Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Comparison of three Frisch methods for errors-in-variables identification ⋆

Mei Hong * **Torsten Söderström** * **Umberto Soverini** ** **Roberto Diversi** **

* *Division of Systems and Control, Department of Information Technology, Uppsala University, P O Box 337, SE-75105 Uppsala, Sweden (e-mail: Mei.Hong@it.uu.se, Torsten.Soderstrom@it.uu.se).*
** *Dipartimento di Elettronica, Informatica e Sistemistica, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy (e-mail: umberto.soverini@unibo.it, roberto.diversi@unibo.it).*

Abstract: The errors–in–variables framework concerns static or dynamic systems whose input and output variables are affected by additive noise. Several estimation methods have been proposed for identifying dynamic errors–in–variables models. One of the more promising approaches is the so–called Frisch scheme. This paper decribes three different estimation criteria within the Frisch context and compares their estimation accuracy on the basis of the asymptotic covariance matrices of the estimates. Some numerical examples support well the theoretical results.

Keywords: System identification, errors–in–variables models, linear systems

1. INTRODUCTION

The Errors–In–Variables (EIV) framework concerns static or dynamic systems whose input and output variables are affected by additive noise. These models play an important role in several engineering applications like, for example, time series modeling, direction–of–arrival estimation, blind channel equalization, and many other signal and image processing problems; see Van Huffel (1997); Van Huffel and Lemmerling (2002). Many different solutions have been proposed for the identification of EIV models. An overview can be found in Söderström (2007b).

Among these methods, the so–called Frisch scheme is one of the more interesting. Its roots are in Frisch (1934), with reference to static problems; the extension to dynamic systems was proposed in Beghelli et al. (1990). One of the main difficulties of the method concerns its application to real cases, when for several reasons (limited number of the available samples, non–linearities in the system, etc.) the assumptions behind the scheme are not exactly satisfied. In these cases the identification procedure does no more lead to a single solution, unless a selection criterion is introduced.

One of the first criteria that has shown remarkable robustness properties was originally proposed in Beghelli et al. (1993) and further in Diversi et al. (2004). This criterion, denoted here as Frisch–SR, relies on the shift properties of time–invariant systems and is based on rank deficiency conditions of the noise–free covariance matrix.

⋆ This research was partially supported by The Swedish Research Council, contract 621-2005-4207, and The Italian Ministry for University and Research.

Another robust criterion, characterized by a high level of estimation accuracy, was proposed in Diversi et al. (2003). This approach, denoted as Frisch–CM, relies on a comparison between the true and estimated statistical properties of the EIV system residuals.

A third criterion, characterized by a good compromise between computational efficiency and estimation accuracy, was introduced in Diversi et al. (2006). This method, denoted as Frisch–YW, relies on the properties of the high order Yule–Walker equations and is characterized by the same asymptotic properties of the Frisch–SR, as shown in Hong et al. (2007).

For the three Frisch alternatives, it is of interest to know which one gives the best estimation accuracy. This paper compares these criteria on the basis of the asymptotic covariance matrices of the estimates.

It will be shown that for moderate signal–to–noise ratios (SNR) no alternative is better than the other, both for the system parameters and noise variances. On the contrary, for high SNR the noise variances can, in general, be better estimated when the Frisch–CM is used.

2. PROBLEM STATEMENT AND NOTATIONS

Consider a linear and single input single output (SISO) system given by

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

where $u_0(t)$ and $y_0(t)$ are the noise–free input and output, respectively. Further, $A(q^{-1})$ and $B(q^{-1})$ are polynomials described as

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \cdots + a_{n_a}\, q^{-n_a} \\ B(q^{-1}) &= b_0 + b_1\, q^{-1} + \cdots + b_{n_b}\, q^{-n_b} \end{aligned} \tag{2}$$

We assume that the observations are corrupted by additive measurement noises $\tilde{u}(t)$ and $\tilde{y}(t)$. The available signals are in discrete time and of the form

$$u(t) = u_0(t) + \tilde{u}(t), \quad y(t) = y_0(t) + \tilde{y}(t). \tag{3}$$

The following assumptions are introduced.

A1. The dynamic system (1) is asymptotically stable, observable and controllable.
A2. The polynomial degrees n_a and n_b are a priori known.
A3. The true input $u_0(t)$ is a zero-mean stationary ergodic random signal, that is persistently exciting of sufficiently high order.
A4. The input noise $\tilde{u}(t)$ and the output noise $\tilde{y}(t)$ are both independent of $u_0(t)$ and mutually independent white Gaussian noise sequences of zero mean, and variances λ_u and λ_y, respectively.

The problem of identifying this EIV system is concerned with consistently estimating the parameter vector $\theta_0 = (a_1 \dots a_{n_a} \quad b_0 \dots b_{n_b})^T$ and the noise variances λ_u and λ_y from the measured noisy data $\{u(t), y(t)\}_{t=1}^{N}$. We introduce the regressor vector

$$\begin{aligned}\varphi(t) &= (-y(t-1) \cdots - y(t-n_a) \quad u(t) \dots u(t-n_b))^T \\ &= (-y_0(t-1) \cdots - y_0(t-n_a) \quad u_0(t) \dots u_0(t-n_b))^T \\ &\quad +(-\tilde{y}(t-1) \cdots - \tilde{y}(t-n_a) \quad \tilde{u}(t) \dots \tilde{u}(t-n_b))^T \\ &\triangleq \varphi_0(t) + \tilde{\varphi}(t), \end{aligned} \tag{4}$$

where $\varphi_0(t)$ and $\tilde{\varphi}(t)$ denote the noise-free part and the noise contribution part of $\varphi(t)$, respectively. For convenience, we utilize the extended regressor $\phi(t)$ and the true extended parameter vector Θ_0 as

$$\phi(t) = (-y(t) \quad \varphi^T(t))^T, \qquad \Theta_0 = (1 \quad \theta_0^T)^T. \tag{5}$$

In a similar way the extended vectors $\phi_0(t)$ and $\tilde{\phi}(t)$ can be defined. Some further expressions are introduced for the regressor vector and the system parameter vector, partitioned as

$$\begin{aligned}&\varphi_y(t) = (-y(t-1) \cdots - y(t-n_a))^T, \\ &\varphi_u(t) = (u(t) \dots u(t-n_b))^T, \quad \phi_y(t) = (-y(t) \ \varphi_y(t)^T)^T, \\ &\theta = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \ \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_{n_a} \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} b_0 \\ \vdots \\ b_{n_b} \end{pmatrix}, \ \bar{\mathbf{a}} = \begin{pmatrix} 1 \\ \mathbf{a} \end{pmatrix}. \end{aligned} \tag{6}$$

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

$$r_x(\tau) = \mathsf{E}(x(t)x(t-\tau)), \quad \tau = 0, \ \pm1, \ \pm2, \ \dots. \tag{7}$$

where E is the expectation operator. Further, the 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

$$R_{xy} = \mathsf{E}x(t)y^T(t), \quad r_{xz} = \mathsf{E}x(t)z(t). \tag{8}$$

3. THREE VARIANTS OF THE FRISCH SCHEME

The Frisch scheme was first proposed by Ragnar Frisch, Frisch (1934). It was developed to identify dynamic EIV systems in Beghelli et al. (1990) and was further elaborated in Beghelli et al. (1993); Diversi et al. (2003, 2004, 2006). Consider the relation

$$\phi_0^T(t)\Theta_0 = -A_0(q^{-1})y_0(t) + B_0(q^{-1})u_0(t) = 0. \tag{9}$$

It follows from (9) that

$$R_{\phi_0\phi_0}\Theta_0 = \mathsf{E}(\phi_0(t)\phi_0(t)^T)\Theta_0 = \mathbf{0}. \tag{10}$$

Hence matrix $R_{\phi_0\phi_0}$ is singular (positive semidefinite), with at least one eigenvalue equal to zero. Since it holds that $R_{\phi\phi} = R_{\phi_0\phi_0} + R_{\tilde{\phi}\tilde{\phi}}$, relation (10) can also be expressed as

$$\left(R_{\phi\phi} - R_{\tilde{\phi}\tilde{\phi}}\right)\Theta_0 = \mathbf{0}, \tag{11}$$

where

$$R_{\tilde{\phi}\tilde{\phi}} = \begin{pmatrix} \lambda_y I_{n_a+1} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{n_b+1} \end{pmatrix}. \tag{12}$$

The relations (11) and (12) are the basis for the Frisch method. They constitute a system of n_a+n_b+2 equations in n_a+n_b+3 unknowns. The solution of this system of equations can thus be expressed, in general, as a function of one variable, for example λ_u. In fact, if an estimate of λ_u is available, the value of λ_y that satisfies (11) is given by

$$\lambda_y = \lambda_{\min}(R_{\phi_y\phi_y} - R_{\phi_y\varphi_u}(R_{\varphi_u\varphi_u} - \lambda_u I_{n_b+1})^{-1}R_{\varphi_u\phi_y}), \tag{13}$$

where $\lambda_{\min}(R)$ denotes the smallest eigenvalue of R (see Beghelli et al. (1990)). Therefore, it can be stated that, in general, the solution of the nonlinear system of equations (11) and (12) is univocally determined if at least one additional equation is introduced. In this paper three alternatives will be considered.

- One choice is to evaluate the Frisch equations for an extended model by using an additional extended regressor $\underline{\varphi}(t)$. This method was proposed in Beghelli et al. (1993); Diversi et al. (2004).
- The second alternative is to compute residuals and compare their statistical properties with what can be predicted from the model, proposed in Diversi et al. (2003).
- The third alternative is to use the Yule-Walker equations, proposed recently in Diversi et al. (2006).

For the *first alternative*, denoted as Frisch–SR, the extended Frisch equation will be

$$(R_{\bar{\phi}\bar{\phi}} - R_{\bar{\tilde{\phi}}\bar{\tilde{\phi}}})\bar{\Theta}_0 = \mathbf{0}, \tag{14}$$

where

$$\bar{\phi} = \begin{pmatrix} \phi(t) \\ \underline{\varphi}(t) \end{pmatrix}, \quad \bar{\tilde{\phi}} = \begin{pmatrix} \tilde{\phi}(t) \\ \tilde{\underline{\varphi}}(t) \end{pmatrix}, \quad \bar{\Theta}_0 = \begin{pmatrix} \Theta_0 \\ \mathbf{0} \end{pmatrix}. \tag{15}$$

The model extension can, for example, mean that an additional A parameter is appended. In that case, $\underline{\varphi}(t) = -y(t-n_a-1)$. Another possibility is to append an additional B parameter, leading to $\underline{\varphi}(t) = u(t-n_b-1)$. Furthermore, it is also possible to let $\underline{\varphi}(t)$ be a vector. The number of new relations derived will be equal to the dimension of $\underline{\varphi}(t)$. In this alternative of the Frisch method, two functions $\lambda_y(\lambda_u)$ of type (13) are evaluated, referring to the nominal model and the extended one, respectively. They correspond to two curves in the (λ_u, λ_y) plane. The curves will ideally have one unique contact point, which defines the estimates, see Figure 1. See Beghelli et al. (1993); Diversi et al. (2004) for details.

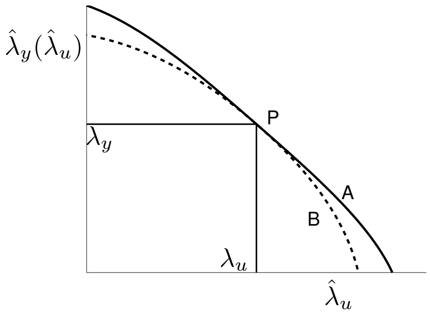

Fig. 1. Illustration of the principle for the Frisch estimation using an extended model.

For the *second alternative*, denoted as Frisch–CM, an additional relation is given as

$$\frac{d}{d\lambda_u} V_N(\lambda_u) \mid_{\lambda_u = \hat{\lambda}_u} = 0. \tag{16}$$

The estimated $\hat{\lambda}_u$ is determined as the minimizing element of the criterion

$$\hat{\lambda}_u = \arg\min_{\lambda_u} V_N(\lambda_u). \tag{17}$$

The criterion $V_N(\lambda_u)$ is defined as

$$V_N(\lambda_u) = \delta^T \Gamma \delta, \tag{18}$$

where Γ is a user chosen, positive definite weighting matrix. The vector δ is

$$\delta = \Big(\hat{r}_\epsilon(1) - \hat{r}_{\epsilon_0}(1) \cdots \hat{r}_\epsilon(m) - \hat{r}_{\epsilon_0}(m) \Big)^T. \tag{19}$$

Note that $\hat{r}_\epsilon(0) - \hat{r}_{\epsilon_0}(0)$ is not used because it automatically equals to zero Söderström (2007a). The maximum lag m used in (19) is to be chosen by the user. In expression (19), $\hat{r}_\epsilon(k)$ are the sample covariance elements

$$\hat{r}_\epsilon(k) = \frac{1}{N} \sum_{t=1}^{N} \epsilon(t, \hat{\theta}) \epsilon(t+k, \hat{\theta}). \tag{20}$$

where the residuals $\epsilon(t, \hat{\theta})$ are defined as

$$\epsilon(t, \hat{\theta}) = \hat{A}(q^{-1}) y(t) - \hat{B}(q^{-1}) u(t). \tag{21}$$

The theoretical covariance elements $\hat{r}_{\epsilon_0}(k)$ are based on the model

$$\epsilon_0(t) = \hat{A}(q^{-1}) \hat{\tilde{y}}(t) - \hat{B}(q^{-1}) \hat{\tilde{u}}(t), \tag{22}$$

where $\hat{\tilde{y}}(t)$ and $\hat{\tilde{u}}(t)$ are zero mean white noise sequences of variances $\hat{\lambda}_y$ and $\hat{\lambda}_u$, respectively.

In the *third variant*, denoted as Frisch–YW, the high order Yule-Walker equations are used Diversi et al. (2006). Similar as the extended model alternative, a regressor vector is introduced as

$$\underline{\varphi}(t) = (\ u(t - n_b - 1) \ \ldots \ u(t - n_b - p)\ )^T. \tag{23}$$

Because of Assumption **A4** and equation (9), we get the following high order Yule-Walker equations

$$(R_{\underline{\varphi}\phi})\Theta_0 = \mathbf{0} \ \Leftrightarrow \ (R_{\underline{\varphi}_0\phi_0})\Theta_0 = \mathbf{0}. \tag{24}$$

The noise variances λ_u and λ_y are then evaluated by searching the minimum of the cost function

$$J(\lambda_u, \lambda_y) = ||R_{\underline{\varphi}\phi}\Theta||^2 = \Theta^T R_{\underline{\varphi}\phi}^T R_{\underline{\varphi}\phi}\Theta, \tag{25}$$

where Θ is a function of λ_y and λ_u through (13) and (14).

We stress that there are different user choices for the three variants of Frisch. For Frisch–SR, we have freedom to choose the ways to extend the system model, such as adding one A or B parameter, or several A and/or B parameters. If the extended model has only one additional parameter, then the number of equations is equal to the number of unknowns. If the extended model has more additional parameter, we will choose not only the parameters included in the extended vector, but also some possible weightings. For Frisch–CM, we will choose the number of the residual lags m. If $m > 1$, we have overdetermined equations and a suitable weighting will also be chosen by us. Similarly, for Frisch–YW, the number and the type of the Yule-Walker equations and the possible weighting are the user choices. In general, for Frisch–CM and Frisch–YW, the number of equations is mostly larger than the number of the unknowns.

4. ALGORITHMIC ASPECTS

A recent analysis in Hong and Söderström (2007) has shown that the equations used in Frisch–SR and in the BELS method Zheng (1998) are equivalent when the same extended model is used. Under Assumption **A4**, the relations used in Frisch–SR equal to the following three equations

$$(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}})\,\theta_0 = r_{\varphi y}, \tag{26}$$

$$r_{y\varphi}\theta_0 = r_y(0) - \lambda_y, \tag{27}$$

$$R_{\underline{\varphi}\varphi}\theta_0 = r_{\underline{\varphi}y}. \tag{28}$$

The first two equations (26) and (27) are coming from the basic Frisch equations (11) and (12), and equation (28) can be derived from equation (14), which is used in Frisch–SR. See Hong and Söderström (2007) for a detailed proof.

In Frisch–YW, besides the basic Frisch equations (11) and (12), we use the equation (24), which can easily be further expressed as

$$(R_{\underline{\varphi}\phi})\Theta_0 = \mathbf{0} \ \Leftrightarrow \ (-r_{\underline{\varphi}y} \ R_{\underline{\varphi}\varphi}) \begin{pmatrix} 1 \\ \theta_0 \end{pmatrix} = \mathbf{0} \ \Leftrightarrow \ R_{\underline{\varphi}\varphi}\theta_0 = r_{\underline{\varphi}y},$$

i.e. equation (24) is also equivalent to equation (28). It follows that Frisch–SR and Frisch–YW are equivalent from the equations point of view providing that the same regressor vector $\underline{\varphi}(t)$ is used.

For the Frisch–CM, no explicit regressor vector $\underline{\varphi}(t)$ is used, and equation (16) can not be rewritten as (28). Frisch–CM is therefore different from Frisch–SR and Frisch–YW.

In all three Frisch methods we have a set of (overdetermined) nonlinear equations to solve. The statistically best way (in terms of covariance matrix of the parameter estimates) for the case of more equations than unknowns, is to solve all equations simultaneously in a weighted sense. The sets of nonlinear equations (26)–(28) or (26), (27), (16) can be written as

$$f(\vartheta) = \mathbf{0}, \tag{29}$$

where

$$\vartheta = (\theta^T, \theta_\lambda^T) \qquad \theta_\lambda = (\lambda_u, \lambda_y)^T. \tag{30}$$

Then the parameter vector ϑ can be estimated by

$$\hat{\vartheta} = \arg\min_{\vartheta} ||f(\vartheta)||_W^2 = \arg\min_{\vartheta} f^T(\vartheta) W f(\vartheta), \tag{31}$$

where W is a positive definite weighting matrix designed by the user. There is an optimal choice of the weighting matrix, see Söderström and Stoica (1989), but the optimal weighting is computationally rather complex to derive explicitly.

In the described Frisch scheme methods, on the contrary, a different approach is followed. In fact, some of the equations are forced to hold exactly and some others approximately. This can be formulated as an optimization problem with equality constraints. In particular, the basic Frisch equations (11)–(12), must hold exactly. The remaining equations, that depend on the specific criterion, will hold approximately.

The previous discussion concludes that the equations used in Frisch–SR and Frisch–YW are equivalent, while Frisch–CM is different. It means that the asymptotic statistical properties of Frisch–SR and Frisch–YW should be the same but differ from Frisch–CM. However, the methods have different performances depending not only on the equations that they use but also on the techniques utilized for finding the solution. For example, the BELS method Zheng (1998), which use a certain iterative algorithm, is not the best way to solve the set of equations. It can have convergence problems when the signal-to-noise ratio (SNR) is low. If the equations are solved using a variable projection algorithm then the performance will be improved, Söderström et al. (2005).

5. COMPARISON AND ANALYSIS OF THE ASYMPTOTIC COVARIANCE MATRICES OF THE ESTIMATES

For the three Frisch alternatives, it is of interest to know which one gives the best estimation accuracy. A statistical analysis of the accuracy can facilitate the evaluation and comparison of the methods. When the data number $N \to \infty$, an asymptotic covariance matrix of the parameter estimates is defined as

$$P \triangleq \lim_{N\to\infty} \left\{ N\mathsf{E}(\hat{\vartheta} - \vartheta_0)(\hat{\vartheta} - \vartheta_0)^T \right\}, \tag{32}$$

where $\hat{\vartheta}$ and ϑ_0 denote the estimate and the true value of ϑ, respectively. Assume that $\hat{\vartheta}$ is close to the true parameter vector ϑ_0 for large N. Then we linearize each equation used in the methods into the generic form

$$\alpha_\theta \tilde{\theta} + \alpha_{\lambda_u} \tilde{\lambda}_u + \alpha_{\lambda_y} \tilde{\lambda}_y \approx \beta. \tag{33}$$

The coefficients $\alpha_\theta, \alpha_{\lambda_u}, \alpha_{\lambda_y}$ are deterministic variables, while β is a random term which has zero mean and a variance that decreases when N increases. Under the given assumptions in Section 2, the estimated parameter $\hat{\vartheta}$ is asymptotically Gaussian distributed

$$\sqrt{N}(\hat{\vartheta} - \vartheta_0) \xrightarrow{dist} \mathcal{N}(0, P), \tag{34}$$

where

$$\begin{aligned} P &= \lim_{N\to\infty} N\mathsf{E}\left\{(\hat{\vartheta} - \vartheta_0)(\hat{\vartheta} - \vartheta_0)^T\right\} \\ &= (G^T W G)^{-1} G^T W Q W G (G^T W G)^{-1}. \end{aligned} \tag{35}$$

The coefficients $\alpha_\theta, \alpha_{\lambda_u}, \alpha_{\lambda_y}$ appear as elements of G. The block elements of Q are covariance matrices of the random terms β. A key step for realizing (35) is to consider $f(\vartheta)$ in (29) near the true value ϑ_0 and approximate it as

$$f(\vartheta) \approx f(\vartheta_0) + \frac{\partial f}{\partial \vartheta}(\hat{\vartheta} - \vartheta_0) \triangleq f(\vartheta_0) + G(\hat{\vartheta} - \vartheta_0). \tag{36}$$

If choosing the weighting matrix W in (31) as

$$W = Q^{-1}, \tag{37}$$

we will get the optimal minimal covariance matrix

$$P_{opt} = (G^T W G)^{-1}. \tag{38}$$

See Söderström and Stoica (1989) for a proof. However, this result is normally complicated to utilize in practice, because the matrix Q, which is parameter ϑ related, needs to be known first before using (37).

The asymptotic covariance matrix P of the estimates for Frisch–CM has been derived in Söderström (2007a). For the equations used in Frisch–SR and Frisch–YW, we already proved that they are equivalent to each other and also equivalent to the equations used in BELS providing the same additional regressor vector $\underline{\varphi}(t)$ is used. If these equation sets are treated in the same way (using the same weighting etc.), the asymptotic covariance matrices of the Frish–SR and Frisch–YW methods will be identical to that of the BELS methods, which has been given in Hong and Söderström (2007). Hence we have the explicit expressions of the asymptotic covariance matrices for all three Frisch methods. For simplicity, only Gaussian data are considered here. The results can be extended to handle more general data as shown in Söderström (2007a) and Hong and Söderström (2007).

In this section, we use the asymptotic theoretical covariance matrices derived in Hong and Söderström (2007) and Söderström (2007a) to numerically analyze the asymptotic estimation accuracy of the Frisch–SR and Frisch–CM methods by means of examples. For the Frisch–SR, we choose the extended vector as $\underline{\varphi}(t) = -y(t - n_a - 1)$. In Frisch–CM, the lag m equals 5 and the weighting matrix Γ is taken as in Diversi et al. (2003) and Söderström (2007a):

$$\Gamma = \text{diag}\ (\, 2m,\ 2(m-1),\ \ldots,\ 2\,). \tag{39}$$

Example 1. Consider a second-order system

$$(1 - 1.5q^{-1} + 0.7q^{-2}) y_0(t) = (2.0q^{-1} + 1.0q^{-2}) u_0(t), \tag{40}$$

where the noise-free input $u_0(t)$ is the ARMA(1,1) process

$$(1 - 0.5q^{-1}) u_0(t) = (1 + 0.7q^{-1}) e(t), \tag{41}$$

and $e(t)$ is a zero-mean white noise with unit variance. The variances of the white measurement noises at the input and output sides are equal to 1 and 4, respectively. Assume $\underline{\varphi}(t) = -y(t-3)$. Then the theoretical normalized asymptotic covariance matrix of the Frisch–SR scheme is

$$P_{\text{Frisch-SR}} = \begin{pmatrix} 0.33 & & & & & \\ -0.26 & 0.22 & & & & \\ -2.03 & 1.27 & 57.47 & & & \\ 3.42 & -2.32 & -58.07 & 69.99 & & \\ -0.14 & 0.25 & -29.88 & 21.36 & 96.54 & \\ -0.07 & -0.01 & 14.75 & -11.51 & -14.49 & 11.36 \end{pmatrix},$$

and Frisch–CM leads to

$$P_{\text{Frisch-CM}} = \begin{pmatrix} 0.44 & & & & & \\ -0.34 & 0.28 & & & & \\ -2.36 & 1.59 & 46.6 & & & \\ 4.45 & -3.17 & -52.5 & 73.3 & & \\ -0.68 & 0.63 & -22.1 & 11.6 & 97.1 & \\ 0.29 & -0.26 & 7.22 & -3.34 & -13.0 & 9.0 \end{pmatrix}.$$

The results of this example show that, for some parameters, using Frisch–CM method gives better estimation accuracy than using Frisch–SR, while for some other parameters Frisch–SR works better instead.

Comparisons with other numerical examples, Hong et al. (2007) have shown that, when both input and output sides have moderate SNR, the accuracies of the Frisch–SR and Frisch–CM estimates differ for all the parameters. No alternative is always better than the other. Depending on the system, the noise-free input signal, comparison criterion etc, one or the other version of the Frisch scheme may be considered to give the best result.

Next we will examine the accuracy properties of the three Frisch methods when both input and output SNR are high. Assume the noise free input $u_0(t)$ is an ARMA process

$$u_0(t) = \frac{C(q^{-1})}{D(q^{-1})} e(t), \tag{42}$$

where $e(t)$ is zero mean white noise with variance equals λ_e. Keeping λ_u and λ_y as constant and letting $\lambda_e \to \infty$, that is both input and output SNR tend to high values, we have the following two lemmas, whose proofs are reported in Hong et al. (2007). For simplicity, no weighting is considered here.

Lemma 1. For Frisch–CM, the G and Q matrices in (35) can be partitioned, according to (30), as follows

$$G_{\text{CM}} = \begin{pmatrix} \lambda_e M_{11} & M_{12} \\ \mathbf{0} & M_{22} \end{pmatrix}, \tag{43}$$

$$Q_{\text{CM}} = \begin{pmatrix} \lambda_e T_{11} + \tilde{T}_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}, \tag{44}$$

where all dependencies on λ_e are as shown. For the asymptotic covariance matrix P, which is expressed as

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{pmatrix}, \tag{45}$$

it follows that P_{11} depends on λ_e, and when λ_e becomes very large, that is for (very) large SNR,

$$\lim_{\lambda_e \to \infty} (\lambda_e P_{11}) = M_{11}^{-1} T_{11} M_{11}^{-T}. \tag{46}$$

For P_{12}, it holds that

$$\lambda_e P_{12} = M_{11}^{-1} T_{12} M_{22}^{-T} - M_{11}^{-1} M_{12} M_{22}^{-1} T_{22} M_{22}^{-T}. \tag{47}$$

Furthermore, the block element P_{22} does not depend at all on λ_e.

Lemma 2. For Frisch–SR and Frisch–YW, the G and Q matrices in (35) can be partitioned as

$$G_{\text{SR/YW}} = \begin{pmatrix} \lambda_e M_{11} & M_{12} \\ \lambda_e M_{21} & M_{22} \end{pmatrix}, \tag{48}$$

$$Q_{\text{SR/YW}} = \begin{pmatrix} \lambda_e T_{11} + \tilde{T}_{11} & \lambda_e T_{12} + \tilde{T}_{12} \\ \lambda_e T_{21} + \tilde{T}_{21} & \lambda_e T_{22} + \tilde{T}_{22} \end{pmatrix}. \tag{49}$$

(Note that the bock matrices M_{ij}, T_{ij} and $\tilde{T}_{ij}$ are not the same in (43), (44) as in (48), (49).) It follows that the block elements P_{11}, P_{12} and P_{22} of matrix (45) all depend on λ_e. When λ_e becomes very large, that is for (very) large SNR,

$$\lim_{\lambda_e \to \infty} (\lambda_e P_{11}) = (V_{11} T_{11} + V_{12} T_{21}) V_{11}^T + (V_{11} T_{12} + V_{12} T_{22}) V_{12}^T, \tag{50}$$

$$\lim_{\lambda_e \to \infty} P_{12} = (V_{11} T_{11} + V_{12} T_{21}) V_{21}^T + (V_{11} T_{12} + V_{12} T_{22}) V_{22}^T, \tag{51}$$

$$\lim_{\lambda_e \to \infty} P_{22} = \infty, \tag{52}$$

where

$$G^{-1} = \frac{1}{\lambda_e} \begin{pmatrix} V_{11} & V_{12} \\ \lambda_e V_{21} & \lambda_e V_{22} \end{pmatrix}. \tag{53}$$

We present an illustrative example to show the performance as stated by the preceding lemmas.

Example 2. Consider a first-order system given by

$$(1 - 0.8q^{-1}) y_0(t) = 2.0 q^{-1} u_0(t). \tag{54}$$

where $u_0(t)$ is the same as in Example 1. We increased the variance of the noise–free input λ_e from 1 to 10^8 and kept the variances of the measurement noises λ_u and λ_y as 1 and 2, respectively. For the asymptotic covariance matrices of the estimated parameters by using Frisch–SR and Frisch–CM, their block elements P_{11}, P_{12} and P_{22} are listed in Table 1 and Table 2. For Frisch–CM, the values of the following items were calculated as

$$M_{11}^{-1} T_{11} M_{11}^{-T} = \begin{pmatrix} 5.78e-02 & 1.74e-01 \\ 1.74e-01 & 2.19e+00 \end{pmatrix},$$

$$M_{11}^{-1} T_{12} M_{22}^{-T} - M_{11}^{-1} M_{12} M_{22}^{-1} T_{22} M_{22}^{-T} = \begin{pmatrix} -2.43e+00 & 7.94e-01 \\ -2.54e+01 & 8.30e+00 \end{pmatrix},$$

$$M_{22}^{-1} T_{22} M_{22}^{-T} = \begin{pmatrix} 9.48e+01 & -2.43e+01 \\ -2.43e+01 & 1.13e+01 \end{pmatrix},$$

and for Frisch–SR we have

$$(V_{11} T_{11} + V_{12} T_{21}) V_{11}^T + (V_{11} T_{12} + V_{12} T_{22}) V_{12}^T = \begin{pmatrix} 6.31e-02 & 2.41e-01 \\ 2.41e-01 & 3.0e+00 \end{pmatrix},$$

$$(V_{11} T_{11} + V_{12} T_{21}) V_{21}^T + (V_{11} T_{12} + V_{12} T_{22}) V_{22}^T = \begin{pmatrix} -2.31e-01 & 9.46e-02 \\ -2.71e+00 & 1.11e+00 \end{pmatrix}.$$

We see that the equations (46)-(47) and (50)-(51) in Lemmas 1 and 2 are well supported by the numerical results.

The preceeding analysis of the asymptotic covariance matrices of Frisch methods shows that the estimates for the system parameter θ for both Frisch–SR and Frisch–CM are good. The variances of the estimates decrease when

Table 1. The asymptotic covariance matrices of Example 2 with different SNR for Frisch–SR method. (Note: the blocks of the matrix, P_{11}, P_{12} and P_{22}, are given.)

λ_e	P_{11}		P_{12}		P_{22}	
1	6.84e-02	2.99e-01	-4.58e-01	1.61e-01	3.52e+01	-4.74e+00
	2.99e-01	3.70e+00	-5.44e+00	1.92e+00	-4.74e+00	5.23e+00
10	6.36e-03	2.47e-02	-2.53e-01	1.01e-01	1.14e+02	-3.69e+01
	2.47e-02	3.07e-01	-2.98e+00	1.19e+00	-3.69e+01	1.85e+01
100	6.32e-04	2.42e-03	-2.33e-01	9.53e-02	8.98e+02	-3.58e+02
	2.42e-03	3.01e-02	-2.73e+00	1.12e+00	-3.58e+02	1.50e+02
1000	6.31e-05	2.41e-04	-2.31e-01	9.47e-02	8.74e+03	-3.57e+03
	2.41e-04	3.00e-03	-2.71e+00	1.11e+00	-3.57e+03	1.47e+03
1e+08	6.31e-10	2.41e-09	-2.31e-01	9.46e-02	8.71e+08	-3.57e+08
	2.41e-09	3.04e-08	-2.71e+00	1.11e+00	-3.57e+08	1.46e+08

Table 2. The asymptotic covariance matrices of Example 2 with different SNR for Frisch–CM method. (Note: the blocks of the matrix, P_{11}, P_{12} and P_{22}, are given.)

λ_e	P_{11}		P_{12}		P_{22}	
1	1.33e-01	9.52e-01	-2.43e+00	7.94e-01	9.48e+01	-2.43e+01
	9.52e-01	1.03e+01	-2.54e+01	8.30e+00	-2.43e+01	1.13e+01
10	6.52e-03	2.52e-02	-2.43e-01	7.94e-02	9.48e+01	-2.43e+01
	2.52e-02	3.00e-01	-2.54e+00	8.30e-01	-2.43e+01	1.13e+01
100	5.85e-04	1.82e-03	-2.43e-02	7.94e-03	9.48e+01	-2.43e+01
	1.82e-03	2.27e-02	-2.54e-01	8.30e-02	-2.43e+01	1.13e+01
1000	5.79e-05	1.75e-04	-2.43e-03	7.94e-04	9.48e+01	-2.43e+01
	1.75e-04	2.20e-03	-2.54e-02	8.30e-03	-2.43e+01	1.13e+01
1e+08	5.78e-10	1.74e-09	-2.43e-08	7.94e-09	9.48e+01	-2.43e+01
	1.74e-09	2.19e-08	-2.54e-07	8.30e-08	-2.43e+01	1.13e+01

the SNR increases and tends to a limit. The two limits of the estimates by Frisch–SR and Frisch–CM are different. For the noise parameters λ_u and λ_y, in general, Frisch–CM gives better estimates than Frisch–SR. In Frisch–CM, the estimates for the noise variances λ_u and λ_y keep the same accuracy when the SNR increases. In Frisch–SR, the variances of the estimates for λ_u and λ_y continuously increase with increasing SNR.

REFERENCES

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

S. Beghelli, P. Castaldi, R. Guidorzi, and U. Soverini. A robust criterion for model selection in identification from noisy data. In *Proc. 9th International Conference on Systems Engineering*, pages 480–484, Las Vegas, Nevada, USA, 1993.

R. Diversi, R. Guidorzi, and U. Soverini. A new criterion in EIV identification and filtering applications. In *Proc. 13th IFAC Symposium on System Identification*, pages 1993–1998, Rotterdam, The Netherlands, August 27-29 2003.

R. Diversi, R. Guidorzi, and U. Soverini. Frisch scheme–based algorithms for EIV identification. In *Proc. 12th IEEE Mediterranean Conference on Control and Automation*, Kusadasi, Turkey, June 2004.

R. Diversi, U. Soverini, and R. Guidorzi. 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 (MTNS)*, pages 391–395, Kyoto, Japan, July 2006.

R. Frisch. Statistical confluence analysis by means of complete regression systems. Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway, 1934.

M. Hong and T. Söderström. Relations between bias–eliminating least squares, the Frisch scheme and extended compensated least squares methods for identifying errors–in–variables systems. 2007. Submitted.

M. Hong, T. Söderström, U. Soverini, and R. Diversi. Comparison of three Frisch methods for errors-in-variables identification. Technical Report 2007-021, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2007. Available as http://www.it.uu.se/research/publications/reports/2007-021.

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

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

T. Söderström and P. Stoica. *System Identification.* Prentice Hall International, Hemel Hempstead, UK, 1989.

T. Söderström, M. Hong, and W. X. Zheng. Convergence properties of bias-eliminating algorithms for errors-in-variables identification. *International Journal of Adaptive Control and Signal Processing*, 19(9):703–722, November 2005.

S. Van Huffel, editor. *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling.* SIAM, Philadelphia, USA, 1997.

S. Van Huffel and Ph. Lemmerling, editors. *Total Least Squares and Errors-in-Variables Modelling. Analysis, Algorithms and Applications.* Kluwer, Dordrecht, The Netherlands, 2002.

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