that in a number of occasions the estimation scheme of the BELS1 method takes longer time to attain good estimates of the noise variances. Consequently, the BELS1 method requires more steps of iteration before a convergence criterion is reached. It is known that one of the primary performance measures of an iterative algorithm like the BELS1 method is its convergence speed, and a fast rate of convergence is always desired in real-world applications. Another observation is that the estimation scheme of the BELS1 method sometimes experiences difficulty to provide desirable estimates of the noise variances when the measurement noise level is relatively high. Estimation accuracy is then affected in the sense that it results in the BELS parameter estimate with relatively high variance though the estimate is still unbiased.

In this paper, a modified estimation scheme for the two noise variances is developed, and a new BELS based algorithm is thus proposed. The elementary idea is to utilize more than one additional delayed noisy output measurements so as to set up more than two linear equations with respect to σ_v^2 and σ_w^2. This can also be considered as that use of more autocovariances and crosscovariances of noisy input and output signals is made in estimation of the noise variances. Solving the resultant over-determined system of linear equations in a least-squares sense will provide better estimates for the two noise variances at the very early stage of the iterative process. This will then enable a better BELS parameter estimate to be obtained in a quicker way. Owing to such a beneficial interplay between the noise variance estimation and the parameter estimation, the proposed BELS based algorithm will attain convergence at a much fast speed, usually within a few number of iterations. Moreover, good estimates of the noise variances given by the modified estimation scheme will simultaneously enhance accuracy of the BELS parameter estimate in the form of comparatively low variance. As such, the new identification algorithm will work much reliably in the presence of high measurement noise.

2 Modified Estimation Scheme for Noise Variances

To begin with, several assumptions on the noisy input-output system (5) are made in order to satisfy the conditions of parametric identifiability [1].

- A1. The transfer function $G(q^{-1})$ is exponentially stable and contains no identical zeros and poles.
- A2. The true input signal $r(t)$ is stationary with a rational spectral density.
- A3. $r(t)$, $v(t)$ and $w(t)$ are independent of each other.
- A4. The model orders (n, m) are given.
- A5. The linear DEV model is with zero initial conditions, namely, $u(t) = 0$, $y(t) = 0$ for $t \leq 0$.

As illustrated before, central to the BELS based methods is estimation of the measurement noise variances σ_v^2 and σ_w^2. For this purpose, the mean LS error

$$J(\boldsymbol{\theta}_{LS}) = E[(y(t) - \boldsymbol{\psi}_t^\top \boldsymbol{\theta}_{LS})^2] \tag{18}$$

is examined. It can be shown that the following asymptotic expression holds [12]:

$$J(\boldsymbol{\theta}_{LS}) = \sigma_w^2 + \boldsymbol{\theta}_{LS}^\top \boldsymbol{D}\boldsymbol{\theta}. \tag{19}$$

If $\boldsymbol{\theta}_{LS}$ is partitioned in the same way as $\boldsymbol{\theta}$, namely

$$\boldsymbol{\theta}_{LS}^\top = [\boldsymbol{a}_{LS}^\top;\ \boldsymbol{b}_{LS}^\top], \quad \boldsymbol{a}_{LS} \in \mathbb{R}^n,\ \boldsymbol{b}_{LS} \in \mathbb{R}^m \tag{20}$$

then $J(\boldsymbol{\theta}_{LS})$ may be expressed explicitly in relation to σ_v^2 and σ_w^2 as

$$J(\boldsymbol{\theta}_{LS}) = \sigma_w^2(1 + \boldsymbol{a}_{LS}^\top \boldsymbol{a}) + \sigma_v^2 \boldsymbol{b}_{LS}^\top \boldsymbol{b}. \quad (21)$$

To arrive at more equations regarding the noise variances, an auxiliary transfer function

$$\mathcal{G}(q^{-1}) = \frac{B(q^{-1})}{\mathcal{A}(q^{-1})} = \frac{b_1 q^{-1} + \ldots + b_m q^{-m}}{1 - a_1 q^{-1} - \ldots - a_{n+d} q^{-(n+d)}} \quad (22)$$

is introduced, and the corresponding auxiliary parameter vector $\boldsymbol{\Theta}$ is defined as

$$\boldsymbol{\Theta}^\top = [\boldsymbol{\theta}^\top;\ \boldsymbol{\alpha}^\top], \quad \boldsymbol{\alpha}^\top = [a_{n+1}\ \ldots\ a_{n+d}] = \mathbf{0}^\top \in \mathbb{R}^d \quad (23)$$

where $d > 1$. It is seen from (23) that the introduced d parameters, $a_{n+1}, \cdots, a_{n+d}$, in $\mathcal{G}(q^{-1})$ all take the true value of zero. So the dynamic system described by $\mathcal{G}(q^{-1})$ should exhibit the same input-output behavior as that described by $G(q^{-1})$. However, one major difference between them is that $\mathcal{G}(q^{-1})$ is now with the increased model orders $(n, m + d)$.

Similarly to (11), the LS estimate of the auxiliary parameter vector $\boldsymbol{\Theta}$ may be obtained as

$$\boldsymbol{\Theta}_{LS} = \boldsymbol{R}_{\Psi\Psi}^{-1} \boldsymbol{R}_{\Psi y} \quad (24)$$

where

$$\boldsymbol{R}_{\Psi\Psi} = E[\boldsymbol{\Psi}_t \boldsymbol{\Psi}_t^\top], \quad \boldsymbol{R}_{\Psi y} = E[\boldsymbol{\Psi}_t y(t)] \quad (25)$$

$$\boldsymbol{\Psi}_t^\top = [\boldsymbol{\psi}_t^\top;\ \boldsymbol{\phi}_t^\top], \quad \boldsymbol{\phi}_t^\top = [y(t-n-1)\ \ldots\ y(t-n-d)]. \quad (26)$$

The definition of $\boldsymbol{\phi}_t$ shows that d actually represents the number of extra delayed noisy output measurements used. By a similar procedure to (13), $\boldsymbol{\Theta}_{LS}$ may be shown to have the asymptotic expression

$$\boldsymbol{\Theta}_{LS} = \boldsymbol{\Theta} + \Delta\boldsymbol{\Theta} \quad (27)$$

where $\Delta\boldsymbol{\Theta}$ is the noise-induced bias given by

$$\Delta\boldsymbol{\Theta} = -\boldsymbol{R}_{\Psi\Psi}^{-1} \boldsymbol{\mathcal{D}} \boldsymbol{\Theta} \quad (28)$$

with

$$\boldsymbol{\mathcal{D}} = \mathrm{diag}\{\boldsymbol{D};\ \sigma_w^2 \boldsymbol{I}_d\}. \quad (29)$$

With the auxiliary transfer function, another d equations regarding the noise variances may be derived. This is stated in the following theorem whose proof is omitted here.

Theorem 1 *The noise variances σ_v^2 and σ_w^2 satisfy*

$$\sigma_w^2 \boldsymbol{R}_{\psi\phi}^\top \boldsymbol{R}_a \boldsymbol{a} + \sigma_v^2 \boldsymbol{R}_{\psi\phi}^\top \boldsymbol{R}_b \boldsymbol{b} = \boldsymbol{R}_{\phi y} - \boldsymbol{R}_{\psi\phi}^\top \boldsymbol{\theta}_{LS} \quad (30)$$

where

$$\boldsymbol{R}_{\psi\phi} = E[\boldsymbol{\psi}_t \boldsymbol{\phi}_t^\top], \quad \boldsymbol{R}_{\phi y} = E[\boldsymbol{\phi}_t y(t)] \quad (31)$$

$$\boldsymbol{R}_a = \boldsymbol{R}_{\psi\psi}^{-1}(:, 1{:}n), \quad \boldsymbol{R}_b = \boldsymbol{R}_{\psi\psi}^{-1}(:, 1{:}m) \quad (32)$$

Putting (21) and (30) together, we now obtain the following set of $d + 1$ linear equations with respect to σ_v^2 and σ_w^2:

$$\begin{bmatrix} \boldsymbol{R}_{\psi\phi}^\top \boldsymbol{R}_a \boldsymbol{a} & \boldsymbol{R}_{\psi\phi}^\top \boldsymbol{R}_b \boldsymbol{b} \\ 1 + \boldsymbol{a}_{LS}^\top \boldsymbol{a} & \boldsymbol{b}_{LS}^\top \boldsymbol{b} \end{bmatrix} \begin{bmatrix} \sigma_w^2 \\ \sigma_v^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{\phi y} - \boldsymbol{R}_{\psi\phi}^\top \boldsymbol{\theta}_{LS} \\ J(\boldsymbol{\theta}_{LS}) \end{bmatrix} \quad (33)$$

which may be used to estimate the noise variances in conjunction with estimating the parameters of the linear DEV model. To sum up, the iterative procedure of the proposed fast BELS (FBELS) algorithm is given as follows.

The Fast BELS Algorithm

Step 0. Algorithm initialization:

(i) Choose the value of d.

(ii) Compute the covariance estimates $\hat{\boldsymbol{R}}_{\psi\psi}$, $\hat{\boldsymbol{R}}_{\psi y}$, $\hat{\boldsymbol{R}}_{\psi\phi}$ and $\hat{\boldsymbol{R}}_{\phi y}$ by use of noisy input-output measurements $\{u(t), y(t), t = 1, \ldots, N\}$, where N denotes the number of data points.

(iii) Evaluate the LS parameter estimate

$$\hat{\boldsymbol{\theta}}_{LS} = \hat{\boldsymbol{R}}_{\psi\psi}^{-1} \hat{\boldsymbol{R}}_{\psi y}. \quad (34)$$

(iv) Calculate the average LS error

$$\hat{J}(\hat{\boldsymbol{\theta}}_{LS}) = \frac{1}{N} \sum_{t=1}^{N} [y(t) - \boldsymbol{\psi}_t^\top \hat{\boldsymbol{\theta}}_{LS}]^2. \quad (35)$$

(v) Set the initial BELS parameter estimate

$$\hat{\boldsymbol{\theta}}_{BELS}(0) = \hat{\boldsymbol{\theta}}_{LS}. \quad (36)$$

Step 1. Compute the noise variance estimates

$$\begin{bmatrix} \hat{\sigma}_w^2(k) \\ \hat{\sigma}_v^2(k) \end{bmatrix} = [\hat{\boldsymbol{H}}^\top(k) \hat{\boldsymbol{H}}(k)]^{-1} \hat{\boldsymbol{H}}^\top(k) \hat{\boldsymbol{h}}(k) \quad (37)$$

where

$$\hat{\boldsymbol{H}}(k) = \begin{bmatrix} \hat{\boldsymbol{R}}_{\psi\phi}^\top \hat{\boldsymbol{R}}_a \hat{\boldsymbol{a}}_{BELS}(k-1) & \hat{\boldsymbol{R}}_{\psi\phi}^\top \hat{\boldsymbol{R}}_b \hat{\boldsymbol{b}}_{BELS}(k-1) \\ 1 + \hat{\boldsymbol{a}}_{LS}^\top \hat{\boldsymbol{a}}_{BELS}(k-1) & \hat{\boldsymbol{b}}_{LS}^\top \hat{\boldsymbol{b}}_{BELS}(k-1) \end{bmatrix} \quad (38)$$

$$\hat{\boldsymbol{h}}(k) = \begin{bmatrix} \hat{\boldsymbol{R}}_{\phi y} - \hat{\boldsymbol{R}}_{\psi\phi}^\top \hat{\boldsymbol{\theta}}_{LS} \\ \hat{J}(\hat{\boldsymbol{\theta}}_{LS}) \end{bmatrix}. \quad (39)$$

Step 2. Evaluate the BELS parameter estimate

$$\hat{\boldsymbol{\theta}}_{BELS}(k) = \hat{\boldsymbol{\theta}}_{LS} + \hat{\boldsymbol{R}}_{\psi\psi}^{-1} \hat{\boldsymbol{D}}(k) \hat{\boldsymbol{\theta}}_{BELS}(k-1) \quad (40)$$

where $\hat{\boldsymbol{D}}(k)$ is as defined in (16).

Step 3. Check whether $\hat{\boldsymbol{\theta}}_{BELS}(k)$ satisfies the convergence criterion

$$\frac{\|\hat{\boldsymbol{\theta}}_{BELS}(k) - \hat{\boldsymbol{\theta}}_{BELS}(k-1)\|}{\|\hat{\boldsymbol{\theta}}_{BELS}(k)\|} < \delta \quad (41)$$

where δ is a small positive number. If not, set $k = k + 1$ and go to step 1; otherwise, terminate the iterative process.

3 Discussion and Remarks

We notice that the difference between the BELS1 method presented in [13] and the FBELS algorithm proposed here only lies in their estimation scheme for the noise variances. Although the FBELS algorithm may be viewed as an extension of the BELS1 method in the sense that (33) will reduce to the one in the BELS1 method if d is selected as $d = 1$, this difference is still very significant. As illustrated before, successful application of the BELS based methods critically depends upon whether the two noise variances σ_v^2 and σ_w^2 can be estimated in an effective way. When extra delayed noisy output measurements are used (i.e. $d > 1$ is chosen), the two noise variances are estimated with more information of the identified linear DEV model. This will result in better estimates of σ_v^2 and σ_w^2 at the very early stage of the iterative process, which, in turn, will accelerate the convergence speed of the proposed FBELS algorithm in terms of the number of iterations required to satisfy the convergence criterion (41). In addition to a fast rate at which the FBELS algorithm will converge, accuracy of the parameter estimates can also be much improved owing to the use of the good estimates of the noise variances, which is conceivable.

The computational aspect of the FBELS algorithm is examined. Except for $\boldsymbol{R}_{\psi\psi}$ and $\boldsymbol{R}_{\psi y}$ that are needed for the LS estimate $\boldsymbol{\theta}_{LS}$, one major increased computational cost in implementation of the FBELS algorithm is evaluation of estimates of the covariance matrix $\boldsymbol{R}_{\psi\phi}$ and the covariance vector $\boldsymbol{R}_{\phi y}$. To this end, it is interesting to take a close look at these covariance matrix and vector. For illustration convenience, assume that $m = n$ and define

$$r_{yy}(j) = E[y(t)y(t-j)], \quad r_{uy}(j) = E[u(t)y(t-j)]. \quad (42)$$

It follows that in calculating $\boldsymbol{R}_{\psi\phi}$ and $\boldsymbol{R}_{\phi y}$, it suffices to compute the following $2d$ covariance functions

$$\begin{aligned} &\{r_{yy}(n+1),\ \ldots,\ r_{yy}(n+d)\} \\ &\{r_{uy}(n),\ \ldots,\ r_{uy}(n+d-1)\} \end{aligned} \quad (43)$$

because all the other components of $\boldsymbol{R}_{\psi\phi}$ and $\boldsymbol{R}_{\phi y}$ may simply make use of appropriate components from the available covariance matrix $\boldsymbol{R}_{\psi\psi}$ and the covariance vector $\boldsymbol{R}_{\psi y}$. It can be envisaged that the computational load thus increased is quite limited. Our simulation studies have shown that usually a small value of d (e.g. $d = 2$) can lead to significantly improved performances in such aspects as rate of convergence and accuracy of estimation. The other computational issue is that because $d > 1$, (33) actually represents an overdetermined set of $d + 1$ linear equations with respect to the two noise variances σ_v^2 and σ_w^2. As shown in (37), it must be solved in a least-squares sense to obtain the estimates $\hat{\sigma}_v^2(k)$ and $\hat{\sigma}_w^2(k)$. Fortunately, since d only needs to take a small value as just mentioned, evaluating the pseudo-inverse of the $(d+1) \times 2$ matrix $\hat{\boldsymbol{H}}^+(k)$ will not cause much added computations. By and large, at the price of a small increase in the computational burden, the modified estimation scheme (37) for the noise variances may be used in return for important performance improvements of the FBELS algorithm. This will be further evidenced by the simulation results given in the next section.

We now compare the proposed FBELS algorithm with the other methods. Like the BELS1 method presented in [13], the FBELS algorithm is superior to the the KL method and the JO method in that it has good robustness against lack of a priori noise information because of its capacity of direct estimation of the variances of the input and output noises. Furthermore, the FBELS algorithm has a computational advantage over the KL method, the JO method and the IV-WSF method since only linear calculations are involved. Unlike the LLS-FD method, the FBELS algorithm is readily applicable to real-time identification. On the other hand, the JO method, the LLS-FD method and the IV-WSF method may be used to identify linear DEV models with colored input and output noises, thereby having wider domain of application than the FBELS algorithm.

4 Simulation Studies

Monte-Carlo stochastic simulation studies are carried out to illustrate the performances of the proposed FBELS algorithm in comparison with the BELS1 method in [13] in terms of estimation accuracy, computational complexity and convergence speed. The estimation accuracy is described by bias and variance of parameter estimates. Moreover, to get an overall description, the relative error (RE) and the normalized root mean squared error (RMSE) of an estimator $\hat{\boldsymbol{\theta}}_k$ with regard to its true value $\boldsymbol{\theta}$ in the kth run over a total of M Monte-Carlo runs are introduced:

$$\text{RE} = \frac{\|\boldsymbol{m}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}\|}{\|\boldsymbol{\theta}\|} \quad (44)$$

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\frac{\|\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^2}} \quad (45)$$

where $\boldsymbol{m}(\hat{\boldsymbol{\theta}})$ denotes the sample mean of $\hat{\boldsymbol{\theta}}_k$. The computational complexity is measured approximately by the MATLAB code `flops` (count of floating point operations), and described by the number of `flops` per run (NFPR). The convergence speed of the iterative procedure of the BELS based algorithms is described by the number of iterations per run (NIPR) according to the convergence criterion (41) with a choice of $\delta = 0.001$. In the following, $M = 200$ Monte-Carlo runs are conducted for all examples considered. Further, the signal-to-noise ratios (SNR) at the system input and the system output are calculated respectively

Table 1: Comparative Estimation Performances
($N = 2000$, SNRI = SNRO $\approx$ 10dB, 200 Monte-Carlo runs)

Parameter	a_1	a_2	b_1	b_2	σ_v^2	σ_w^2	RE	RMSE	NFPR	NIPR
True value	-1.6	-0.9	1.2	0.7	0.2	0.4				
LS	-0.9449 ± 0.0383	-0.3266 ± 0.0321	1.0126 ± 0.0231	0.1204 ± 0.0346	—	—	46.15%	46.23%	56289	—
BELS1	-1.6315 ± 0.2743	-0.9273 ± 0.2284	1.1438 ± 0.6512	0.7209 ± 0.3037	0.0604 ± 0.3053	0.4403 ± 0.5118	3.17%	24.99%	89129	16.000
FBELS d=2	-1.6001 ± 0.0194	-0.8999 ± 0.0175	1.1991 ± 0.0613	0.6956 ± 0.0417	0.1915 ± 0.0868	0.4020 ± 0.0383	0.19%	3.42%	95567	5.895
FBELS d=3	-1.6006 ± 0.0133	-0.9004 ± 0.0122	1.1980 ± 0.0588	0.6973 ± 0.0372	0.1913 ± 0.0828	0.4021 ± 0.0352	0.14%	3.12%	103688	5.705
FBELS d=4	-1.6012 ± 0.0131	-0.9008 ± 0.0126	1.1956 ± 0.0609	0.6982 ± 0.0349	0.1875 ± 0.0859	0.4035 ± 0.0365	0.21%	3.12%	111860	5.755
FBELS d=5	-1.6011 ± 0.0129	-0.9007 ± 0.0124	1.1953 ± 0.0604	0.6981 ± 0.0347	0.1870 ± 0.0867	0.4036 ± 0.0366	0.22%	3.13%	120017	5.745
FBELS d=6	-1.6009 ± 0.0125	-0.9006 ± 0.0121	1.1945 ± 0.0618	0.6980 ± 0.0343	0.1854 ± 0.0893	0.4040 ± 0.0372	0.25%	3.17%	128174	5.735
FBELS d=7	-1.6009 ± 0.0136	-0.9005 ± 0.0124	1.1944 ± 0.0618	0.6979 ± 0.0359	0.1853 ± 0.0892	0.4041 ± 0.0371	0.26%	3.21%	136310	5.655
FBELS d=8	-1.6012 ± 0.0122	-0.9008 ± 0.0112	1.1940 ± 0.0603	0.6987 ± 0.0348	0.1853 ± 0.0863	0.4041 ± 0.0360	0.27%	3.12%	144462	5.635
FBELS d=9	-1.6007 ± 0.0124	-0.9004 ± 0.0115	1.1947 ± 0.0612	0.6978 ± 0.0351	0.1860 ± 0.0871	0.4037 ± 0.0358	0.24%	3.16%	152614	5.620
FBELS d=10	-1.6009 ± 0.0118	-0.9006 ± 0.0112	1.1937 ± 0.0603	0.6984 ± 0.0344	0.1849 ± 0.0857	0.4042 ± 0.0355	0.28%	3.11%	160772	5.620

as

$$\text{SNRI} = 10\log_{10}\frac{\sum_{t=1}^{N} r(t)^2}{\sum_{t=1}^{N} v(t)^2}\ \text{(dB)} \tag{46a}$$

$$\text{SNRO} = 10\log_{10}\frac{\sum_{t=1}^{N} x(t)^2}{\sum_{t=1}^{N} w(t)^2}\ \text{(dB)}. \tag{46b}$$

Example 1. The transfer function of a linear DEV model to be identified is given by

$$G(q^{-1}) = \frac{1.2q^{-1} + 0.7q^{-2}}{1 + 1.6q^{-1} + 0.9q^{-2}} \tag{47}$$

which means that $a_1 = -1.6$, $a_2 = -0.9$, $b_1 = 1.2$, and $b_2 = 0.7$. The true input signal $r(t)$ is generated as a first-order autoregressive process

$$r(t) = \frac{1}{1 - 0.7q^{-1}}\omega(t) \tag{48}$$

where $\omega(t)$ is zero-mean white noise with unit variance. The variances of the input noise $v(t)$ and the output noise $w(t)$ are selected as $\sigma_v^2 = 0.2$ and $\sigma_v^2 = 0.4$, respectively, so that the corresponding SNRI and SNRO both are approximately equal to 10dB. By setting the sample length $N = 2000$, the underlying linear DEV model is identified by using the LS method, the BELS1 method in [13] and the proposed FBELS algorithm. In particular, the different values of d (i.e. the number of extra delayed noisy output measurements used), ranging from $d = 2$ to $d = 10$, are chosen to examine its effect on the performances of the FBELS algorithm. The sample means and standard deviations of the estimates, RE, RMSE, NFPR and NIPR are displayed in Table 1.

In agreement with the analysis given in Section 1, the LS estimate is seriously biased with a high RMSE value due to the presence of input and output noises. The BELS1 method and the FBELS method both have produced good results. However, with the use of more delayed noisy output measurements, the rate of convergence of the FBELS algorithm is significantly accelerated, which is at least 2.7 times faster than that of the BELS1 method in terms of NIPR for all the chosen values of d. Moreover, the FBELS algorithm has much smaller RMSE values than the BELS1 method, thus being a more accurate estimator. As expected, such greatly improved performances are achieved at an increased numerical cost. For instance, in comparison with the BELS1 method, the computational load of the FBELS method in terms of NFPR is increased by roughly 7% when $d = 2$ and by roughly 80% when $d = 10$. But it is important to note that the other performances (such as RE, RMSE and NIPR) of the FBELS algorithm with $d = 2$ only differ slightly from those of the FBELS algorithm with the other values of d. These observations may lead to the conclusion that the choice of $d = 2$ will be appropriate for the FBELS algorithm since this may produce the best trade-off between the accelerated rate of convergence and the increased amount of computations while at almost no sacrifice of estimation accuracy.

Example 2. To study the effect of the input and output noises on the proposed FBELS algorithm, Example 1 is again considered but the different noise levels are examined:

$$\begin{array}{lll} \text{(a)}\ \sigma_v^2=0.63, & \sigma_w^2=1.26 & \Longrightarrow \text{SNRI}=\text{SNRO}\approx 5\text{dB} \\ \text{(b)}\ \sigma_v^2=0.2, & \sigma_w^2=0.4 & \Longrightarrow \text{SNRI}=\text{SNRO}\approx 10\text{dB} \\ \text{(c)}\ \sigma_v^2=0.063, & \sigma_w^2=0.126 & \Longrightarrow \text{SNRI}=\text{SNRO}\approx 15\text{dB} \end{array}$$

The FBELS algorithm with $d = 2$ is applied to the simulated data of a sample length $N = 4000$, together with the LS method and the BELS1 method. The numerical results are obtained based on an ensemble average of 200 Monte-Carlo runs, but are omitted here due to limited space.

With the noise level being reduced from SNR = 5dB to SNR = 15dB, the LS method still gives nonconsistent parameter estimates whereas the two BELS based algorithms always yield unbiased parameter estimates. Similarly to what has been observed in Example 1, with a slightly increase (i.e. around 8%) in the numerical cost, the FBELS algorithm exhibits better estimation accuracy than the BELS1 method in both low and high SNR environments. An interesting observation is that the FBELS algorithm converges at a moderately fast rate (around 1.5 times faster than the BELS1 method) in the case of low SNR value (SNR = 5dB) but at a considerably fast rate (around 3.3 times faster than the BELS1 method) in the case of high SNR value (SNR = 15dB).

5 Conclusions

An effective estimation scheme for the measurement noise variances has been presented. It is based on the use of d additional delayed noisy output measurements so that good estimates of the noise variances are readily obtainable. The FBELS algorithm has then been established for unbiased identification of linear DEV models. The attractive features of the proposed algorithm are its accelerated rate of convergence during the iterative process and its enhanced accuracy of estimation in the face of high measurement noise. In particular, with the choice of $d = 2$, the proposed FBELS algorithm can be implemented at a computational cost just slightly more than that of the previous BELS1 method, thereby being most cost-effective with respect to the achieved important performance improvements. The proposed identification algorithm has been validated by the numerical examples.

Acknowledgements—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, Nepean, Australia.

References

[1] B. D. O. Anderson and M. Deistler, "Identifiability in dynamic errors-in-variables models," *J. Time Series Analysis*, vol.5, pp.1-13, 1984.

[2] M. H. A. Davis and R. B. Vinter, *Stochastic Modeling and Control*. London: Chapman and Hall, 1985.

[3] M. Deistler and B. D. O. Anderson, "Linear dynamic errors-in-variables models: some structure theory," *J. Econometrics*, vol.41, pp.39-63, 1989.

[4] K. V. Fernando and H. Nicholson, "Identification of linear systems with input and output noise: the Koopmans-Levin method," *IEE Proc. Control Theory & Applications*, vol.132, pp.30-36, 1985.

[5] C. W. J. Granger and P. Newbold, *Forecasting Economic Time Series*. Princeton, NJ: Academic Press, 1986.

[6] P. Guillaume, R. Pintelon and J. Schoukens, "Robust parametric transfer function estimation using complex logarithmic frequency response data," *IEEE Trans. Automatic Control*, vol.40, pp.1180-1190, 1995.

[7] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1996.

[8] W. Scherrer and M. Deistler, "A structure theory for linear dynamic errors-in-variables models," *SIAM J. Control & Optimization*, vol.36, pp.2148-2175, 1998.

[9] J. Schoukens and R. Pintelon, *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*. Oxford, U.K.: Pergamon Press, 1991.

[10] T. Söderström, "Identification of stochastic linear systems in presence of input noise," *Automatica*, vol.17, pp.713-725, 1981.

[11] P. Stoica, M. Cedervall and A. Eriksson, "Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems," *IEEE Trans. Signal Processing*, vol.43, pp.2386-2397, 1995.

[12] W. X. Zheng, "Transfer function estimation from noisy input and output data," *Int. J. Adaptive Control & Signal Processing*, vol.12, pp.365-380, 1998.

[13] W. X. Zheng, "On least-squares identification of stochastic linear systems with noisy input-output data," *Int. J. Adaptive Control & Signal Processing*, vol.13, pp.131-143, 1999.