Relations between Bias-Eliminating Least Squares, the Frisch scheme and Extended Compensated Least Squares methods for identifying errors-in-variables systems

Mei Hong, Torsten Söderström

Division of Systems and Control,
Department of Information Technology, Uppsala University,
P O Box 337, SE-75105 Uppsala, Sweden

March 6, 2008

Abstract

There are many methods for identifying errors-in-variables systems. Among them Bias-Eliminating Least Squares (BELS), the Frisch scheme and Extended Compensated Least Squares (ECLS) methods are attractive approaches because of their simplicity and good estimation accuracy. These three methods are all based on a bias-compensated least-squares (BCLS) principle. In this report, the relations between them are considered. In particular, the nonlinear equations utilized in these three methods are proved to be equivalent under different noise conditions. It is shown that BELS, Frisch and ECLS methods have the same asymptotic estimation accuracy providing the same extended vector is used.

1 Introduction

A system with errors or noises on both input and output measurements is called an errors-in-variables (EIV) system. Such models have broad applications and appear in many practical data analysis situations. During the past decades, many solutions to the EIV system identification problem have been proposed. An overview and some comparisons of different approaches have been made in [8] and [10].

It is well known that the standard Least Squares (LS) method does not yield consistent estimates for EIV situations due to the noise on both the input and the output sides. Assuming the noise parameters are known or could be estimated, a bias-compensated least-squares (BCLS) scheme can be used to remove the noise-contribution part of the normal equations to get consistent estimates. Various approaches designed for the bias compensated scheme exist. Among them, Bias-Eliminating Least Squares (BELS) methods [12], [14], Frisch schemes [1], [2] and Extended Compensated Least Squares (ECLS) methods [4], [5] are well known for their quite good estimation accuracy and for a modest computational cost. In order to estimate the noise parameters in BCLS scheme methods, besides compensated normal equations, additional equations are needed. In BELS, Frisch and ECLS methods different approaches are used to built these equations. It is interesting to analyze the relations of these three algorithms. In some numerical examples we have noticed that BELS and Frisch do have (very) similar estimates, but till now, no papers have theoretically analyzed the relations between BELS and Frisch schemes and their connections with the ECLS methods. In this report, we focus on comparing the equations used by these three algorithms and prove that, from the equations point of view, BELS and Frisch are equivalent to the ECLS method with the same extended model.

The report is organized as follows. In Section 2 we describe the background and introduce notations. The main idea of BELS, Frisch and ECLS methods are reviewed in Section 3. In Sections 4, we compare the equations of these three methods under different noise conditions. Finally, a numerical example is given in Section 5 before we draw conclusions in Section 6.

2 Background and Notations

Consider the noise-free input and output processes $u_0(t)$ and $y_0(t)$ which are linked by a linear dynamic system

$$A(q^{-1})\, y_0(t) = B(q^{-1})\, u_0(t), \tag{2.1}$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \ldots + a_{n_a}\, q^{-n_a} \\ B(q^{-1}) &= b_0 + b_1\, q^{-1} + \ldots + b_{n_b}\, q^{-n_b} \end{aligned} \tag{2.2}$$

are polynomials in the backward shift operator q^{-1}.

For EIV systems, the input and the output are measured with additive noise $\tilde{u}(t)$ and $\tilde{y}(t)$, respectively. The available signals are thus of the form:

$$u(t) \;=\; u_0(t) + \tilde{u}(t), \quad y(t) = y_0(t) + \tilde{y}(t). \tag{2.3}$$

The problem of identifying the errors-in-variables system is concerned with consistently estimating the system parameter vector

$$\theta = (a_1 \ldots a_{n_a} \quad b_0 \ldots b_{n_b})^T \tag{2.4}$$

and the noise parameters (such as noise variances) from the measured noisy data $\{u(t), y(t)\}_{t=1}^{N}$.

The following are general assumptions:

A1. The dynamic system (2.1) is asymptotically stable, and $A(z)$ and $B(z)$ have no common factors.

A2. The polynomial degrees n_a and n_b are a priori known.

A3. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually independent white noise sequences, of zero mean and variances λ_u and λ_y, respectively. Both noise signals are independent of $u_0(t)$.

A4. The true input $u_0(t)$ is a zero-mean stationary ergodic random signal, that is persistently exciting of sufficiently high order.

We introduce the regressor vector

$$\begin{aligned}\varphi(t) &= (-y(t-1)\ldots - y(t-n_a) \quad u(t)\ldots u(t-n_b))^T \\ &= (-y_0(t-1)\ldots - y_0(t-n_a)\; u_0(t)\ldots u_0(t-n_b))^T \\ &\quad +(-\tilde{y}(t-1)\ldots - \tilde{y}(t-n_a) \quad \tilde{u}(t)\ldots \tilde{u}(t-n_b))^T \\ &\triangleq \varphi_0(t)+\tilde{\varphi}(t), \end{aligned} \tag{2.5}$$

where $\varphi_0(t)$ and $\tilde{\varphi}(t)$ denote the noise-free part and noise contribution part of $\varphi(t)$, respectively.

For convenience, we utilize the extended regressor $\phi(t)$ and the extended parameter vector Θ as

$$\phi(t) = (-y(t) \quad \varphi^T(t))^T, \qquad \Theta = (1 \quad \theta^T)^T. \tag{2.6}$$

We introduce some further expressions for the regressor vector and the system parameter vector in partitioned form as

$$\varphi_u(t)=\begin{pmatrix} u(t) \\ \vdots \\ u(t-n_b) \end{pmatrix}, \quad \varphi_y(t)=\begin{pmatrix} -y(t-1) \\ \vdots \\ -y(t-n_a) \end{pmatrix}, \quad \phi_y(t)=\begin{pmatrix} -y(t) \\ \varphi_y(t) \end{pmatrix}. \tag{2.7}$$

For a general random process $x(t)$, we define its empirical covariance function $r_x(\tau)$ as:

$$r_x(\tau) = \frac{1}{N}\sum_{t=1}^{N}(x(\tau)x(t-\tau)), \quad \tau = 0,\ \pm 1,\ \pm 2,\ \ldots. \tag{2.8}$$

Further, the empirical cross-covariance matrix between two random vectors $x(t)$ and $y(t)$ and the cross-covariance vector between random vector $x(t)$ and random variable $z(t)$ are denoted as

$$R_{xy} = \frac{1}{N}\sum_{t=1}^{N} x(t)y^T(t), \quad r_{xz} = \frac{1}{N}\sum_{t=1}^{N} x(t)z(t). \tag{2.9}$$

The EIV system described by (2.1)-(2.3) can be expressed as a linear regressor model

$$y(t) = \varphi^T(t)\theta + \varepsilon(t), \tag{2.10}$$

where $\varepsilon(t) = \tilde{y}(t) - \tilde{\varphi}^T(t)\theta$.

For the standard least squares (LS) method, the estimate θ_{LS} is given by the solution to the normal equation, which for $N \to \infty$ can be written as

$$\begin{aligned} &\left(E\varphi(t)\varphi^T(t)\right)\hat{\theta}_{LS} = E\varphi(t)y(t) \\ \Rightarrow\ &\left(E\varphi_0(t)\varphi_0^T(t) + E\tilde{\varphi}(t)\tilde{\varphi}^T(t)\right)\hat{\theta}_{LS} = E\varphi_0(t)y_0(t) + E\tilde{\varphi}(t)\tilde{y}(t). \end{aligned} \tag{2.11}$$

According to the assumption **A3**, $E\tilde{\varphi}(t)\tilde{\varphi}^T(t) > \mathbf{0}$ and $E\tilde{\varphi}(t)\tilde{y}(t) = \mathbf{0}$. Also it holds that $y_0(t) = \varphi_0^T(t)\theta$. Hence

$$\hat{\theta}_{LS} \neq \theta, \qquad i.e. \quad \hat{\theta}_{LS} \text{ is biased.} \tag{2.12}$$

To get unbiased estimates, a bias-compensated least-squares (BCLS) scheme [11] can be introduced as

$$\hat{\theta}_{BCLS} = \left(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}}\right)^{-1}\left(r_{\varphi y} - r_{\tilde{\varphi}\tilde{y}}\right). \tag{2.13}$$

The basic idea of BCLS is to remove the noise contributions from the covariance matrix $R_{\varphi\varphi}$ and the vector $r_{\varphi y}$. In the compensated normal equation (2.13), $R_{\tilde{\varphi}\tilde{\varphi}}$ and $r_{\tilde{\varphi}\tilde{y}}$ are constructed by estimated noise covariances of both the input and the output sides.

When both the input noise $\tilde{u}(t)$ and the output noise $\tilde{y}(t)$ are white (**A3**), the covariance vector $r_{\tilde{\varphi}\tilde{y}}$ equals to zero and the covariance matrix of the noise vector $\tilde{\varphi}(t)$ is simplified as

$$R_{\tilde{\varphi}\tilde{\varphi}} = \begin{pmatrix} \lambda_y I_{n_a} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{n_b+1} \end{pmatrix}. \tag{2.14}$$

The unknown noise parameters are only λ_u and λ_y.

When $\tilde{y}(t)$ needs to model not only measurement and/or sensor noise but also effects of process disturbances, assumption **A3** should change to $\tilde{u}(t)$ white and $\tilde{y}(t)$ colored. Then, the detailed expressions of the matrix $R_{\tilde{\varphi}\tilde{\varphi}}$ and vector $r_{\tilde{\varphi}\tilde{y}}$ are

$$\begin{aligned} R_{\tilde{\varphi}\tilde{\varphi}} &= \begin{pmatrix} R_{\tilde{y}\tilde{y}} & \mathbf{0} \\ \mathbf{0} & R_{\tilde{u}\tilde{u}} \end{pmatrix} = \begin{pmatrix} r_{\tilde{y}}(0) & \dots & r_{\tilde{y}}(n_a-1) & \\ \vdots & \ddots & \vdots & \mathbf{0} \\ r_{\tilde{y}}(n_a-1) & \dots & r_{\tilde{y}}(0) & \\ & \mathbf{0} & & \lambda_u I_{n_b+1} \end{pmatrix}, & (2.15) \\ r_{\tilde{\varphi}\tilde{y}} &= \begin{pmatrix} r_{\tilde{y}}(1) & \dots & r_{\tilde{y}}(n_a) & \mathbf{0}_{1\times(n_b+1)} \end{pmatrix}^T. & (2.16) \end{aligned}$$

The corresponding unknown noise parameter vector is in this case defined as

$$\rho = (\lambda_u, \rho_y^T)^T = (\lambda_u, r_{\tilde{y}}(0), r_{\tilde{y}}(1), \dots, r_{\tilde{y}}(n_a))^T. \tag{2.17}$$

In order to determine the noise covariances (and hence $R_{\tilde{\varphi}\tilde{\varphi}}$ and $r_{\tilde{\varphi}\tilde{y}}$), more relations are needed in addition to the equations (2.13). BELS, Frisch and ECLS use different techniques to build these additional equations. In the following section, we will briefly review these three methods.

3 Review of BELS, Frisch and ECLS methods

3.1 BELS methods

The BELS methods were first proposed by Zheng in [12], [14] for identifying EIV systems with white input and output measurement noise sequences. In order to determine the noise variances (λ_u, λ_y) (and hence $R_{\tilde{\varphi}\tilde{\varphi}}$), at least two more relations are needed in addition to the equations (2.13).

In BELS, the first such relation is derived from the minimal value of the least squares criterion:

$$V_{LS} \;=\; \frac{1}{N}\sum_{t=1}^{N}\Big(y(t)-\varphi^T(t)\hat{\theta}_{LS}\Big)^2 = \lambda_y + \hat{\theta}_{LS}^T R_{\tilde{\varphi}\tilde{\varphi}}\theta. \tag{3.1}$$

See [12] for details in deriving this equality. To get also a second relation for λ_y and λ_u, an extended model structure was considered for the BELS methods. For this purpose we introduce extended versions of $\varphi(t)$ and θ as

$$\bar{\varphi}(t) = \begin{pmatrix} \varphi^T(t) & \underline{\varphi}^T(t) \end{pmatrix}^T, \quad \bar{\theta} = \begin{pmatrix} \theta^T & \mathbf{0} \end{pmatrix}^T. \tag{3.2}$$

The model extension can, for example, mean that additional A or B (or both) parameters in (2.10) are appended. Consider least squares estimation in the extended model

$$y(t) \;=\; \bar{\varphi}^T(t)\bar{\theta} + \varepsilon(t), \tag{3.3}$$

which leads to $R_{\bar{\varphi}\bar{\varphi}}\hat{\bar{\theta}}_{LS} = r_{\bar{\varphi}y}$. Recall that $y(t) = y_0(t) + \tilde{y}(t)$ and $\bar{\varphi}(t) = \bar{\varphi}_0(t) + \tilde{\bar{\varphi}}(t)$. Hence

$$R_{\bar{\varphi}\bar{\varphi}}\hat{\bar{\theta}}_{LS} \;=\; r_{\bar{\varphi}_0 y_0} + r_{\tilde{\bar{\varphi}}\tilde{y}} = R_{\bar{\varphi}_0\bar{\varphi}_0}\bar{\theta} = (R_{\bar{\varphi}\bar{\varphi}} - R_{\tilde{\bar{\varphi}}\tilde{\bar{\varphi}}})\bar{\theta}. \tag{3.4}$$

Set $H = (0,\ldots,0,1)$, $\quad J = \begin{pmatrix} I_{n_a+n_b} & \mathbf{0} \end{pmatrix}^T$, $\quad \bar{\theta} = J\theta$, and recalling that $H\bar{\theta} = 0$, we find that (3.4) imply

$$H\hat{\bar{\theta}}_{LS} \;=\; HR_{\bar{\varphi}\bar{\varphi}}^{-1}(R_{\bar{\varphi}\bar{\varphi}} - R_{\tilde{\bar{\varphi}}\tilde{\bar{\varphi}}})\bar{\theta} = -HR_{\bar{\varphi}\bar{\varphi}}^{-1}R_{\tilde{\bar{\varphi}}\tilde{\bar{\varphi}}}J\theta. \tag{3.5}$$

To sum up, the BELS algorithm consists of three equations:

- the compensated normal equations (2.13),
- the minimal value of the loss function (3.1),
- the additional equation (3.5) obtained by using an extended model.

In the classical BELS methods, we initially set $\hat{\theta} = \hat{\theta}_{LS}$, and then solve equations (3.1), (3.5) to get the estimates of the noise variance λ_u and λ_y. Next, a new value of $\hat{\theta}$ is computed by using the compensated normal equation (2.13). This process is iterated until convergence.

The BELS methods were extended to colored output noise conditions in [13]. In this case, the minimal value of the loss function becomes

$$\begin{aligned} V_{LS} \;&=\; \frac{1}{N}\sum_{t=1}^{N}\Big(y(t)-\varphi^T(t)\hat{\theta}_{LS}\Big)^2 \\ &=\; r_{\tilde{y}}(0) + \hat{\theta}_{LS}^T R_{\tilde{\varphi}\tilde{\varphi}}\theta - (\theta + \hat{\theta}_{LS})^T r_{\tilde{\varphi}\tilde{y}}, \end{aligned} \tag{3.6}$$

(see equation (43) in [13]). In order to get the second relation, the system model is extended by appending an additional B parameter vector

$$\underline{\varphi}(t) = \begin{pmatrix} u(t-n_b-1) & \ldots & u(t-n_b-p) \end{pmatrix}^T. \tag{3.7}$$

The second relation becomes

$$R_{\varphi\underline{\varphi}}^T R_{\varphi\varphi}^{-1}(R_{\tilde{\varphi}\tilde{\varphi}}\theta - r_{\tilde{\varphi}\tilde{y}}) = r_{\underline{\varphi}y} - R_{\varphi\underline{\varphi}}^T \hat{\theta}_{LS}, \tag{3.8}$$

instead of (3.5). See *Theorem 2* in [13] for details. In [6], the equations (3.5) and (3.8) were further simplified. It was shown that each can be equivalently substituted by

$$R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y}. \tag{3.9}$$

3.2 Frisch schemes

The Frisch scheme was developed to identify dynamic EIV systems in [1] and was further elaborated in [2]. Consider the case of the white input and output measurement noise (**A3**). The Frisch scheme utilizes the relation, *cf.* (2.6),

$$\phi_0^T(t)\Theta_0 \;=\; -A_0(q^{-1})y_0(t) + B_0(q^{-1})u_0(t) = 0. \tag{3.10}$$

Further it holds that $R_{\phi\phi} = R_{\phi_0\phi_0} + R_{\tilde{\phi}\tilde{\phi}}$. It follows from (3.10) that

$$R_{\phi_0\phi_0}\Theta_0 \;=\; \frac{1}{N}\sum_{t=1}^{N}(\phi_0(t)\phi_0^T(t)\Theta_0) = \mathbf{0}. \tag{3.11}$$

Hence the matrix $R_{\phi_0\phi_0}$ is singular (positive semidefinite), with at least one eigenvalue equal to zero. The relation (3.11) is the basis for the Frisch method, and can be expressed as

$$\left(R_{\phi\phi} - R_{\tilde{\phi}\tilde{\phi}}\right)\Theta \;=\; \mathbf{0}, \tag{3.12}$$

where $R_{\tilde{\phi}\tilde{\phi}} = \begin{pmatrix} \lambda_y I_{n_a+1} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{n_b+1} \end{pmatrix}$.

Assume for a moment that an estimate of λ_u is available. Then the output noise variance λ_y is determined so that the matrix in (3.12) is singular. More specifically, see [1]:

$$\lambda_y = \lambda_{\min}(R_{\phi_y\phi_y} - R_{\phi_y\varphi_u}(R_{\varphi_u\varphi_u} - \lambda_u I_{n_b+1})^{-1}R_{\varphi_u\phi_y}), \tag{3.13}$$

where $\lambda_{\min}(R)$ denotes the smallest eigenvalue of R. To determine λ_u, several alternatives have been proposed to get the needed second relation. Similar to the BELS methods, one choice is to evaluate the Frisch equation (3.13) for an extended model with an additional regressor $\underline{\varphi}(t)$. This method was proposed in [1]. For this alternative the extended Frisch equation will be

$$(R_{\bar{\phi}\bar{\phi}} - R_{\bar{\tilde{\phi}}\bar{\tilde{\phi}}})\bar{\Theta} = \mathbf{0}, \tag{3.14}$$

where $\bar{\phi} = \begin{pmatrix} \phi(t) \\ \underline{\varphi}(t) \end{pmatrix}, \quad \bar{\tilde{\phi}} = \begin{pmatrix} \tilde{\phi}(t) \\ \underline{\tilde{\varphi}}(t) \end{pmatrix}, \quad \bar{\Theta} = \begin{pmatrix} \Theta \\ \mathbf{0} \end{pmatrix}$.

Therefore the function $\lambda_y(\lambda_u)$ is evaluated both for the nominal model and for the extended model by utilizing the relations (3.12) and (3.14), respectively. The two functions correspond to two curves in the (λ_u, λ_y) plane. The curves will ideally have one unique common point, which defines the estimate.

Besides using the extended model, two other approaches have been proposed to build the new relations in the Frisch methods. One alternative is to compute residuals and compare their statistical properties with what can be predicted from the model [2]. The second alternative is to use the overdetermined Yule-Walker equations [3]. In this report, we focus on analyzing the Frisch methods which utilize the extended model.

Recently, the Frisch scheme was extended for identifying EIV systems with correlated output noise [9]. The additional equations were obtained by using the cross correlation of residuals and past input,

$$\begin{aligned} & \frac{1}{N}\sum_{t=1}^{N}\begin{pmatrix} u(t-n_b-1) \\ \vdots \\ u(t-n_b-p) \end{pmatrix}(A(q^{-1})y(t)-B(q^{-1})u(t))=\mathbf{0}, \\ \Leftrightarrow \quad & \frac{1}{N}\sum_{t=1}^{N}\underline{\varphi}(t)\varepsilon(t) = \mathbf{0}, \\ \Leftrightarrow \quad & R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y}. \end{aligned} \tag{3.15}$$

3.3 The ECLS methods

Extended Compensated Least Squares (ECLS) is a relatively new method proposed by Ekman in [4]. The main idea of the ECLS methods is to extend the bias eliminating concept to the instrumental variable methods. The methods are based on the following overdetermined system of equations

$$\left(R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho)\right)\theta = r_{zy} - r_{\tilde{z}\tilde{y}}(\rho). \tag{3.16}$$

The choice of the entries in the instrumental vector $z(t)$ will determine the structure of $R_{\tilde{z}\tilde{\varphi}}(\rho)$ and $r_{\tilde{z}\tilde{y}}(\rho)$. To estimate the system parameter vector θ as well as the noise parameter vector ρ, we will choose at least some of the entries in $z(t)$ to be correlated with the system disturbance $\varepsilon(t)$. In ECLS, the equation (3.16) is solved by considering it as a nonlinear LS problem. The estimates of the system parameters θ and the noise parameters ρ are given as the minimal point of

$$(\hat{\theta}, \hat{\rho}) \quad = \quad \arg\min_{\theta,\rho} f(\theta, \rho), \tag{3.17}$$

where $f(\theta, \rho)$ is the loss function

$$f(\theta, \rho) \quad = \quad ||r_{zy} - r_{\tilde{z}\tilde{y}}(\rho) - \left(R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho)\right)\theta||^2. \tag{3.18}$$

Algorithms for the ECLS methods utilize the fact that the optimization problem (3.17) is separable so that θ and ρ can be estimated separately. If the loss function (3.18) is minimized with respect to θ, then, for a given ρ the minimization is achieved by

$$\hat{\theta}(\rho) = \left(R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho)\right)^{\dagger}\left(r_{zy} - r_{\tilde{z}\tilde{y}}(\rho)\right), \tag{3.19}$$

where $(R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho))^\dagger$ denotes the pseudo-inverse of $(R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho))$. Substituting (3.19) into (3.17), the estimates of the noise vector $\hat{\rho}$ is found from solving a variable projection minimization problem

$$\begin{aligned}\hat{\rho} \quad = \quad & \arg\min_{\rho} ||r_{zy} - r_{\tilde{z}\tilde{y}}(\rho) - (R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho)) \\ & \qquad \times (R_{z\varphi} - R_{\tilde{z}\tilde{\varphi}}(\rho))^\dagger (r_{zy} - r_{\tilde{z}\tilde{y}}(\rho))||^2. \end{aligned} \tag{3.20}$$

Then a consistent estimate of the system parameters $\hat{\theta}_{ECLS}$ can be obtained by replacing ρ by $\hat{\rho}$ in (3.19).

4 Comparisons of the equations in BELS, Frisch and ECLS

It is interesting to compare the equations utilized in BELS, Frisch and ECLS methods. An important user choice in BELS and Frisch methods is the additional extended vector $\underline{\varphi}(t)$, which has large effects on the performance of the estimates. For ECLS, in the key equation (3.16) the entries of the vector $z(t)$ can be designed in many different ways. In particular we choose here $z(t)$ as

$$z(t) \quad = \quad \begin{pmatrix} y(t) & \varphi^T(t) & \underline{\varphi}^T(t) \end{pmatrix}^T. \tag{4.1}$$

Insert such $z(t)$ into equation (3.16). For both $\tilde{u}$ and $\tilde{y}$ being white noise, we will obtain the following three equations:

$$r_{y\varphi}\theta = r_y(0) - r_{\tilde{y}}(0). \tag{4.2}$$

$$(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\,\theta = r_{\varphi y}. \tag{4.3}$$

$$\begin{aligned} & \left(R_{\underline{\varphi}\varphi} - R_{\underline{\tilde{\varphi}}\tilde{\varphi}}\right)\theta = r_{\underline{\varphi}y} - r_{\underline{\tilde{\varphi}}\tilde{y}}, \\ \Leftrightarrow \quad & R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y}. \end{aligned} \tag{4.4}$$

It is obvious that equation (4.3) is just the compensated normal equation used in both BELS and the Frisch schemes. For the remaining two equations, (4.2) and (4.4), assume that (4.3) holds, we then have the following two lemmas.

Lemma 1 Equation (3.1) in BELS and equation (3.12) used in the Frisch scheme are equivalent to equation (4.2).

Lemma 2 Equation (3.5) in BELS and the Frisch equation (3.14) are equivalent to equation (4.4).

Proof: See Appendices A and B, respectively.

We see that, under the white measurement noises case, all the equations used in BELS and Frisch are identical to ECLS when the instrumental vector $z(t)$ is chosen as in (4.1).

For the correlated output noise case, in contrast to the white output noise, the covariance vector $r_{\tilde{\varphi}\tilde{y}}$ will not equal to zero and the unknown noise parameters are collected in the vector ρ, (2.17), instead of only the noise variances λ_u and λ_y. From the descriptions in Section 3, if the extended vector $\underline{\varphi}(t)$ is designed by using the past inputs and the instrumental vector $z(t)$ is chosen as in (4.1), we summarize the equations used in the three algorithms as

$$(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\,\theta = r_{\varphi y} - r_{\tilde{\varphi}\tilde{y}}, \tag{4.5}$$

$$R_{\underline{\varphi}\varphi}\,\theta = r_{\underline{\varphi}y}, \tag{4.6}$$

$$\text{BELS:} \quad V_{LS} = r_{\tilde{y}}(0) + \hat{\theta}_{LS}^T R_{\tilde{\varphi}\tilde{\varphi}}\theta - (\theta + \hat{\theta}_{LS})^T r_{\tilde{\varphi}\tilde{y}}, \tag{4.7}$$

$$\text{Frisch:} \quad r_{\tilde{u}}(0) = \lambda_{min}(R_{\varphi_u\varphi_u} - R_{\varphi_u\bar{\varphi}_y}(R_{\bar{\varphi}_y\bar{\varphi}_y} - R_{\tilde{\bar{\varphi}}_y\tilde{\bar{\varphi}}_y}(\rho_y))^{-1}R_{\bar{\varphi}_y\varphi_u}), \tag{4.8}$$

$$\text{ECLS:} \quad (r_{y\varphi} - r_{\tilde{y}\tilde{\varphi}})\,\theta = r_y(0) - r_{\tilde{y}}(0). \tag{4.9}$$

The detailed derivation for equation (4.8) was given in [9]. The number of equations above are $n_a + n_b + 1$ for (4.5), p for (4.6) and 1 for each of (4.7)-(4.9). The number of the total equations must be larger than or equal to the number of unknowns, so the size of the additional extended vector $\underline{\varphi}(t)$ as in (3.7) should meet $p \geq n_a + 1$ to guarantee the computability. If p is chosen as $p = n_a + n_b + 1$, (4.6) itself gives exactly the basic instrumental variable estimates of θ.

It can be seen that, for identifying EIV systems in the correlated output noise case, equations (4.5) and (4.6) are exactly same for BELS, Frisch and ECLS method. For the equations (4.7)-(4.9), we have the following lemma.

Lemma 3 Equations (4.7)-(4.9) are equivalent.

Proof: See Appendix C.

So again, all the equations used in BELS and Frisch are equivalent to those of ECLS.

Extensions of these three methods to the colored noises at both sides are more complicated. In this case using the extended model to get additional equations does not work except if the noises can be parametrically modeled as ARMA processes [5].

We know that different methods can have different performances depending not only on the equations that they use but also depending on the techniques utilized to solve these equations. However, the asymptotic statistical properties should rely only on the set of defining equations. Summing up so far, we get the following result.

Result: Although BELS, Frisch and ECLS methods have different implementations, their asymptotic estimation accuracy will be identical provided the same extended vector $\underline{\varphi}(t)$ is used.

5 Numerical illustrations

A simple numerical example will be used to illustrate the previous theoretical analysis. Consider a first-order system

$$(1-0.8q^{-1})y_0(t) \quad = \quad 1.0q^{-1}u_0(t). \tag{5.1}$$

The white measurement noises at the input and output sides both have unit variance. The undisturbed input $u_0(t)$ is an AR(1) process

$$(1-0.5q^{-1})u_0(t) = e(t), \tag{5.2}$$

where $e(t)$ is a zero-mean white noise with unit variance. We will compare the normalized asymptotic covariance matrix of the estimated parameters

$$P = \lim_{N\to\infty} E\mathrm{N}(\hat{\vartheta}-\vartheta_0)(\hat{\vartheta}-\vartheta_0)^T, \tag{5.3}$$

where $\hat{\vartheta}$ and ϑ_0 denote the estimate and the true value of the total parameter vector ϑ, which is defined as: $\vartheta = (a_1 \dots a_{na} \quad b_0 \dots b_{nb}; \quad \lambda_y \quad \lambda_u)$.

Assume first that we choose the extended vector $\underline{\varphi}(t) = -y(t-n_a-1)$. Then, for data number $N = 10^4$ and 200 realizations, the normalized covariance matrix of BELS estimates is

$$P_{\mathrm{BELS}} = \begin{pmatrix} 0.32 & & & \\ 1.12 & 6.81 & & \\ -0.72 & -4.31 & 7.42 & \\ 1.08 & 6.31 & -4.43 & 11.59 \end{pmatrix}.$$

The normalized covariance matrix of the parameters estimated by the Frisch estimator is

$$P_{\mathrm{Frisch}} = \begin{pmatrix} 0.32 & & & \\ 1.17 & 7.12 & & \\ -0.83 & -5.13 & 8.12 & \\ 1.05 & 5.99 & -5.31 & 10.07 \end{pmatrix},$$

and the normalized covariance matrix of the parameters estimated by the ECLS is

$$P_{\mathrm{ECLS}} = \begin{pmatrix} 0.31 & & & \\ 1.05 & 6.67 & & \\ -0.68 & -4.45 & 7.07 & \\ 0.93 & 6.10 & -4.07 & 10.59 \end{pmatrix}.$$

As can be seen, the normalized asymptotic covariance matrix for these three methods are very similar when N is large. The theoretical proofs in the previous section are well supported by this numerical example. During the simulations, we also notice that when signal to noise ratio at both input and output sides are moderate, the normalized asymptotic covariance matrix for these three methods are very similar even when the data number is not very large. The computational load of ECLS is largest among these three methods.

6 Conclusions

In this report identification of linear errors-in-variables system has been studied. Three estimators, bias-eliminating least squares, the Frisch scheme and the extended compensated least squares methods have been analyzed. We have focused on comparing the nonlinear equations used in the algorithms. It has been proved that the defining equations are equivalent when the same additional extended vector $\underline{\varphi}(t)$ is used by the three methods. Despite of the different techniques used to solve these equations, BELS, Frisch and ECLS methods have the same asymptotic estimation accuracy providing the same extended model is used.

Acknowledgments: The authors are grateful to Professor Umberto Soverini for the helpful comments.

A Proof of Lemma 1

V_{LS}, the minimal value of the loss function used in BELS, can be further expressed as:

$$\begin{aligned} V_{LS} &= \frac{1}{N}\sum_{t=1}^{N}(y(t)-\varphi^T(t)\hat{\theta}_{LS})^2 \\ &= \frac{1}{N}\sum_{t=1}^{N}\big(y(t)-\varphi^T(t)R_{\varphi\varphi}^{-1}r_{\varphi y})(y(t)-\varphi^T(t)R_{\varphi\varphi}^{-1}r_{\varphi y}\big) \\ &= r_y(0)-r_{\varphi y}^T R_{\varphi\varphi}^{-1}r_{\varphi y}. \end{aligned} \tag{A.1}$$

Insert (A.1) into equation (3.1) in BELS, and use the relations $\hat{\theta}_{LS} = R_{\varphi\varphi}^{-1}r_{\varphi y}$ and $r_{\varphi_0 y_0} = R_{\varphi_0\varphi_0}\theta$. We get that (3.1) is equivalent to

$$\begin{aligned} & r_y(0)-r_{\varphi y}^T R_{\varphi\varphi}^{-1}r_{\varphi y} = r_{\tilde{y}}(0)+r_{\varphi y}^T R_{\varphi\varphi}^{-1}R_{\tilde{\varphi}\tilde{\varphi}}\theta \\ \Leftrightarrow\quad & r_y(0)-r_{\tilde{y}}(0) = r_{\varphi y}^T R_{\varphi\varphi}^{-1}(r_{\varphi y}+R_{\tilde{\varphi}\tilde{\varphi}}\theta) \\ \Leftrightarrow\quad & r_y(0)-r_{\tilde{y}}(0) = r_{y\varphi}R_{\varphi\varphi}^{-1}(r_{\varphi_0 y_0}+R_{\tilde{\varphi}\tilde{\varphi}}\theta) \\ \Leftrightarrow\quad & r_y(0)-r_{\tilde{y}}(0) = r_{y\varphi}R_{\varphi\varphi}^{-1}(R_{\varphi_0\varphi_0}\theta+R_{\tilde{\varphi}\tilde{\varphi}}\theta) \\ \Leftrightarrow\quad & r_y(0)-r_{\tilde{y}}(0) = r_{y\varphi}R_{\varphi\varphi}^{-1}R_{\varphi\varphi}\theta \\ \Leftrightarrow\quad & r_y(0)-r_{\tilde{y}}(0) = r_{y\varphi}\theta, \end{aligned} \tag{A.2}$$

which is (4.2).

Utilizing the normal equation for the white noise case, (4.3), the Frisch equation (3.12) can be further expressed as

$$\begin{aligned} & (R_{\phi\phi}-R_{\tilde{\phi}\tilde{\phi}})\Theta = \mathbf{0} \\ \Leftrightarrow\quad & \begin{pmatrix} r_y(0)-r_{\tilde{y}}(0) & -r_{y\varphi} \\ -r_{\varphi y} & R_{\varphi\varphi}-R_{\tilde{\varphi}\tilde{\varphi}} \end{pmatrix}\begin{pmatrix} 1 \\ \theta \end{pmatrix} = \mathbf{0} \\ \Leftrightarrow\quad & \begin{pmatrix} r_y(0)-r_{\tilde{y}}(0) & -r_{y\varphi} \\ -(R_{\varphi\varphi}-R_{\tilde{\varphi}\tilde{\varphi}})\theta & R_{\varphi\varphi}-R_{\tilde{\varphi}\tilde{\varphi}} \end{pmatrix}\begin{pmatrix} 1 \\ \theta \end{pmatrix} = \mathbf{0} \\ \Leftrightarrow\quad & r_y(0)-r_{\tilde{y}}(0) = r_{y\varphi}\theta, \end{aligned} \tag{A.3}$$

which is again (4.2).

B Proof of Lemma 2

Equation (3.5) can be simplified as

$$HR_{\bar{\varphi}\bar{\varphi}}^{-1}(r_{\bar{\varphi}y} + R_{\tilde{\bar{\varphi}}\tilde{\bar{\varphi}}}J\theta) = 0, \tag{B.1}$$

where we assume $R_{\bar{\varphi}\bar{\varphi}}^{-1}$ to be exist. Equation (B.1) can be further expressed as

$$\begin{aligned}
\Leftrightarrow \quad & \begin{pmatrix} \mathbf{0}_{1\times(n_a+n_b+1)} & 1 \end{pmatrix} \begin{pmatrix} R_{\varphi\varphi} & R_{\varphi\underline{\varphi}} \\ R_{\underline{\varphi}\varphi} & R_{\underline{\varphi}\underline{\varphi}} \end{pmatrix}^{-1} \begin{pmatrix} r_{\varphi y} + R_{\tilde{\varphi}\tilde{\varphi}}\theta \\ r_{\underline{\varphi}y} + R_{\underline{\tilde{\varphi}}\tilde{\varphi}}\theta \end{pmatrix} = 0 \\
\Leftrightarrow \quad & (R_{\underline{\varphi}\underline{\varphi}} - R_{\underline{\varphi}\varphi}R_{\varphi\varphi}^{-1}R_{\varphi\underline{\varphi}})^{-1} \\
& \times\begin{pmatrix} -R_{\underline{\varphi}\varphi}R_{\varphi\varphi}^{-1} & I_{(n_a+n_b+1)} \end{pmatrix} \begin{pmatrix} r_{\varphi y} + R_{\tilde{\varphi}\tilde{\varphi}}\theta \\ r_{\underline{\varphi}y} \end{pmatrix} = 0 \\
\Leftrightarrow \quad & R_{\underline{\varphi}\varphi}R_{\varphi\varphi}^{-1}(r_{\varphi y} + R_{\tilde{\varphi}\tilde{\varphi}}\theta) = r_{\underline{\varphi}y} \\
\Leftrightarrow \quad & R_{\underline{\varphi}\varphi}R_{\varphi\varphi}^{-1}(R_{\varphi_0\varphi_0} + R_{\tilde{\varphi}\tilde{\varphi}})\theta = r_{\underline{\varphi}y} \\
\Leftrightarrow \quad & R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y},
\end{aligned} \tag{B.2}$$

which is (4.4).

The first alternative of the extended Frisch equation, (3.14), can be described as

$$\begin{aligned}
& (R_{\bar{\phi}\bar{\phi}} - R_{\tilde{\bar{\phi}}\tilde{\bar{\phi}}})\bar{\Theta} = \mathbf{0} \\
\Leftrightarrow \quad & \begin{pmatrix} r_y(0) - r_{\tilde{y}}(0) & -r_{y\varphi} & -r_{y\underline{\varphi}} \\ -r_{\varphi y} & R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}} & R_{\varphi\underline{\varphi}} \\ -r_{\underline{\varphi}y} & R_{\underline{\varphi}\varphi} & R_{\underline{\varphi}\underline{\varphi}} - R_{\underline{\tilde{\varphi}}\underline{\tilde{\varphi}}} \end{pmatrix} \begin{pmatrix} 1 \\ \theta \\ 0 \end{pmatrix} = \mathbf{0} \\
\Leftrightarrow \quad & \begin{pmatrix} r_y(0) - r_{\tilde{y}}(0) - r_{y\varphi}\theta \\ -r_{\varphi y} + (R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\theta \\ -r_{\underline{\varphi}y} + R_{\underline{\varphi}\varphi}\theta \end{pmatrix} = \mathbf{0}.
\end{aligned} \tag{B.3}$$

Using the relation (A.3) and the compensated normal equations (4.3), the equation (B.3) gives

$$\begin{aligned}
\Leftrightarrow \quad & \begin{pmatrix} 0 \\ \mathbf{0} \\ -r_{\underline{\varphi}y} + R_{\underline{\varphi}\varphi}\theta \end{pmatrix} = \mathbf{0} \\
\Leftrightarrow \quad & R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y},
\end{aligned} \tag{B.4}$$

which is again (4.4).

C Proof of Lemma 3

Equation (4.7) used in BELS can be further expressed as

$$V_{LS} = E(y(t) - \varphi^T(t)\hat{\theta}_{LS})^2 = r_y(0) - r_{\varphi y}^T R_{\varphi\varphi}^{-1} r_{\varphi y}. \tag{C.1}$$

From (3.6), we also get the relation

$$\begin{aligned}
V_{LS} &= r_y(0) - r_{\varphi y}^T R_{\varphi\varphi}^{-1} r_{\varphi y} \\
&= r_{\tilde{y}}(0) + r_{\varphi y}^T R_{\varphi\varphi}^{-1} R_{\tilde{\varphi}\tilde{\varphi}}\theta - \theta^T r_{\tilde{\varphi}\tilde{y}} - r_{\varphi y}^T R_{\varphi\varphi}^{-1} r_{\tilde{\varphi}\tilde{y}}.
\end{aligned} \tag{C.2}$$

We find that (C.2) can be rewritten as

$$
\begin{aligned}
\Leftrightarrow \quad & r_y(0) - r_{\varphi y}^T R_{\varphi\varphi}^{-1} r_{\varphi y} \\
& = r_{\tilde{y}}(0) + r_{\varphi y}^T R_{\varphi\varphi}^{-1} R_{\tilde{\varphi}\tilde{\varphi}}\theta - \theta^T r_{\tilde{\varphi}\tilde{y}} - r_{\varphi y}^T R_{\varphi\varphi}^{-1} r_{\tilde{\varphi}\tilde{y}} \\
\Leftrightarrow \quad & r_y(0) - r_{\tilde{y}}(0) = r_{\varphi y}^T R_{\varphi\varphi}^{-1}(r_{\varphi y} + R_{\tilde{\varphi}\tilde{\varphi}}\theta - r_{\tilde{\varphi}\tilde{y}}) - \theta^T r_{\tilde{\varphi}\tilde{y}} \\
\Leftrightarrow \quad & r_y(0) - r_{\tilde{y}}(0) = r_{y\varphi} R_{\varphi\varphi}^{-1}(r_{\varphi_0 y_0} + R_{\tilde{\varphi}\tilde{\varphi}}\theta) - r_{\tilde{y}\tilde{\varphi}}\theta \\
\Leftrightarrow \quad & r_y(0) - r_{\tilde{y}}(0) = r_{y\varphi} R_{\varphi\varphi}^{-1}(R_{\varphi_0\varphi_0}\theta + R_{\tilde{\varphi}\tilde{\varphi}}\theta) - r_{\tilde{y}\tilde{\varphi}}\theta \\
\Leftrightarrow \quad & r_y(0) - r_{\tilde{y}}(0) = r_{y\varphi} R_{\varphi\varphi}^{-1} R_{\varphi\varphi}\theta - r_{\tilde{y}\tilde{\varphi}}\theta \\
\Leftrightarrow \quad & r_y(0) - r_{\tilde{y}}(0) = r_{y\varphi}\theta - r_{\tilde{y}\tilde{\varphi}}\theta,
\end{aligned}
\tag{C.3}
$$

which is identical to (4.9).

For equation (3.13), if we assume that the output noise parameter vector ρ_y is available, then the input noise variance is determined so that the matrix $(R_{\phi\phi} - R_{\tilde{\phi}\tilde{\phi}})$ appearing in (3.12) is singular. Please refer to Lemma 3.1 in [7] for details. The Frisch equation (4.8) can be presented as (3.12), $(R_{\phi\phi} - R_{\tilde{\phi}\tilde{\phi}})\Theta = \mathbf{0}$, which can be further expressed as

$$
\begin{aligned}
\Leftrightarrow \quad & \begin{pmatrix} r_y(0) - r_{\tilde{y}}(0) & -(r_{y\varphi} - r_{\tilde{y}\tilde{\varphi}}) \\ -(r_{\varphi y} - r_{\tilde{\varphi}\tilde{y}}) & R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}} \end{pmatrix} \begin{pmatrix} 1 \\ \theta \end{pmatrix} = \mathbf{0} \\
\Leftrightarrow \quad & \begin{pmatrix} r_y(0) - r_{\tilde{y}}(0) & -(r_{y\varphi} - r_{\tilde{y}\tilde{\varphi}}) \\ -(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\theta & R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}} \end{pmatrix} \begin{pmatrix} 1 \\ \theta \end{pmatrix} = \mathbf{0} \\
\Leftrightarrow \quad & \begin{pmatrix} r_y(0) - r_{\tilde{y}}(0) - (r_{y\varphi} - r_{\tilde{y}\tilde{\varphi}})\theta \\ -(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\theta + (R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\theta \end{pmatrix} = \mathbf{0} \\
\Leftrightarrow \quad & \begin{pmatrix} r_y(0) - r_{\tilde{y}}(0) - (r_{y\varphi} - r_{\tilde{y}\tilde{\varphi}})\theta \\ \mathbf{0} \end{pmatrix} = \mathbf{0} \\
\Leftrightarrow \quad & r_y(0) - r_{\tilde{y}}(0) = (r_{y\varphi} - r_{\tilde{y}\tilde{\varphi}})\theta.
\end{aligned}
\tag{C.4}
$$

It is again the same equation as (4.9).

References

[1] S. Beghelli, R. P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26:171–176, 1990.

[2] R. Diversi, R. Guidorzi, and U. Soverini. A new criterion in EIV identification and filtering applications. In *Proc. 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August 27-29 2003.

[3] R. Diversi, R. Guidorzi, and U. Soverini. Yule-walker equations in the Frisch scheme solution of errors-in-variables identification problems. In *Proc. 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, July 24-28 2006.

[4] M. Ekman. Identification of linear system with errors in variables using separable nonlinear least-squares. In *Proc. 16th IFAC World Congress*, Prague, Czech Republic, July 4-8 2005.

[5] M. Ekman, M. Hong, and T. Söderström. Identification of linear system with errors in variables using separable nonlinear least-squares. In *Proc. 14th IFAC Symposium on System Identification, SYSID 2006*, Newcastle, Australia, March 29-31 2006.

[6] M. Hong, T. Söderström, and W. X. Zheng. A simplified form of the bias-eliminating least squares method for errors-in-variables identification. *IEEE Transactions on Automatic Control*, 52(9):1754–1756, September 2007.

[7] T. Söderström. Accuracy analysis of the Frisch estimates for identifying errors-in-variables systems. *IEEE Transactions on Automatic Control*, 52(6):985–997, June 2007.

[8] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, June 2007. Survey paper.

[9] T. Söderström. Extending the Frisch scheme for errors-in-variables identification to correlated output noise. *International Journal of Adaptive Control and Signal Processing*, 22(1):55–73, February 2008.

[10] T. Söderström, U. Soverini, and K. Mahata. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing*, 82(8):1139–1154, August 2002.

[11] P. Stoica and T. Söderström. Bias correction in least–squares identification. *International Journal of Control*, 35(3):449–457, 1982.

[12] W. X. Zheng. Transfer function estimation form noisy input and output data. *International Journal of Adaptive Control and Signal Processing*, 12:365–380, 1998.

[13] W. X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47:1142–1147, 2002.

[14] W. X. Zheng and C. B. Feng. Unbiased parameter estimation of linear systems in presence of input and output noise. *International Journal of Adaptive Control and Signal Processing*, 3:231–251, 1989.