# ERRORS-IN-VARIABLES METHODS IN SYSTEM IDENTIFICATION

**Torsten Söderström**

*Div. of Systems and Control, Dept. of Information Technology, Uppsala University, Uppsala, Sweden, email: ts@it.uu.se*

Abstract: The paper gives a survey of errors-in-variables methods in system identification. Background and motivation are given, and examples illustrate why the identification problem can be difficult. Under general weak assumptions, the systems are not identifiable, but can be parameterized using one degree of freedom. Examples where identifiability is achieved under additional assumptions are also provided. A number of approaches for parameter estimation of errors-in-variables models are presented. The underlying assumptions and principles for each approach are highlighted.

Keywords: system identification, errors-in-variables, parameter estimation, identifiability, consistency

## 1. INTRODUCTION

Many different solutions have been presented for system identification of linear dynamic systems from noise–corrupted output measurements see, for example, Ljung (1999), Söderström and Stoica (1989). On the other hand, estimation of the parameters for linear dynamic systems where also the input is affected by noise is recognized as a more difficult problem. Representations where errors or measurement noises are present on both inputs and outputs are usually called "errors–in–variables" (EIV) models. They play an important role when the purpose is determination of the physical laws that describe the process, rather than the prediction of its future behavior.

When searching on the internet for 'errors-in-variables' Google gives more than 80000 hits. Various publication data bases such as Science Citation Index or Elsevier Science Direct give a few hundred different references each on the subject. The area is therefore quite extensive. The vast majority of papers are written from an application perspective, and can deal with biomedicine, chemistry, chemical engineering, earth sciences, econometrics, managements, mechanical engineering, finance, ecology, geoscience, image systems, time series analysis, etc. Most of the papers published in the automatic control journals and conference proceedings focus on methodology. So do also some of the many papers in various statistical journals. 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, Fuller (1987), Van Schuppen (1989), Scherrer and Deistler (1998).

Errors-in-variables models can be motivated in several situations. One such case is the modeling of the dynamics between the noise-free input and noise-free output. The reason can be to have a better understanding of the underlying relations, rather than to make a good prediction from noisy data. This is the 'classical' motivation used in econometrics and some other areas. In some applications, perhaps typically so in non-technical areas such as biology, economics, environment, it may be useful to regard the identification experiment as designed by somebody else, and the modeler has to work with given recorded input and output data. Another situation is when a high-dimensional data vector is to be approximated by a small number of factors, which is the standard

when the user lacks enough information to classify the available signals into inputs and outputs, and prefer to use a 'symmetric' system model. This is closely connected to the behavioral approach to modeling, Willems (1986), Heij et al. (1997), Markovsky et al. (2005), Markovsky et al. (2006b). We will return to the issue when EIV problems occur in Section 2.

With reference to these systems, the assumptions (*prejudices*) which lie behind the identification procedure have been thoroughly analyzed in Kalman (1982a), Kalman (1982b) with particular attention to the Frisch scheme, Frisch (1934). This scheme assumes that each variable is affected by an unknown amount of additive noise and each noise component is independent of every other noise component and of every variable. As a consequence, in this case the solution is constituted by a whole family of models compatible with the set of noisy data, unlike other traditional approaches where the solution is characterized by a single model. Kalman's work relates to the static, mainly multivariable case. He uses the term 'prejudice' for non-errors-in-variables approaches of modelling dynamic systems.

There is also a rich literature dealing with the EIV problem for the static case, which falls outside the scope of this paper. Some classical work on EIV include Adcock (1877), Adcock (1878), Frisch (1934), Koopmans (1937), Reiersøl (1950) and others. Extensive analysis is given in Anderson (1984). The topic is also well treated from different points of view in the books Cheng and Ness (1999), Fuller (1987). The Frisch scheme, Frisch (1934), has been analyzed (for the static case) in Guidorzi (1991), Beghelli and Soverini (1992) and Guidorzi (1995). Other works deal with 'errors-in-variables filtering'. This refers to filtering problems, where both input and output measurements are contaminated by noise. This topic is treated, for example, in Guidorzi et al. (2003), Markovsky et al. (2002).

The paper is organized as follows. The problem is described in the next section. Identifiability is discussed from different points of view in Sections 3-5. The Cramér-Rao bound for unbiased estimators is discussed in Section 6, and basic notations can be found in Section 8. A general characterization of approaches is given in Sections 7 and 9-15. Some concluding remarks appear in Section 16.

## 2. THE ERRORS-IN-VARIABLES PROBLEM FOR DYNAMIC SYSTEMS

As a typical model example, consider the linear single-input single-output (SISO) system depicted in Figure 1 with noise-corrupted input and output measurements. The paper is mainly mentioned when extensions to multi-input multi-output (MIMO) systems are possible and straightforward.

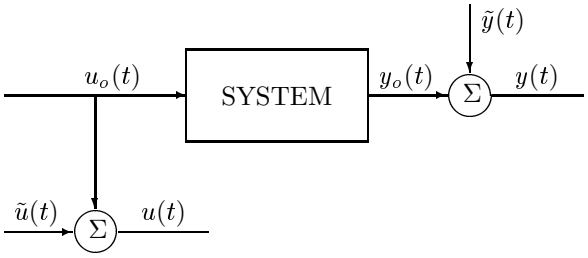


Fig. 1. The basic setup for a dynamic errors-in-variables problem.

The noise–free input is denoted by $u_o(t)$ and the undisturbed output by $y_o(t)$. 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

$$\begin{aligned} u(t) &= u_o(t) + \tilde{u}(t), \\ y(t) &= y_o(t) + \tilde{y}(t). \end{aligned} \tag{1}$$

The general problem is to determine the system characteristics, such as the system transfer function.

In order to proceed, some further assumptions must be introduced. To some degree they can be made more or less restrictive. In the coming subsections we therefore introduce a number of assumptions, that are partly alternative ones. Then it will be discussed what assumptions that are necessary for different results.

Concerning the system, the following assumption will be imposed.

**AS1**. The system is single-input single-output (SISO), linear and asymptotically stable. □

In many, but not all, cases we also use

**AS2**. The system is causal, so $y_o(t)$ depends on $u_o(s)$ for $s \leq t$, but not on future values of $u_o(\cdot)$. □

In most cases $t$ will denote time, and then it is most natural to require the system to be causal as in **AS2**. However, the estimation techniques can also be applied if $t$ has the meaning, say, of a spatial variable, and the model describes some cross-directional property of a material. For such cases, non-causal models make perfect sense. Also in the case of a 'symmetric' approach, assumption **AS2** can be dispensed with. In this case, rather than using the causality between $u_o(t)$ and $y_o(t)$ implied by Figure 1, write the measured data as

$$z(t) = z_o(t) + \tilde{z}_o(t), \qquad z_o(t) = \begin{pmatrix} y_o(t) \\ u_o(t) \end{pmatrix}. \tag{2}$$

The noise-free data, $z_o(t)$ is assumed to fulfil a relation

where $H(q^{-1})$ is an asymptotically stable filter. However, we do not require that $y_o(t)$ can be solved as a function of past values of the noise-free input $u_o(s), s \leq t$, using (3).

Next introduce assumptions on the noise and the noise-free input.

**AN1**. The noise sequences $\tilde{u}(t)$, $\tilde{y}(t)$ are stationary processes, with zero mean values and spectra $\phi_{\tilde{u}}(\omega)$ and $\phi_{\tilde{y}}(\omega)$, respectively. Further, $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually uncorrelated. [In a few cases to be specified later, the noise sequences are allowed to be cross-correlated.] □

**AI1**. The true input $u_o(t)$ is a stationary process of zero mean, with spectral density $\phi_{u_o}(\omega)$. The input $u_o(t)$ is assumed to be persistently exciting of a suitable order, which means that $\phi_{u_o}(\omega) > 0$ for a sufficient number of frequencies. Further, $u_o(t)$ is uncorrelated with the measurement noise sources $\tilde{u}(t)$ and $\tilde{y}(t)$. □

The assumption of zero mean is a weak one, and rather made for convenience. A situation where $u_o(t)$ and $y_o(t)$ have nonzero means is in fact easier from an identifiability point of view, as the static gain of the system can then be determined separately.

Introduce also an assumption about the experimental condition.

**AE1**. The data comes from one (single) experiment, where **AN1** and **AI1** apply. □

When do errors-in-variables problems occur? Consider the following extension of Figure 1.

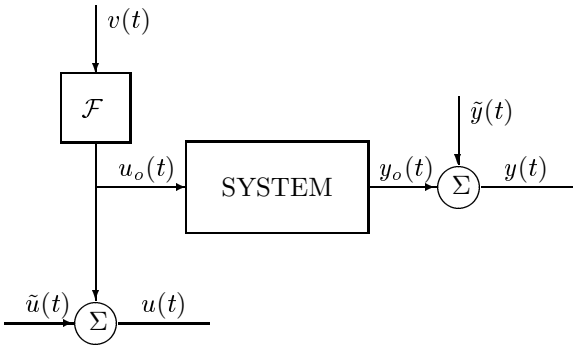


Fig. 2. The basic setup, including input generation.

One option is to assume one experiment only, and that $u_o(t)$ cannot be affected by the user. The experiment is 'arranged' by nature, or the considered system is just a part of a larger system and excited at some other point. Possibly the true input $u_o(t)$ can be modeled as a stationary stochastic process with rational spectrum, i.e. as an ARMA process. This means that in Figure 2, $\mathcal{F}$ is a finite order, unknown linear filter, and $v(t)$ for this option assumption **AE1** applies.

Another option is to assume that the signal $v(t)$ is fully accessible to the user, but that the filter $\mathcal{F}$ is an unknown and possibly nonlinear filter, so that $u_o(t)$ can neither be chosen freely, nor computed. Nevertheless, in such scenarios it is possible to make repeated experiments with the same $v(t)$, and hence with the same $u_o(t)$. In such cases the assumption **AE2b** applies, see Section 5 for further details.

A further situation is the one depicted in Figure 3.

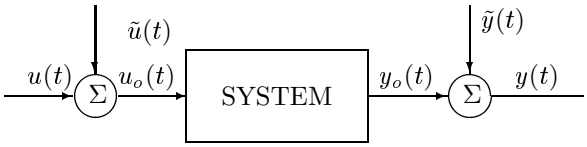


Fig. 3. A false errors-in-variables problem.

Here $u(t)$ is the designed input but $\tilde{u}(t)$ is added (due to distortions or other unavoidable reasons) before the input reach the system as $u_o(t)$. It is important to realize that *this is not* an EIV problem. Note for example that in contrast to the situation in Figure 1.

- The dynamics between $u_o(t)$ and $y_o(t)$ is the same as between $u(t)$ and $y_o(t)$ (hence the presence of $\tilde{u}(t)$ is not so problematic here as in Figure 1.
- $\tilde{u}(t)$ effects also the output measurements $y(t)$.

For the situation in Figure 3, it is appropriate to regard $\tilde{u}(t)$ as a form of process disturbance. The total effect of $\tilde{u}(t)$ and $\tilde{y}(t)$ can be modelled as a single autocorrelated disturbance on the output side.

## 3. IDENTIFIABILITY ASPECTS

Introduce the following assumptions.

**AN2**. The measurement noises are Gaussian distributed. □

**AI2**. The noise-free input is Gaussian distributed. □

Under the assumptions **AN2** and **AI2** only (first and) second order moments carry information about the distribution, and higher-order moments do not bring further information. We may alternatively say that the study for the time being is limited to infer information from second-order statistics.

It turns out that in such cases, without introducing more explicit assumptions, it is not possible

Deistler (1984), Anderson and Deistler (1987).

As an illustration of the identifiability difficulties, consider the following situation. Let the measurement noises be auto-correlated, with spectral densities $\phi_{\tilde{u}},\ \phi_{\tilde{y}}$, respectively. The system is assumed to have a transfer function $G$. Denote the observations as

$$z(t) = \begin{pmatrix} y(t) \\ u(t) \end{pmatrix}. \tag{4}$$

Their spectrum is

$$\phi_z = \begin{pmatrix} GG^* & G \\ G^* & 1 \end{pmatrix} \phi_{u_o} + \begin{pmatrix} \phi_{\tilde{y}} & 0 \\ 0 & \phi_{\tilde{u}} \end{pmatrix}. \tag{5}$$

The variables in (5) are all functions of the angular frequency $\omega$. For briefness use the symbols

$$G = G(\mathrm{e}^{\mathrm{i}\omega}), \qquad G^* = [G(\mathrm{e}^{\mathrm{i}\omega})]^* = G^\top(\mathrm{e}^{-\mathrm{i}\omega}) \tag{6}$$

in (5) and in what follows. To be more explicit, let the estimates of the aforementioned variables be denoted by $\hat{G},\ \hat{\phi}_{u_o},\ \hat{\phi}_{\tilde{u}},\ \hat{\phi}_{\tilde{y}}$. The equations determining the estimates are

$$\phi_z = \begin{pmatrix} \hat{G}\hat{G}^* & \hat{G} \\ \hat{G}^* & 1 \end{pmatrix} \hat{\phi}_{u_o} + \begin{pmatrix} \hat{\phi}_{\tilde{y}} & 0 \\ 0 & \hat{\phi}_{\tilde{u}} \end{pmatrix}. \tag{7}$$

Note that based on (5) and (7), for each frequency there are 3 equations with 4 unknowns. There is hence one degree of freedom (for each frequency) in the solution.

The relations are easily extended to the MIMO case, where (5) and (7) are replaced by

$$\begin{pmatrix} G \\ I \end{pmatrix} \phi_{u_o} \begin{pmatrix} G^* & I \end{pmatrix} + \begin{pmatrix} \phi_{\tilde{y}} & 0 \\ 0 & \phi_{\tilde{u}} \end{pmatrix} = \begin{pmatrix} \hat{G} \\ I \end{pmatrix} \hat{\phi}_{u_o} \begin{pmatrix} \hat{G}^* & I \end{pmatrix} + \begin{pmatrix} \hat{\phi}_{\tilde{y}} & 0 \\ 0 & \hat{\phi}_{\tilde{u}} \end{pmatrix}. \tag{8}$$

When the system has $n_u$ inputs and $n_y$ outputs, it turns out that equation (8) has $n_u(n_u+1)/2$ degrees-of-freedom (the number of unknowns minus the number of equations). For $n_u = 1$ this leads to one degree of freedom as already seen above.

Now introduce the following assumptions.

**AS3**. The system transfer function has no zeros mirrored in the unit circle, that is, if $G(z_1) = 0$, then $G(z_1^{-1}) \neq 0$. □

**AS4**. If the system is noncausal, then $G$ has no poles mirrored in the unit circle, that is, $p_1$ and $p_1^{-1}$ cannot both be poles of $G(z)$. □

Under the assumptions of **AS2** and **AS3**, or **AS3** and **AS4**, the solution to (5) and (7) can be written with a scalar, frequency independent degree of freedom. In fact, under the stated assumptions, the non-diagonal elements of (5), (7) give

$$G\phi_{u_o} = \hat{G}\hat{\phi}_{u_o}. \tag{9}$$

order. Consider the left hand side of (9). All zeros and poles of the factor $\phi_{u_o}(\omega)$ appears in mirrored pairs. Due to Assumptions **AS1** and **AS3** no poles or zeros of $G$ appear in mirrored pairs. Therefore, all mirrored zeros and poles must be attributed to the factor $\phi_{u_o}$. Considering only models that satisfy Assumptions **AS1** and **AS3**, one can then conclude

$$\hat{G}(\mathrm{e}^{\mathrm{i}\omega}) = G(\mathrm{e}^{\mathrm{i}\omega})\frac{1}{\alpha}, \tag{10}$$

where $\alpha$ is a constant. Proceeding further leads to

$$\hat{\phi}_{\tilde{u}}(\omega) = \phi_{\tilde{u}}(\omega) + \phi_{u_o}(\omega)\,(1-\alpha)\,, \tag{11}$$

$$\hat{\phi}_{\tilde{y}}(\omega) = \phi_{\tilde{y}}(\omega) + |G(\mathrm{e}^{\mathrm{i}\omega})|^2\phi_{u_o}(\omega)\left(1 - \frac{1}{\alpha}\right) \tag{12}$$

The parameter $\alpha$ is bounded as

$$\alpha \leq \alpha^* \triangleq \inf_\omega \frac{\phi_{u_o}(\omega) + \phi_{\tilde{u}}(\omega)}{\phi_{u_o}(\omega)}, \tag{13}$$

$$\alpha \geq \alpha_* \triangleq \sup_\omega \frac{|G(\mathrm{e}^{\mathrm{i}\omega})|^2\phi_{u_o}(\omega)}{\phi_{\tilde{y}}(\omega) + |G(\mathrm{e}^{\mathrm{i}\omega})|^2\phi_{u_o}(\omega)}. \tag{14}$$

Introduce the frequency specific signal-to-noise ratios, $S_u(\omega)$ and $S_y(\omega)$ on the input and output sides, respectively as

$$S_u(\omega) = \frac{\phi_{u_o}(\omega)}{\phi_{\tilde{u}}(\omega)}, \qquad S_y(\omega) = \frac{\phi_{y_o}(\omega)}{\phi_{\tilde{y}}(\omega)}. \tag{15}$$

Then

$$\alpha^* = \inf_\omega \frac{S_u(\omega)+1}{S_u(\omega)} \geq 1. \tag{16}$$

If $S_u(\omega)$ for some frequency is (very) large, then it follows that the upper bound $\alpha^*$ is (very) close to 1. Similarly,

$$\alpha_* = \sup_\omega \frac{S_y(\omega)}{1+S_y(\omega)} \leq 1. \tag{17}$$

If $S_y(\omega)$ is (very) large, possibly for some other frequency, then the lower bound $\alpha_*$ is (very) close to 1. Hence, if $S_u$ and $S_y$ are large for some frequencies, then the possible interval $[\alpha_*, \alpha^*]$ for $\alpha$ becomes small.

The one degree of freedom solution (10)-(12) with the bounds (13), (14) can be derived using an explicit spectral factorization of the spectra, see Agüero et al. (2005), Agüero (2005), Agüero and Goodwin (2006) for details.

A more detailed identifiability analysis is carried out in Anderson (1985), Deistler (1986), Deistler and Anderson (1989). In general, there are $n_z + 1$ degrees of freedom for characterizing the class of systems that match a given spectrum of the measured signals. Here $n_z$ is the number of non-minimum phase zeros. Identifiability of multivariable systems is treated in Green and Anderson (1986). An identifiability analysis covering also noncausal systems can be found in Anderson and

scription of all observationally equivalent systems is given in Scherrer and Deistler (1998). In the analysis leading to (10), we have $n_z = 0$ due to Assumption **AS3**, and the solution set hence has one degree of freedom as already seen.

To summarize so far, there is a fundamental lack of identifiability in the errors-in-variables problem, as long as only quite general assumptions are imposed. There are different possibilities for how to reflect about the situation.

(1) One possibility is to 'accept' the status, and not make any further assumptions. Instead of looking for a unique estimate, one has to deal with the whole set of estimates (10) where the constant $\alpha$ is bounded as in (13), (14). This choice is in the spirit set membership estimation, Milanese and Vicino (1991), where one characterizes the set of all possible models. In Agüero et al. (2005), the specific choice

$$\hat{\alpha} = (\alpha_* + \alpha^*)/2 \tag{18}$$

is presented. It minimizes the $H^\infty$ norm of the estimation error $\hat{G}(e^{i\omega}) - G(e^{i\omega})$.

(2) Another option is to modify at least one of the assumptions **AN2**, **AI2** on Gaussian distributed data. When the data are not Gaussian distributed, higher order statistics can be employed to gain additional information about the system. This option was discussed in Deistler (1986) and used in Tugnait (1992). Using higher order statistics tends to be time-consuming and may not lead to accurate estimates.

(3) A third option is to impose more detailed models for the measurement noises and for the noise-free input $u_o(t)$. Typically, $\tilde{u}(t)$, $\tilde{y}(t)$ and $u_o(t)$ are then modeled as ARMA processes of specified orders. Then the decomposition of the spectrum $\phi_z(\omega)$ in (7) into one part depending on the noise-free input and a second part due to the measurement noise may have a unique solution in some cases. More details about this option are presented in Section 4.

(4) A fourth option applies if more than one experiment can be used. This happens in some applications, say when the user can control the signal $v(t)$ in Figure 2. It is then needed that the noise-free input spectrum $\phi_{u_o}(\omega)$ differs between the different experiments, while the measurement noise properties remain the same. Another possibility is that the noise-free input $u_o(t)$ is (well) correlated between experiments, but the measurement noises $\tilde{y}(t)$, $\tilde{u}(t)$ are uncorrelated between experiments. For details, see Section 5.

to mention. It may not be trivial to define an appropriate (discrete-time) transfer function that describes the behavior between discrete-time data of $u_o(t)$ and $y_o(t)$. Most physical systems operate in continuous-time. In case the inter-sample behaviour is known, one can determine a corresponding discrete-time model description. However, here when the noise-free input signal $u_o(t)$ is not measurable, it is not obvious how to model this signal.

One approach is to formulate the whole errors-in-variables problem using continuous-time models. Some initial examinations using this idea are given by Söderström et al. (2006), Mahata and Garnier (2005), Sagara et al. (1991).

A second approach has been pointed out by Pintelon and Schoukens (2005). Let the system have a continuous-time transfer function $G_\mathrm{c}$ and assume that $v(t)$ in Figure 2 is a zero order hold signal. Under this assumption one can sample exactly the system from $v(t)$ to $y_o(t)$ as well as the system from $v(t)$ to $u_o(t)$, leading to the discrete-time models of the form

$$y_o(t) = \frac{B_y(q^{-1})}{A_F(q^{-1})A_S(q^{-1})}v(t), \tag{19}$$

$$u_o(t) = \frac{B_u(q^{-1})}{A_F(q^{-1})}v(t). \tag{20}$$

Combining (19) and (20) gives a sampled, discrete-time model of the system with

$$G_\mathrm{d}(q^{-1}) = \frac{B_y(q^{-1})}{A_S(q^{-1})B_u(q^{-1})}. \tag{21}$$

Hence, under the assumption of $v(t)$ being a zero order hold signal, there is a *unique* discrete-time transfer function $G_\mathrm{d}$ describing how $y_o(t)$ and $u_o(t)$ are related. Note though, that the transfer function $G_\mathrm{d}(q^{-1})$ has higher order than the continuous-time system $G_\mathrm{c}(s)$, contains no sampling delay, and depends also on the unknown filter $\mathcal{F}$.

Another attempt is to regard the noise-free signal $z_o(t)$ as a stationary process. Note that in continuous-time the spectrum of $z_o(t)$ is

$$\phi_{z_o}^{(\mathrm{c})} = \begin{pmatrix} G_\mathrm{c} \\ I \end{pmatrix} \phi_{\tilde{u}} \begin{pmatrix} G_\mathrm{c}^* & I \end{pmatrix}, \tag{22}$$

which apparently has rank equal to one.. According to Poisson's summation formula, the sampled data spectrum of the signal is

$$\phi_{z_o}^{(\mathrm{d})}(e^{\imath\omega h}) = \sum_{j=-\infty}^{\infty} \phi_{z_o}^{(\mathrm{c})}(\omega + j\frac{2\pi}{h}), \tag{23}$$

where $h$ is the sampling interval. Due to the folding effects caused by all terms with $j \neq 0$ in (23), $\phi_{z_o}^{(\mathrm{d})}$ will get rank 2 and there is no *no causal* discrete-time transfer function $G_\mathrm{d}(q^{-1})$ such that

$$\phi_{z_o}(\omega) = \begin{pmatrix} G_d(q^{-1}) \\ I \end{pmatrix} \phi_{u_o}(\omega) \begin{pmatrix} G_d(q^{-1}) & I \end{pmatrix}. \quad (24)$$

## 4. IDENTIFIABILITY ANALYSIS: PARAMETRIC MODELS

It is important to realize that the errors $\tilde{u}(t)$ and $\tilde{y}(t)$ can have several causes. One possible cause is pure measurement errors. It seems often realistic to assume such errors to be uncorrelated in time, and therefore relevant to model as white noise processes. However, the output error $\tilde{y}(t)$ must also accommodate effects of process disturbances and modeling errors. Both these types of contributions are typically autocorrelated in time. Therefore, it is natural to model the output error as an ARMA process.

As an ARMA process is a general model for describing a stationary, or quasistationary, process with a rational spectra, Ljung (1999), such a model may also be postulated for describing the noise-free input $u_o(t)$.

Assuming the maximum lags to be fixed and finite, we hence now have a *parametric* problem. The system is modeled as a finite order one. In the most general form it is also assumed that the noise-free input $u_o(t)$ as well as the input noise $\tilde{u}(t)$ and output noise $\tilde{y}(t)$ are ARMA processes. The total model is described in Figure 4.

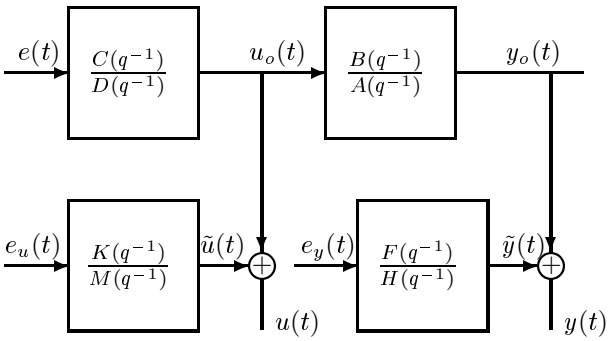


Fig. 4. Modeling a finite order system, with $u_o(t)$, $\tilde{u}(t)$ and $\tilde{y}(t)$ as ARMA processes.

More specifically, let the system transfer function be described as

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}, \quad (25)$$

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}, \\ B(q^{-1}) &= b_1 + \cdots + b_{nb} q^{-nb+1}, \end{aligned} \quad (26)$$

As described in the end of Section 3, it is reasonable to assume that $G(q^{-1})$ has no internal delay, but a direct term ($b_1$ in this case). Further, in Figure 4, the noise-free input signal is the ARMA process

$$\begin{aligned} C(q^{-1}) &= 1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}, \\ D(q^{-1}) &= 1 + d_1 q^{-1} + \cdots + d_{nd} q^{-nd}, \\ Ee(t)e(s) &= \lambda_e \delta_{t,s}. \end{aligned} \quad (27)$$

The output noise model is

$$\begin{aligned} H(q^{-1})\tilde{y}(t) &= F(q^{-1})e_y(t), \\ F(q^{-1}) &= 1 + f_1 q^{-1} + \cdots + f_{nf} q^{-nf}, \\ H(q^{-1}) &= 1 + h_1 q^{-1} + \cdots + h_{nh} q^{-nh}, \\ Ee_y(t)e_y(s) &= \lambda_y \delta_{t,s}. \end{aligned} \quad (28)$$

and the input noise model is

$$\begin{aligned} M(q^{-1})\tilde{u}(t) &= K(q^{-1})e_u(t), \\ K(q^{-1}) &= 1 + k_1 q^{-1} + \cdots + k_{nk} q^{-nk}, \\ M(q^{-1}) &= 1 + m_1 q^{-1} + \cdots + m_{nm} q^{-nm}, \\ Ee_u(t)e_u(s) &= \lambda_u \delta_{t,s}. \end{aligned} \quad (29)$$

The parameter vector to be estimated is

$$\begin{aligned} \theta = (&a_1 \dots\ a_{na}\ b_1 \dots\ b_{nb}\ c_1 \dots\ c_{nc}, \\ &d_1 \dots\ d_{nd}, f_1 \dots\ f_{nf},\ h_1 \dots\ h_{nh}, \\ &k_1 \dots\ k_{nk}, m_1 \dots\ m_{nm},\ \lambda_e\ \lambda_y\ \lambda_u)^\top. \end{aligned} \quad (30)$$

The identifiability problem will then be whether or not the parameter vector $\theta$ can be uniquely recovered from the spectrum $\phi_z(\omega)$, (5), of the measured input-output data. Assuming for simplicity that the polynomial degrees are known, and using $\hat{}$ to denote the estimated quantities, the equations determining the identifiability properties will be

$$\frac{\hat{B}\hat{B}^*}{\hat{A}\hat{A}^*}\frac{\hat{C}\hat{C}^*}{\hat{D}\hat{D}^*}\hat{\lambda}_e + \hat{\phi}_{\tilde{y}} \equiv \frac{BB^*}{AA^*}\frac{CC^*}{DD^*}\lambda_e + \phi_{\tilde{y}}, \quad (31)$$

$$\frac{\hat{B}}{\hat{A}}\frac{\hat{C}\hat{C}^*}{\hat{D}\hat{D}^*}\hat{\lambda}_e \equiv \frac{B}{A}\frac{CC^*}{DD^*}\lambda_e, \quad (32)$$

$$\frac{\hat{C}\hat{C}^*}{\hat{D}\hat{D}^*}\hat{\lambda}_e + \hat{\phi}_{\tilde{u}} \equiv \frac{CC^*}{DD^*}\lambda_e + \phi_{\tilde{u}}. \quad (33)$$

The identities in (31)-(33) are to hold for all frequencies. Trivially, the true value of the parameter vector will satisfy the identities. The system is identifiable if there is no other value of the parameter vector that satisfies the identities.

We will consider some different scenarios, describing different special cases.

**AN3a**. Both $\tilde{y}(t)$ and $\tilde{u}(t)$ are ARMA processes, as in (28) and (29). □

**AN3b**. The output noise $\tilde{y}(t)$ is an ARMA process, while the input noise $\tilde{u}(t)$ is white. This means that $nk = nm = 0$ in (29). □

**AN3c**. Both $\tilde{y}(t)$ and $\tilde{u}(t)$ are white noise sequences. This means that $nf = nh = 0$ in (28) and $nk = nm = 0$ in (29). □

Of the assumptions **AN3a**, **AN3b**, **AN3c**, obviously **AN3a** is most general and **AN3c** is most restrictive. In practice, the output noise $\tilde{y}(t)$ should model not only sensor noise but also effects of process disturbances. As this does not apply for

**AN3b** is a fairly realistic one.

The parametric identifiability problem has been dealt with by several authors. An extensive analysis with various previous results as special cases are given in Agüero (2005), Agüero and Goodwin (2006). In the frequency domain an essential assumption for identifiability is that the noisy input–output signals $u(t)$, $y(t)$ have a rational spectrum, Castaldi and Soverini (1996). In this case the identifiability of the EIV system is ensured even if the orders of the processes are not *a priori* known, provided that no zero/pole cancellation occurs between the transfer function $G(q^{-1})$ and the ARMA model of the noise–free input $u_0(t)$, and all the ARMA processes involved in the EIV representation do not share common poles.

Other results are more specific for various special cases. For example, identifiability under the noise assumption **AN3c** is investigated in Castaldi and Soverini (1996), Söderström (2003), Stoica and Nehorai (1987). The situation of the output noise being an ARMA process and the input noise being white, Assumption **AN3b**, is treated in Söderström (1980), Solo (1986). The case of $\tilde{y}(t)$ being a general stochastic process is dealt with by Hsiao (1977). The more general case **AN3a** where both $\tilde{u}(t)$ and $\tilde{y}(t)$ are ARMA processes, is coped with in Agüero (2005), Agüero and Goodwin (2006), Nowak (1985), Nowak (1993), Castaldi and Soverini (1996), Anderson and Deistler (1984). Generalization to the multivariate case is considered in Nowak (1992).

In order to give some insight, a simple case is presented first before the most general result is given.

**Example 4.1.** Assume that Assumption **AN3b** applies and further that

$$nd > nc. \tag{34}$$

holds. In this case, it is not difficult to find out that the system is (uniquely) identifiable. Note that here $\phi_{\tilde{u}} = \lambda_u$. As the polynomials $C$ and $D$, as well as the estimates $\hat{C}$ and $\hat{D}$ have all zeros inside the unit circle, it follows by considering the denominators in (33) that $\hat{D} = D$. Further, (33) then implies

$$\hat{C}\hat{C}^*\hat{\lambda}_e + DD^*\hat{\lambda}_u \equiv CC^*\lambda_e + DD^*\lambda_u.$$

The terms here consist of sums over $\mathrm{e}^{\mathrm{i}k\omega}$, with $k$ ranging from $-nd$ to $nd$. Examining the specific terms with $k = nd$ and invoking (34) it follows that $\hat{\lambda}_u = \lambda_u$. Then spectral factorization, see, *e.g.*, Söderström (2002), gives $\hat{C} = C$, $\hat{\lambda}_e = \lambda_e$. Hence the spectrum of the measured input can be uniquely decomposed in the effect of the input noise and the spectra of the noise-free input. It while (31) finally gives $\hat{\phi}_{\tilde{y}} = \phi_{\tilde{y}}$. ■

In the above example, the degree condition (34) is crucial. When this is not fulfilled the analysis becomes a bit more complicated, and one cannot use the identifiability equations just one by one as in Example 4.1.

It is possible to extend the identifiability analysis though. The most general result known is due to Agüero (2005), Agüero and Goodwin (2006), and runs as follows.

**Result 4.1.** Let the noise-free input be an ARMA process as in (27), the output noise an ARMA process as in (28) and the input noise an ARMA process as (29). Assume that

- $B(z)$ has no zero that is mirrored in the unit circle (that is, it is not allowed to be also a zero of $B(z^{-1})$),
- $B(z)$ has not a zero that is also a zero of $D(z^{-1})$,
- $A(z)$ has not a zero that is also a zero of $C(z)$.

Then the system is identifiable, if any of the following additional assumptions holds:

(1) There exists at least one zero of $D(z)$ that is not a zero of $M(z)$.
(2) There exists at least one zero of $A(z)$ that is not a zero of $H(z)$.
(3) There exists at least on zero of $D(z)$ that is not a zero of $H(z)$.
(4) The polynomial degrees satisfy

$$nm - nk > nd - nc. \tag{35}$$

(5) The polynomial degrees satisfy

$$nh - nf > (nd - nc) + (na - nb). \tag{36}$$

Note that the expressions in the inequalities (35) and (36) are expressed in terms of the pole excess of various filters. ■

## 5. IDENTIFIABILITY: USE OF MULTIPLE EXPERIMENTS

There is in general a fundamental lack of identifiability for EIV systems, unless some additional assumption or condition is added. Here two cases are considered. Both are discussed in the literature, where data are available from two or more experiments, and these experiments have some features to exploit.

Should the unperturbed input signal, $u_o(t)$, be the same in all experiments, and the experiments of equal length, then concatenating the measurements will indeed produce periodic data.

which the noise-free input has *different* character. Such a situation is treated in Markovsky et al. (2006a), although it is there described as that $u_o(t)$ changes character at some point of time in a single experiment.

Now impose the following assumption.

**AE2a**. There are more than one experiment. The spectrum of the noise-free input is different in the different experiments, while the measurement noise properties remain the same. □

It is straightforward to show that the system is identifiable under Assumption **AE2a**. ■

**Example 5.2.** Another scenario with more than one experiment where the system becomes identifiable is based on the following assumptions.

**AE2b**. There is more than one experiment. The measurement noises $\tilde{u}(t)$, $\tilde{y}(t)$ are uncorrelated between different experiments. The true noise-free input $u_o(t)$ is correlated between the experiments. □

Such a situation may occur when making repeated experiments for determining some system properties as explained in Schoukens et al. (1997), Pintelon and Schoukens (2001). The system is then a mechanical setup for determining material properties from wave propagation experiments. The noise-free input signal $u_o(t)$ is in this case the applied force to the test structure. The user has access to a command signal $v(t)$, cf Figure 2, that through a shaker $\mathcal{F}$ produces the force $u_o(t)$. However, the shaker is an electromechanical device with unknown dynamics, and its movement is also influenced by the movement in the system. The filter $\mathcal{F}$ may therefore not only be unknown but also nonlinear. By choosing the command signal $v(t)$ as periodic, it is possible to ensure that also the true input $u_o(t)$ is periodic. ■

## 6. CRAMÉR-RAO BOUNDS ON PARAMETER ESTIMATES

To assess the statistical efficiency of a parametric estimator, it is imperative to know the Cramér-Rao lower bound (CRLB). This bound gives a lower bound for the covariance matrix of the parameter estimates,

$$\operatorname{cov}(\hat{\theta} - \theta_o) \geq \text{CRLB} = J^{-1}, \tag{37}$$

$$J = E\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^{\top}\left(\frac{\partial \log L(\theta)}{\partial \theta}\right), \tag{38}$$

where $L(\theta)$ is the likelihood function. The matrix $J$ is the Fisher information matrix, Söderström and Stoica (1989). The computation of the CRLB would also delineate the set of poorly identifiable systems, as such systems should lead to a large estimator. When applied in a system identification context, the considered estimates are consistent but in general not unbiased. Then (37) applies asymptotically for large data sets, i.e. when $N \to \infty$.

Computing the CRLB for the errors-in-variables problem, one essential aspect is that the noise-free input signal must be parameterized in some way, see also Section 4. Assuming that the data are Gaussian distributed (Assumptions **AN2** and **AI2**), there are at least two ways of computing the asymptotic CRLB, see Karlsson et al. (2000) and Söderström (2006) for details. Note in particular, that the CRLB is equal to the matrix $P_{\text{ML}}$ given in (123), see Section 14.

## 7. CLASSIFICATION OF ESTIMATORS BASED ON DATA COMPRESSION

Estimation methods can be classified and organized in different ways. It may be useful to group methods together, based on an initial data compression step. After a first 'pre-processing' of the data, some reduced information is set up and used for the final computation of the parameter estimates. In case the condensed information is really a sufficient statistics, one would even be able to achieve statistical efficiency in the final step. Also when this is not the case, such an estimation scheme can still be useful, for example due to low computational complexity.

The two steps of the estimators, with an initial data compression, are illustrated in the Figure 5 below.

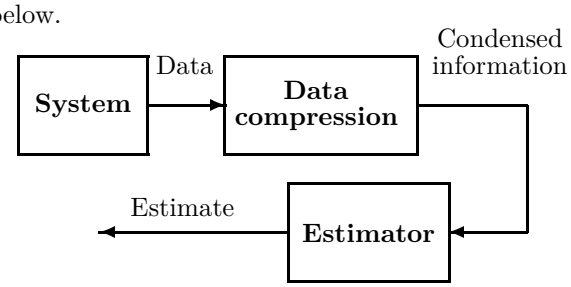


Fig. 5. Classification of the data compression prior to the estimation step.

The different groups of methods that will be discussed in the sections to follow, differ in the way the data compression is carried out. The following cases will be treated.

a) Using a covariance matrix. This case includes instrumental variables, (Section 9), bias-eliminating least squares, (Section 10), the Frisch scheme, (Section 11), total least squares, (Section 12). In all these cases the condensed information is a small set of estimated covariance elements $\{\hat{r}_u(\tau)\}$, $\{\hat{r}_{yu}(\tau)\}$ and $\{\hat{r}_y(\tau)\}$.

quency domain data, (Section 13). As compared to case a), we basically have a large set of estimated covariance elements here.

c) Using the original time-series data. This case includes the use of prediction error and maximum likelihood techniques, (Section 14). No data compression takes place.

## 8. BASIC SETUP AND NOTATIONS

In this section several general notations to be used in what follows are introduced.

Assume that the system is of finite order, and can be described by a linear difference equation:

**AS5**. The system is described as

$$A(q^{-1})y_o(t) = B(q^{-1})u_o(t), \tag{39}$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}, \\ B(q^{-1}) &= b_1 + \cdots + b_{nb} q^{-nb+1}. \end{aligned} \tag{40}$$

Introduce the parameter vector

$$\theta = \begin{pmatrix} a_1 \dots \ a_{na} \ b_1 \dots \ b_{nb} \end{pmatrix}^\top \triangleq (a^\top \ b^\top)^\top, \tag{41}$$

and the extended parameter vector

$$\overline{\theta} = \begin{pmatrix} 1 \\ \theta \end{pmatrix}. \tag{42}$$

Similarly introduce the regressor vector

$$\begin{aligned} \varphi(t) = \big( &- y(t-1) \dots \ - y(t-na) \\ &u(t) \dots \ u(t-nb+1) \big)^\top. \end{aligned} \tag{43}$$

Then the system can be written as a linear regression

$$y(t) = \varphi^\top(t)\theta + \varepsilon(t), \tag{44}$$

where the noise effects are collected in the term

$$\varepsilon(t) = A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \tag{45}$$

Introduce further the extended regressor vector

$$\begin{aligned} \overline{\varphi}(t) &= \big( - y(t) \dots \ - y(t-na) \\ &\qquad u(t) \dots \ u(t-nb+1) \big)^\top \\ &= \big( - y(t) \ \ \varphi^\top(t) \big)^\top. \end{aligned} \tag{46}$$

Further, use the conventions:

- $\theta_o$ denotes the true parameter vector, and $\hat{\theta}$ denotes its estimate.
- $\varphi_o(t)$ denotes the noise-free part of the regressor vector.
- $\tilde{\varphi}(t)$ denotes the noise-contribution to the regressor vector.

Using the above notations, and the system description (39) it follows that

$$\overline{\varphi}_o^\top(t)\overline{\theta}_o = \big( - y_o(t) \ \ \varphi_o^\top(t) \big) \begin{pmatrix} 1 \\ \theta_o \end{pmatrix} \tag{47}$$

$$= -A(q^{-1})y_o(t) + B(q^{-1})u_o(t) = 0. \tag{48}$$

$$R_\varphi = E[\varphi(t)\varphi^\top(t)], \tag{49}$$

and their estimates from finite data as

$$\hat{R}_\varphi = \frac{1}{N}\sum_{t=1}^{N} \varphi(t)\varphi^\top(t). \tag{50}$$

Similarly, let $r_{\varphi y}$ and $\hat{r}_{\varphi y}$ denote the true and estimated, respectively, covariance between $\varphi(t)$ and $y(t)$.

## 9. INSTRUMENTAL VARIABLE-BASED METHODS

Instrumental variable (IV) methods are often used as computationally simple estimators that replace least squares estimators to avoid biased estimates. The IV estimates are frequently used in EIV problems, for example in econometric applications. The IV estimator tracks its roots to Reiersøl (1941), Reiersøl (1950). For a general treatment see Söderström and Stoica (1983).

The underlying model is the linear regression (44), (45). Assume that there exist a vector $z(t)$ such that

- $z(t)$ is uncorrelated with the noise term $\varepsilon(t)$,
- $z(t)$ is well correlated with the regressor $\varphi(t)$.

Then the IV estimate can be defined as

$$\hat{\theta}_{\text{IV}} = \left( \frac{1}{N}\sum_{t=1}^{N} z(t)\varphi^\top(t) \right)^{-1} \left( \frac{1}{N}\sum_{t=1}^{N} z(t)y(t) \right). \tag{51}$$

To make the treatment more general, proceed a little differently in what follows. Introduce a vector $z(t)$ of dimension $na + nb$ or larger, satisfying

$$E[z(t)\tilde{\overline{\varphi}}^\top(t)] = 0. \tag{52}$$

Now write, using (48) and (52),

$$E[z(t)\overline{\varphi}^\top(t)]\overline{\theta}_o = E[z(t)\left(\overline{\varphi}_o^\top(t) + \tilde{\overline{\varphi}}^\top(t)\right)]\overline{\theta}_o = 0, \tag{53}$$

which we write for short as

$$R_{z\overline{\varphi}}\overline{\theta}_o = 0. \tag{54}$$

The matrix in (54) can easily be estimated from the data as

$$\hat{R}_{z\overline{\varphi}} = \frac{1}{N}\sum_{t=1}^{N} z(t)\overline{\varphi}^\top(t), \tag{55}$$

and due to the relation (54) one can derive several estimators from the approximate relation

$$\hat{R}_{z\overline{\varphi}}\hat{\overline{\theta}} \approx 0. \tag{56}$$

One possible approach for defining the estimate, leading to an instrumental variable (IV) estimator

(1989)) is to partition the matrix in (56) as

$$\left( \hat{r}_{zy} \;\; \hat{R}_{z\varphi} \right) \begin{pmatrix} 1 \\ \hat{\theta} \end{pmatrix} = \hat{r}_{zy} + \hat{R}_{z\varphi}\hat{\theta} \approx 0. \qquad (57)$$

In case $\dim z(t) = na+nb$, the relation (57) gives a linear system of equations with an exact solution, namely the basic IV estimator (51).

When the vector $z(t)$ has higher dimension, (57) gives an overdetermined system, and has in general no exact solution. A weighted least squares estimator may be taken, which gives the extended IV estimator, see Söderström and Stoica (1983),

$$\hat{\theta}_{\rm EIV} = -(\hat{R}_{z\varphi}^\top W \hat{R}_{z\varphi})^{-1}(\hat{R}_{z\varphi}^\top W \hat{r}_{zy}), \qquad (58)$$

where $W$ is a positive definite weighting matrix.

An alternative approach for exploiting (56) is to solve it in a total least squares (TLS) sense, Van Huffel and Vandewalle (1991), see also Section 12. When $\dim z(t) > na + nb$ holds, the IV and TLS estimators are not identical. It was shown experimentally in Van Huffel and Vandewalle (1989) that they have very similar behavior, and in Söderström and Mahata (2002) that they can have the same asymptotic covariance matrix in the following sense. It holds that for large $N$

$$\sqrt{N}(\hat{\theta}_{\rm IV} - \theta_o) \stackrel{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_{\rm IV}), \qquad (59)$$

$$\sqrt{N}(\hat{\theta}_{\rm TLS} - \theta_o) \stackrel{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_{\rm TLS}), \qquad (60)$$

where

$$P_{\rm IV} = \sigma S^\top(W) {\rm cov}\left(L(q^{-1})z(t)\right) S(W), \qquad (61)$$

$$P_{\rm TLS} = \sigma S^\top(I) {\rm cov}\left(L(q^{-1})z(t)\right) S(I), \qquad (62)$$

$$S(W) = W R_{z\varphi}(R_{z\varphi}^\top W R_{z\varphi})^{-1}, \qquad (63)$$

and where $\sigma$ and the filter $L(q^{-1})$ are given by the spectral factorization

$$\sigma L(z)L^*(z) = A(z)A^*(z)\phi_{\tilde{y}}(z) + B(z)B^*(z)\phi_{\tilde{u}}(z). \qquad (64)$$

The covariance matrix $P_{\rm IV}$, (61) apparently depends on the weighting matrix $W$. $P_{\rm IV}$ is minimized with the choice

$$W = \left({\rm cov}\left(L(q^{-1})z(t)\right)\right)^{-1}. \qquad (65)$$

*Remark 1.* One of the principal advantages of the IV method is its applicability under fairly general noise conditions, **AN3b**. It is inexpensive from a computational point of view. □

*Remark 2.* However, a poor accuracy of the parameter estimates is obtained: $P_{\rm IV}$ is often much larger than the the Cramér-Rao lower bound. □

*Remark 3.* The matrix $R_{z\overline{\varphi}}$ has to be full rank for the IV based estimates to exist. This is a persistence-of-excitation like condition to be satisfied by the noise-free input signal. □

rithms employing state space models are proposed in Chou and Verhaegen (1997). Their method is based on the noise assumption **AN3b** ($\tilde{u}(t)$ white, $\tilde{y}(t)$ ARMA), but allows the noise terms to be correlated. These algorithms can be applied to multivariable systems operating in open or closed loop, where one has to account for the process noise also. □

*Remark 5.* The main idea of the IV technique has been subsequently developed and generalized in several ways, *e.g.* by combining it with a weighted subspace fitting approach Stoica et al. (1995a). The noise assumption **AN3b** applies (and can be somewhat weakened: one can let $\tilde{u}(t)$ be an MA process, that is also finitely correlated with $\tilde{y}(t)$.) This combined instrumental variable weighted subspace fitting has a much improved statistical performance. □

## 10. BIAS-COMPENSATION APPROACHES

In this section, consider the case when the system is modeled as the linear regression, cf **AS5**, written as

$$y(t) = \varphi^\top(t)\theta + \varepsilon(t). \qquad (66)$$

Here, the regressor $\varphi(t)$ is defined in (43) and the parameter vector $\theta$ in (41). The term $\varepsilon(t)$ is the equation error.

The least squares (LS) estimate of $\theta$ using the model (66) is

$$\hat{\theta}_{\rm LS} = \hat{R}_\varphi^{-1}\hat{r}_{\varphi y} \to R_\varphi^{-1} r_{\varphi y}, \qquad N \to \infty. \qquad (67)$$

Assume for simplicity that the measurement noises $\tilde{u}(t)$ and $\tilde{y}(t)$ are white, **AN3c**. Then it holds that

$$R_\varphi = R_{\varphi_o} + R_{\tilde{\varphi}}, \quad r_{\varphi y} = r_{\varphi_o y_o} = R_{\varphi_o}\theta_o. \qquad (68)$$

Using (67)-(68) gives (for $N \to \infty$)

$$R_\varphi \hat{\theta}_{\rm LS} = [R_\varphi - R_{\tilde{\varphi}}]\theta, \qquad (69)$$

and $\hat{\theta}_{\rm LS}$ is biased due to the term $R_{\tilde{\varphi}}$. The principle for bias-compensated least squares (BCLS) is to adjust the least squares estimate for this effect. The adjusted estimate will be

$$\hat{\theta}_{\rm BCLS} = \left[\hat{R}_\varphi - \hat{R}_{\tilde{\varphi}}\right]^{-1} \hat{r}_{\varphi y}, \qquad (70)$$

where the noise term $\hat{R}_{\tilde{\varphi}}$ has to be estimated in some way.

Under Assumption **AN3c** the matrix $R_{\tilde{\varphi}}$ becomes

$$R_{\tilde{\varphi}} = \begin{pmatrix} \lambda_y I_{na} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{nb} \end{pmatrix}, \qquad (71)$$

and contains two different parameters to be determined. An exception is the case when the ratio $\lambda_y/\lambda_u$ is known. For such a case *one single* scaling

cation algorithms for this (considerably simpler!) case of EIV problem have been proposed by Koopmans (1937), Guidorzi (1981), Levin (1964), Aoki and Yue (1970), Eising et al. (1983), Fernando and Nicholson (1985). A further analysis of such 'one degree of freedom' bias-compensating schemes has shown that they can be interpreted as a form of weighted instrumental variable methods, Stoica et al. (1995b), Söderström et al. (1999), Garnier et al. (2000), Gilson and Van den Hof (2001).

When $\tilde{y}(t)$ and $\tilde{u}(t)$ both are white but with unknown variances, the modified normal equations (70) must be complemented with (at least) two more equations to determine also the two unknown noise variances $\lambda_y$ and $\lambda_u$. This can in fact be done in several ways. Here, we present an approach by Zheng (1998).

One such relation can be derived from the minimal value of the least squares criterion. One can show that

$$\begin{aligned} V_{\mathrm{LS}} &= \min_{\theta} E[y(t) - \varphi^T(t)\theta]^2 \\ &= \lambda_y + \theta_0^T R_{\bar{\varphi}} \hat{\theta}_{\mathrm{LS}}. \end{aligned} \tag{72}$$

Note that (72) can be seen as a linear equation in $\lambda_y$ and $\lambda_u$.

To get a second relation for $\lambda_y$ and $\lambda_u$, an extended model structure is considered, *cf.* the Frisch scheme in Section 11. For this purpose introduce extended versions of $\varphi(t)$, $\theta$ and $\theta_0$ as

$$\bar{\varphi}(t) = \begin{pmatrix} \varphi(t) \\ \underline{\varphi}(t) \end{pmatrix}, \quad \bar{\theta} = \begin{pmatrix} \theta \\ \underline{\theta} \end{pmatrix}, \quad \bar{\theta}_0 = \begin{pmatrix} \theta_0 \\ 0 \end{pmatrix}, \tag{73}$$

which are used in this section, rather than the previous conventions (42) and (46).

The model extension can, for example, mean that an additional $A$ parameter is appended:

$$\underline{\varphi}(t) = -y(t - na - 1), \qquad \underline{\theta} = a_{na+1}. \tag{74}$$

Another possibility is to append an additional $B$ parameter, leading to

$$\underline{\varphi}(t) = u(t - nb), \qquad \underline{\theta} = b_{nb+1}. \tag{75}$$

For simplicity assume $\underline{\varphi}(t)$ being a scalar (the vector case is also feasible though).

Next consider least squares estimation in the extended linear regression model

$$y(t) = \bar{\varphi}^T(t)\bar{\theta} + \varepsilon(t), \tag{76}$$

which leads to

$$R_{\bar{\varphi}} \hat{\bar{\theta}}_{\mathrm{LS}} = r_{\bar{\varphi} y}. \tag{77}$$

Similarly to (69), it holds that

$$R_{\bar{\varphi}} \hat{\bar{\theta}}_{\mathrm{LS}} = r_{\bar{\varphi}_0 y_0} + r_{\bar{\varphi}\tilde{y}} = R_{\bar{\varphi}_0}\bar{\theta}_0 = (R_{\bar{\varphi}} - R_{\tilde{\varphi}})\bar{\theta}_0. \tag{78}$$

with $\lambda_y$ and $\lambda_u$. Set

$$H = (\mathbf{0} \quad 1), \quad J = \begin{pmatrix} I_{na+nb} \\ 0 \end{pmatrix}, \quad \bar{\theta}_0 = J\theta_0. \tag{79}$$

Observe that $H\bar{\theta}_0 = 0$. Eq. (78) implies

$$H\hat{\bar{\theta}}_{\mathrm{LS}} = HR_{\bar{\varphi}}^{-1}(R_{\bar{\varphi}} - R_{\tilde{\varphi}})\bar{\theta}_0 = -HR_{\bar{\varphi}}^{-1}R_{\tilde{\varphi}}J\theta_0. \tag{80}$$

As shown in Hong et al. (2006b) one can equivalently substitute (80) by

$$\boldsymbol{E}\underline{\varphi}(t)[y(t) - \varphi^\top(t)\theta_o] = 0 \tag{81}$$

which leads to a set of *linear* equations,

$$\hat{r}_{\underline{\varphi} y} = \hat{R}_{\underline{\varphi}\varphi}\hat{\theta}. \tag{82}$$

Set

$$\begin{aligned} \lambda &= (\lambda_y \quad \lambda_u)^\top, \\ \vartheta &= (\theta^\top \quad \lambda^\top)^\top. \end{aligned} \tag{83}$$

Summing up so far we have derived the following equations for determining $\theta$ and $\lambda$: (69), (72), (80). These equations turn out to be bilinear in the unknowns $\theta$ and $\lambda$. That is, they are linear in $\theta$ and linear in $\lambda$. There are different ways to solve the equations, Zheng (1998), Söderström et al. (2005). Assume that the equations (69), (72), (80) are solved. The statistical distribution of the estimation error $\hat{\vartheta} - \vartheta_o$ due to the deviation of estimated covariance elements from the true covariance elements, is as follows.

**Result 10.1.** Under the given assumptions the parameter estimates $\hat{\vartheta}$ are asymptotically Gaussian distributed,

$$\sqrt{N}(\hat{\vartheta} - \vartheta_o) \xrightarrow{\mathrm{dist}} \mathcal{N}(0, P_{\mathrm{B}}). \tag{84}$$

A detailed expression for the matrix $P_{\mathrm{B}}$ is given in Hong et al. (2006a). ■

As the equations determining the bias-compensated estimate are determined from a small set of estimated covariance elements, and these covariance elements can easily be computed recursively in time, it is fairly natural that the estimate $\hat{\theta}_{\mathrm{BCLS}}$ can be arranged as a recursive algorithm. For algorithmic details, see Zheng and Feng (1989), Wada et al. (1990) and Feng and Zheng (1991).

There are many further possible varieties of the BCLS algorithm, though. Other approaches for deriving the required two equations to be used in addition to (70) have been proposed by Wada et al. (1990), Jia et al. (2001), Ikenoue et al. (2005) and Zheng (1999a). BCLS approaches using prefilters of the input are launched in Zheng and Feng (1989) and Zheng (1999b).

An interesting generalization of the standard BCLS method is proposed in Mahata (2006). It is shown that basing the estimates on more general

filtering, a significantly improved performance is achieved, that is close to the Cramér-Rao lower bound.

The BCLS principle can be extended to handle also the case of a general output noise as in Assumption **AN3b** ($\tilde{u}(t)$ white, $\tilde{y}(t)$ an ARMA process), Zheng (2002), and Zheng and Feng (1992). The basic idea is as follows. Introduce a noise parameter vector as

$$\eta = \begin{pmatrix} r_{\tilde{y}}(0) & r_{\tilde{y}}(1) & \dots & r_{\tilde{y}}(na) & \lambda_u \end{pmatrix}^\top. \qquad (85)$$

Consider LS estimation in two ARX models where the number of $B$ parameters is $nb$ and $na+nb+1$, respectively. From the LS estimates it is possible to derive BCLS type of equations that determine both $\theta$ and $\eta$.

Related approaches have been devised by Ekman and colleagues, Ekman (2005), Ekman et al. (2006).

In a general setting the situation can be described as follows, as far as the 'design' of BCLS methods goes.

- As a first step, the user has to provide the structure of a noise model. More specifically, the type of input and output noise is to be provided. Two natural options here are Assumption **AN3c** (both $\tilde{u}(t)$ and $\tilde{y}(t)$ white noise) or Assumption **AN3b** ($\tilde{u}(t)$ is white noise, and $\tilde{y}(t)$ is an ARMA process).
- A second part is to set up the underlying equations. In the case above, the set of equations are (69), (72), (80). While the normal equations for a standard least squares estimation typically are kept, the other equations may be substituted. It is also possible to use more equations than unknowns. Due to the nature of the problem, the set of equations will always be nonlinear, although they have a considerable amount of structural properties.
- The third item to consider is the numerical algorithm to use for solving the system of equations.

To formulate the situation in a general way, the system of equations may be written as

$$f(\hat{r}_N, \hat{\theta}) \approx 0. \qquad (86)$$

In (86) $\hat{r}_N$ is a vector of a number of covariance elements based on $N$ data points. In the standard BCLS case (86) is a concise notation for (69), (72), (80). In case the number of equations and the number of unknowns are equal, strict equality holds in (86). The algorithm to solve (86) is written as

$$\hat{\theta}_k = g(\hat{\theta}_{k-1}, \hat{r}_N), \qquad (87)$$

where $k$ is an iteration number, and $g$ is a function that is tied to $f$. The scheme (87) is to be iterated be updated recursively in time, it is possible to substitute (87) by a recursive algorithm, cf. Ljung and Söderström (1983).

*Consistency* can be examined by investigating if the solution to (86) converges to the true parameter vector, as the number of data points grows to infinity. In mathematical terms, one examines if the possible implication

$$f(\hat{r}_\infty, \theta) \approx 0 \Rightarrow \theta = \theta_o? \qquad (88)$$

holds true.

*Convergence* of the iterations is mostly considered in the asymptotic case, that is for $N = \infty$. This means that one considers the discrete-time deterministic nonlinear system

$$\hat{\theta}_k = g(\hat{\theta}_{k-1}, \hat{r}_\infty), \qquad (89)$$

and explores whether or not the desired solution $\theta = \theta_o$ is a stable solution. This stability analysis can be done both locally, and (which is more difficult) globally. For the analysis of 'local' stability, it is enough to examine the linearized system. More particularly, the matrix

$$\left.\frac{\partial g(\theta, \hat{r}_\infty)}{\partial \theta}\right|_{\theta=\theta_o}$$

must have all eigenvalues inside the unit circle to guarantee (local) stability.

Finally, one can also examine the *statistical accuracy*, as expressed by the asymptotic distribution of the parameter estimates. Assume that the iterative scheme (87) is chosen so that the iterations do converge. The properties of the asymptotic ($k \to \infty$) estimate will then not depend on the used algorithm for solving the equations (86). When $N$ is large but finite, the estimate will deviate somewhat from the true parameter vector. Under weak assumptions the deviation is asymptotically Gaussian distributed in the sense

$$\sqrt{N}(\hat{\theta} - \theta_o) \longrightarrow \mathcal{N}(0, P), \qquad N \to \infty. \qquad (90)$$

Here, the normalized asymptotic covariance matrix $P$ will depend on the system, the noise-free input spectrum, the signal-to-noise ratios and the estimation method. The dependence on these quantities can be rather involved.

## 11. THE FRISCH SCHEME

The Frisch scheme has its roots in a classical algebraic estimation problem, see Frisch (1934). The term 'Frisch scheme' has also been used in situations where the noise has uncorrelated components, but is correlated in time. Extensions of the Frisch scheme to identification of dynamic models appeared in Beghelli et al. (1990), Scherrer and Deistler (1998). Later refinements and

Guidorzi (1996), Söderström et al. (2002), Diversi et al. (2003). The Frisch scheme may be interpreted as a special form of bias-compensated least squares.

The Frisch estimation method is based on the assumption of white input and output measurement noise, **AN3c**. First note that

$$\overline{\varphi}_o^\top(t)\overline{\theta}_o = -A_o(q^{-1})y_o(t) + B_o(q^{-1})u_o(t) = 0. \tag{91}$$

Further it holds that

$$R_\varphi = R_{\varphi_o} + R_{\tilde{\varphi}}, \qquad R_{\overline{\varphi}} = R_{\overline{\varphi}_o} + R_{\tilde{\overline{\varphi}}}. \tag{92}$$

It follows from (91) that

$$R_{\overline{\varphi}_o}\overline{\theta}_o = E\left[\overline{\varphi}_o\overline{\varphi}_o^\top\overline{\theta}_o\right] = \mathbf{0}. \tag{93}$$

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 **AI1** 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

$$R_{\tilde{\overline{\varphi}}} = \begin{pmatrix} \lambda_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{nb} \end{pmatrix}. \tag{94}$$

The relation (93) is the basis for the Frisch method. The idea is to have appropriate estimates of the noise variances and then determine the parameter vector $\theta$ from

$$\left(\hat{R}_{\overline{\varphi}} - \hat{R}_{\tilde{\overline{\varphi}}}\right)\hat{\overline{\theta}} = \mathbf{0}. \tag{95}$$

Assume for the time being that some estimate $\hat{\lambda}_u$ of the input noise variance is available. Then the output noise variance $\lambda_y$ is determined so that the matrix appearing in (95) is singular. More specifically, see Söderström (2005):

$$\hat{\lambda}_y = \lambda_{\min}\left(\hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y\varphi_u}\left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb}\right)^{-1}\hat{R}_{\varphi_u\overline{\varphi}_y}\right) \tag{96}$$

where $\lambda_{\min}(R)$ denotes the smallest eigenvalue of $R$. The estimate of the parameter vector $\theta$ is determined by solving

$$\left(\hat{R}_\varphi - \begin{pmatrix} \hat{\lambda}_y I_{na} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix}\right)\hat{\theta} = \hat{r}_{\varphi y}, \tag{97}$$

which is indeed the BCLS equations (69), (95). By (96), $\hat{\lambda}_y$ will be a function of $\hat{\lambda}_u$.

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

- In Beghelli et al. (1990), 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 two functions correspond to curves in the $(\hat{\lambda}_u, \hat{\lambda}_y)$ plane. The curves will ideally intersect in one unique $\hat{R}_{\overline{\varphi}}$ would be replaced by its true value $R_{\overline{\varphi}}$ a situation as displayed in Figure 6 would be obtained. Curve A corresponds to the true

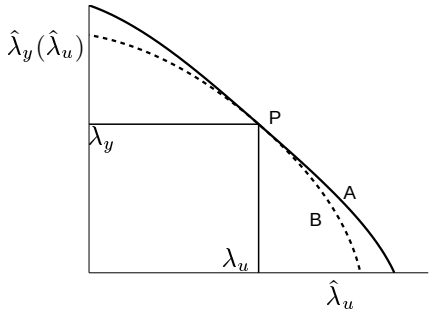


Fig. 6. Illustration of the principle for Frisch estimation.

model order, while curve B applies for the increased model order. The coordinates of the common point P give precisely the true noise variances $\lambda_u,\ \lambda_y$.

For a finite data set the situation is less ideal, and there is not a distinct point P where the curves A and B share a common point. We refer to Beghelli et al. (1990), Soverini and Söderström (2000) for more detailed aspects on how this type of the Frisch scheme can be implemented.

- Another alternative is to compute residuals, and compare their statistical properties with what can be predicted from the model. This alternative was proposed in Diversi et al. (2003). It is presented here in a slightly more general form. Define the residuals

$$\varepsilon(t,\hat{\theta}) = \hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t), \tag{98}$$

and compute sample covariance elements

$$\hat{r}_\varepsilon(k) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\hat{\theta})\varepsilon(t+k,\hat{\theta}). \tag{99}$$

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

$$\varepsilon_o(t) = \hat{A}(q^{-1})\hat{\tilde{y}}(t) - \hat{B}(q^{-1})\hat{\tilde{u}}(t), \tag{100}$$

where $\hat{\tilde{y}}(t)$ and $\hat{\tilde{y}}(t)$ are independent white noise sequences with

$$E\left[\hat{\tilde{y}}^2(t)\right] = \hat{\lambda}_y, \qquad E\left[\hat{\tilde{u}}^2(t)\right] = \hat{\lambda}_u. \tag{101}$$

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

$$V_N(\hat{\lambda}_u) = \delta^\top W\delta, \tag{102}$$

where $W$ is a user chosen, positive definite weighting matrix, and the vector $\delta$ is

$$\delta = \begin{pmatrix} \hat{r}_\varepsilon(1) - \hat{r}_{\varepsilon_o}(1) \\ \vdots \\ \hat{r}_\varepsilon(m) - \hat{r}_{\varepsilon_o}(m) \end{pmatrix}. \tag{103}$$

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

$$\hat{\lambda}_u = \arg\min_{\lambda_u} V_N(\lambda_u). \tag{104}$$

In summary the Frisch scheme algorithm consists of the equations (96), (97) and (104). In its implementation, there is an optimization over one variable, $\hat{\lambda}_u$, in (104). Possibly minimization of the criterion (102) is substituted by some other form of searching for the point P in Figure 6. In the evaluation of the loss function $V_N(\hat{\lambda}_u)$, also (96) and (97) are used to get $\hat{\lambda}_y$ and $\hat{\theta}$, respectively.

The following result describes the asymptotic distribution of the parameter estimates for the Frisch scheme.

**Result 11.1.** Under the given assumptions the parameter estimates $\hat{\vartheta}$ are asymptotically Gaussian distributed,

$$\sqrt{N}(\hat{\vartheta} - \vartheta_o) \xrightarrow{\text{dist}} \mathcal{N}(0, P_{\text{F}}). \tag{105}$$

where $\vartheta$ is given by (83). A detailed expression for the matrix $P_{\text{F}}$ is given in Söderström (2005). ■

In its basic form the Frisch scheme is somewhat restrictive in that white output noise is assumed. Extensions of the Frisch scheme to situations where $\tilde{y}(t)$ is autocorrelated (that is, using Assumption **AN3b** rather than **AN3c**) are proposed in Söderström (2006).

## 12. TOTAL LEAST SQUARES

The total least squares (TLS) method is in principle a method for treating overdetermined linear systems of equations where both the coefficient matrix and the right hand side are subject to errors. Consider the overdetermined system of equations

$$Ax \approx b, \tag{106}$$

where it is assumed that the $m \times n$ matrix $A$ has full rank $n$. The least squares solution to (106) is

$$x_{\text{LS}} = (A^\top A)^{-1} A^\top b \tag{107}$$

and can also be formulated as the solution to the following optimization problem

$$\min \| \Delta b \|^2 \text{ subject to } Ax_{\text{LS}} = b + \Delta b. \tag{108}$$

The TLS problem is different, and takes into account also uncertainties in the coefficient matrix $A$. The TLS problem tracks it roots to orthogonal regression, Adcock (1877), Adcock (1878) and have been given a detailed treatment in Golub and Van Loan (1980), Van Huffel and Vandewalle (1991). There are several connections between TLS and the EIV problems, as manifested in the workshop proceedings Van Huffel (1997), Van can be formulated as, compare (108),

$$\min \| [\Delta A \ \Delta b] \|_F^2 \text{ s. t. } (A + \Delta A)x_{\text{TLS}} = b + \Delta b. \tag{109}$$

Note that in (109) it is important to use the Frobenius norm, while for (108) the Frobenius norm and the common 2-norm coincide as $b$ is a vector.

Computationally, the TLS solution is obtained from a singular value decomposition of $[\Delta A \ \Delta b]$. Indeed the right singular vector associated with the smallest singular value of this matrix can after scaling be written as $\begin{pmatrix} 1 & -x_{\text{TLS}}^\top \end{pmatrix}^\top$.

If the errors in the various $A$ and $b$ elements are independent and identically distributed, the TLS solution to the problem (106) coincides with the maximum likelihood estimate, Gleser (1981). Further, both consistency and the asymptotic distribution of the estimates are examined under this condition. Unfortunately, the analysis described above is not of much help, when TLS is applied for identification of dynamic systems. In fact, in many cases there are couplings between various elements in $A$ and $b$. The matrix may be constructed to be Toeplitz of Hankel for example. For such cases one can apply a structured total least squares (STLS) or constrained total least squares solution, that takes couplings into account. Consider a simple example, and let the system dynamics be given by (39), with $\varphi(t)$ as in (43). Repeating (39) for various values of $t$ gives

$$\begin{pmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{pmatrix} \theta = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix}. \tag{110}$$

As the matrix in (110) is block Hankel (with equal elements along the block diagonals), the structured TLS solution is more relevant than the basic TLS solution in general.

The statistical properties of the solution to a structured TLS problem is considered in several papers, for example Kukush et al. (2005). A quite general TLS situation is examined. In the analysis, however, it is assumed that the covariance structure may vary from row to row, but that the total covariance structure is known up to a scalar factor. This corresponds in the BCLS framework in Section 11 to the simpler, but not so realistic, case of a one degree of freedom problem.

## 13. FREQUENCY DOMAIN METHODS

The methods described in the previous sections are all based on time domain techniques. It is, however, also possible to work in the frequency domain. Then, typically, as a first step the spectrum of the observations is determined.

satisfies in the SISO case, see (2), (4), (5),

$$\phi_z = \begin{pmatrix} G \\ 1 \end{pmatrix} \begin{pmatrix} G^* & 1 \end{pmatrix} \phi_{u_o} + \begin{pmatrix} \lambda_y & 0 \\ 0 & \lambda_u \end{pmatrix}. \quad (111)$$

If the spectral density matrix $\phi_z$ is known, and an appropriate diagonal matrix is subtracted, then one would get a rank 1 matrix, corresponding to the first term of (111), for all frequencies $\omega$. In case the decomposition as in (111) can be carried out, the first term would easily lead to estimates of the transfer function $G(e^{i\omega})$ and the true input spectrum $\phi_{u_o}(\omega)$.

The idea here is similar to that of the Frisch scheme, with the important difference that the full spectral information in the data is used instead of the covariance matrix (55).

As the first term in (111) is singular, it must hold for each frequency $\omega_k, k = 1, 2, \ldots$, that

$$[\phi_y(\omega_k) - \lambda_y][\phi_u(\omega_k) - \lambda_u] - |\phi_{yu}(\omega_k)|^2 = 0. \quad (112)$$

This relation is exploited in Söderström et al. (2003) as a linear regression with $\lambda_y$, $\lambda_u$, $\lambda_y\lambda_u$ as three unknowns, to derive an estimate of the noise variances. Once estimates of $\lambda_y$ and $\lambda_u$ are available, it is straightforward to estimate the transfer function, for example as

$$\hat{G}(e^{i\omega_k}) = \phi_{yu}(\omega_k)/[\phi_u(\omega_k) - \hat{\lambda}_u]. \quad (113)$$

In this approach the spectrum $\phi_z$ is estimated as a first step. In order to guarantee that the estimated spectrum $\hat{\phi}_z(\omega)$ indeed is positive definite, it is important to use a more sophisticated spectrum estimator than a straightforward Discrete Fourier Transform.

The above idea is developed in Beghelli et al. (1997) using an FFT-based spectral estimator. The quality of the estimate is moderate. A more promising approach, based on a black-box modeling of the spectral density $\phi_z(\omega)$, is described in Söderström et al. (2003). A nonparametric approach is considered, where the spectrum $\phi_z(\omega)$ (111) is modeled using a two-dimensional ARMA model and a two step procedure for fitting its parameters. Two parametric estimators with further improved performances are also derived. In one of them a parametric model is fitted to the nonparametric one. The other parametric method is exploiting information in the two-dimensional ARMA model in a more direct way. In qualitative terms, the performance of the two parametric estimators is at least as good as typical IV-based estimators.

In Castaldi et al. (2002) another Frisch domain approach is used. The problem is formulated as a bilinear matrix inequality. An $H^\infty$ estimate is proposed in Agüero et al. (2005), see also (18). The degree of freedom is used such that the $H^\infty$ quency domain IV method for multivariable systems is developed in Hsiao and Robinson (1978). In general, an iterative procedure is needed to tope with nonlinear dependencies, but it is shown how consistent and efficient parameter estimates can be found using a three step procedure.

A general treatment of frequency domain estimators for the EIV problem is given in Pintelon et al. (1994). As in many of these methods it is assumed that the noise-free input $u_o(t)$ is periodic. More details are provided in Section 15.

## 14. PREDICTION ERROR AND MAXIMUM LIKELIHOOD METHODS

In this approach the errors–in–variable model of Figure 4 is regarded as a multivariable system with both $y(t)$ and $u(t)$ as outputs, Söderström (1981). Of crucial importance for this approach is the assumption **AI3**, *i.e.*, the noise–free input $u_o(t)$ is characterized by a rational spectrum. It can thus be represented by its innovations form, described as an ARMA process of the type (27). In this way, the whole errors–in–variables model can be considered as a system with a two–dimensional output vector $z(t) = (y(t)\, u(t))^\top$ and three mutually uncorrelated white noise sources $e(t)$, $\tilde{y}(t)$ and $\tilde{u}(t)$:

$$\begin{pmatrix} y(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} \dfrac{B(q^{-1})C(q^{-1})}{A(q^{-1})D(q^{-1})} & 1 & 0 \\ \dfrac{C(q^{-1})}{D(q^{-1})} & 0 & 1 \end{pmatrix} \begin{pmatrix} e(t) \\ \tilde{y}(t) \\ \tilde{u}(t) \end{pmatrix}. \quad (114)$$

Thus the model $C(q^{-1})/D(q^{-1})$ of the undisturbed input is a part of the errors–in–variables representation and its coefficients must be estimated together with the parameters of $A(q^{-1})$ and $B(q^{-1})$.

The model (114) is transformed to a general state-space model, which in turn is transformed into innovations form, obtained from the Kalman filter, see for example, Söderström (2002). The prediction errors $\varepsilon(t, \theta) = z(t) - \hat{z}(t|t-1; \theta)$ depend on the data and the model matrices. Let $Q$ denote the covariance matrix of the prediction errors. The parameter vector $\theta$ is estimated from a data sequence $\{z(t)\}_{t=1}^N$ by minimizing a loss function:

$$\hat{\theta}_N = \arg\min_\theta V_N(\theta). \quad (115)$$

Assume that $V_N(\theta)$ is a (sufficiently) smooth function of $\theta$, and that $V_N(\theta)$ converges (uniformly on compact subsets) as $N \to \infty$. Then $\hat{\theta}_N$ is consistent.

There is no unique way of defining a prediction error method (PEM) criterion that penalizes the

to take

$$V_N(\theta) = \det\left(\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta)\varepsilon^\top(t,\theta)\right). \quad (116)$$

One has then also to make use of the minimal value of the criterion, $V_N(\hat{\theta}_N)$, Söderström (2006). The estimation error is asymptotically Gaussian distributed, Ljung (1999), Söderström and Stoica (1989). In fact,

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \xrightarrow{\text{dist}} \mathbf{N}(0, P_{\text{PEM}}), \quad (117)$$

with

$$P_{\text{PEM}} = \left[E\{\psi^\top(t)Q^{-1}(\theta_o)\psi(t)\}\right]^{-1}. \quad (118)$$

In (118), $\psi(t)$ denotes

$$\psi(t,\theta) = -\left(\frac{\partial\varepsilon(t,\theta)}{\partial\theta}\right)^\top. \quad (119)$$

The maximum likelihood (ML) estimate (for Gaussian distributed data, **AN2**, **AI2**) minimizes

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\ell(\varepsilon(t,\theta),\theta,t), \quad (120)$$

with

$$\ell(\varepsilon,\theta,t) = \frac{1}{2}\log\det Q(\theta) + \frac{1}{2}\varepsilon^\top(t,\theta)Q^{-1}(\theta)\varepsilon(t,\theta). \quad (121)$$

In principle it is straightforward to extend the PEM and ML estimates to MIMO EIV systems.

For the EIV problem considered the ML estimate will be more accurate than the PEM estimate, in the sense that it gives a smaller covariance matrix for the parameter errors. The reason is that in (121), the innovations covariance matrix $Q(\theta)$ and the prediction errors $\{\varepsilon(t,\theta)\}$ are parameterized with joint parameters (in contrast to 'standard identification' problems).

Also the ML estimate is asymptotically Gaussian distributed, Söderström (2006),

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \xrightarrow{\text{dist}} \mathbf{N}(0, P_{\text{ML}}), \quad (122)$$

with

$$P_{\text{ML}} = \left[E\{\psi^\top(t)Q^{-1}(\theta_o)\psi(t)\} + \overline{P}\right]^{-1}, \quad (123)$$

where

$$\overline{P}_{ij} = \frac{1}{2}\text{tr}[Q^{-1}Q_iQ^{-1}Q_j], \qquad Q_i = \frac{\partial Q}{\partial\theta_i}. \quad (124)$$

From (118), (123), it is obvious that the presence of the term $\overline{P}$ in (123) implies that the ML estimate is more accurate than the PEM estimate. Furthermore, the Cramér-Rao bound turns out to be equal to the matrix $P_{\text{ML}}$, (123), for Gaussian data.

in the frequency domain, Pintelon and Schoukens (2006), even if there are differences in how transient effects are handled. The inherent spectral factorization may be easier to carry out in the frequency domain.

The main disadvantage of PEM and ML for the EIV problem is that the numerical optimization procedure is, in general, quite complex since at every iteration a Riccati equation must be solved in order to find the innovations $\varepsilon(t)$ used in (116) or (120). The procedure may fail to give good results if only poor initial parameter estimates are available.

## 15. METHODS DESIGNED FOR PERIODIC DATA

The methods described in this section are partly tied to periodic data. However, we also include the case when there is more than one experiment. Should the unperturbed input signal, $u_o(t)$, be the same in all experiments, and the experiments of equal length, then concatenating the measurements will indeed produce periodic data.

First recall from Section 5, see Examples 5.1 and 5.2, that under mild conditions the system will be identifiable if the noise-free input signal $u_o(t)$ is periodic.

A straightforward way to handle multiple experiments using time-domain data is the following, which is an instrumental variable estimator. Consider the linear regression model

$$y(t) = \varphi^\top(t)\theta + \varepsilon(t), \quad (125)$$

where $\varepsilon(t)$ denotes the equation error. Assume that more than one data set is available, so that

$$y^{(i)}(t) = \varphi^{(i)T}(t)\theta + \varepsilon^{(i)}(t), \quad i = 1, 2, \ldots \quad (126)$$

The true parameter vector fits perfectly the models when undisturbed data are used:

$$y_o^{(i)}(t) = \varphi_o^{(i)T}(t)\theta_o. \quad (127)$$

Assume now that **AE2b** applies. Hence, the noise is independent in the different data sets, and the unperturbed regressor vector $\varphi_o^{(i)}(t)$ is (well) correlated in the different data sets. Using two data sets, one gets

$$\begin{aligned} &E[\varphi^{(1)}(t)\varphi^{(2)T}(t)]\theta_o - E[\varphi^{(1)}(t)y^{(2)}(t)] \\ &= E\varphi_o^{(1)}(t)[\varphi_o^{(2)T}(t)\theta_o - y_o^{(2)}(t)] = 0. \end{aligned} \quad (128)$$

Assume that the matrix $E[\varphi_o^{(1)}(t)\varphi_o^{(2)T}(t)]$ is non-singular. This is partly a condition on the inputs being persistently exciting. It is also a condition on sufficient correlation between the data sets. The consequence is that from *two* data sets, it

mator as

$$\left[\frac{1}{N}\sum_{t=1}^{N}\varphi^{(1)}(t)\varphi^{(2)T}(t)\right]\hat{\theta}=\frac{1}{N}\sum_{t=1}^{N}\varphi^{(1)}(t)y^{(2)}(t), \tag{129}$$

which is indeed an instrumental variable estimator, Söderström and Stoica (1989).

It is though also possible to apply the estimator (129) in other situations, and the 'experiments' can be allowed to be overlapping, as long as the basic assumptions are satisfied. More extended versions of this idea including the use of optimal weightings are described in Söderström and Hong (2005). The principle to use IV estimation using cross-correlation between the different experiments has earlier been reported by van den Bos (1992) and Pintelon et al. (1994).

In some situations it happens that one can have repeated experiments where the noise-free regressor vector remains the same, that is $\varphi_o^{(i)}(t)$ does not change from one experiment to another. This case is treated in the frequency domain in Schoukens et al. (1997), Guillaume et al. (1995), and in the time domain with TLS techniques in Forsell et al. (1999). With periodic excitation it is possible to separate the driving signals and the disturbances. A nonparametric noise model is applied, and the noise can be allowed to have an arbitrary correlation structure.

For the (frequency-domain) sample maximum likelihood method (SML) in Schoukens et al. (1997), see also Pintelon and Schoukens (2001), it is assumed that at least four full periods of the data (or four independent experiments) are available. The method has a quite good statistical performance, and is close to statistically efficient. The achieved covariance matrix is $(M-2)/(M-3)$ times the Cramér-Rao lower bound, where $M \geq 6$ is the number of periods in the experiment. The data are first pre-processed to estimate the noise covariance functions (or equivalently, the frequency dependent variances of the DFT of the input and output noise). Note that the SML method can be extended to handle also cases where the input noise $\tilde{u}(t)$ is correlated with the output noise $\tilde{y}(t)$. Let $U(\omega_k)$ and $Y(\omega_k)$, with $\omega_k = 2\pi k/N$, denote the discrete Fourier transforms of the input and output measurements, respectively, and assume that the transfer function is $G(z) = B(z)/A(z)$. The ML criterion in the frequency domain used in Schoukens et al. (1997) can be written as

$$V_{\mathrm{ML}}(\theta, Z) = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|B(\mathrm{e}^{\mathrm{i}\omega_k},\theta)U(\omega_k)-A(\mathrm{e}^{\mathrm{i}\omega_k},\theta)Y(\omega_k)\right|^2}{D(\omega_k)} \tag{130}$$

$$D(\omega)=\sigma_u^2(\omega)|B(\mathrm{e}^{\mathrm{i}\omega},\theta)|^2+\sigma_y^2(\omega)|A(\mathrm{e}^{\mathrm{i}\omega},\theta)|^2 -2\mathrm{Re}\,[\sigma_{yu}^2(\omega)A(\mathrm{e}^{\mathrm{i}\omega},\theta)B(\mathrm{e}^{-\mathrm{i}\omega},\theta)] \tag{131}$$

and $\sigma_u^2(\omega_k)$ is the variance of the DFT $\tilde{U}(\omega_k)$ of the input noise $\tilde{u}(t)$ at frequency $\omega_k$, etc.

The method described in Markovsky et al. (2006a) is also based on two experiments. Formally, though, it is presented as that there is a clustering in time of the data, so that the statistical properties of the noise-free input, such as its spectrum $\phi_{u_o}(\omega)$, change at some point of time, cf. **AE2a**. The same type of idea is discussed in Wald (1940) under the name of grouping.

## 16. CONCLUDING REMARKS

It has been demonstrated in the paper that EIV systems in general suffer from an identifiability problem. There are several ways to add some extra condition to achieve identifiability. Various parameter estimation methods are designed to work under such an additional condition.

An attempt to compare some of the principal approaches, in terms of underlying additional assumption, computational complexity, and statistical accuracy (asymptotic variances of the parameter estimates), is presented in Table 1. One extreme case is basic IV, which is very simple computationally, and gives only crude estimates. The other extreme is the ML estimator which is based on a complex optimization problem, and can give very accurate estimates. It is fairly natural that PEM and ML have high computational complexity, as they involve a nonlinear optimization problem, where each function evaluation requires a considerable amount of calculations. There is also a 'middle group' of methods that are computationally much simpler than PEM and ML, but on the other hand give much more accurate estimates than the basic IV. It is also worth stressing that for basic IV only the system parameters are estimated. For both ML and PEM model parameters for describing the true input are estimated as well. For the 'middle group' of estimators parameters describing the measurement noises are estimated.

It is important to read the comparison with some caution. The statements are certainly qualitative, not quantitative. The precise results will depend on the actual coding, and also on the system, the frequency properties of the true input, the signal-to-noise ratios on the input and output sides, etc.

As illustration two examples are provided.

**Example 16.1.** We compare the computational load in terms of Matlab flops for basic IV, the

identification problem. *): In the basic version $\tilde{y}(t)$ is assumed to be white. Extensions to an arbitrarily correlated $\tilde{y}(t)$ are available. **): SML method defined by (130). ***): Note that in this case $\tilde{u}(t)$ and $\tilde{y}(t)$ can be allowed to be cross-correlated.

| Method | Noise condition | Experimental condition | Computational complexity | Statistical accuracy |
|---|---|---|---|---|
| Basic IV | $\tilde{u}(t)$ MA **AN3a** | - | very low | low |
| IV + WSF | $\tilde{u}(t)$ MA **AN3a** | - | medium | medium-high |
| BCLS | $\tilde{u}(t)$ white, $\tilde{y}(t)$ *) **AN3bc** | - | low | medium-high |
| Frisch | $\tilde{u}(t)$ white, $\tilde{y}(t)$ *) **AN3bc** | - | low | medium-high |
| TLS | $\tilde{u}(t)$, $\tilde{y}(t)$ white **AN3c** and $\lambda_y/\lambda_u$ known, or | repeated exp's **AE2a** | medium | medium-high |
| Frequency domain **) | $\tilde{u}(t)$, $\tilde{y}(t)$ ARMA **AN3a** | repeated exp's **AE2b** ***) | medium-high | very high |
| PEM | $\tilde{u}(t)$, $\tilde{y}(t)$ ARMA **AN3a** | - | high | high |
| ML | $\tilde{u}(t)$, $\tilde{y}(t)$ ARMA **AN3a** | - | high | very high |

Frisch scheme, and PEM. These methods were applied to a first and a second order system, disturbed with white measurement noises, and where the noise-free input is a first order ARMA process. The effective number of estimated parameters differ between the methods. The mean values of the number of Matlab flops from 10 realizations, each with $N = 200$ data points, are presented in Figure 7. The figure supports the statements above on how the load differs between the methods. ■

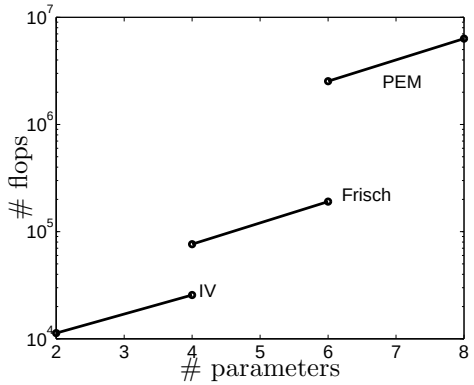


Fig. 7. Illustration of the computational loads.

**Example 16.2.** The theoretical asymptotic variances of the system parameter estimates were examined for some methods. In other papers, these theoretical expressions are shown to quite well predict what happens in simulations. Consider a second order system with the true parameter values $a_1 = -1.5,\ a_2 = 0.7,\ b_1 = 2,\ b_2 = 1$. The measurement noises were white with $\lambda_u = 1, \lambda_y = 10$. The noise-free input was an ARMA(1,2) process with the parameters $c_1 = 0.7,\ c_2 = 0.2,\ d_1 = -0.5$. The noise variance $\lambda_e$ was varied and the asymptotic normalized variance of the parameter estimates were computed. The results are plotted in Figure 8. The plots demonstrate clearly that the basic IV method produces quite crude estimates. No method produces as accurate estimates as the ML method. Sometimes (say for $b_1$) there is a quite significant difference in performance between PEM and ML. There is no uniformly valid ordering between Frisch, BCLS and PEM: no method is always better than the other. ■

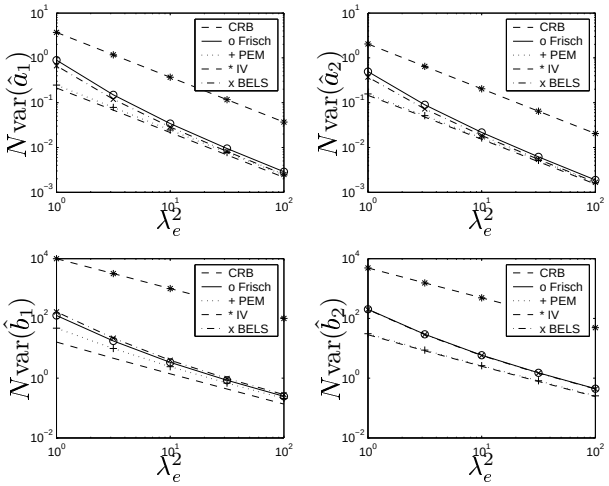


Fig. 8. Normalized variances of $\hat{a}_1,\ \hat{a}_2,\ \hat{b}_1$ and $\hat{b}_2$, for the Frisch scheme, PEM, BCLS, basic IV and the Cramer-Rao lower bound.

It is also worth stressing that the experimental condition and a priori information can have profound implications on the estimation results. For example, if there are repeated experiments, the frequency domain methods described in Section 13 exploit this knowledge in an efficient way and can produce very accurate results.

Another case is when the noise variance ratio $\lambda_y/\lambda_u$ is known. When this happens, the fundamental loss of identifiability vanishes. Further, the TLS based estimators become perfectly feasible due to the possibility for rescaling variables, so that scaled input and output measurements have equal levels of noise. It is also possible to derive a maximum likelihood estimator in this case, where all values of the noise-free input $u_o(t), t = 1, \ldots, N$, are treated as additional nuisance parameters, Diversi et al. (2006). The resulting esti-

consistent.

**Acknowledgments**. This work has been partly funded by the Swedish Research Council under contract 621-2002-4671.

It is a pleasure to express my gratitude to many colleagues with whom I over the years have discussed, learned from, and published work together with on errors-in-variables problems. These colleagues include Juan Carlos Agüero, Theodore Anderson, Mats Cedervall, Han-Fu Chen, Manfred Deistler, Roberto Diversi, Mats Ekman, Graham Goodwin, Roberto Guidorzi, Christiaan Heij, Mei Hong, Erlendur Karlsson, Alexander Kukush, Erik K. Larsson, Kaushik Mahata, Ivan Markovsky, Magnus Mossberg, Rik Pintelon, Wolfgang Scherrer, Johan Schoukens, Virginija Šimonytė, Joachim Sorelius, Umberto Soverini, Petre Stoica, Sabine Van Huffel, Kiyoshi Wada and Wei Xing Zheng. I also thank the reviewers for helpful and valuable suggestions.

## References

R. J. Adcock. Note on the methods of least squares. *The Analyst*, 4(6):183–184, 1877.

R. J. Adcock. A problem in least squares. *The Analyst*, 5(2):53–54, 1878.

J. C. Agüero. *System identification methodologies incorporating constraints*. PhD thesis, The University of Newcastle, Callaghan, NSW, Australia, August 2005.

J. C. Agüero and G. C. Goodwin. Identifiability of errors in variables dynamic systems. In *IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

J. C. Agüero, G. C. Goodwin, and M. E. Salgado. On the optimal estimation of errors in variables models for robust control. In *16th IFAC World Congress*, Prague, Czech Republic, July 4-8 2005.

B. D. O. Anderson. Identification of scalar errors-in-variables models with dynamics. *Automatica*, 21:709–716, 1985.

B. D. O. Anderson and M. Deistler. Identifiability of dynamic errors-in-variables models. *J. Time Series Analysis*, 5:1–13, 1984.

B. D. O. Anderson and M. Deistler. Dynamic errors-in-variables systems in three variables. *Automatica*, 23:611–616, 1987.

T. W. Anderson. Estimating linear statistical relationships. *The Annals of Mathematical Statistics*, 12(1):1–45, 1984.

M. Aoki and P. C. Yue. On *a priori* error estimate of some identification methods. *IEEE Transactions on Automatic Control*, AC-15:541–548, October 1970.

S. Beghelli and U. Soverini. Identification of linear relations from noisy data: Geometrical 339–346, 1992.

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

S. Beghelli, P. Castaldi, and U. Soverini. A frequential approach for errors-in-variables models. In *Proc. European Control Conference, ECC '97*, Brussels, Belgium, July 1–4 1997.

P. Castaldi and U. Soverini. Identification of dynamic errors-in-variables models. *Automatica*, 32(4):631–636, April 1996.

P. Castaldi, M.Montanari, and A. Tilli. Induction motor model identification via frequency-domain Frisch scheme. In *IFAC World Congress*, Barcelona, Spain, 2002.

C. L. Cheng and J. W. Van Ness. *Kendall's library of statistics: Statistical Regression With Measurement Error*, volume 6. Edward Arnold, 1999.

C. T. Chou and M. Verhaegen. Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, 33(10):1857–1869, October 1997.

M. Deistler. Linear dynamic errors-in-varibles models. *Journal of Applied Probability*, 23A: 23–39, 1986.

M. Deistler and B. D. O. Anderson. Linear dynamic errors-in-varibles models. Some structure theory. *Journal of Econometrics*, 41:39–63, 1989.

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.

R. Diversi, R. Guidorzi, and U. Soverini. Maximum likelihood identification of noisy input-output models. 2006. Submitted for publication.

R. Eising, N. Linssen, and H. Rietbergen. System identification from noisy measurements of inputs and outputs. *Systems and Control Letters*, 2:348–353, 1983.

M. Ekman. Identification of linear systems with errors in variables using separable nonlinear least squares. In *16th IFAC World Congress*, Prague, Czech Republic, July 04-08 2005.

M. Ekman, M. Hong, and T. Söderström. A separable nonlinear least squares approach for identification of linear systems with errors in variables. In *14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

C. B. Feng and W. X. Zheng. Robust identification of stochastic linear systems with correlated noise. *IEE Proceedings, Part D*, 138(5):484–492, September 1991.

K. V. Fernando and H. Nicholson. Identification of linear systems with input and output noise: the

*Part D*, 132(1):30–36, January 1985.

U. Forsell, F. Gustafsson, and T. McKelvey. Time-domain identification of dynamic errors-in-variables systems using periodic excitation signals. In *Proc. IFAC 14th World Congress*, Beijing, P.R.China, 1999.

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

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

H. Garnier, M. Gilson, and W. X. Zheng. A bias-eliminated least-squares method for continuous-time model identification of closed-loop systems. *International Journal of Control*, 71:38–48, 2000.

M. Gilson and P. Van den Hof. On the relation between a bias-eliminated least-squares (BELS) and an IV estimator in closed-loop identification. *Automatica*, 37:1593–1600, 2001.

L. J. Gleser. Estimation in a multivariable "errors in variables" regression model: large sample results. *The Annals of Statistics*, 9(1):24–44, 1981.

G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numerical Analysis*, 17:883–893, 1980.

M. Green and B. D. O. Anderson. Identification of multivariable errors in variable models with dynamics. *IEEE Transactions on Automatic Control*, 31(5):467–471, May 1986.

R. Guidorzi. Invariants and canonical forms for systems structural and parametric identification. *Automatica*, 17(1):117–133, January 1981.

R. Guidorzi. Certain models from uncertain data: The algebraic case. *Systems and Control Letters*, 17(6):415–424, December 1991.

R. Guidorzi. Identification of the maximal number of linear relations from noisy data. *Systems and Control Letters*, 24(3):159–165, February 1995.

R. Guidorzi. Identification of multivariable processes in the Frisch scheme context. In *MTNS 96*, St Louis, USA, 1996.

R. Guidorzi, R. Diversi, and U. Soverini. Optimal errors-in-variables filtering. *Automatica*, 39: 281–289, 2003.

P. Guillaume, R. Pintelon, and J. Schoukens. Robust parametric transfer function estimation using complex logarithmic frequency response data. *IEEE Transactions on Automatic Control*, 40:1180–1190, 1995.

C. Heij, W. Scherrer, and M. Deistler. System identification by dynamic factor models. *SIAM Journal on Control and Optimization*, 35(6): 1924–1951, November 1997.

M. Hong, T. Söderström, and W. X. Zheng. Accuracy analysis of bias-eliminating least squares estimates for errors-in-variables identification. *tion*, Newcastle, Australia, March 29-31 2006a.

M. Hong, T. Söderström, and W. X. Zheng. A simplified form of the bias-eliminating least squares method for errors-in-variables identification. Technical Report 2006-000, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2006b. Available as http://www.it.uu.se/research/publications/reports/2006-000.

C. Hsiao. Identification for a linear dynamic simultaneous error-shock model. *International Economic Review*, 18(1):181–194, February 1977.

C. Hsiao and P. M. Robinson. Efficient estimation of dynamic error-shock model. *International Economic Review*, 19(2):467–479, June 1978.

M. Ikenoue, S. Kanae, Z.-J. Yand, and K. Wada. Identification of noisy input-output system using bias-compensated least-squares method. In *IFAC 16th World Congress*, Prague, Czech Republic, July 2005.

L. J. Jia, M. Ikenoue, C. Z. Jin, and K. Wada. On bias compensated least squares method for noisy input-output system identification. In *Proc. of 40th IEEE Conference on Decision and Control*, pages 3332–3337, Orlando, Florida, USA, 2001.

R. E. Kalman. Identification from real data. In M. Hazewinkel and A. H. G. Rinnoy Kan, editors, *Current Developments in the Interface: Economics, Econometrics, Mathematics*, Dordrecht, The Netherlands, 1982a. D. Reidel.

R. E. Kalman. System identification from noisy data. In A. R. Bednarek and L. Cesari, editors, *Dynamical Systems II*, New York, N.Y., 1982b. Academic Press.

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.

T. J. Koopmans. *Linear Regression Analysis of Economic Time Series*. N. V. Haarlem, The Netherlands, 1937.

A. Kukush, I. Markovsky, and S. Van Huffel. Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *Journal of Statistical Planning and Inference*, 133:315–358, 2005.

M. J. Levin. Estimation of a system pulse transfer function in the presence of noise. *IEEE Transactions on Automatic Control*, AC–9:229–235, July 1964.

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

L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA, USA, 1983.

proach for errors-in-variables model identification. 2006. Submitted for publication.

K. Mahata and H. Garnier. Direct identification of continuous-time errors-in-variables models. In *14th IFAC World Congress*, Prague, Czech Republic, July 3-8 2005.

I. Markovsky, J. C. Willems, and B. De Moor. Continuous-time errors-in-variables filtering. In *41st Conference on Decision and Control (CDC 2002)*, pages 2576–2581, Las Vegas, Nevada, December 2002.

I. Markovsky, J. C. Willems, S. Van Huffel, B. De Moor, and R. Pintelon. Application of structured total least squares for system identification and model reduction. *IEEE Transactions on Automatic Control*, 50(10):1490–1500, 2005. Special issue on system identification: linear vs. nonlinear.

I. Markovsky, A. Kukush, and S. Van Huffel. On errors-in-variables with unknown noise variance ratio. In *IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006a.

I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor. *Exact and Approximate Modeling of Linear Systems. A Behavioral Approach*. SIAM, Philadelphia, PA, 2006b.

M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27 (6):997–1009, November 1991.

E. Nowak. Global identification of the dynamic shock-error model. *Journal of Econometrics*, 27:211–219, 1985.

E. Nowak. Identifiability in multivariate dynamic linear errors-in-variables models. *Journal of the American Statistical Association*, 87(419):713–723, September 1992.

E. Nowak. The identification of multivariate linear dynamic errors-in-variables models. *Journal of Econometrics*, 59:213–227, 1993.

R. Pintelon and J. Schoukens. *System Identification. A Frequency Domain Approach*. IEEE Press, New York, NY, USA, 2001.

R. Pintelon and J. Schoukens. Personal communication. 2005.

R. Pintelon and J. Schoukens. Frequency domain maximum likelihood estimation of linear dynamic errors-in-variables models. 2006. Submitted for publication.

R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van hamme. Parametric identification of transfer functions in the frequency domain – a survey. *IEEE Transactions on Automatic Control*, 39 (11):2245–2260, November 1994.

O. Reiersøl. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9:1–24, 1941.

between variables which are subject to error. *Econometrica*, 18(4):375–389, 1950.

S. Sagara, Z.J. Yang, and K. Wada. Identification of continuous systems from noisy sampled input-output data. In *IFAC Symposium on System Identification and Parameter Estimation*, pages 1622–1627, Budapest, Hungary, 1991.

W. Scherrer and M. Deistler. A structure theory for linear dynamic errors-in-variables models. *SIAM Journal on Control and Optimization*, 36 (6):2148–2175, November 1998.

J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume. Frequency domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, 33:1073–1086, 1997.

T Söderström. *Discrete-time Stochastic Systems - Estimation and Control, 2nd edition*. Springer-Verlag, London, UK, 2002.

T. Söderström. Why are errors-in-variables problems often tricky? In *European Control Conference, ECC 2003*, Cambridge, UK, September 1-4 2003.

T. Söderström. Accuracy analysis of the Frisch estimates for identifying errors-in-variables systems. In *44th IEEE CDC/ECC 2005*, Seville, Spain, December 12-15 2005.

T. Söderström. On computing the Cramer-Rao bound and covariance matrices for PEM estimates in linear state space models. In *14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

T. Söderström. Spectral decomposition with application to identification. In F. Archetti and M. Gugiani, editors, *Numerical Techniques for Stochastic Systems*, NorthHolland, Amsterdam, 1980.

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

T. Söderström. Extending the Frisch scheme for errors-in-variables identification to correlated noise. Technical Report 2006-032, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2006. Available as http://www.it.uu.se/research/publications/reports/2006-032.

T. Söderström and M. Hong. Identification of dynamic errors-in-variables systems with periodic data. In *Proc. 16th IFAC World Congress*, Prague, Czech Republic, July 04-08 2005.

T. Söderström and K. Mahata. On instrumental variable and total least squares approaches for identification of error-in-variables models. *International Journal of Control*, 75:709–718, April 2002.

T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Springer-Verlag, Berlin, 1983.

*tion*. Prentice Hall International, Hemel Hempstead, UK, 1989.

T. Söderström, W. X. Zheng, and P. Stoica. Comments on 'On a least-squares-based algorithm for identification of stochastic linear systems'. *IEEE Transactions on Signal Processing*, 47: 1395–1396, May 1999.

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

T. Söderström, K. Mahata, and U. Soverini. Identification of dynamic errors-in-variables model: Approaches based on two-dimensional ARMA modelling of the data. *Automatica*, 39(5):929–935, May 2003.

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.

T. Söderström, E. K. Larsson, K. Mahata, and M. Mossberg. Using continuous-time modeling for errors-in-variables identification. In *14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

V. Solo. Identifiability of time series models with errors in variables. *Journal of Applied Probability*, 23A:63–71, 1986.

U. Soverini and T. Söderström. Identification methods of dynamic systems in presence of input noise. In *Proc. SYSID 2000, IFAC 12th Symposium on System Identification*, Santa Barbara, California, June 21–23 2000.

P. Stoica and A. Nehorai. On the uniqueness of prediction error models for systems with noisy input-output data. *Automatica*, 23(4):541–543, 1987.

P. Stoica, M. Cedervall, and A. Eriksson. Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. *IEEE Transactions on Signal Processing*, 43:2386–2397, 1995a.

P. Stoica, T. Söderström, and V. Šimonytė. Study of a bias-free least squares parameter estimator. *IEE Proc. Control Theory Appl.*, 142:1–6, 1995b.

J. K. Tugnait. Stochastic system identification with noisy input using cumulant statistics. *IEEE Transactions on Automatic Control*, AC-37:476–485, 1992.

A. van den Bos. Identification of continuous-time systems using multiharmonic test signals. In N. K. Sinha and G. P. Rao, editors, *Identification of Continuous-Time Systems*. Kluwer Academic, Dordrecht, The Netherlands, 1992.

S. Van Huffel, editor. *Recent Advances in Total Least Squares Techniques and Errors-in-*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.

S. Van Huffel and J. Vandewalle. Comparison of total least squares and instrumental variable methods for parameter estimation of transfer function models. *International Journal of Control*, 50:1039–1056, 1989.

S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computation Aspects and Analysis*. SIAM, Philadelphia, USA, 1991.

J. H. Van Schuppen. Stochastic realization problems. In N. Nijmeijer and J. M. Schumacher, editors, *Three Decades of Mathematical System Theory*, Berlin, 1989. Springer-Verlag.

K. Wada, M. Eguchi, and S. Sagara. Estimation of pulse transfer function via bias-compensated least-squares method in the presence of input and output noise. *System Sciences*, 16(3):57–70, 1990.

A. Wald. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, Vol. 11, Series B(3): 284–300, 1940.

J. Willems. From time series to linear systems. *Automatica*, 37:561–580, 1986.

W. X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7):1142–1147, July 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.

W. X. Zheng. On least-squares identification of stochastic linear systems with noisy input-output data. *International Journal of Adaptive Control and Signal Processing*, 13:131–143, 1999a.

W. X. Zheng. Parametric identification of noisy closed-loop linear systems. In *Proc. 14th IFAC World Congress*, Beijing, P.R.China, July 1999b.

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.

W. X. Zheng and C. B. Feng. Identification of a class of dynamic errors–in–variables models. *International Journal of Adaptive Control and Signal Processing*, 6:431–440, 1992.