Received: 10 January 2016 | Revised: 29 September 2016 | Accepted: 6 December 2016
DOI 10.1002/acs.2751

WILEY

RESEARCH ARTICLE

Identification of errors-in-variables systems: An asymptotic approach

Xin Liu | Yucai Zhu

State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, China

Correspondence
Yucai Zhu, State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China.
Email: yczhu@iipc.zju.edu.cn

Funding information
National Science Foundation of China, Grant/Award Number: (grant 61273191); 973 Program of China, Grant/Award Number: (grant 2012CB720500)

Summary
This work studies the identification of errors-in-variables (EIV) systems. An asymptotic method (ASYM) is developed for the EIV system. Firstly, an auto regressive with exogeneous (ARX) model estimation method is proposed, which is consistent for EIV systems. Then the asymptotic variance expression of the estimated high-order ARX model is derived, which forms the basis of the ASYM method. In parameter estimation, the ASYM starts with a high-order ARX model estimation followed by a frequency domain weighted model reduction. The obtained model is consistent, and its efficiency needs to be investigated. Besides parameter estimation, a criterion for model order selection is proposed, which is based on frequency domain considerations, and the frequency domain error bound is established that can be used for model validation. Simulations and comparisons with other methods are used to illustrate the performance of the method.

KEYWORDS
ARX model, asymptotic method, Box-Jenkins model, errors-in-variables models, input noise variance

1 | INTRODUCTION

System identification is the field of mathematical modeling of systems using test data. Most identification methods, for example, the class of prediction error methods and subspace methods, assume that the true input signals are available, which is true in control system applications. In noncontrol applications, the inputs are obtained through measurement, which are corrupted by measurement noises. Examples of noncontrol applications are population growth modeling, infectious disease transmission modeling, fault monitoring and diagnosis, and environment modeling. Systems where measurement noises and/or unmeasured disturbances are present on both inputs and outputs are called errors-in-variables (EIV) systems, see Soderstrom.[1] There are 2 major causes of model errors. On the one hand, the disturbances cause estimation errors and also cause difficulties in model order selection and parameter estimation. On the other hand, the limited amount of data leads to inaccurate description of the system. In EIV system identification, understanding the properties of these errors is of great importance for increasing the model accuracy.

During the past few decades, many researches have studied the identification of EIV systems. One study[2] proposes the Koopmans-Levin method, which only treats the case of the white input and white output noises, and it also assumes that the ratio of the variances of the input and output noises is known. The joint-output method has been analyzed in on study,[3] which treats the measured input and output signals as outputs of a multivariable stochastic system. The method must solve a Riccati equation in every evaluation of the loss function, and it has a problem that it may converge to false minima. The method may not work well when the input is not a true auto regressive moving average (ARMA) process. The Frisch scheme is a classical algebra estimation problem, see Frisch,[4] and many researchers have extended it to dynamic model estimation, see 6 studies.[5–10] The Frisch estimation method is based on the assumption of white input and white output measurement noise. The idea of Frisch estimation method is to have appropriate estimates of the noise variances and then determine the parameters. By using the instrumental-variable

(IV) principle, an improved IV method called the combined IV and weighted subspace fitting (IV-WSF) method is proposed, see 2 studies[11,12]; it gives a consistent estimate, but an optimal weighting matrix must be chosen first, which costs extended calculations. In one study,[13] the author puts several methods into a general framework, resulting in a Generalized Instrumental Variable Estimator. The bias-eliminated least squares (BELS) method has been proposed in 1 study[14]; it treats the case that the input noise and output noise are both white; in this method, a known prefilter is connected to the input of the system, and the original system parameter estimation needs to be extracted from that of the augmented system parameters. Later, a new version of the BELS method called the BELS-II method was developed, which can be applied in the white input noise and colored output noise cases, see 1 study.[15] The so-called total least squares method considers both input and output noises in system identification, and this class of methods assume that the output disturbance is white noise, see, eg, 2 studies.[16,17]

Compared to traditional prediction error methods and subspace methods, EIV system identification is in an immature stage and basic problems such as efficient estimation, order selection, and model validation still need to be resolved.

In 1 study,[18] an asymptotic method (ASYM) was developed for noise-free input systems. The method provides systematic solutions to the 4 basic problems of system identification: test signal design, parameter estimation, order selection, and model validation. It gives an efficient estimate for the frequency response, and it performs very well in industry control applications. In parameter estimation, the first step of the ASYM is a high-order ARX model estimation; then, based on the so called asymptotic theory, see Ljung[19] and Zhu,[20] a frequency domain weighted model reduction is used to obtain a reduced order model that is asymptotically efficient. This work will extend the ASYM of Zhu[18] to EIV systems. The existing ASYM cannot be applied to EIV systems because of 2 obstacles: (1) The linear least-squares (LS) estimation of an ARX EIV system is biased and (2) the existing asymptotic theory for the noise-free input system does not hold for EIV systems. This work will solve these 2 problems and then extend the ASYM of Zhu[18] to EIV systems.

The paper is organized as follows. Section 2 is the problem statement. In Section 3, a method of ARX model estimation of EIV systems is proposed that gives consistent parameter estimates. In Section 4, the asymptotic properties of high-order ARX models are analyzed, and the asymptotic variance expression of the model is derived. In Section 5, the ASYM is developed for EIV systems with a Box-Jenkins structure. In Section 6, numerical examples and comparisons are given to illustrate the performance of the method. Section 7 is the conclusion.

2 | PROBLEM STATEMENT

Consider an open-loop single-input single-output system with Box-Jenkins model structure, which is depicted in Figure 1.

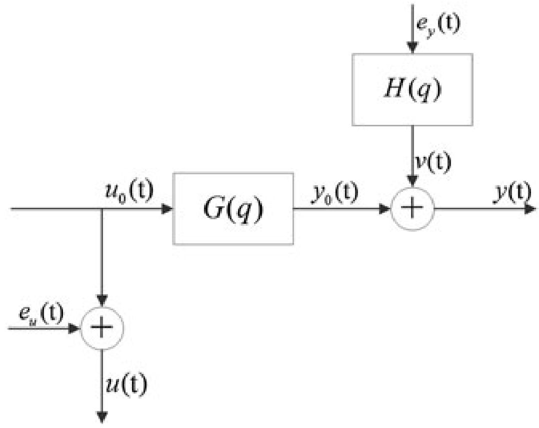

FIGURE 1 Block diagram of the errors-in-variables system

The process model and the disturbance model are denoted by $G(q)$ and $H(q)$, respectively. The noise-free input and output signals are denoted by $u_0(t)$ and $y_0(t)$; measured (noisy) input and output signals are denoted by $u(t)$ and $y(t)$. Under the Box-Jenkins model structure, the system can be written as

$$y(t) = \frac{B_0(q)}{A_0(q)} u_0(t) + v(t), \tag{1}$$

$$u(t) = u_0(t) + e_u(t), \tag{2}$$

where

$$A_0(q) = 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na} \tag{3}$$

$$B_0(q) = b_1 q^{-1} + \cdots + b_{nb} q^{-nb}. \tag{4}$$

Here q^{-1} is the backward shift operator, ie, $q^{-1}u_0(t) = u_0(t-1)$. $e_u(t)$ is a white noise signal with 0 mean and variance $\lambda_{eu} = E[e_u(t)^2]$, where $E[\cdot]$ denotes the expectation operator. $v(t)$ is the output noise, which is filtered white noise

$$v(t) = \frac{C_0(q)}{D_0(q)} e_y(t), \tag{5}$$

where

$$C_0(q) = 1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}, \tag{6}$$

$$D_0(q) = 1 + d_1 q^{-1} + \cdots + d_{nd} q^{-nd}, \tag{7}$$

and $e_y(t)$ is the white noise with 0 mean and variance λ_{ey}.

The parameter vector to be estimated is

$$\theta_{BJ} = [a_1, \cdots, a_{na}, b_1, \cdots, b_{nb}, c_1, \cdots, c_{nc}, d_1, \cdots, d_{nd}]^{\mathrm{T}}. \tag{8}$$

Denote the available data set as $Z_N = \{y(1), u(1), \cdots y(N), u(N)\}$, where N is the number of sampled data used for identification. Then the identification problem treated in this work is three folds: (1) to estimate the parameters θ_{BJ} from the data set; (2) to determine the model order; and (3) to quantify the model error for model validation.

3 | ARX MODEL ESTIMATION OF EIV SYSTEMS

The ASYM of Zhu[18] starts with a high-order ARX model estimation using the well-known linear least-squares method, which is consistent for noise-free input systems. However, for an EIV ARX system, the LS method will give a biased estimate. Here, a method of EIV ARX model estimation is developed, which is consistent for EIV ARX systems, see 1 study.[21]

3.1 | The LS estimate of the ARX model

In this section, for the moment, we assume that the system is with the normal ARX model structure, see Figure 2. Then, we have

$$A_0(q)y(t) = B_0(q)u_0(t) + e_y(t), \tag{9}$$

where $A_0(q)$ and $B_0(q)$ are polynomials as

$$A_0(q) = 1 + a_1q^{-1} + a_2q^{-2} + \cdots + a_{na}q^{-na}, \tag{10}$$

$$B_0(q) = b_1q^{-1} + b_2q^{-2} + \cdots + b_{nb}q^{-nb}, \tag{11}$$

and the system is represented by

$$y(t) = \frac{B_0(q)}{A_0(q)}u_0(t) + \frac{1}{A_0(q)}e_y(t), \tag{12}$$

$$u(t) = u_0(t) + e_u(t). \tag{13}$$

The parameter vector to be estimated is

$$\theta_0 = [a_1, a_2, \cdots, a_{na}, b_1, b_2, \cdots, b_{nb}]^{\mathrm{T}}. \tag{14}$$

Introduce the data vectors

$$\psi_t = [-y(t-1), \cdots, -y(t-na), u(t-1), \cdots, u(t-nb)]^{\mathrm{T}}, \tag{15}$$

$$\varepsilon_t = [\underbrace{0, \cdots, 0}_{na}, e_u(t-1), \cdots, e_u(t-nb)]^{\mathrm{T}}. \tag{16}$$

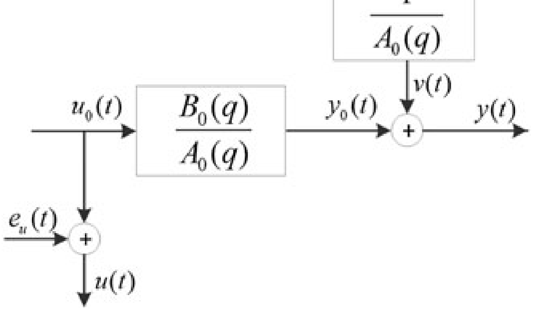

FIGURE 2 ARX model of the errors-in-variables system

Then the system Equations 12 and 13 can be rewritten in linear regression form

$$y(t) = \psi_t^{\mathrm{T}}\theta + \varepsilon(t), \tag{17}$$

where

$$\varepsilon(t) = e_y(t) - \varepsilon_t^{\mathrm{T}}\theta. \tag{18}$$

Concerning the system, the following assumptions are imposed.

A1. The system is stable.
A2. na and nb of the system are known.
A3. There is no common factor in $A_0(q)$ and $B_0(q)$.
A4. $e_u(t)$ and $e_y(t)$ are white noise sequences with 0 means, and they are mutually uncorrelated.
A5. The noise-free input $\{u_0(t)\}$ is a stationary ergodic random sequence with 0 mean and is uncorrelated with $\{e_u(t)\}$ and $\{v(t)\}$, which implies that the data are collected from an open loop test.

The linear LS method for the EIV system can be obtained by minimizing the loss function

$$\hat{\theta}_{LS}(N) = \arg\min_{\theta \in D} V_N(\theta), \tag{19}$$

where

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t)^2 = \frac{1}{N}\sum_{t=1}^{N}[y(t) - \psi_t^{\mathrm{T}}\theta]^2. \tag{20}$$

Thus, the LS estimate for the system parameter vector can be obtained as

$$\hat{\theta}_{LS}(N) = \left[\frac{1}{N}\sum_{t=1}^{N}\psi_t\psi_t^{\mathrm{T}}\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N}\psi_t y(t)\right]. \tag{21}$$

If the input noise does not exist, it can be shown that the estimate is consistent and when the white noise is Gaussian, it is also efficient. However, under EIV condition, the estimate is biased and we will derive the asymptotic bias expression. Using Equations 17, 18, and 21 we get

$$\left[\frac{1}{N}\sum_{t=1}^{N}\psi_t\psi_t^{\mathrm{T}}\right]\left[\hat{\theta}_{LS}(N) - \theta_0\right] = \frac{1}{N}\sum_{t=1}^{N}\psi_t e_y(t) - \frac{1}{N}\sum_{t=1}^{N}\psi_t\varepsilon_t^{\mathrm{T}}\theta. \tag{22}$$

Following the assumptions A4 and A5, one has

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\psi_t e_y(t) = E\{\psi_t e_y(t)\} = 0, \tag{23}$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\psi_t\varepsilon_t^{\mathrm{T}}\theta = E\{\psi_t\varepsilon_t^{\mathrm{T}}\theta\} = D\theta, \tag{24}$$

where the matrix D is defined as

$$D = diag\{O_{na}; \lambda_{eu}I_{nb}\}, \tag{25}$$

where O_{na} denotes a 0 matrix with dimension na, I_{nb} denotes a unit matrix with dimension nb. Then we have

$$\lim_{N\to\infty}\hat{\theta}_{LS}(N)-\theta=-\lim_{N\to\infty}\left[\frac{1}{N}\sum_{t=1}^{N}\psi_t\psi_t^{\mathrm{T}}\right]^{-1}D\theta. \quad (26)$$

Solving the above equation, one obtains

$$\theta=[I_{na+nb}-R_{\psi\psi}^{-1}D]^{-1}\hat{\theta}_{LS}(N), \quad (27)$$

where

$$R_{\psi\psi}=\lim_{N\to\infty}\left[\frac{1}{N}\sum_{t=1}^{N}\psi_t\psi_t^{\mathrm{T}}\right]=E[\psi_t\psi_t^{\mathrm{T}}]. \quad (28)$$

For a finite data, assuming that the input noise variance is known, one can obtain a consistent estimation of the parameter vector

$$\hat{\theta}_{ARX}=[I_{na+nb}-\hat{R}_{\psi\psi}^{-1}\hat{D}]^{-1}\hat{\theta}_{LS}(N). \quad (29)$$

Another form, which is equal to Equation 29, can be obtained after simple calculations

$$\hat{\theta}_{ARX}=\left[\frac{1}{N}\sum_{t=1}^{N}\psi_t\psi_t^{\mathrm{T}}-\hat{D}\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N}\psi_t y(t)\right]. \quad (30)$$

Note that this last equation does not contain the LS estimate of the system. In practice, however, we have no knowledge about the exact input noise variance. Then a natural way to consistent ARX model estimation is to estimate the variance of the input noise λ_{eu}.

3.2 | Error criterion for input noise variance estimation

An error criterion will be developed that can be used to estimate the input noise variance. Denote the model estimate as $\hat{B}(q)/\hat{A}(q)$ and output error of the model estimate as

$$\varepsilon_{oe}(t)=y(t)-\frac{\hat{B}(q)}{\hat{A}(q)}u(t). \quad (31)$$

When the input signal is noise free and the test is in open loop, one can use the output error criterion (the sum of the squares of the output error) in parameter estimation and the estimate will be consistent. In EIV system identification, because of the impact of the input noise, one cannot obtain a consistent estimate using the output error criterion. The idea here is to revise the output error criterion of the input noise-free case for the ARX model of EIV systems so that a consistent estimate can be obtained. For eliminating the influence of the input noise, we need to develop a criterion that is independent of the input noise. Because the output signal is not correlated with the input noise, one can take away the influence of the input noise by correlating the output signal with the output error. Therefore, we introduce the following loss function:

$$\begin{aligned}V_\lambda(N)=\frac{1}{N-p}\sum_{t=p+1}^{N}\Big\{&[\varepsilon_{oe}(t-p)\cdot y(t)]^2+\cdots\\&+[\varepsilon_{oe}(t-1)\cdot y(t)]^2+[\varepsilon_{oe}(t)\cdot y(t)]^2\\&+[\varepsilon_{oe}(t+1)\cdot y(t)]^2+\cdots+[\varepsilon_{oe}(t+p)\cdot y(t)]^2\Big\},\end{aligned} \quad (32)$$

where p is the maximum lag in calculating the cross correlations. In applications, the value of p should cover most of the system transient time and can be set to the length of the system impulse response. If the user does not know the length of the impulse response, use a longer estimate is better than a shorter one. This loss function can be called correlated output error criterion. The motivation of using this criterion is simply to average out the influence of the input noise influence in the output error criterion. Considering the $(p-k+1)$th term in the sum on the right hand side of Equation 32, according to Equation 12, one has

$$\begin{aligned}\varepsilon_{oe}(t-k)\cdot y(t)=&\left[\frac{B_0(q)}{A_0(q)}u_0(t-k)+v(t-k)-\frac{\hat{B}(q)}{\hat{A}(q)}u_0(t-k)\right.\\&\left.-\frac{\hat{B}(q)}{\hat{A}(q)}e_u(t-k)\right]\left[\frac{B_0(q)}{A_0(q)}u_0(t)+v(t)\right].\end{aligned} \quad (33)$$

Denote

$$G_0=\frac{B_0(q)}{A_0(q)},\quad \hat{G}=\frac{\hat{B}(q)}{\hat{A}(q)},\quad \Delta G=\frac{B_0(q)}{A_0(q)}-\frac{\hat{B}(q)}{\hat{A}(q)} \quad (34)$$

for convenience. Then one has

$$\begin{aligned}(\varepsilon_{oe}(t-k)\cdot y(t))^2=&\,G_0^2\Delta G^2u_0^2(t)u_0^2(t-k)+G_0^2u_0^2(t)v^2(t-k)\\&+\Delta G^2u_0^2(t-k)v^2(t)\\&+G_0^2\hat{G}^2u_0^2(t)e_u^2(t-k)+\hat{G}^2e_u^2(t-k)v^2(t)\\&+2G_0u_0(t)v(t)v^2(t-k)\\&+2G_0^2\Delta Gu_0^2(t)u_0(t-k)v(t-k)\\&+2G_0\Delta G^2u_0(t)u_0^2(t-k)v(t)\\&-2G_0^2\hat{G}\Delta Gu_0^2(t)u_0(t-k)e_u(t-k)\\&-2G_0^2\hat{G}u_0^2(t)e_u(t-k)v(t-k)\\&-4G_0\hat{G}\Delta Gu_0(t)u_0(t-k)e_u(t-k)v(t)\\&+2\Delta Gu_0(t-k)v^2(t)v(t-k)\\&+4G_0\Delta Gu_0(t)u_0(t-k)v(t)v(t-k)\\&+2G_0\hat{G}^2u_0(t)e_u^2(t-k)v(t)\\&-2\hat{G}\Delta Gu_0(t-k)e_u(t-k)v^2(t)\\&-4G_0\hat{G}u_0(t)e_u(t-k)v(t)v(t-k)\\&-2\hat{G}e_u(t-k)v^2(t)v(t-k)+v^2(t)v^2(t-k).\end{aligned} \quad (35)$$

When $N\to\infty$, one has

$$\lim_{N\to\infty}\frac{1}{N-p}\sum_{t=p+1}^{N}[\varepsilon_{oe}(t-k)\cdot y(t)]^2=E[\varepsilon_{oe}(t-k)\cdot y(t)]^2. \quad (36)$$

Taking the expectation and following the assumptions A4 and A5, ie, $\{u_0(t)\}$, $\{e_u(t)\}$ and $\{v(t)\}$ are mutually uncorrelated one obtains

$$E[\varepsilon_{oe}(t-k)\cdot y(t)]^2 = (G_0^2\Delta G^2)\cdot E[u_0^2(t)u_0^2(t-k)] + E[v^2(t)v^2(t-k)]. \tag{37}$$

Similar procedure can be applied to $(p+k+1)$th term in the sum on the right hand side of Equation 32 and then it follows

$$E[\varepsilon_{oe}(t)\cdot y(t-k)]^2 = (G_0^2\Delta G^2)\cdot E[u_0^2(t)u_0^2(t-k)] + E[v^2(t)v^2(t-k)]. \tag{38}$$

In Equations 37 and 38, for any $0 \leqslant k \leqslant p$, G_0, $E[u_0^2(t)u_0^2(t-k)]$ and $E[v^2(t)v^2(t-k)]$ are all constants, hence $E\{[\varepsilon_{oe}(t-p)\cdot y(t)]^2 + \cdots + [\varepsilon_{oe}(t-1)\cdot y(t)]^2 + [\varepsilon_{oe}(t)\cdot y(t)]^2 + [\varepsilon_{oe}(t+1)\cdot y(t)]^2 + \cdots + [\varepsilon_{oe}(t+p)\cdot y(t)]^2\}$ is minimized when $\Delta G = 0$, which implies consistent estimate in terms of transfer function. Based on this observation, one can estimate the input noise variance as follows

$$\begin{aligned}\hat{\lambda}_{eu} = \arg\min_{\lambda_{eu}} \frac{1}{N-p}\sum_{t=p+1}^{N} \{&[\varepsilon_{oe}(t-p)\cdot y(t)]^2 + \cdots \\ &+ [\varepsilon_{oe}(t-1)\cdot y(t)]^2 \\ &+ [\varepsilon_{oe}(t)\cdot y(t)]^2 + [\varepsilon_{oe}(t)\cdot y(t-1)]^2 + \cdots \\ &+ [\varepsilon_{oe}(t)\cdot y(t-p)]^2\}.\end{aligned} \tag{39}$$

3.3 | Consistent ARX model estimation

Based on Equation 39, one can estimate the input noise variance and model parameters as follows.

Algorithm 1

Step 1. Apply the usual LS method to the available data set $\{y(t), u(t), t = 1, 2, \cdots, N\}$. Obtain the biased parameter estimate $\hat{\theta}_{LS}(N)$.

Step 2. Determine the range of the input noise variance. Denote λ_{min} and λ_{max} as the minimum and maximum value of the input noise variance λ_{eu}, respectively. If there is no knowledge about the range of the input noise variance, λ_{min} can be set to 0 and λ_{max} set to the variance of the measured input. If the range is known, smaller range can be used, which will reduce the cost of computation.

Step 3. Set the input noise variance step size $h = (\lambda_{max} - \lambda_{min})/M$, where $10 < M < 100$ is an integer. Then vary the input noise variance using the step size h in the interval $[\lambda_{min}, \lambda_{max}]$. With each variance $\hat{\lambda}_{eu}(i) = h\times i, i = 1, 2, \ldots, M$, calculate the related estimate $\hat{\theta}_{ARX}(i)$ using equations

$$\hat{D}(i) = diag\{O_{na}; \hat{\lambda}_{eu}(i)I_{nb}\}, \tag{40}$$

$$\hat{\theta}_{ARX}(i) = [I_{na+nb} - \hat{R}_{\psi\psi}^{-1}\hat{D}(i)]^{-1}\hat{\theta}_{LS}(N). \tag{41}$$

then calculate the corresponding loss function $V_\lambda(N)$ with each $\hat{\theta}_{ARX}(i)$. Select the variance λ_{eu}^* when the loss function

$$\begin{aligned}\frac{1}{N-p}\sum_{t=p+1}^{N} \{&[\varepsilon_{oe}(t-p)\cdot y(t)]^2 + \cdots + [\varepsilon_{oe}(t-1)\cdot y(t)]^2 \\ &+ [\varepsilon_{oe}(t)\cdot y(t)]^2 + [\varepsilon_{oe}(t)\cdot y(t-1)]^2 + \cdots \\ &+ [\varepsilon_{oe}(t)\cdot y(t-p)]^2\}\end{aligned} \tag{42}$$

is minimal. Finally, the parameter estimation is given as

$$\theta_{ARX}^* = [I_{na+nb} - \hat{R}_{\psi\psi}^{-1} diag\{O_{na}; \lambda_{eu}^* I_{nb}\}]^{-1}\hat{\theta}_{LS}(N). \tag{43}$$

Because only the input noise variance needs to be determined in the minimization problem, and the range of the input noise variance is often known, the global minimum can be easily obtained using the proposed search algorithm, which costs little computation.

One may say that the ARX model estimation here is closely related to the well-known bias correction method of one study.[15] Equation 43 may be interpreted as a kind of bias correction method, but they are very different because the bias correction method estimates the bias of the LS estimate and subtract it from the LS estimate; the ARX method here estimates the input noise variance and achieves a consistent ARX model estimate using Equation 43. Moreover, the equivalent Equation 30 is not related to the LS estimate at all, and hence, the link to the bias correction method is not really necessary.

The consistency of the ARX model estimation can be stated as follows.

Theorem 1. Consider the system that satisfies Equations 9 to 13 with assumptions A1, A2, A3, A4, and A5 hold. Then the estimate in Equation 43 is consistent, namely,

$$\hat{\theta}_{ARX} \xrightarrow{w.p.1} \theta_0 \quad as \quad N \to \infty. \tag{44}$$

Proof. Based on A1, A2, A4, and A5, when the loss function Equation 32 is minimized, it follows that $\Delta G = 0$ or

$$\hat{G}(q) = G_0(q). \tag{45}$$

Moreover, assumption A3 says that there is no common factor in $A_0(q)$ and $B_0(q)$, then Equation 45 means that

$$\hat{A}(q) = A_0(q), \quad \hat{B}(q) = B_0(q). \tag{46}$$

This leads to Equation 44. End of proof. □

4 | ASYMPTOTIC PROPERTIES OF THE ARX MODEL

After obtaining an consistent ARX model estimation of EIV systems, to follow the asymptotic method of identification, one needs to extend the asymptotic theory for noise-free input systems to EIV systems. For us to keep the work simple, only the asymptotic properties of the ARX model are studied. Denote the ARX model parameter vector as

$$\theta^n = [a_1, b_1, \cdots, a_n, b_n]^{\mathrm{T}}, \tag{47}$$

where n is the model order. Note that, for the convenience of the analysis, the parameters are ordered differently from that in Section 3.

Using the ARX model estimation of Section 3, one can obtain the ARX model parameters

$$\hat{\theta}^n_{ARX} = \left[\frac{1}{N}\sum_{t=n+1}^{N}\varphi_t\varphi_t^{\mathrm{T}} - \hat{D}^n\right]^{-1}\left[\frac{1}{N}\sum_{t=n+1}^{N}\varphi_t y(t)\right], \tag{48}$$

where

$$\varphi_t = [-y(t-1), u(t-1), \cdots, -y(t-n), u(t-n)]^{\mathrm{T}}, \tag{49}$$

$$\hat{D}^n = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & \hat{\lambda}_{eu} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \hat{\lambda}_{eu} \end{bmatrix}_{2n\times 2n}. \tag{50}$$

Note that the data vector is also ordered differently from that in Section 3. For the parameter estimate $\hat{\theta}^n_{ARX}$, the frequency response estimates are

$$\begin{aligned}\hat{G}^n_N(e^{i\omega}) &= G(e^{i\omega}, \hat{\theta}^n_{ARX}),\\ \hat{H}^n_N(e^{i\omega}) &= H(e^{i\omega}, \hat{\theta}^n_{ARX}).\end{aligned} \tag{51}$$

For the identifiability of the system, it is assumed that the noise-free input signal has a spectrum that is non-0 at all frequencies. Some further conditions on the system and the model order are stated as follows.

C1. $n(N) \to \infty$, as $N \to \infty$.
C2. $n(N)^{3+\delta}/N \to 0$, for some $\delta > 0$, as $N \to \infty$.
C3. $\sqrt{N}\left[\sum_{k=n(N)+1}^{\infty} |a_k| + |b_k|\right] \to \infty$, as $N \to \infty$.

Remark 1. The key idea that leads to the asymptotic theory is to allow the model order to increase as the number of data increases, this is given in C1. However, the above conditions also mean that one does not use too high ARX model order compared to the number of observations, this is given by C2. Condition C3 depends on the ARX model and gives a lower bound on the order of the fitted model.

Now we have the following results.

Lemma 1. Given the ARX model estimation in Equation 48. Under the assumptions A1, A4, A5, C1, C2, and C3, further assume that in the ARX model estimation a given consistent estimate of input noise variance is used. Then, we have

$$\sqrt{N}(\hat{\theta}^n_{ARX} - \theta^n) \in AsN(0, P_{ARX}), \tag{52}$$

where

$$P_{ARX} = \lambda_{ey} \cdot [R^n(N) - D^n]^{-1} R^n(N) [R^n(N) - D^n]^{-1} \tag{53}$$

and

$$R^n(N) = \frac{1}{N}\sum_{t=n+1}^{N}\varphi_t\varphi_t^{\mathrm{T}}. \tag{54}$$

Remark 2. For simplicity, this result is developed for fixed input noise variance $\hat{\lambda}_{eu}$, it would be technically very complex to include the variance of $\hat{\lambda}_{eu}$ in the analysis.

The proof is given in Appendix A.

Theorem 2. Consider the frequency response estimates of the process model $\hat{G}^n_N(e^{i\omega})$ and the disturbance model $\hat{H}^n_N(e^{i\omega})$. Assume that the system is under open-loop test and conditions A1, A4, A5, C1, C2, and C3 hold, also assume that in the ARX model estimation a given consistent estimate of input noise variance is used. Then, the errors of the 2 frequency responses follow Gaussian distributions with 0 means as $N \to \infty$, and their variances are given as

$$\operatorname{var} \hat{G}^n_N(e^{i\omega}) \sim \frac{n}{N}\frac{\Phi_v(\omega)\Phi_u(\omega)}{[\Phi_u(\omega) - \lambda_{eu}]^2}, \tag{55}$$

$$\operatorname{var} \hat{H}^n_N(e^{i\omega}) \sim \frac{n}{N}\frac{\Phi_v(\omega)}{\lambda_{ey}} = \frac{n}{N}|H(e^{i\omega})|^2, \tag{56}$$

where $\Phi_v(\omega)$ and $\Phi_u(\omega)$ are the spectra of $v(t)$ and $u(t)$, respectively.

The proof is given in Appendix B.

5 | ASYMPTOTIC METHOD OF IDENTIFICATION OF EIV SYSTEMS

It is now ready to present the ASYM for EIV systems in the Box-Jenkins structure. When the input-output data set is given, the ASYM aims to solve 3 basic problems: (1) parameter estimation, (2) order selection, and (3) model validation. The estimation part consists of 2 steps: a high-order ARX model estimation and then a frequency weighted model reduction. In parameter estimation, the idea is to first estimate a simple but high-order ARX model to obtain a practically unbiased frequency response of the more complex but low order Box-Jenkins system, then model reduction is used to reduce the variance error of the high order ARX model.

5.1 | Estimate a high order ARX model

Using the method of ARX model estimation in Section 3, estimate an ARX model with an order much higher than the Box-Jenkins model. For systems with order less than 10, the order of the high-order ARX model can be in the range of [10, 50]. Experience has shown that the obtained final model after the model reduction step is not sensitive to the order of the ARX model. A theoretically sound approach of model order selection remains a research topic. The aim here is to arrive at a consistent estimate of the frequency response of the process.

Denote $\hat{G}^n_{ARX}(q)$ as the estimate of the process and $\hat{H}^n_{ARX}(q)$ of the disturbance.

$$\hat{G}^n_{ARX}(q) = \frac{\hat{B}^n_{ARX}(q)}{\hat{A}^n_{ARX}(q)} \qquad \hat{H}^n_{ARX}(q) = \frac{1}{\hat{A}^n_{ARX}(q)}, \tag{57}$$

where $\hat{A}^n_{ARX}(q)$ and $\hat{B}^n_{ARX}(q)$ are the polynomials of the modified estimate, they are both with order n.

5.2 | Model reduction

Employing the high-order ARX model structure in Equation 57, the process model $G(q)$ is often over-parameterized, thus leads to high variance owing to the high order n. Here, we intend to reduce the variance by performing model reduction on the high order model.

Denote the model parameters with reduced order as

$$\theta^l = [a_1, \cdots, a_{la}, b_1, \cdots, b_{lb}]^{\mathrm{T}}, \tag{58}$$

where la and lb are the reduced order of the process polynomials, respectively.

The asymptotic theory of Section 4 shows that in frequency domain, the high-order model follows approximate Gaussian distribution. The frequency response of the high-order model can be viewed as data with the Gaussian distribution and known variance. Hence, the maximum likelihood can be used in the calculation of the reduced order model, see 1 study.[22] With the asymptotic results of Equation 55, we obtain the asymptotic negative log-likelihood function of the process model as follows:

$$V = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{G}^n_{ARX}(e^{i\omega}) - \hat{G}(e^{i\omega})|^2 \times \frac{[\Phi_u(\omega) - \lambda_{eu}]^2}{\Phi_v(\omega)\Phi_u(\omega)} d\omega. \tag{59}$$

The reduced model $\hat{G}(q)$ can be obtained by minimizing the asymptotic negative log-likelihood function for a fixed order

$$\begin{aligned} \hat{\theta}^l = \arg\min_{\theta^l} \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{G}^n_{ARX}(e^{i\omega}) - \hat{G}(e^{i\omega})|^2 \\ \times \frac{[\Phi_u(\omega) - \lambda_{eu}]^2}{\Phi_v(\omega)\Phi_u(\omega)} d\omega. \end{aligned} \tag{60}$$

The required spectra $\Phi_u(\omega)$ and $\Phi_v(\omega)$ can be estimated using input data $u(t)$ and output error residual $\hat{v}(t)$, and the variance λ_{eu} can be estimated with the method in Section 3.2. The same model reduction procedure can be done for the disturbance model $\hat{H}(q)$.

Firstly, the frequency response of the high-order ARX model and the related power spectra are calculated for $m > n$ frequency points uniformly distributed in $[0, \pi]$. Then the frequency domain data are used to minimize the above loss function. Many numerical optimization methods can be used in the minimization of the above loss function, for example, the well-known Gauss-Newton method. Good initial estimate is important for the global convergence of the optimization. Here, we can suggest at least 2 methods: (1) obtain the initial estimate using a balance model reduction on the high order ARX model; (2) simulate the high-order ARX model using the measured input and then estimate an initial model using an instrumental variable method.

Compared with bias correction methods and instrumental variable methods, the ASYM developed here will be much more costly in computation. However, there is big benefit with this high cost, namely, the ASYM is consistent and probably also efficient (minimum variance). The reason for the consistency is that the high-order ARX model is consistent in view of system transfer function; the reason for asymptotic efficiency is that model reduction is a maximum likelihood estimation, which is efficient, and the high-order ARX model is probably also efficient. When both steps are efficient, then the final model will be efficient. More detailed discussion on this can be found in 1 study.[22]

5.3 | Model order selection using asymptotic criterion

The order of the reduced model is determined using a frequency domain criterion asymptotic criterion as in 1 study.[18] The basic idea of the criterion is to choose the best order, which makes the frequency domain difference between the high order model and the reduced model approximately equal to the variance of the high order model, ie,

$$|\hat{G}^n_{ARX}(e^{i\omega}) - \hat{G}(e^{i\omega})|^2 \approx \frac{n}{N}\frac{\Phi_v(\omega)\Phi_u(\omega)}{[\Phi_u(\omega) - \lambda_{eu}]^2}. \tag{61}$$

Then the best order can be get through minimizing the following asymptotic criterion:

$$\begin{aligned} ASYC = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Bigg| |\hat{G}^n_{ARX}(e^{i\omega}) - \hat{G}(e^{i\omega})|^2 \\ - \frac{n}{N}\frac{\Phi_v(\omega)\Phi_u(\omega)}{[\Phi_u(\omega) - \lambda_{eu}]^2}\Bigg| d\omega. \end{aligned} \tag{62}$$

The idea of this criterion is to obtain a model with small error in the frequency domain. It remains a heuristic criterion, and there is no analysis done about it.

5.4 | Model validation using upper error bound

Based on the asymptotic theory of Section 4, the errors of the high-order model frequency response follow an asymptotic Gaussian distribution with a given variance as in Equation 55. Hence, a 3δ upper bound of the errors of the high-order model can be derived as

$$\begin{aligned} &\left|G_0(e^{i\omega}) - \hat{G}^n_{ARX}(e^{i\omega})\right| \\ &\leqslant 3\sqrt{\frac{n}{N}\frac{\Phi_v(\omega)\Phi_u(\omega)}{[\Phi_u(\omega) - \lambda_{eu}]^2}} \quad w.p.\ 99.99\%. \end{aligned} \tag{63}$$

As the reduced model $\hat{G}(e^{i\omega})$ has a smaller variance than $\hat{G}^n_{ARX}(e^{i\omega})$, we can also use this upper bound for $\hat{G}(e^{i\omega})$ as follows:

$$\left|G_0(e^{i\omega}) - \hat{G}(e^{i\omega})\right| \leqslant 3\sqrt{\frac{n}{N}\frac{\Phi_v(\omega)\Phi_u(\omega)}{[\Phi_u(\omega)-\lambda_{eu}]^2}} \quad w.p. \ \geqslant 99.99\%. \tag{64}$$

6 | SIMULATIONS AND COMPARISONS

Example 1, ARX model estimation

Firstly, the accuracy of the ARX model estimation will be verified. This example gives a numerical illustration for the ARX model estimation method proposed in Section 3. Consider a second-order ARX EIV system

$$y(t) = \frac{q^{-1}+0.5q^{-2}}{1-1.6q^{-1}+0.7q^{-2}}u_0(t) + \frac{1}{1-1.6q^{-1}+0.7q^{-2}}e_y(t), \tag{65}$$

$$u(t) = u_0(t) + e_u(t). \tag{66}$$

The noise-free input signal $u_0(t)$ is a generalized binary noise (GBN) sequence (see 1 study[23]) with 0 mean and variance of 1.00, and the mean switching time of the GBN signal is 10. The noise-free output has a variance of 179.4793. The input noise $e_u(t)$ is a white noise sequence with variance 0.20, and the output noise is a colored noise sequence with variance 35.90. The noise-to-signal ratios at the input and at the output are both about 20% in power. Hence, one can say that the measured data are very noisy. In total, 100 simulation runs are conducted, and models are identified. The number of data samples of each simulation is N = 3000 each. During the model estimation procedure, we use traversal operators to search the input variance, and the step size is set $h = 0.01$, and the traversal times i varies from 1 to 100. The parameter $P = 20$ in evaluating the loss function (32).

For comparison, ARX models using noise-free input data are also estimated.

Table 1 gives the results obtained from 100 simulations. In the table, the first term of a parameter is the mean value of the 100 estimations of that parameter, the second term is the standard deviation of the 100 estimations.

One can see that the proposed ARX model estimation method gives much better estimate than the common least squares method in terms of bias; its variance is close to the ARX model using the noise-free input data, which is minimum variance estimate.

Figure 3 shows the curve of the parameter error as a function of the input noise variance and the loss function value as a function of the input noise variance. The parameter error is defined as

$$\frac{\|\hat{\theta}_{ARX}(i)-\theta_0\|_2}{\|\theta_0\|_2}. \tag{67}$$

One can see that the input noise variance with which the loss function value gets to its minimum is very close to the value that minimizes the parameter estimation error.

In the following, we will compare the ASYM with 2 well-known methods in the literature.

Example 2, Box-Jenkins model and comparison with the BELS-II

This example gives a numerical illustration for the ASYM of a general EIV system. Only parameter estimation is tested. Consider a second-order EIV system with Box-Jenkins model structure

$$y(t) = \frac{q^{-1}+0.5q^{-2}}{1-1.5q^{-1}+0.7q^{-2}}u_0(t) + \frac{1-0.6q^{-1}+0.4q^{-2}}{1-1.9q^{-1}+0.95q^{-2}}e_y(t), \tag{68}$$

$$u(t) = u_0(t) + e_u(t). \tag{69}$$

In this example, the noise-free input signal $u_0(t)$ is also a GBN sequence with 0 mean, variance of 1.00, and mean switching

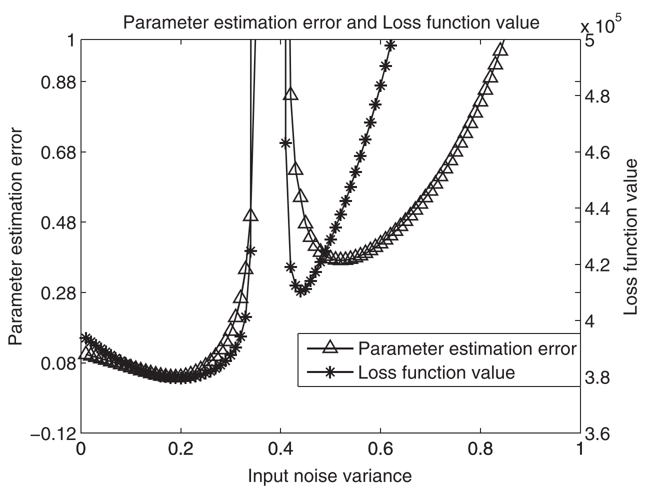

FIGURE 3 Parameter error curve and loss function curve of Example 1

TABLE 1 Simulation results of Example 1

Method	a_1	a_2	b_1	b_2	λ_{eu}
LS method	-1.6294 ± 0.0091	0.7228 ± 0.0081	0.7839 ± 0.0335	0.4844 ± 0.0351	/
ARX method	-1.6002 ± 0.0108	0.7004 ± 0.0095	1.0018 ± 0.0691	0.4996 ± 0.0696	0.1976 ± 0.0255
LS without input noise	-1.6002 ± 0.0083	0.7004 ± 0.0074	0.9982 ± 0.0415	0.5042 ± 0.0483	/
True value	-1.6000	0.7000	1.0000	0.5000	0.2000

Abbreviation: LS, least squares.

TABLE 2 Simulation results of Example 2

Method	a_1	a_2	b_1	b_2	λ_{eu}	per run time(s)
BELS-II method	-1.5005 ± 0.0386	0.7009 ± 0.0298	1.0109 ± 0.1991	0.4910 ± 0.1153	/	0.07
ASYM method	-1.5006 ± 0.0107	0.6999 ± 0.0093	0.9886 ± 0.0388	0.4947 ± 0.0495	0.1823 ± 0.0117	1.21
True value	-1.5000	0.7000	1.0000	0.5000	0.2000	/

Abbreviations: BELS, bias-eliminated least squares; ASYM, asymptotic method.

time of 10. The input noise $e_u(t)$ is also a white noise sequence with variance 0.20. The noise-free output has variance 59.64, and the output noise is a colored noise sequence with variance 11.93. So the noise-to-signal ratios at the input and at the output are both 20% in power. One hundred simulation runs are conducted, and the number of samples is N = 4000. The parameter $P = 15$ in evaluating the loss function (32).

In ASYM parameter estimate, firstly, a high-order ARX model with order 20 is estimated. In the input noise variance estimation, we also use traversal operators to search the input variance, where the step size is $h = 0.01$ and the traversal times i varies from 1 to 100. Then the model reduction is performed by minimizing the loss function (59). For comparison, we have implemented the so-called BELS-II method in 1 study[15] and verified it with the examples in 1 study[15]; we estimated the model using the same simulation data here.

Table 2 shows the results of the ASYM and the BELS-II method. For each parameter in the table, the first term is the mean value of the 100 estimations of the parameter and the second term is the standard deviation of the 100 estimations. One can see that the ASYM gives much better estimate than the BELS-II method in view of the mean values and standard deviations. Considering the computational load, the ASYM is 16 times slower than the BELS-II method, which is a disadvantage.

Figure 4 shows the impulse response parameter error of the estimated high-order ARX model as a function of the input noise variance and the loss function value as a function of the input noise variance. The impulse response parameter error is defined as

$$\frac{\|\hat{\theta}^n_{FIR}(i) - \theta^0_{FIR}\|_2}{\|\theta^0_{FIR}\|_2}, \tag{70}$$

where $\hat{\theta}^n_{FIR}(i)$ is the finite impulse response model parameter of the high-order ARX model and θ^0_{FIR} is the finite impulse response model parameter of the true process. The truncated time of both parameters is set to be 50, which covers the transient time of the process.

One can see that the input noise variance with which the loss function value gets to its minimum is very close to the value that minimizes the impulse response parameter estimation error.

Example 3, Comparison with IV-WSF method

To compare the ASYM method and the so-called instrumental variable with subspace fitting (IV-WSF) method of 1 study,[12] we applied the ASYM to the numerical example 1 of 1 study[12] as follows. Consider an ARMA system

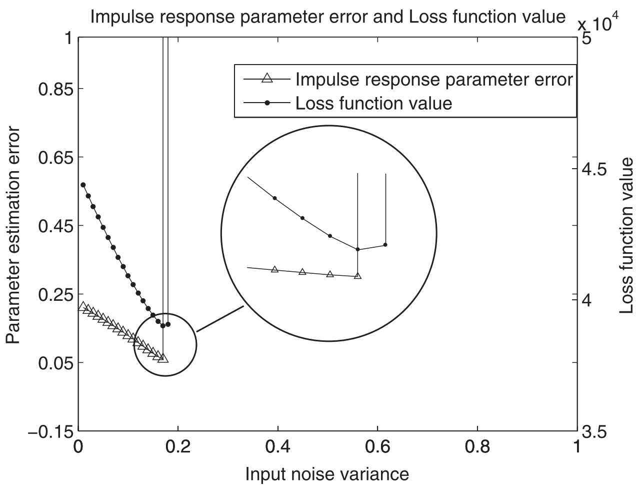

FIGURE 4 Impulse response parameter error curve and loss function curve of Example 2

$$A(q^{-1}) = 1 - 1.5q^{-1} + 0.7q^{-2} \tag{71}$$

$$B(q^{-1}) = 1 - q^{-1} + 0.5q^{-2} \tag{72}$$

In 1 study,[12] the noise-free input signal is generated as the following moving average:

$$u_0(t) = (0.7 + 0.8q^{-1} + 0.9q^{-2} + q^{-3} + 0.9q^{-4} + 0.8q^{-5} + 0.7q^{-6})e_x(t), \tag{73}$$

where $e_x(t)$ is a white noise of unit variance. And the input and output noises are the following first-order moving averages, respectively:

$$w(t) = (1 + 0.5q^{-1})e_w(t), \tag{74}$$

$$v(t) = (1 + 0.5q^{-1})e_v(t), \tag{75}$$

where $e_w(t)$ and $e_v(t)$ are white noises that are uncorrelated with each other. The variances of $e_w(t)$ and $e_v(t)$ are selected so as to get an signal-to-noise ratio (SNR) equal to 20 dB for both the input and output signals. The input SNR is defined as $SNR_i = 10log\{E[u_0^2(t)]/E[w^2(t)]\}$ and an analogous definition is used for the output SNR: $SNR_o = 10log\{E[y_0^2(t)]/E[v^2(t)]\}$. Note that the input noise is auto-correlated with lag 1, which is not exactly white noise but close.

Using the same conditions as in 1 study,[12] the parameters of the above system are estimated using the ASYM and compared with the IV-WSF method. One hundred Monte-Carlo simulations are used to determine the standard deviations (stds) of the various parameter estimates. Figure 5 shows the stds of the 2 methods:

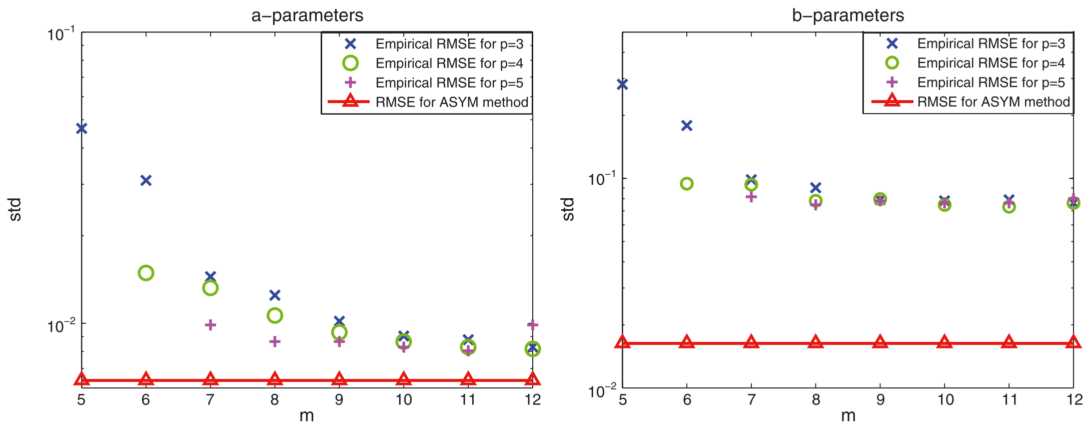

FIGURE 5 Root mean square error (RMSE) values for IV-WSF method and asymptotic method. std, standard deviation [Colour figure can be viewed at wileyonlinelibrary.com]

$$std(\hat{\alpha}) = [std(\hat{\alpha}_1) + std(\hat{\alpha}_2)]/2 \quad (76)$$

and similarly for the *b*-parameters. Here, in this figure *P* is a tuning parameter of the IV-WSF method, which is the maximum delay in calculating certain correlation function.

From the result one can see that the stds of a-parameters and b-parameters of the ASYM model are considerably smaller than those of the IV-WSF model, which implies that the ASYM model is much more accurate than the IV-WSF model.

7 | CONCLUSIONS

The problem of EIV systems identification is studied where the system has a Box-Jenkins model structure. For given input-output data, 3 basic identification problems are solved in this work: (1) parameter estimation, (2) model order selection, and (3) model validation. An ASYM is developed to address the problems. First, a consistent and probably efficient method of ARX model estimation of EIV systems is developed. Then the asymptotic theory of noise-free input system is extended to EIV systems. Based on these 2 results, the ASYM for noise-free input systems can be naturally extended to EIV systems. The obtained ASYM model is consistent. The high accuracy of the ASYM has been demonstrated using simulation examples. The cost of computation of the ASYM is much higher than the compared methods, which is a disadvantage. One theoretical problem remains to be solved in this approach: An order selection criterion for the high-order ARX model. This problem is under investigation. The efficiency of the ASYM raises another interesting theoretical question, which needs to be investigated.

ACKNOWLEDGMENTS

The authors thank the reviewers for their comments and suggestions that have helped to improve the work. This work is supported by the National Science Foundation of China (grant 61273191) and by 973 Program of China (grant 2012CB720500).

REFERENCES

1. Torsten S. Errors-in-variables methods in system identification. *Autom* 2007; 43(6):939–958.
2. Fernando, Vince K, Nicholson H. Identification of linear systems with input and output noise: the Koopmans-Levin method. *IEE Proc D (IET Control Theory Appl)* 1989; 132(1):30–36.
3. Torsten S. Identification of stochastic linear systems in presence of input noise. *Autom* 1981; 17(5):713–725.
4. Frisch R. Statistical confluence analysis by means of complete regression sys-tems. In Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway, 1934.
5. Guidorzi R Identification of multivariable process in the Frisch scheme context. *MTNS 96*, St Louis, USA, 1996.
6. Torsten S, Soverini U, Mahata K. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Process* 2002; 82(8):1139–1154.
7. Diversi R, Guidorzi R, Soverini U A new criterion in EIV identification and filtering applications. *In 13th IFAC symposium on system identification*, Rotterdam, The Netherlands, August 27-29, 2003.
8. Diversi R, Guidorzi R, Soverini U. The Frisch scheme in algebraic and dynamic identification problems. *Kybernetika* 2008; 44(5):585–616.
9. Söderström T. Extending the Frisch scheme for errors-in-variables identification to correlated output noise. *Int J Adapt Control Signal Process* 2008; 22(1):55–73.
10. Diversi R, Guidorzi R, Soverini U. Identification of ARX and ARARX models in the presence of input and output noises. *Eur J Control* 2010; 16(3):242–255.
11. Cedervall M, Stoica P. System identification from noisy measurements by using instrumental variables and subspace fitting. *Circuits Syst Signal Process* 1996; 15(2):275–290.
12. Stoica P, Cedervall M, Eriksson A. Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. *IEEE Trans Signal Process* 1995; 43(10):2386–2397.
13. Soderstrom T. A generalized instrumental variable estimation method for errors-in-variables identification problems. *Autom* 2011; 47(8): 1656–1666.
14. Zheng W-X, Feng C-B. Unbiased parameter estimation of linear systems in the presence of input and output noise. *Int J Adapt Control Signal Process* 1989; 3(3):231–251.
15. Zheng WX. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Trans Autom Control* 2002; 47(7):1142–1147.
16. Shen Y, Li B, Chen Y. An iterative solution of weighted total least-squares adjustment. *J Geod* 2011; 85(4):229–238.
17. Pešta M. Unitarily invariant errors-in-variables estimation. *Stat Pap* 2016; 57(4) :1–17.

18. Zhu YC. Multivariable process identification for MPC: the asymptotic method and its applications. *J Process Control* 1998; 8(2):101–115.
19. Ljung L. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans Autom Control* 1985; 30(9):834–844.
20. Zhu YC. Black-box identification of MIMO transfer functions: asymptotic properties of prediction error models. *Int J Adapt Control Signal Process* 1989; 3:357–373.
21. Liu X, Zhu YC. A Method of ARX Model Estimation of Errors-in-Variables Systems. *IFAC-PapersOnLine* 2015; 48(28):302–306.
22. Wahlberg B. Model reductions of high-order estimated models: the asymptotic ML approach. *Int J Control* 1989; 49(1):169–192.
23. Tulleken H. Generalized binary noise test-signal concept for improved identification-experiment design. *Autom* 1990; 26(1):37–49.
24. Ljung L, Bo W. Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Adv Appl Probab* 1992; 24(2):412–440.
25. Ljung L, Yuan ZD. Asymptotic properties of black-box identification of transfer functions. *IEEE Trans Automat Contr* 1985; 30:514–530.
26. Yuan Z, Ljung L. Black-box identification of multivariable transfer functions-Asymptotic properties and optimal input design. *Int J Contr* 1984; 40(2):233–256.

How to cite this article: Liu X, Zhu Y. Identification of errors-in-variables systems: An asymptotic approach. *Int J Adapt Control Signal Process*. 2017. doi:10.1002/acs.2751

APPENDIX A

PROOF OF LEMMA 1

Denote

$$\phi_t\phi_t^{\mathrm{T}} = \sum_{t=n+1}^{N} \varphi_t\varphi_t^{\mathrm{T}} \qquad \phi_t Y = \sum_{t=n+1}^{N} \varphi_t y(t) \tag{A1}$$

The estimated high-order ARX model parameters can be rewritten as

$$\hat{\theta}^n_{ARX} = (\varphi_t\varphi_t^{\mathrm{T}} - N\hat{D}^n)^{-1}\varphi_t Y. \tag{A2}$$

For convenience we denote

$$\xi_t = [0, e_u(t-1), \cdots, 0, e_u(t-n)]^{\mathrm{T}} \tag{A3}$$

Under ARX model structure, combine Equations 47, 49, and A3 we can obtain that

$$y(t) = \varphi_t^{\mathrm{T}}\theta^n - \xi_t^{\mathrm{T}}\theta^n + e_y(t). \tag{A4}$$

Put Equation A4 into Equation 48 ,we have

$$\left[\frac{1}{N}\sum_{t=n+1}^{N}\varphi_t\varphi_t^{\mathrm{T}} - \hat{D}^n\right]\hat{\theta}^n_{ARX} = \left[\frac{1}{N}\sum_{t=n+1}^{N}\varphi_t\varphi_t^{\mathrm{T}} - \frac{1}{N}\sum_{t=n+1}^{N}\varphi_t\xi_t^{\mathrm{T}}\right]\theta^n + \frac{1}{N}\sum_{t=n+1}^{N}\varphi_t e_y(t). \tag{A5}$$

Under the assumptions A4 and A5, and we further assume that in the ARX model estimation a given consistent estimate of input noise variance is used, then we obtain

$$\hat{\theta}^n_{ARX} - \theta^n = \left[\frac{1}{N}\sum_{t=n+1}^{N}\varphi_t\varphi_t^{\mathrm{T}} - D^n\right]^{-1}\left[\frac{1}{N}\sum_{t=n+1}^{N}\varphi_t e_y(t)\right]. \tag{A6}$$

With the definition of the covariance matrix, we have

$$\operatorname{cov}\hat{\theta}^n_{ARX} = E\left\{\left[\hat{\theta}^n_{ARX} - E(\hat{\theta}^n_{ARX})\right]\left[\hat{\theta}^n_{ARX} - E(\hat{\theta}^n_{ARX})\right]^{\mathrm{T}}\right\}. \tag{A7}$$

As mentioned in previous sections, the high-order ARX model estimation is consistent, ie, $E(\hat{\theta}^n_{ARX}) = \theta^n$, then after simple computation we obtain that

$$\operatorname{cov}\hat{\theta}^n_{ARX} = \left(\phi_t\phi_t^{\mathrm{T}} - ND^n\right)^{-1}\phi_t E\left(\varepsilon_y\varepsilon_y^{\mathrm{T}}\right)\phi_t^{\mathrm{T}}\left(\phi_t\phi_t^{\mathrm{T}} - ND^n\right)^{-1}, \tag{A8}$$

where

$$\varepsilon_y = [e_y(n+1), e_y(n+2), \cdots, e_y(N)]^{\mathrm{T}}. \tag{A9}$$

With the properties of the white noise $e_y(t)$ and after simple computations, see also Theorem 7.1 and Theorem 7.2 of 1 study,[24] we have

$$\sqrt{N}(\hat{\theta}^n_{ARX} - \theta^n) \in AsN(0, P_{ARX}) \tag{A10}$$

and

$$P_{ARX} = \lambda_{ey}\cdot[R^n(N) - D^n]^{-1}R^n(N)[R^n(N) - D^n]^{-1} \tag{A11}$$

End of proof. □

APPENDIX B

PROOF OF THEOREM 2

For convenience, denote $\hat{A}^n_{ARX}(e^{i\omega})$ and $\hat{B}^n_{ARX}(e^{i\omega})$ as the polynomials of the estimated high-order ARX model. Similarly, $A^n(e^{i\omega})$ and $B^n(e^{i\omega})$ are the polynomials of the high-order ARX model denoted by θ^n.

Denote

$$W_n(\omega) = [e^{-i\omega}I_2, e^{-i2\omega}I_2, \cdots, e^{-in\omega}I_2] \tag{B1}$$

With simple computation, we have

$$\begin{bmatrix}\hat{A}^n_{ARX}(e^{i\omega}) - A^n(e^{i\omega})\\ \hat{B}^n_{ARX}(e^{i\omega}) - B^n(e^{i\omega})\end{bmatrix} = W_n(\omega)\left(\hat{\theta}^n_{ARX} - \theta^n\right), \tag{B2}$$

Under the condition of Lemma 1, we obtain that

$$\sqrt{\frac{N}{n}}\begin{bmatrix}\hat{A}^n_{ARX}(e^{i\omega}) - A^n(e^{i\omega})\\ \hat{B}^n_{ARX}(e^{i\omega}) - B^n(e^{i\omega})\end{bmatrix} \in AsN(0, P_W), \tag{B3}$$

where

$$P_W = \frac{1}{n}\lambda_{ey}\cdot W_n(\omega)[R^n(N)-D^n]^{-1} \times R^n(N)[R^n(N)-D^n]^{-1}W_n(-\omega)^{\mathrm{T}}. \tag{B4}$$

Denote

$$M(e^{i\omega}) = H(e^{i\omega})\begin{bmatrix} -G(e^{i\omega}) & 1 \\ -H(e^{i\omega}) & 0 \end{bmatrix} \tag{B5}$$

We have

$$\begin{bmatrix} \hat{G}^n_{ARX}(e^{i\omega}) - G(e^{i\omega}) \\ \hat{H}^n_{ARX}(e^{i\omega}) - H(e^{i\omega}) \end{bmatrix} = M(e^{i\omega})\begin{bmatrix} \hat{A}^n_{ARX}(e^{i\omega}) - A^n(e^{i\omega}) \\ \hat{B}^n_{ARX}(e^{i\omega}) - B^n(e^{i\omega}) \end{bmatrix}. \tag{B6}$$

Similarly, we obtain

$$\sqrt{\frac{N}{n}}\begin{bmatrix} \hat{G}^n_{ARX}(e^{i\omega}) - G(e^{i\omega}) \\ \hat{H}^n_{ARX}(e^{i\omega}) - H(e^{i\omega}) \end{bmatrix} \in AsN(0, P_T), \tag{B7}$$

where

$$P_T = \frac{1}{n}\lambda_{ey}\cdot M(e^{i\omega})W_n(\omega)[R^n(N)-D^n]^{-1}R^n(N)[R^n(N)-D^n]^{-1}, W_n(-\omega)^{\mathrm{T}}M(e^{i\omega})^{\mathrm{T}}. \tag{B8}$$

As $R^n(N) - D^n$ and $R^n(N)$ are both block Toeplitz covariance matrices. According to the properties of block Toeplitz covariance matrices, see 3 studies,[19,20,24] we have

$$\frac{1}{n}W_n(\omega)[R^n(N)-D^n]^{-1}W_n(-\omega)^{\mathrm{T}} = \begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu_0}(\omega) \\ -\Phi_{u_0y}(\omega) & \Phi_{u_0}(\omega) \end{bmatrix}^{-1}, \tag{B9}$$

$$\frac{1}{n}W_n(\omega)R^n(N)W_n(-\omega)^{\mathrm{T}} = \begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu}(\omega) \\ -\Phi_{uy}(\omega) & \Phi_u(\omega) \end{bmatrix}. \tag{B10}$$

Combine Equations 49, 50 and 54, and assumptions A4 and A5 hold, as $N \to \infty$, we obtain

$$R^n(N) = \begin{bmatrix} \chi_{1,1} & \chi_{1,2} & \cdots & \chi_{1,n} \\ \chi_{2,1} & \chi_{2,2} & \cdots & \chi_{2,n} \\ \vdots & \vdots & & \vdots \\ \chi_{n,1} & \chi_{n,2} & \cdots & \chi_{n,n} \end{bmatrix}, \tag{B11}$$

where the k, l-block element is

$$\chi_{k,l} = \begin{bmatrix} r_{yy}(l-k) & -r_{yu}(l-k) \\ -r_{uy}(l-k) & r_{uu}(l-k) \end{bmatrix} \tag{B12}$$

and $r_*(\tau)$ denotes the correlation function of variables, ie, $r_{yy}(l-k) = \frac{1}{N}\sum_{t=n+1}^{N} y(t-k)y(t-l)$ and $r_{yu}(l-k) = \frac{1}{N}\sum_{t=n+1}^{N} y(t-k)u(t-l)$.

Similarly, denote the s, p-block element of $[R^n(N) - D^n]^{-1}$ is $\zeta_{s,p}$

$$[R^n(N)-D^n]^{-1} = \begin{bmatrix} \zeta_{1,1} & \zeta_{1,2} & \cdots & \zeta_{1,n} \\ \zeta_{2,1} & \zeta_{2,2} & \cdots & \zeta_{2,n} \\ \vdots & \vdots & & \vdots \\ \zeta_{n,1} & \zeta_{n,2} & \cdots & \zeta_{n,n} \end{bmatrix}. \tag{B13}$$

For convenience, denote the product of $[R^n(N) - D^n]^{-1}$ and $R^n(N)$ as Q^n, with simple computation, we obtain the s, l-block element

$$Q^n(s,l) = \sum_{m=1}^{n} \zeta_{s,m}\chi_{m,l} \tag{B14}$$

Combine Equations B11, B12, and B13, the element of matrix Q^n is the convolution of the correlation function of processes $[-y(t)u_0(t)]$ and $[-y(t)u(t)]$. With the same computation, it is obtained that the element of the product of Q^n and $[R^n(N) - D^n]^{-1}$ (denote as T^n) also has the convolution form in time domain.

$$T^n(i,j) = \sum_{p=1}^{n}\sum_{m=1}^{n} \zeta_{i,m}\chi_{m,p}\zeta_{p,j} = \sum_{p=1}^{n} Q^n(i,p)\zeta_{p,j}. \tag{B15}$$

Follow the steps of Lemma 1 of Zhu,[20] see also Lemma 4.2 of Ljung[25] and Lemma 4.3 ofYuan,[26] we have

$$\begin{aligned} &\frac{1}{n}W_n(\omega)[R^n(N)-D^n]^{-1}R^n(N)[R^n(N)-D^n]^{-1}W_n(-\omega)^{\mathrm{T}} \\ &= \begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu_0}(\omega) \\ -\Phi_{u_0y}(\omega) & \Phi_{u_0}(\omega) \end{bmatrix}^{-1}\begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu}(\omega) \\ -\Phi_{uy}(\omega) & \Phi_u(\omega) \end{bmatrix}\begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu_0}(\omega) \\ -\Phi_{u_0y}(\omega) & \Phi_{u_0}(\omega) \end{bmatrix}^{-1}. \end{aligned} \tag{B16}$$

From Equations 1, 2, and 5

$$\begin{aligned} \begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu_0}(\omega) \\ -\Phi_{u_0y}(\omega) & \Phi_{u_0}(\omega) \end{bmatrix}^{-1} &= |H(e^{i\omega})|^2 M(e^{i\omega})^{-1}\begin{bmatrix} \Phi_{u_0}(\omega) & \Phi_{u_0e_y}(\omega) \\ \Phi_{e_yu_0}(\omega) & \lambda_{ey} \end{bmatrix}^{-1} M^{\mathrm{T}}(e^{i\omega})^{-1} \\ &=: |H(e^{i\omega})|^2 M(e^{i\omega})^{-1}\Gamma_1 M^{\mathrm{T}}(e^{i\omega})^{-1}, \end{aligned} \tag{B17}$$

$$\begin{aligned} \begin{bmatrix} \Phi_y(\omega) & -\Phi_{yu}(\omega) \\ -\Phi_{uy}(\omega) & \Phi_u(\omega) \end{bmatrix} &= |H(e^{i\omega})|^{-2} M^{\mathrm{T}}(e^{i\omega})\begin{bmatrix} \Phi_u(\omega) & \Phi_{ue_y}(\omega) \\ \Phi_{e_yu}(\omega) & \lambda_{ey} \end{bmatrix} M(e^{i\omega}) \\ &=: |H(e^{i\omega})|^{-2} M^{\mathrm{T}}(e^{i\omega})\Gamma_2 M(e^{i\omega}). \end{aligned} \tag{B18}$$

Under the condition that the system is an open loop test, we have

$$P_T = \lambda_{ey}|H(e^{i\omega})|^2\Gamma_1\Gamma_2\Gamma_1 = \Phi_v(\omega)\begin{bmatrix} \Phi_{u_0}^{-2}(\omega)\Phi_u(\omega) & 0 \\ 0 & \lambda_{ey}^{-1} \end{bmatrix}, \quad \text{(B19)}$$

i,e,

$$\operatorname{cov} \hat{G}_N^n(e^{i\omega}) \sim \frac{n}{N}\frac{\Phi_v(\omega)\Phi_u(\omega)}{[\Phi_u(\omega) - \lambda_{eu}]^2} \quad \text{(B20)}$$

$$\operatorname{cov} \hat{H}_N^n(e^{i\omega}) \sim \frac{n}{N}\frac{\Phi_v(\omega)}{\lambda_{ey}} = \frac{n}{N}|H(e^{i\omega})|^2 \quad \text{(B21)}$$

End of proof. □