



HAL
open science

Identification of continuous-time errors-in-variables models

Kaushik Mahata, Hugues Garnier

► **To cite this version:**

Kaushik Mahata, Hugues Garnier. Identification of continuous-time errors-in-variables models. *Automatica*, 2006, 42 (9), pp.1470-1490. 10.1016/j.automatica.2006.04.012 . hal-00119875

HAL Id: hal-00119875

<https://hal.science/hal-00119875>

Submitted on 12 Dec 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification of continuous-time errors-in-variables models¹

K. MAHATA[†], H. GARNIER[‡]

Abstract : A novel direct approach for identifying continuous-time linear dynamic errors-in-variables models is presented in this paper. The effects of the noise on the state-variable filter outputs are analyzed. Subsequently, a few algorithms to obtain consistent continuous-time parameter estimates in the errors-in-variables framework are derived. It is also possible to design search-free algorithms within our framework. The algorithms can be used for nonuniformly sampled data. The asymptotic distributions of the estimates are derived. The performances of the proposed algorithms are illustrated with some numerical simulation examples.

Keywords: Errors-in-variables, system identification, continuous-time models, linear estimation.

[†] Centre for Complex Dynamic Systems and Control, School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia. kaushik.mahata@newcastle.edu.au

[‡] Centre de Recherche en Automatique de Nancy, Nancy-Université, CNRS, BP 239, 54506 Vandœuvre-lès-Nancy, France. hugues.garnier@cran.uhp-nancy.fr

¹This report is related to the paper published in *Automatica*, Vol. 42 (9), pp. 1470-1490, September 2006.

1 Introduction

Consider a single-input, single-output, linear, time-invariant, continuous-time system having input $\hat{u}(t)$ and output $\hat{y}(t)$ linked by the differential equation

$$\begin{aligned} a_n \frac{d^n \hat{y}(t)}{dt^n} + \dots + a_1 \frac{d \hat{y}(t)}{dt} + a_0 \hat{y}(t) \\ = b_n \frac{d^n \hat{u}(t)}{dt^n} + \dots + b_1 \frac{d \hat{u}(t)}{dt} + b_0 \hat{u}(t). \end{aligned} \quad (1)$$

In this paper we address the problem of identifying the system parameters from sampled and noise corrupted versions of $\hat{u}(t)$ and $\hat{y}(t)$. Without any loss of generality we use the normalization² $a_0 = 1$. It is assumed that the input and the output signals are sampled at time-instants $\{t_k\}_{k=0}^N$, not necessarily uniformly spaced. The sampled signals are denoted by $\{\hat{u}(t_k); \hat{y}(t_k)\}$. It is further assumed that the measurements are noise corrupted. The observed sampled data $\{u(t_k)\}_{k=0}^N$ and $\{y(t_k)\}_{k=0}^N$ are given by

$$u(t_k) = \hat{u}(t_k) + \tilde{u}(t_k), \quad y(t_k) = \hat{y}(t_k) + \tilde{y}(t_k).$$

We assume that $\{\tilde{u}(t_k)\}_{k=0}^N$ and $\{\tilde{y}(t_k)\}_{k=0}^N$ are mutually independent, zero mean, *discrete-time* white noise sequences with variances σ_I^2 and σ_O^2 , respectively. The noise variances are assumed to be unknown. Our task is to identify the continuous-time parameters $\{a_j\}_{j=1}^n$ and $\{b_j\}_{j=0}^n$ from the noise corrupted observations of the input-output sampled data $\{u(t_k)\}_{k=0}^N$ and $\{y(t_k)\}_{k=0}^N$. Note that the degree of the numerator of the system transfer function is the same as that of the denominator. This helps us keep the notations simple. However, all the subsequent analysis in this paper can readily be generalized for strictly proper systems.

The model under consideration is often referred to as the continuous-time errors-in-variables (EIV) model. Many methods have been proposed to solve the related problem in discrete-time. The popular approaches can be classified in a few broad categories [24]; namely, the bias compensating least squares methods [3, 4, 40], prediction error and maximum likelihood methods [20, 11], instrumental variable based approaches [26, 23] and frequency-domain methods based on non-parametric noise modeling [18]. Unless we impose certain assumptions on the signal and noise models, it is well-known that the *general* EIV model is not uniquely identifiable from second order statistics [1]. This motivates the approaches based on higher order statistics [29, 30].

Most physical systems are continuous-time in nature. Thus like any other disciplines in system identification, continuous-time EIV problem is a relevant and important research topic. To our best knowledge, the case of continuous-time EIV model identification has not received appropriate attention so far. A first attempt to solve the EIV filtering problem for continuous-time models has been recently proposed in [14]. Here our object is to identify a transfer function model. In comparison with the discrete-time model identification, direct continuous-time model identification raises several technical issues. Unlike the difference equation model, the differential equation model (1) contains time-derivative terms that may be required and are not normally available for measurement. Various methods have been devised to deal with the need to reconstruct these time-derivatives [38, 31, 32, 33, 19, 39, 8, 9]. Each method is characterized by specific advantages, such as mathematical convenience, simplicity in numerical implementation and computation, handling of initial conditions, physical insight, and accuracy. Most methods are largely deterministic, in the sense that they do not explicitly model the additive noise process nor attempt to

²Another usual practice is to choose $a_n = 1$. This helps in improving the conditioning of the normal equations arising in least squares-based identification, particularly when n is large.

quantify the statistical properties of the parameter estimates. Instead, consistent estimates were obtained by using basic instrumental variable (IV) methods.

Such a deterministic approach, however, is insufficient for an errors-in-variables model. Since every available measurement is corrupted by additive noise, it is impossible to avoid bias in the parameter estimates even if a standard IV approach [9] is used. In fact there is no known way of constructing a suitable instrument vector to ensure consistency for continuous-time errors-in-variables problem. In order to achieve consistency, it is needed to study and quantify the asymptotic contribution of the additive noise in the second order moments of the reconstructed derivatives. In this paper we consider a traditional deterministic approach, known as the state-variable filter (SVF) method [35, 36, 37], and analyze the effect of additive noise therein. The analysis is applicable to non-uniformly sampled data. As a next step we use this novel characterization to develop a consistent estimator for the continuous-time EIV identification problem introduced above. We also study the asymptotic distribution of the estimates. The results in this paper can be readily extended for discrete-time systems as well.

The paper is structured as follows. In the next section we discuss briefly the basic principles behind least squares (LS) based direct continuous-time model identification, and emphasize the problems we face when the observed data are noisy. In Section 3 we study the noise effects on SVF. An identification algorithm for errors-in-variables models is discussed in Section 4. Section 5 gives the detailed statistical analysis leading to the estimation technique, and the asymptotic distribution of the estimates. The analytical results are illustrated using numerical simulations in Section 6, followed by the conclusions in Section 7.

2 Least squares identification

In this section we present a brief review of direct least squares-based identification of continuous-time models [8, 9, 7] when the data are noise-free. Next, we examine the additional issues involved in EIV problems. The background provided in this section will be used in the subsequent analysis to develop the EIV estimation algorithm.

2.1 Noise-free case

A crucial step in direct continuous-time model identification is the appropriate reconstruction of the time-derivatives from the sampled data. It is well-known that the computation of derivatives from sampled noisy data is an ill posed problem. The traditional SVF approach mitigates this problem by passing both input/output signals through an all-pole filter of minimum order n . To explain the idea, consider the case when both the input and output data are noise-free. Let us prefilter³ the *continuous-time* input-output data as follows to obtain the signals $w(t)$ and $r(t)$:

$$f_0 \frac{d^{n+1}w(t)}{dt^{n+1}} + \dots + f_n \frac{dw(t)}{dt} + f_{n+1}w(t) = y(t), \quad (2)$$

$$f_0 \frac{d^{n+1}r(t)}{dt^{n+1}} + \dots + f_n \frac{dr(t)}{dt} + f_{n+1}r(t) = u(t) \quad (3)$$

³Here we have applied a prefilter of order $n + 1$ to identify a system of order n , while a prefilter of order n is generally employed. This is done to keep the notations simple in the later parts of the paper. Using a prefilter of order $n + 1$ is not a must for the proposed approach to work. All the following analysis can be extended readily for a prefilter of order n .

Without any loss of generality we assume $f_0 = 1$. Then it is readily verified from (1) that the filtered signals $w(t)$ and $r(t)$ satisfy the differential equation

$$\begin{aligned} a_n \frac{d^n w(t)}{dt^n} + \dots + a_1 \frac{dw(t)}{dt} + a_0 w(t) \\ = b_n \frac{d^n r(t)}{dt^n} + \dots + b_1 \frac{dr(t)}{dt} + b_0 r(t) \end{aligned}$$

The equation above can be written alternatively as

$$\phi'(t)\bar{\theta} = 0 \quad \Rightarrow \quad \psi'(t)\theta = w(t), \quad (4)$$

where the regressor vectors $\phi(t)$, $\psi(t)$ and the parameter vectors θ are given by

$$\begin{aligned} \phi_I(t) &:= \left[\frac{d^n r(t)}{dt^n} \quad \dots \quad \frac{dr(t)}{dt} \quad r(t) \right]', \\ \phi_O(t) &:= \left[\frac{d^n w(t)}{dt^n} \quad \dots \quad \frac{dw(t)}{dt} \quad w(t) \right]', \\ \phi(t) &:= \left[\phi_I'(t) \quad -\phi_O'(t) \right]' = \left[\psi'(t) \quad -w(t) \right]'; \end{aligned} \quad (5)$$

$$\bar{\theta} := \left[b_n \quad \dots \quad b_0 \quad a_n \quad \dots \quad a_0 \right]' = \left[\theta' \quad 1 \right]' \quad (6)$$

Provided that it is possible to compute $\{\phi(t_k)\}_{k=0}^N$ by using a suitable numerical technique, we can stack the equations derived from (4) corresponding to different values of k in a matrix, we get an overdetermined system of equations. This overdetermined system of equations leads us to

$$\Phi\bar{\theta} = 0, \quad \Phi = \sum_{k=0}^N \phi(t_k)\phi'(t_k). \quad (7)$$

Due to numerical errors introduced in the digital simulation of the continuous-time filtering, (7) does not hold exactly⁴. Therefore, it is required to solve (7) [see also (2) and (3)] in a LS or total least squares sense.

2.2 Errors-in-variables scenario

Here we consider the case when the data are noise contaminated. Since the data are corrupted by additive measurement noise, any linear function of the data, therefore, has two components. The first component is the contribution of the noise-free part of the data, while the second part is the contribution of the measurement noise. Throughout the paper, for any linear function D of the data we use \mathring{D} to denote the noise-free part of D , and \tilde{D} to denote the noise contribution in D .

It is well-known that in presence of additive noise in either or both of the input and output measurements, the conventional least squares method gives biased estimates. If the input data are noise-free, one can however obtain consistent estimates by using an instrumental variable estimator even if the output measurements are noise corrupted. Unfortunately, in presence of additive noise in the input measurements the instrumental variable based methods [9] fail to achieve parameter consistency.

When we consider the LS method, the relationship (7) no longer holds in presence of noise. This is because Φ suffers from noise contamination. If we use SVF to compute the regressor $\phi(t)$, then it turns out that, see Section 3, we can split Φ in its noise-free and noisy parts:

$$\Phi = \mathring{\Phi} + \tilde{\Phi}$$

⁴We assume here that the simulation errors can be neglected in comparison with the estimation errors introduced by the presence of additive noise on the input-output data. Therefore, we assume the numerical simulation errors to be negligible, and concentrate in the sequel on the effect of the additive noise on the SVF outputs.

for sufficiently large N . Since the noise-free part satisfies $\mathring{\Phi}\bar{\theta} = 0$, according to the discussion above, we have

$$\Phi\bar{\theta} = \tilde{\Phi}\bar{\theta}. \quad (8)$$

Therefore solving (7) leads to serious bias in the parameter estimates. If we know $\tilde{\Phi}$, then we can solve the problem by finding the generalized eigenvectors of the pair $(\Phi, \tilde{\Phi})$ nearest to unity. But unfortunately, $\tilde{\Phi}$ depends upon the unknown noise variances σ_I^2 and σ_O^2 . Hence, we cannot use this approach. However, as we see below, we can get partial information about $\tilde{\Phi}$, which can be used to solve the identification problem.

3 Effect of measurement noise on SVF

Recall that $\phi_I(t)$ is obtained by passing $u(t)$ through a continuous-time linear time-invariant filter. Similarly, $\phi_O(t)$ is obtained by passing $y(t)$ through the same continuous-time filter. However, we have access to the sampled data $\{u(t_k)\}_{k=0}^N$ and $\{y(t_k)\}_{k=0}^N$ only. Hence to reconstruct $\phi(t_k)$ at the sampling instants $\{t_k\}_{k=0}^N$ it is required to use a standard discretization technique for implementing the continuous-time filtering operation using discrete-time input data. The resulting discrete-time representation of the filter depends on the assumptions made on the input inter-sample behaviour. The popular assumptions considered here are zero order hold (ZOH) and first order hold (FOH). In the ZOH assumption the sampled signal is assumed to remain constant between the sampling instants, while for the FOH assumption the input signal is assumed to vary linearly between sampling instants⁵. We assume that the sampling is rapid enough so that the resulting systematic errors are insignificant. Typically, the sampling frequency should be chosen greater than 20 times the bandwidth of the system to be identified. This is not a strict requirement, since the commercial A/D converters are capable of sampling fast enough to meet this requirement. A detailed systematic error analysis has been carried out in [34] assuming the continuous-time signals are bandlimited, see also [17, pp. 356] for a related discussion.

In the classical continuous-time model identification algorithms SVF part is treated as a black box. It is not required to know the digital implementation details in the subsequent stages. This is not the case in an EIV environment. Because the observed data being noise corrupted, we also pass the measurement noise along with the true signal through the SVF. So it is necessary to characterize the statistical properties of the SVF output at a later stage. In this section we give a brief description of the discrete-time state space model employed to implement the SVF. Although our discussion is limited to the ZOH and FOH assumptions, the following theory can be extended to higher order hold assumptions.

Let us define the $(n+1) \times (n+1)$ matrix A and $(n+1) \times 1$ vector b as

$$A = \begin{bmatrix} -f_1 & \cdots & -f_n & -f_{n+1} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (9)$$

Then it is straightforward to verify from the theory of the state space models in controllable canonical form that

$$\frac{d\phi_I(t)}{dt} = A\phi_I(t) + bu(t), \quad \frac{d\phi_O(t)}{dt} = A\phi_O(t) + by(t) \quad (10)$$

⁵We point out that we do not assume the signals to be actually ZOH or FOH in between the sampling instants. These assumptions are required at this stage only for the digital implementation of the continuous-time SVF.

Let us define

$$h_k := t_{k+1} - t_k. \quad (11)$$

Then for the ZOH assumption one can show that the SVF for the input side can be implemented using the sampled data in the following state space form [15, p.33]:

$$\phi_I(t_{k+1}) = e^{Ah_k} \phi_I(t_k) + [e^{Ah_k} - I]A^{-1}bu(t_k).$$

Since the SVF is asymptotically stable by construction, A is invertible. Under the FOH assumption, one can show that the *sampled* input regressor vector $\phi_I(t_k)$ satisfies [15, p.33]

$$\phi_I(t_{k+1}) = e^{Ah_k} \phi_I(t_k) + \bar{b}_k u(t_{k+1}) + \check{b}_k u(t_k),$$

where \bar{b}_k and \check{b}_k are defined as

$$\begin{aligned} \check{b}_k &= \{e^{Ah_k} A^{-1} - (1/h_k)(e^{Ah_k} - I)A^{-2}\}b, \\ \bar{b}_k &= \{(1/h_k)(e^{Ah_k} - I)A^{-2} - A^{-1}\}b. \end{aligned}$$

Therefore, the numerical implementation of the SVF to obtain the input regressor $\phi_I(t_k)$ for $k = 1, \dots, N$ takes the following form:

$$\begin{aligned} x_I(t_{k+1}) &= A_I(k)x_I(t_k) + b_I(k)\bar{u}(t_k), \\ \phi_I(t_k) &= c_I x_I(t_k), \end{aligned} \quad (12)$$

where if we use the ZOH assumption for computing the input regressor ϕ_I we have

$$x_I(t_k) = \phi_I(t_k), \quad \bar{u}(t_k) = u(t_k), \quad (13)$$

$$A_I(k) = e^{Ah_k}, \quad c_I = I, \quad b_I(k) = [e^{Ah_k} - I]A^{-1}b. \quad (14)$$

If the FOH assumption is used to compute ϕ_I , we have an augmented state vector and time delayed input signal:

$$x_I(t_k) = \begin{bmatrix} \phi_I(t_k) \\ u(t_k) \end{bmatrix}, \quad \bar{u}(t_k) = u(t_{k+1}), \quad (15)$$

$$A_I(k) = \begin{bmatrix} e^{Ah_k} & \check{b}_k \\ 0 & 0 \end{bmatrix}, \quad b_I(k) = \begin{bmatrix} \bar{b}_k \\ 1 \end{bmatrix}, \quad c_I = [I \ 0]. \quad (16)$$

The numerical steps for computing the output regressor vector $\phi_O(t_k)$ for $k = 0, \dots, N$ are identical to that of $\phi_I(t_k)$. We implement the SVF for the output side in the state space form:

$$\begin{aligned} x_O(t_{k+1}) &= A_O(k)x_O(t_k) + b_O(k)\bar{y}(t_k), \\ \phi_O(t_k) &= c_O x_O(t_k), \end{aligned} \quad (17)$$

where if we use the ZOH assumption for computing ϕ_O we use

$$x_O(t_k) = \phi_O(t_k), \quad \bar{y}(t_k) = y(t_k), \quad (18)$$

$$A_O(k) = e^{Ah_k}, \quad c_O = I, \quad b_O(k) = [e^{Ah_k} - I]A^{-1}b. \quad (19)$$

If the FOH assumption is used to compute ϕ_O , we use

$$x_O(t_k) = \begin{bmatrix} \phi_O(t_k) \\ y(t_k) \end{bmatrix}, \quad \bar{y}(t_k) = y(t_{k+1}), \quad (20)$$

$$A_O(k) = \begin{bmatrix} e^{Ah_k} & \check{b}_k \\ 0 & 0 \end{bmatrix}, \quad b_O(k) = \begin{bmatrix} \bar{b}_k \\ 1 \end{bmatrix}, \quad c_O = [I \ 0]. \quad (21)$$

Since it is common to use different inter-sample behaviours for the input and the output. In such a situation $A_I(k) \neq A_O(k)$, $b_I(k) \neq b_O(k)$, and $c_I(k) \neq c_O(k)$.

The result given in the following theorem plays a crucial role in the subsequent analysis. The proof of this theorem is given in Section 5.2 for better readability.

Theorem 1. *Let t_k be a sequence of monotonically increasing positive numbers such that $\lim_{k \rightarrow \infty} t_k = \infty$, and*

$$0 < h_* := \min_k t_{k+1} - t_k$$

exists. Assume that the discrete-time white noise processes $\tilde{u}(t_k)$ and $\tilde{y}(t_k)$ have bounded fourth order moments. Let us define the matrices $\mathring{\Phi}$ and $\check{\Phi}$ as

$$\mathring{\Phi} = \sum_{k=0}^N \mathring{\phi}(t_k) \mathring{\phi}'(t_k), \quad \check{\Phi} = \sum_{k=0}^N \mathcal{E} \{ \check{\phi}(t_k) \check{\phi}'(t_k) \}.$$

Then

$$\lim_{N \rightarrow \infty} \frac{1}{t_N} \{ \Phi - \mathring{\Phi} - \check{\Phi} \} = 0$$

with probability one (w.p.1).

Proof: See Section 5.2. ■

Remark 1. *For uniform sampling the discrete-time system in (16) is stable and time-invariant, then the discrete-time process $\check{\phi}(t_k)$ is asymptotically stationary. In that case the above theorem is well-known [13, 25], and widely used. ■*

Next, we quantify the noise contribution $\check{\Phi}$. Since $\tilde{y}(t_k)$ and $\tilde{u}(t_k)$ are mutually independent, it is straightforward to verify that $\check{\Phi}$ is a block diagonal matrix of the form

$$\check{\Phi} = \begin{bmatrix} \sigma_I^2 \check{\Phi}_I & 0 \\ 0 & \sigma_O^2 \check{\Phi}_O \end{bmatrix} = \check{\Phi} \Sigma, \quad (22)$$

$$\Sigma = \begin{bmatrix} \sigma_I^2 I & 0 \\ 0 & \sigma_O^2 I \end{bmatrix}, \quad \check{\Phi} = \begin{bmatrix} \check{\Phi}_I & 0 \\ 0 & \check{\Phi}_O \end{bmatrix}.$$

Furthermore, the matrices $\check{\Phi}_O$ and $\check{\Phi}_I$ can be computed using the knowledge of state-space implementation of SVF in the input and the output side [2]:

$$\check{\Phi}_I = \sum_{k=0}^N c_I \bar{\Phi}_I(k) c_I', \quad \check{\Phi}_O = \sum_{k=0}^N c_O \bar{\Phi}_O(k) c_O',$$

where the matrices $\bar{\Phi}_I(k)$ $\bar{\Phi}_O(k)$ are given recursively as the solution to the time-varying Lyapunov equations

$$\bar{\Phi}_I(k+1) = \begin{cases} A_I(k) \bar{\Phi}_I(k) A_I'(k) + b_I(k) b_I'(k), & k > 0 \\ b_I(1) b_I'(1), & k = 0. \end{cases}$$

$$\bar{\Phi}_O(k+1) = \begin{cases} A_O(k) \bar{\Phi}_O(k) A_O'(k) + b_O(k) b_O'(k), & k > 0 \\ b_O(1) b_O'(1), & k = 0. \end{cases}$$

If we have uniformly sampled data, the process $\tilde{\phi}(t_k)$ is asymptotically stationary. As a result $\bar{\Phi}_I(k) \rightarrow \bar{\Phi}_I(\infty)$, and $\bar{\Phi}_O(k) \rightarrow \bar{\Phi}_O(\infty)$ as $k \rightarrow \infty$. Also the matrices $\bar{\Phi}_I(\infty)$ and $\bar{\Phi}_O(\infty)$ satisfy the time-invariant discrete-time Lyapunov equations

$$\begin{aligned}\bar{\Phi}_I(\infty) &= A_I \bar{\Phi}_I(\infty) A_I' + b_I b_I', \\ \bar{\Phi}_O(\infty) &= A_O \bar{\Phi}_O(\infty) A_O' + b_O b_O'.\end{aligned}$$

Remark 2. Partition Φ into four $n \times n$ blocks as

$$\Phi = \begin{bmatrix} \Phi_{II} & \Phi_{IO} \\ \Phi_{OI} & \Phi_{OO} \end{bmatrix}, \quad (23)$$

Assume $N \rightarrow \infty$. Then the limiting matrices satisfy

$$\dot{\Phi} = \begin{bmatrix} \Phi_{II} - \sigma_I^2 \check{\Phi}_I & \Phi_{IO} \\ \Phi_{OI} & \Phi_{OO} - \sigma_O^2 \check{\Phi}_O \end{bmatrix},$$

which is a positive semidefinite matrix. Given Φ , the largest admissible value of σ_I^2 is the minimum eigenvalue associated to the generalized eigenvalue problem

$$[\Phi_{II} - \Phi_{IO} \Phi_{OO}^{-1} \Phi_{OI}]v = \lambda \check{\Phi}_I v.$$

This fact is a straightforward extension of [3, Theorem 3]. Similarly the largest admissible value of σ_I^2 is the minimum eigenvalue associated to the generalized eigenvalue problem

$$[\Phi_{OO} - \Phi_{OI} \Phi_{II}^{-1} \Phi_{IO}]v = \lambda \check{\Phi}_O v.$$

This observation can be used to carry an identifiability analysis of the problem under consideration. In particular, all the results for discrete-time EIV problem derived in [3] can be readily extended if we consider an SVF of order $n + 2$ instead of order $n + 1$ in (2) and (3). ■

4 Consistent EIV model identification

Theorem 1 does not provide any information about the rate at which the convergence takes place. Calculations similar to the proof of Theorem 2, see Section 5.1, can be used to conclude that the rate of convergence is like $1/\sqrt{t_N}$. However, for our purpose it is enough to assume that the convergence is like γ_N , which is a sequence converging to zero as $N \rightarrow \infty$. In the following we use $O(\gamma_N)$ to denote a stochastic term which converges *almost surely* (i.e with probability one) to zero as $N \rightarrow \infty$, and the convergence is at the same rate as γ_N . Then using Theorem 1 in the previous section we have for large enough t_N that [see (8) and (22)]

$$\check{\Phi}^{-1} \Phi \bar{\theta} = \Sigma \bar{\theta} + O(\gamma_N) \quad (24)$$

The first term in the right hand side of the above equation is responsible for the asymptotic bias in LS-based estimation methods. In the literature we can find several different approaches to *compensate* for this term. One popular way is to estimate this term [40] using an iterative algorithm (known as BELS method). In the Frisch scheme [3], two different equations of the form (24) are derived, one corresponding to model order n and the other corresponding to the model order $n + 1$. Finally a 1-D search algorithm

is used to estimate the parameter vector θ . BELS iteration may not converge unless the signal to noise ratio is large [22]. However, when the signal to noise ratio is large, the contribution of the first term in the right hand side of (24) is very small, and a bias correction may not be necessary at all. The Frisch scheme is known to give reliable estimates when σ_I and σ_O are of the same order, and the accuracy deteriorates significantly if the ratio σ_I/σ_O is not close to unity [24]. In addition, for continuous-time case (24) is subject to numerical approximation errors. Using a simulation study, it was noticed that the numerical approximation error in SVF stage tends to get amplified when a bias compensation using Frisch scheme is employed. Therefore we adopt a different approach in the following.

The main idea is to use two different prefilters in (9) and (10). As a result we get two different equations of the form (24). In the rest of the paper $\phi_j(t_k)$ will denote the $\phi(t_k)$ vector for the j -th prefilter ($j = 1, 2$). Similarly, the set of variables $\{A_{Ij}, A_{Oj}, b_{Ij}, b_{Oj}, c_{Ij}, c_{Oj}\}$ will be used to denote the variables $\{A_I, A_O, b_I, b_O, c_I, c_O\}$ associated to the j -th prefilter for $j = 1, 2$. Let us denote

$$\Phi_{ij} = \sum_{k=0}^N \phi_i(t_k) \phi_j'(t_k).$$

Then from Theorem 1 we can write [similar to (24)]

$$\begin{aligned} \frac{1}{t_N} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \begin{bmatrix} \bar{\theta} & 0 \\ 0 & \bar{\theta} \end{bmatrix} \\ = \frac{1}{t_N} \begin{bmatrix} \check{\Phi}_{11} & \check{\Phi}_{12} \\ \check{\Phi}_{21} & \check{\Phi}_{22} \end{bmatrix} \begin{bmatrix} \Sigma \bar{\theta} & 0 \\ 0 & \Sigma \bar{\theta} \end{bmatrix} + O(\gamma_N), \end{aligned} \quad (25)$$

which is equivalently written as

$$\Psi \bar{\theta} = \check{\Psi} \Sigma \bar{\theta} + O(\gamma_N). \quad (26)$$

Here we define

$$\begin{aligned} \Psi &= \frac{1}{t_N} [\Phi'_{11} \quad \Phi'_{12} \quad \Phi'_{21} \quad \Phi'_{22}]', \\ \check{\Psi} &= \frac{1}{t_N} [\check{\Phi}'_{11} \quad \check{\Phi}'_{12} \quad \check{\Phi}'_{21} \quad \check{\Phi}'_{22}]'. \end{aligned}$$

There are many ways to extract consistent estimate of θ from (26). Here we solve (26) in LS sense. We obtain the estimates of θ , σ_I and σ_O as

$$[\hat{\theta}' \quad \hat{\sigma}_I^2 \quad \hat{\sigma}_O^2]' = \arg \min_{\theta, \sigma_I^2, \sigma_O^2} \| \Psi \bar{\theta} - \check{\Psi} \Sigma \bar{\theta} \|^2. \quad (27)$$

Note that (26) is bilinear in Σ and $\bar{\theta}$. Therefore, we can solve the optimization problem (27) using a cyclic minimization. The optimization is initialized at a suitable $\hat{\theta}_{(0)}$. The minimization algorithm is given below.

1. Choose a suitable initial value $\hat{\theta}_{(0)}$ and set $k = 1$.
2. Set $\bar{\theta}_{(k-1)} = [\hat{\theta}'_{(k-1)} \quad 1]'$.
3. Minimize (27) with respect to σ_I^2 and σ_O^2 by setting $\bar{\theta} = \bar{\theta}_{(k-1)}$. This is a standard LS problem. Denote argument minimizers by $\hat{\sigma}_{I,(k)}^2$ and $\hat{\sigma}_{O,(k)}^2$. Set

$$\hat{\Sigma}_{(k)} = \begin{bmatrix} \hat{\sigma}_{I,(k)}^2 I & 0 \\ 0 & \hat{\sigma}_{O,(k)}^2 I \end{bmatrix}.$$

4. Minimize (27) with respect to θ by setting $\Sigma = \hat{\Sigma}_{(k)}$. This also is a standard LS problem. Denote argument minimizer by $\hat{\theta}_{(k)}$.
5. If $k > 1$ and

$$\left\| \begin{bmatrix} \hat{\theta}_{(k)} - \hat{\theta}_{(k-1)} \\ \hat{\sigma}_{I,(k)}^2 - \hat{\sigma}_{I,(k-1)}^2 \\ \hat{\sigma}_{O,(k)}^2 - \hat{\sigma}_{O,(k-1)}^2 \end{bmatrix} \right\|$$

is less than some predetermined threshold then stop; other wise $k = k + 1$ and go to step 2.

It is guaranteed to achieve convergence to a *local minimum* using the above approach [27, pp. 181]. If the initial guess $\hat{\theta}_{(0)}$ is good enough the convergence to the global minimum is ensured. Next we give a procedure to obtain a good starting value $\hat{\theta}_{(0)}$. The idea is to use a consistent estimate of θ . This idea is in fact not new, see for example [28]. Let us denote

$$\begin{bmatrix} \check{\Phi}_{11} & \check{\Phi}_{12} \\ \check{\Phi}_{21} & \check{\Phi}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}.$$

The matrix inverse in the last equality exists because the pair (A, b) [see (9)] is controllable for the prefilters. Then it is readily verified from (25) that

$$\Gamma\bar{\theta} = O(\gamma_N), \quad \Gamma = \begin{bmatrix} \Gamma_{12} \\ \Gamma_{21} \end{bmatrix}. \quad (28)$$

We can now estimate $\hat{\theta}_{(0)}$ by solving the LS problem

$$\hat{\theta}_{(0)} = \arg \min_{\theta} \|\Gamma\bar{\theta}\|^2. \quad (29)$$

Note that $\hat{\theta}_{(0)}$ is a consistent estimate of θ as well.

There is another interesting way to estimate θ without any numerical search. Note from (25) that

$$[\check{\Phi}_{11}^{-1}\Phi_{11} - \check{\Phi}_{22}^{-1}\Phi_{22}]\bar{\theta} = O(\gamma_N). \quad (30)$$

Therefore we can obtain another estimate $\hat{\theta}_a$ of θ by solving the linear LS problem

$$\hat{\theta}_a = \arg \min_{\theta} \|[\check{\Phi}_{11}^{-1}\Phi_{11} - \check{\Phi}_{22}^{-1}\Phi_{22}]\bar{\theta}\|^2. \quad (31)$$

Results from simulation study suggest that the estimates $\hat{\theta}$ and $\hat{\theta}_a$ perform similarly as far as accuracy is concerned.

Remark 3. *The pair of SVFs used in the algorithm is chosen by the user. In simulations it is noticed that the estimation accuracy is better when the filter bandwidths include the system bandwidth to be identified. This is a well-known fact when the SVF approach is used for noisy output signals only (see [9] for example). The cut-off frequencies of the two SVF filters should be chosen in order to emphasize the frequency band of interest and it is advised in general to choose them a little bit larger than the frequency bandwidth of the system to be identified. However, if the bandwidths of either or both of the SVFs are significantly larger than the system bandwidth, then the estimation accuracy deteriorates because of the amplification of the out-of-band noise. Another crucial point is that the two SVFs should not have any common pole. The performance improves if the poles are well separated. Note that it is possible to have a pair of filters with similar frequency response functions, see the illustrative example in Section 6. ■*

5 Statistical analysis

In this section we examine the asymptotic distribution of $\hat{\theta}_a$ in (31) assuming uniform sampling. Next we give a proof of Theorem 1 when the sampling is nonuniform. Throughout this section we assume that the discrete-time noise sequences $\tilde{y}(t_k)$ and $\tilde{u}(t_k)$ have bounded fourth order moments, and

$$\mathcal{E} \{ \tilde{u}^4(t_k) \} = \mu_I, \quad \mathcal{E} \{ \tilde{y}^4(t_k) \} = \mu_O.$$

5.1 Asymptotic distribution of the estimates

In this section we examine the asymptotic second order statistical properties of the estimates $\hat{\theta}_a$. The covariance matrix of the estimate $\hat{\theta}_{(0)}$ can also be derived using a similar approach, and we do not discuss that here. The derivation of the asymptotic distribution of $\hat{\theta}$, however involves more complex calculations, which is beyond the scope of this paper. Another reason for analyzing only $\hat{\theta}_a$ in this context is because $\hat{\theta}_a$ is as accurate as $\hat{\theta}$ in most cases, see Section 6. We also assume that the sampling is uniform, and the corresponding sampling interval is h . In principle, it is possible to derive the asymptotic properties of the estimates even if the sampling is non-uniform, but the associated calculations are quite involved. It is also difficult to obtain a final closed form expression for the asymptotic covariance matrix.

We use the notations (due to two different prefilters) used in Section 4. First partition Φ_{jj} [see (25)] as

$$\Phi_{jj} = [\Gamma_j \quad \gamma_j], \quad j = 1, 2,$$

where γ_j is the last column of Φ_{jj} . Then (30) implies

$$\epsilon := D\theta + d = O(\gamma_N), \tag{32}$$

where

$$D = \check{\Phi}_{11}^{-1}\Gamma_1 - \check{\Phi}_{22}^{-1}\Gamma_2, \quad d = \check{\Phi}_{11}^{-1}\gamma_1 - \check{\Phi}_{22}^{-1}\gamma_2.$$

Instead of minimizing the simple norm in (31), one can also minimize a weighted norm by solving a weighted least squares (WLS) problem. Then for any positive definite matrix W we see that

$$\hat{\theta}_a := -[D^\top W D]^{-1} D^\top W d = \theta + O(\gamma_N).$$

Therefore $\hat{\theta}_a$ is a consistent estimator of the parameter vector θ for any positive definite W . We can use the standard theory of WLS estimation to derive the asymptotic covariance matrix of $\hat{\theta}$. In that goal let us introduce $\mathring{\Gamma}_j$ and $\mathring{\gamma}_j$ such that

$$\mathring{\Phi}_{jj} = [\mathring{\Gamma}_j \quad \mathring{\gamma}_j], \quad j = 1, 2.$$

Then as a consequence of Theorem 1 it follows for large N that, see also (22)

$$\check{\Phi}_{jj}^{-1} \Phi_{jj} = \check{\Phi}_{jj}^{-1} \mathring{\Phi}_{jj} + \Sigma + O(\gamma_N).$$

Thus, for large N we get

$$D = \bar{D} + O(\gamma_N)$$

where

$$\bar{D} = \check{\Phi}_{11}^{-1} \mathring{\Gamma}_1 - \check{\Phi}_{22}^{-1} \mathring{\Gamma}_2.$$

Then following a few steps of standard calculations [25] it is straightforward to verify that, see (32)

$$\theta - \hat{\theta}_a = [\bar{D}^\top W \bar{D}]^{-1} \bar{D}^\top W \epsilon,$$

after neglecting the higher order stochastic contributions in γ_N^2 . Now by denoting

$$Q = \text{cov}(\epsilon) = \text{cov}\{(\check{\Phi}_{11}^{-1}\Phi_{11} - \check{\Phi}_{22}^{-1}\Phi_{22})\bar{\theta}\}, \quad (33)$$

it follows right away that

$$\text{cov}(\hat{\theta}_a) = [\bar{D}^\top W \bar{D}]^{-1} \bar{D}^\top W Q W \bar{D} [\bar{D}^\top W \bar{D}]^{-1}.$$

The covariance matrix is minimized if we choose the positive definite weighting matrix $W = W_\star$ where

$$W_\star = Q^{-1},$$

and the resulting covariance matrix of the associated optimal estimate $\hat{\theta}_\star$ is given by

$$\text{cov}(\hat{\theta}_\star) = [\bar{D}^\top Q^{-1} \bar{D}]^{-1}.$$

Next, we give a closed form expression for the residual covariance matrix Q when the sampling is uniform.

The process $\check{\phi}(t_k)$ is stationary as for large k in case of uniform sampling. Recall that in the estimation algorithm, we use two prefilters both in the input and the output. The state space realization of the input side prefilters are denoted by (A_{I1}, b_{I1}, c_I) , (A_{I2}, b_{I2}, c_I) . Note that c_I does not depend on the prefilter transfer function. Similarly, the state space realization of the output side prefilters will be denoted by (A_{O1}, b_{O1}, c_O) , (A_{O2}, b_{O2}, c_O) . Let us define

$$H_{Ij}(\ell) = c_I A_{Ij}^\ell b_{Ij}, \quad H_{Oj}(\ell) = c_O A_{Oj}^\ell b_{Oj}, \quad (34)$$

and

$$\left. \begin{aligned} \bar{P}_O^{jk} &:= \sum_{\ell=0}^{\infty} A_{Oj}^\ell b_{Oj} b'_{Ok} A'_{Ok}{}^\ell, \\ \bar{P}_I^{jk} &:= \sum_{\ell=0}^{\infty} A_{Ij}^\ell b_{Ij} b'_{Ik} A'_{Ik}{}^\ell, \end{aligned} \right\} j, k = 1, 2. \quad (35)$$

Note that we can compute \bar{P}_O^{jk} and \bar{P}_I^{jk} by solving the Sylvester equations

$$\begin{aligned} \bar{P}_O^{jk} &= A_{Oj} \bar{P}_O^{jk} A'_{Ok} + b_{Oj} b'_{Ok}, \\ \bar{P}_I^{jk} &= A_{Ij} \bar{P}_I^{jk} A'_{Ik} + b_{Ij} b'_{Ik}. \end{aligned}$$

It is well-known [2] for $\tau \rightarrow \infty$ and any ℓ that

$$\begin{aligned} \mathcal{E} \{ \check{\phi}^j(t_{\tau+\ell}) \check{\phi}^{k'}(t_\tau) \} &= P^{jk}(\ell) \Sigma; \\ P^{jk}(\ell) &:= \begin{bmatrix} P_I^{jk}(\ell) & 0 \\ 0 & P_O^{jk}(\ell) \end{bmatrix} \end{aligned} \quad (36)$$

where we define $P_I^{jk}(\ell)$ as

$$P_I^{jk}(\ell) = \begin{cases} c_I A_{Ij}^\ell \bar{P}_I^{jk} c_I' & \ell \geq 0 \\ c_I \bar{P}_I^{jk} A'_{Ik}{}^{-\ell} c_I' & \ell < 0. \end{cases} \quad (37)$$

The matrix $P_O^{jk}(\ell)$ is defined in a similar way as $P_I^{jk}(\ell)$ by replacing subscripts as before. We point out that (36) can also be verified readily using a calculation similar to (54). The difference here is that we compute the cross-correlation between the outputs of two different filters. We note by passing that for uniform sampling for large number of samples

$$\check{\Phi}_{jj} = N P^{jj}(0). \quad (38)$$

With slight abuse of notations, we use $P^{jk}(0) = P^{jk}$, $P_I^{jk}(0) = P_I^{jk}$ and $P_O^{jk}(0) = P_O^{jk}$ for short.

The main result is stated in the following theorem, proof of which will be deferred to Appendix A.

Theorem 2. Assume that the true input signal $\mathring{u}(t)$ is a zero mean ergodic process so that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N \mathring{\phi}_j(t_{\ell+\tau}) \mathring{\phi}'_k(t_\ell) = R^{jk}(\tau).$$

Let the fourth order moments of the measurement noise sequences be $\mathcal{E} \{\tilde{u}^4(t_k)\} = \mu_I$ and $\mathcal{E} \{\tilde{y}^4(t_k)\} = \mu_O$.

Let us define the block diagonal matrix M with $(n+1) \times (n+1)$ diagonal blocks as

$$M = \begin{bmatrix} \mu_I I & 0 \\ 0 & \mu_O I \end{bmatrix}.$$

Then as the number of data samples $N \rightarrow \infty$ then ϵ is asymptotically Gaussian with covariance matrix

$$Q = \frac{1}{N} \sum_{j=1}^2 \sum_{k=1}^2 (-1)^{j-k} [P^{jj}(0)]^{-1} X^{jk} [P^{kk}(0)]^{-1}.$$

where the matrix X^{jk} is defined as

$$\begin{aligned} X^{jk} &= \sum_{\ell=-\infty}^{\infty} \bar{\theta}' P^{jk}(\ell) \Sigma \bar{\theta} \{R^{jk}(\ell) + P^{jk}(\ell) \Sigma\} \\ &\quad + \sum_{\ell=-\infty}^{\infty} P^{jk}(\ell) \Sigma \bar{\theta} \bar{\theta}' \Sigma P^{jk}(\ell) \\ &\quad + (M - 3\Sigma^2) S^{jk} \end{aligned}$$

for $i, j \in \{1, 2\}$. We define S^{jk} as

$$S^{jk} = \begin{bmatrix} S_I^{jk} & 0 \\ 0 & S_O^{jk} \end{bmatrix}.$$

where

$$\begin{aligned} S_I^{jk} &= \sum_{\ell=0}^{\infty} \left[\{H_{I_j}(\ell) H'_{I_j}(\ell)\} \theta_I \theta'_I \{c_I A_{I_k}^\ell P_I^{kk}(0) A_{I_k}^{\prime\ell} c'_I\} \right. \\ &\quad \left. + \{c_I A_{I_j}^\ell P_I^{jj}(0) A_{I_j}^{\prime\ell} c'_I\} \theta_I \theta'_I \{H_{I_k}(\ell) H'_{I_k}(\ell)\} \right]. \end{aligned} \quad (39)$$

and S_O^{jk} is obtained by replacing the subscript I by O in (39).

Proof: See Appendix A. ■

5.2 Convergence analysis for nonuniform sampling

In this section we give a proof of Theorem 1 for non-uniform sampling case. Here we do not distinguish between two different prefilters used for EIV estimation algorithms described in Section 4. We stick to the general notations used in Section 3. The following results constitute the basic tools in the analysis, the proof of which are given in the appendix.

Lemma 1. Let t_k be a sequence of monotonically increasing positive numbers such that $\lim_{k \rightarrow \infty} t_k = \infty$, and

$$0 < h_\star := \min_k t_{k+1} - t_k$$

exists. Let $\{v(t_k)\}_{k>0}$ be a sequence of zero-mean random variables such that

$$|\mathcal{E}\{v(t_{k+\tau})v(t_k)\}| \leq Br^{|\tau|},$$

for some $0 < B < \infty$ and $0 < r < 1$. Let $\{\varphi(t_k)\}_{k>0}$ be a bounded ($|\varphi(t_k)| < B_1 < \infty$, say) real-valued sequence. Then

$$\lim_{N \rightarrow \infty} \frac{1}{t_N} \sum_{k=0}^N \varphi(t_k)v(t_k) = 0 \quad \text{w.p.1.}$$

Proof: See Appendix B. ■

Lemma 2. Under the assumptions of Theorem 1 the following hold:

1. $|\mathring{\phi}_I(t_k)|$ is bounded for all k .
2. There exists $0 < B_0 < \infty$ such that

$$\mathcal{E}\{\tilde{\phi}_I(t_k)\tilde{\phi}'_I(t_\ell)\} \preceq B_0 e^{-\lambda h_* |k-\ell|},$$

where $-\lambda$ is the real part of the eigenvalue of A nearest to the imaginary axis.

Proof: See Appendix C. ■

Lemma 3. Let $-\lambda$ be the real part of the eigenvalue of A nearest to the imaginary axis. Then under the assumptions of Theorem 1 there exists $0 < B_1 < \infty$ such that

$$\mathcal{E}[\alpha(t_k) \odot \alpha(t_\ell)] \preceq B_1 e^{-\lambda |t_k - t_\ell|}, \quad (40)$$

where \odot denotes the matrix Hadamard product (element wise multiplication) operator, and we define

$$\alpha(t_k) = \tilde{\phi}_I(t_k)\tilde{\phi}'_I(t_k) - \mathcal{E}\{\tilde{\phi}_I(t_k)\tilde{\phi}'_I(t_k)\}.$$

Proof: See Appendix D. ■

Proof of Theorem 1

Consider the partition of Φ in (23). It is required to establish the convergence of each of the three individual blocks Φ_{II} , Φ_{IO} and Φ_{OO} to prove the asymptotic convergence of Φ in Theorem 1. Here we present the proof for the block Φ_{II} . The proof for Φ_{OO} is exactly identical if we replace the subscripts I by O . The proof for the Φ_{IO} is also similar but simpler. Now

$$\begin{aligned} \Phi_{II} &= \sum_{k=0}^N \phi_I(t_k)\phi'_I(t_k) \\ &= \sum_{k=0}^N [\mathring{\phi}_I(t_k)\mathring{\phi}'_I(t_k) + \mathring{\phi}_I(t_k)\tilde{\phi}'_I(t_k) \\ &\quad + \tilde{\phi}_I(t_k)\mathring{\phi}'_I(t_k) + \tilde{\phi}_I(t_k)\tilde{\phi}'_I(t_k)] \\ &= \mathring{\Phi}_{II} + \sum_{k=0}^N [\mathring{\phi}_I(t_k)\tilde{\phi}'_I(t_k) + \tilde{\phi}_I(t_k)\mathring{\phi}'_I(t_k) \\ &\quad + \tilde{\phi}_I(t_k)\tilde{\phi}'_I(t_k)], \end{aligned}$$

where $\mathring{\Phi}_{II}$ is defined in the same way as $\mathring{\Phi}$. It suffices to show that

$$\lim_{N \rightarrow \infty} \frac{1}{t_N} \sum_{k=0}^N [\mathring{\phi}_I(t_k) \tilde{\phi}'_I(t_k) + \tilde{\phi}_I(t_k) \mathring{\phi}'_I(t_k)] = 0, \quad \text{w.p.1} \quad (41)$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{t_N} \sum_{k=0}^N \tilde{\phi}_I(t_k) \tilde{\phi}'_I(t_k) - \tilde{\Phi}_{II} = 0, \quad \text{w.p.1} \quad (42)$$

where

$$\tilde{\Phi}_{II} = \frac{1}{t_N} \sum_{k=0}^N \mathcal{E} \{ \tilde{\phi}_I(t_k) \tilde{\phi}'_I(t_k) \}.$$

Now (41) follows from Lemma 1 and Lemma 2, while (42) follows from Lemma 1 and Lemma 3. With this the proof is complete.

6 A case study

In the simulations we consider the system [5]

$$\frac{d^2 \mathring{y}(t)}{dt^2} + 2 \frac{d \mathring{y}(t)}{dt} + \mathring{y}(t) = \frac{d \mathring{u}(t)}{dt} - \mathring{u}(t).$$

The input $\mathring{u}(t)$ is chosen as a multi-sine signal:

$$\mathring{u}(t) = \sin(t) + \sin(1.9t) + \sin(2.1t) + \sin(2.3t).$$

The input-output signals are sampled uniformly with a sampling interval 0.05 s. The number of samples is $N = 1000$, *i.e.* the observation time is $t_N = 50$ s. We adjust the noise variances σ_I^2 and σ_O^2 to control the signal to noise ratio of the observed data. The true system was simulated by solving the *autonomous* ordinary differential equation associated to the joint input-output data using a state-space approach with proper initial conditions⁶.

We need two prefilters of the form (2) and (3) for EIV identification algorithm. For both filters we use $f_0 = 1$. For the first filter we have chosen

$$f_1 = 3.25, \quad f_2 = 4.5, \quad f_3 = 2.5.$$

For the second filter these values are

$$f_1 = 2.75, \quad f_2 = 6.5, \quad f_3 = 3.75.$$

In the identification process, we use the FOH assumption for both the input and output signals. In Table 1 we show the estimation results from a Monte Carlo simulation of 100 independent runs for a signal-to-noise ratio (SNR) of 10 dB for both input and output signals. For this SNR the variance of the input measurement noise σ_I^2 equals 0.1950, while the variance of the output measurement noise σ_O^2 equals 0.0532. The measurement noise sequences are Gaussian distributed. In Table 1 we show the mean and empirical standard deviation of the conventional LS-based SVF estimate (denoted by $\hat{\theta}_{ls}$) [9], $\hat{\theta}_{(0)}$ obtained as in (29), $\hat{\theta}$ obtained as in (27) and $\hat{\theta}_a$ as in (31). As can be seen in Table 1, the ordinary LS-based SVF

⁶We emphasize that the solution to any autonomous linear differential equation can be computed exactly by using matrix exponentials.

		b_1	b_0	a_2	a_1
	True value	1	-1	1	2
$\hat{\theta}_{ls}$	Mean	0.8743	-0.8170	0.8822	1.7142
	Std Dev (empirical)	0.0417	0.0515	0.0336	0.0860
$\hat{\theta}_{(0)}$	Mean	1.0434	-1.0661	1.0371	2.0816
	Std Dev (empirical)	0.1818	0.2549	0.1508	0.3186
$\hat{\theta}$	Mean	1.0131	-1.0170	1.0058	2.0296
	Std Dev (empirical)	0.0807	0.1112	0.0624	0.1593
$\hat{\theta}_a$	Mean	1.0138	-1.0242	1.0123	2.0408
	Std Dev (empirical)	0.0836	0.1195	0.0730	0.1834
	Std Deviation (analytical)	0.0831	0.1173	0.0768	0.1768

Table 1: Simulation results for uniformly sampled data with SNR= 10 dB.

estimate is biased. The estimate $\hat{\theta}$ is significantly more accurate than $\hat{\theta}_{(0)}$. However, $\hat{\theta}_a$ performs as good as $\hat{\theta}$ in most cases. The analytical standard deviation derived via Theorem 2 is quite well in agreement with the empirical standard deviation of $\hat{\theta}_a$.

A similar result for 5 dB SNR is shown in Table 2. In this case $\sigma_I^2 = 0.6165$ and $\sigma_O^2 = 0.1683$. As expected, with increase in the measurement noise level, the bias in the LS based SVF estimate is now significant. On the other hand, estimates proposed in the paper are unbiased. However, the estimation accuracy of $\hat{\theta}_{(0)}$ deteriorates significantly. On the other hand $\hat{\theta}$ is comparatively more accurate. It is clear that $\hat{\theta}_a$ is also as reliable as $\hat{\theta}$. In this case also the empirical standard deviation of $\hat{\theta}_a$ is quite close to the analytical prediction provided by Theorem 2.

In our final example we test the algorithm with non-uniformly sampled data. We use the same multi-sine signal as input. The data are sampled non-uniformly where the sampling interval is uniformly distributed in $[0.03, 0.07]$. The SNR in this case is 10 dB. The estimation results from a Monte Carlo simulation of 100 independent runs are shown in Table 3. The estimation results using non-uniform sampling follows a pattern that is similar to the uniform sampling case in Table 1.

Noise variance estimation results obtained in the above examples are shown in Table 4.

7 Conclusions

In this paper we have addressed the problem of identifying a continuous-time dynamic EIV models using a direct identification approach. In that goal, we have presented a new framework for analyzing the effect of noise on state variable filtering. The results therefrom are applied to develop a set of algorithms for identification of continuous-time dynamic EIV model. The proposed methods employ two state variable filters in order to circumvent the problem of additive noise, and resulting bias in the conventional LS estimate. The computation of $\hat{\theta}_{(0)}$ and $\hat{\theta}_a$ does not require any numerical search and is computationally

		b_1	b_0	a_2	a_1
	True value	1	-1	1	2
$\hat{\theta}_{ls}$	Mean	0.6770	-0.5443	0.7124	1.2655
	Std Dev (empirical)	0.0490	0.0720	0.0482	0.1062
$\hat{\theta}_{(0)}$	Mean	1.0549	-1.0971	1.0583	2.0918
	Std Dev (empirical)	0.3137	0.4679	0.2858	0.5782
$\hat{\theta}$	Mean	0.9952	-1.0018	0.9984	1.9930
	Std Dev (empirical)	0.1398	0.2032	0.1191	0.2906
$\hat{\theta}_a$	Mean	1.0083	-0.9868	1.0028	1.9980
	Std Dev (empirical)	0.1483	0.2241	0.1427	0.3391
	Std Dev (analytical)	0.1482	0.2089	0.1369	0.3151

Table 2: Simulation results for uniformly sampled data with SNR= 5 dB.

		b_1	b_0	a_2	a_1
	True value	1	-1	1	2
$\hat{\theta}_{ls}$	Mean	0.8608	-0.7980	0.8727	1.6913
	Std Dev (empirical)	0.0432	0.0495	0.0358	0.0893
$\hat{\theta}_{(0)}$	Mean	1.0097	-1.0099	1.0049	2.0221
	Std Dev (empirical)	0.1501	0.2027	0.1284	0.2801
$\hat{\theta}$	Mean	1.9953	-0.9905	1.9947	1.9990
	Std Dev (empirical)	0.0873	0.1072	0.0697	0.1792
$\hat{\theta}_a$	Mean	1.0093	-1.0047	1.0040	2.0184
	Std Dev (empirical)	0.0933	0.1189	0.0809	0.2015

Table 3: Simulation results for non-uniformly sampled data with SNR= 10 dB.

sampling	SNR		σ_I^2	σ_O^2
uniform	10dB	True value	0.1950	0.0532
		Mean	0.1265	0.0531
		Std Dev (empirical)	0.6422	0.1474
	5dB	True value	0.6165	0.1683
		Mean	0.5321	0.1397
		Std Dev (empirical)	1.0045	0.2256
nonuniform	10dB	True value	0.1950	0.0532
		Mean	0.1808	0.0489
		Std Dev (empirical)	0.5250	0.0878

Table 4: Noise variance estimation results.

efficient. This is in contrast with other bias elimination methods for identification of EIV model. The estimate $\hat{\theta}$, however, needs a numerical search, but the convergence to the global minimum is ensured in almost all the cases. The proposed methods can also be applied for non-uniformly sampled data. The estimate $\hat{\theta}_{(0)}$ is like an instrumental variable approach. Like other instrumental variable estimators, $\hat{\theta}_{(0)}$ is unable to provide good result when the SNR is low. One can derive a refined estimate $\hat{\theta}$, which is more reliable in low signal to noise ratio. However, it has been seen in simulation experiments that $\hat{\theta}_a$ is as accurate as $\hat{\theta}$. We have derived the asymptotic distribution of $\hat{\theta}_a$ in this paper under assumption of uniform sampling. However, this analysis might be extended to derive better estimates. For example, it is not yet clear how user can choose the prefilters in order to use the information most efficiently. It is also interesting to investigate how (25) should be processed to yield more accurate results. Finally we emphasize that the proposed algorithms can also be extended for identifying also the discrete-time EIV models in a straightforward manner.

A Proof of Theorem 2

In this appendix we prove Theorem 2. Recall from (33) that

$$Q = \sum_{j=1}^2 \sum_{k=1}^2 (-1)^{j-k} \check{\Phi}_{jj}^{-1} \mathcal{E} \{ \Phi_{jj} \bar{\theta} \bar{\theta}' \Phi_{kk}' \} [\check{\Phi}_{kk}']^{-1}. \quad (43)$$

Consider the term within the curly braces in the last equation. By definition of Φ_{jj} it follows that

$$\begin{aligned} \Phi_{jj} \bar{\theta} &= \sum_{\ell=0}^N \phi_j(t_\ell) \phi_j'(t_\ell) \bar{\theta} \\ &= \sum_{\ell=0}^N \phi_j(t_\ell) \tilde{\phi}_j'(t_\ell) \bar{\theta} \\ &= \sum_{\ell=0}^N \hat{\phi}_j(t_\ell) \tilde{\phi}_j'(t_\ell) \bar{\theta} + \sum_{\ell=0}^N \check{\phi}_j(t_\ell) \tilde{\phi}_j'(t_\ell) \bar{\theta}, \end{aligned}$$

where in the last equality we use $\mathring{\phi}'\bar{\theta} = 0$, see (4) in the noise-free case. Now using Ljung's central limit theorem [25, p.550], see also [12, 13], it is readily verified the ϵ is asymptotically Gaussian. It also follows straightaway that

$$\begin{aligned} & \mathcal{E} \{ \Phi_{jj} \bar{\theta} \bar{\theta}' \Phi_{kk}' \} \\ &= \mathcal{E} \left\{ \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathring{\phi}_j(t_{\ell_1}) \mathring{\phi}_j'(t_{\ell_1}) \bar{\theta} \bar{\theta}' \tilde{\phi}_k(t_{\ell_2}) \tilde{\phi}_k'(t_{\ell_2}) \right. \\ & \quad \left. + \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \tilde{\phi}_j(t_{\ell_1}) \tilde{\phi}_j'(t_{\ell_1}) \bar{\theta} \bar{\theta}' \tilde{\phi}_k(t_{\ell_2}) \tilde{\phi}_k'(t_{\ell_2}) \right\}. \end{aligned} \quad (44)$$

Note the third moment terms vanish in (44) because $\mathcal{E} \mathring{\phi}_j(t_\ell) = 0$ [since $\hat{u}(t)$ is assumed to have zero mean] and $\tilde{\phi}_k(t_{\ell_1})$ is independent of $\mathring{\phi}_j(t_{\ell_2})$ for all j, k, ℓ_1 and ℓ_2 .

In all the following calculation we assume $N \rightarrow \infty$. To simplify the notations, often we use this without any explicit mention. Next we consider the two terms in the right hand side of (44) separately. Note that we assume the true input signal $\hat{u}(t)$ to be ergodic. Consequently we write

$$\mathcal{E} \left\{ \mathring{\phi}_j(t_{\ell_1}) \mathring{\phi}_k'(t_{\ell_2}) \right\} = R^{jk}(\ell_1 - \ell_2)$$

Then using the statistical independence of $\mathring{\phi}_j(t_k)$ and $\tilde{\phi}_\ell(t_k)$ for all $j, k = 1, 2$ we get

$$\begin{aligned} & \mathcal{E} \left\{ \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathring{\phi}_j(t_{\ell_1}) \tilde{\phi}_j'(t_{\ell_1}) \bar{\theta} \bar{\theta}' \tilde{\phi}_k(t_{\ell_2}) \mathring{\phi}_k'(t_{\ell_2}) \right\} \\ &= \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N R^{jk}(\ell_1 - \ell_2) \bar{\theta}' P^{jk}(\ell_1 - \ell_2) \Sigma \bar{\theta} \\ &= N \sum_{\ell=-\infty}^{\infty} R^{jk}(\ell) \bar{\theta}' P^{jk}(\ell) \Sigma \bar{\theta}, \end{aligned} \quad (45)$$

where we have used (36). Next we consider the second term in the right hand side of (44). This is a fourth order moment term. We partition it into the blocks as

$$\begin{aligned} & \mathcal{E} \left\{ \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \tilde{\phi}_j(t_{\ell_1}) \tilde{\phi}_j'(t_{\ell_1}) \bar{\theta} \bar{\theta}' \tilde{\phi}_k(t_{\ell_2}) \tilde{\phi}_k'(t_{\ell_2}) \right\} \\ &= \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}. \end{aligned}$$

The matrices Q_{11} , Q_{12} , Q_{21} and Q_{22} are given below, where we show only the terms which give a nonzero

contribution:

$$\begin{aligned}
Q_{11} &= \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathcal{E} \left[\tilde{\phi}_{I_j}(t_{\ell_1}) \tilde{\phi}'_{I_j}(t_{\ell_1}) \theta_I \theta'_I \tilde{\phi}_{Ik}(t_{\ell_2}) \tilde{\phi}'_{Ik}(t_{\ell_2}) \right. \\
&\quad \left. + \tilde{\phi}_{I_j}(t_{\ell_1}) \tilde{\phi}'_{O_j}(t_{\ell_1}) \theta_O \theta'_O \tilde{\phi}_{Ok}(t_{\ell_2}) \tilde{\phi}'_{Ik}(t_{\ell_2}) \right], \\
Q_{22} &= \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathcal{E} \left[\tilde{\phi}_{O_j}(t_{\ell_1}) \tilde{\phi}'_{O_j}(t_{\ell_1}) \theta_O \theta'_O \tilde{\phi}_{Ok}(t_{\ell_2}) \tilde{\phi}'_{Ok}(t_{\ell_2}) \right. \\
&\quad \left. + \tilde{\phi}_{O_j}(t_{\ell_1}) \tilde{\phi}'_{I_j}(t_{\ell_1}) \theta_I \theta'_I \tilde{\phi}_{Ik}(t_{\ell_2}) \tilde{\phi}'_{Ok}(t_{\ell_2}) \right], \\
Q_{12} &= \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathcal{E} \left[\tilde{\phi}_{I_j}(t_{\ell_1}) \tilde{\phi}'_{I_j}(t_{\ell_1}) \theta_I \theta'_O \tilde{\phi}_{Ok}(t_{\ell_2}) \tilde{\phi}'_{Ok}(t_{\ell_2}) \right. \\
&\quad \left. + \tilde{\phi}_{I_j}(t_{\ell_1}) \tilde{\phi}'_{O_j}(t_{\ell_1}) \theta_O \theta'_I \tilde{\phi}_{Ik}(t_{\ell_2}) \tilde{\phi}'_{Ok}(t_{\ell_2}) \right], \\
Q_{21} &= \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathcal{E} \left[\tilde{\phi}_{O_j}(t_{\ell_1}) \tilde{\phi}'_{O_j}(t_{\ell_1}) \theta_O \theta'_I \tilde{\phi}_{Ik}(t_{\ell_2}) \tilde{\phi}'_{Ik}(t_{\ell_2}) \right. \\
&\quad \left. + \tilde{\phi}_{O_j}(t_{\ell_1}) \tilde{\phi}'_{I_j}(t_{\ell_1}) \theta_I \theta'_O \tilde{\phi}_{Ok}(t_{\ell_2}) \tilde{\phi}'_{Ik}(t_{\ell_2}) \right].
\end{aligned}$$

Due to statistical independence of $\tilde{\phi}_{Ok}$ and $\tilde{\phi}_{I_j}$ it is fairly straightforward to evaluate Q_{21} and Q_{12} . We use (36) and the definition of Σ . A calculation similar to (45) gives by taking $N \rightarrow \infty$

$$\begin{aligned}
Q_{12} &= \sigma_I^2 \sigma_O^2 N^2 P_I^{jj}(0) \theta_I \theta'_O P_O^{kk}(0) \\
&\quad + N \sigma_I^2 \sigma_O^2 \sum_{\ell=-\infty}^{\infty} P_I^{jk}(\ell) \theta_I \theta'_O P_O^{jk}(\ell), \tag{46}
\end{aligned}$$

$$\begin{aligned}
Q_{21} &= \sigma_I^2 \sigma_O^2 N^2 P_O^{jj}(0) \theta_O \theta'_I P_I^{kk}(0) \\
&\quad + N \sigma_I^2 \sigma_O^2 \sum_{\ell=-\infty}^{\infty} P_O^{jk}(\ell) \theta_O \theta'_I P_I^{jk}(\ell). \tag{47}
\end{aligned}$$

Computation of Q_{11} and Q_{22} involves identical sequence of calculations. First consider Q_{11} . From the definition we have following a calculation similar to (45)

$$\begin{aligned}
Q_{11} &= \sum_{\ell_1=0}^N \sum_{\ell_2=0}^N \mathcal{E} \left[\tilde{\phi}_{I_j}(t_{\ell_1}) \tilde{\phi}'_{I_j}(t_{\ell_1}) \theta_I \theta'_I \tilde{\phi}_{Ik}(t_{\ell_2}) \tilde{\phi}'_{Ik}(t_{\ell_2}) \right] \\
&\quad + N \sum_{\ell=-\infty}^{\infty} \sigma_I^2 \sigma_O^2 P_I^{jk}(\ell) \theta'_O P_O^{jk}(\ell) \theta_O. \tag{48}
\end{aligned}$$

Computation of the fourth order moment in the first term in above involves a few standard but tedious steps for computing the fourth order moments of filtered white noise sequences [25, p.549]. It can be shown

for large ℓ_1 and ℓ_2 that [25, p.549]

$$\begin{aligned}
& \mathcal{E} \left[\tilde{\phi}_{I_j}(t_{\ell_1}) \tilde{\phi}'_{I_j}(t_{\ell_1}) \theta_I \theta'_I \tilde{\phi}_{I_k}(t_{\ell_2}) \tilde{\phi}'_{I_k}(t_{\ell_2}) \right] \\
&= \sigma_I^4 \left\{ \sum_{\ell=0}^{\infty} H_{I_j}(\ell) H'_{I_j}(\ell) \right\} \theta_I \theta'_I \left\{ \sum_{\ell=0}^{\infty} H_{I_k}(\ell) H'_{I_k}(\ell) \right\} \\
&+ \sigma_I^4 \left\{ \sum_{\ell=0}^{\infty} H_{I_j}(\ell) H'_{I_k}(\ell_1 - \ell_2 + \ell) \right\} \theta_I \theta'_I \times \\
&\quad \left\{ \sum_{\ell=0}^{\infty} H_{I_j}(\ell) H'_{I_k}(\ell_1 - \ell_2 + \ell) \right\} \\
&+ \sigma_I^4 \theta'_I \left\{ \sum_{\ell=0}^{\infty} H_{I_j}(\ell) H'_{I_k}(\ell_1 - \ell_2 + \ell) \right\} \theta_I \times \\
&\quad \left\{ \sum_{\ell=0}^{\infty} H_{I_j}(\ell) H'_{I_k}(\ell_1 - \ell_2 + \ell) \right\} \\
&+ [\mu_I - 3\sigma_I^2] \sum_{\ell=0}^{\infty} \left\{ H_{I_j}(\ell_1 - \ell_2 + \ell) H'_{I_j}(\ell_1 - \ell_2 + \ell) \times \right. \\
&\quad \left. \theta_I \theta'_I H_{I_k}(\ell) H'_{I_k}(\ell) \right\} \\
&= \sigma_I^4 \left[P_I^{jj}(0) \theta_I \theta'_I P_I^{jj}(0) + P_I^{jk}(\ell_1 - \ell_2) \theta_I \theta'_I P_I^{jk}(\ell_1 - \ell_2) \right. \\
&\quad \left. + \theta'_I P_I^{jk}(\ell_1 - \ell_2) \theta_I P_I^{jk}(\ell_1 - \ell_2) \right] \\
&\quad + [\mu_I - 3\sigma_I^2] \sum_{\ell=0}^{\infty} \left\{ H_{I_j}(\ell_1 - \ell_2 + \ell) H'_{I_j}(\ell_1 - \ell_2 + \ell) \times \right. \\
&\quad \left. \theta_I \theta'_I H_{I_k}(\ell) H'_{I_k}(\ell) \right\}, \tag{49}
\end{aligned}$$

where in the second equality we have used (34), (35) and (37). The contribution of the last term in the right hand side of (49) in (48) is given by (we omit the factor $\mu_I - 3\sigma_I^4$ for convenience)

$$\begin{aligned}
& N \sum_{\ell=0}^N \sum_{\tau=-N}^N H_{I_j}(\tau + \ell) H'_{I_j}(\tau + \ell) \theta_I \theta'_I H_{I_k}(\ell) H'_{I_k}(\ell) \\
&= N \sum_{\ell=0}^N \sum_{\tau=0}^{\infty} [H_{I_j}(\tau + \ell) H'_{I_j}(\tau + \ell) \theta_I \theta'_I H_{I_k}(\ell) H'_{I_k}(\ell) \\
&\quad + H_{I_j}(\ell) H'_{I_j}(\ell) \theta_I \theta'_I H_{I_k}(\ell + \tau) H'_{I_k}(\ell + \tau)] \\
&= N \sum_{\ell=0}^N \left[c_I A_{I_j}^{\ell} P_I^{jj}(0) A_{I_j}^{\prime \ell} c_I \theta_I \theta'_I H_{I_k}(\ell) H'_{I_k}(\ell) \right. \\
&\quad \left. + H_{I_j}(\ell) H'_{I_j}(\ell) \theta_I \theta'_I c_I A_{I_k}^{\ell} P_I^{kk}(0) A_{I_k}^{\prime \ell} c_I \right] \\
&= N S_I^{jk}
\end{aligned}$$

We emphasize that the terms in the summation above are exponentially decaying, and it is straightforward

to verify that the summation does converge. Combining this with (49) and (48) we get

$$\begin{aligned}
Q_{11} &= \sigma_I^4 N^2 P_I^{jj}(0) \theta_I \theta'_I P_I^{kk}(0) \\
&+ N \sum_{\ell=-\infty}^{\infty} \sigma_I^4 P_I^{jk}(\ell) \theta'_I P_I^{jk}(\ell) \theta_I \\
&+ N \sum_{\ell=-\infty}^{\infty} \sigma_I^4 P_I^{jk}(\ell) \theta_I \theta'_I P_I^{jk}(\ell) \\
&+ N \sum_{\ell=-\infty}^{\infty} \sigma_I^2 \sigma_O^2 P_I^{jk}(\ell) \theta'_O P_O^{jk}(\ell) \theta_O + N(\mu_I - 3\sigma_I^2) S_I^{jk} \\
&= \sigma_I^4 N^2 P_I^{jj}(0) \theta_I \theta'_I P_I^{kk}(0) \\
&+ N \sigma_I^2 \sum_{\ell=-\infty}^{\infty} \left\{ \sigma_O^2 \theta'_O P_O^{jk}(\ell) \theta_O + \sigma_I^2 \theta'_I P_I^{jk}(\ell) \theta_I \right\} P_I^{jk}(\ell) \\
&+ N \sigma_I^4 \sum_{\ell=-\infty}^{\infty} P_I^{jk}(\ell) \theta_I \theta'_I P_I^{jk}(\ell) \\
&+ N(\mu_I - 3\sigma_I^2) S_I^{jk}. \tag{50}
\end{aligned}$$

Following exactly similar steps, but by interchanging the subscripts O and I , we can show that

$$\begin{aligned}
Q_{22} &= \sigma_O^4 N^2 P_O^{jj}(0) \theta_O \theta'_O P_O^{kk}(0) \\
&+ N \sigma_O^2 \sum_{\ell=-\infty}^{\infty} \left\{ \sigma_O^2 \theta'_O P_O^{jk}(\ell) \theta_O + \sigma_I^2 \theta'_I P_I^{jk}(\ell) \theta_I \right\} P_O^{jk}(\ell) \\
&+ N \sigma_O^4 \sum_{\ell=-\infty}^{\infty} P_O^{jk}(\ell) \theta_O \theta'_O P_O^{jk}(\ell) \\
&+ N(\mu_O - 3\sigma_O^2) S_O^{jk}.
\end{aligned}$$

Finally, by combining (46), (47), (50) and (51) we get

$$\begin{aligned}
&\mathcal{E} \left\{ \sum_{\ell_1=1}^N \sum_{\ell_2=1}^N \tilde{\phi}_j(t_{\ell_1}) \tilde{\phi}'_j(t_{\ell_1}) \bar{\theta} \bar{\theta}' \tilde{\phi}_k(t_{\ell_2}) \tilde{\phi}'_k(t_{\ell_2}) \right\} \\
&= N^2 P^{jj}(0) \Sigma \bar{\theta} \bar{\theta}' \Sigma P^{kk}(0) \\
&+ N \sum_{\ell=-\infty}^{\infty} \bar{\theta}' P^{jk}(\ell) \Sigma \bar{\theta} P^{jk}(\ell) \Sigma \\
&+ N \sum_{\ell=-\infty}^{\infty} P^{jk}(\ell) \Sigma \bar{\theta} \bar{\theta}' \Sigma P^{jk}(\ell) \\
&+ N(M - 3\Sigma^2) S^{jk},
\end{aligned}$$

which along with (44) and (45) gives

$$\begin{aligned}
\mathcal{E} \{ \Phi_j \bar{\theta} \bar{\theta}' \Phi'_k \} &= N^2 P^{jj}(0) \Sigma \bar{\theta} \bar{\theta}' \Sigma P^{kk}(0) \\
&+ N \sum_{\ell=-\infty}^{\infty} \bar{\theta}' P^{jk}(\ell) \Sigma \bar{\theta} \{ R^{jk}(\ell) + P^{jk}(\ell) \Sigma \} \\
&+ N \sum_{\ell=-\infty}^{\infty} P^{jk}(\ell) \Sigma \bar{\theta} \bar{\theta}' \Sigma P^{jk}(\ell) \\
&+ N(M - 3\Sigma^2) S^{jk}. \tag{51}
\end{aligned}$$

Now combining (38), (43) and (51) we get the result in Theorem 2 after a few steps of straightforward calculations.

B Proof of lemma 1

By Kronecker's lemma [6, p.129], it is sufficient to prove that χ_N is a Cauchy sequence with probability one, where

$$\chi_N = \sum_{k=0}^N \frac{1}{t_k} \varphi_k v_k.$$

In that goal, using Tchebychev's inequality [16] we have

$$\begin{aligned} & \text{Prob} \{ |\chi_{k+\ell} - \chi_k| \geq \epsilon \} \\ & \leq \frac{1}{\epsilon^2} \mathcal{E} (\chi_{k+\ell} - \chi_k)^2 \\ & = \frac{1}{\epsilon^2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \frac{\varphi_{k+i} \varphi_{k+j}}{t_{k+i} t_{k+j}} \mathcal{E} \{ v_{k+i} v_{k+j} \} \\ & \leq \frac{B_1^2 B}{\epsilon^2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \frac{r^{|i-j|}}{t_{k+i} t_{k+j}}. \end{aligned}$$

Now t_k is a sequence of monotonically increasing positive numbers such that $\lim_{k \rightarrow \infty} t_k = \infty$, and

$$0 < h_* := \min_k t_{k+1} - t_k$$

exists. Consequently, $t_{k+i} > t_k + ih_*$, and we have

$$\begin{aligned} & \text{Prob} \{ |\chi_{k+\ell} - \chi_k| \geq \epsilon \} \\ & \leq \frac{B_1^2 B}{\epsilon^2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \frac{r^{|i-j|}}{(t_k + ih_*)(t_k + jh_*)} \\ & < \frac{B_1^2 B(1+r)}{\epsilon^2(1-r)} \sum_{i=1}^{\ell} \frac{1}{(t_k + ih_*)^2}. \end{aligned}$$

The last inequality follows because $\frac{1+r}{1-r}$ is the maximum eigenvalue of the symmetric Toeplitz matrix T , where $T_{ij} = r^{|i-j|}$. This fact follows, since $\frac{1+r}{1-r}$ is the supremum of the spectrum of the process having an autocorrelation $r^{|k|}$ at lag k [21, p.45], see also [10]. Now for any given ϵ and ℓ the right hand side of (52) can be made arbitrarily small by choosing a sufficiently large k . Therefore, the lemma follows.

C Proof of Lemma 2

Since the initial conditions decay exponentially with time, they do not affect the asymptotic results. Hence we assume that the initial conditions are zero without any loss of generality. Then by applying (12) repeatedly it is easy to verify that

$$\phi_I(t_k) = \sum_{\ell=0}^{k-1} c_I B_I(k, \ell + 1) b_I(\ell) u(t_\ell), \quad (52)$$

where we define

$$B_I(k, \ell) = \begin{cases} A_I(k-1)A_I(k-2)\cdots A_I(\ell), & k > \ell; \\ I, & k = \ell \end{cases}$$

In exactly similar manner we can also derive an analogous relationship from (17) for the output side regressor $\phi_O(t_k)$. This can also be done by replacing the subscript I by O in the above equations. To simplify the notations we denote

$$\begin{aligned} G_I(k, \ell) &:= c_I B_I(k, \ell + 1) b_I(\ell), \\ G_O(k, \ell) &:= c_O B_O(k, \ell + 1) b_O(\ell). \end{aligned}$$

If the ZOH assumption is used in the input side it is straightforward to verify that [see (14)]

$$B_I(k, \ell) = e^{A(t_k - t_\ell)},$$

while for the FOH assumption from (16) it follows that

$$B_I(k, \ell) = \begin{bmatrix} e^{A(t_k - t_\ell)} & e^{A(t_k - t_{\ell+1})} \check{b}_\ell \\ 0 & 0 \end{bmatrix}.$$

Recall that $-\lambda$ is the real part of the eigenvalue of A [see (9)] located nearest to the imaginary axis. Note that $\lambda > 0$. Then for both the FOH and the ZOH assumptions there exists a vector m_I of bounded and positive real numbers such that (see also the statement of Theorem 1)

$$|G_I(k, \ell)| \preceq m_I e^{-\lambda(t_k - t_\ell)} \preceq m_I e^{-\lambda h_\star(k - \ell)}, \quad k \geq \ell, \quad (53)$$

where we use the notation \preceq to denote the element-wise inequality between two matrices, *i.e.*

$$\alpha \preceq \beta \quad \Leftrightarrow \quad |\alpha_{jk}| \leq |\beta_{jk}|.$$

Lemma 2 is the immediate consequences of (53). Firstly, (53) ensures that $|\dot{\phi}(t_k)|$ is a bounded sequence for all k provided $\dot{u}(t_k)$ and $\dot{y}(t_k)$ are bounded. Secondly, it ensures that the auto correlation sequence of $\tilde{\phi}'_I(t_k)$ decays exponentially. To see the details, consider $\tau > 0$. We have

$$\begin{aligned} & \mathcal{E} \{ \tilde{\phi}_I(t_{k+\tau}) \tilde{\phi}'_I(t_k) \} \\ &= \sum_{\ell_1=0}^{k+\tau-1} \sum_{\ell_2=0}^{k-1} G_I(k+\tau, \ell_1) \mathcal{E} \{ \tilde{u}(t_{\ell_1}) \tilde{u}(t_{\ell_2}) \} G'_I(k, \ell_2) \\ &= \sigma_I^2 \sum_{\ell_1=0}^{k+\tau-1} \sum_{\ell_2=0}^{k-1} G_I(k+\tau, \ell_1) \delta_{\ell_1, \ell_2} G'_I(k, \ell_2) \\ &= \sigma_I^2 \sum_{\ell=0}^{k-1} G_I(k+\tau, \ell) G'_I(k, \ell). \end{aligned} \quad (54)$$

Now using (53) we see that

$$\begin{aligned} \left| \mathcal{E} \{ \tilde{\phi}_I(t_{k+\tau}) \tilde{\phi}'_I(t_k) \} \right| &\leq \sigma_I^2 m_I m'_I e^{-\lambda h_\star(2k+\tau)} \sum_{\ell=0}^{k-1} e^{2\lambda h_\star \ell} \\ &= \frac{\sigma_I^2 m_I m'_I e^{-\lambda h_\star \tau}}{e^{2\lambda h_\star} - 1} \left[1 - e^{-2\lambda h_\star k} \right] \\ &\leq < \frac{\sigma_I^2 m_I m'_I}{e^{2\lambda h_\star} - 1} e^{-\lambda h_\star \tau}. \end{aligned} \quad (55)$$

With this the proof is complete.

D Proof of Lemma 3

In this appendix we need some new notations. The element at j -th row and ℓ -th column of the matrix $\alpha(t_k)$ is denoted by $\alpha_{j\ell}(k)$. The element at the j -th row of the vector $\phi_I(t_k)$ is denoted by $z_j(t_k)$. Similarly, the element at the j -th row of the time-varying impulse response sequence $c_I B_I(k, \tau + 1) b_I(\tau)$ is denoted by $g_k^{(j)}(\tau)$. Now from the definition of $\alpha_{j\ell}(k)$ we note that

$$\begin{aligned} & \mathcal{E} \{ \alpha_{j\ell}(k + \tau) \alpha_{j\ell}(k) \} \\ &= \mathcal{E} \{ z_j(t_{k+\tau}) z_\ell(t_{k+\tau}) z_j(t_k) z_\ell(t_k) \} \\ & \quad - \mathcal{E} \{ z_j(t_{k+\tau}) z_\ell(t_{k+\tau}) \} \mathcal{E} \{ z_j(t_k) z_\ell(t_k) \}. \end{aligned} \quad (56)$$

We evaluate the two terms in the right hand side of (56) separately. We need (52) in that goal. Using the definition of $g^{(j)}(k)$ for the second term in the right hand side of (56) we get, see (54),

$$\begin{aligned} & \mathcal{E} \{ z_j(t_{k+\tau}) z_\ell(t_{k+\tau}) \} \mathcal{E} \{ z_j(t_k) z_\ell(t_k) \} \\ &= \sigma_I^4 \sum_{\tau_1=0}^{k+\tau-2} \sum_{\tau_2=0}^{k+\tau-2} \sum_{\tau_3=0}^{k-2} \sum_{\tau_4=0}^{k-2} \left\{ g_{k+\tau}^{(j)}(\tau_1) g_{k+\tau}^{(\ell)}(\tau_2) \times \right. \\ & \quad \left. g_k^{(j)}(\tau_3) g_k^{(\ell)}(\tau_4) \delta_{\tau_1, \tau_2} \delta_{\tau_3, \tau_4} \right\}. \end{aligned} \quad (57)$$

Now we consider the first term in the right hand side of (56). We use (52) again. We have

$$\begin{aligned} & \mathcal{E} \{ z_j(t_{k+\tau}) z_\ell(t_{k+\tau}) z_j(t_k) z_\ell(t_k) \} \\ &= \sum_{\tau_1=0}^{k+\tau-2} \sum_{\tau_2=0}^{k+\tau-2} \sum_{\tau_3=0}^{k-2} \sum_{\tau_4=0}^{k-2} g_{k+\tau}^{(j)}(\tau_1) g_{k+\tau}^{(\ell)}(\tau_2) \times \\ & \quad g_k^{(j)}(\tau_3) g_k^{(\ell)}(\tau_4) \mathcal{E} \{ \tilde{u}(t_{k+\tau-\tau_1-1}) \tilde{u}(t_{k+\tau-\tau_2-1}) \times \\ & \quad \tilde{u}(t_{k-\tau_3-1}) \tilde{u}(t_{k-\tau_4-1}) \}. \end{aligned} \quad (58)$$

Now, we evaluate the fourth order moment in the right hand side of (58). We use a standard technique [25, p.572]:

$$\begin{aligned} & \mathcal{E} \{ \tilde{u}(t_{k+\tau-\tau_1-1}) \tilde{u}(t_{k+\tau-\tau_2-1}) \tilde{u}(t_{k-\tau_3-1}) \tilde{u}(t_{k-\tau_4-1}) \} \\ &= \sigma_I^4 [\delta_{\tau_1, \tau_2} \delta_{\tau_3, \tau_4} + \delta_{\tau_1-\tau, \tau_3} \delta_{\tau_2-\tau, \tau_4} + \delta_{\tau_1-\tau, \tau_4} \delta_{\tau_2-\tau, \tau_3}] \\ & \quad + [\mu_I - 3\sigma_I^4] \delta_{\tau_1-\tau, \tau_4} \delta_{\tau_2-\tau, \tau_4} \delta_{\tau_3, \tau_4}. \end{aligned} \quad (59)$$

Next we combine (54), (57), (58) and (59). Straightforward algebraic manipulation gives

$$\begin{aligned} & \mathcal{E} \{ \alpha_{j\ell}(k + \tau) \alpha_{j\ell}(k) \} \\ &= \sum_{\tau_1=0}^{k+\tau-2} \sum_{\tau_2=0}^{k+\tau-2} \sum_{\tau_3=0}^{k-2} \sum_{\tau_4=0}^{k-2} g_{k+\tau}^{(j)}(\tau_1) g_{k+\tau}^{(\ell)}(\tau_2) g_k^{(j)}(\tau_3) \times \\ & \quad g_k^{(\ell)}(\tau_4) \left\{ \sigma_I^4 \delta_{\tau_1-\tau, \tau_3} \delta_{\tau_2-\tau, \tau_4} + \sigma_I^4 \delta_{\tau_1-\tau, \tau_4} \delta_{\tau_2-\tau, \tau_3} \right. \\ & \quad \left. + [\mu_I - 3\sigma_I^4] \delta_{\tau_1-\tau, \tau_4} \delta_{\tau_2-\tau, \tau_4} \delta_{\tau_3, \tau_4} \right\} \\ &= T_1 + T_2 + T_3, \end{aligned} \quad (60)$$

$$(61)$$

where the terms T_1 , T_2 and T_3 corresponds to the first, second and the third term inside the curly braces in the right hand side of the first equality in (61). Using a calculation similar to (55) it is straightforward

to show that

$$|T_1| \leq \sigma_I^4 \left\{ \sum_{\tau_3=0}^{k-2} |g_{k+\tau}^{(j)}(\tau + \tau_3)g_k^{(j)}(\tau_3)| \right\} \times \left\{ \sum_{\tau_4=0}^{k-2} |g_{k+\tau}^{(\ell)}(\tau + \tau_4)g_k^{(\ell)}(\tau_4)| \right\} < \bar{B}_a e^{-2\lambda h_* \tau}.$$

for some bounded positive-valued \bar{B}_a . Similarly, for some bounded \bar{B}_b we have

$$|T_2| \leq \sigma_I^4 \left\{ \sum_{\tau_4=0}^{k-2} |g_{k+\tau}^{(j)}(\tau + \tau_4)g_k^{(\ell)}(\tau_4)| \right\} \times \left\{ \sum_{\tau_3=0}^{k-2} |g_{k+\tau}^{(\ell)}(\tau + \tau_3)g_k^{(j)}(\tau_3)| \right\} < \bar{B}_b e^{-2\lambda h_* \tau}.$$

Also, there exists a bounded real number \bar{B}_c such that

$$|T_3| \leq [\mu_I - 3\sigma_I^2] \times \sum_{\tau_4=0}^{k-2} |g_{k+\tau}^{(j)}(\tau + \tau_4)g_{k+\tau}^{(\ell)}(\tau + \tau_4)g_k^{(j)}(\tau_4)g_k^{(\ell)}(\tau_4)| < \bar{B}_c e^{-2\lambda h_* \tau}.$$

Combining the above observations we see that

$$|\mathcal{E} \{ \alpha_{j\ell}(k + \tau)\alpha_{j\ell}(k) \}| < B_d e^{-2\lambda h_* \tau}$$

for some bounded positive valued B_d , and the required result is proved.

References

- [1] B. D. O. Anderson and M. Deistler. Identifiability of dynamic errors-in-variables models. *J. Time Series Analysis*, 5:1–13, 1984.
- [2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, New Jersey, 1979.
- [3] S. Beghelli, R.P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26:171–176, 1990.
- [4] P. Castaldi and U. Soverini. Identification of dynamic errors-in-variables models. *Automatica*, 32(4):631–636, April 1996.
- [5] C.T. Chou, M. Verhaegen, and R. Johansson. Continuous-time identification of SISO systems using Laguerre functions. *IEEE Transactions on Signal Processing*, 47(2):349–362, 1999.
- [6] K. L. Chung. *A course in probability theory*. Academic Press, San Diego, CA, USA, 3 edition, 2001.
- [7] H. Garnier, G.C. Goodwin, and J. Welsh. A time-domain approach to continuous-time model identification of highly resonant wide-band systems. In *3rd IFAC Symposium on Mechatronic Systems*, Sydney (Australia), September 2004.

- [8] H. Garnier, M. Mensler, and A. Richard. Continuous-time model identification from sampled data. Implementation issues and performance evaluation. *International Journal of Control*, 76:13:1337–1357, 2003.
- [9] H. Garnier and P.C. Young. Time-domain approaches for continuous-time model identification from sampled data. In *Invited tutorial paper for the American Control Conference (ACC'2004)*, Boston, MA (USA), June 2004.
- [10] R. M. Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Transactions on Information Theory*, IT-18:6:725–730, 1972.
- [11] E. Karlsson, T. Söderström, and P. Stoica. The Cramér-Rao lower bound for noisy input output systems. *Signal Processing*, 80(11):2421–2447, November 2000.
- [12] L. Ljung. Some limit results for functionals of stochastic processes. Technical report, LiTH-ISY-I-0167, Department of Electrical Engineering, Linköping University, Sweden, 1977.
- [13] L. Ljung. *System Identification - Theory for the User, 2nd edition*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.
- [14] I. Markovski, J.C. Willems, and B. De Moor. Continuous-time errors-in-variables filtering. In *41st IEEE Conference on Decision and Control (CDC'02)*, pages 2576–2581, Las-Vegas, Nevada (USA), December 2002.
- [15] R. H. Middleton and G. C. Goodwin. *Digital Control and Estimation: A unified approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1990.
- [16] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
- [17] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, New York, 2001.
- [18] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume. Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, 33:6:1073–1086, June 1997.
- [19] N. K. Sinha and G.P. Rao, editors. *Identification of continuous-time systems: Methodology and computer implementation*. Kluwer Academic Publishers, Dordrecht, 1991.
- [20] T. Söderström. Identification of stochastic linear systems in presence of input noise. *Automatica*, 17:713–725, 1981.
- [21] T. Söderström. *Discrete-time Stochastic Systems, 2 ed.* Springer-Verlag, London, 2002.
- [22] T. Söderström, M. Hong, and W. X. Zheng. Convergence properties of bias-eliminating algorithms for errors-in-variables identification. *International Journal of Adaptive Control and Signal Processing*, 19:703–722, November 2005.
- [23] T. Söderström and K. Mahata. On instrumental variable and total least squares approaches for identification of noisy systems. *International Journal of Control*, 75(6):381–389, April 2002.

- [24] T. Söderström, U. Soverini, and K. Mahata. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing*, 82:1139–1154, 2002.
- [25] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, Hemel Hempstead, UK, 1989.
- [26] P. Stoica, M. Cedervall, and A. Eriksson. Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. *IEEE Transactions on Signal Processing*, 43:2386–2397, 1995.
- [27] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 2004.
- [28] P. Stoica, R. Moses, B. Friedlander, and T. Söderström. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37:3:378–392, March 1989.
- [29] J. K. Tugnait. Stochastic system identification with noisy input using cumulant statistics. *IEEE Transactions on Automatic Control*, AC-37:476–485, 1992.
- [30] J. K. Tugnait and Y. Ye. Stochastic system identification with noisy input-output measurement using polyspectra. *IEEE Transactions on Automatic Control*, AC-40:670–683, 1995.
- [31] H. Unbehauen and G.P. Rao. *Identification of continuous systems*. North-Holland, Amsterdam, 1987.
- [32] H. Unbehauen and G.P. Rao. Continuous-time approaches to system identification - a survey. *Automatica*, 26:1:23–35, 1990.
- [33] H. Unbehauen and G.P. Rao. A review of identification in continuous-time systems. *Annual reviews in Control*, 22:145–171, 1998.
- [34] P. Van Hamme, R. Pintelon, and J. Schoukens. *Identification of continuous-time systems. Methodology and computer implementation*, chapter Discrete-time modeling and identification of continuous-time systems: a general framework, pages 17–77. Kluwer Academic Publishers, N.K. Sinha and G.P. Rao edition, 1991.
- [35] P. C. Young. In flight dynamic checkout - a discussion. *IEEE Transactions on Aerospace*, 2:1106–1111, 1964.
- [36] P. C. Young. The determination of the parameters of a dynamic process. *Radio and Electronic Engineering (Journal of IERE)*, 29:345–361, 1965.
- [37] P. C. Young. Theory of self adaptive control systems. In P. H. Hammond, editor, *Process parameter estimation and self adaptive control*, pages 118–140. Plenum Press, 1966.
- [38] P. C. Young. Parameter estimation for continuous-time models - a survey. *Automatica*, 17:1:23–39, 1981.
- [39] P.C. Young and H. Garnier. Identification and estimation of continuous-time data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling and Software*, Accepted for publication, to be published in 2006.

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