WHY ARE ERRORS-IN-VARIABLES PROBLEMS OFTEN TRICKY?

Torsten Söderström

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

Keywords: system identification, errors-in-variables

Abstract

There are several identification methods designed for the errors-in-variables problem. In this paper we focus on discussing the underlying assumptions for several of these methods. Some assumptions are shown to have far reaching consequences. For example, if the noise-free input happens to be periodic, simple estimators that give consistent parameter estimates of the system parameters can easily be designed. If the variances of the input and output noises are unknown, some structural assumption must be added for the system dynamics to be identifiable. On the other hand, should the ratio between output noise variance and input noise variance be known, it is possible not only to estimate the system parameters consistently, but also to combine this with a reasonable estimate of the unperturbed input.

1 Introduction

Many different solutions have been presented for system identification of linear dynamic systems from noise–corrupted output measurements see, for example, [6], [11]. On the other hand, estimation of the parameters for linear dynamic systems when 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" models. They play an important role when the purpose is determination of the inner laws that describe the process, rather than the prediction of its future behavior.

The class of scientific disciplines which make use of such representations is very broad, as proved by the several applications collected in [14], [15], such as time series modelling, array signal processing for direction–of–arrival estimation, blind channel equalization, multivariate calibration in analytical chemistry, image processing, astronomical data reduction, etc. In case of static systems, errors–in–variables representations are closely related to other well–known topics such as *latent variables* models and *factor* models [3].

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

The focus here is on the role of the assumptions. In 'classical identification' problems where the input signal is measured without errors, some assumptions on the noise properties and the input properties have mostly little impact on the consistency on the parameter estimates, but more on their accuracy. In the errors-in-variables context the assumptions can on the other hand often have far reaching consequences. The assessment of what assumptions that are realistic in practice may vary.

The paper is organized as follows. The next section presents the basic setup, while the principal identifiability problem is dealt with in Section 3. There it is shown that it is not possible to get identifiability from one single experiment without imposing some (structural) assumptions, that cannot be verified from the data. In Section 4 we describe bias-compensated versions of the least-squares method, while other estimators based on repeated and correlated experiments are treated in Section 5. It can be appealing to let an identification scheme include estimation of the unknown unperturbed input. This idea is discussed in Section 6, where we show that it is feasible only if we have some appropriate *a priori* information about the noise variances. We show in Section 7 how the system can be identified when a *parametric* model of the unperturbed input is applied, and summarizing conclusions are given in Section 8. Several of the sections are ended by summarizing different results, where the role of the key assumptions are highlighted.

2 Basic setup and assumptions

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

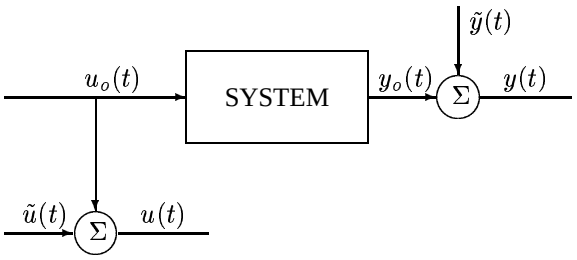

Figure 1: The basic setup for a dynamic error-in-variables problem.

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

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

where $A(q^{-1})$ and $B(q^{-1})$ are polynomials

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \cdots + a_n\, q^{-n} \\ B(q^{-1}) &= b_1\, q^{-1} + \cdots + b_n\, q^{-n} \end{aligned} \tag{2}$$

and q^{-1} is the backward shift operator, *i.e.* $q^{-1}\,x(t) = x(t-1)$. It is not restrictive to assume that the polynomials $A(q^{-1})$, $B(q^{-1})$ have equal degree n, which represents the order of the system.

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

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

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

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

In 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. We will then discuss what assumptions that are necessary for different results, and what the consequences will be if the assumptions are relaxed.

2.1 Assumptions on the system

A1a. The dynamic system (1) is asymptotically stable, *i.e.* $A(z)$ has all zeros outside the unit circle.

A1b. All the system modes are observable and controllable, *i.e.* $A(z)$ and $B(z)$ have no common factors.

A1c. The order n of the system is *a priori* known.

Out of the above assumptions, **A1c** is the most restrictive in practice. It implies that we do not treat the undermodelling case.

2.2 Assumptions on the measurement noise

A2a. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually uncorrelated, and uncorrelated with the noise–free signals $u_o(t)$ and $y_o(t)$.

A2b. The sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero–mean white noises with variances λ_u and λ_y, respectively.

The assumption **A2b** may be regarded as a bit restrictive. If $\tilde{y}(t)$ should model not only measurement noise, but also account for process noise and modelling errors, then it is likely correlated and *not* white.

We can have different situations concerning to what degree the noise variances are known.

A2c The noise variances λ_u and λ_y are unknown.

A2d The noise variances λ_u and λ_y are both known.

A2e The ratio λ_y/λ_u is known.

Obviously, the condition **A2c** is more general (and more realistic) than **A2e**. It will turn out in the coming analysis, that when **A2e** is implied, much stronger results will hold than under **A2c**. Of course, assumption **A2d** is still more restrictive than **A2e**.

It is sometimes possible to allow for weaker assumptions. The measurement noises may be somewhat correlated. Some estimation methods can be extended to handle that $\tilde{u}(t)$ is a moving average process, $\tilde{y}(t)$ is an ARMA process, or that $\tilde{u}(t)$ and $\tilde{y}(t)$ are correlated.

2.3 Assumptions on the true input

A3a. The true input $u_o(t)$ is a zero–mean stationary ergodic random signal.

We may also have

A3b. The spectral density of $u_o(t)$ is a rational function. Note that this implies that $u_o(t)$ can be modelled as an ARMA process.

For some identification methods, see Section 7, but certainly not all, the assumptions **A3a** and **A3b** have to be imposed.

2.4 Assumptions on the experiment

A4a. One experimental data set is available.

A4b. More than one experimental data set are available. The unperturbed input data $u_o(t)$ from the different data sets are *correlated*.

When **A4b** holds, a simple estimation algorithm can be applied, see Section 5. However, it is not obvious when the assumption can be realistically applied.

To discuss these assumptions, consider the following extension of Figure 1.

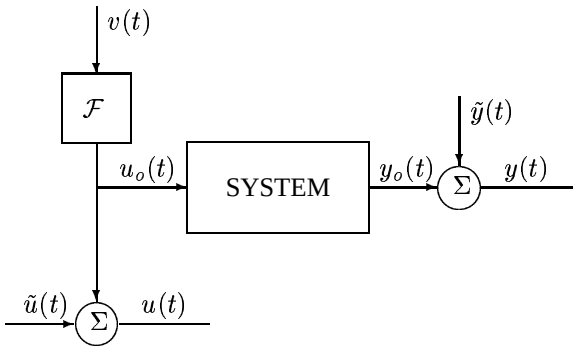

Figure 2: The basic setup, including input generation.

One option is to assume one experiment only, and that $u_o(t)$ cannot be effected 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. The true input $u_o(t)$ is modelled as a stationary stochastic process with rational spectrum. This means that in Figure 2, $\mathcal{F}$ is a finite order, unknown linear filter, and $v(t)$ is an (unmeasurable) white noise sequence. Hence for this option assumptions **A3b** and **A4a** apply.

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 (t), and hence with the same $u_o(t)$. In such cases the assumption **A4b** applies.

2.5 Notations

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

$$\theta = (a_1 \dots a_n \; b_1 \dots b_n)^T. \tag{5}$$

Similarly we introduce the regressor vector

$$\varphi(t) = (-y(t-1) \cdots - y(t-n) \; u(t-1) \dots u(t-n))^T. \tag{6}$$

Further, we will use the conventions:

- θ_o denotes the true parameter vector, and $\hat{\theta}$ denotes its estimate.
- $\varphi_o(t)$ denotes the noisefree part of the regressor vector. Hence

$$\varphi_o(t) = (-y_o(t-1) \dots u_o(t-1) \dots u_o(t-n))^T. \tag{7}$$

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

$$\tilde{\varphi}(t) = (-\tilde{y}(t-1) \cdots - \tilde{y}(t-n) \; \tilde{u}(t-1) \dots \tilde{u}(t-n))^T. \tag{8}$$

3 Identifiability and verifiability

Consider the case of Gaussian distributed data. This means that only (first and) second order moments carry information about the distribution, and that higher-order moments do not bring further information. We may alternatively say that we limit our study for the time being to infer information from second-order statistics.

It turns out that in such cases, without introducing more explicit assumptions, such as **A2e**, it is not possible to uniquely identify the system. In fact, when only second–order statistics are exploited the identification of errors–in–variables models cannot, in general, admit a single solution, [1]. As illustration, consider the following example.

Example 3.1. Let the measurement noises be auto-correlated, with spectral densities $\phi_{\tilde{u}}, \; \phi_{\tilde{y}}$, respectively. (Note that this is weaker than **A2b**.) Set

$$z(t) = \begin{pmatrix} u(t) \\ y(t) \end{pmatrix}. \tag{9}$$

Its spectrum is

$$\phi_z = \begin{pmatrix} 1 & G^* \\ G & GG^* \end{pmatrix} \phi_{u_o} + \begin{pmatrix} \phi_{\tilde{u}} & 0 \\ 0 & \phi_{\tilde{y}} \end{pmatrix}. \tag{10}$$

With ϕ_z given and $G, \; \phi_{u_o}, \; \phi_{\tilde{u}}, \; \phi_{\tilde{y}}$ as unknowns, there is no unique solution to (10).

To be more explicit, let the estimates of the aforementioned variables be denoted by $\hat{G}, \; \hat{\phi}_{u_o}, \; \hat{\phi}_{\tilde{u}}, \hat{\phi}_{\tilde{y}}$. Choose a spectrum $\hat{\phi}_{u_o}$ (so far arbitrarily). One can with straightforward calculations derive

$$\hat{G} = G \frac{\phi_{u_o}}{\hat{\phi}_{u_o}} \tag{11}$$

$$\hat{\phi}_{\tilde{u}} = \phi_{\tilde{u}} + \phi_{u_o} - \hat{\phi}_{u_o} \tag{12}$$

$$\hat{\phi}_{\tilde{y}} = \phi_{\tilde{y}} + GG^* \left(\phi_{u_o} - \frac{\phi_{u_o}^2}{\hat{\phi}_{u_o}} \right). \tag{13}$$

Requiring that $\hat{\phi}_{\tilde{u}}$ and $\hat{\phi}_{\tilde{y}}$ are positive definite function will give the possible values of $\hat{\phi}_{u_o}$. Some straightforward calculations give

$$-\frac{\phi_{u_o}\phi_{\tilde{y}}}{\phi_{u_o}GG^* + \phi_{\tilde{y}}} \leq \hat{\phi}_{u_o} - \phi_{u_o} \leq \phi_{\tilde{u}}. \tag{14}$$

This shows clearly that except for the trivial and true solution $\hat{\phi}_{u_o} = \phi_{u_o}$, many more solutions are possible. ■

Results:

1. Without assumptions on the structures (the parameterizations), there will *not* be a unique solution to the identification problem.
2. It is *not* possible to verify from the data, whether or not these structural assumptions are verified. ■

4 Bias-compensation of the least-squares estimate

When estimating the parameter vector θ from the linear regression model

$$y(t) = \varphi(t)\theta + \varepsilon(t), \tag{15}$$

the normal equations will be (in the asymptotic case of an infinite number of data points)

$$[E\varphi(t)\varphi^T(t)]\hat{\theta}_{LS} = E\varphi(t)y(t). \tag{16}$$

As the data contain one contribution due to the unperturbed input $u_o(t)$ and another due to the noise, the normal equations can be written as

$$\begin{array}{l} [E\varphi_o(t)\varphi_o^T(t) + E\tilde{\varphi}(t)\tilde{\varphi}^T(t)]\hat{\theta}_{LS} \\ = E\varphi_o(t)y_o(t) + E\tilde{\varphi}(t)\tilde{y}(t) \end{array}. \tag{17}$$

Under assumption **A2b** $E\tilde{\varphi}(t)\tilde{y}(t) = 0$ holds. As the true parameter vector θ_o satisfies

$$[E\varphi_o(t)\varphi_o^T(t)]\theta_o = E\varphi_o(t)y_o(t) \tag{18}$$

it follows that there will be a bias,

$$\hat{\theta}_{LS} \neq \theta_o. \tag{19}$$

The bias is caused by the matrix

$$E\tilde{\varphi}(t)\tilde{\varphi}^T(t) = R_{\tilde{\varphi}\tilde{\varphi}} = \begin{pmatrix} \lambda_y I & 0 \\ 0 & \lambda_u I \end{pmatrix}. \quad (20)$$

There are many identification schemes available that aim to compensate for the bias effect that is induced by the noise sources. As examples, see [4], [5], [12], [16], [17]. Assuming the noise variances are known or estimated, a bias-compensated least-squares scheme can be constructed as

$$\hat{\theta}_{BCLS} = \left[\hat{R}_{\varphi\varphi} - \begin{pmatrix} \hat{\lambda}_y I & 0 \\ 0 & \hat{\lambda}_u I \end{pmatrix}\right]^{-1} \hat{r}_{\varphi y}\ , \quad (21)$$

where

$$\hat{R}_{\varphi\varphi} = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t),\ \ \hat{r}_{\varphi y} = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t). \quad (22)$$

Note that it is crucial in most of these schemes that the measurement noise sequences are white (assumption **A2b**). A wide class of bias-compensated least squares methods can in fact be interpreted as specific variants of the instrumental variable method, see [13].

5 Use of repeated and correlated experiments

Consider the linear regression model

$$y(t) = \varphi^T(t)\theta + \varepsilon(t) \quad (23)$$

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

$$y^{(i)}(t) = \varphi^{(i)T}(t)\theta + \varepsilon^{(i)}(t), \quad i = 1, 2, \ldots \quad (24)$$

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

$$y_o^{(i)}(t) = \varphi_o^{(i)T}(t)\theta_o. \quad (25)$$

Assume now that

1. the noise is independent in the different data sets
2. the unperturbed regressor vector $\varphi_o^{(i)}(t)$ is (well) correlated in the different data sets.

Using two data sets, we then get

$$\begin{aligned} & E[\varphi^{(1)}(t)\varphi^{(2)T}(t)]\theta_o - E[\varphi^{(1)}(t)y^{(2}(t))] \\ = \ & E\varphi_o^{(1)}(t)[\varphi_o^{(2)T}(t)\theta_o - y_o^{(2)}(t)] = 0. \end{aligned} \quad (26)$$

Assume that the matrix $E[\varphi_o^{(1)}(t)\varphi_o^{(2)T}(t)]$ is nonsingular. 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 is possible to derive a consistent parameter estimator as

$$\left[\frac{1}{N}\sum_{t=1}^{N}\varphi^{(1)}(t)\varphi^{(2)T}(t)\right]\hat{\theta} = \left[\frac{1}{N}\sum_{t=1}^{N}\varphi^{(1)}(t)y^{(2)}(t)\right]. \quad (27)$$

This is indeed an instrumental variable estimator, [11].

A special case is mentioned in [2], [7], where some specific measurement situations allow for repeated experiments where the noise-free regressor vector remain the same, that is $\varphi_o^{(i)}(t)$ does not vary from one experiment to another. It is though also possible to apply the estimator (27) in other situations, and the 'experiments' can be allowed to be overlapping, as long as the basic assumptions are satisfied. In particular two subsets of a single experiments may be used, as illustrated in Figure 3.

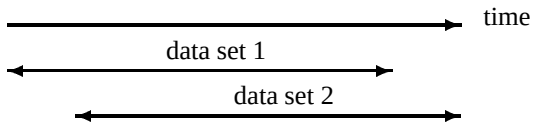

Figure 3: Split of a total data set into two overlapping ones.

Results:

1. If the total data set can be split into parts, where the measurement noise $\tilde{y}(t),\ \tilde{\varphi}(t)$ in the different parts are uncorrelated, but the noisefree regressors $\varphi_o(t)$ are correlated, it is possible to derive simple and consistent estimates of the parameter vector θ.
2. The estimates are IV variants, with delayed input-output data as instruments.

6 Nonparametric modeling of the input

A tempting approach is to treat the signal values of the undisturbed input, $u_o(t),\ t = 1, \ldots, N$ as auxiliary unknowns to be estimated. It seems though not to be a pertinent approach in general as shown in the following. When assumption **A2e** holds it is though fully feasible.

Example 6.1. Consider a case as simple as possible, namely a purely static system:

$$\begin{array}{rcl} y(t) & = & bu_o(t) + \tilde{y}(t) \\ u(t) & = & u_o(t) + \tilde{u}(t) \end{array}. \quad (28)$$

We regard

$$\theta = (b\ \lambda_y\ \lambda_u)^T, \qquad U_o = (u_o(1)\ldots u_o(N))^T \quad (29)$$

as the unknowns. Assume the data to be Gaussian distributed. The negative loglikehood function will then be (neglecting a constant term and skipping a factor of $N/2$)

$$\begin{aligned} L(\theta, U_o) = & \frac{1}{N\lambda_y}\sum_{t=1}^{N}[y(t) - bu_o(t)]^2 \\ & + \frac{1}{N\lambda_u}\sum_{t=1}^{N}[u(t) - u_o(t)]^2 + \log\lambda_y + \log\lambda_u\ . \end{aligned} \quad (30)$$

The loss function $L(\theta, U_o)$ is easily minimized with respect to U_o giving

$$u_o(t) = \frac{b\lambda_u y(t) + \lambda_y u(t)}{b^2\lambda_u + \lambda_y},\ t = 1, \ldots, N\ . \quad (31)$$

The concentrated loss function becomes, after inserting (31) into (30) and some straightforward calculations

$$\begin{aligned} V(\theta) &= \min_{U_o} L(\theta, U_o) \\ &= \frac{1}{(b^2\lambda_u + \lambda_y)} \left(\hat{R}_y + b^2 \hat{R}_u - 2b\hat{R}_{yu} \right) \\ &\quad + \log \lambda_y + \log \lambda_u, \end{aligned} \tag{32}$$

where the covariance elements are

$$\hat{R}_y = \frac{1}{N} \sum_{t=1}^{N} y^2(t) \tag{33}$$

etc. One can consider some different cases, based on what assumptions on noise variances apply. The results below are proved in [9].

Results:

1. The estimate $\hat{b}$ will not be consistent in the cases
   - λ_u and λ_y are both unknown, (**A2c**).
   - One of the variances λ_u and λ_y is known and the other is unknown.
2. The estimate $\hat{b}$ will be consistent, if either
   - both the noise variances λ_y and λ_u are known, (**A2d**).
   - their ratio λ_y/λ_u is known, (**A2e**).
3. Under any of the assumptions **A2d** or **A2e** it is in fact possible to consistently estimate the gain b_o, even if there are $2N$ data values, and $N + 2$ (or $N + 1$) unknowns. It is hence not the fact that the number of unknowns grows as fast as the number of data points, that causes the consistency problem in the other cases. ■

Extension of the above results to the dynamic case are possible, and will be reported elsewhere.

7 Parametric modelling of the input

In this method the errors–in–variable model (1), (3) is regarded as a multi-variable system with both $u(t)$ and $y(t)$ as outputs, [8]. Of crucial importance for this approach is the assumption **A3b**,

$$u_o(t) = \frac{C(q^{-1})}{D(q^{-1})}\ e(t)\ , \tag{34}$$

where $e(t)$ is a white noise with variance λ_e and the polynomials $C(q^{-1})$, $D(q^{-1})$ are of the form

$$\begin{aligned} C(q^{-1}) &= 1 + c_1\, q^{-1} + \cdots + c_p\, q^{-p} \\ D(q^{-1}) &= 1 + d_1\, q^{-1} + \cdots + d_p\, q^{-p} \end{aligned} \quad , \tag{35}$$

with known degree p; moreover, $C(z)$ and $D(z)$ are relatively prime and asymptotically stable polynomials. In this way, the whole errors-in-variables model can be considered as a system with a two–dimensional output vector $z(t) = (u(t)\, y(t))^T$ and three mutually uncorrelated white noise sources $e(t)$, $\tilde{u}(t)$ and $\tilde{y}(t)$:

$$\begin{pmatrix} u(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \frac{C(q^{-1})}{D(q^{-1})} & 1 & 0 \\ \frac{B(q^{-1})C(q^{-1})}{A(q^{-1})D(q^{-1})} & 0 & 1 \end{pmatrix} \begin{pmatrix} e(t) \\ \tilde{u}(t) \\ \tilde{y}(t) \end{pmatrix} . \tag{36}$$

The model (36) can be converted into the innovations form

$$z(t) = S(q^{-1})\, \varepsilon(t)\ , \tag{37}$$

where $S(z)$ is the square (2×2) canonical (stable and minimum–phase) spectral factor of $\phi_z(\omega)$ and $\varepsilon(t)$ is the two–dimensional innovations vector with identity covariance matrix. Since the innovations $\varepsilon(t)$ of model (37) can also be interpreted as the one step ahead prediction errors, a prediction error method can be applied for the identification of the system parameters. The parameter vector θ includes the coefficients of polynomials $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$, $D(q^{-1})$ and the variances λ_e, λ_y, λ_u. The solution to the identification problem is then determined by minimizing the loss function

$$J(\theta) = \det \left(\frac{1}{N} \sum_{t=1}^{N} \varepsilon(t)\, \varepsilon^T(t) \right)\ , \tag{38}$$

where N is the number of input–output samples.

Results:

1. This method yields consistent parameter estimates under quite mild conditions. See [8].
2. There are (also other) ways to introduce parametric models for the unperturbed input $u_o(t)$, that leads to consistent parameter estimators.

8 Conclusions

For errors-in-variables identification, it is often crucial what general assumptions on noise properties and on the unperturbed input that are imposed. Some assumptions of these types need to be made in order to get consistent estimates of the system parameters. Furthermore, it is not possible to verify the validity of these assumptions from the same experimental data. In contrast, in the case of no input errors ('classical identification' situations) these assumptions are not so critical. Then they normally do not influence the consistency properties, but the accuracy of the estimates.

Concerning the assumptions on the experimental conditions, the consequences can be summarized as follows.

1. If the unperturbed input is known to be *periodic*, it is easy to derive simple and consistent estimators. These estimators can also be applied to the case of a more general input, if the recorded data sets can be splitted into two (possibly overlapping) data sets, where the unperturbed inputs are correlated, and the noise contribution to the data are uncorrelated.

2. Another situation occurs when the unperturbed input is regarded as *completely unknown*, and estimated along with the system parameters. In the general case, it is not possible to get consistent parameter estimates. However, if the ratio between the output noise variance and the input noise variance is known, this approach is fully feasible. It leads to consistent estimates of the system parameters, despite that simultaneously all values of the unperturbed input are estimated.

3. Another possibility is to model the unperturbed input in a *parametric* way, for example that it is coming from a stationary process with a rational spectrum. This means that a form of joint input-output estimation process can be applied. This will be a prediction error method tied to a particular parameterization of the measured data, and gives consistent parameter estimates under weak assumptions.

Acknowledgement

This work was partially supported by Swedish Research Council for Engineering Sciences under contract 98-654.

References

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

[2] 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.

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

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

[5] 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.

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

[7] 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.

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

[9] T. Söderström. Why are errors-in-variables problems often tricky? Technical Report 2002-040, Department of Information Technology, Uppsala University, Uppsala, Sweden, November 2002.

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

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

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

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

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

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

[16] 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.

[17] W. X. Zheng. On least–squares identification of stochastic linear systems with noisy input–output data. *International Journal of Adaptive Control and Signal Processing*, 13:11–143, 1999.