Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Accuracy analysis of the Frisch estimates for identifying errors-in-variables systems

Torsten Söderström
Division of Systems and Control,
Department of Information Technology, Uppsala University.
P O Box 337, SE-751 05 Uppsala, Sweden. Email: ts@it.uu.se

Abstract— **Several estimation methods have been proposed for identifying errors-in-variables systems, where both input and output measurements are corrupted by noise. One of the promising approaches is the so called Frisch scheme. This paper provides an accuracy analysis of the Frisch scheme applied to system identification. The estimates of the system parameters and the noise variances are shown to be asymptotically Gaussian distributed. An explicit expression for the covariance matrix of the asymptotic distribution is given as well. Numerical simulations support the theoretical results. A comparison with the Cramer-Rao lower bound is also given in examples, and it is shown that the Frisch scheme gives a performance close to the Cramer-Rao bound for large signal-to-noise ratios.**

I. INTRODUCTION

Many different solutions have been presented for system identification of linear dynamic systems from noise–corrupted output measurements see, for example, [6], [11]. Estimation of the parameters for linear dynamic systems when also the input is affected by noise ('errors–in–variables' models) is recognized as a more difficult problem.

The class of scientific disciplines which makes use of such representations is very broad, as proved by the several applications collected in [12], [13], such as time series modelling, array signal processing for direction–of–arrival estimation, blind channel equalization, multivariate calibration in analytical chemistry, image processing, astronomical data reduction, etc. In case of static systems, errors–in–variables representations are closely related to other well–known topics such as *latent variables* models and *factor* models [4].

Some comparisons between different approaches for errors-in-variables modelling are given in [10] and references therein.

The so called Frisch scheme is one of the more interesting approaches for the errors-in-variables identification. It has its roots in [3], where a regression problem was treated. It has been proposed for identifying dynamic systems in [1] and was further elaborated in [2]. So far, theoretical analysis has been limited to consistency. The aim of this paper is to provide such an analysis concerning the accuracy of the estimates obtained using the Frisch scheme.

This work was supported by The Swedish Research Council under contract 621-2002-4671.

II. PROBLEM STATEMENT AND NOTIONAL SETUP

A. Setup

As a typical model example, consider the system depicted in Figure 1 with noise-corrupted input and output measurements.

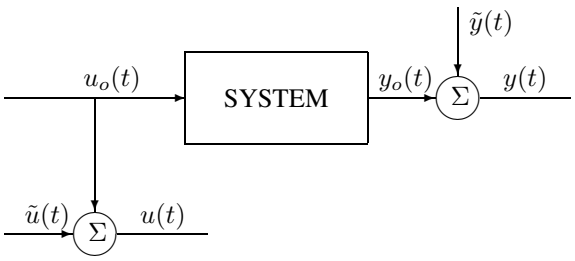

Fig. 1. The basic setup for a error-in-variables problem.

The noise–free input is denoted by $u_o(t)$ and the undisturbed output by $y_o(t)$. They are linked through the linear difference equation

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

where $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the backward shift operator q^{-1}, *i.e.* $q^{-1}\, x(t) = x(t-1)$ *etc.* More precisely,

$$\begin{array}{rcl} A(q^{-1}) & = & 1 + a_1 q^{-1} + \ldots + a_{na} q^{-na} \\ B(q^{-1}) & = & b_1 q^{-1} + \ldots + b_{nb} q^{-nb} \end{array} . \tag{2}$$

We assume that the observations are corrupted by additive measurement noises $\tilde{u}(t)$ and $\tilde{y}(t)$ of zero mean and variances λ_u^2 and λ_y^2, respectively. The available signals are of the form

$$\begin{array}{rcl} u(t) & = & u_o(t) + \tilde{u}(t) \\ y(t) & = & y_o(t) + \tilde{y}(t) \end{array} . \tag{3}$$

The general problem is to determine the system characteristics, *i.e.* the transfer function

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}. \tag{4}$$

In other words, the estimation problem is as follows. Given the noisy input-output data $u(1), y(1), \ldots u(N), y(N)$, determine an estimate of the (total) parameter vector

$$\vartheta = (a_1 \ldots a_{na}\ b_1 \ldots b_{nb}\ \lambda_y\ \lambda_u)^T . \tag{5}$$

0-7803-9568-9/05/$20.00 ©2005 IEEE

There are several estimation methods that can produce an estimate of ϑ:

- The bias-eliminating least squares method (BELS) is described in [15], [14].
- The Frisch scheme goes back to [3]. Its application to system identification is presented in [1]. An alternative implementation was proposed in [2]. This is the method that will be considered in this paper.
- A prediction error method or a maximum likelihood method can be applied. Its use for the errors-in-variables problem is described in [7], [10].

Once an estimation method is specified, it is of interest to examine its statistical properties. In this paper we will focus on the asymptotic covariance matrix

$$P = \lim_{N\to\infty} E\left\{N\mathrm{cov}(\hat{\vartheta}-\vartheta_o)(\hat{\vartheta}-\vartheta_o)^T\right\} \tag{6}$$

where $\hat{\vartheta}$ is the estimate of ϑ, and ϑ_o denotes the true value.

B. Assumptions

In order to proceed, some further assumptions must be introduced.

A1. The dynamic system (1) is asymptotically stable, *i.e.* $A(z)$ has all zeros outside the unit circle.

A2. All the system modes are observable and controllable, *i.e.* $A(z)$ and $B(z)$ have no common factors.

A3. The polynomial degrees na and nb are *a priori* known.

A4. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually uncorrelated, and uncorrelated with the noise–free signals $u_o(t)$ and $y_o(t)$.

A5. The sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero–mean Gaussian white noise sequences with variances λ_u and λ_y, respectively.

A6. The true input $u_o(t)$ is a zero–mean stationary ergodic random signal, that is persistently exciting at least of order $na+nb$.

C. Notations

The following notations will be convenient. The system parameter vector to be estimated is

$$\theta = (a_1 \dots a_{na}\ b_1 \dots b_{nb})^T. \tag{7}$$

Similarly we introduce the regressor vector

$$\varphi(t) = (-y(t-1)\dots -y(t-na)\ u(t-1)\dots u(t-nb))^T. \tag{8}$$

Further, we will use the conventions:

- θ_o denotes the true parameter vector, and $\hat{\theta}$ its estimate.
- Similarly, we let $A_o(q^{-1})$, $B_o(q^{-1})$, λ_u^o, λ_y^o, ϑ_o denote the true values of $A(q^{-1})$, $B(q^{-1})$, λ_u, λ_y, ϑ, respectively.
- $\varphi_o(t)$ denotes the noise-free part of the regressor vector:

$$\varphi_o(t) = (-y_o(t-1)\dots -y_o(t-na)\ u_o(t-1)\dots u_o(t-nb))^T. \tag{9}$$

- $\tilde{\varphi}(t)$ denotes the noise-contribution to the regressor vector. This means that

$$\tilde{\varphi}(t) = (-\tilde{y}(t-1)\dots -\tilde{y}(t-na)\ \tilde{u}(t-1)\dots \tilde{u}(t-nb))^T. \tag{10}$$

Sometimes it is very convenient to add a leading element to θ and to φ. For this reason we also introduce the extended regressor vector as

$$\overline{\varphi}(t) = \begin{pmatrix} -y(t) \\ \varphi(t) \end{pmatrix}, \tag{11}$$

and the extended parameter vector

$$\overline{\theta} = \begin{pmatrix} 1 \\ \theta \end{pmatrix}. \tag{12}$$

At other times it is useful to work with partitioned parameter and regression vectors. For this reason we introduce also $\mathbf{a}$ and $\mathbf{b}$ through

$$\theta = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_{na} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{nb} \end{pmatrix}, \tag{13}$$

and

$$\begin{aligned} \varphi(t) &= \begin{pmatrix} \varphi_y(t) \\ \varphi_u(t) \end{pmatrix}, \quad \varphi_y(t) = \begin{pmatrix} -y(t-1) \\ \vdots \\ -y(t-na) \end{pmatrix}, \\ \varphi_u(t) &= \begin{pmatrix} u(t-1) \\ \vdots \\ u(t-nb) \end{pmatrix}. \end{aligned} \tag{14}$$

Extended versions of the partioned vectors will also be handy:

$$\overline{\theta} = \begin{pmatrix} \overline{\mathbf{a}} \\ \mathbf{b} \end{pmatrix}, \quad \overline{\mathbf{a}} = \begin{pmatrix} 1 \\ \mathbf{a} \end{pmatrix}, \tag{15}$$

$$\overline{\varphi}(t) = \begin{pmatrix} \overline{\varphi}_y(t) \\ \varphi_u(t) \end{pmatrix}, \quad \overline{\varphi}_y(t) = \begin{pmatrix} -y(t) \\ \varphi_y(t) \end{pmatrix}. \tag{16}$$

Cross-covariance matrices between two vectors $x(t)$ and $y(t)$ are denoted

$$R_{xy} = E\left\{x(t)y^T(t)\right\}, \tag{17}$$

and their natural estimates are denoted as

$$\hat{R}_{xy} = \frac{1}{N}\sum_{t=1}^{N} x(t)y^T(t). \tag{18}$$

The covariance matrices are often partioned in a way compatible with the partioning of the vectors. For example,

$$\hat{R}_{\overline{\varphi}} = \begin{pmatrix} \hat{R}_{\overline{\varphi}_y} & \hat{R}_{\overline{\varphi}_y\varphi_u} \\ \hat{R}_{\varphi_u\overline{\varphi}_y} & \hat{R}_{\varphi_u} \end{pmatrix}. \tag{19}$$

III. THE FRISCH SCHEME

A. Basic Relations

The relations presented in this subsection are fundamental when deriving the Frisch estimator.

First we note that

$$\overline{\varphi}_o^T(t)\overline{\theta}_o = -A_o(q^{-1})y_o(t) + B_o(q^{-1})u_o(t) = 0. \quad (20)$$

Further it holds that

$$R_\varphi = R_{\varphi_o} + R_{\tilde{\varphi}}, \qquad R_{\overline{\varphi}} = R_{\overline{\varphi}_o} + R_{\tilde{\overline{\varphi}}}. \quad (21)$$

It follows from (20) that

$$R_{\overline{\varphi}_o}\overline{\theta}_o = E\left\{\overline{\varphi}_o\overline{\varphi}_o^T\overline{\theta}_o\right\} = \mathbf{0}. \quad (22)$$

Hence the matrix $R_{\overline{\varphi}_o}$ is singular (positive semidefinite), with at least one eigenvalue equal to zero. The corresponding eigenvector is $\overline{\theta}_o$. One can show that under the general assumptions **A2** and **A6**, the matrix $R_{\overline{\varphi}_o}$ will in fact have only one eigenvalue in the origin.

The noise covariance matrix has a simple structure, as

$$R_{\tilde{\overline{\varphi}}} = \begin{pmatrix} \lambda_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{nb} \end{pmatrix}. \quad (23)$$

The relation (22) is the basis for the Frisch method. The idea is to have appropriate estimates of the noise variances and then determine the parameter vector θ from

$$\left(\hat{R}_{\overline{\varphi}} - \hat{R}_{\tilde{\overline{\varphi}}}\right)\hat{\overline{\theta}} = \mathbf{0}. \quad (24)$$

B. Determining $\hat{\lambda}_y$ and $\hat{\theta}$

Assume for the time being that an estimate $\hat{\lambda}_u$ of the input noise variance is available. Then the output noise variance λ_y is determined so that the matrix appearing in (24) is singular. More specifically, we have the following result.

Lemma 3.1. Let the estimate $\hat{\lambda}_u$ satisfy

$$0 \le \hat{\lambda}_u \le \lambda_{\min}\left(\hat{R}_{\varphi_u} - \hat{R}_{\varphi_u\overline{\varphi}_y}\hat{R}_{\overline{\varphi}_y}^{-1}\hat{R}_{\overline{\varphi}_y\varphi_u}\right), \quad (25)$$

where $\lambda_{\min}(C)$ denotes the minimal eigenvalue of the symmetric matrix C. Define

$$\hat{\lambda}_y = \lambda_{\min}\left(\hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y\varphi_u}\left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb}\right)^{-1}\hat{R}_{\varphi_u\overline{\varphi}_y}\right). \quad (26)$$

Then the matrix

$$C = \begin{pmatrix} \hat{R}_{\overline{\varphi}_y} & \hat{R}_{\overline{\varphi}_y\varphi_u} \\ \hat{R}_{\varphi_u\overline{\varphi}_y} & \hat{R}_{\varphi_u} \end{pmatrix} - \begin{pmatrix} \hat{\lambda}_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix} \quad (27)$$

is positive semidefinite with one eigenvalue in the origin.

Proof: See [9]. ∎

An essential part of the Frisch algorithm is based on Lemma 3.1. Assume that an estimate $\hat{\lambda}_u$ of the input noise variance is available (how this estimate is to be found will be described in the Section III-C). The estimate $\hat{\lambda}_y$ is then found from (26). The estimate of the parameter vector θ is next determined by solving equations $2, \ldots, na+nb+1$ of

$$C\hat{\overline{\theta}} = \mathbf{0}, \quad (28)$$

where the matrix C is given by (27). As C by construction is singular, this means that $\hat{\theta}$ is the solution to

$$\left(\hat{R}_\varphi - \begin{pmatrix} \hat{\lambda}_y I_{na} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix}\right)\hat{\theta} = \hat{r}_{\varphi y}. \quad (29)$$

C. Determination of $\hat{\lambda}_u$

What remains is to determine $\hat{\lambda}_u$. Different alternatives have been proposed:

- In [1], the function $\hat{\lambda}_y(\hat{\lambda}_u)$ is evaluated both for the nominal model and for an extended model, adding one A or one B parameter (or both). The functions correspond to curves in the $(\hat{\lambda}_u, \hat{\lambda}_y)$ plan. The curves will ideally intersect in one unique point, which defines the estimates.
- Another alternative is to compute residuals, and compare their statistical properties with what can be predicted from the model. This alternative was proposed in [2] and is the option analysed in this paper. It is described below.

Define the residuals

$$\varepsilon(t,\hat{\theta}) = \hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t) \quad (30)$$

and compute sample covariance elements

$$\hat{r}_\varepsilon(k) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\hat{\theta})\varepsilon(t+k,\hat{\theta}). \quad (31)$$

Compute also theoretical covariance elements $\hat{r}_{\varepsilon_o}(k)$ based on the model

$$\varepsilon_o(t) = \hat{A}(q^{-1})\hat{\tilde{y}}(t) - \hat{B}(q^{-1})\hat{\tilde{u}}(t), \quad (32)$$

where

$$E\left\{\hat{\tilde{y}}^2(t)\right\} = \hat{\lambda}_y, \qquad E\left\{\hat{\tilde{u}}^2(t)\right\} = \hat{\lambda}_u.$$

Next, define a criterion for comparing $\{\hat{r}_\varepsilon(k)\}$ and $\{\hat{r}_{\varepsilon_o}(k)\}$. A fairly general way to do this is to take

$$V_N(\hat{\lambda}_u) = \delta^T W \delta \quad (33)$$

where W is a user chosen, positive definite weighting matrix and the vector δ is

$$\delta = \begin{pmatrix} \hat{r}_\varepsilon(1) - \hat{r}_{\varepsilon_o}(1) \\ \vdots \\ \hat{r}_\varepsilon(m) - \hat{r}_{\varepsilon_o}(m) \end{pmatrix}. \quad (34)$$

The maximum lag m used in (34) is to be chosen by the user. The estimate $\hat{\lambda}_u$ is determined as the minimizing element of the criterion

$$\hat{\lambda}_u = \arg\min_{\lambda_u} V_N(\lambda_u). \quad (35)$$

We hence have

$$\frac{d}{d\lambda_u}V_N(\lambda_u)\Big|_{\lambda_u=\hat{\lambda}_u} = 0. \quad (36)$$

Remark 1 As shown in [9], it holds that $\hat{r}_\varepsilon(0) = \hat{r}_{\varepsilon_o}(0)$. There is hence no point including the time argument 0 in (34). □

Remark 2. A criterion such as (33) was first proposed in [2] although the description followed a slightly different form. Further, a leading element of δ with time argument 0 was included, and the weighting matrix was chosen as

$$W = \mathrm{diag}(m+1, 2m, 2(m-1), \ldots, 2) \tag{37}$$

□

In summary the Frisch scheme algorithm consists of the equations (26), (29) and (35). In its implementation, there is an optimization over one variable, $\hat{\lambda}_u$, in (35). In the evaluation of the loss function $V_N(\hat{\lambda}_u)$, also (26) and (29) are used to get $\hat{\lambda}_y$ and $\hat{\theta}$, respectively.

IV. ASYMPTOTIC DISTRIBUTION

The main result of the paper is as follows.

Theorem 4.1. Under the given assumptions of Section 2.2 the parameter estimates $\hat{\vartheta}$ are asymptotically Gaussian distributed

$$\sqrt{N}(\hat{\vartheta} - \vartheta_o) \xrightarrow{\text{dist}} \mathcal{N}(0, P), \tag{38}$$

where

$$P = R^{-1} Q R^{-T} \tag{39}$$

and

$$R = \begin{pmatrix} R_{\varphi_o} & \begin{pmatrix} -\mathbf{a}_o \\ \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} \\ -\mathbf{b}_o \end{pmatrix} \\ \mathbf{0} & \overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o & \mathbf{b}_o^T \mathbf{b}_o \\ \mathbf{0} & \sum_{k=1}^m \gamma_k \sum_i a_i^o a_{i+k}^o & \sum_{k=1}^m \gamma_k \sum_i b_i^o b_{i+k}^o \end{pmatrix}, \tag{40}$$

$$Q = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix}. \tag{41}$$

In (40) the numbers $\{\gamma_k\}_{k=1}^m$ are defined by the vector-matrix relation

$$\gamma = W\beta$$

where W is introduced in (33) and the vector β is given elementwise as

$$\beta_k = -\sum_i b_i^o b_{i+k}^o + \sum_i a_i^o a_{i+k}^o \frac{\mathbf{b}_o^T \mathbf{b}_o}{\overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o}, \qquad k = 1, \ldots, m. \tag{42}$$

The blocks in the symmetric matrix Q are as follows

$$\begin{aligned} Q_{11} &= \sum_\tau R_{\varphi_o}(\tau) r_\varepsilon(\tau) \\ &\quad + \sum_\tau \left[R_{\tilde{\varphi}}(\tau) r_\varepsilon(\tau) + r_{\tilde{\varphi}\varepsilon}(\tau) r_{\tilde{\varphi}\varepsilon}^T(-\tau) \right], \end{aligned} \tag{43}$$

$$Q_{12} = 2\sum_\tau r_{\tilde{\varphi}\varepsilon}(\tau) r_\varepsilon(\tau), \tag{44}$$

$$\begin{aligned} Q_{13} &= \sum_{k=1}^m \gamma_k \left\{ \sum_\tau [r_{\tilde{\varphi}\varepsilon}(\tau) r_\varepsilon(\tau - k) \right. \\ &\quad \left. + r_{\tilde{\varphi}\varepsilon}(\tau - k) r_\varepsilon(\tau)] \right\}, \end{aligned} \tag{45}$$

$$Q_{22} = 2\sum_\tau r_\varepsilon^2(\tau), \tag{46}$$

$$Q_{23} = 2\sum_{k=1}^m \gamma_k \sum_\tau [r_\varepsilon(\tau) r_\varepsilon(\tau - k)], \tag{47}$$

$$\begin{aligned} Q_{33} &= \sum_{k=1}^m \sum_{j=1}^m \gamma_k \gamma_j \sum_\tau [r_\varepsilon(\tau) r_\varepsilon(\tau + k - j) \\ &\quad + r_\varepsilon(\tau - j) r_\varepsilon(\tau + k)]. \end{aligned} \tag{48}$$

The covariance elements satisfy

$$r_\varepsilon(k) = \lambda_y^o \textstyle\sum_i a_i^o a_{i+k}^o + \lambda_u^o \sum_i b_i^o b_{i+k}^o \ , \tag{49}$$

$$r_{\tilde{\varphi}\varepsilon}(k) = -\begin{pmatrix} \lambda_y^o \begin{pmatrix} a_{1-k}^o \\ \vdots \\ a_{na-k}^o \end{pmatrix} \\ \lambda_u^o \begin{pmatrix} b_{1-k}^o \\ \vdots \\ b_{nb-k}^o \end{pmatrix} \end{pmatrix}. \tag{50}$$

Note that in (40), (42), (49) and (50) the conventions

$$a_i^o = \begin{cases} 1 & i = 0 \\ 0 & i > na, \ i < 0 \end{cases}, \tag{51}$$

$$b_i^o = 0 \quad i > nb, \ i \le 0, \tag{52}$$

are applied. In particular, all the covariance elements in (49) and (50) are zero as soon as

$$|k| > \max(na, nb - 1). \tag{53}$$

The summations over τ in (43) – (48) are over all values making the terms nonzero. Due to the condition in (53), each sum will have only a modest number of nonzero terms.

Proof: The main idea of the proof is to linearize equations (26), (29), (36) for large N. Assuming $\hat{\vartheta}$ is close to ϑ_o, each of the three equations is linearized into the generic form

$$A_\theta(\hat{\theta} - \theta_o) + A_y(\hat{\lambda}_y - \lambda_y^o) + A_u(\hat{\lambda}_u - \lambda_u^o) \approx A_s \tag{54}$$

The 'coefficients' $A_\theta,\ A_y,\ A_u$ appear as block elements of R. The block elements of Q are the covariance matrices of the random terms A_s. For all details of the proof, see [9]. ■

V. NUMERICAL ILLUSTRATION

We first illustrate the findings in the paper by comparing the theoretical expressions of the covariance matrix with simulations. Next, we compare the theoretical expressions with the Cramer-Rao lower bound. We will consider two systems.

System S1. This system is of first order and given by

$$(1 - 0.8q^{-1})y_o(t) = 2.0q^{-1}u_o(t). \tag{55}$$

Further, the noise-free input is assumed to be an ARMA(1,1) process

$$\begin{aligned}(1 - 0.5q^{-1})u_o(t) &= (1 + 0.7q^{-1})e(t),\\ E\{e(t)e(s)\} &= \lambda_v^o\delta_{t,s}.\end{aligned} \tag{56}$$

The noise levels are

$$\lambda_v^o = 1, \qquad \lambda_u^o = 1, \qquad \lambda_y^o = 1, \tag{57}$$

resulting in signal-to-noise ratios on the input and output sides, respectively

$$\mathrm{SNR}_u = 5.82 \text{ dB}, \qquad \mathrm{SNR}_y = 10.55 \text{ dB} \tag{58}$$

□

System S2. This system is of second order, given by

$$(1-1.5q^{-1}+0.7q^{-2})y_o(t) = (2.0q^{-1}+1.0q^{-2})u_o(t), \tag{59}$$

and the noise-free input $u_o(t)$ is still assumed to be described by (56). The noise levels are

$$\lambda_u^o = 1, \qquad \lambda_y^o = 4 \tag{60}$$

resulting in signal-to-noise ratios on the input and output sides, respectively

$$\mathrm{SNR}_u = 5.82 \text{ dB}, \qquad \mathrm{SNR}_y = 14.40 \text{ dB}. \tag{61}$$

□

Example 1. We first compare the theoretical expression (39) for the covariance matrix with sample covariance matrices obtained from a Monte Carlo simulation, In this study the number of data points was $N = 10000$, and we used $M = 100$ realizations. We used $m = 1$ in the criterion (33), and hence W has no significance in this particular case. The following results were obtained.

For System S1, the theoretical normalized covariance matrix is

$$P = \begin{pmatrix} 0.105 & & & \\ 0.68 & 7.4 & & \\ -1.6 & -16 & 53 & \\ 0.55 & 5.7 & -16 & 8.4 \end{pmatrix}. \tag{62}$$

The corresponding result from simulations was

$$P_{\mathrm{sim}} = \begin{pmatrix} 0.106 & & & \\ 0.75 & 8.7 & & \\ -1.8 & -20 & 64 & \\ 0.64 & 6.7 & -17 & 8.2 \end{pmatrix}. \tag{63}$$

As can be seen, there is a good agreement between theory and simulations. According to Appendix B.9 of [11] a deviation of 20 % is certainly reasonable when P_{sim} is based on 100 realizations.

The procedure was repeated for system S2. Then the results were

$$P = \begin{pmatrix} 0.43 & & & & & \\ -0.33 & 0.27 & & & & \\ -2.5 & 1.7 & 46 & & & \\ 4.4 & -3.2 & -53 & 73 & & \\ -0.3 & 0.3 & -20 & 12 & 92 & \\ 0.1 & -0.1 & 6 & -4 & -10 & 7.5 \end{pmatrix},$$

$$P_{\mathrm{sim}} = \begin{pmatrix} 0.41 & & & & & \\ -0.30 & 0.24 & & & & \\ -2.9 & 1.8 & 54 & & & \\ 4.7 & -3.1 & -61 & 82 & & \\ 0.1 & -0.1 & -27 & 18 & 86 & \\ -0.2 & 0.1 & 9 & -7 & -9 & 7.6 \end{pmatrix}.$$

Again, the agreement between theory and simulation is fairly good. □

Example 2. In this example we compare the theoretical covariance expressions with the Cramer-Rao lower bound. The latter gives a lower bound on the covariance matrix, that seems only achievable with a full maximum likelihood approach, which is quite costly computationally.

For the two given systems S1 and S2 the covariance matrices P, given by (39) and the Cramer-Rao bound P_{CRB} were evaluated numerically. The Cramer-Rao bound can be computed using either a polynomial-based framework, see [5], or using a state-space based formalism, with details given in [8]. (Both approaches give identical results).

As an illustration, we show in Figures 2 and 3 how the true transfer function as well as its theoretical standard deviations for the Frisch scheme and the Cramer-Rao lower bound vary with frequency. More precisely, we plotted the *normalized* relative error

$$\frac{\sigma_G}{|G|} = \frac{\sqrt{E\{|\Delta G|^2\}}}{|G|} \tag{64}$$

versus angular frequency ω. In (64) the error ΔG is defined as

$$\Delta G = G(e^{i\omega}, \hat{\theta}) - G(e^{i\omega}, \theta_o) \tag{65}$$

and is assumed to be small. Note that the relative error is obtained by dividing the expression (64) with $\sqrt{N}$. It is seen that the statistical error when the Frisch scheme is used is indeed larger than the Cramer-Rao lower bound, but the difference is rather small, in particular for low frequencies. □

Example 3. We next examined how the parameter variances vary with the signal-to-noise-ratio. More specifically, we varied λ_v^o and thereby the variance of the unperturbed input, $E\{u_o^2(t)\}$. We show in Figures 4 and 5 how the diagonal elements of P varies with λ_v^o. Again, the results confirm that the Frisch scheme gives estimates that are less accurate than

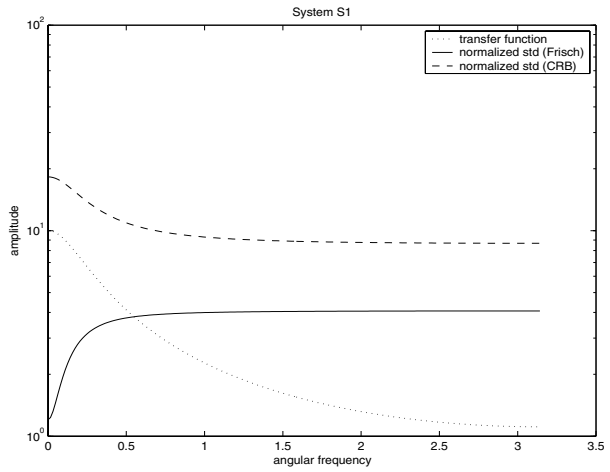

Fig. 2. Frequency response of the transfer function (dotted) for System S1, standard deviation of the normalized relative error using the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed).

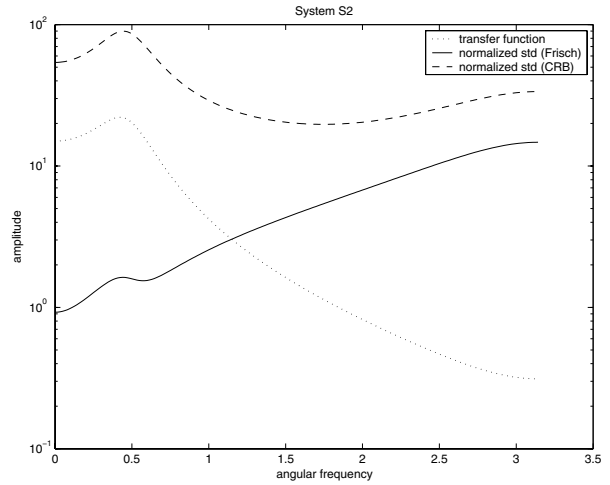

Fig. 3. Frequency response of the transfer function (dotted) for System S2, standard deviation of the normalized relative error using the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed).

the Cramer-Rao bound. The difference can be significant for low values of λ_v^o, that is for low signal-to-noise ratios, and is typically quite small for large values of λ_v^o. □

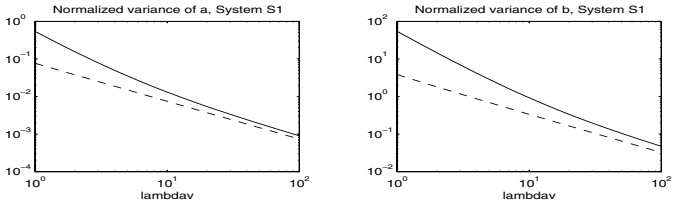

Fig. 4. Normalized variances of $\hat{a}$ and $\hat{b}$, System S1: the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed). (The true variances are obtained after division by N.)

VI. CONCLUSIONS

The Frisch approach for identifying error-in-variables systems has been analysed. In particular, the asymptotic co-

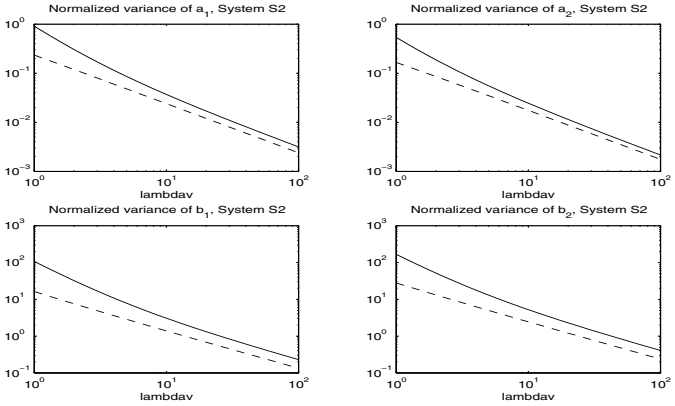

Fig. 5. Normalized variances of $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$ and $\hat{b}_2$, System S2: the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed). (The true variances are obtained after division by N.)

variance matrix of the parameter estimates has been derived. This matrix has also been compared to the Cramer-Rao lower bound, and it has been shown that the differences are small when the signal-to-noise ratio is high.

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 *13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August 27-29 2003.

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

[4] W. A. Fuller. *Measurement Error Models*. Wiley, New York, NY, 1987.

[5] E Karlsson, T Söderström, and P Stoica. Computing the Cramér-Rao lower bound for noisy input output systems. *Signal Processing*, 80(11):2421–2447, November 2000.

[6] L. Ljung. *System Identification - Theory for the User, 2nd edition*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[7] T. Söderström. Identification of stochastic linear systems in presence of input noise. *Automatica*, 17:713–725, 1981.

[8] T Söderström. On computing the asymptotic accuracy of estimated linear state space models. Technical report, UPTEC 97031 R, School of Engineering, Uppsala University, Uppsala, Sweden, February 1997.

[9] T. Söderström. Accuracy analysis of the Frisch scheme for identifying errors-in-variables systems. Technical Report 2005-005, Department of Information Technology, Uppsala University, Uppsala, Sweden, February 2005.

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

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

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

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

[15] 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.