Least Squares Parameter Estimation of Continuous-Time ARX Models from Discrete-Time Data

Torsten Söderström, *Fellow, IEEE*, H. (Howard) Fan, *Senior Member, IEEE*, Bengt Carlsson, *Member, IEEE*, and Stefano Bigi

***Abstract*—When modeling a system from discrete-time data, a continuous-time parameterization is desirable in some situations. In a direct estimation approach, the derivatives are approximated by appropriate differences. For an ARX model this lead to a linear regression. The well-known least squares method would then be very desirable since it can have good numerical properties and low computational burden, in particular for fast or nonuniform sampling. It is examined under what conditions a least squares fit for this linear regression will give adequate results for an ARX model. The choice of derivative approximation is crucial for this approach to be useful. Standard approximations like Euler backward or Euler forward cannot be used directly. The precise conditions on the derivative approximation are derived and analyzed. It is shown that if the highest order derivative is selected with care, a least squares estimate will be accurate. The theoretical analysis is complemented by some numerical examples which provide further insight into the choice of derivative approximation.**

***Index Terms*— Autoregressive processes, bias compensation, continuous-time stochastic models, derivative approximation, least squares method, linear regression, time series.**

I. Introduction

PARAMETER estimation or system identification of continuous-time systems is an important subject which has numerous applications ranging from control and signal processing, to astrophysics and economics [2], [3], [5]–[8], [11], [16], [19]. This is because most physical systems or phenomena are continuous time in nature, for example in many control applications [5], [11], [16], in astrophysics [8], and in economics [3], [7]. Due to the advent of digital computers, research for control and identification of these continuous-time systems and processes has concentrated on their discretized models with samples from the underlying continuous-time system inputs and outputs. Recently, interest in identification of continuous-time systems and processes has arisen (see the above references). One particularly interesting and practical scenario is the identification of continuous-time systems using discrete data.

Manuscript received September 8, 1995; revised June 10, 1996. Recommended by Associate Editor, V. Solo. This work was supported in part by the Swedish Research Council for Engineering Sciences under Contract 95-800. The fourth author was supported by the Office of Naval Research under Grant N00014-90-J-1017 and by the National Science Foundation under Grant INT-9321821.

T. Söderström, B. Carlsson, and S. Bigi are with the Systems and Control Group, Department of Technology, Uppsala University, Uppsala, Sweden.

H. Fan is with the Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, OH 45221 USA.

Publisher Item Identifier S 0018-9286(97)03436-3.

The objective of this paper is to investigate the identification of continuous-time ARX processes by least squares (LS) using discrete data and a "direct approach." In such an approach the derivatives are substituted by discrete-time differences and the model casted into a (discrete-time) linear regression. Some possible advantages of a direct approach include the following.

- It may have good numerical properties, especially for fast sampling [17].
- The principle may be modified to be used in the case of nonuniform sampling also.
- It is computationally efficient, in particular if the sampling period is nonuniform and/or an order-recursive scheme may be used.

An alternative way of identifying the process might be, for a fixed sampling interval h, to obtain an equivalent discrete-time model. Due to the sampling effects, the disturbance part of the sampled model will be an autoregressive moving average (ARMA) process [12]. All the ARMA parameters could be estimated by using a prediction error method (PEM) [15]. This common methodology can be interpreted as a maximum-likelihood technique; its implementation may be done using a Kalman filter representation. Alternatively, a modified Yule–Walker method might be used to estimate the AR parameters of the ARMA process. Once the AR parameters are known, the corresponding continuous-time parameters could be found by reversing the sampling transformation. For the case of short sampling intervals, this approach will give an ill-conditioned numerical problem, as all discrete-time poles cluster around the point $z = 1$ [6]. The method under study in this paper has a limit as $h \to 0$, which is not the case for the PEM approach. In addition, it is simple to apply. The associated optimization problem (linear LS) has a unique minimum point which can be expressed in analytical form. In contrast, the PEM approach relies on a numerical search procedure with a risk to get stuck in local minima. It will also require a larger amount of computation. Still, a possible approach for identifying the system would be to set up a linear continuous-time state-space model, where the system matrices depend on the unknown ARX parameters. Next an extended and *nonlinear* model can be formulated by adding the parameters as additional states. A state estimator, such as an extended Kalman filter, may then be applied to estimate the parameters from the discrete-time data records. Note, however, that as the underlying state-space model is nonlinear, there is no straightforward way to analyze convergence or other properties of the estimates. A further motivation for the study is that we provide a careful analysis of an approach that is available anyway in the literature.

In this paper we analyze an LS fit to estimate the parameters of a continuous-time ARX process when a direct approach is used and the process operates in open loop. In [4] and [13], we considered and analyzed the use of an instrumental variable method for the parameter estimation. It turned out that by using appropriately delayed outputs as instrumental variables, it is possible to get accurate parameter estimates in the case of fast sampling. The choice of derivative approximation has a significant influence on the estimates. In contrast, for the LS method, it turns out that the conditions on the derivative approximations for the estimates to become reasonable are crucial, as standard approximations *cannot* be used for the highest derivative. However, it is still possible to use some simple schemes and achieve accurate results. Based on the analysis, we present some simple LS schemes which give accurate parameter estimates.

The paper is organized as follows. In the next section, we present the necessary setup and introduce key notions. Section III is devoted to statements of the problem to be examined. Analysis, under which the bias of the parameter estimates is negligible for short sampling periods, is examined in Section IV. In Section V we describe noise level estimation and a scheme for model-order determination. It is possible to arrange the computations in a numerically efficient scheme that works order-recursively. This is done in Section VI. Finally, several numerical examples illustrating the results of the paper can be found in Section VII, while conclusions are given in Section VIII. Some further details and results related to the paper can be found in the technical report [14].

II. Background

Consider a continuous-time ARX process

$$\begin{aligned}(p^n + a_1 p^{n-1} + \cdots + a_n) y(t) &= (b_1 p^{n-1} + \cdots + b_n)\\ &\quad \cdot u(t) + e(t)\\ Ee(t)e(s) &= \sigma^2 \delta(t-s)\end{aligned} \tag{1}$$

where p denotes the differentiation operator, $y(t)$ is the output, $u(t)$ the input, and $e(t)$ a (continuous-time) white noise source. The time series is observed in discrete-time at $t = h, 2h, 3h, \cdots Nh$. The model order n is supposed to be known. In Section V, we will discuss a way to estimate n. It is of interest to estimate the parameter vector

$$\theta = (a_1 \cdots a_n b_1 \cdots b_n)^T \tag{2}$$

from available data. Possibly, the noise intensity σ^2 is to be estimated. If this can be done, a parametric estimate of the spectral density of the noise part of $y(t)$ is obtained as

$$\hat{\phi}(\omega) = \frac{\hat{\sigma}^2}{|(i\omega)^n + \hat{a}_1 (i\omega)^{n-1} + \cdots + \hat{a}_n|^2}. \tag{3}$$

As a general linear approximation of the differentiation operator p, we will consider

$$p^k f(t) \approx D^k f(t) \triangleq \frac{1}{h^k} \sum_j \beta_{k,j} f(t + jh) \tag{4}$$

where $p^k f(t)$ denotes the kth-order derivative of a smooth function $f(t)$ and $\{\beta_{k,j}\}$ are some weights. The summation limits in (4) may depend on k. Needless to say, some conditions must be imposed on the weights $\{\beta_{k,j}\}$ to make the approximations meaningful.

We introduce

$$D^k f(t) = p^k f(t) + O(h), \qquad k = 0, \cdots, n \tag{5}$$

which we will refer to as the *natural conditions*. Assume that $f(t)$ is $k+1$ times differentiable and consider the case of a short sampling period h. By series expansion, we then have

$$\begin{aligned}D^k f(t) &= \frac{1}{h^k} \sum_j \beta_{k,j} f(t + jh)\\ &= \frac{1}{h^k} \sum_j \beta_{k,j} \left[\sum_{\nu=0}^{k} \frac{1}{\nu!} j^\nu h^\nu p^\nu f(t) + O(h^{k+1}) \right]\\ &= \sum_{\nu=0}^{k} \frac{1}{\nu!} p^\nu f(t) \left[\sum_j \beta_{k,j} j^\nu \right] h^{\nu-k} + O(h).\end{aligned} \tag{6}$$

The *natural conditions* (5) can now be expressed as

$$\sum_j \beta_{k,j} j^\nu = \begin{cases} 0, & \nu = 0, \cdots, k-1 \\ k!, & \nu = k. \end{cases} \tag{7}$$

The frequency function of the filter will be

$$\begin{aligned}D^k(e^{i\omega}) &= \frac{1}{h^k} \sum_j \beta_{k,j} e^{ij\omega}\\ &= \frac{1}{h^k} \sum_j \beta_{k,j} \sum_{\nu=0}^{\infty} \frac{(i\omega)^\nu}{\nu!} j^\nu\\ &= (i\omega)^k + O(|\omega|^{k+1}).\end{aligned} \tag{8}$$

The low-frequency asymptote of the filter frequency function is hence $(i\omega)^k$.

The minimal number of terms in (7) is apparently $k+1$. In such a case D^k will be a high-pass filter. If the number of $\beta_{k,j}$ coefficients is increased, the gained degrees of freedom can be used to decrease the high frequency gain of the filter D^k, for example by imposing

$$D^k(e^{i\omega})_{|\omega=\pi} = 0. \tag{9}$$

In addition to (9), it is possible to require some derivatives of $D^k(e^{i\omega})$ to vanish at $\omega = \pi$. Such measures would make the filter bandpass instead of high-pass and hence more robust to unmodeled wideband noise.

After substituting derivatives by approximations in (1), the following model:

$$\begin{aligned}w(t) &= \varphi^T(t)\theta + \varepsilon(t)\\ w(t) &= D^n y(t)\\ \varphi^T(t) &= [-D^{n-1}y(t) \cdots - D^0 y(t) \quad D^{n-1}u(t)\\ &\qquad \cdots D^0 u(t)]\end{aligned} \tag{10}$$

can be constructed, where $\varepsilon(t)$ is an equation error. This model can be viewed as a (discrete-time) linear regression.

In [4] and [13], we analyzed the instrumental variable (IV) estimate (in the pure time series case with no input present)

$$\hat{\theta}_N = [\Sigma z(t)\varphi^T(t)]^{-1}[\Sigma z(t)w(t)] \tag{11}$$

where the IV vector $z(t)$ consists of delayed observations of $y(t)$. It was shown that the parameter estimate $\hat{\theta}_N$ is close to the true value if:

- the number of data is large;
- the sampling period is small;
- the output values appearing in $z(t)$ are appropriately delayed. (The minimal delay depends on the summation index chosen in (4).)

It is required that the sampling period is relatively short. This is a fairly natural condition as the approximation of the derivatives (4) is constructed to be reasonable for the case $h \to 0$. Further, it turned out that even if the choice of weights $\{\beta_{k,j}\}$ can be made freely, subject to the natural conditions and matched to the instrumental delay in $z(t)$, the resulting accuracy is highly influenced by such choices.

It is of interest to examine in some depth if and how an LS approach

$$\hat{\theta}_N = [\Sigma\varphi(t)\varphi^T(t)]^{-1}[\Sigma\varphi(t)w(t)] \tag{12}$$

for estimating θ can give feasible results. This is the topic of the current paper. We will consider the asymptotic case when the number of data, N, tends to infinity. Then the parameter estimate becomes, c.f., the ergodicity results of [15]

$$\hat{\theta} \triangleq \lim_{N\to\infty} \hat{\theta}_N = [E\varphi(t)\varphi^T(t)]^{-1}[E\varphi(t)w(t)] \quad \text{w.p.1.} \tag{13}$$

Note that Zhao *et al.* have studied the LS methods for a similar problem with a slightly different model (the output error type); see [20] and [21], using an integral filter approach. Since their LS estimates are biased due to noise, an additional debiasing step is needed after the LS estimation. As will be shown later, our method guarantees asymptotic consistency, so no debiasing is needed. Furthermore, one can scale the elements of (13) by defining

$$\overline{D}^i f(t) = h^{n-1}D^i f(t) = h^{n-1-i}\sum_j \beta_{i,j} f(t+jh) \tag{14}$$

so that (13) becomes

$$\hat{\theta} = [E\overline{\varphi}(t)\overline{\varphi}^T(t)]^{-1}[E\overline{\varphi}(t)\overline{w}(t)] \quad \text{w. p. 1} \tag{15}$$

where

$$\begin{aligned}\overline{\varphi}(t) &= [-\overline{D}^{n-1}y(t)\cdots - \overline{D}^0 y(t) \quad \overline{D}^{n-1}u(t)\cdots\overline{D}^0u(t)]\\ \overline{w}(t) &= \overline{D}^n y(t).\end{aligned} \tag{16}$$

Now the only difference between (14) and the integral filter of [21] is that (14) is a high-pass filter and that of [21] is a low-pass filter. Since in our model the noise enters y in a low-pass fashion, the high-pass nature of (14) does not amplify the noise. Nevertheless, for the robustness of our approach, we can increase the length of our filter (14) as described previously. Any unmodeled wideband noise on y can be suppressed by the new filter. This type of filter may be called integral-differential filters. Since incorporation of the additional linear constraint (9) is straightforward, in the sequel we will only concentrate on (7) to make our presentation clearer.

In the literature on continuous-time identification, a common approach is to rewrite (1) into an equivalent integral equation and to approximate the integrals [10], [16]. For the approximation, the use of block-pulse functions is popular. As noted in [10], this would correspond to the substitution

$$p \to \frac{2}{h}\cdot\frac{q-1}{q+1}. \tag{17}$$

It will hence appear as a *special case* of the general framework given here. In particular, for small h, it will behave as an Euler forward method, c.f., Example 3.2.

Now rewrite the output $y(t)$ as a sum of a deterministic and a stochastic term

$$y(t) = \frac{B(p)}{A(p)}u(t) + \frac{1}{A(p)}e(t) \triangleq y_u(t) + y_e(t). \tag{18}$$

We assume the process operates in open loop, so the two terms in (18) are independent. Similarly, we decompose the regressor vector $\varphi(t)$

$$\begin{aligned}\varphi(t) &= (-D^{n-1}y(t)\cdots - D^0y(t) \quad D^{n-1}u(t)\cdots D^0u(t))^T\\ &= (-D^{n-1}y_u(t)\cdots - D^0y_u(t) \quad D^{n-1}u(t)\cdots D^0u(t))^T\\ &\quad + (-D^{n-1}y_e(t)\cdots - D^0y_e(t) \quad 0\cdots 0)^T\\ &\triangleq \varphi_u(t) + \varphi_e(t).\end{aligned} \tag{19}$$

Assuming that $u(t)$ is sufficiently differentiable, the regressor $\varphi_u(t)$ has a limit as $h \to 0$

$$\begin{aligned}\lim_{h\to 0}\varphi_u(t) &= (-p^{n-1}y_u(t)\cdots - y_u(t) \; p^{n-1}u(t)\cdots u(t))^T\\ &\triangleq \tilde{\varphi}_u(t).\end{aligned} \tag{20}$$

Hence for the deterministic part we have

$$\lim_{h\to 0}\lim_{N\to\infty}\hat{\theta} = \lim_{h\to 0}[E\varphi_u(t)\varphi_u^T(t)]^{-1}[E\varphi_u(t)w(t)] \tag{21}$$

$$\begin{aligned}&= [E\tilde{\varphi}_u(t)\tilde{\varphi}_u^T(t)]^{-1}[E\tilde{\varphi}_u(t)p^n y_u(t)]\\ &= [E\tilde{\varphi}_u(t)\tilde{\varphi}_u^T(t)]^{-1}[E\tilde{\varphi}_u(t)(\tilde{\varphi}_u^T(t)\theta)]\\ &= \theta.\end{aligned} \tag{22}$$

This means that the LS estimate is close to the true value of the parameter vector for large data sets and small sampling periods.

The stochastic part is more intricate to analyze. A main reason is that the derivative $p^n y_e(t)$ does not exist in a mean-square sense (it will not have finite variance). As will be shown later in the paper, by a careful choice of the weights $\{\beta_{k,j}\}$ though it is possible to estimate θ without any significant bias.

In order to satisfy the desired condition (22), it is sufficient that this condition applies to the stochastic part $y_e(t)$. In order to simplify the treatment, we therefore restrict the analysis to the stochastic part. Note that such a case is of interest by itself in continuous-time AR models of time series analysis. For examples in economics and astrophysics, see [3] and [8].

III. Statement of the Problem

As motivated above we will consider the stochastic part of the process (1). In order to show that an analysis of the LS estimate (13) is not a trivial extension, neither in the deterministic case nor the corresponding IV estimates, we repeat some simple examples from [4] and [13].

Example 3.1: Consider a first-order process, $n = 1$

$$(p + a)y(t) = e(t).$$

Now use an Euler backward approximation, or the backward delta operator

$$Dy(t) = \frac{1}{h}[y(t) - y(t - h)], \quad D^0 y(t) = y(t).$$

The asymptotic parameter estimate becomes

$$\begin{aligned}\hat{\theta} &= [E\varphi(t)\varphi^T(t)]^{-1}[E\varphi(t)w(t)] \\ &= -\frac{\frac{1}{h}E[y(t) - y(t-h)]y(t)}{Ey^2(t)} \\ &= \frac{1}{h}\left[-1 + \frac{r(h)}{r(0)}\right]\end{aligned}$$

where

$$r(\tau) = Ey(t + \tau)y(t) \tag{23}$$

denotes the covariance function of the process. For a first-order process

$$r(\tau) = \frac{\sigma^2}{2a}e^{-a|\tau|}$$

see, for example, [12]. Hence

$$\hat{\theta} = \frac{1}{h}[-1 + e^{-ah}] \approx -a + O(h).$$

The estimate is thus far from its true value, even for large N and small h. □

Example 3.2: Let us reconsider the first-order process of Example 3.1 but using an Euler *forward* approximation, that is, the forward delta operator

$$Dy(t) = \frac{1}{h}[y(t + h) - y(t)] \quad D^0 y(t) = y(t).$$

In this case

$$\begin{aligned}\hat{\theta} &= \frac{-\frac{1}{h}E[y(t+h) - y(t)]y(t)}{Ey^2(t)} = \frac{1}{h}\left[1 - \frac{r(h)}{r(0)}\right] \\ &= a + O(h).\end{aligned}$$

Hence the estimate is accurate for sufficiently large N and sufficiently small h. □

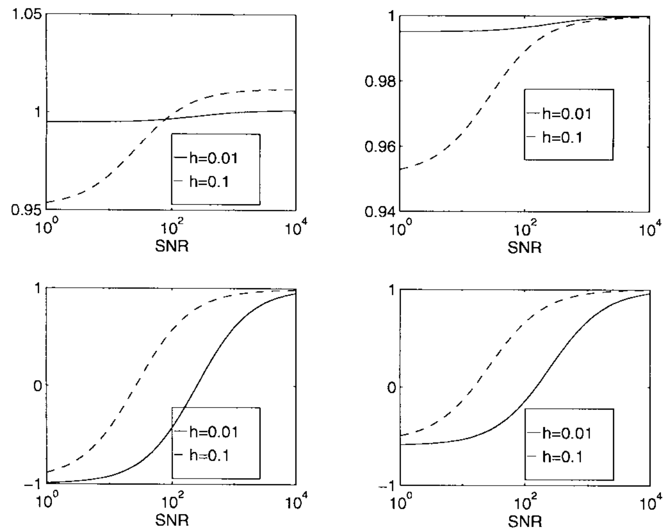

Fig. 1. Parameter estimates, Example 3.4 versus SNR. Parameter values used are $a = 1, b = 1, \omega = 0.5$. Upper diagrams—Euler forward, lower diagrams—Euler backward. Left diagrams—$\hat{a}$, right diagrams—$\hat{b}$.

Example 3.3: One might believe that a general recipe is to use Euler forward approximations. However, for a second-order system using

$$\begin{aligned}D^2 y(t) &= \frac{1}{h^2}[y(t + 2h) - 2y(t + h) + y(t)] \\ Dy(t) &= \frac{1}{h}[y(t + h) - y(t)] \quad D^0 y(t) = y(t)\end{aligned}$$

leads to (see [14] for technical details)

$$\begin{aligned}\hat{a}_1 &= \tfrac{2}{3}a_1 + O(h) \\ \hat{a}_2 &= a_2 + O(h).\end{aligned}$$

Hence, good quality of the estimates is not guaranteed by using Euler forward approximations. □

The above three examples illustrate that it is not trivial to find out if and how a LS approach can be used successfully. A more systematic treatment appears necessary to find out if the LS approach has any potentials.

In the following example we will show the impact of Euler forward and backward approximations for a system with input.

Example 3.4: Consider the first-order system

$$(p + a)y(t) = bu(t) + e(t)$$

where the input is a sinusoid of angular frequency ω, and $e(t)$ is continuous-time white noise. The asymptotic $(N \to \infty)$ parameter estimates of a and b were computed for Euler forward and Euler backward approximations. Two values of the sampling period were considered. The estimates were computed for several values of the signal-to-noise ratio (SNR), $\text{SNR} = Ey_u^2(t)/Ey_e^2(t)$. The results are displayed in Fig. 1.

It is clear from Fig. 1 that the Euler forward approximation gives estimates with a small bias. The bias decreases with decreasing h. When an Euler backward approximation is used, the bias is still small for high enough SNR but becomes very significant for moderate and small SNR. Note that the observations are completely in line with the findings in Examples 3.1 and 3.2. (In particular, $\hat{a} \approx 1$ for Euler forward, with $h = 0.01$, SNR small, and $\hat{a} \approx -1$ for Euler backward, with $h = 0.01$, SNR small.) □

We will examine (the stochastic part of) the properties of the LS estimate in several respects. Note that in the light of Example 3.3, the use of an Euler forward approximation does not give an accurate estimate for ARX models of order two or higher. The following problems will be managed.

1) First, we will investigate the properties of the limiting estimates (13). It is clearly influenced by the weights $\{\beta_{k,j}\}$. It is a modest requirement that the bias of the estimate vanishes when the sampling period approaches zero. This condition is mathematically expressed as

$$\lim_{h\to 0} \hat{\theta} = \lim_{h\to 0}[E\varphi(t)\varphi^T(t)]^{-1}E\varphi(t)w(t) = \theta_o. \quad (24)$$

Assuming $\hat{\theta}$ permits a series expansion in h, we can write this as

$$\hat{\theta} = \theta_o + O(h) \quad (25)$$

It is *a priori* not certain if any set of weights can be found such that this condition is satisfied. If (24) can be fulfilled, it is of interest both to characterize for what weights it is possible and to find simple ways to choose weights that satisfy the condition.

2) The asymptotic bias can be reduced further (for short sampling periods) if (25) can be substituted by

$$\hat{\theta} = \theta_o + O(h^2). \quad (26)$$

Again, it is of interest to characterize the set of possible weights and simple rules for selecting such weights.

3) A further problem concerns the possible estimation of σ^2. If this can be done, a parametric estimate of the (continuous-time) spectral density can be found; see (3).

4) Finally, it would be of interest to develop an order-estimation scheme, assuming that the true value of n is not known. Needless to say, in practice n is seldom *a priori* known.

IV. Analysis of the Asymptotic Bias

In this section, we will cope with the asymptotic bias. In particular, the behavior of the limiting estimate $\hat{\theta}$ for small sampling periods will be of interest. The detailed calculations are given in the Appendix, while the main results are stated here.

We will first state a technical result which is fundamental in the analysis.

Lemma 4.1: Let $y(t)$ be a continuous $\mathrm{AR}(n)$ process. Then, its covariance function $r(t)$ satisfies

$$p^{2k+1}r(t)_{|t=0^+} = 0 \qquad k = 0, 1, \cdots, n-2 \quad (27)$$

$$p^{2n-1}r(t)_{|t=0^+} = (-1)^n\frac{\sigma^2}{2}. \quad (28)$$

We then have the following result.

Lemma 4.2: Consider a stable continuous-time AR process. The following holds:

$$\hat{\theta} = \theta_o + O(h)$$

for arbitrary process parameters, if the weights $\{\beta_{k,j}\}$ satisfy the natural conditions (7), and

$$\sum_j \beta_{n-1,j}\sum_\ell \beta_{n,\ell}[|\ell - j|^{2n-1} - (\ell - j)^{2n-1}] = 0. \quad (29)$$

Proof: See Appendix A. □

Next, it is relevant to discuss if, and how, (29) can be satisfied. An obvious possibility would be to require

$$\ell \geq j. \quad (30)$$

This means that *all* measurements used when forming $D^n y(t)$ are at least as recent as those used when forming $D^{n-1}y(t)$. As an illustration, let us recapitulate Example 3.3, where $n = 2$, and choose now as before

$$D^0 y(t) = y(t), \qquad Dy(t) = \frac{1}{h}[y(t+h) - y(t)].$$

A simple modification, as compared to Example 3.3, is to further choose

$$D^2 y(t) = \frac{1}{h^2}[y(t+3h) - 2y(t+2h) + y(t+h)].$$

In such a way $\{\ell\} = \{1, 2, 3\}$ and $\{j\} = \{0, 1\}$. Hence, $\ell \geq j$ holds, and Lemma 4.2 is applicable. Needless to say, there are more possibilities to choose differentiation approximations to satisfy (30).

Note that the standard choices of integral filters and block pulse functions will *not* satisfy (29). These methods will hence suffer from the same bias problem as methods based on standard derivative approximation. However, by using (29), the user choices in these methods may be modified appropriately.

So far, we have examined how to choose the weights $\{\beta_{k,j}\}$ in order to achieve a bias of order $O(h)$. To reduce the bias further for short sampling periods, it can be of interest to explore the possibilities to achieve a bias of order $O(h^2)$. We then have the following result.

Lemma 4.3: Consider a stable continuous-time AR process. The following holds:

$$\hat{\theta} = \theta_o + O(h^2)$$

for arbitrary process parameters, if the weights $\{\beta_{k,j}\}$ satisfy (29) and

$$\sum_j \beta_{n-2,j}\sum_\ell \beta_{n,\ell}[|\ell - j|^{2n-1} - (\ell - j)^{2n-1}] = 0 \quad (31)$$

$$\sum_j \beta_{n-1,j}\sum_\ell \beta_{n-1,\ell}[|\ell - j|^{2n-1} - (\ell - j)^{2n-1}] = 0 \quad (32)$$

as well as the *extended* natural conditions

$$\sum_\ell \beta_{i,\ell}\ell^m = \begin{cases} 0, & \text{if } m < i \\ i!, & \text{if } m = i, \qquad i = 0, \cdots, n \\ 0, & \text{if } m = i+1. \end{cases} \quad (33)$$

Proof: See Appendix A. □

Remark: Note that under the extended natural conditions (33)

$$\sum_j \beta_{n-2,j} \sum_\ell \beta_{n,\ell}(\ell-j)^{2n-1} = 0$$
$$\sum_j \beta_{n-1,j} \sum_\ell \beta_{n-1,\ell}(\ell-j)^{2n-1} = 0.$$

□

The two conditions, (31) and (32), have different characteristics. It is easy to satisfy (31). A simple way is to choose the weights so that $\ell \geq j$ (measurements used for forming $D^n y(t)$ are all more recent than those employed for $D^{n-2}y(t)$). However, (32) is more tricky as it is a *quadratic* constraint. It may or may not have solutions. We illustrate this fact by an example.

Example 4.1: Consider the case $n-1=1$. To simplify notations, we drop the first index and write the extended natural conditions as

$$\sum_\ell \beta_\ell = 0 \quad \sum \ell\beta_\ell = 1 \quad \sum \ell^2\beta_\ell = 0 \tag{34}$$

while (32) becomes

$$\sum_j \sum_\ell \beta_j\beta_\ell|\ell-j|^3 = 0. \tag{35}$$

As there are four equations, we must have (at least) four different unknowns. Set

$$\beta = \begin{pmatrix} \beta_k \\ \beta_{k+1} \\ \beta_{k+2} \\ \beta_{k+3} \end{pmatrix}$$

where k is an integer (positive or negative) that remains to be chosen. The extended natural conditions (34) can be written as

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ k & k+1 & k+2 & k+3 \\ k^2 & (k+1)^2 & (k+2)^2 & (k+3)^2 \end{pmatrix} \beta = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

By tedious but straightforward calculations, it can be shown that the general solution to (34) can be written as

$$\beta = \frac{1}{20}\begin{pmatrix} -21 \\ 13 \\ 17 \\ -9 \end{pmatrix} + \frac{k}{2}\begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} + \gamma\begin{pmatrix} 1 \\ -3 \\ 3 \\ -1 \end{pmatrix} \tag{36}$$

where γ is arbitrary. Inserting (36) into (35) gives, again after some tedious calculations, the following equation for γ:

$$12\gamma^2 + 8.8\gamma + 5.53 + 15k + 5k^2 = 0. \tag{37}$$

It turns out to have (real-valued) solutions, if and only if $k=-1$ or $k=-2$. Furthermore, the solutions are

$$k=-1, \quad \gamma=-1.0787, \quad \beta = \begin{pmatrix} -1.6287 \\ 3.3860 \\ -2.8860 \\ 1.1287 \end{pmatrix}$$
$$k=-1, \quad \gamma=0.3453, \quad \beta = \begin{pmatrix} -0.2047 \\ -0.8860 \\ 1.3860 \\ -0.2953 \end{pmatrix}$$
$$k=-2, \quad \gamma=-1.0787, \quad \beta = \begin{pmatrix} -1.1287 \\ 2.8860 \\ -3.3860 \\ 1.6287 \end{pmatrix}$$
$$k=-2, \quad \gamma=0.3453, \quad \beta = \begin{pmatrix} 0.2953 \\ -1.3860 \\ 0.8860 \\ 0.2047 \end{pmatrix}.$$

Note the close relations between the first and the third solutions as well as between the second and the fourth. □

More general cases can, in principle, be approached in the same way. The extended natural conditions (34) can be rephrased as an underdetermined linear system of equations, $A\beta = b$, with the solution $\beta = \beta^{(0)} + B\gamma$. Here, B is a full rank matrix whose columns span the nullspace of A, and γ is an arbitrary vector $(\dim\gamma = \dim\beta - n + 1)$. Then, (35) can be rewritten into a quadratic condition on γ. It may or may not have solutions and must be examined for each case separately.

Remark: Reconsidering the proofs of Lemmas 4.2 and 4.3, we can now give the following interpretations of the conditions of these lemmas. Assume that the weights $\{\beta_{k,j}\}$ satisfy the natural conditions. Then

$$ED^n y(t)D^{n-1}y(t) = \begin{cases} (-1)^{n-1}p^{2n-1}r_{|t=0^+} + O(h), \\ \quad \text{if (29) holds} \\ (-1)^{n-1}p^{2n-1}r_{|t=0^+} + O(h^2), \\ \quad \text{if (29) and (33) hold} \end{cases}$$
$$ED^n y(t)D^{n-2}y(t) = \begin{cases} (-1)^{n-2}p^{2n-2}r_{|t=0^+} + O(h) \\ (-1)^{n-2}p^{2n-2}r_{|t=0^+} + O(h^2), \\ \quad \text{if (31) and (33) hold} \end{cases}$$
$$ED^{n-1} y(t)D^{n-1}y(t) = \begin{cases} (-1)^{n-1}p^{2n-2}r_{|t=0^+} + O(h) \\ (-1)^{n-1}p^{2n-2}r_{|t=0^+} + O(h^2), \\ \quad \text{if (32) and (33) hold} \end{cases}$$

and for $j+k < 2n-2, j \leq n, k \leq n$

$$ED^j y(t)D^k y(t) = \begin{cases} (-1)^{\min(j,k)}p^{j+k}r_{|t=0^+} + O(h) \\ (-1)^{\min(j,k)}p^{j+k}r_{|t=0^+} + O(h^2), \quad \text{if (33) holds.} \end{cases}$$

It also holds for the latter set of j and k values that

$$p^{j+k}r_{|t=0^+} = (-1)^{\min(j,k)}Ep^j y(t)p^k y(t).$$

As a consequence of the above relations, we may in the estimator replace

$$\frac{1}{N}\sum_{t=1}^N D^{n-1}y(t)D^{n-1}y(t)$$

by

$$-\frac{1}{N}\sum_{t=1}^{N} D^n y(t) D^{n-2} y(t)$$

as these two expressions coincide up to $O(h)$ or $O(h^2)$. The point is that (32) then can be circumvented; see also (4) in [9]. □

V. Noise Level Estimation and Model Order Selection

The problems addressed in this section are those of estimating the noise level (σ^2) and the model order n. For this purpose, let n denote the true model order and $\hat{n}$ the one estimated. The natural conditions as well as the constraint (29) on the weights $\{\beta_{k,j}\}$ are then assumed to hold for $\hat{n}$.

In light of (28) and the remark ending Section IV, it will be interesting to examine the following quantity:

$$\begin{aligned} S &\triangleq D^{\hat{n}} y(t) D^{\hat{n}-1} y(t) \\ &= \frac{1}{h^{2\hat{n}-1}} \sum_j \beta_{\hat{n}-1,j} \sum_k \beta_{\hat{n},k} y(t+jh) y(t+kh). \end{aligned} \tag{38}$$

It holds that

$$\begin{aligned} ES &= \frac{1}{h^{2\hat{n}-1}} \sum_j \beta_{\hat{n}-1,j} \sum_k \beta_{\hat{n},k} r(|k-j|h) \\ &= \frac{1}{h^{2\hat{n}-1}} \sum_j \beta_{\hat{n}-1,j} \sum_k \beta_{\hat{n},k} \sum_{\nu=0}^{2\hat{n}-1} \frac{|k-j|^\nu}{\nu!} h^\nu p^\nu r + O(h) \\ &= \frac{1}{h^{2\hat{n}-1}} \sum_j \sum_k \beta_{\hat{n}-1,j}\beta_{\hat{n},k} \sum_{\nu=0}^{2\hat{n}-1} (k-j)^\nu \frac{h^\nu}{\nu!} p^\nu r \\ &\quad + \frac{1}{h^{2\hat{n}-1}} \sum_j \sum_k \beta_{\hat{n}-1,j}\beta_{\hat{n},k} \\ &\quad \cdot \sum_{\nu=0}^{2\hat{n}-1} [|k-j|^\nu - (k-j)^\nu] \frac{h^\nu}{\nu!} p^\nu r + O(h) \\ &\triangleq T_1 + T_2 + O(h). \end{aligned} \tag{39}$$

Taking the natural conditions into account, the first term can be evaluated as

$$\begin{aligned} T_1 &= \frac{1}{h^{2\hat{n}-1}} \sum_j \sum_k \beta_{\hat{n}-1,j}\beta_{\hat{n},k} \\ &\quad \cdot \binom{2\hat{n}-1}{\hat{n}} k^{\hat{n}} (-j)^{\hat{n}-1} \frac{h^{2\hat{n}-1}}{(2\hat{n}-1)!} p^{2\hat{n}-1} r_{|t=0^+} \\ &= (-1)^{\hat{n}-1} p^{2\hat{n}-1} r_{|t=0^+}. \end{aligned} \tag{40}$$

We now separate between three different cases.

Case 1 ($\hat{n} < n$): As all odd derivatives up to $p^{2n-3}r$ vanish (c.f., Lemma A.2) and $2\hat{n}-1 \le 2n-3$, it follows that $T_1 \equiv 0$. Further, $T_2 \equiv 0$ holds (terms with even values of ν vanish; terms with odd values of ν vanish as well since $p^\nu r = 0$ in such a case). Hence, for Case 1

$$ES = O(h). \tag{41}$$

Case 2 ($\hat{n} = n$): Now

$$T_1 = (-1)^{n-1} p^{2n-1} r_{|t=0^+} \tag{42}$$

while the only possible contributing term for T_2 would be $\nu = 2n-1$. However, this term vanishes due to the constraint (29). Hence, in this case

$$ES = (-1)^{n-1} p^{2n-1} r_{|t=0^+} + O(h). \tag{43}$$

Case 3 ($\hat{n} > n$): The term T_1 is given by (40). The term T_2 is evaluated as

$$\begin{aligned} T_2 &= \frac{1}{h^{2\hat{n}-1}} \sum_j \sum_k \beta_{\hat{n}-1,j} \hat{\beta}_{\hat{n},k} \\ &\quad \cdot \sum_{\nu=2n-1}^{2\hat{n}-3} [|k-j|^\nu - (k-j)^\nu] \frac{h^\nu}{\nu!} p^\nu r. \end{aligned} \tag{44}$$

In the particular (and natural!) case when (29) is satisfied by choosing $k \ge j$, all terms in (44) will vanish and we obtain

$$ES = (-1)^{\hat{n}-1} p^{2\hat{n}-1} r_{|t=0^+} + O(h). \tag{45}$$

The results so far indicate an approach for model-order selection. Apply the model estimation scheme with successively increasing values of $\hat{n}$. Evaluate the statistic S. When it has increased from a "small value" [of order $O(h)$] to a "medium sized" value [corresponding to (43)], the appropriate model order has been found. Note, further, that the statistic S is available when computing the parameter estimate as

$$\hat{S} = \left[\frac{1}{N} \sum_{t=1}^{N} \varphi(t) w(t) \right]_1 \tag{46}$$

that is, the first component of the vector in brackets; compare with (12).

It will be of interest to further examine the derivative appearing in (43). The following result applies.

Lemma 5.1: Consider the autoregressive process (1). Assume that $\hat{n} = n$. Then

$$ES = -\frac{\sigma^2}{2} + O(h). \tag{47}$$

Proof: See Appendix A. □

Remark: The important and interesting conclusion is that the quantity S can be used to estimate the noise level. According to the lemma, it is natural to estimate the noise level as

$$\hat{\sigma}^2 = -2\hat{S}. \tag{48}$$

□

VI. Order-Recursive Implementation

One major advantage of the LS approach in the discrete time is the possibility of avoiding inverting the matrix, such as the one in (12), by using efficient order-recursive algorithms such as the celebrated Levinson's algorithm. So if possible we would like to do the same in (12). In addition, (46), the order-estimation scheme based on S, makes an order-recursive algorithm especially attractive for this work.

However, the conventional Levinson's algorithm which is based on the shift operator breaks down as $h \rightarrow 0$; see [17]. In [17] the authors proposed using the δ operator as an approximation to the differentiation, which is nothing but the forward Euler operator as in Example 3.2. Since (29) is not satisfied with the forward Euler operator, the δ Levinson-type algorithm does not converge (as $h \rightarrow 0$) to the continuous-time AR (CAR) process parameters, as demonstrated in our Example 3.2 and in [17]. On the other hand, in [9] the authors proposed an order-recursive algorithm for estimating continuous-time AR process parameters based on *continuous-time data and their derivatives*. This algorithm is rearranged and presented in Table V. Now in combination with (29) or (31)–(33), we can use it to estimate the CAR process order and parameters *consistently* from its *discrete-time data*. The amount of computation, without counting differentiation approximation, has dropped from $O(n^3)$ as in (12) to $O(n^2)$ as in Table V.

Note that in Table V we have utilized the correlation properties of the derivatives mentioned at the end of Section IV, to save computation. In fact, appropriate combinations of (7), (29), and (31)–(33) guarantee consistent approximations of various correlations of the derivatives from the discrete-time data as described at the end of Section IV. More precisely, compared to a standard LS formulation, we have replaced terms

$$\frac{1}{N}\sum D^i y(t) D^j y(t) \qquad i+j \text{ odd}, i+j < 2n-1$$

by zero. This will introduce a deviation of the order $O(h/N)$ or $O(h^2/N)$ from the exact LS estimates. It is our experience that this modification not only simplifies the computation but also has a beneficial influence on the result. Since the order-recursive algorithm in Table V is not a rearrangement of (12) but basically that of [9], it does not have strictly an LS interpretation as (12). Rather, the LS interpretation would now be in the asymptotic sense (as $N \rightarrow \infty$), similar to the argument in [9, eq. (1.8)].

It should also be noted that unlike the discrete-time case, there is no obvious lattice structure associated with this order-recursive algorithm in the continuous-time parameters. In addition, the order determination statistic S as given in (38) does not depend on an estimate of the process parameters. It can be used to estimate the process order for both the full-matrix version (12) and the order-recursive version in Table V.

VII. Numerical Illustrations

In this section, we will illustrate the theoretical results on the LS method of the previous sections by means of numerical simulations.

For the LS estimator (12), the central design variables are the derivative approximations (4). Based on simulation studies, we will illustrate the effect of these design variables on the estimates. Also, we will compare the LS method with other estimation methods.

A. Illustration of Lemmas 4.1 and 4.2

Lemmas 4.1 and 4.2 can be considered as a basic guideline on how to choose a sensible derivative approximation (4). Hence, it is of interest to verify their validity experimentally.

Data were generated after (instantaneously) sampling the second-order process

$$p^2 y(t) + a_1 p y(t) + a_2 y(t) = e(t) \tag{49}$$

where $e(t)$ is a continuous-time white Wiener process with unit incremental variance. The parameters in (49) were chosen as $a_1 = 2, a_2 = 2$. Note that (49) is the same system as used in [4] and [13] for numerical simulations.

The type of derivative approximations used in the estimations are summarized in Table I.

Remarks:

1) Derivatives B and C satisfy (29) of Lemma 4.1. Hence, they will give a bias of order $O(h)$. The derivative B is based on the Euler backward approximation, while the derivative C on the central difference. Notice that a forward factor has been inserted in the approximation of the second-order derivative in order to satisfy condition (29).
2) Derivatives ZF1 and ZF2 satisfy all the conditions of Lemma 4.2. Hence, they will give a bias of order $O(h^2)$. The name Zero Forcing recalls the fact that the bias term depending on h has been forced to be zero. Note that for a second-order process, only four combinations of the first- and second-order derivative approximation are possible in order to satisfy Lemma 4.2, (see Example 4.1).

For the above derivative approximations, the experimental and theoretical bias are compared in Figs. 2 and 3. To reduce the variance contribution and highlight the bias part of the estimation error $\hat{\theta} - \theta_o$, the number of data points used in the simulation was chosen large, $N = 10\,000$. The sampling period h varies between $h = 0.005$ and $h = 0.3$. The value of the bias obtained experimentally has been averaged in over 150 realizations of the process.

B. Comparison of LS and IV Estimates

The same data, generated from (49) as in the previous subsection, were used here. The sampling period has been chosen as $h = 0.08$. The data length is $N = 2000$.

The LS estimator (12) has been calculated for different choices of the derivative approximations. The LS estimators are compared with an IV-estimator and a prediction error estimator based on a discrete-time ARMA model, respectively.

The IV-estimator corresponds to (11), where the IV vector consists of observations of $y(t)$ delayed one sampling period. The derivatives are approximated by the use of the central difference. In [4] and [13] it is shown that this choice gives a good IV-estimator.

The ARMA estimator is obtained by implementing an indirect method based on a discrete-time ARMA (2,2) model. First, a PEM is used in order to estimate the ARMA parameters. Then the continuous-time AR parameters are estimated by

TABLE I
DERIVATIVE APPROXIMATIONS

Short name	Description	Derivative Approximations
B	Euler Backward	$Dy(t) = \frac{y(t)-y(t-h)}{h}$ $D^2y(t) = \frac{y(t+2h)-2y(t+h)+y(t)}{h^2}$
C	Central Difference	$Dy(t) = \frac{y(t+h)-y(t-h)}{2h}$ $D^2y(t) = \frac{y(t+5h)-2y(t+3h)+y(t+h)}{4h^2}$
ZF1	Zero Forcing #1	$Dy(t) = \frac{0.2047y(t+h)+0.8860y(t)-1.3860y(t-h)+0.2953y(t-2h)}{h}$ $D^2y(t) = \frac{-2y(t+4h)+7y(t+3h)-8y(t+2h)+3y(t+h)}{h^2}$
ZF2	Zero Forcing #2	$Dy(t) = \frac{-0.2953y(t+2h)+1.3860y(t+h)-0.8860y(t)-0.2047y(t-h)}{h}$ $D^2y(t) = \frac{-3y(t+5h)+10y(t+4h)-11y(t+3h)+4y(t+2h)}{h^2}$

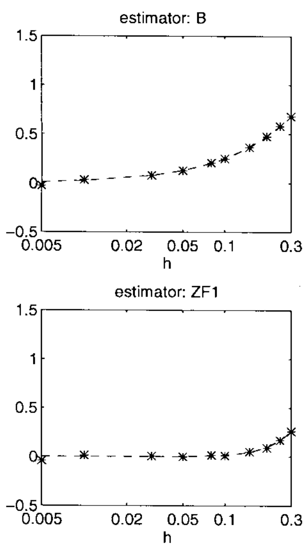

Fig. 2. Experimental (*) and theoretical (dashed) bias of the estimate of parameter $a_1 = 2$. The experimental bias is averaged in over 150 realizations of the process.

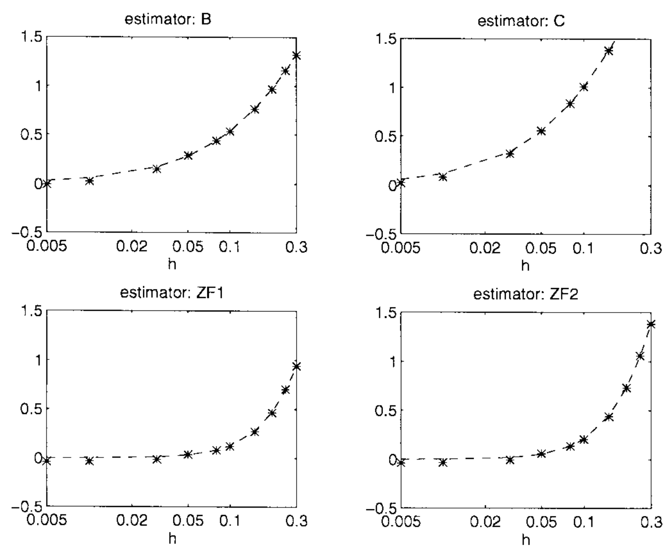

Fig. 3. Experimental (*) and theoretical (dashed) bias of the estimate of parameter $a_2 = 2$. The experimental bias is averaged in over 150 realizations of the process.

transforming the discrete-time poles to continuous time using the relation $z_d = e^{z_c h}$, where z_d and z_c are the discrete- and continuous-time poles, respectively.

The sample mean and standard deviation of the parameter estimates from 150 realizations of the process were calculated for each estimator. The results from the Monte Carlo simulations are shown in Table II. Observe that in Table II, derivative approximations satisfying Lemma 4.1, thus giving bias of order $O(h)$, have been grouped in the first four rows. Those satisfying Lemma 4.2, thus giving bias of order $O(h^2)$, have been grouped in the second four rows.

Remarks:

1) Derivative C1 is based on the central difference. It uses the approximation used by estimator C, shifted one sampling period backward. See also Table I.
2) Derivative CF uses the central difference for the first-order derivative and the Euler forward approximation for the second-order derivative. The central difference is shifted one sampling period backward in order to satisfy (29).
3) Derivatives ZF3 and ZF4 have been introduced in Example 4.1. The first derivative approximation corresponds

TABLE II
PERFORMANCE OF LS, IV, AND ARMA ESTIMATORS. THE NUMBER OF DATA POINTS, $N = 2000$, AND THE SAMPLING PERIOD, $h = 0.08$

short	Parameter $a_1 = 2$				Parameter $a_2 = 2$			
name	mean	std	bias	mse	mean	std	bias	mse
B	1.824	0.139	0.176	0.050	1.575	0.181	0.425	0.214
C	1.716	0.120	0.284	0.095	1.165	0.134	0.835	0.715
C1	1.620	0.121	0.380	0.159	1.193	0.137	0.807	0.670
CF	1.784	0.144	0.216	0.067	1.455	0.167	0.545	0.325
ZF1	2.030	0.224	-0.030	0.051	1.939	0.231	0.061	0.057
ZF2	2.013	0.282	-0.013	0.080	1.880	0.228	0.120	0.066
ZF3	1.997	0.263	0.003	0.069	1.987	0.237	0.013	0.057
ZF4	1.942	0.361	0.058	0.134	1.833	0.222	0.167	0.077
IV	2.034	0.237	-0.034	0.057	2.031	0.238	-0.031	0.057
ARMA	2.036	0.239	-0.036	0.058	2.031	0.246	-0.031	0.061

TABLE III
PARAMETER ESTIMATION OF $a_1, \cdots, a_4, h = 0.01, N = 50\,000$ POINTS

	True value	Estimate
a_1	11	11.2152
a_2	58	56.7432
a_3	112	112.9398
a_4	64	64.6856

TABLE IV
ORDER ESTIMATION, $h = 0.01, N = 50\,000$ POINTS

$\hat{n}$	Theoretical value	$-S(\hat{n})$
1	0	-9.3769×10^{-8}
2	0	-3.1939×10^{-8}
3	0	1.2472×10^{-4}
4	0.5	0.4982

to the choices $k = -2, \gamma = -1.0787$, and $k = -1, \gamma = 1.0787$, respectively (see Example 4.1). The second derivative approximation is given by the unique approximation with four coefficients $\beta_{2,i}$, which satisfies the extended natural conditions (33).

The conclusions from the Monte Carlo simulations are as follows.

1) LS estimators based on derivative approximations satisfying Lemma 4.1, thus giving bias of order $O(h)$, have in general small variance, but considerable bias. In particular, observe that the estimate of a_2 has a rather large bias.

2) LS estimators based on derivative approximations satisfying Lemma 4.2, thus giving bias of order $O(h^2)$, have good performances. In particular, estimator ZF1 gives performance comparable to the ARMA estimator. Hence, the guidelines given by Lemma 4.2 seem to be meaningful.

In conclusion, estimator ZF1 seems to be a good variant of the LS estimator. Its performance is as good as that of any other studied method. The reason for this, as a general guideline in choosing the derivative approximation, is that in addition to satisfying the requirements for $O(h^2)$ accuracy, all the derivative approximates should span the least amount of time shift and should collectively have time shifts as symmetric as possible around t. This tends to give smaller biases and smaller mean-square errors. Notice that the implementation of such an estimator is much simpler than the indirect method. Also, it does not suffer from the limitations described in the introduction.

TABLE V
AN ORDER-RECURSIVE ALGORITHM FOR ESTIMATING CONTINUOUS-TIME AR COEFFICIENTS WITH ORDER DETERMINATION

Model: $p^M y(t) + a_1 p^{M-1} y(t) + \cdots + a_M y(t) = \sigma e(t)$

Definition:

$$\begin{aligned}
\mathbf{a}_{even} &= [a_M, a_{M-2}, \cdots, a_4, a_2]^T \quad if\ M\ even.\\
&= [a_{M-1}, a_{M-3}, \cdots, a_4, a_2]^T \ if\ M\ odd.\\
\mathbf{a}_{odd} &= [a_{M-1}, a_{M-3}, \cdots, a_3, a_1]^T \ if\ M\ even\\
&= [a_M, a_{M-2}, \cdots, a_3, a_1]^T \quad if\ M\ odd.
\end{aligned}$$

Order recursive parameters:

$$\begin{aligned}
\mathbf{a}_k &= [a_{k,k}, a_{k,k-2}, \cdots, a_{k,4}, a_{k,2}]^T\\
\mathbf{a}'_{k+1} &= [a_{k+1,k}, a_{k+1,k-2}, \cdots, a_{k+1,4}, a_{k+1,2}]^T.
\end{aligned}$$

Note: $\mathbf{a}_M = \mathbf{a}_{even}$ *if M even and* $\mathbf{a}'_M = \mathbf{a}_{even}$ *if M odd.*

Initialization: $v_0 = \frac{1}{N}\sum y^2(t)$, $A_1 = \frac{1}{N}\sum y(t)Dy(t)$.

If $-A_1$ > threshold, then

$$M = 1,\ a_1 = -(A_1/v_0),\ \sigma^2 = -2A_1.$$

stop.

otherwise continue.

$$\begin{aligned}
v_1 &= -\frac{1}{N}\sum y(t)D^2y(t).\dagger\\
A_2 &= \frac{1}{N}\sum Dy(t)D^2y(t).
\end{aligned}$$

If $-A_2$ > threshold, then

$$\begin{aligned}
M &= 2,\ \mathbf{a}_2 = a_{2,2} = (v_1/v_0),\\
a_{2,1} &= -(A_2/v_1),\ \sigma^2 = -2A_2.
\end{aligned}$$

stop.

otherwise continue.

$$\mathbf{r}_1'^T = v_1,\ \mathbf{a}_2 = (v_1/v_0), \begin{bmatrix} \mathbf{a}'_1 \\ 1 \end{bmatrix} = 1$$

C. Illustration of the Order-Recursive Algorithm and Estimation of Noise Intensity

The order-recursive algorithm as summarized in Table V has been tested for a CAR process of order four. The results are presented in Tables III–IV. The CAR process is driven by a continuous-time white noise process with a unit incremental variance and is sampled with the sampling period h. The samples are then used in the order-recursive algorithm to estimate the CAR process parameters. The transfer function

TABLE V (*Continued*)

Algorithm: For $k = 2, 4, 6, \cdots$, do the following:

$$\begin{aligned}
\mathbf{r}_k^T &= [-\mathbf{r}'^T_{k-1},\ -\frac{1}{N}\sum D^{k-1}y(t)D^{k+1}y(t)].\dagger \\
v_k &= \mathbf{r}_k^T \begin{bmatrix}\mathbf{a}_k \\ 1\end{bmatrix}. \\
\mathbf{a}'_{k+1} &= \mathbf{a}_k + (v_k/v_{k-1}) \begin{bmatrix}\mathbf{a}'_{k-1} \\ 1\end{bmatrix}. \\
A_{k+1} &= \frac{1}{N}\sum D^k y(t) D^{k+1} y(t).
\end{aligned}$$

If $-A_{k+1}$ > threshold go to *Step 1*.

Otherwise continue.

$$\begin{aligned}
\mathbf{r}'^T_{k+1} &= \begin{bmatrix}\textit{The last } \frac{k}{2}+1 \textit{ elements of} \\ [-\mathbf{r}_k^T,\ -\frac{1}{N}\sum D^k y(t) D^{k+2} y(t)]\end{bmatrix}.\dagger \\
v_{k+1} &= \mathbf{r}'^T_{k+1}\begin{bmatrix}\mathbf{a}'_{k+1} \\ 1\end{bmatrix}. \\
\mathbf{a}_{k+2} &= \begin{bmatrix}0 \\ \mathbf{a}'_{k+1}\end{bmatrix} + (v_{k+1}/v_k)\begin{bmatrix}\mathbf{a}_k \\ 1\end{bmatrix}. \\
A_{k+2} &= \frac{1}{N}\sum D^{k+1} y(t) D^{k+2} y(t).
\end{aligned}$$

If $-A_{k+2}$ > threshold, go to *Step 2*.
Otherwise continue.

End of k loop. (If order M is known then the tests $-A_{k+i}$ > threshold becomes testing $k+i = M$, $i = 1, 2$).

Step 1. In this case $M = k+1 =$ *odd*.

$$\begin{aligned}
\mathbf{a}_{even} &= \mathbf{a}'_{k+1} = \mathbf{a}'_M\ ,\ \mathbf{a}_{odd} = -(A_{k+1}/v_k)\begin{bmatrix}\mathbf{a}_k \\ 1\end{bmatrix} = -(A_M/v_{M-1})\begin{bmatrix}\mathbf{a}_{M-1} \\ 1\end{bmatrix}, \\
\sigma^2 &= -2A_M.
\end{aligned}$$

Step 2. In this case, $M = k+2 =$ *even*.

$$\begin{aligned}
\mathbf{a}_{even} &= \mathbf{a}_{k+2} = \mathbf{a}_M\ ,\ \mathbf{a}_{odd} = (-A_{k+2}/v_{k+1})\begin{bmatrix}\mathbf{a}'_{k+1} \\ 1\end{bmatrix} = -(A_M/v_{M-1})\begin{bmatrix}\mathbf{a}'_{M-1} \\ 1\end{bmatrix}, \\
\sigma^2 &= -2A_M.
\end{aligned}$$

†: The summation $-\frac{1}{N}\sum D^{k-1}y(t)D^{k+1}y(t)$ is used to avoid (4.4) and to achieve $O(h^2)$ accuracy. Equations (4.1), (4.3), and (4.5) should be met at every k.

For $O(h)$ accuracy, $-\frac{1}{N}\sum D^{k-1}y(t)D^{k+1}y(t)$ should be replaced by $\frac{1}{N}\sum [D^k y(t)]^2$ to save computation. Equations (2.7) and (4.1) should be met at every k.

The approximation $D^{k+1}y(t)$ can be different in calculating A_{k+1} and A_{k+2} to meet the above requirement.

of the CAR model used in the simulation is

$$A(p) = \frac{1}{p^4 + 11p^3 + 58p^2 + 112p + 64}$$

which has poles at $-1, -2$, and $4 \pm j4$. The threshold is set at 0.25 which is halfway between $-E\{S\}|_{\hat{n}<\text{trueorder}} = 0$ and $-E\{S\}|_{\hat{n}=\text{trueorder}} = 0.5$. The coefficients $\beta_{k,\ell}$ of the derivative approximation D^k in Table V are obtained by solving the linear system (31) for each k and then appropriately shifted to satisfy (27) and (29), similar to ZF1 and ZF3. The algorithm correctly detected the order and gave good parameter estimates, provided h and the total number of data points are chosen appropriately. Table III shows the parameter estimates for ensemble average of 20 trials. Table IV shows $-S$ as a function of $\hat{n}$, also the ensemble average of 20 trials.

It needs to be pointed out, however, that as the order increases, the data points needed to obtain reasonable estimates need also increase. In addition, the performance of the algorithm becomes more sensitive to the choice of h. Too small an h may give erroneous estimates if the number of data points is not chosen appropriately. This is not unexpected since instantaneous sampling is used here, which is known to be prone to numerical error for small h [18].

VIII. Conclusions

LS estimation of continuous-time ARX models using discrete-time approximations of the derivatives has been analyzed. It was shown that the stochastic part causes difficulties, and standard approximations such as Euler (delta) forward or Euler (delta) backward cannot be used in general. The estimation of continuous-time autoregressive processes was considered in some detail.

When is it feasible to use the LS approach? According to our findings (which combines a theoretical analysis and practical, numerical experience), the LS approach can often give very low mean-square errors of the parameter estimates, in particular for a moderately small sampling period. It turns out to be crucial how the true derivatives of the process are approximated by finite differences. Standard methods for derivative approximation cannot be used directly. So far, variants where the theoretical parameter bias declines as $O(h^2), h$ being the sampling period, seem to work best. In particular, it is important how the two highest derivatives are approximated. Based on a complete analysis, simple schemes are provided which yield parameter estimates with a bias term that declines with the sampling period. The LS approach has the potential of being extended to work also for nonuniform sampling, by modifying the derivative approximations accordingly. It is a clear alternative to an instrumental variable method.

An order-recursion implementation scheme has been presented, and also illustrated numerically. As an important byproduct we have provided means to estimate the continuous-time noise intensity also and can hence completely characterize the spectrum of the signal by the estimated parameters.

Appendix
Proofs and Derivations

For the analysis of the asymptotic bias, several preliminary results will be helpful.

First it would be easier to examine the residual

$$\varepsilon \triangleq \frac{1}{N}\Sigma\varphi(t)w(t) - \frac{1}{N}\Sigma\varphi(t)\varphi^T(t)\theta_o \tag{50}$$

than the estimation error $\hat{\theta} - \theta_o$. The following lemma shows that both errors tend to zero *at the same rate*.

Lemma A.1: Let $A_N \to A, b_N \to b, A$ be nonsingular and set $x_N = A_N^{-1}b_N, x = A^{-1}b$. Then

$$x_N - x = -A_N^{-1}(A_N x - b_N) \approx -A^{-1}(A_N x - b_N) \tag{51}$$

and hence $x_N - x$ and $A_N x - b_N$ tend to zero at the same rate.

Proof: Trivial and omitted. □

Lemma A.2: The covariance function satisfies the Yule–Walker equation

$$(p^n + a_1 p^{n-1} + \cdots + a_n) r(t) = 0, \qquad t \geq 0^+. \tag{52}$$

Proof: This is a well-known result and the proof is omitted. □

Corollary: Set $a_o = 1$ and let $j \geq 0$. Then

$$\sum_{i=0}^{n} a_{n-i} p^{i+j} r(t) = 0, \qquad \text{for} \quad t \geq 0^+. \tag{53}$$

Proof of Lemma 4.1: Set $A(p) = \Sigma_{i=0}^n a_i p^{n-i}$. Then the spectral density is

$$\phi(\omega) = \frac{\sigma^2}{|A(i\omega)|^2}.$$

The covariance function can be written as

$$r(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\sigma^2}{|A(i\omega)|^2} e^{i\omega\tau}\, d\omega.$$

By differentiation

$$p^j r(\tau) = \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty} \frac{(i\omega)^j}{|A(i\omega)|^2} e^{i\omega\tau}\, d\omega$$
$$p^j r(\tau)_{|\tau=0^+} = \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty} \frac{(i\omega)^j}{|A(i\omega)|^2}\, d\omega.$$

The integral exists if $j \leq 2n-2$ and vanishes if j is odd (as the integrand then becomes an odd function). This concludes the proof of (27).

To prove (28), represent the process in state-space form as

$$\dot{x} = Ax + Be$$
$$y = Cx$$
$$Ee(t)e(s) = \delta(t-s).$$

The first n Markov parameters of this model satisfy

$$CA^jB = \begin{cases} 0; & j = 0, \cdots, n-2 \\ \sigma; & j = n-1. \end{cases} \tag{54}$$

The covariance function can be written as

$$r(t) = Ce^{At}PC^T, \qquad t \geq 0 \tag{55}$$

where P is the unique, nonnegative definite solution to the Lyapunov function

$$0 = AP + PA^T + BB^T. \tag{56}$$

Differentiating (55) gives

$$p^{2n-1} r_{|t=0^+} = CA^{2n-1}PC^T. \tag{57}$$

Multiplying (56) from the left by CA^{n-1} and from the right by $A^{T^{n-1}}C^T$ gives with (54)

$$0 = CA^n P A^{T^{n-1}} C^T + CA^{n-1} P A^{T^n} C^T + \sigma^2$$

or (as all terms are scalars)

$$0 = 2CA^n P A^{T^{n-1}} C^T + \sigma^2. \tag{58}$$

Next, multiply (56) by CA^j $(0 \leq j \leq n-2)$ and use (54) to conclude

$$0 = CA^{j+1}P + CA^j P A^T, \qquad 0 \leq j \leq n-2$$

and hence

$$0 = PA^{T^{j+1}}C^T + APA^{j^T}C^T, \qquad 0 \leq j \leq n-2.$$

Using this relation repeatedly, we find that

$$PA^{T^{n-1}}C^T = (-1)^{n-1}A^{n-1}PC^T. \tag{59}$$

Finally, combining (57)–(59) gives

$$\begin{aligned} p^{2n-1}r &= CA^{2n-1}PC^T = CA^n(-1)^{n-1}PA^{T^{n-1}}C^T \\ &= (-1)^{n-1}(-1)\frac{\sigma^2}{2} = (-1)^n\frac{\sigma^2}{2} \end{aligned}$$

which is (28). □

Proof of Lemma 4.2: We examine the asymptotic residual error

$$\varepsilon = E\varphi(t)w(t) - E\varphi(t)\varphi^T(t)\theta_o. \tag{60}$$

According to Lemma A.1, ε and the parameter estimation error $\hat{\theta} - \theta_o$ converge to zero at the same rate when $h \to 0$ (assuming they go to zero). Set $a_o = 1$ for convenience and let k $(1 \le k \le n)$ be arbitrary. We have

$$\begin{aligned} \varepsilon_k &= ED^{n-k}y(t)[w(t) - \varphi^T(t)\theta_o] \\ &= ED^{n-k}y(t)\sum_{i=0}^{n} a_{n-i}D^i y(t) \\ &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}E\sum_j \beta_{n-k,j}y(t+jh) \\ &\quad \cdot \sum_\ell \beta_{i,\ell}y(t+\ell h) \\ &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}\sum_j \beta_{n-k,j}\sum_\ell \beta_{i,\ell}r(\ell h - jh). \end{aligned} \tag{61}$$

Next we make a series expansion around $t = 0^+$ of the covariance function. It is arbitrarily continuously differentiable if we stick to positive arguments. Note, though, that at $t = 0$ it is differentiable only a finite number of times (in fact, up to $2n - 2$).

Hence

$$\begin{aligned} \varepsilon_k &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}\sum_j \beta_{n-k,j}\sum_\ell \beta_{i,\ell}r(|\ell - j|h) \\ &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}\sum_j \beta_{n-k,j}\sum_\ell \beta_{i,\ell} \\ &\quad \cdot \left[\sum_{\nu=0}^{n-k+i} |\ell - j|^\nu \frac{h^\nu}{\nu!}p^\nu r_{|t=0^+} + O(h^{n-k+i+1})\right] \\ &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}\sum_j \beta_{n-k,j}\sum_\ell \beta_{i,\ell} \\ &\quad \cdot \sum_{\nu=0}^{n-k+i} |\ell - j|^\nu \frac{h^\nu}{\nu!}p^\nu r_{|t=0^+} + O(h). \end{aligned} \tag{62}$$

To proceed, it will turn out useful if we have

$$|\ell - j|^\nu p^\nu r_{|t=0^+} = (\ell - j)^\nu p^\nu r_{|t=0^+}. \tag{63}$$

Clearly, this is always true for an even value of ν. Further, by Lemma 4.1, it holds for $\nu = 1, 3, \cdots, 2n-3$. The only remaining case is $\nu = 2n - 1$ which precisely corresponds to

$$i = n, \qquad k = 1. \tag{64}$$

Assume for a moment that (63) applies generally. Then

$$\begin{aligned} \varepsilon_k &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}\sum_j \beta_{n-k,j}\sum_\ell \beta_{i,\ell} \\ &\quad \cdot \sum_{\nu=0}^{n-k+i}\frac{h^\nu}{\nu!}p^\nu r_{|t=0^+}\sum_{m=0}^{\nu}\binom{\nu}{m}\ell^m(-j)^{\nu-m} + O(h) \\ &= \sum_{i=0}^{n} a_{n-i}\frac{1}{h^{n-k+i}}\sum_{\nu=0}^{n-k+i}\frac{h^\nu}{\nu!}p^\nu r_{|t=0^+}\sum_{m=0}^{\nu}\binom{\nu}{m} \\ &\quad \cdot \left[\sum_j \beta_{n-k,j}(-j)^{\nu-m}\right]\left[\sum_\ell \beta_{i,\ell}\ell^m\right] + O(h). \end{aligned} \tag{65}$$

Now, by the natural condition

$$\sum_\ell \beta_{i,\ell}\ell^m = \begin{cases} 0, & \text{if } m < i \\ i!, & \text{if } m = i \end{cases}$$

$$\sum_j \beta_{n-k,j}(-j)^{\nu-m} = \begin{cases} 0, & \text{if } \nu - m < n - k \\ (-1)^{\nu-m}(n-k)!, & \text{if } \nu - m = n - k. \end{cases}$$

Hence, the only way the products of these sums become nonzero is

$$m \ge i \quad \text{and} \quad \nu - m \ge n - k.$$

However, as $\nu \le n - k + i$, this can only happen if

$$m = i, \qquad \nu - m = n - k. \tag{66}$$

Thus

$$\begin{aligned} \varepsilon_k &= \sum_{i=0}^{n} a_{n-i}\frac{1}{(i+n-k)!}p^{i+n-k}r \\ &\quad \cdot \binom{i+n-k}{i}(-1)^{n-k}(n-k)!i! + O(h) \\ &= (-1)^{n-k}\sum_{i=0}^{n} a_{n-i}p^{i+n-k}r + O(h) \\ &= O(h) \end{aligned} \tag{67}$$

where Lemma A.2 was used in the last equality.

It remains to cope with (63) when $\nu = 2n-1, i = n, k = 1$; see (64). This only occurs for the following term of ε_1, which can be rewritten as

$$\begin{aligned} &a_0\frac{1}{h^{2n-1}}\sum_j \beta_{n-1,j}\sum_\ell \beta_{n,\ell}|\ell - j|^{2n-1} \\ &\quad \cdot \frac{h^{2n-1}}{(2n-1)!}p^{2n-1}r \\ &= a_0\frac{1}{(2n-1)!}p^{2n-1}r\sum_j \beta_{n-1,j} \\ &\quad \cdot \sum_\ell \beta_{n,\ell}(\ell - j)^{2n-1} \end{aligned}$$

by (29). We then proceed as in (65)–(67) to conclude that $\varepsilon_1 = O(h)$. This completes the proof. □

Proof of Lemma 4.3: We now modify (62) into

$$\varepsilon_k = \sum_{i=0}^{n} a_{n-i} \frac{1}{h^{n-k+i}} \sum_j \beta_{n-k,j} \sum_\ell \beta_{i,\ell} \cdot \sum_{\nu=0}^{n-k+i+1} |\ell - j|^\nu \frac{h^\nu}{\nu!} p^\nu r + O(h^2). \tag{68}$$

Assume for a while that

$$|\ell - j|^\nu p^\nu r = (\ell - j)^\nu p^\nu r \quad \nu = 0, \cdots, n-k+i+1. \tag{69}$$

The only troublesome case is $\nu = 2n-1$, which now corresponds to

$$i = n, k = 2; \quad i = n-1, k = 1; \quad \text{and} \quad i = n, k = 1. \tag{70}$$

The last case is taken care of using (52). The other two cases are taken care of by (54) and (55). Similarly to (65) we have

$$\varepsilon_k = \sum_{i=0}^{n} a_{n-i} \frac{1}{h^{n-k+i}} \sum_{\nu=0}^{n-k+i+1} \frac{h^\nu}{\nu!} p^\nu r \sum_{m=0}^{\nu} \binom{\nu}{m} \cdot \left[\sum_j \beta_{n-k,j} (-j)^{\nu-m}\right] \left[\sum_\ell \beta_{i,\ell} \ell^m\right] + O(h^2). \tag{71}$$

Then, the products of the brackets of (71) will be nonzero precisely when $m = i$ and $\nu - m = n - k$, c.f., (33).

Paralleling (67), we can then conclude that $\varepsilon_k = O(h^2)$, which completes the proof. □

Proof of Lemma 5.1: The statement follows directly by combining (43) and (28). □

REFERENCES

[1] K. J. Åström, *Introduction to Stochastic Control Theory*. New York: Academic, 1970.
[2] A. R. Bergström, Ed., *Statistical Inference in Continuous Time Economic Models*. Amsterdam, The Netherlands: North-Holland, 1976.
[3] ——, *Continuous Time Econometric Modeling*. Oxford Univ., 1990.
[4] S. Bigi, T. Söderström, and B. Carlsson, "An IV scheme for estimating continuous-time stochastic models from discrete-time data," in *Proc. 10th IFAC Symp. Syst. Identification,* Copenhagen, Denmark, 1994.
[5] W. L. Luyben, *Process Modeling, Simulation, and Control for Chemical Engineers*. Tokyo: McGraw Hill, 1990.
[6] R. H. Middleton and G. C. Goodwin, *Digital Control and Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[7] E. Parzen, Ed., *Time Series Analysis of Irregularly Observed Data*. New York: Springer-Verlag, 1984.
[8] M. S. Phadke and S. M. Wu, "Modeling of continuous stochastic processes from discrete observations with application to sunspots data," *J. American Stat. Assoc.,* vol. 69, no. 346, June 1974.
[9] D. T. Pham and A. Le Breton, "Levinson Durbin-type algorithms for continuous-time autoregressive models and applications," *Math. Contr., Signals Syst.,* vol. 4, pp. 69–79, 1991.
[10] S. Sagara, Z.-J. Yang, K. Wada, and T. Tsuji, "Parameter identification and adaptive control of continuous systems with zero-order hold," in *Proc. 12th IFAC World Congr.,* Sydney, Australia, 1993.
[11] N. K. Sinha and G. P. Rao, *Identification of Continuous-Time Systems*. Dordrecht, The Netherlands: Kluwer Academic, 1991.
[12] T. Söderström, *Discrete-Time Stochastic Systems: Estimation and Control*. Hemel Hempstead, U.K.: Prentice-Hall Int., 1994.
[13] T. Söderström, B. Carlsson, and S. Bigi, "On estimating continuous-time stochastic models from discrete-time data," *Rep. UPTEC 92104R*, Dept. Technol., Uppsala Univ., Sweden, 1992.
[14] T. Söderström, H. Fan, S. Bigi, and B. Carlsson, "When is a least-squares fit feasible for modeling continuous-time autoregressive processes from discrete-time data?" *Rep. UPTEC 95003R*, Dept. Technol., Uppsala Univ., Uppsala, Sweden, 1995.
[15] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, U.K.: Prentice-Hall Int., 1989.
[16] H. Unbehauen and G. P. Rao, *Identification of Continuous Systems*. Amsterdam: North-Holland, 1987.
[17] R. Vijayan, H. V. Poor, J. B. Moore, and G. C. Goodwin, "A Levinson-type algorithm for modeling fast-sampled data," *IEEE Trans. Automat. Contr.*, vol. 36, no. 3, pp. 314–321, 1991.
[18] B. Wahlberg, L. Ljung and T. Söderström, "On sampling of continuous time stochastic processes," *Contr. Theory Advanced Technol.*, vol. 9, no. 1, pp. 99–112, 1993.
[19] P. Young, "Parameter estimation for continuous-time models—A survey," *Automatica*, vol. 17, pp. 23–29, 1981.
[20] Z. Y. Zhao, S. Sagara, and M. Tomizuka, "A new bias-compensating LS method for continuous system identification in the presence of colored noise," *Int. J. Contr.*, vol. 56, no. 6, pp. 1441–1452, 1992.
[21] Z. Y. Zhao, S. Sagara, and K. Wada, "Bias-compensating least squares method for identification of continuous-time systems from sampled data," *Int. J. Contr.*, vol. 53, no. 2, pp. 445–461, 1991.

Torsten Söderström (M'76–SM'83–F'92) was born in Malmö, Sweden, in 1945. He received the M.Sc. degree ("civilingenjör") in engineering physics in 1969 and the Ph.D. degree in automatic control in 1973, both from Lund Institute of Technology, Lund, Sweden.

Since 1974, he has been at the Department of Technology, Uppsala University, Uppsala, Sweden, where he is a professor of automatic control, and he has been the Head of the Systems and Control Group since 1975. His main research interests are in the fields of system identification, signal processing, and adaptive systems. He is the author or coauthor of many technical papers and four books, the most recent being *Discrete-Time Stochastic Systems* (Englewood Cliffs, NJ: Prentice-Hall, 1994).

Dr. Söderström was, with coauthors, given an Automatica Prize Paper Award in 1981. Since 1992 he has been an editor for *Automatica*, in the area of System Parameter Estimation.

H. (Howard) Fan (S'82–M'85–SM'90) received the B.S. degree from Guizhou University, China, in 1976, and the M.S. and Ph.D. degrees from the University of Illinois, Champaign-Urbana, in 1982 and 1985, respectively, all in electrical engineering.

From 1977 to 1978, he worked as a Research Engineer at the Provincial Standard Laboratory and Bureau of Guizhou Province in China. In 1978, he entered the Graduate School of the University of Science and Technology of China and then transferred to the University of Illinois where he was a Teaching Assistant and a Research Assistant from 1982 to 1985. He joined the Electrical and Computer Engineering Department of the University of Cincinnati in 1985 as an Assistant Professor, where he is now a Professor. His research interests are in the general fields of systems and signal processing, in particular signal representation and reconstruction, adaptive signal processing, signal processing for communications, and system identification. He spent six months in 1994 with the Systems and Control Group, Uppsala University, Sweden, as a Visiting Researcher.

Dr. Fan received the first ECE Departmental Distinguished Progress in Teaching Excellence Award at the University of Cincinnati, in 1987, and served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1991 to 1994. He is a member of Tau Beta Pi and Phi Kappa Phi.

Bengt Carlsson (M'92) was born in Uppsala, Sweden, in 1958. He received the M.S. degree in engineering physics in 1984 and the Ph.D. degree in automatic control in 1989, all from the Institute of Technology, Uppsala University, Uppsala, Sweden.

In 1988, he was a Visiting Research Associate at the Department of Electrical Engineering, University of Newcastle, Australia. He is with the Systems and Control Group at Uppsala University, where he was an Assistant Professor (1989–1994) and since 1994 an Associate Professor. Since 1991 he has been a project leader dealing with modeling and control of wastewater treatment plants. His research interests also include modeling and controlling of environmental processes.

In 1996 Dr. Carlsson received the Best Teacher Award from the engineering students at Uppsala University.

Stefano Bigi was born in 1964. He graduated in electrical engineering from Padova University, Italy, in 1990 and received the Licentiate Degree in automatic control from the System and Control Group, Uppsala University, Sweden, in 1995.

He is now with S.I.M.A.C.S.p.A., Italy, where he has a management position in the Sales and Marketing sector. His research interests include system identification and its use for design purposes.