IT Licentiate theses
2002-005

Identification of Dynamic Errors-in-Variables Models

KAUSHIK MAHATA

UPPSALA UNIVERSITY
Department of Information Technology

UPPSALA
UNIVERSITET

Identification of Dynamic Errors-in-Variables Models

BY
KAUSHIK MAHATA

May 2002

DEPARTMENT OF SYSTEMS AND CONTROL
INFORMATION TECHNOLOGY
UPPSALA UNIVERSITY
UPPSALA
SWEDEN

Dissertation for the degree of Licentiate of Philosophy in Signal processing
at Uppsala University 2002

Identification of Dynamic Errors-in-Variables Models

Kaushik Mahata
Kaushik.Mahata@SysCon.uu.se

Department of Systems and Control
Information Technology
Uppsala University
Box 337
SE-751 05 Uppsala
Sweden

http://www.it.uu.se/

ISSN 1404-5117
ISBN
Printed by the University Printers, Ekonomikum, Uppsala, Sweden

Abstract

The problem of identifying dynamic errors-in-variables models is of fundamental interest in many areas like process control, array signal processing, astronomical data reduction. In recent years, this field has received increased attention of the research community. In this thesis, some time domain and frequency domain approaches for identifying the errors-in-variables model is studied. The first chapter gives an overview of various methods for identifying dynamic errors-in-variables systems. Several approaches are classified and a qualitative comparison of different existing methods is also presented. The second chapter deals with instrumental variables based approaches. The least squares and the total least squares methods of solving the Yule-Walker equation is of central interest here. The methods are compared from the view point of asymptotic performance, numerical robustness and computation. The method presented in the third chapter uses prefiltered data. The input-output data is passed through a pair of user defined prefilters and the output data from the prefilters is subjected to a least-squares like algorithm. Compared to the IV approach, the proposed method shows a significant improvement in the small-sample properties of the MA parameter estimates, without any increase in the computational load. In the fourth chapter, we show that the two-dimensional process composed of the input-output data admits a finite order ARMA representation. Then we propose a parametric identification algorithm and another non-parametric identification method based on the ARMA representation.

List of publications

Chapter 1 of the thesis is based on
T. Söderström, U. Soverini and K. Mahata, Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing*, to appear.

The second chapter is a part of
T.Söderström and K.Mahata, On instrumental variable and total least squares approaches for identification of noisy systems. *International Journal of Control*, vol 75, No 6, 381-389.

The contents of the third chapter can also be found in
K. Mahata and T. Söderström, identification of dynamic errors-in-variables model using prefiltered data. *Proc. 15th IFAC World Congress*, Barcelona, Spain, July 21-26, 2002, to appear.

Chapter 4 of this thesis is based on
T. Söderström, K. Mahata and U. Soverini, Identification of dynamic errors-in-variables model using a frequency domain Frisch scheme. *Proc. 15th IFAC World Congress*, Barcelona, Spain, July 21-26, 2002, to appear.

Acknowledgments

I take the opportunity to express my sincere gratitude to my supervisor, Professor Tosten Söderström for sharing his profound knowledge and for his guidance. I also thank my co-author Professor Umberto Soverini for interesting discussions. Finally, thanks to all my colleagues in SysCon for your pleasant company and cooperation.

Contents

Chapter 1

Introduction

1.1 Outline

The central theme of this licentiate thesis is identification of dynamic errors-in-variables models. It comprises four main contributions.

- Firstly, in this chapter, the problem is introduced formally after a brief background. Some difficulties inherent in the problem are analyzed. Subsequently, a bird's eye view of different estimation methods is presented.

- Next, in the second chapter, instrumental variables based methods of solving the same is considered. The convensional approach is compared with the total least squares method of solving the Yule-Walker equations.

- In third chapter, we present a computationally efficient method where the data is prefiltered and then subjected to a least squares type machinery to obtain consistent estimates.

- Finally, in the fourth chapter, we introduce a data model where the input-output data vector is expressed as a two dimensional ARMA process. This ARMA model is exploited in two different ways to identify the unknown transfer function.

1.2 Background

Identification of linear dynamic systems from noise-corrupted output measurements is a fundamental research problem, which has been investigated in the past few decades. Many different solutions have been presented, see, for example, (Ljung 1999, Ljung and Söderström 1983, Söderström and Stoica 1989). On the other hand, the estimation of the parameters for linear dynamic systems when the input also is affected by noise is recognized as a more difficult problem which has received increasing attention recently. Representations where errors or measurement noises are present on both the inputs and outputs are usually called "errors-in-variables" models and play an important role when the identification purpose is the determination of the inner laws that describe the process, rather than the prediction of its future behaviour.

As proved by several applications collected in (Van-Huffel 1997), a very broad class of scientific disciplines makes use of such representations, such as time series modeling, 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 (Fuller 1987, Schuppen 1989). The extension to the dynamic case has been considered in the literature by several authors with different approaches. In fact, the problem can be examined from different point of views. For example, we can distinguish between time–domain (Ljung 1999, Söderström and Stoica 1989), and frequency-domain methods (Schoukens *et al.* 1997); approaches exploiting deterministic signals (e.g. periodic inputs (Forsell *et al.* 1999)) or stochastic processes, methods that make use of second–order statistics only or based on higher order cumulant statistics (Tugnait 1992, Tugnait and Ye 1995). When only second–order statistics are exploited it is a well–known result that, in general, the identification of errors–in–variables models cannot admit a single solution (Anderson 1985, Anderson and Deistler 1984, Anderson and Deistler 1987).

The stochastic input–output noises can also be represented in different ways by assuming different model structures. For example, in (Beghelli *et al.* 1990) the analysis of the Frisch scheme, originally related to the static case, has been extended to linear dynamic SISO systems under the assumption of white and uncorrelated additive noises with unknown variances. In (Beghelli *et al.* 1997) the identification of SISO errors–in–variables models has been analyzed by means of frequency–domain techniques. In (Castaldi and Soverini 1996) the identifiability conditions have been studied in case of SISO errors–in–variables model with coloured input–output noises with unknown rational spectra.

Some comparison between different approaches are given in (Söderström 1981, Soverini and Söderström 2000). A more complete description can be found in (Söderström *et al.* 2002*c*).

1.3 Problem statement

As a typical model example, consider the following system with noise-corrupted input and output measurements.

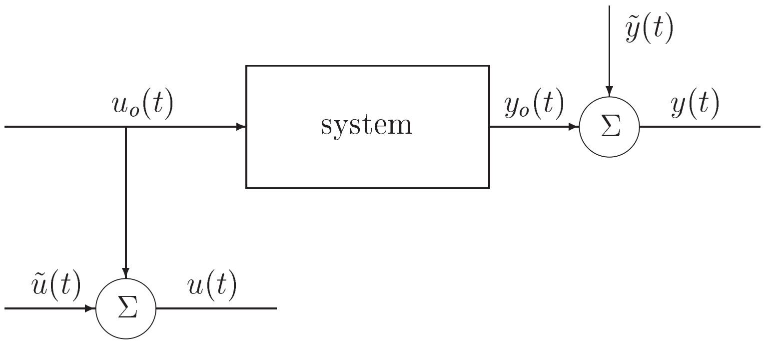


Figure 1.1: The basic setup for a dynamic error-in-variables problem.

The noise–free input $u_0(t)$ and the undisturbed output $y_0(t)$ are linked through the linear difference equation

$$A(q^{-1})\, y_0(t) = B(q^{-1})\, u_0(t). \tag{1.1}$$

$A(q^{-1})$ and $B(q^{-1})$ are polynomials of the type

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \cdots + a_n\, q^{-n} \\ B(q^{-1}) &= b_1\, q^{-1} + \cdots + b_n\, q^{-n} \end{aligned} \tag{1.2}$$

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 $A(q^{-1})$, $B(q^{-1})$ are of equal degree n, the order of the system.

In an errors–in–variables environment we assume that the observations are corrupted by additive measurement noises $\tilde{u}(t)$ and $\tilde{y}(t)$, at the input and output respectively. Therefore, the available signals are of the form

$$\begin{array}{rcl} u(t) & = & u_0(t) + \tilde{u}(t) \\ y(t) & = & y_0(t) + \tilde{y}(t) \end{array} . \tag{1.3}$$

The following assumptions are introduced.

A1. The dynamic system (1) is asymptotically stable, i.e. $A(z)$ has all zeros outside the unit circle; all the system modes are observable and controllable, i.e. $A(z)$ and $B(z)$ have no common factors. Moreover, the order n of the system is *a priori* known.

A2. The true input $u_0(t)$ is a zero–mean stationary ergodic random signal.

We may also have

A2'. The spectral density of $u_0(t)$ is a rational function.

A3. The sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero–mean white noises with unknown variances λ_u and λ_y, respectively. These processes are mutually uncorrelated and uncorrelated with the noise–free signals $u_0(t)$ and $y_0(t)$.

The problem under investigation can thus be stated as follows.

Problem. Let the available measurements $\{u(t)\}_{t=1}^{N}$ and $\{y(t)\}_{t=1}^{N}$ be generated in accordance with the previous assumptions. Determine the unknown parameter vector

$$\boldsymbol{\theta} = \left[\begin{array}{cccccc} a_1 & \dots & a_n & b_1 & \dots & b_n \end{array}\right]^T \tag{1.4}$$

or equivalently the transfer function

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} \tag{1.5}$$

and the variances λ_u and λ_y of the noise sequences. ■

The assumptions can sometimes be relaxed. In a few cases additional assumptions are required to be imposed.

- 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.
- Most of the methods can be extended to multi variable systems.
- The problem will be significantly simplified if the noise variances λ_u and λ_y are known, or even only the ratio λ_y/λ_u.
- A few approaches assume the model parameterization **A2'** of the unperturbed input signal $u_0(t)$ also. Maximum likelihood or prediction error approach is an example of such methods. Note that Assumption **A2'** implies that $u_0(t)$ can be modeled as an ARMA process.

In order to keep the description reasonably simple, we will stick to the assumptions stated earlier. The essential ingredients in the problem are still kept with this formulation.

1.4 What difficulties are inherent in the problem?

The aim of this section is to present two natural ways of coping with the estimation problem. As we will see, both fail to solve the problem in a successful way. We will offer some explanations and comment on some aspects on the identifiability properties.

1.4.1 Rewriting the dynamics into an ARMAX model structure

The idea here is to rewrite the total dynamics into an ARMAX form as follows. Consider the stationary process $x(t)$ defined as

$$\begin{aligned} x(t) &= A(q^{-1})y(t) - B(q^{-1})u(t) \\ &= A(q^{-1})[y_0(t) + \tilde{y}(t)] - B(q^{-1})[u_0(t) + \tilde{u}(t)] \\ &= A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t) \ . \end{aligned} \tag{1.6}$$

As the right hand side is the difference of two moving average processes, one of order n and the other of order $n-1$, one can apply a spectral factorization to get an equivalent description, in the sense that the second-order statistical properties of $x(t)$ remain the same. Therefore, introduce a monic polynomial $C(q^{-1})$ and a positive scalar λ such that

$$\lambda C(z)C(z^{-1}) = \lambda_y A(z)A(z^{-1}) + \lambda_u B(z)B(z^{-1}) \ . \tag{1.7}$$

Then we can replace the noise term in (1.7) by *one* moving average process as

$$x(t) = C(q^{-1})v(t) = A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t) \ . \tag{1.8}$$

where $v(t)$ is a white noise sequence of variance λ.

Thus, it is tempting to try a prediction error approach, (Ljung 1999, Söderström and Stoica 1989), where one is estimating the (extended) parameter vector

$$\boldsymbol{\theta}_e = \left[\hat{a}_1 \ldots \hat{a}_n \ \hat{b}_1 \ldots \hat{b}_n \ \hat{c}_1 \ldots \hat{c}_n\right]^T \tag{1.9}$$

in the straightforward way by minimizing the loss function

$$V_e(\boldsymbol{\theta}_e) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t, \boldsymbol{\theta}_e). \tag{1.10}$$

The prediction errors $\varepsilon(t, \boldsymbol{\theta}_e)$ are computed from the model structure

$$\hat{C}(q^{-1})\varepsilon(t, \boldsymbol{\theta}_e) = \hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t) \ . \tag{1.11}$$

The asymptotic estimate, as the number of data points N tends to infinity, is given by the minimizing vector of the asymptotic loss function

$$\begin{aligned}
V_e(\boldsymbol{\theta}_e) &= E\varepsilon^2(t, \boldsymbol{\theta}_e) \\
&= E\left[\frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})}y(t) - \frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})}u(t)\right]^2 \\
&= E\left[\frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})}\left(\frac{B(q^{-1})}{A(q^{-1})}u_0(t) + \tilde{y}(t)\right) - \frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})}\Big(u_0(t) + \tilde{u}(t)\Big)\right]^2 \\
&= E\left[\frac{\hat{A}(q^{-1})B(q^{-1}) - A(q^{-1})\hat{B}(q^{-1})}{A(q^{-1})\hat{C}(q^{-1})}u_0(t)\right]^2 \\
&\quad + E\left[\frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})}\tilde{y}(t)\right]^2 + E\left[\frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})}\tilde{u}(t)\right]^2 .
\end{aligned} \tag{1.12}$$

One can show that the gradient of the criterion *does not* vanish for the true parameter vector, that is

$$V_e'(\boldsymbol{\theta}_{e0}) \neq 0 \tag{1.13}$$

unless the input is measured exactly (so that $\lambda_u = 0$). Details are shown in the appendix A.1. This means that the estimator will not be consistent.

1.4.2 Nonparametric estimation of the true input signal

Another tempting approach is to treat the signal values of the undisturbed input, $\{u_0(t)\}_{t=1}^N$ as auxiliary unknowns to be estimated. Even if this can be easily done in practice, it seems though not to be a pertinent approach as shown in the following.

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

$$\begin{array}{rcl} y(t) & = & bu_0(t) + \tilde{y}(t) \\ u(t) & = & u_0(t) + \tilde{u}(t) \end{array} . \tag{1.14}$$

We regard

$$\boldsymbol{\theta}_i = [b \ \lambda_y \ \lambda_u]^T , \qquad U_0 = [u_0(1) \dots u_0(N)]^T \tag{1.15}$$

as the unknowns. Assume the data to be Gaussian distributed. The negative likelihood function will then be (neglecting a constant term)

$$L(\boldsymbol{\theta}_i, U_0) \ = \ \frac{1}{N} \sum_{t=1}^{N} \left\{ \frac{[y(t) - bu_0(t)]^2}{\lambda_y} + \frac{[u(t) - u_0(t)]^2}{\lambda_u} \right\} + \log{(\lambda_y \lambda_u)} \tag{1.16}$$

The loss function $L(\boldsymbol{\theta}_i, U_0)$ is easily first minimized with respect to U_0 giving

$$u_0(t) = \frac{b\lambda_u y(t) + \lambda_y u(t)}{b^2\lambda_u + \lambda_y}, \ t = 1, \dots, N \ . \tag{1.17}$$

It is shown in the appendix A.2 that the concentrated loss function

$$V_i(\boldsymbol{\theta}_i) = \min_{U_0} L(\boldsymbol{\theta}_i, U_0) \tag{1.18}$$

is *not* minimized for the true parameter values.

1.4.3 Concluding comments on identifiability

Based on the analysis of Section 1.4.1 and Section 1.4.2, we make the following claim.

Conjecture 1.1. Including all the values of the undisturbed input as additional parameter estimates, will not make it possible to estimate the parameter vector $\boldsymbol{\theta}$ consistently. ■

Note that the conjecture refers to a situation of nonparametric treatment of the undisturbed input. It does not matter if the estimation is carried out in the time domain or in the frequency domain. However, in case $u_0(t)$ is described by a parametric model (say with a fixed finite number of parameters), then the situation is different. See Section 1.5.4 for more details.

Conjecture 1.2. Some structural assumptions (such as model order, correlation structure of measurement noise) are needed for the estimation to be uniquely solved, but these assumptions cannot be verified from data. Expressed differently, the given data cannot be expressed in an unique way as due to measurement noise and the unknown input. ■

The following example supports and illustrates Conjecture 1.2.

Example 1.1. Let the measurement noises be auto-correlated, with spectral densities $\Phi_{\tilde{u}},\ \Phi_{\tilde{y}}$, respectively. Set the input output data vector

$$\mathbf{z}(t) = \left[\begin{array}{cc} y(t) & u(t) \end{array}\right]^T . \tag{1.19}$$

In the Gaussian case all information in the data is represented by the second order statistics, notably in the spectrum

$$\Phi_{\mathbf{z}} = \left[\begin{array}{cc} GG^* & G \\ G^* & 1 \end{array}\right] \Phi_{u_0} + \left[\begin{array}{cc} \Phi_{\tilde{y}} & 0 \\ 0 & \Phi_{\tilde{u}} \end{array}\right] . \tag{1.20}$$

With $\Phi_{\mathbf{z}}$ given and $G,\ \Phi_{u_0},\ \Phi_{\tilde{u}},\ \Phi_{\tilde{y}}$ as unknowns, there is no unique solution to (1.21). Recall that due to symmetry, the non diagonal elements of $\Phi_{\mathbf{z}}$ brings identical information.

To be more explicit, let the estimates of the aforementioned variables be denoted by $\hat{G},\ \hat{\Phi}_{u_0},\ \hat{\Phi}_{\tilde{u}},\ \hat{\Phi}_{\tilde{y}}$. Equating the elements of (1.21) leads to

$$\begin{aligned} \hat{\Phi}_{u_0} + \hat{\Phi}_{\tilde{u}} &= \Phi_{u_0} + \Phi_{\tilde{u}} && (1.21) \\ \hat{G}\hat{\Phi}_{u_0} &= G\Phi_{u_0} && (1.22) \\ \hat{G}\hat{G}^*\hat{\Phi}_{u_0} + \hat{\Phi}_{\tilde{y}} &= GG^*\Phi_{u_0} + \Phi_{\tilde{y}} . && (1.23) \end{aligned}$$

Choose a spectrum $\hat{\Phi}_{u_0}$ (so far arbitrarily). One can with straightforward calculations derive

$$\begin{aligned}
\hat{G} &= G\frac{\Phi_{u_0}}{\hat{\Phi}_{u_0}} && (1.24)\\
\hat{\Phi}_{\tilde{u}} &= \Phi_{\tilde{u}} + \Phi_{u_0} - \hat{\Phi}_{u_0} && (1.25)\\
\hat{\Phi}_{\tilde{y}} &= \Phi_{\tilde{y}} + GG^*\left(\Phi_{u_0} - \frac{\Phi_{u_0}^2}{\hat{\Phi}_{u_0}}\right) . && (1.26)
\end{aligned}$$

Requiring that $\hat{\Phi}_{\tilde{u}}$ and $\hat{\Phi}_{\tilde{y}}$ are positive definite function will give the possible values of $\hat{\Phi}_{u_0}$. Some straightforward calculations give

$$-\frac{\Phi_{u_0}\Phi_{\tilde{y}}}{\Phi_{u_0}GG^* + \Phi_{\tilde{y}}} \leq \hat{\Phi}_{u_0} - \Phi_{u_0} \leq \Phi_{\tilde{u}} \; . \qquad (1.27)$$

This shows clearly that except for the trivial and true solution $\hat{\Phi}_{u_0} = \Phi_{u_0}$, many more solutions are possible. ■

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.

In the paper (Scherrer and Deistler 1998), the problem of identifying multi variable EIV models from the population second moments of the observations is considered. Starting from the fact that this knowledge is not sufficient for uniquely identifying the underlying system, a class of observationally equivalent systems is derived and the continuity property of mapping relating this class to the spectral densities of the observations is investigated. A particular attention is turned to the models characterized by the maximum number of outputs.

1.5 Estimation Methods

Estimation methods can be classified and organized in different ways. Here, we have found it 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 1.2 below.

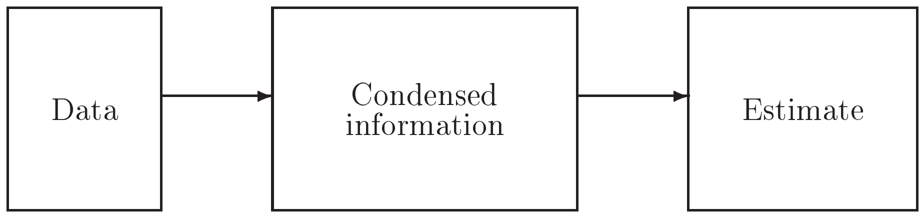


Figure 1.2: Classification of the data compression prior to the estimation step.

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.

- Using a symmetric covariance matrix
- Using a non symmetric covariance matrix
- Using the input-output spectrum
- Using the time-series data

At this point let us introduce a few notations, which has been followed in the rest of the thesis. Any scalar or vector quantity $\alpha(t)$, which is a function of the data $\{u(t)\}_{t=1}^{N}$, $\{y(t)\}_{t=1}^{N}$ is composed of two parts, i.e.

$$\alpha(t) = \alpha_0(t) + \tilde{\alpha}(t),$$

where we denote the contribution of noise free part of the data, i.e. $\{u_0(t)\}_{t=1}^{N}$ and $\{y_0(t)\}_{t=1}^{N}$ by $\alpha_0(t)$, while $\tilde{\alpha}(t)$ is used to denote the contribution of the measurement noise sequences $\{\tilde{u}(t)\}_{t=1}^{N}$ and $\{\tilde{y}(t)\}_{t=1}^{N}$. For example, consider the regressor vector

$$\begin{aligned}\boldsymbol{\psi}(t) &= \begin{bmatrix} -y(t-1) & \dots & -y(t-n) & u(t-1) & \dots & u(t-n) \end{bmatrix}^T & (1.28)\\ &= \boldsymbol{\psi}_0(t) + \tilde{\boldsymbol{\psi}}(t), & (1.29)\end{aligned}$$

where the corresponding noise-free and noisy parts of $\boldsymbol{\psi}(t)$ are given by

$$\begin{aligned}\boldsymbol{\psi}_0(t) &= \begin{bmatrix} -y_0(t-1) & \dots & -y_0(t-n) & u_0(t-1) & \dots & u_0(t-n) \end{bmatrix}^T & (1.30)\\ \tilde{\boldsymbol{\psi}}(t) &= \begin{bmatrix} -\tilde{y}(t-1) & \dots & -\tilde{y}(t-n) & \tilde{u}(t-1) & \dots & \tilde{u}(t-n) \end{bmatrix}^T & (1.31)\end{aligned}$$

We define the extended regrassor $\boldsymbol{\phi}(t)$ and the extended parameter vector $\overline{\boldsymbol{\theta}}$ as

$$\boldsymbol{\phi}(t) = \begin{bmatrix} -y(t) & \boldsymbol{\psi}^T(t) \end{bmatrix}^T, \quad \overline{\boldsymbol{\theta}} = \begin{bmatrix} 1 & \boldsymbol{\theta}^T \end{bmatrix}^T. \tag{1.32}$$

It is straight forward to see

$$\begin{aligned}y_0(t) &= \boldsymbol{\psi}_0(t)\boldsymbol{\theta} \quad \Rightarrow \; x_0(t) = \boldsymbol{\phi}_0(t)\overline{\boldsymbol{\theta}} = 0, & (1.33)\\ \Rightarrow \tilde{x}(t) &= x(t) = C(q^{-1})v(t) = -\tilde{\boldsymbol{\phi}}(t)\overline{\boldsymbol{\theta}}, & (1.34)\end{aligned}$$

where $x(t)$ is defined in (1.7). The innovation from of $x(t)$ given by (1.9) plays an important role in this problem the polynomial $C(q^{-1})$, the white sequence $v(t)$ and its variance λ appears very often in relevant expressions.

For any two stationary vectors $\alpha(t)$ and $\beta(t)$ we denote

$$R_{\alpha\beta} = E\alpha(t)\beta^T(t). \tag{1.35}$$

The estimate of $R_{\alpha\beta}$ is given by

$$\hat{R}_{\alpha\beta} = \frac{1}{N}\sum_{t=1}^{N}\alpha(t)\beta^T(t). \tag{1.36}$$

With this notational background we proceed to the next section where we describe bias compensating schemes.

1.5.1 Bias-compensating schemes

The classical least squares approach gives a biased parameter estimates for errors-in-variables models. To investigate this consider the least squares estimate $\hat{\boldsymbol{\theta}}_{ls}$, which satisfies as $N \to \infty$

$$\begin{aligned}R_{\boldsymbol{\psi\psi}}\hat{\boldsymbol{\theta}}_{ls} &= R_{\boldsymbol{\psi}y} = R_{\boldsymbol{\psi}_0 y_0} = R_{\boldsymbol{\psi}_0\boldsymbol{\psi}_0}\boldsymbol{\theta} = \left\{R_{\boldsymbol{\psi\psi}} - R_{\tilde{\boldsymbol{\psi}}\tilde{\boldsymbol{\psi}}}\right\}\boldsymbol{\theta}\\ \Rightarrow \hat{\boldsymbol{\theta}}_{ls} &= \boldsymbol{\theta} - R_{\boldsymbol{\psi\psi}}^{-1}R_{\tilde{\boldsymbol{\psi}}\tilde{\boldsymbol{\psi}}}\boldsymbol{\theta}. & (1.37)\end{aligned}$$

Note that, we have used the mutual independence of $\boldsymbol{\psi}_0(t)$,$\tilde{\boldsymbol{\psi}}(t)$, $y_0(t)$ and $\tilde{y}(t)$, while the third equality is straightforward from (1.34). One approach to identify the errors-in-variables model is to compensate for this asymptotic bias. This is possible if the noise variance estimates are available.

The Frisch scheme can be viewed as an advanced bias-compensating scheme. Note that, from (1.34) the matrix $R_{\boldsymbol{\phi}_0\boldsymbol{\phi}_0}$ is singular (more specifically, positive semidefinite with one eigenvalue equal to zero) and $\overline{\boldsymbol{\theta}}$ belongs to the corresponding null space. Now let $\hat{\lambda}_u$ be some estimate of λ_u. The estimated output noise variance $\hat{\lambda}_y$ is then constructed so that

$$\hat{R}_{\boldsymbol{\phi}\boldsymbol{\phi}} - \begin{bmatrix} \hat{\lambda}_y I_{n+1} & 0 \\ 0 & \hat{\lambda}_u I_n \end{bmatrix} \tag{1.38}$$

is singular. Note that, we denote the identity matrix of dimension n by I_n. Hence $\hat{\lambda}_y$ will be an implicit function of $\hat{\lambda}_u$, (Beghelli *et al.* 1990). The proce-

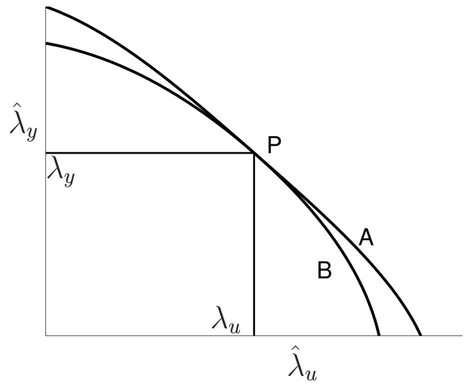


Figure 1.3: Illustration of the principle for Frisch estimation.

dure can be repeated with the model order n increased by one unit. In case $\hat{R}_{\boldsymbol{\phi}\boldsymbol{\phi}}$ would be replaced by its true value $R_{\boldsymbol{\phi}\boldsymbol{\phi}}$ we would then get a situation as displayed in Figure 1.3.

In Figure 1.3 curve A corresponds to the true model order, while curve B applies for the increased model order. The coordinates of the common point P give precisely the true noise variances λ_u, λ_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 implementation related aspects of the Frisch scheme.

The BELS method (Zheng 1999*b*, Zheng and Feng 1989, Zheng and Feng 1992) is another bias-compensation approach, where the noise variances and the parameter vector are estimated iteratively. Let us consider the model with over-parameterized numerator polynomial. Define

$$\boldsymbol{\theta}^* = \begin{bmatrix} \mathbf{a}^{*T} & \mathbf{b}^{*T} \end{bmatrix}^T, \tag{1.39}$$

where $\mathbf{a}^* = \begin{bmatrix} a_1 & \dots & a_n & 0 \end{bmatrix}^T$ and $\mathbf{b}^* = \begin{bmatrix} b_1 & \dots & b_n \end{bmatrix}^T$, the corresponding regressor vector being

$$\boldsymbol{\psi}^*(t) = \begin{bmatrix} -y(t-1) & \dots & -y(t-n-1) & u(t-1) & \dots & u(t-n) \end{bmatrix}^T. \tag{1.40}$$

The biased least squares estimate $\hat{\boldsymbol{\theta}}^*_{ls}$ with necessary modifications in notations in (1.38) satisfies

$$\lim_{N\to\infty} \hat{\boldsymbol{\theta}}^*_{ls} = \boldsymbol{\theta}^* - R^{-1}_{\boldsymbol{\psi}^*\boldsymbol{\psi}^*} R_{\tilde{\boldsymbol{\psi}}^*\tilde{\boldsymbol{\psi}}^*} \boldsymbol{\theta}^*. \tag{1.41}$$

One can also establish that (refer (Zheng 1998) for details),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \left[y(t) - \boldsymbol{\psi}^{*T}(t) \hat{\boldsymbol{\theta}}^*_{ls} \right]^2 = \lim_{N\to\infty} \left[\lambda_y \left(1 + \mathbf{a}^{*T} \hat{\mathbf{a}}^*_{ls} \right) + \lambda_u \mathbf{b}^{*T} \hat{\mathbf{b}}^*_{ls} \right] \tag{1.42}$$

$$- \lim_{N\to\infty} \mathbf{h}^T \hat{\mathbf{a}}^*_{ls} = \lambda_y \mathbf{h}^T R_{11} \mathbf{a}^* + \lambda_u \mathbf{h}^T R_{12} \mathbf{b}^*, \tag{1.43}$$

where we introduced the $(n+1)$ dimensional vector $\mathbf{h} = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}^T$. The matrix R_{11} and R_{12} are $(n+1) \times (n+1)$ and $(n+1) \times n$, respectively, and formed by the partition of the matrix $R^{-1}_{\boldsymbol{\psi}^*\boldsymbol{\psi}^*}$ as follows,

$$R^{-1}_{\boldsymbol{\psi}^*\boldsymbol{\psi}^*} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}. \tag{1.44}$$

The iteration is initialized using $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}^*_{ls}$ and one can solve approximately for λ_u and λ_y using (1.44) and (1.45). These values of λ_u and λ_y are used in (1.43) to get a new estimate of $\hat{\boldsymbol{\theta}}^*$. The steps are repeated until convergence. Similar procedure can be derived by overparameterising the numerator polynomial also. The method can also be extended for coloured output noise. We refer to (Zheng 1998, Zheng 1999*a*) for a detailed analysis.

Comments on early work in the field:

- One of the first solutions for identification of dynamic errors–in–variables models using a bias-compensated scheme was presented in the works of Koopmans and Levin (Koopmans 1937, Levin 1964) (see also (Aoki and Yue 1970)). In more recent years the Koopmans–Levin method has been re-proposed by several authors with slightly different variants (Cadzow and Solomon 1986, Fernando and Nicholson 1985, Van-Huffel and Vandewalle 1991). Koopmans–Levin method is constituted by the fact that the input and output noises are assumed white and the ratio of their variances is *a priori* known. These restrictive assumptions drastically reduce the practical use of the method.

- The Frisch scheme approach takes its name from a classical algebraic estimation problem, see (Frisch 1934), which was extended to the dynamic case in in (Beghelli *et al.* 1990). In this work it was developed a procedure for identifying linear dynamic systems in presence of additive and uncorrelated white noises with unknown (and possibly different) variances.

- The compensated-bias least-squares techniques (Stoica and Söderström 1982), based on suitable pre-filtering elaborations of the available data, are even more flexible since they can work under more general noise conditions than those of assumption **A2**. The appealing properties of these approaches have been deeply investigated in recent years and several recursive and off–line algorithms have been proposed (Zheng 1999*b*, Zheng and Feng 1989, Zheng and Feng 1992).

1.5.2 Instrumental variables methods

The instrumental variable method is an useful and computationally efficient tool for the estimation problems with complex structure. It is possible to develop instrumental variable based estimators for errors-in-variables problem also. In this section we shall introduce them. We shall maintain the notations introduced in earlier Sections.

From (1.34) we see that, for any data-dependent vector $\boldsymbol{\zeta}(t)$ of dimension $p \geq 2n$, which is correlated with $\boldsymbol{\phi}_0(t)$, satisfies

$$R_{\boldsymbol{\zeta}\boldsymbol{\psi}_0}\boldsymbol{\theta} = R_{\boldsymbol{\zeta}y_0} \Rightarrow R_{\boldsymbol{\zeta}\boldsymbol{\phi}_0}\overline{\boldsymbol{\theta}} = 0. \tag{1.45}$$

Now, if it is possible to construct $\boldsymbol{\zeta}(t)$ in such a way that

$$R_{\boldsymbol{\zeta}\tilde{\boldsymbol{\phi}}} \;=\; 0, \tag{1.46}$$

then

$$R_{\boldsymbol{\zeta\psi}}\boldsymbol{\theta} \;=\; R_{\boldsymbol{\zeta}y} \Rightarrow R_{\boldsymbol{\zeta\phi}}\overline{\boldsymbol{\theta}} = 0 \tag{1.47}$$

will hold and we can construct an instrumental variable estimate $\hat{\boldsymbol{\theta}}_{iv}$ of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}_{iv} = \hat{R}^{\dagger}_{\boldsymbol{\zeta\psi}}\hat{R}_{\boldsymbol{\zeta}y}. \tag{1.48}$$

Note that, as $N \to \infty$, (1.49) is equivalent to a search for the single dimensional null-space of $R^T_{\boldsymbol{\zeta\phi}}$. This motivates another consistent estimate $\hat{\boldsymbol{\theta}}_{tls}$ of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}_{tls} = \hat{\mathbf{w}}_{2:2n+1}/\hat{\mathbf{w}}_1 \tag{1.49}$$

in Matlab notation, where $\hat{\mathbf{w}}$ is the right singular vector of $\hat{R}_{\boldsymbol{\zeta\phi}}$. For a detailed comparison between $\hat{\boldsymbol{\theta}}_{tls}$ and $\hat{\boldsymbol{\theta}}_{iv}$, we refer the reader to (Söderström and Mahata 2002) and Chapter 2 of this thesis.

The main idea of the IV technique, based on well–known properties of the correlation functions, has been subsequently developed and generalized in several ways, e.g. by combining it with the concept of weighted subspace fitting (Stoica *et al.* 1995). This combined instrumental variable weighted subspace fitting, or IV-WSF, can be viewed as an extension to the approach discussed above. Here the regressor vector $\boldsymbol{\phi}(t)$ in (1.33) corresponds to a model order q such that $q \geq n$. Let the null vector of dimension $q - n$ be denoted by $\mathbf{o}$ and

$$\mathbf{a} \;=\; \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}^T, \quad \mathbf{b} = \begin{bmatrix} b_1 & \dots & b_n \end{bmatrix}^T, \tag{1.50}$$

$$\boldsymbol{\eta}_q(\boldsymbol{\theta}) \;=\; \begin{bmatrix} 1 & \mathbf{a}^T & \mathbf{o}^T & \mathbf{b}^T & \mathbf{o}^T \end{bmatrix}^T. \tag{1.51}$$

Consider the Toeplitz matrix $W_q(\boldsymbol{\theta})$ having $\boldsymbol{\eta}_q(\boldsymbol{\theta})$ as its first column and the first row identical to that of I_{q-n+1}. Then it can be shown that (1.48) gets modified to

$$R_{\boldsymbol{\zeta\phi}}W_q(\boldsymbol{\theta}) = 0. \tag{1.52}$$

An approach to exploit (1.53) is to fit $W_q(\boldsymbol{\theta})$ to span the null space of $R^T_{\boldsymbol{\zeta\phi}}$. Let the columns of $\tilde{W}_q$ be the right singular vectors corresponding to the least $q - n + 1$ singular values of $\hat{R}_{\boldsymbol{\zeta\phi}}$. Then one way to determine an estimate $\hat{\boldsymbol{\theta}}_{wsf}$ of $\boldsymbol{\theta}$ is to minimize the distance between $W_q(\hat{\boldsymbol{\theta}}_{wsf})$ and the span of $\tilde{W}_q$, i.e.

$$\hat{\boldsymbol{\theta}}_{wsf} = \arg\min_{\boldsymbol{\theta}} \left(\min_{C} \frac{1}{2} \left[\mathrm{vec}(W_q(\boldsymbol{\theta}) - \tilde{W}_q\Gamma) \right]^T Q \left[\mathrm{vec}(W_q(\boldsymbol{\theta}) - \tilde{W}_q\Gamma) \right] \right) \tag{1.53}$$

where Γ is a nonsingular $(q-n+1) \times (q-n+1)$ matrix, Q is a positive definite weighting matrix and may be data dependent, and $\text{vec}(\cdot)$ denotes the vectorization operation. For the details of this optimization steps, we refer the reader to (Stoica *et al.* 1995). The accuracy of the estimate $\hat{\boldsymbol{\theta}}_{wsf}$ can be optimized by choosing an optimal weight Q. The optimal weight determination is computationally expensive. The details of the computationally economical sub-optimal IV-WSF can be found in (Cedervall and Stoica 1996).

Comments:

- One of the principal advantages of IV the method is its applicability under fairly general noise conditions.
- It is rather inexpensive from computational point of view.
- However, with this procedure a poor accuracy of the parameter estimates is often obtained.
- $R_{\boldsymbol{\zeta\phi}}$ 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.
- IV based subspace algorithms employing the state space models are proposed in (Chou and Verhaegen 1997). These algorithms can be applied to MIMO systems operating in open or closed loop, where one has to account for the process noise also.

1.5.3 Frequency domain estimators

The idea here is similar to that of the Frisch scheme, with the important difference that we use the full spectral information in the data and not only the covariance matrix $R_{\boldsymbol{\phi\phi}}$.

To be more specific, consider the two dimensional input-output data model introduced in (1.20). The spectral density $\Phi_{\mathbf{z}}$ of the input-output data satisfies

$$\Phi_{\mathbf{z}}(\omega) = \begin{bmatrix} G(e^{-i\omega}) \\ 1 \end{bmatrix} \begin{bmatrix} G^*(e^{-i\omega}) & 1 \end{bmatrix} \Phi_{u_0}(\omega) + \begin{bmatrix} \lambda_y & 0 \\ 0 & \lambda_u \end{bmatrix}, \qquad \forall \omega \quad (1.54)$$

If the spectral density matrix $\Phi_{\mathbf{z}}$ is known, and an appropriate diagonal matrix is subtracted, then one would get a rank 1 matrix, corresponding to the first term of (1.55), for all frequencies ω. In case the decomposition as in (1.55) 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 $\Phi_{u_0}(\omega)$.

The above idea is developed in (Beghelli *et al.* 1997) using an FFT-based spectral estimator. The quality of the estimates is moderate. A more promising approach, based on a black-box modeling of the spectral density $\Phi_{\mathbf{z}}$ is described in (Söderström *et al.* 2002*b*). In fact, in (Söderström *et al.* 2002*b*) and (Söderström *et al.* 2002*a*) we consider a nonparametric approach, where the spectrum $\Phi_{\mathbf{z}}(\omega)$ (1.55) is modeled using a two-dimensional ARMA model and a two step procedure for fitting its parameters. Further, 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.

Other frequency domain identification techniques which are employing periodic noise-free input signal are proposed in (Guillaume *et al.* 1995), (Schoukens *et al.* 1997), see also (Pintelon and Schoukens 2001). It is essential for the techniques to work that the signal $u_0(t)$ is periodic, which has strong consequences on what information that can be retrieved from the data.

At least four full periods of the data (or four independent experiments) are assumed to be available in (Schoukens *et al.* 1997). Here, the data is first pre-processed to estimate the individual noise variances. The method can be extended to handle also cases where the noise is correlated. Once the noise variances can be regarded as known, it basically remains to estimate $G(e^{i\omega})$ from the first term in (1.55), which can be done in several ways. The resulting estimates will be consistent.

For the method treated in (Guillaume *et al.* 1995), there is no explicit estimation of the noise variances. Instead, a specific loss function which depends on the logarithmic values of the transfer function is applied. This approach yields 'practically consistent' estimates, i. e. the estimates will be biased, but the bias is insignificant.

1.5.4 Maximum likelihood and prediction error approach

In this method the errors-in-variables model (1.1), (1.3) is regarded as a multi variable system with both $u(t)$ and $y(t)$ as outputs (Söderström 1981). The crucial step for this approach is the assumption **A2'**, i.e. the noise–free input $u_0(t)$ is characterized by a rational spectrum. The signal $u_0(t)$ can thus be represented as an ARMA process

$$u_0(t) = \frac{F(q^{-1})}{D(q^{-1})}\; e(t) \;, \tag{1.55}$$

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

$$\begin{array}{rcl} F(q^{-1}) & = & 1 + f_1\, q^{-1} + \cdots + f_m\, q^{-m} \\ D(q^{-1}) & = & 1 + d_1\, q^{-1} + \cdots + d_m\, q^{-m} \end{array} \;, \tag{1.56}$$

with known degree m; moreover, $F(z)$ and $D(z)$ are relatively prime and asymptotically stable polynomials. In this way, the input-output data vector $\mathbf{z}(t)$ can be represented as

$$\mathbf{z}(t) = \begin{bmatrix} \frac{B(q^{-1})C(q^{-1})}{A(q^{-1})D(q^{-1})} & 0 & 1 \\ \frac{F(q^{-1})}{D(q^{-1})} & 1 & 0 \end{bmatrix} \begin{bmatrix} e(t) \\ \tilde{u}(t) \\ \tilde{y}(t) \end{bmatrix} \;. \tag{1.57}$$

Thus, the model $F(q^{-1})/D(q^{-1})$ 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 $B(q^{-1})$.

Using well–known techniques the model (1.58) can be converted into the innovations form

$$\mathbf{z}(t) = \mathcal{G}(q^{-1})\, \boldsymbol{\varepsilon}(t) \;, \tag{1.58}$$

where the 2×2 matrix polynomial $\mathcal{G}(z)$ is the unique stable and minimum–phase spectral factor of $\Phi_{\mathbf{z}}(\omega)$ and $\boldsymbol{\varepsilon}(t)$ is the two–dimensional white innovations vector. One way to compute (1.59) from (1.58) is to convert the model (1.58) into a state space form. Then, the innovations form (1.59) can be computed from the Kalman filter after solving a Riccati equation. Since the innovations $\boldsymbol{\varepsilon}(t)$ of model (1.59) 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 $\boldsymbol{\theta}_{pem}$ includes the coefficients of polynomials $A(q^{-1})$, $B(q^{-1})$, $F(q^{-1})$, $D(q^{-1})$ and the variances λ_e, λ_y, λ_u. The solution of the identification problem is then determined by minimizing the loss function

$$J(\boldsymbol{\theta}_{pem}) = \det \left(\frac{1}{N} \sum_{t=1}^{N} \boldsymbol{\varepsilon}(t)\, \boldsymbol{\varepsilon}^T(t) \right) \;, \tag{1.59}$$

where N is the number of input–output samples.

Comments and experiences:

- As it always happens for identification of multi-variable systems, the choice of a suitable parameterization of the model is a crucial point of the approach.
- In (Söderström 1979, Söderström 1981) it was proved that this method yields consistent parameter estimates under quite mild conditions.
- Its main disadvantage is that the numerical optimization procedure is, in general, quite involved since at every step a Riccati equation must be solved in order to find the innovations $\boldsymbol{\varepsilon}(t)$ used in (1.60).
- The procedure may fail to give good results when only poor initial parameter estimates are available.
- This technique can work with rather arbitrarily correlated noises, for which an extended system representation by including also models for the noises is needed.

Chapter 2

Instrumental variables and total least squares approaches for dynamic errors-in-variables model estimation

2.1 Introduction

Identification of linear dynamic systems from noise corrupted input-output data is a fundamental research problem and we can find its application in many practical fields. The prediction error method (PEM) is known as the statistically efficient tool to handle the problem (Ljung 1999, Söderström and Stoica 1989), in the sense that PEM is equivalent to the maximum-likelihood estimation when the data follow Gaussian distribution. The most appealing property of PEM is that the best possible statistical accuracy of the parameter estimates is achieved. On the other hand, it involves a numerical minimization of a nonlinear cost function of the parameters. This minimization procedure can often be time consuming, in particular when the input data is noisy (Söderström 1981). In such cases it might be useful to device some method which is computationally more economical even if the high statistical accuracy is not maintained. Instrumental Variable method (IVM) (Söderström and Stoica 1983, Young 1970, Young 1984) can be seen as an attempt to achieve this compromise.

Apart from the conventional IVM we also have a total least squares (TLS) algorithm (Stoica *et al.* 1992, Van-Huffel 1997), which also is computationally efficient. The goal of this chapter is to compare the conventional IVM with the TLS approach from different view points. We shall show that the approaches have the same asymptotic accuracy, in the sense that the parameter estimates have the same asymptotic distribution. In particular, we shall derive the explicit expressions for the asymptotic covariance matrices of the parameter estimates for both the approaches. In finite samples the approaches differ though. We shall discuss their relative performances in terms of numerical robustness and computational complexity.

2.2 The IV and TLS estimator

Recall from the discussion in Section 1.5.2 that, if it possible to construct a data dependent instrument vector $\boldsymbol{\zeta}(t)$ of dimension $p \geq 2n$ such that

- $\boldsymbol{\zeta}(t)$ is correlated with $\boldsymbol{\phi}_0(t)$,
- $R_{\boldsymbol{\zeta}\tilde{\boldsymbol{\phi}}} = 0$,

we have

$$\Gamma R_{\boldsymbol{\zeta\phi}}\boldsymbol{\theta} = 0 \quad \Rightarrow \quad \Gamma R_{\boldsymbol{\zeta\psi}}\boldsymbol{\theta} = \Gamma R_{\boldsymbol{\zeta}y}\boldsymbol{\theta}. \tag{2.1}$$

For example, for a second order system, the vector

$$\boldsymbol{\zeta}^T(t) = [u(t-3)\dots u(t-p-2)],$$

can be used as an instrument. Note that, we have introduced an arbitrary weighting matrix Γ, of rank $\dim\{\boldsymbol{\zeta}(t)\}$. It can be rectangular in general. Then the equation (2.1) can be used in two different ways to construct two different consistent estimators as

$$\hat{\boldsymbol{\theta}}_{iv} = \left(\Gamma \hat{R}_{\boldsymbol{\zeta\psi}}\right)^{\dagger} \Gamma \hat{R}_{\boldsymbol{\zeta}y} \tag{2.2}$$

and

$$\hat{\boldsymbol{\theta}}_{tls} = \hat{\mathbf{w}}_{2:2n+1}/\hat{\mathbf{w}}_1 \tag{2.3}$$

in Matlab notation, where $\hat{\mathbf{w}}$ is the right singular vector of $\Gamma\hat{R}_{\boldsymbol{\zeta\phi}}$ associated to its least singular value. Note that, $\hat{\mathbf{w}}$ is a consistent estimate of $\mathbf{w}$, the right singular vector of $\Gamma R_{\boldsymbol{\zeta\phi}}$ associated to its least singular value.

Comments:

- Note that, for $\hat{\boldsymbol{\theta}}_{iv}$ or $\hat{\boldsymbol{\theta}}_{tls}$ to exist, it is necessary to have $R_{\boldsymbol{\zeta\psi}}$ full rank. This is a persistence of excitation like condition on the noise-free input

$u_0(t)$. It is possible to construct $u_0(t)$ such that this condition is not satisfied. But this assumption is not restrictive in the sense that most inputs satisfy this condition, see the discussion on generic consistency in (Söderström and Stoica 1983, Söderström and Stoica 1989, Stoica *et al.* 1995).

- The estimator (2.3) can in fact be interpreted as the total least squares (TLS) (Golub and Loan 1989, Van-Huffel 1997), solution to the equation (2.1). Solving (2.1) in a TLS sense means that a deviation of the full matrix $\hat{R}_{\boldsymbol{\zeta}\boldsymbol{\phi}}$ is sought, as small as possible in Frobenious norm, so that the modified system of equations has an exact solution. This solution will then be given precisely by (2.3), see (Van-Huffel 1997).

- For the case when $\dim\{\Gamma\boldsymbol{\zeta}(t)\} = 2n$ the estimators $\hat{\boldsymbol{\theta}}_{iv}$ and $\hat{\boldsymbol{\theta}}_{tls}$ coincide. This is easily seen as follows. First note that in this case $\Gamma\hat{R}_{\boldsymbol{\zeta}\boldsymbol{\phi}}$ is of dimension $(2n|2n{+}1)$ and is hence singular. The smallest singular value is zero, and the associated singular vector turns out to be $\begin{bmatrix} 1 & \hat{\boldsymbol{\theta}}_{iv}^T \end{bmatrix}^T$:

$$\Gamma\hat{R}_{\boldsymbol{\zeta}\boldsymbol{\phi}} \begin{bmatrix} 1 \\ \hat{\boldsymbol{\theta}}_{iv} \end{bmatrix} = \begin{bmatrix} -\Gamma\hat{R}_{\boldsymbol{\zeta}y} & \Gamma\hat{R}_{\boldsymbol{\zeta}\boldsymbol{\psi}} \end{bmatrix} \begin{bmatrix} 1 \\ (\Gamma\hat{R}_{\boldsymbol{\zeta}\boldsymbol{\psi}})^{-1}\Gamma\hat{R}_{\boldsymbol{\zeta}y} \end{bmatrix} = 0. \tag{2.4}$$

- For the case $\dim\{\Gamma\boldsymbol{\zeta}(t)\} > 2n$ the two estimators differ.

We have the following main result of this chapter.

Theorem 2.1. If the noise processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are Gaussian, then both the estimates $\boldsymbol{\theta}_{iv}$ and $\boldsymbol{\theta}_{tls}$ are asymptotically Gaussian distributed as

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta}) \quad \stackrel{N\to\infty}{\longrightarrow} \quad \mathcal{N}(0, P), \tag{2.5}$$

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{iv} - \boldsymbol{\theta}) \quad \stackrel{N\to\infty}{\longrightarrow} \quad \mathcal{N}(0, P), \tag{2.6}$$

$$P = \lambda(R_{\boldsymbol{\zeta}\boldsymbol{\psi}}^T W R_{\boldsymbol{\zeta}\boldsymbol{\psi}})^{-1} R_{\boldsymbol{\zeta}\boldsymbol{\psi}}^T W \operatorname{cov}\left\{C(q^{-1})\boldsymbol{\zeta}(t)\right\} W R_{\boldsymbol{\zeta}\boldsymbol{\psi}} (R_{\boldsymbol{\zeta}\boldsymbol{\psi}}^T W R_{\boldsymbol{\zeta}\boldsymbol{\psi}})^{-1}, \tag{2.7}$$

where λ and $C(q^{-1})$ are defined in (1.8) and $W = \Gamma^T\Gamma$.

Proof. See Appendix B.1. ■

Note that the covariance matrices $\boldsymbol{\theta}_{iv}$ and $\boldsymbol{\theta}_{tls}$ have the same asymptotic distribution. A similar analysis when the data model is sum of sine waves in noise instead of a linear dynamic system (2.1) was given in (Stoica *et al.* 1992). It turned out that also in that case the two types of estimators, corresponding to (3.12) and (3.13), have the same asymptotic distribution.

The covariance matrix P apparently depend upon the weighting matrix W. It can be shown in the same way as other instrumental variable estimators, see (Söderström and Stoica 1989, Söderström and Stoica 1983), that there is an optimal choice of W. The matrix P is minimized with the choice

$$W_{opt} = \left[\text{cov}\{C((q^{-1})\boldsymbol{\zeta}(t)\}\right]^{-1}. \tag{2.8}$$

The corresponding choice of Γ can be obtained by the Cholesky factorization of W_{opt}.

We mention by passing that the above theorem is not restricted to errors-in-variables model. It has been shown in (Söderström and Mahata 2002) that the above is true for ARMAX model also.

2.3 Numerical illustration

To illustrate the above-mentioned results, we shall consider identification of a second order open loop ARMAX model and of an Errors in Variables model.

Let the individual parameter estimates in m independent realizations be denoted by $\{\hat{\boldsymbol{\theta}}_i\}_{i=1}^m$. Then the sample covariance matrix $\hat{P}$ is given by

$$\hat{P} = \frac{1}{m-1}\sum_{i=1}^{m}\left[\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right]\left[\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right]^T. \tag{2.9}$$

In order to compare $\hat{P}$, with the theoretical covariance matrix P, obtained in (2.7) the following holds with 95 % probability, see Appendix B9 of (Söderström and Stoica 1989),

$$\frac{|\hat{P}_{ij} - P_{ij}|}{|P_{ij}|} \leq \frac{3}{\sqrt{m}}. \tag{2.10}$$

P_{ij} is the element at i th row and j th column of P. The relation (2.10) applies in particular to diagonal elements ($i = j$) and to non-diagonal elements where $|P_{ij}|$ is not too small. It is also possible to set up a scalar criterion as follows. Introduce

$$V = tr\left(\left[\hat{P} - P\right]Q\left[\hat{P} - P\right]\right), \tag{2.11}$$

where Q is a positive definite matrix. Note that,

$$V \geq 0. \tag{2.12}$$

Then, it is possible to establish the following result.

Theorem 2.2. For the above-mentioned definitions of $\hat{P}$, P and V

$$EV = \frac{1}{m-1}\left[tr(PQP) + tr(P)tr(QP)\right]. \tag{2.13}$$

Proof. See Appendix B.2 ■

Putting $Q = P^{-2}$, in (2.11) and denoting the corresponding criterion by V_1 we have

$$EV_1 = \frac{1}{m-1}[2n + tr(P)tr(P^{-1})]. \tag{2.14}$$

Another possibility is to use $Q = I$. Denoting the associated criterion by V_2, this gives

$$EV_2 = \frac{1}{m-1}[tr(P^2) + (tr(P))^2]. \tag{2.15}$$

The criteria V_1 and V_2 follow approximately a χ^2 like distribution. One would expect that the value of V obtained from the simulations is not much larger than the corresponding value of EV.

Example 1. Consider the following second order ARMAX model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + H(q^{-1})e(t), \tag{2.16}$$

where $u(t)$ and $e(t)$ are mutually independent, unity-variance white noise sequences. The polynomials are given by

$$\begin{aligned} A(q^{-1}) &= 1 - 1.5q^{-1} + 0.7q^{-2}, & (2.17)\\ B(q^{-1}) &= 1.0q^{-1} + 0.5q^{-2}, & (2.18)\\ H(q^{-1}) &= 1 - 1.0q^{-1} + 0.2q^{-2}. & (2.19) \end{aligned}$$

The system was identified using delayed inputs as instruments, i.e.

$$\boldsymbol{\zeta}^T(t) = [u(t-1)\ldots u(t-p)]\,.$$

Five hundred independent Monte-Carlo simulations were performed with $p = 10$. For each run, IV and TLS estimates of the parameter vector were obtained using $N = 300$ data points. For both the IV and TLS method, the weighting matrix W was taken as the identity matrix. The results are summarized in the Table 2.1, where the parameter values are shown along with the corresponding standard deviations. The theoretical values of EV_1 and EV_2 are 3.09 and 0.11 respectively. The values of V_1 and V_2 obtained from

	Theoretical values	IV estimate	TLS estimate
a_1	-1.5 ± 0.0237	-1.48 ± 0.0261	-1.49 ± 0.0257
a_2	0.7 ± 0.0211	0.687 ± 0.0233	0.693 ± 0.0228
b_1	1 ± 0.0639	1.00 ± 0.0626	1.01 ± 0.0636
b_2	0.5 ± 0.0810	0.516 ± 0.0867	0.515 ± 0.0855

Table 2.1: Identification results for ARMAX model

	Theoretical values	IV estimate	TLS estimate
a_1	-1.5 ± 0.0196	-1.503 ± 0.0213	-1.501 ± 0.0223
a_2	0.7 ± 0.0202	0.704 ± 0.0206	0.702 ± 0.0220
b_1	1 ± 0.3056	0.945 ± 0.2768	0.967 ± 0.2950
b_2	0.5 ± 0.3236	0.550 ± 0.2900	0.531 ± 0.3015

Table 2.2: Identification results for Errors in Variables model

the simulations are 4.76 and 0.30 respectively, for the IV method. Those for the TLS method are 8.29 and 0.50.

Example 2. The Errors in Variables model treated is described as follows,

$$A(q^{-1})y_0(t) = B(q^{-1})u_0(t), \tag{2.20}$$

where

$$\begin{aligned} y(t) &= y_0(t) + \tilde{y}(t), \\ u(t) &= u_0(t) + \tilde{u}(t), \\ u_0(t) &= \frac{1 + 2q^{-1} + q^{-2}}{1 - 1.8q^{-1} + 0.9q^{-2}} e(t) \end{aligned} \tag{2.21}$$

and $e(t), \tilde{y}(t)$ and $\tilde{u}(t)$ are mutually independent white noise sequences, with individual variances 0.25, 16 and 16, respectively. Estimation methods which do not account for the presence of the input noise would lead to biased estimates for these noise levels. The polynomials $A(q^{-1})$and $B(q^{-1})$ are the same as in (2.18) and (2.19). In this case, the instruments used are given by

$$\boldsymbol{\zeta}^T(t) = [u(t-3) \ldots u(t-p-2)] \, .$$

Results obtained from 500 independent Monte-Carlo simulations with $p = 10$ and $N = 1000$, are summarized in Table 2.2. We rejected results obtained from some of the simulations where the estimated parameter vector deviates abnormally from the true parameter vector. We reject the estimate $\hat{\theta}$, if

	Theoretical values	IV estimate	TLS estimate
a_1	-1.5 ± 0.0196	-1.497 ± 0.0205	-1.499 ± 0.0207
a_2	0.7 ± 0.0164	0.697 ± 0.0171	0.699 ± 0.0173
b_1	1 ± 0.0587	0.992 ± 0.0591	0.996 ± 0.0595
b_2	0.5 ± 0.0284	0.497 ± 0.0284	0.497 ± 0.0285

Table 2.3: Identification results for Errors in Variables model when the true input having a flat spectrum

$|\hat{\theta} - \theta_0| > 1.5$. The theoretical values of EV_1 and EV_2 can be found to be, 225 and 0.19 respectively. The values of V_1 and V_2 obtained from the simulations are 163 and 0.58 respectively, for IV method. Those for TLS method are 175 and 0.78.

Example 3.
In this example, we consider the same system as in Example 2, except that the true input signal is given by

$$u_0(t) = \frac{1}{1 - 0.2q^{-1} + 0.5q^{-2}} e(t), \tag{2.22}$$

where $e(t)$ is a white noise with variance 16. Note that, in this case the input signal has a flat spectrum. In order to maintain the same SNR at the input and the output as in Example 2, the input noise variance λ_u is taken as 1.7 while the output noise variance λ_y is set to 0.25. Results from 500 independent Monte-Carlo simulations are shown in Table 2.3. The number of data points N in each simulation has been 1000. In this example, none of the estimates were rejected. The theoretical values of EV_1 and EV_2 can be found to be, 1.10 and 8×10^{-8} respectively. The values of V_1 and V_2 obtained from the simulations are 1.18 and 2×10^{-8} respectively, for IV method. Those for TLS method are 1.14 and 4×10^{-8}.

To compare the computational load of the IV and the TLS approaches, we evaluated the Matlab flops used to obtain $\hat{\boldsymbol{\theta}}_{iv}$ and $\hat{\boldsymbol{\theta}}_{tls}$ respectively from $\hat{R}_{\boldsymbol{\zeta\phi}}$. It turned out in this example that $\hat{\boldsymbol{\theta}}_{iv}$ requires 3212 flops, while for $\hat{\boldsymbol{\theta}}_{tls}$ 1212 flops were enough.

2.4 Discussion

As can be seen from the Examples 1, 2 and 3, the numerical simulation results are well in accordance with the theoretical covariance matrices derived in (2.7). From the numerical examples we can also note that the asymptotic equivalence of the two approaches does not ensure analogous small-sample properties. It is also of interest to compare the IV and TLS estimators in terms of robustness and noise sensitivity on one hand, and computational complexity on the other.

The conventional IV method has better small sample performance, and the TLS approach is somewhat less robust for some type of input signals. This can be explained as follows. Let $R_{\boldsymbol{\zeta\phi}}$ have ordered singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{2n} \geq \sigma_{2n+1} = 0$. The true parameter vector obtained from the singular vector $\mathbf{w}$ associated with σ_{2n+1}. Let the estimated matrix $\hat{R}_{\boldsymbol{\zeta\phi}}$ have the singular values $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \ldots \geq \hat{\sigma}_{2n} \geq \hat{\sigma}_{2n+1}$. If $\hat{\sigma}_{2n}$ and $\hat{\sigma}_{2n+1}$ are close, it may occasionally happen that it is the singular vector corresponding to $\hat{\sigma}_{2n}$, rather than that associated to $\hat{\sigma}_{2n+1}$, that is close to $\mathbf{w}$. For a given SNR of the input and output data, this is likely to occur if the condition number σ_1/σ_{2n} is large, which happens if the input signal has a peaky spectrum or if the system to be identified has poles close to the unit circle.

In terms of the computational complexity, we first note that forming the matrix $\hat{R}_{\boldsymbol{\zeta\phi}}$ from the input output data is a common step for both the IV and TLS estimators. For large data record (N large), this will be the dominating part of the computation. In what follows, we discuss the difference in the computational load when computing the estimators from $\hat{R}_{\boldsymbol{\zeta\phi}}$.

To compute $\hat{\boldsymbol{\theta}}_{iv}$ we have to compute SVD of the matrix $\Gamma\hat{R}_{\boldsymbol{\zeta\phi}}$. A slightly more economical way would be to use the QR factorization, as we must anyway assume $\Gamma\hat{R}_{\boldsymbol{\zeta\phi}}$ to have full rank (Golub and Loan 1989). On the contrary, $\hat{\boldsymbol{\theta}}_{tls}$ is based on SVD of $\Gamma\hat{R}_{\boldsymbol{\zeta\phi}}$, which has one further column compared to $\Gamma\hat{R}_{\boldsymbol{\zeta\psi}}$. It is important to note though that a full SVD is not required as only the right singular vector associated to the smallest singular value is needed. There are economical ways to compute this singular vector avoiding a full SVD computation, (Van-Huffel 1997).

The numerical findings of Example 3 support and illustrate the above reasoning that $\hat{\boldsymbol{\theta}}_{tls}$ is less costly computationally than $\hat{\boldsymbol{\theta}}_{iv}$.

2.5 Conclusions

For identifying noise-corrupted systems, we have discussed how certain Yule-Walker type of equations can be used for computing parameter estimates.

A standard choice is the classical instrumental variable (IV) estimator. We described and analyzed, also an alternative total least squares (TLS) approach. The two approaches were shown to be identical in simple cases with a minimum number of equations, and to be different but having the same asymptotic distribution in the general case. The analysis is carried out also for the errors-in-variables case, where the input signal is subjected to measurement errors. The theoretical results are further illustrated and supported by Monte-Carlo simulations. For substantially high noise levels, the IV approach seemed to be the more robust alternative, while the TLS approach is computationally more economical.

Chapter 3

Identification of dynamic errors-in-variables model using prefiltered data

3.1 Introduction

As discussed in chapter 2, one of the computationally efficient solutions of the dynamic errors-in-variables problem is the instrumental variable method and related subspace fitting approaches (Stoica *et al.* 1995, Söderström and Mahata 2002). The accuracy of the parameter estimates obtained by IV based methods is often poor, particularly when the system poles are close to the unit circle or the input signal spectrum is peaky. This poor accuracy of IV based methods can mainly be attributed to the fact that they use the estimates of correlation of the data at 'high' lags. In this chapter, we propose an approach where the data is prefiltered by a pair of user defined filters. The design of the prefilters enables us to use a least-squares like method, where correlation estimates at lower lags are used. We also demonstrate that significant improvement in the estimate accuracy is achieved for the same computational cost.

3.2 Estimation using prefiltered data

The often encountered poor accuracy of the instrumental variable estimator is a direct consequence of (1.47). The elements of $\boldsymbol{\zeta}(t)$ must be sufficiently delayed with respect to the elements of $\boldsymbol{\psi}(t)$ in order to satisfy (1.47). For instance, we can use $\boldsymbol{\zeta}(t) = \boldsymbol{\psi}(t-\tau)$ for $\tau \geq n$. Since it is difficult to accurately estimate the correlation of the data at higher lags, often poor estimates of $R_{\boldsymbol{\zeta}\boldsymbol{\psi}}$ and $R_{\boldsymbol{\zeta}y}$ may result, which leads to poor accuracy of the estimates. Here, an attempt will be made to remedy this problem by passing the data through a pair of user defined filters.

Let us assume that the input-output data are passed through a pair of FIR filters $F_1(q^{-1})$ and $F_2(q^{-1})$, each of order $m > n$. Let us denote for $i = 1, 2$

$$u_i(t) = F_i(q^{-1})u(t), \tag{3.1}$$
$$y_i(t) = F_i(q^{-1})y(t), \tag{3.2}$$
$$x_i(t) = F_i(q^{-1})x(t), \tag{3.3}$$
$$\boldsymbol{\psi}_i(t) = F_i(q^{-1})\boldsymbol{\psi}(t), \tag{3.4}$$
$$\boldsymbol{\phi}_i(t) = F_i(q^{-1})\boldsymbol{\phi}(t). \tag{3.5}$$

It is straightforward to see from (1.1) and (1.31) that for $i = 1, 2$

$$\boldsymbol{\psi}_{i0}^T(t)\boldsymbol{\theta} = y_{i0}(t). \tag{3.6}$$

Thus, if it is possible to design $F_1(q^{-1})$ and $F_2(q^{-1})$ such that

$$R_{\tilde{\boldsymbol{\phi}}_1\tilde{\boldsymbol{\phi}}_2} = 0, \tag{3.7}$$

we can obtain two estimators of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}_1 = \hat{R}^{\dagger}_{\boldsymbol{\phi}_1\boldsymbol{\psi}_2}\hat{R}_{\boldsymbol{\phi}_1 y_2}, \tag{3.8}$$
$$\hat{\boldsymbol{\theta}}_2 = \hat{R}^{\dagger}_{\boldsymbol{\phi}_2\boldsymbol{\psi}_1}\hat{R}_{\boldsymbol{\phi}_2 y_1}, \tag{3.9}$$

provided $R_{\boldsymbol{\phi}_{10}\boldsymbol{\psi}_{20}}$ and $R_{\boldsymbol{\phi}_{20}\boldsymbol{\psi}_{10}}$, respectively, are full rank. We have the following result concerning the design of $F_1(q^{-1})$ and $F_2(q^{-1})$ in order to satisfy (3.7).

Theorem 3.1. Let us define the polynomial $H(q)$ as

$$H(q) = q^m F_1(q^{-1})F_2(q) = \sum_{i=0}^{2m} h_i q^i. \tag{3.10}$$

Then (3.7) is satisfied if

$$h_i = 0, \quad m - n \le i \le m + n. \tag{3.11}$$

Proof. See Appendix C.1. ∎

Thus, we can start with $H(q)$, the polynomial of order $2m$ satisfying (3.11) and factor it accordingly to obtain $F_1(q^{-1})$ and $F_2(q^{-1})$. We have the following asymptotic result for the accuracy of the estimates $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$.

Theorem 3.2. Define the polynomials

$$
\begin{aligned}
C(q^{-1})F_1(q^{-1})F_2(q^{-1}) &= \mathcal{C}(q^{-1}), && (3.12)\\
A^2(q^{-1})H^2(q^{-1}) &= \sum_{i=0}^{2m+2n} \beta_i^{(aa)} q^{-i}, && (3.13)\\
A(q^{-1})B(q^{-1})H^2(q^{-1}) &= \sum_{i=0}^{2m+2n} \beta_i^{(ab)} q^{-i}, && (3.14)\\
B^2(q^{-1})H^2(q^{-1}) &= \sum_{i=0}^{2m+2n} \beta_i^{(bb)} q^{-i}. && (3.15)
\end{aligned}
$$

If the measurement noise sequences are Gaussian, $\hat{\boldsymbol{\theta}}_i$, $i = 1, 2$ are asymptotically Gaussian distributed as

$$
\begin{aligned}
&\sqrt{N}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_i), && (3.16)\\
&P_i = R^{\dagger}_{\boldsymbol{\phi}_i\boldsymbol{\psi}_j} Q R^{\dagger T}_{\boldsymbol{\phi}_i\boldsymbol{\psi}_j}, i \neq j \in \{1, 2\} && (3.17)
\end{aligned}
$$

where

$$
Q = \lambda \mathrm{cov}\left[\mathcal{C}(q^{-1})\boldsymbol{\phi}(t)\right] + Q_e \qquad (3.18)
$$

and Q_e is given by (C.27)-(C.33) in appendix.

Proof. See Appendix C.2. ■

3.3 Numerical illustration

To illustrate numerically the identification methods introduced in the previous section, we shall consider the second order system with

$$
\begin{aligned}
A(q^{-1}) &= 1 - 1.5q^{-1} + 0.7q^{-2}, && (3.19)\\
B(q^{-1}) &= 1.0q^{-1} + 0.5q^{-2}. && (3.20)
\end{aligned}
$$

The true input $u_0(t)$ is an ARMA process given by

$$
u_0(t) = \frac{1 + 2q^{-1} + q^{-2}}{1 - 1.8q^{-1} + 0.9q^{-2}} e(t), \qquad (3.21)
$$

Estimate		a_1	a_2	b_1	b_2
$\hat{\boldsymbol{\theta}}_1$	Mean	-1.50	0.70	1.00	0.49
	S.D(n)	0.037	0.029	0.165	0.208
	S.D(t)	0.031	0.024	0.120	0.156
$\hat{\boldsymbol{\theta}}_2$	Mean	-1.50	0.70	1.00	0.49
	S.D(n)	0.036	0.028	0.127	0.162
	S.D(t)	0.031	0.022	0.120	0.156
$\hat{\boldsymbol{\theta}}_{iv}$ $(p = 6)$	Mean	-1.50	0.70	0.88	0.64
	S.D(n)	0.022	0.029	0.792	0.944
	S.D(t)	0.015	0.022	0.556	0.661
$\hat{\boldsymbol{\theta}}_{iv}$ $(p = 10)$	Mean	-1.50	0.70	0.99	0.51
	S.D(n)	0.014	0.013	0.239	0.274
	S.D(t)	0.013	0.013	0.215	0.246

Table 3.1: Simulation results form 500 Monte-Carlo simulations.

Estimate	$\hat{\boldsymbol{\theta}}_1$	$\hat{\boldsymbol{\theta}}_2$	$\hat{\boldsymbol{\theta}}_{iv}$ $(p = 6)$	$\hat{\boldsymbol{\theta}}_{iv}$ $(p = 10)$
	30579	30579	31741	49173

Table 3.2: Computational load involved in different algorithms, in terms of Matlab flops.

where $e(t)$ is a zero mean white noise sequence with variance 0.25. The variance of both the measurement noise sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ is 16. As an illustrative example, we consider the polynomial $H(q^{-1})$ as

$$H(q^{-1}) = 1 - q^{-6}. \tag{3.22}$$

Note that, the value of m is set to 3 and (3.7) is satisfied with this choice of $H(q^{-1})$. Among several possibilities of factorizing $H(q^{-1})$ into a pair of polynomials of order 3, see (3.10), we have arbitrarily taken $F_1(q^{-1})$ and $F_2(q^{-1})$ as

$$F_1(q^{-1}) = 1 + 2q^{-1} + 2q^{-2} + q^{-3}, \tag{3.23}$$

$$F_2(q^{-1}) = 1 - 2q^{-1} + 2q^{-2} - q^{-3}. \tag{3.24}$$

As candidate algorithms, along with the proposed algorithm using prefiltered data, we consider the IV algorithms with the instrument vector given by

$$\boldsymbol{\zeta}(t) = \begin{bmatrix} u(t-3) & \dots & u(t-p-2) \end{bmatrix}^T \tag{3.25}$$

for different values of p, the dimension of the instrument vector. We have summarized the simulation results in Table 3.1. The results are based on 500 independent Monte-Carlo simulations. The number of data points N in each simulation was 300. For each method we have given the sample mean and sample standard deviation denoted by SD(n). The standard deviations of the parameters obtained from the Monte-Carlo simulations are compared with the corresponding asymptotic standard deviation denoted by SD(t) calculated using Theorem 2.1 and Theorem 3.2.

In Table 3.2, we have compared the number of Matlab floating point operations involved in each of the algorithms.

3.4 Discussion

From Table 3.1, we see that the IV estimates of the A parameters are reasonably accurate, but the corresponding accuracy of the B parameters are very poor for $p = 6$. Increasing the instrument dimension from 6 to 10, an improvement in the B parameter estimation can be noticed. On the other hand, using the approach using the prefiltered data, the improvement in the B parameter estimation is significant. One demerit of the proposed algorithm is the slight loss of accuracy in A parameter estimation. If we compare the computation involved in the algorithms in Table 3.2, we see that increasing the dimension of the instrument from 6 to 10, the computational load is increased significantly, while the proposed algorithm is of less computational complexity.

In the illustrative example above, the estimates $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ have been designed in such a way that they have the same asymptotic properties. But it is interesting to note that their small sample properties differ. As can be seen from Table 3.1, $\hat{\boldsymbol{\theta}}_2$ performs better than $\hat{\boldsymbol{\theta}}_1$. This can be attributed to the fact that $\hat{\boldsymbol{\theta}}_2$ uses $\boldsymbol{\psi}_1(t)$ as the regressor vector and $\boldsymbol{\phi}_2(t)$ as the instrument vector. Since in the example considered here, the system to be identified is low pass and $F_1(q^{-1})$ is also a low pass filter, the system properties are excited better using $\boldsymbol{\psi}_1(t)$ as the regressor vector.

3.5 Conclusions

In this work, we have presented a computationally economical algorithm to identify the dynamic errors-in-variables model. The algorithm shows significant improvement in the accuracy of the system parameter estimates over the instrumental variable method. The method can also be used, when the measurement noise sequences satisfy more generalized constraints. The same algorithm can handle the case when $\tilde{u}(t)$ and $\tilde{y}(t)$ are correlated with each other. In case the sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are moving average processes of finite order, we can increase the parameter m to account for that. The polynomial $H(q^{-1})$ is to be designed by the user. In the example in the last section $H(q^{-1})$ was selected in an ad-hoc manner. But it is possible to make it data dependent, i.e. one can think of an iterative procedure of adjusting the coefficients of the polynomial $H(q^{-1})$, based on the last estimates of the system parameters, to arrive at statistically more accurate estimates. For example, the expression (3.17) can be used to optimize the covariance matrix of the estimates in some sense to derive such a scheme.

Chapter 4

Identification of dynamic errors-in-variables model: Approaches based on two-dimensional ARMA modeling of the data

4.1 Introduction

In this chapter, we combine the ideas of the Frisch scheme with the estimation of the joint input-output spectrum using multi-variable ARMA modeling. First, we show that under the assumption **A2'**, the two-dimensional process composed of the input-output data admits an ARMA representation of finite order. Subsequently, we propose some parametric and non-parametric ways of identifying the SISO errors-in-variables model using this ARMA model. The non-parametric method uses the ARMA structure to estimate of the joint input-output spectrum. The Frisch scheme is then employed to extract an estimate of the joint spectrum of the noise free input-output signals from the ARMA spectrum. The spectrum of the noise-free input-output leads us to a non-parametric estimate of the transfer function of the system. A similar attempt can be found in (Beghelli *et al.* 1997) which uses a periodogram estimate of the spectrum. The algorithm presented in this work outperforms the method introduced in (Beghelli *et al.* 1997). The parametric method exploits the ARMA structure in a more direct way.

4.2 Modeling of the data

As mentioned in the previous section, the unperturbed input $u_0(t)$ will be assumed to be an ARMA process in this chapter, i.e.

$$u_0(t) \;=\; \frac{F(q^{-1})}{D(q^{-1})}e(t), \tag{4.1}$$

where the polynomials $F(q^{-1})$ and $D(q^{-1})$ are given by

$$\begin{aligned} F(q^{-1}) &= 1 + f_1\,q^{-1} + \cdots + f_m\,q^{-m} \\ D(q^{-1}) &= 1 + d_1\,q^{-1} + \cdots + d_m\,q^{-m} \end{aligned} \tag{4.2}$$

and $e(t)$ is a zero-mean white noise with variance λ_e. The order m is assumed to be known apriori. As a starting point, we show that the input-output data vector $\mathbf{z}(t)$ admits an ARMA representation. Let us introduce

$$\boldsymbol{\Delta}(q^{-1}) = I + \Delta_1 q^{-1} + \cdots + \Delta_K q^{-K} = \begin{bmatrix} A(q^{-1}) & -B(q^{-1}) \\ 0 & D(q^{-1}) \end{bmatrix}, \tag{4.3}$$

for which one can easily show using (1.1), (1.2) and (1.3) that

$$\boldsymbol{\Delta}(q^{-1})\mathbf{z}(t) = \begin{bmatrix} A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t) \\ F(q^{-1})e(t) + D(q^{-1})\tilde{u}(t) \end{bmatrix}. \tag{4.4}$$

From (4.4) we see that $\boldsymbol{\Delta}(q^{-1})\mathbf{z}(t)$ is clearly a moving average process in the sense that its covariance function vanishes after $\max(m,n)$ lags. Therefore, $\mathbf{z}(t)$ admits the two dimensional ARMA representation

$$\boldsymbol{\Delta}(q^{-1})\mathbf{z}(t) = \boldsymbol{\Gamma}(q^{-1})\boldsymbol{\epsilon}(t). \tag{4.5}$$

where $\boldsymbol{\Gamma}(q^{-1})$ is a 2×2 matrix polynomial of order $K = \max(m,n)$ with the identity matrix as the leading coefficient, i.e.

$$\boldsymbol{\Gamma}(q^{-1}) \;=\; I + \Gamma_1 q^{-1} + \cdots + \Gamma_K q^{-K}. \tag{4.6}$$

Further, in (4.5) $\boldsymbol{\epsilon}(t)$ is the two dimensional innovation of $\mathbf{z}(t)$, a zero mean white noise with covariance matrix $\boldsymbol{\Lambda}$.

Although the ARMA representation of a multidimensional process is non-unique in general (Gevers and Wertz 1984), the special structure of the polynomial $\boldsymbol{\Delta}(q^{-1})$ enables us to derive also the parametric estimates of the polynomials $A(q^{-1})$, $B(q^{-1})$, $F(q^{-1})$ and $D(q^{-1})$ along with the noise variances λ_y, λ_u and λ_e from $\boldsymbol{\Delta}(q^{-1})$, $\boldsymbol{\Gamma}(q^{-1})$ and $\boldsymbol{\Lambda}$.

4.3 Estimation methods

The ARMA structure (4.4) motivates different types of identification approaches. The matrix polynomial $\mathbf{\Delta}(q^{-1})$ contains information of the system. We can think of some nonparametric approaches too. But the starting point would be to identify the ARMA model (4.4).

4.3.1 Estimating the two-dimensional ARMA model

In order to estimate the ARMA model in (4.4), we adopt a well-known procedure, previously used in (Durbin 1959, Mayne and Firoozan 1982), etc. The idea is to first estimate the innovation process $\{\boldsymbol{\epsilon}(t)\}_{t=1}^{N}$ and then use it as a second input, thus using a linear regression model to estimate $\mathbf{\Delta}(q^{-1})$ and $\mathbf{\Gamma}(q^{-1})$.

In order to estimate the innovation $\boldsymbol{\epsilon}(t)$, we fit an AR model of large order L to the data $\mathbf{z}(t)$, i.e.

$$\left[I + A_1 q^{-1} + \cdots + A_L q^{-L}\right] \mathbf{z}(t) = \boldsymbol{\epsilon}(t). \tag{4.7}$$

The model order L should be sufficiently large, in order to ensure a good approximation of the ARMA model by the AR model. The AR parameters $\{A_k\}_{k=1}^{L}$ are estimated using least squares. Let us introduce

$$\begin{aligned} \Theta_1 &= \begin{bmatrix} A_1 & \ldots & A_L \end{bmatrix}, \\ Z_1(t) &= \begin{bmatrix} -\mathbf{z}^T(t-1) & \ldots & -\mathbf{z}^T(t-L) \end{bmatrix}^T. \end{aligned}$$

Then we have

$$\begin{aligned} \widehat{\Theta}_1 &= \arg\min_{\Theta_1} \sum_{t=1}^{N} \|\mathbf{z}(t) - \Theta_1 Z_1(t)\|^2 \\ &= \left[\sum_{t=1}^{N} \mathbf{z}(t) Z_1^T(t)\right] \left[\sum_{t=1}^{N} Z_1(t) Z_1^T(t)\right]^{-1}. \end{aligned} \tag{4.8}$$

Hence, we get an estimate of the innovation $\boldsymbol{\epsilon}(t)$ as

$$\hat{\boldsymbol{\epsilon}}(t) = \mathbf{z}(t) - \widehat{\Theta}_1 Z_1(t). \tag{4.9}$$

We also obtain an estimate of $\mathbf{\Lambda}$ as

$$\hat{\mathbf{\Lambda}} = \frac{1}{N} \sum_{t=1}^{N} \hat{\boldsymbol{\epsilon}}(t) \hat{\boldsymbol{\epsilon}}^T(t). \tag{4.10}$$

Next, we consider $\hat{\boldsymbol{\epsilon}}(t)$ as an additional input to estimate the ARMA model (4.4). Note that the order of ARMA model $K = \max(n, m)$ is known due to the assumptions **A1** and **A2'**. Hence we have the model as

$$\left[I + \Delta_1 q^{-1} + \cdots + \Delta_K q^{-K}\right] \mathbf{z}(t) = \left[I + \Gamma_1 q^{-1} + \cdots + \Gamma_K q^{-K}\right] \boldsymbol{\epsilon}(t). \quad (4.11)$$

We shall use least squares again to estimate the ARMA model. Let us introduce

$$\begin{aligned} \Theta_2 &= \left[\begin{array}{cccccc} \Delta_1 & \ldots & \Delta_K & \Gamma_1 & \ldots & \Gamma_K \end{array}\right], \\ Z_2(t) &= \left[\begin{array}{cccccc} -\mathbf{z}^T(t-1) & \ldots & -\mathbf{z}^T(t-K) & \hat{\boldsymbol{\epsilon}}^T(t-1) & \ldots & \hat{\boldsymbol{\epsilon}}^T(t-K) \end{array}\right]^T. \end{aligned}$$

Note that, the special structure of $\boldsymbol{\Delta}(q^{-1})$ must be imposed on the corresponding elements of Θ_2 in order to ensure the uniqueness of the identified model. Using least squares we have

$$\begin{aligned} \widehat{\Theta}_2 &= \arg\min_{\Theta_2} \sum_{t=1}^{N} \|\mathbf{z}(t) - \Theta_2 Z_2(t)\|^2 \\ &= \left[\sum_{t=1}^{N} \mathbf{z}(t) Z_2^T(t)\right] \left[\sum_{t=1}^{N} Z_2(t) Z_2^T(t)\right]^{-1}. \end{aligned} \quad (4.12)$$

Form $\widehat{\Theta}_2$ we get the spectrum estimate of $\mathbf{z}(t)$ as

$$\hat{\Phi}_{\mathbf{z}}(\omega) = \hat{\boldsymbol{\Delta}}^{-1}(e^{-i\omega})\hat{\boldsymbol{\Gamma}}(e^{-i\omega})\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Gamma}}^T(e^{i\omega})\hat{\boldsymbol{\Delta}}^{-T}(e^{i\omega}), \quad (4.13)$$

where

$$\hat{\boldsymbol{\Delta}}(q^{-1}) = I + \hat{\Delta}_1(q^{-1}) + \cdots + \hat{\Delta}_K(q^{-K}), \quad (4.14)$$

$$\hat{\boldsymbol{\Gamma}}(q^{-1}) = I + \hat{\Gamma}_1(q^{-1}) + \cdots + \hat{\Gamma}_K(q^{-K}). \quad (4.15)$$

4.3.2 Frisch based non-parametric identification (FNP)

In Section 1.4.3 we introduced the spectrum $\Phi_{\mathbf{z}}$ of the process $\mathbf{z}(t)$, which is given by

$$\Phi_{\mathbf{z}}(\omega) = \Phi_{\mathbf{z}_0}(\omega) + \Phi_{\tilde{\mathbf{z}}}(\omega), \quad (4.16)$$

where $\Phi_{\mathbf{z}_0}$ and $\Phi_{\tilde{\mathbf{z}}}$ are given by

$$\Phi_{\mathbf{z}_0}(\omega) = \left[\begin{array}{cc} G(e^{-i\omega})G^*(e^{-i\omega}) & G(e^{-i\omega}) \\ G^*(e^{-i\omega}) & 1 \end{array}\right] \Phi_{u_0}(\omega), \quad (4.17)$$

$$\Phi_{\tilde{\mathbf{z}}}(\omega) = \left[\begin{array}{cc} \lambda_y & 0 \\ 0 & \lambda_u \end{array}\right]. \quad (4.18)$$

The idea here is to use the spectrum estimate in (4.13) and follow the methodology of the Frisch scheme. we first estimate the noise variances λ_u and λ_y using the fact from (4.17) that $\Phi_{\mathbf{z_0}}(\omega)$ is singular for all ω. Subsequently, the estimates of λ_u and λ_y are used to obtain the non-parametric estimate of $G(e^{-i\omega})$ using (4.16) and (4.17). We refer to this estimate as the FNP estimate. Since $\Phi_{\mathbf{z}_0}$ is singular, using (4.16) and (4.18) we get

$$\{\Phi_{yy}(\omega) - \lambda_y\}\{\Phi_{uu}(\omega) - \lambda_u\} = |\Phi_{yu}(\omega)|^2. \tag{4.19}$$

Rearranging we get

$$\begin{bmatrix} \Phi_{yy}(\omega) & \Phi_{uu}(\omega) & -1 \end{bmatrix} \Theta_3 = \Phi_{yy}(\omega)\Phi_{uu}(\omega) - |\Phi_{yu}(\omega)|^2, \tag{4.20}$$

where

$$\Theta_3 = [\ \lambda_u \quad \lambda_y \quad \lambda_y\lambda_u\ ]^T \tag{4.21}$$

is a vector with *three* unknown parameters. We can derive an overdetermined system of linear equations in Θ_3 by repeating this relation for a large number of frequencies. The solution of this system gives an estimate of Θ_3.

A more sophisticated way to obtain $\widehat{\Theta}_3$ would be to impose the constraint, cf (4.21),

$$\Theta_3^{(1)}\Theta_3^{(2)} = \Theta_3^{(3)}, \tag{4.22}$$

where we denote the i th element of Θ_3 by $\Theta_3^{(i)}$. Using the constraint (4.22), it is possible to set up a non-linear least squares problem in *two* unknowns. One can convert it to a nonlinear search in one parameter only.

Once the estimates $\hat{\lambda_u}$ and $\hat{\lambda_y}$ (and hence $\hat{\Phi}_{\tilde{\mathbf{z}}}$) are obtained, we have the estimate of $\Phi_{\mathbf{z_0}}(\omega)$ as

$$\hat{\Phi}_{\mathbf{z_0}}(\omega) = \hat{\Phi}_{\mathbf{z}}(\omega) - \hat{\Phi}_{\tilde{\mathbf{z}}} \tag{4.23}$$

A simple way to estimate $G(e^{-i\omega})$ from $\hat{\Phi}_{\mathbf{z_0}}(\omega)$ is

$$\hat{G}(e^{-i\omega}) = \frac{\hat{\Phi}_{\mathbf{z}_0}^{(12)}}{\hat{\Phi}_{\mathbf{z}_0}^{(22)}}. \tag{4.24}$$

A more sophisticated estimate may be derived by using also the remaining elements of $\hat{\Phi}_{\mathbf{z_0}}(\omega)$. This can be done by posing the following set of equations

$$\hat{\Phi}_{\mathbf{z}_0}^{(11)} = G(e^{-i\omega})G(e^{i\omega})\Phi_{u_0}, \tag{4.25}$$

$$\hat{\Phi}_{\mathbf{z}_0}^{(12)} = G(e^{-i\omega})\Phi_{u_0}, \tag{4.26}$$

$$\hat{\Phi}_{\mathbf{z}_0}^{(22)} = \Phi_{u_0} \tag{4.27}$$

and solving a nonlinear least squares problem. Again the dependence on Φ_{u_0} is linear, so the problem is separable and can be reduced to an optimization problem over $G(e^{-i\omega})$. As $G(e^{-i\omega})$ is complex, the solution will be reduced to a two dimensional search, treating the real and the imaginary parts of $G(e^{-i\omega})$ as two unknowns.

4.3.3 Frisch based parametric identification (FP)

The NFP estimate discussed in the previous section will be modified into a parametric estimate in this section. We shall use the FNP estimate of $G(e^{i\omega})$ to derive the estimate $\hat{\boldsymbol{\theta}}_{fp}$ of $\boldsymbol{\theta}$. With a little modification of notations we define

$$G(e^{-i\omega}, \boldsymbol{\theta}) = \frac{b_1 e^{-i\omega} + \cdots + b_n e^{-in\omega}}{1 + a_1 e^{-i\omega} + \cdots + a_n e^{-in\omega}}. \tag{4.28}$$

One possible way to estimate $\boldsymbol{\theta}$ is to minimize the weighted spectral distance between $G(e^{-i\omega}, \boldsymbol{\theta})$ and $\hat{G}(e^{-i\omega})$ given by (4.24), i.e.

$$\hat{\boldsymbol{\theta}}_{fp} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{M} W_k |G(e^{-i\omega_k}, \boldsymbol{\theta}) - \hat{G}(e^{-i\omega_k})|^2 \tag{4.29}$$

where $\{\omega_k\}_{k=1}^{M} \subset [0, \pi]$ are user defined discrete frequency points and W_k is a user chosen weight corresponding to frequency ω_k, can be data dependent.

4.3.4 Parametric method using the ARMA structure (PARMA)

As mentioned earlier, parametric estimates of $A(q^{-1})$, $B(q^{-1})$ and $D(q^{-1})$ can also be obtained from $\hat{\boldsymbol{\Delta}}(q^{-1})$ using (4.3). Let us introduce the process

$$\mathbf{v}(t) = \boldsymbol{\Delta}(q^{-1})\mathbf{z}(t) = \boldsymbol{\Gamma}(q^{-1})\boldsymbol{\epsilon}(t) \tag{4.30}$$

and its correlation sequence

$$\mathbf{r}_\tau = \mathbf{E}\mathbf{v}(\mathbf{t})\mathbf{v}^{\mathbf{T}}(\mathbf{t} - \tau). \tag{4.31}$$

The sequence $\{\mathbf{r}_\tau\}_{\tau=1}^{K}$ can then be used to estimate $F(q^{-1})$, λ_u, λ_y and λ_e. There are two ways to estimate $\mathbf{r}_\tau$. We can estimate $\mathbf{v}(t)$ and then estimate $\mathbf{r}_\tau$ as follows.

$$\hat{\mathbf{v}}(t) = \hat{\boldsymbol{\Delta}}(q^{-1})\mathbf{z}(t), \tag{4.32}$$

$$\hat{\mathbf{r}}_\tau = \frac{1}{N}\sum_{t=1}^{N} \hat{\mathbf{v}}(t)\hat{\mathbf{v}}^T(t - \tau). \tag{4.33}$$

There is an alternative approach, too. From (4.30) and (4.31) it is straightforward to derive

$$\mathbf{r}_\tau = \sum_{i=\tau}^{K} \Gamma_i \mathbf{\Lambda} \Gamma_{i-\tau}^T, \tag{4.34}$$

where $\Gamma_0 = I$. Thus, the previously obtained estimates of Λ and $\{\Gamma_i\}_{i=1}^K$ can be substituted in (4.34) to obtain $\hat{\mathbf{r}}_\tau$.

To explore how $F(q^{-1})$, λ_u, λ_y and λ_e are determined, let us examine $\mathbf{r}_\tau$ element wise. Combining (4.3), (4.30) and (4.31), we have

$$\begin{aligned}
\mathbf{r}_\tau^{(11)} &= E\left\{A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t)\right\}\left\{A(q^{-1})\tilde{y}(t-\tau) - B(q^{-1})\tilde{u}(t-\tau)\right\} \\
&= \lambda_y \sum_{i=0}^{n-\tau} a_i a_{i+\tau} + \lambda_u \sum_{i=1}^{n-\tau} b_i b_{i+\tau}, \qquad 0 \le \tau \le n; && (4.35) \\
\mathbf{r}_\tau^{(12)} &= E\left\{A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t)\right\}\left\{F(q^{-1})e(t-\tau) + D(q^{-1})\tilde{u}(t-\tau)\right\} \\
&= -\lambda_u \sum_{i=0}^{n-\tau} d_i b_{i+\tau}, \qquad 0 \le \tau \le n; && (4.36) \\
\mathbf{r}_\tau^{(21)} &= E\left\{F(q^{-1})e(t) + D(q^{-1})\tilde{u}(t)\right\}\left\{A(q^{-1})\tilde{y}(t-\tau) - B(q^{-1})\tilde{u}(t-\tau)\right\} \\
&= -\lambda_u \sum_{i=1}^{m-\tau} b_i d_{i+\tau}, \qquad 0 \le \tau \le m-1; && (4.37)
\end{aligned}$$

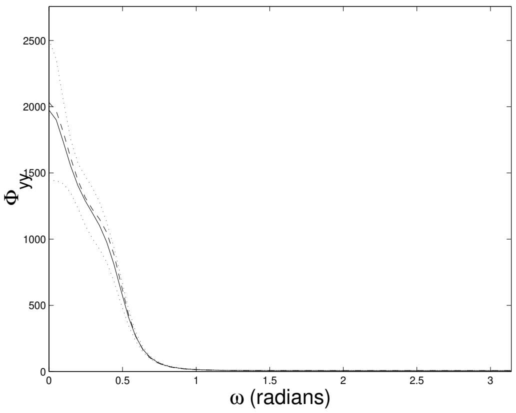
where $b_0 = 0$ and $a_0 = d_0 = 1$. Note that estimates of $A(q^{-1})$, $B(q^{-1})$ and $D(q^{-1})$ are already known. Hence (4.35), (4.36) and (4.37) can be used for different values of $\tau \le K$ to get an overdetermined system of equations in λ_y and λ_u. This system can be solved to obtain the estimates of the unknowns. Now we can use the remaining element of $\mathbf{r}_\tau$ to estimate $F(q^{-1})$. We have

$$\begin{aligned}
\mathbf{r}_\tau^{(22)} = \; & E\left\{F(q^{-1})e(t) + D(q^{-1})\tilde{u}(t)\right\} \\
& \left\{F(q^{-1})e(t-\tau) + D(q^{-1})\tilde{u}(t-\tau)\right\}, \qquad (4.38)
\end{aligned}$$

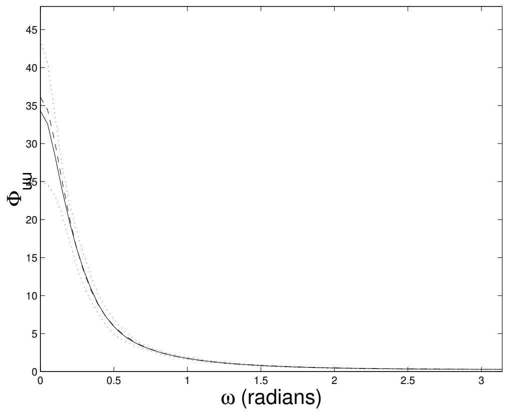
and

$$\sum_{\tau=-m}^{m} \mathbf{r}_\tau^{(22)} q^{-\tau} - \lambda_u D(q^{-1})D(q) = \lambda_e F(q^{-1})F(q). \tag{4.39}$$

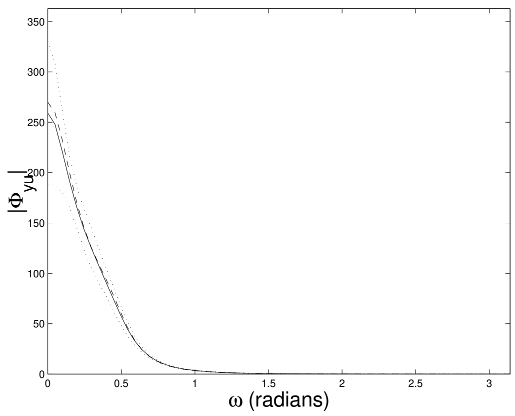
The polynomial at the left hand side of (4.39) can be factorized to obtain $\hat{F}(q^{-1})$. Since $\hat{\mathbf{r}}_\tau^{(22)}$ and $\hat{\lambda}_u$ somewhat deviate from their true values, it may happen that the left hand side cannot be exactly factorized. This problem is encountered precisely when the polynomial is not positive definite.



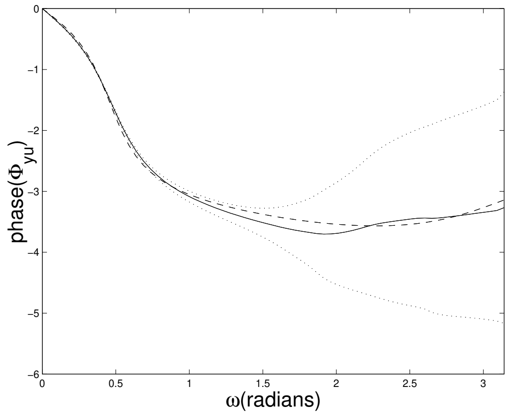
(a) Φ_{yy}



(b) Φ_{uu}



(c) Magnitude Φ_{yu}



(d) Phase Φ_{yu}

Figure 4.1: Spectrum estimates at 15 dB input–output SNR and $L = 20$. The mean value from Monte Carlo simulations (solid line), the mean $\pm$ one standard deviation (dotted line) and the true value (dashed line).

4.4 Numerical illustration

To illustrate numerically the identification methods introduced in the previous sections, we shall consider the system transfer function as follows

$$A(q^{-1}) = 1 - 1.5q^{-1} + 0.7q^{-2}, \tag{4.40}$$

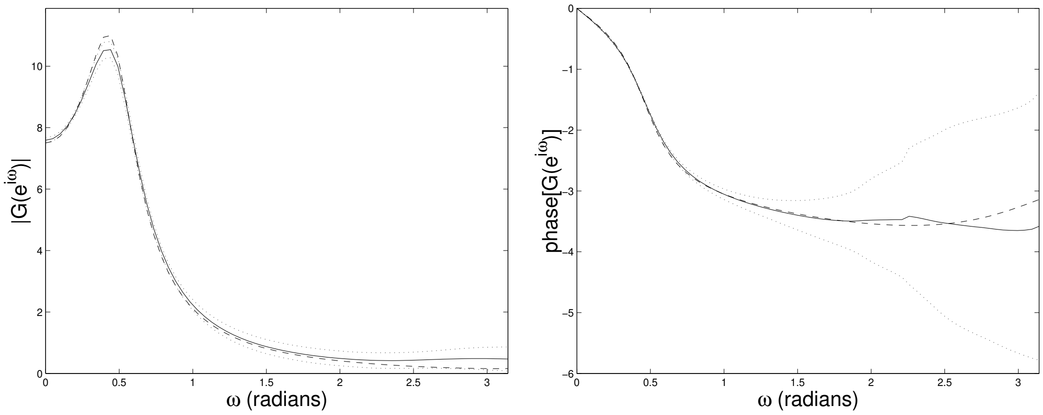
$$B(q^{-1}) = 1.0q^{-1} + 0.5q^{-2}, \tag{4.41}$$

$$F(q^{-1}) = 1 + 0.2q^{-1}, \tag{4.42}$$

$$D(q^{-1}) = 1 - 0.8q^{-1}. \tag{4.43}$$

The process $e(t)$ is zero mean white noise and its variance equals unity. We shall consider different combinations of λ_u and λ_y to illustrate the performance of the methods at different noise levels. The number of data points N was 1000 in each Monte-Carlo simulation.

In Figure 4.1, we have compared the mean value and the standard deviation of the spectrum estimate $\hat{\Phi}_{\mathbf{z}}$ from 50 independent Monte-Carlo simulations with the true spectrum $\Phi_{\mathbf{z}}$. The variance of the input noise λ_u has been taken as 0.11, while the output noise variance is 7.5. With these choices the SNR at both the input and the output are 15 dB approximately. The order L of the AR model involved in estimation of the innovation $\boldsymbol{\epsilon}(t)$ in (4.7), has been taken as 20. The order K of the ARMA model in (4.4) is 2. We have used 128 equally spaced discrete frequency points in the interval $[0, \pi]$ to set up the optimization problem in order to obtain $\hat{\boldsymbol{\theta}}_{fp}$ defined in (4.29).

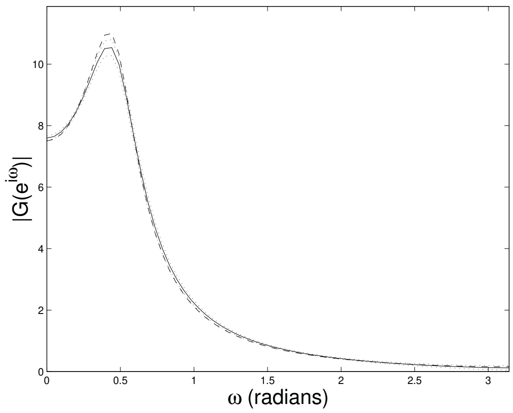


(a) Magnitude (b) Phase

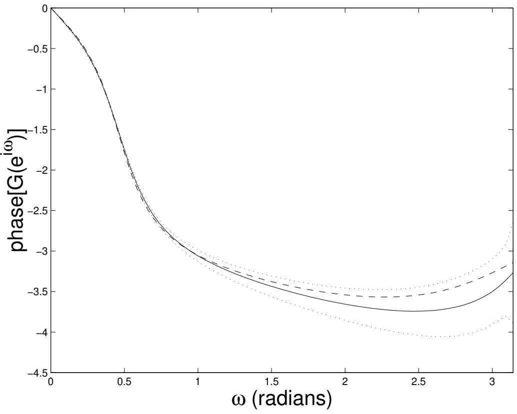
Figure 4.2: FNP estimates of the transfer function at 15 dB input–output SNR and $L = 20$. The mean value from Monte Carlo simulations (solid line), the mean $\pm$ one standard deviation (dotted line) and the true value (dashed line).

The FNP estimates of the transfer function $\hat{G}(e^{-i\omega})$ are given in Figure 4.2. We have compared the true transfer function $G(e^{-i\omega})$ with the mean value and the standard deviation of $\hat{G}(e^{-i\omega})$ obtained form the Monte-Carlo runs.

Since the FNP transfer function estimate given by (4.24) is expected to be more accurate at the frequencies for which the data SNR is large, the natural



(a) Magnitude



(b) Phase

Figure 4.3: FP estimates of the transfer function at 15 dB input–output SNR and $L = 20$. The mean value from Monte Carlo simulations (solid line), the mean $\pm$ one standard deviation (dotted line) and the true value (dashed line).

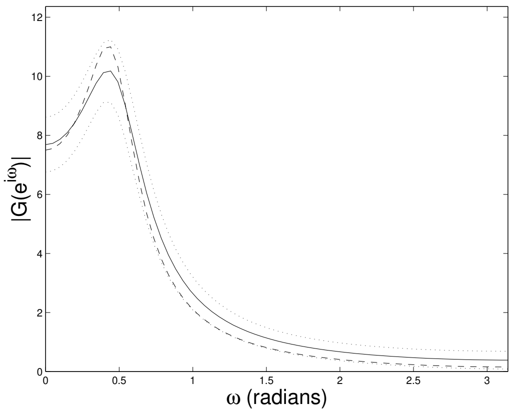
choice of the weighting vector W_k in (4.29) while deriving the FP estimate would be

$$W_k = \hat{\Phi}_{u_0 u_0}(\omega_k). \tag{4.44}$$

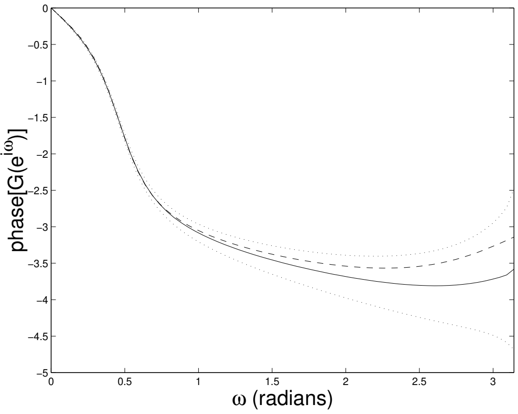
Using this choice of W_k during the optimization in (4.29), the FNP estimate of the transfer function at a particular frequency is assigned a weight directly proportional to the SNR at that frequency. The FP estimate of the transfer function $G(e^{-i\omega})$ obtained from (4.29) is shown in Figure 4.3.

Let us turn back to the parametric estimation method described in Section 4.4. In Figure 4.4, we have depicted the results of identification when the system is identified using PARMA method.

It can be observed in Figure 4.1 that the estimates of the spectrum obtained using the two-dimensional ARMA model is reasonably good for the noise level considered. The degradation of the performance in the high frequency region is predictable, since the SNR is significantly low at high frequencies for this example. We shall come back to this point later in this section. It can also be seen in Figure 4.1(d) that the performance of the cross-spectrum phase estimator is sensitive to noise. The same inference can be drawn from Figures 4.2 and 4.3 too. The way the spectrum $\hat{\Phi}_{\mathbf{z}}$ is estimated has a large



(a) Magnitude



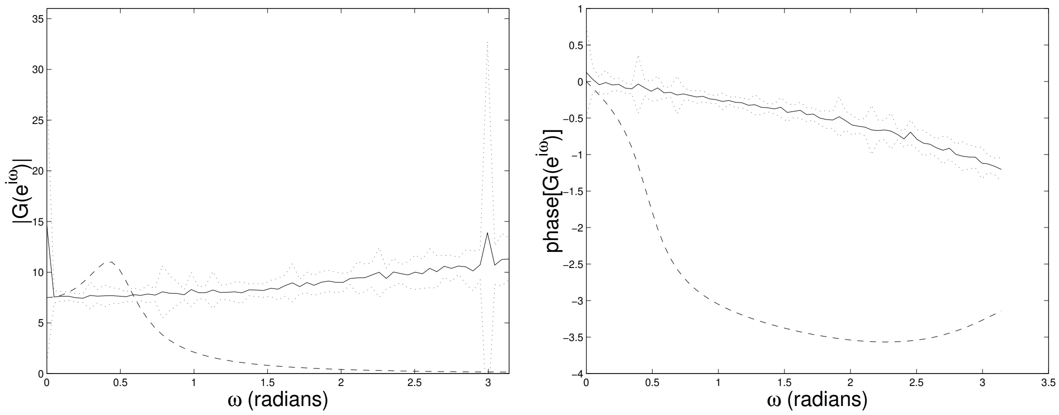
(b) Phase

Figure 4.4: Parametric transfer function estimates at 15 dB input and output SNR using PARMA method. $L = 20$. The mean from Monte Carlo simulations (solid line), the mean $\pm$ one standard deviation (dotted line) and the true value (dashed line).

impact on the estimate of the transfer function. As an illustration, consider the periodogram approach as attempted in (Beghelli *et al.* 1997). The results are displayed in Figure 4.5. It can be seen that the transfer function estimate based on the periodogram spectrum estimator is poor compared to the transfer function estimate computed by the proposed algorithm.

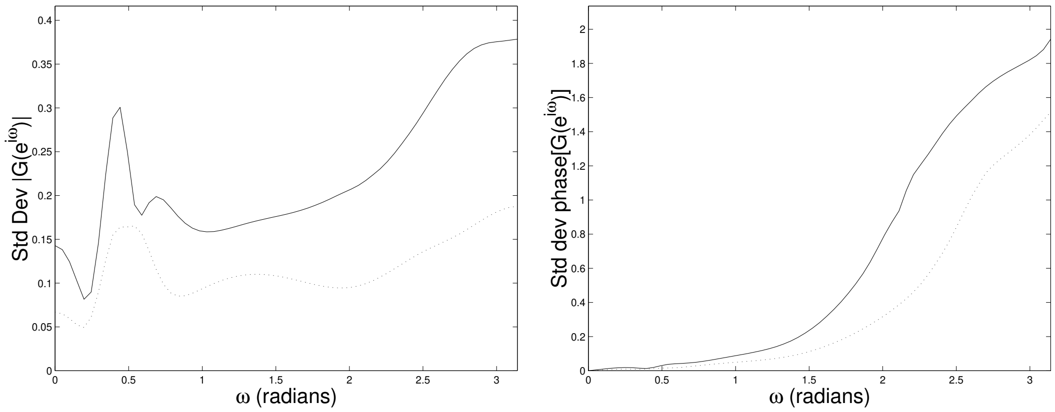
Since the non-parametric estimates of the transfer function is better for larger SNR of the signal, it is reasonable to make the choice of W_k in (4.44) while estimating $\boldsymbol{\theta}$ as (4.29). Estimating $\boldsymbol{\theta}_{fp}$ is a computationally expensive operation but the corresponding improvement in the estimate of the transfer function is also impressive as shown in Figure 4.3 compared to Figure 4.2.

The noise sensitivity of the algorithms has been examined by further Monte-Carlo simulations at lower noise level. The variance of the input noise λ_u has been taken as 0.037, while the output noise variance is set to 2.37. The SNR at both the input and the output were 20 dB approximately. The value of L has been taken equal to 20. The standard deviations of the FNP estimate of the transfer function obtained from 50 Monte-Carlo simulations are compared with that obtained for 15 dB SNR in Figure 4.6. We can notice the significant improvement in the estimation accuracy with an increase in SNR.



(a) Magnitude (b) Phase

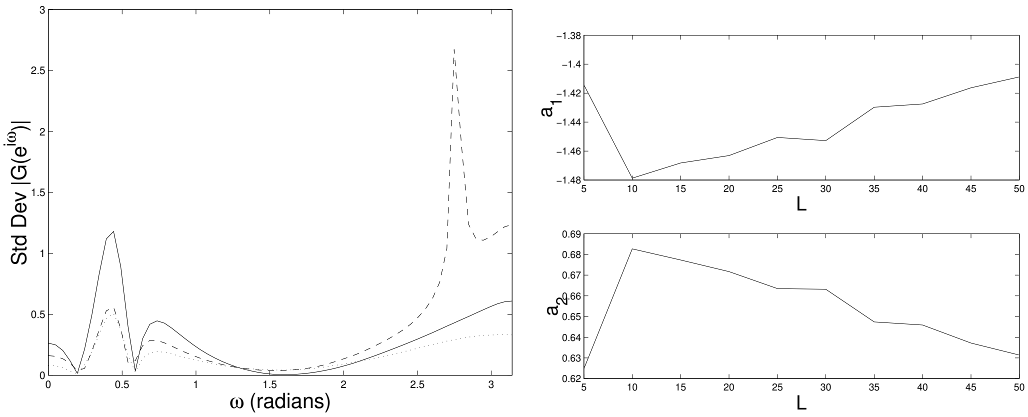
Figure 4.5: Transfer function estimates at 15 dB SNR at the input and output using periodogram approach. The mean value from Monte Carlo simulations (solid line), the mean $\pm$ standard deviation (dotted line) and the true value (dashed line).



(a) S.D of Magnitude (b) S.D of Phase

Figure 4.6: S.D of the NFP Transfer function estimates at 15 dB (solid lines) and 20 dB (dotted lines) SNR at the input and output.

In order to examine the effect of the AR model order L in (4.4), we have repeated the simulations with a different value of L with $N = 1000$ and SNR$= 15$ dB. In Figure 4.7(a), we have shown the absolute deviation of the



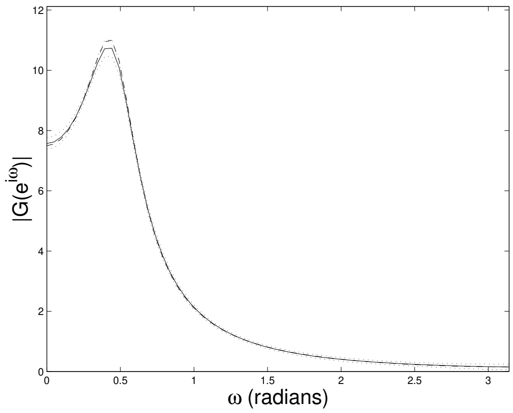
(a) Absolute deviation of FNP magnitude estimate

(b) PARMA estimates of a_1 and a_2

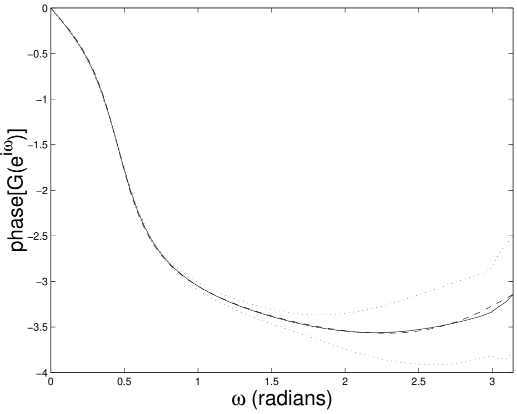
Figure 4.7: Effect of L. (a) The absolute deviation of the sample mean of the FNP magnitude function from the true magnitude function as a function of frequency for $L = 6$ (dashed line), $L = 10$ (dotted line) and $L = 25$ (solid line). (b) PARMA estimates of a_1 and a_2 as functions of L. (Recall that the true values are $a_1 = -1.5$ and $a_2 = 0.7$.)

sample mean of the magnitude of the FNP transfer function from the corresponding true values as a function of frequency for three different values of L: 6, 10 and 25. As we can see, the bias is minimum for $L = 10$. Similar observations can be made from Figure 4.7(b), where we have shown the sample mean of the PI estimate of a_1 and a_2 as a function of L. If L is too small, the degradation of the estimation performance can be attributed to the fact that the AR model is under-parameterized and the resulting $\hat{\boldsymbol{\epsilon}}(t)$ is coloured. On the contrary, the performance is degraded with the increase in L, because the model is practically overparameterised and we need to estimate too many AR coefficients. One way to chose a reasonably good value of L is to apply a whiteness test, see for example (Söderström and Stoica 1989), on the estimated innovation $\hat{\boldsymbol{\epsilon}}(t)$. For the system described in (4.40) - (4.43), a good choice of L for parametric estimation is 10.

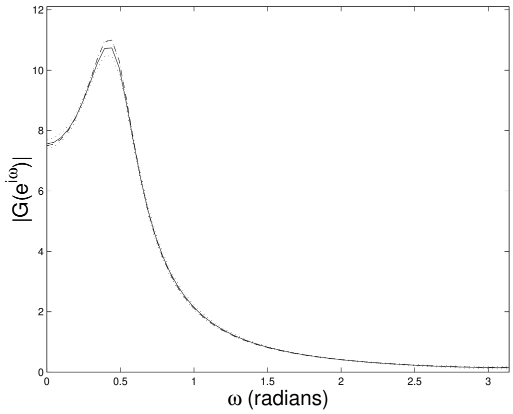
The impact of the true input spectrum will be examined next. In Figure 4.8, we have shown the parametric and the non-parametric transfer function estimates for an input signal having a more flat spectrum. Here the polynomials $F(q^{-1})$ and $D(q^{-1})$



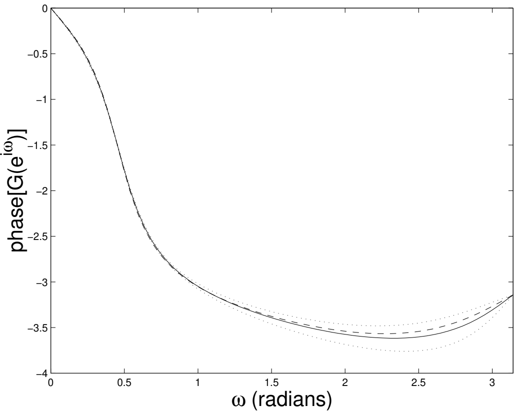
(a) Magnitude (FNP)



(b) Phase (FNP)



(c) Magnitude (FP)



(d) Phase (FP)

Figure 4.8: FNP and FP transfer function estimates at 15 dB input and output SNR with input signal having more flat spectrum and $L = 20$. The mean from Monte Carlo simulations (solid line), the mean $\pm$ standard deviation (dotted line) and the true value (dashed line).

are given by

$$F(q^{-1}) = 1 + 0.1q^{-1}, \tag{4.45}$$

$$D(q^{-1}) = 1 - 0.2q^{-1}. \tag{4.46}$$

The system under consideration remains unchanged, i.e. the polynomials $A(q^{-1})$ and $B(q^{-1})$ are given by (4.40) and (4.41) respectively. The input

	True Values	Mean (PARMA)	Mean (FP)	Std Dev (PARMA)	Std Dev (FP)	CRB
a_1	-1.5	-1.45	-1.44	0.031	0.032	0.027
a_2	0.7	0.66	0.65	0.021	0.024	0.023
b_1	1	0.93	0.96	0.84	0.087	0.135
b_2	0.5	0.70	0.67	0.71	0.160	0.156
c_1	0.2	0.33	-	0.434	-	0.182
d_1	-0.8	-0.80	-	0.022	-	0.031
λ_u	0.11	0.19	0.12	0.393	0.040	0.343
λ_y	7.5	7.6	6.2	0.434	5.53	0.012

Table 4.1: Parametric identification results from 50 Monte Carlo Simulations. Input and output SNR = 15 dB. $L = 20$ and $N = 1000$.

noise variance has been taken as 0.01, while the output noise variance is set to 0.3 in order to maintain the input and output SNR at 20 dB. The value of L has been taken equal to 20. Comparing Figure 4.8 with Figures 4.2 and 4.3 we can notice a significant improvement in both the parametric and the non-parametric transfer function estimates particularly in the high frequency region when the spectrum of the input signal is more flat.

Finally, in Table 4.1, we have summarized the parametric estimation results for FP and PARMA method. The value of L was taken as 20, while the number of data point N is 1000 in each simulation. The output and input SNR are maintained at 15 dB. The standard deviation obtained has been compared with the Cramer-Rao bound (Karlsson *et al.* 2000).

4.5 Conclusions

Two parametric algorithms and one non-parametric algorithm based on a frequency domain approach for dynamic errors-in-variables system are proposed. The algorithms provide reasonably good estimates with low computational cost. The PARMA method is computationally economical, because all the operations involved with it are linear. The non-parametric FNP method given by (4.24) is also fast because of the same reason. It is though sensitive to noise. Hence it is sometimes required to modify the non-parametric estimate using another parametric method FP given by (4.29). This modification involves non-linear optimization and is computationally expensive.

Appendix A

Appendix to Chapter 1

A.1 Gradient of the loss function (1.11)

We write the loss function (1.13) as

$$V_e(\boldsymbol{\theta}_e) = V_1(\boldsymbol{\theta}_e) + V_2(\boldsymbol{\theta}_e), \tag{A.1}$$

where

$$\begin{aligned} V_1(\boldsymbol{\theta}_e) &= E\left[\frac{\hat{A}(q^{-1})B(q^{-1}) - A(q^{-1})\hat{B}(q^{-1})}{A(q^{-1})\hat{C}(q^{-1})}u_0(t)\right]^2, & \text{(A.2)}\\ V_2(\boldsymbol{\theta}_e) &= E\left[\frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})}\tilde{y}(t)\right]^2 + E\left[\frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})}\tilde{u}(t)\right]^2. & \text{(A.3)}\end{aligned}$$

As $V_1(\boldsymbol{\theta}_e) \geq V_1(\boldsymbol{\theta}_{e0}) = 0$, it follows that $V_1(\boldsymbol{\theta}_e)$ has a minimum for the true parameter vector $\boldsymbol{\theta}_{e0}$, and hence

$$V_1'(\boldsymbol{\theta}_{e0}) = 0. \tag{A.4}$$

We then focus on evaluating the gradient of $V_2(\boldsymbol{\theta}_e)$ at the true parameter vector. Let j be an integer satisfying $1 \leq j \leq n$. Then

$$
\begin{aligned}
\frac{\partial V_2(\boldsymbol{\theta}_e)}{\partial a_j} &= 2E\left[\frac{A(q^{-1})}{C(q^{-1})}\tilde{y}(t)\right]\left[\frac{q^{-j}}{C(q^{-1})}\tilde{y}(t)\right], && \text{(A.5)}\\
\frac{\partial V_2(\boldsymbol{\theta}_e)}{\partial b_j} &= 2E\left[\frac{B(q^{-1})}{C(q^{-1})}\tilde{u}(t)\right]\left[\frac{q^{-j}}{C(q^{-1})}\tilde{u}(t)\right], && \text{(A.6)}\\
\frac{\partial V_2(\boldsymbol{\theta}_e)}{\partial c_j} &= -2E\left[\frac{A(q^{-1})}{C(q^{-1})}\tilde{y}(t)\right]\left[\frac{q^{-j}A(q^{-1})}{C^2(q^{-1})}\tilde{y}(t)\right] - 2E\left[\frac{B(q^{-1})}{C(q^{-1})}\tilde{u}(t)\right]\left[\frac{q^{-j}B(q^{-1})}{C^2(q^{-1})}\tilde{u}(t)\right]\\
&= -2\frac{1}{2\pi i}\oint\left[\frac{A(z)}{C(z)}\frac{z^{-j}A(z^{-1})}{C^2(z^{-1})}\lambda_y + \frac{B(z)}{C(z)}\frac{z^{-j}B(z^{-1})}{C^2(z^{-1})}\lambda_u\right]\frac{dz}{z}\\
&= -2\frac{1}{2\pi i}\oint\frac{z^{-j}}{C(z^{-1})}\frac{A(z)A(z^{-1})\lambda_y + B(z)B(z^{-1})\lambda_u}{C(z)C(z^{-1})}\frac{dz}{z}\\
&= -2\frac{1}{2\pi i}\oint\frac{z^{-j}}{C(z^{-1})}\lambda\frac{dz}{z} = \frac{2\lambda}{2\pi i}\oint\frac{z^{j-1}}{C(z)}dz = 0. && \text{(A.7)}
\end{aligned}
$$

The last equality in (A.7) follows as the integral has no poles inside the unit circle. On the other hand, the elements in (A.5), (A.6) are in general nonzero.

A.2 The concentrated loss function

Let us denote

$$
\hat{R}_y = \frac{1}{N}\sum_{t=1}^{N} y^2(t), \qquad \hat{R}_u = \frac{1}{N}\sum_{t=1}^{N} u^2(t), \qquad \hat{R}_{yu} = \frac{1}{N}\sum_{t=1}^{N} y(t)u(t)\ . \tag{A.8}
$$

Then inserting (1.18) into (1.17) and after some straightforward algebra the concentrated loss function in (1.19) reduces to

$$
V_i(\boldsymbol{\theta}_i) = \frac{1}{(b^2\lambda_u + \lambda_y)}\left(\hat{R}_y + b^2\hat{R}_u - 2b\hat{R}_{yu}\right) + \log\left(\lambda_y\lambda_u\right). \tag{A.9}
$$

The minimum point of $V_i(\boldsymbol{\theta})$ will obviously be some function of the covariance elements in (A.8). This is alarming from a consistency point of view, as for the asymptotic case ($N \to \infty$) we have

$$
\begin{aligned}
R_y &= Ey^2(t) = b_0^2 R_{u_0} + \lambda_y^0\\
R_{yu} &= b_0 R_{u_0}\\
R_u &= R_{u_0} + \lambda_u^0\ .
\end{aligned} \tag{A.10}
$$

These covariance elements depend not only on the three true values of the $\boldsymbol{\theta}_i$ elements: b_0, λ_u^0, λ_y^0, but also on R_{u_0}. Expressed differently, (A.10) as a system of equations with b_0, λ_u^0, λ_y^0, R_{u_0} as unknowns will lead to an under determined system.

To be more specific, setting the gradient of $V_i(\boldsymbol{\theta}_i)$ to zero gives:

$$\begin{aligned}
V_b: \quad & (b^2\lambda_u + \lambda_y)(2b\hat{R}_u - 2\hat{R}_{yu}) - 2b\lambda_u(\hat{R}_y + b^2\hat{R}_u - 2b\hat{R}_{yu}) = 0 \\
& b^2\lambda_u\hat{R}_{yu} + b(\lambda_y\hat{R}_u - \lambda_u\hat{R}_y) - \lambda_y\hat{R}_{yu} = 0 \ . && \text{(A.11)} \\
V_{\lambda_u}: \quad & -\frac{b^2}{(b^2\lambda_u + \lambda_y)^2}(\hat{R}_y + b^2\hat{R}_u - 2b\hat{R}_{yu}) + \frac{1}{\lambda_u} = 0 \ . && \text{(A.12)} \\
V_{\lambda_y}: \quad & -\frac{1}{(b^2\lambda_u + \lambda_y)^2}(\hat{R}_y + b^2\hat{R}_u - 2b\hat{R}_{yu}) + \frac{1}{\lambda_y} = 0 \ . && \text{(A.13)}
\end{aligned}$$

Dividing (A.12) by (A.13) gives

$$\lambda_y = b^2\lambda_u \ . \qquad \text{(A.14)}$$

Inserting (A.14) into (A.11) gives

$$b^2\lambda_u\hat{R}_{yu} + b\lambda_u(b^2\hat{R}_u - \hat{R}_y) - b^2\lambda_u\hat{R}_{yu} = 0$$

leading to (except for the degenerate solutions $b = 0$ and $\lambda_u = 0$)

$$b = \pm\sqrt{\hat{R}_y/\hat{R}_u} \ . \qquad \text{(A.15)}$$

This clearly shows the lack of uniqueness, and hence of loss of identifiability. A comparison of (A.15) with (A.10) reveals that the loss function is not minimized at the true parameters even when $N \to \infty$.

Appendix B

Appendix to Chapter 2

B.1 Proof of theorem 2.1

In this section, we denote $R_{\boldsymbol{\zeta}\boldsymbol{\psi}}$ by R and $R_{\boldsymbol{\zeta}y}$ by r respectively. The outline of the proof is as follows.

- We will first show that

$$\hat{R}^T W \hat{R}(\hat{\boldsymbol{\theta}}_{iv} - \boldsymbol{\theta}) = \hat{R}^T W \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{\zeta}(t) x(t), \tag{B.1}$$

  where $x(t)$ is defined in (1.7).

- Next, we will prove

$$\hat{R}^T W \hat{R}(\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta}) = \hat{R}^T W \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{\zeta}(t) x(t). \tag{B.2}$$

- From (B.1) and (B.2), it is clear that $\boldsymbol{\theta}_{iv}$ and $\boldsymbol{\theta}_{tls}$ have the same asymptotic distribution. It follows from ergodicity results, see (Söderström and Stoica 1989), that

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{iv} - \boldsymbol{\theta}) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P), \tag{B.3}$$

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta}) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P), \tag{B.4}$$

$$P = (R^T W R)^{-1} R^T W \mathcal{P} W R (R^T W R)^{-1}, \tag{B.5}$$

  where the newly introduced matrix $\mathcal{P}$ is given by

$$\mathcal{P} = \lim_{N\to\infty} E\left\{ \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} \boldsymbol{\zeta}(t) x(t) x(s) \boldsymbol{\zeta}^T(s) \right\}. \tag{B.6}$$

- As a final step we will show that

$$\mathcal{P} = \lambda \text{cov}\left\{C(q^{-1})\boldsymbol{\zeta}(t)\right\}, \tag{B.7}$$

where $C(q^{-1})$ and λ are defined by (1.8).

To show (B.7), consider

$$\begin{aligned}
\hat{R}^T W \hat{R}(\hat{\boldsymbol{\theta}}_{iv} - \boldsymbol{\theta}) &= \hat{R}^T W \left\{\hat{r} - \hat{R}\boldsymbol{\theta}\right\} = \hat{R}^T W \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\zeta}(t)\left\{y(t) - \boldsymbol{\psi}^T(t)\boldsymbol{\theta}\right\} \\
&= -\hat{R}^T W \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\zeta}(t)\boldsymbol{\phi}^T\boldsymbol{\theta} = \hat{R}^T W \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\zeta}(t)x(t).
\end{aligned}$$

To show (B.2) is not as straightforward as (B.1). Let us denote $S = \Gamma R_{\boldsymbol{\zeta}\boldsymbol{\phi}}$ and $\hat{S} = \Gamma \hat{R}_{\boldsymbol{\zeta}\boldsymbol{\phi}}$. Recall that the unit singular vector $\mathbf{w}$ of S corresponding to its least singular value is related to the parameter vector $\boldsymbol{\theta}$ as

$$J\mathbf{w} = \frac{\boldsymbol{\theta}}{\sqrt{1 + \boldsymbol{\theta}^T\boldsymbol{\theta}}}, \tag{B.8}$$

where $J = \begin{bmatrix} 0 & I_{2n} \end{bmatrix}$, a $(2n)|(2n{+}1)$ dimensional matrix. We derive the perturbation in $\boldsymbol{\theta}$ due to a small perturbation in $\mathbf{w}$ by Taylor's series expansion. Neglecting the higher order terms we can relate $\hat{\mathbf{w}} - \mathbf{w}$ to $\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta}$

$$\begin{aligned}
J(\hat{\mathbf{w}} - \mathbf{w}) &= \frac{\partial}{\partial \boldsymbol{\theta}}\left[\frac{\boldsymbol{\theta}}{\sqrt{1 + \boldsymbol{\theta}^T\boldsymbol{\theta}}}\right]\left[\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta}\right] \\
&= \frac{1}{\sqrt{1 + \boldsymbol{\theta}^T\boldsymbol{\theta}}}\left[I_{2n} - \frac{\boldsymbol{\theta}\boldsymbol{\theta}^T}{1 + \boldsymbol{\theta}^T\boldsymbol{\theta}}\right]\left[\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta}\right] \\
\Rightarrow \hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta} &= \sqrt{1 + \boldsymbol{\theta}^T\boldsymbol{\theta}}\left(I_{2n} + \boldsymbol{\theta}\boldsymbol{\theta}^T\right) J\left(\hat{\mathbf{w}} - \mathbf{w}\right),
\end{aligned} \tag{B.9}$$

where in the last equality we have used the matrix inversion lemma. This step is justified only for large N, when $\hat{\boldsymbol{\theta}}_{tls}$ would be sufficiently close to $\boldsymbol{\theta}$. Again $\mathbf{w}$ is unit eigen vector of $S^T S$ and the associated eigen value is zero, i.e. $S^T S$ is symmetric and singular matrix with the eigen decomposition

$$S^T S = 0 \times \mathbf{w}\mathbf{w}^T + UDU^T \tag{B.10}$$

where $D^{1/2}$ is $2n|2n$ diagonal matrix composed of the non-zero singular values of S and the columns of U are the associated right singular vectors. For large

N, $\hat{S}$ deviates marginally from S. For small perturbation of S^TS we have the following perturbation expansion of $\mathbf{w}$, (Stewart 1974, Sun 1998):

$$
\begin{aligned}
\hat{\mathbf{w}} - \mathbf{w} &= -UD^{-1}U^T\left(\hat{S}^T\hat{S} - S^TS\right)\mathbf{w} \\
\Rightarrow \hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta} &= -\sqrt{1+\boldsymbol{\theta}^T\boldsymbol{\theta}}\left(I_{2n} + \boldsymbol{\theta}\boldsymbol{\theta}^T\right)JUD^{-1}U^T\left(\hat{S}^T\hat{S} - S^TS\right)\mathbf{w} \\
&= -\left(I_{2n} + \boldsymbol{\theta}\boldsymbol{\theta}^T\right)JUD^{-1}U^T\left(\hat{S}^T\hat{S} - S^TS\right)\overline{\boldsymbol{\theta}}. \qquad \text{(B.11)}
\end{aligned}
$$

Note that, we have used $\mathbf{w}\sqrt{1+\boldsymbol{\theta}^T\boldsymbol{\theta}} = \overline{\boldsymbol{\theta}}$, (see(1.33)). Since $S\overline{\boldsymbol{\theta}} = 0$, we have

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_{tls} - \boldsymbol{\theta} &\approx -\left(I_{2n} + \boldsymbol{\theta}\boldsymbol{\theta}^T\right)JUD^{-1}U^T\hat{S}^T\left(\hat{S} - S\right)\overline{\boldsymbol{\theta}} \\
&\approx -\left(I_{2n} + \boldsymbol{\theta}\boldsymbol{\theta}^T\right)JUD^{-1}U^TS^T\Gamma\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\zeta}(t)\boldsymbol{\phi}^T(t)\overline{\boldsymbol{\theta}} \\
&= \left(I_{2n} + \boldsymbol{\theta}\boldsymbol{\theta}^T\right)JUD^{-1}U^TS^T\Gamma\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\zeta}(t)x(t), \qquad \text{(B.12)}
\end{aligned}
$$

where we have used (1.35) in the last equality. Let S has the singular value decomposition $S = XD^{1/2}U^T$ (the term corresponding to the zero singular value is omitted), Then we have

$$
\begin{aligned}
& JUD^{-1}U^TS^T = JUD^{-1}U^TUD^{1/2}X^T = JUD^{-1/2}X^T = JS^{\dagger} \\
=\ & J\left(\Gamma R_{\boldsymbol{\zeta}\boldsymbol{\phi}}\right)^{\dagger} = J\left(\Gamma\begin{bmatrix} R\boldsymbol{\theta} & R \end{bmatrix}\right)^{\dagger} = J\left(\Gamma R\begin{bmatrix} \boldsymbol{\theta} & I_{2n} \end{bmatrix}\right)^{\dagger} \\
=\ & J\begin{bmatrix} \boldsymbol{\theta} & I_{2n} \end{bmatrix}^{\dagger}(\Gamma R)^{\dagger} \\
=\ & \begin{bmatrix} 0 & I_{2n} \end{bmatrix}\begin{bmatrix} \boldsymbol{\theta}^T \\ I_{2n} \end{bmatrix}\left(\boldsymbol{\theta}\boldsymbol{\theta}^T + I_{2n}\right)^{-1}\left(R^T\Gamma^T\Gamma R\right)^{-1}R^T\Gamma^T \\
=\ & \left(\boldsymbol{\theta}\boldsymbol{\theta}^T + I_{2n}\right)^{-1}\left(R^TWR\right)^{-1}R^T\Gamma^T. \qquad \text{(B.13)}
\end{aligned}
$$

Thus (B.2) follows straightaway from (B.12) and (B.13). It remains to show (B.7). Now, (B.6) can be written as

$$
\begin{aligned}
\mathcal{P} = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}E\Big\{&\boldsymbol{\zeta}_0(t)x(t)x(s)\boldsymbol{\zeta}_0^T(s) + \boldsymbol{\zeta}_0(t)x(t)x(s)\tilde{\boldsymbol{\zeta}}^T(s) \\
&+ \tilde{\boldsymbol{\zeta}}(t)x(t)x(s)\boldsymbol{\zeta}_0^T(s) + \tilde{\boldsymbol{\zeta}}(t)x(t)x(s)\tilde{\boldsymbol{\zeta}}^T(s)\Big\}. \qquad \text{(B.14)}
\end{aligned}
$$

Since $\boldsymbol{\zeta}_0(t)$ is independent of $\tilde{\boldsymbol{\zeta}}(s)$ and $x(s)$ for all possible combinations of t and s, the second and the third term in the right hand side of (B.14) will vanish. It follows from the derivation of the asymptotic distribution of IV estimates (noise-free input case) that (Söderström and Stoica 1989), the first term, denoted by T_1, is given by

$$T_1 = \lambda \mathrm{cov}\left\{C(q^{-1})\boldsymbol{\zeta}_0(t)\right\}. \tag{B.15}$$

Denote the last term of (B.14) by T_4. Then by the property of jointly Gaussian random vectors

$$\begin{aligned} T_4 &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\Big[E\left\{\tilde{\boldsymbol{\zeta}}(t)x(t)\right\}E\left\{x(s)\tilde{\boldsymbol{\zeta}}^T(s)\right\}+ \\ &\quad E\left\{\tilde{\boldsymbol{\zeta}}(t)x(s)\right\}E\left\{x(t)\tilde{\boldsymbol{\zeta}}^T(s)\right\}+E\left\{x(t)x(s)\right\}E\left\{\tilde{\boldsymbol{\zeta}}(t)\tilde{\boldsymbol{\zeta}}^T(s)\right\}\Big] \end{aligned} \tag{B.16}$$

Note that, by construction of instruments, $\tilde{\boldsymbol{\zeta}}(t)$and $x(t)$ are uncorrelated and $\boldsymbol{\zeta}(t)$ is constructed from delayed input and output values. Thus the first two terms of (B.16) will vanish. Let $\{c_i\}_{i=1}^n$ denote the impulse response sequence of $C(q^{-1})$ in (1.9). Then for the remaining term in (B.16), using (1.9) and the fact that $v(t)$ is a white noise with variance λ

$$\begin{aligned} T_4 &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}E\left\{\tilde{\boldsymbol{\zeta}}(t)\tilde{\boldsymbol{\zeta}}^T(s)\right\}E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jv(t-i)v(s-j)\right\} \\ &= \lim_{N\to\infty}\frac{\lambda}{N}E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\sum_{t=1}^{N}\sum_{s=1}^{N}\tilde{\boldsymbol{\zeta}}(t)\tilde{\boldsymbol{\zeta}}^T(s)\delta_{t-s-i+j,0}\right\}, \end{aligned} \tag{B.17}$$

where we use $\delta_{i,j}$ to denote Kronecker delta function. By the stationarity of $\boldsymbol{\zeta}(t)$ we get

$$\begin{aligned} T_4 &= \lim_{N\to\infty}\frac{\lambda}{N}E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\sum_{t=1}^{N}\tilde{\boldsymbol{\zeta}}(t)\tilde{\boldsymbol{\zeta}}^T(t-i+j)\right\} \\ &= \lambda\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jE\left\{\tilde{\boldsymbol{\zeta}}(t-i)\tilde{\boldsymbol{\zeta}}^T(t-j)\right\} = \lambda\mathrm{cov}\left\{C(q^{-1})\tilde{\boldsymbol{\zeta}}(t)\right\} \end{aligned} \tag{B.18}$$

Now, by (B.14), (B.15) and (B.18) and due to the independence of $\boldsymbol{\zeta}_0(t)$ and $\tilde{\boldsymbol{\zeta}}(t)$ we have (B.7). This observation completes the proof.

B.2 Proof of theorem 2.2

Let $\tilde{P} = \hat{P} - P$. Then it can be shown that, see Appendix B9 of (Söderström and Stoica 1989),

$$E\tilde{P}_{ij}\tilde{P}_{kl} = \frac{1}{m-1}[P_{ik}P_{jl} + P_{il}P_{jk}]. \tag{B.19}$$

But

$$V = tr\left(\tilde{P}^T Q \tilde{P}\right) = tr\left(Q\tilde{P}\tilde{P}^T\right). \tag{B.20}$$

Now

$$\begin{aligned}
E\left(\tilde{P}\tilde{P}^T\right)_{lk} &= \sum_{j=1}^{2n} E\left(\tilde{P}_{lj}\tilde{P}_{jk}\right) \\
&= \sum_{j=1}^{2n} \frac{1}{m-1}[P_{lk}P_{jj} + P_{lj}P_{jk}] \\
&= \frac{1}{m-1}[tr(P)P_{lk} + (P^2)_{lk}] \\
\Rightarrow E\left(\tilde{P}\tilde{P}^T\right) &= \frac{1}{m-1}[tr(P)P + (P^2)]. \qquad \text{(B.21)}
\end{aligned}$$

Hence by (B.19)

$$\begin{aligned}
EV &= \frac{1}{m-1} tr[Q\{tr(P)P + P^2\}] \\
&= \frac{1}{m-1}[tr(PQP) + tr(P)tr(QP)]. \qquad \text{(B.22)}
\end{aligned}$$

Appendix C

Appendix to Chapter 3

C.1 Proof of theorem 3.1

We need the following proposition.

Proposition: Consider the stationary vector processes $\boldsymbol{\mu}(t)$, $\boldsymbol{\mu}_1(t)$ and $\boldsymbol{\mu}_2(t)$ such that

$$\begin{aligned}
\boldsymbol{\mu}_i(t) &= \mathcal{G}_i(q^{-1})\boldsymbol{\mu}(t), \quad i = 1, 2; && \text{(C.1)} \\
\mathbf{r}_{\boldsymbol{\mu}}(\tau) &= E\boldsymbol{\mu}(t)\boldsymbol{\mu}^T(t-\tau), && \text{(C.2)} \\
\mathbf{r}_{12}(\tau) &= E\boldsymbol{\mu}_1(t)\boldsymbol{\mu}_2^T(t-\tau), && \text{(C.3)}
\end{aligned}$$

where $\mathcal{G}_1(q^{-1})$ and $\mathcal{G}_2(q^{-1})$ are FIR filters of orders l and k, respectively. Let us assume

$$q^{-l}\mathcal{G}_1(q)\mathcal{G}_2(q^{-1}) = \mathcal{G}(q^{-1}) = \sum_{i=0}^{l+k} g_i q^{-i}. \tag{C.4}$$

Then

$$\mathbf{r}_{12}(\tau) = \sum_{i=0}^{l+k} g_i \mathbf{r}_{\boldsymbol{\mu}}(l - \tau - i). \tag{C.5}$$

Proof: Follows by straightforward calculation.

As a special case of (C.1), if

$$\begin{aligned}
\boldsymbol{\mu}(t) = \tilde{\mathbf{z}}(t) &= \begin{bmatrix} \tilde{y}(t) & \tilde{u}(t) \end{bmatrix}^T, && \text{(C.6)} \\
\boldsymbol{\mu}_i(t) = \tilde{\mathbf{z}}_i(t) &= \begin{bmatrix} \tilde{y}_i(t) & \tilde{u}_i(t) \end{bmatrix}^T, i = 1, 2; && \text{(C.7)}
\end{aligned}$$

then we get from (C.5) and (3.10)

$$\mathbf{r}_{\boldsymbol{\mu}}(\tau) = \begin{cases} R_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}, & \text{if } \tau = 0 \\ 0, & \text{if } \tau \neq 0 \end{cases}, \tag{C.8}$$

$$\mathbf{r}_{12}(\tau) = \begin{cases} R_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} h_{m-\tau}, & \text{if } |\tau| \leq m \\ 0, & \text{if } |\tau| > m \end{cases}. \tag{C.9}$$

Note that, in this specific case, $R_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$ is a diagonal matrix, but it need not be diagonal for (C.9) to hold. Since $R_{\tilde{\boldsymbol{\phi}}_1\tilde{\boldsymbol{\phi}}_2}$ in (3.7) is composed of the elements of $\{\mathbf{r}_{12}(\tau)\}_{\tau=-n}^{n}$, it is sufficient to have (3.10) satisfied in order (3.7) to hold.

C.2 Proof of theorem 3.2

Form (3.6), we get after some steps

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta} &= \hat{R}^{\dagger}_{\boldsymbol{\phi}_1\boldsymbol{\psi}_2} \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{\phi}_1(t) \left[\tilde{y}_2(t) - \tilde{\boldsymbol{\psi}}_2^T(t)\boldsymbol{\theta} \right] \\ &= \hat{R}^{\dagger}_{\boldsymbol{\phi}_1\boldsymbol{\psi}_2} \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{\phi}_1(t) x_2(t). \end{aligned} \tag{C.10}$$

Hence it follows from ergodicity results, see (Söderström and Stoica 1989), that $\hat{\boldsymbol{\theta}}_1$ is asymptotically Gaussian distributed as,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_1), \tag{C.11}$$

where

$$P_1 = R^{\dagger}_{\boldsymbol{\phi}_1\boldsymbol{\psi}_2} Q R^{\dagger T}_{\boldsymbol{\phi}_1\boldsymbol{\psi}_2} \tag{C.12}$$

and

$$Q = \lim_{N\to\infty} E\left\{ \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} \boldsymbol{\phi}_1(t) x_2(t) x_2(s) \boldsymbol{\phi}_1^T(s) \right\} \tag{C.13}$$

It remains to compute Q. In this aim, we decompose $\boldsymbol{\phi}_1(t)$ in (C.13) into its noisy and noise-free parts and retain only the non-zero terms to get

$$Q = Q_a + Q_b, \tag{C.14}$$

where

$$
\begin{aligned}
Q_a &= \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\boldsymbol{\phi}_{10}(t)x_2(t)x_2(s)\boldsymbol{\phi}_{10}^T(s)\right\} \\
&= \lambda\mathrm{cov}\left[\mathcal{C}(q^{-1})\boldsymbol{\phi}_0(t)\right]; && \text{(C.15)} \\
Q_b &= \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\tilde{\boldsymbol{\phi}}_1(t)x_2(t)x_2(s)\tilde{\boldsymbol{\phi}}_1^T(s)\right\}. && \text{(C.16)}
\end{aligned}
$$

Note that, in (C.15) we have used (3.12). For the detailed derivation of (C.15), we refer the reader to (Söderström and Stoica 1983). To simplify Q_b, let us apply the property of jointly Gaussian distributed random vectors to get

$$
Q_b = Q_c + Q_d + Q_e, \tag{C.17}
$$

where

$$
\begin{aligned}
Q_c &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}E\left\{\tilde{\boldsymbol{\phi}}_1(t)x_2(t)\right\}E\left\{x_2(s)\tilde{\boldsymbol{\phi}}_1^T(s)\right\} \\
&= 0, && \text{(C.18)} \\
Q_d &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}E\left\{\tilde{\boldsymbol{\phi}}_1(t)\tilde{\boldsymbol{\phi}}_1^T(s)\right\}E\left\{x_2(t)x_2(s)\right\} \\
&= \lambda\mathrm{cov}\left[\mathcal{C}(q^{-1})\tilde{\boldsymbol{\phi}}(t)\right], && \text{(C.19)} \\
Q_e &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}E\left\{\tilde{\boldsymbol{\phi}}_1(t)x_2(s)\right\}E\left\{x_2(t)\tilde{\boldsymbol{\phi}}_1^T(s)\right\}. && \text{(C.20)}
\end{aligned}
$$

Note that, (C.18) follows from (3.7) and in (C.19) we have used (3.12). The derivation of (C.19) is similar to that of (C.15) and (B.17-B.18) in Appendix B.2, can also be found in (Söderström and Mahata 2002). Combining (C.14)-(C.20) we get

$$
Q = Q_e + Q_f, \tag{C.21}
$$

where

$$
Q_f = \lambda\mathrm{cov}\left[\mathcal{C}(q^{-1})\boldsymbol{\phi}(t)\right]. \tag{C.22}
$$

To examine Q_e closely, let us first define the polynomials

$$\begin{aligned} L_A(q^{-1}) &= A(q^{-1})F_2(q^{-1}), &\text{(C.23)}\\ L_B(q^{-1}) &= B(q^{-1})F_2(q^{-1}), &\text{(C.24)}\\ \sum_{i=0}^{2m+n} \Gamma_i q^{-i} &= A(q^{-1})H(q^{-1}), &\text{(C.25)}\\ \sum_{i=0}^{2m+n} \Delta_i q^{-i} &= B(q^{-1})H(q^{-1}). &\text{(C.26)}\end{aligned}$$

Note also that, $\tilde{\boldsymbol{\phi}}_1(t)$ is composed of the delayed and filtered measurement noise sequences $\tilde{y}_1(t)$ and $\tilde{u}_1(t)$. In the aim of evaluating $E\left\{\tilde{\boldsymbol{\phi}}_1(t)x_2(s)\right\}$, let us examine

$$\begin{aligned} & E\tilde{y}_1(t)x_2(t-\tau) \\ =\ & E\left\{F_1(q^{-1})\tilde{y}(t)\right\}\left\{L_A(q^{-1})\tilde{y}(t-\tau)\right\} \\ =\ & \lambda_y \Gamma_{m-\tau}. \end{aligned} \tag{C.27}$$

Note that, in the last equality we have used the proposition in the Appendix C.1, (C.25) and (3.10). Similarly by (C.5), (C.26) and (3.10) we have

$$E\tilde{u}_1(t)x_2(t-\tau) = \lambda_u \Delta_{m-\tau}. \tag{C.28}$$

Note that Q_e is a $(2n+1)\times(2n+1)$ matrix, which can be slitted into block matrices

$$Q_e = \begin{bmatrix} Q^{(11)} & Q^{(12)} \\ Q^{(21)} & Q^{(22)} \end{bmatrix}, \tag{C.29}$$

where $Q^{(11)}$ is a $(n+1)\times(n+1)$ matrix and $Q^{(22)}$ is a $n\times n$ matrix. Consider $Q^{(11)}_{\mu\nu}$, which is the element at μth row and νth column of $Q^{(11)}$. From (C.27) for $1\le\mu\le n+1$ and $1\le\nu\le n+1$ we get

$$\begin{aligned} Q^{(11)}_{\mu\nu} &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N} E\left\{\tilde{y}(t-\mu+1)x_2(s)\right\}E\left\{\tilde{y}(s-\nu+1)x_2(t)\right\} \\ &= \lim_{N\to\infty}\frac{\lambda_y^2}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\Gamma_{m-t+s+\mu-1}\Gamma_{m-s+t+\nu-1} \\ &= \lambda_y^2\sum_{i=0}^{2m+n}\Gamma_i\Gamma_{2m+\mu+\nu-2-i} \\ &= \lambda_y^2\beta^{(aa)}_{2m+\mu+\nu-2}, \end{aligned} \tag{C.30}$$

where we used the convention $\Gamma_i = 0$ for $i < 0$ and $i > 2m+n$. Note that in the last equality, we have used (3.13). Similarly, for the other blocks of Q_e we can show using (C.27), (C.28), (3.14) and (3.15) that

$$
\begin{aligned}
Q^{(12)}_{\mu\nu} &= \lambda_y \lambda_u \beta^{(ab)}_{2m+\mu+\nu-1}, \quad 1 \le \mu \le n+1, \; 1 \le \nu \le n; && \text{(C.31)} \\
Q^{(21)}_{\mu\nu} &= Q^{(12)}_{\nu\mu}, \quad 1 \le \nu \le n+1, \; 1 \le \mu \le n; && \text{(C.32)} \\
Q^{(22)}_{\mu\nu} &= \lambda_u^2 \beta^{(bb)}_{2m+\mu+\nu}, \quad 1 \le \mu \le n, \; 1 \le \nu \le n. && \text{(C.33)}
\end{aligned}
$$

Thus, (C.22), (C.29)-(C.33) together define the matrix Q. This completes the derivation of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_1$. The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_2$ can be obtained from the expression of P_1 by interchanging the subscripts 1 and 2.

Bibliography

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

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

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

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

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

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

Cadzow, J. A. and O. M. Solomon (1986). Algebraic approach to system identification. *IEEE Transactions on Signal Processing* **ASSP-34**(3), 462–469.

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

Cedervall, M and P Stoica (1996). System identification from noisy measurements by using instrumental variables and subspace fitting. *Circuits, Systems and Signal Processing* **15:2**, 275–290.

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

Durbin, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika* **46**, 306–316.

Fernando, K. V. and H. Nicholson (1985). Identification of linear systems with input and output noise: the Koopmans–Levin method. *IEE Proceedings, Part D* **132**(1), 30–36.

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

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

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

Gevers, M. and V. Wertz (1984). Uniquely identifiable state-space and arma parameterizations for multivariable linear systems. *Automatica* **20**(3), 333–347.

Golub, G. H. and C. F. Van Loan (1989). *Matrix Computations, 2nd ed.* John Hopkins University Press. Baltimore, MD, USA.

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

Karlsson, E, T Söderström and P Stoica (2000). The Cramér-Rao lower bound for noisy input output systems. *Signal Processing* **80**(11), 2421–2447.

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

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

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

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

Mayne, D. Q. and F. Firoozan (1982). Linear identification of ARMA processes. *Automatica* **18**, 461–466.

Pintelon, R. and J. Schoukens (2001). *System Identification: A Frequency Domain Approach*. IEEE Press. New York.

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

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

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

Söderström, T (1979). Some methods for identification of linear systems with noisy input/output data. In: *5th IFAC/IFORS Symposium on Identification and System Parameter Estimation*. Darmstadt, Germany.

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

Söderström, T. and K. Mahata (2002). On instrumental variable and total least squares approaches for identification of noisy systems. *International Journal of Control* **75**(6), 381–389.

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

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

Söderström, T., K. Mahata and U. Soverini (2002*a*). Identification of dynamic errors–in–variables model: Approaches based on two dimensional arma modelling of the data. Submitted to Automatica.

Söderström, T., K. Mahata and U. Soverini (2002*b*). Identification of dynamic errors-in-variables model using a frequency domain Frisch scheme.. In: *15 th IFAC World Congress*. Bercelona. To appear.

Söderström, T., U. Soverini and K. Mahata (2002*c*). Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing*. To appear.

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

Stewart, G W (1974). *Introduction to Matrix Computations*. Academic Press. New York, NY, USA.

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

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

Stoica, P, T Söderström and S Van Huffel (1992). "On the equivalence of SVD and TLS based Yule-Walker approaches to frequency estimation". In: *EUSIPCO'92*. Brussels, Belgium.

Sun, Ji-guang (1998). Stability and Accuracy. Perturbation Analysis of Algebraic Eigenproblems (book manuscript). Technical Report UMINF 98-07. Umeå University, Department of Computing Science.

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

Tugnait, J. K. and Y. Ye (1995). Stochastic system identification with noisy input-output measurement using polyspectra. *IEEE Transactions on Automatic Control* **AC-40**, 670–683.

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

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

Young, P C (1970). An instrumental variable method for real-time identification of a noisy process. *Automatica* **6**, 271–287.

Young, P C (1984). *Recursive Estimation and Time-Series Analysis.* Springer. Berlin.

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

Zheng, W. X. (1999*a*). An efficient algorithm for stochastic system identification with noisy input. In: *Proc. 38th IEEE Conference on Decision and Control*. Phoenix, AZ. pp. 3657–3662.

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

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

Zheng, W. X. and C. B. Feng (1992). Identification of a class of dynamic errors–in–variables models. *International Journal of Adaptive Control and Signal Processing* **6**, 41–440.

Recent licentiate theses from the Department of Information Technology

2001-009 Johan Bengtsson: *Efficient Symbolic State Exploration of Timed Systems: Theory and Implementation*

2001-010 Johan Edlund: *A Parallel, Iterative Method of Moments and Physical Optics Hybrid Solver for Arbitrary Surfaces*

2001-011 Pär Samuelsson: *Modelling and control of activated sludge processes with nitrogen removal*

2001-012 Per Åhgren: *Teleconferencing, System Identification and Array Processing*

2001-013 Alexandre David: *Practical Verification of Real-time Systems*

2001-014 Abraham Zemui: *Fourth Order Symmetric Finite Difference Schemes for the Wave Equation*

2001-015 Stefan Söderberg: *A Parallel Block-Based PDE Solver with Space-Time Adaptivity*

2001-016 Johan Furunäs Åkesson: *Interprocess Communication Utilising Special Purpose Hardware*

2002-001 Eva Mossberg: *Higher Order Finite Difference Methods for Wave Propagation Problems*

2002-002 Anders Berglund: *On the Understanding of Computer Network Protocols*

2002-003 Emad Abd-Elrady: *Harmonic Signal Modeling Based on the Wiener Model Structure*

2002-004 Martin Nilsson: *Iterative Solution of Maxwell's Equations in Frequency Domain*

2002-005 Kaushik Mahata: *Identification of Errors in Variables Models*

UPPSALA
UNIVERSITET

University Printers, Ekonomikum, Uppsala, Sweden