CONTINUOUS-TIME AR PROCESS PARAMETER ESTIMATION FROM DISCRETE-TIME DATA[1]

H. Fan[2], T. Söderström[3], M. Mossberg[3], B. Carlsson[3], Y. Zou[2]

[2]Department of ECECS
University of Cincinnati
Cincinnati, OH 45221-0030, U.S.A.

[3]Systems and Control Group
School of Engineering, Uppsala University
P O Box 27, S-751 03 Uppsala, Sweden

ABSTRACT

The problem of estimating continuous-time autoregressive process parameters from discrete-time data is considered. The basic approach used here is based on replacing the derivatives in the model by discrete-time differences, forming a linear regression and using the least squares method. It is known, however, that all standard approximations of the highest order derivative give a biased least squares estimate even as the sampling interval tends to zero. Some of our previous approaches to overcome this problem are briefly reviewed. Then two new methods are presented. One of them, termed bias compensation, can be easily implemented efficiently in an order recursive manner. Comparative simulation results are also presented.

1. INTRODUCTION

Identification of continuous-time models and continuous-time time series is a problem of considerable importance in various disciplines such as automatic control, signal processing, astrophysics and economics, [1] – [3]. For example, in economics, since variables in most models are the result of a large number of microeconomic decisions at different points of time, they may be regarded as continuous functions of time [1]. Another example is in astrophysics when modeling sunspot data [3]. Continuous-time autoregressive (CAR) processes are therefore common. On the other hand, observations in practice are often made at discrete-time instants. Thus estimation of CAR process parameters from discrete-time data is a practically important problem.

In this paper, two new algorithms for identification of CAR models using direct methods is analyzed. In the direct methods, the derivatives in the CAR model are substituted by discrete-time differences and the model is thus transformed into a (discrete-time) linear regression, parameterized with the CAR parameters. A derivative approximation of particular interest is the delta approximation which has received considerable attention in the control and signal processing literature, see [4]-[6]. The approach used is the least squares, which is numerically well-behaved, computationally efficient and amenable to order-recursive implementations. It is therefore of particular interest to the signal processing community.

In [7], it was shown that standard approximations such as repeated use of the forward delta or backward delta operators do not give an estimate which converges to the true parameter vector as the number of data points goes to infinity and the sampling interval goes to zero. A certain shift structure needs to be imposed on the highest order derivative approximation to give an unbiased estimate as the sampling interval goes to zero, see [7]. These crucial findings are briefly reviewed and further developed in this paper. Specifically, two new methods coping with the bias problem caused by the highest order derivative approximation will be presented. Both methods avoid the shift used in [7]. Furthermore, in the second method repeated use of the delta operators is allowed, resulting in computationally efficient order-recursive algorithms. Simulation results show that the new least squares methods give smaller bias than our previous results using the shift structure in [7], at least for the examples that have been tested.

2. GENERAL SETTING AND PREVIOUS RESULTS

In this section we set up the problem and review our previous results. Consider a continuous-time autoregressive (CAR) process

$$\begin{aligned}(p^n + a_1 p^{n-1} \quad + \quad \cdots + a_n)y(t) &= e(t) \\ Ee(t)e(s) \quad = \quad \sigma^2\delta(t-s) &\end{aligned} \tag{2.1}$$

where p denotes the differentiation operator and $e(t)$ is a (continuous-time) white noise source. The time series is observed at discrete-time instants $t = h, 2h, 3h, \cdots, Nh$. The model order n is supposed to be known. It is of interest to estimate the autoregressive parameters $\theta = [a_1, \cdots, a_n]^T$ from the available discrete-time data.

In the direct approach one would approximate the differentiation operator p as

$$p^k f(t) \approx D^k f(t) \triangleq \frac{1}{h^k}\sum_j \beta_{k,j} f(t+jh) \tag{2.2}$$

where $\{\beta_{k,j}\}$ are some weights. The *natural conditions* ensuring $D^k f(t) = p^k f(t) + O(h)$ must be imposed and are

[1] Work supported by the US Office of Naval Research under Grant N00014-96-1-0241, and by Swedish Research Council for Engineering Sciences under contract 95-800.

given by

$$\sum_j \beta_{k,j} j^\nu = \begin{cases} 0 & \nu = 0, \cdots, k-1 \\ k! & \nu = k \end{cases} \tag{2.3}$$

These conditions are satisfied, for example, by using δ and δ_b approximations where $\delta = \frac{q-1}{h}$ and $\delta_b = \frac{1-q^{-1}}{h}$ with $qf(t) = f(t+1)$ and $q^{-1}f(t) = f(t-1)$. After substituting derivatives by approximations in (2.1), the following model can be constructed

$$\begin{aligned} w(t) &= \varphi^T(t)\theta + \varepsilon(t) \\ w(t) &= D^n y(t) \\ \varphi^T(t) &= [-D^{n-1}y(t), \cdots, -D^0 y(t)] \end{aligned} \tag{2.4}$$

where $\epsilon(t)$ is an equation error. This model can be viewed as a (discrete-time) linear regression. Thus the standard least-squares method can be applied. The corresponding normal equation is

$$[E\varphi(t)\varphi^T(t)]\theta = [E\varphi(t)w(t)] \tag{2.5}$$

and therefore the parameters are estimated by

$$\hat{\theta}_N = [\sum_t \varphi(t)\varphi^T(t)]^{-1}[\sum_t \varphi(t)w(t)] \tag{2.6}$$

We will consider the asymptotic case when the number of data, N, tends to infinity ($\hat{\theta}_N \to \hat{\theta}$). It is shown in [7] that the least-squares estimate of (2.6) will be biased even as the sampling interval tends to zero. A derivative approximation method which solves the bias problem is derived in [7] and is presented in the following lemma.

Lemma 2.1. Let the weights $\{\beta_{k,j}\}$ satisfy the natural conditions (2.3), and in addition

$$\sum_j \beta_{n-1,j} \sum_\ell \beta_{n,\ell}[|\ell - j|^{2n-1} - (\ell - j)^{2n-1}] = 0 \tag{2.7}$$

Then

$$\hat{\theta} = \theta + O(h) \tag{2.8}$$

for arbitrary CAR process parameters θ. □

The condition (2.7) can be satisfied if

$$\ell \geq j \tag{2.9}$$

holds. This means that all measurements used when forming $D^n y(t)$ should be more recent than those used when forming $D^{n-1}y(t)$. A solution to the bias problem is therefore obtained in [7] by shifting the data when approximating $p^n y(t)$.

The shift structure of (2.9) has some drawbacks. Firstly, a large span of data is needed by (2.9) which increases the "end effect" for a finite data set and reduces its effective length. Secondly, in calculating the correlations between $D^n y(t)$ and $D^{n-1}y(t)$, that of the previous stage can not be used due to (2.9). This increases the computational burden. In the next two sections, the least squares approach will be further developed to overcome this shift requirement to achieve better and faster algorithms for CAR parameter estimation.

3. BIAS REMOVAL WITH NO SHIFT

Assuming the order n is known, to avoid the shift one needs to go back to (2.7) and (2.3). The idea is to choose $\beta_{n,\ell}$ with a minimal enlarged time span from that of β_{n-1} such that (2.3) and (2.7) are satisfied, instead of imposing (2.9). To choose $\beta_1, \beta_2, \cdots, \beta_{n-1}$ which satisfy (2.3) is easy. For example, repeated use of the δ or δ_b operators will do. The time span is minimal using δ or δ_b, and is $i+1$ samples for β_i. Once β_{n-1} is chosen, (2.3) and (2.7) for β_n form a set of $n+2$ linear equations. Thus using $n+2$ elements, instead of the minimum number $n+1$ for the δ or δ_b operator, in β_n one can solve for β_n uniquely. Since no shift is required here, one can add one more sample to both ends of β_{n-1} to form the time span for β_n, thus achieving a minimal total time span and maximal symmetry. We then have the following algorithm.

Algorithm 1:

1. Choose $i+1$ elements of β_i that satisfy (2.3), $i = 0, 1, \cdots, n-1$.
2. Choose $n+2$ elements of β_n to satisfy (2.3) and (2.7).
3. Solve the normal equation (2.5) or (2.6), using for example the recursive procedure of [7].

The following example illustrates this method.

Example 3.1 Consider the second order case and using the δ operator. It is obvious that $\beta_{1,0} = -1$ and $\beta_{1,1} = 1$. To find β_2, (2.3) and (2.7) give

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \\ 14 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{2,-1} \\ \beta_{2,0} \\ \beta_{2,1} \\ \beta_{2,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \end{bmatrix} \tag{3.1}$$

(3.1) gives a unique solution for β_2, which is listed in Table 1 together with β_1 and β_0. □

j	-1	0	1	2
$\beta_{0,j}$	0	1	0	0
$\beta_{1,j}$	0	-1	1	0
$\beta_{2,j}$	-0.25	1.75	-2.75	1.25

Table 1: *Weights of derivative approximations, 2nd order, no shift*

4. BIAS REMOVAL BY COMPENSATION

Another possible approach in circumventing the bias problem while avoiding the shift (2.9) is to ignore the requirement (2.7) in solving the normal equation (2.6), and to compensate the bias afterwards, again assuming that the order n is known. Ignoring (2.7), the only requirement then is (2.3), which is satisfied by, for example, the δ or δ_b operators. Under this condition the normal equation (2.5) or (2.6) is *exactly* the same as that of [4] and [5]. Therefore the Levinson-type recursions of [4] and [5] can be immediately applied. With the δ or δ_b operators $\delta^i y(t)$, for example, can be recursively calculated from $\delta^{i-1}y(t)$, whereas in general $D^i y(t)$ can not be obtained from $D^{i-1}y(t)$. Therefore, the

application of [4] and [5] in this proposed approach yields the most computationally efficient algorithm in CAR parameter estimation.

The price for such computational efficiency is of course the bias in the resulting parameter estimate as studied in [7]. Condition (2.7) or (2.9) and the increased computation as used in [7] are precisely designed to correct this bias. Now we show that this bias can also be easily compensated by a simple scaling. To this end we need to analyze the generic matrix element in (2.5).

Lemma 4.1. Under the natural conditions (2.3) the following holds for a CAR(n) process.

$$E\{D^i y(t) D^j y(t)\} = (-1)^{i-1} p^{i+j} r(0^+) + O(h)$$
$$i = 0, \cdots, n-1, \quad j = 0, \cdots, n, \quad i+j < 2n-1 \quad (4.1)$$

$$E\{D^{n-1} y(t) D^n y(t)\} = \xi_n (-1)^{n-1} p^{2n-1} r(0^+) + O(h) \quad (4.2)$$

where

$$\xi_n = \frac{(-1)^{n-1}}{(2n-1)!} \sum_\ell \sum_m \beta_{n,\ell} \beta_{n-1,m} |\ell - m|^{2n-1} \quad (4.3)$$

and $p^k r(0^+)$ denotes the k:th derivative of the covariance function $r(\tau)$ at $\tau = 0^+$. In addition,

$$p^{2k-1} r(0^+) = 0; \quad k = 1, \cdots, n-1 \quad (4.4)$$

□

See [8] for a proof. As a consequence, we can (for $h \to 0$) write the normal equation (2.5) as

$$\begin{bmatrix} (-1)^{n-1} p^{2n-2} r(0^+) & \cdots & \\ (-1)^{n-1} p^{2n-3} r(0^+) & & \vdots \\ \vdots & -p^2 r(0^+) & p^1 r(0^+) \\ & -p^1 r(0^+) & p^0 r(0^+) \end{bmatrix} \theta$$
$$= \begin{bmatrix} \xi_n (-1)^n p^{2n-1} r(0^+) \\ (-1)^{n-1} p^{2n-2} r(0^+) \\ \vdots \\ (-1)^1 p^{2n-n} r(0^+) \end{bmatrix} \quad (4.5)$$

In case $\xi_n = 1$, there would be no bias, as due to the continuous-time Yule-Walker equation

$$(p^n + a_1 p^{n-1} + \cdots + a_n) r(\tau) = 0 \quad \text{for} \quad \tau > 0 \quad (4.6)$$

Note that $\xi_n = 1$ for all n if (2.7) is satisfied. Otherwise $\xi_n \neq 1$, and the solution to (4.5) will have a bias.

Observe that ξ_n is the only factor in (4.5) that causes the bias. Thus one way to eliminate the bias is to scale the first entry of the right hand side vector of (4.5) or (2.5) *before* solving them. But this is not a computationally efficient approach. Thus let us further observe that every other entry in (4.5) is zero due to (4.4). Thus (4.5) can be decomposed into two subsystems as

$$\begin{bmatrix} p^{2n-2} r(0^+) & p^{2n-4} r(0^+) & \cdots \\ p^{2n-4} r(0^+) & p^{2n-6} r(0^+) & \cdots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} a_1 \\ a_3 \\ \vdots \end{bmatrix}$$
$$= - \begin{bmatrix} \xi_n p^{2n-1} r(0^+) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.7)$$

$$\begin{bmatrix} p^{2n-4} r(0^+) & p^{2n-6} r(0^+) & \cdots \\ p^{2n-6} r(0^+) & p^{2n-8} r(0^+) & \cdots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} a_2 \\ a_4 \\ \vdots \end{bmatrix}$$
$$= - \begin{bmatrix} p^{2n-2} r(0^+) \\ p^{2n-4} r(0^+) \\ \vdots \end{bmatrix} \quad (4.8)$$

Obviously the even indexed a's are not biased, and all odd indexed a's are biased by the same factor ξ_n. As ξ_n can be precalculated, it is easy to scale the odd indexed a's *after* they are obtained from (4.7) or (4.5) by the factor $1/\xi_n$ to remove the bias. Therefore for $h > 0$ one can correspondingly solve $\hat{\theta}_N$ from (2.6) with only the natural condition (2.3) imposed, and scale the odd indexed a's afterwards to achieve $O(h)$ accuracy without additional bias. For illustration, we display in Table 2 the numerical values of ξ_n for $n = 1, \cdots, 10$ when the forward delta operator is used. The following algorithm, which is based on [4] using the δ-operator, therefore gives computationally efficient and asymptotically unbiased estimates:

Algorithm 2:

Data Generation:

$$\delta^i y(t) = \frac{1}{\Delta} [\delta^{i-1} y(t+1) - \delta^{i-1} y(t)], \quad i = 1, 2, \cdots$$

Initialization:

$$\mathbf{a}_0 = 1, \quad \tilde{\gamma}_0 = -\frac{1}{\Delta^2}, \quad \tilde{\pi}_0 = \frac{1}{N} \Sigma_t y^2(t),$$
$$\tilde{\alpha}_0 = \frac{1}{N} \Sigma_t y(t) \delta y(t) + \frac{\tilde{\pi}_0}{\Delta}$$
$$\tilde{\gamma}_1 = \frac{\tilde{\alpha}_0}{\tilde{\pi}_0}, \quad \mathbf{a}_1 = [\, 1, \quad \frac{1}{\Delta}(1 - \Delta \tilde{\gamma}_1) \,]^T$$

For $i = 1, 2, \cdots, n-1$:

$$\tilde{\alpha}_i = \frac{1}{N} \Sigma_t y(t) [\, \delta^{i+1} y(t), \quad \cdots, \quad \delta y(t) \,] \mathbf{a}_i$$
$$\tilde{\pi}_i = \frac{1}{N} \Sigma_t \delta^i y(t) [\, \delta^i y(t), \quad \cdots, \quad y(t) \,] \mathbf{a}_i$$
$$\tilde{\gamma}_{i+1} = \frac{\tilde{\alpha}_i}{\tilde{\pi}_i}$$
$$\mathbf{a}_{i+1} = \begin{bmatrix} \mathbf{a}_i \\ 0 \end{bmatrix} + \frac{1}{\Delta}(1 + \frac{\tilde{\gamma}_{i+1}}{\Delta \tilde{\gamma}_i}) \begin{bmatrix} 0 \\ \mathbf{a}_i \end{bmatrix}$$
$$- \frac{\tilde{\alpha}_i}{\Delta \tilde{\alpha}_{i-1}} \left(\Delta \begin{bmatrix} 0 \\ \mathbf{a}_{i-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \mathbf{a}_{i-1} \end{bmatrix} \right)$$

Finalization:

$$\hat{\theta}_N = \left[\mathbf{0} \vdots diag\{ \frac{1}{\xi_n}, \quad 1, \quad \frac{1}{\xi_n}, \quad 1, \quad \cdots \} \right] \mathbf{a}_n$$

n	1	2	3	4	5
ξ_n	1	2/3	0.55	0.4794	0.4304
n	6	7	8	9	10
ξ_n	0.3939	0.3654	0.3422	0.3230	0.3067

Table 2: *Values of ξ_n for the forward delta approximation*

5. SIMULATION EXAMPLE

In this section the two new LS approaches that remove the bias problem are compared with the LS shift method of [7] in a simulation study using the forward delta approximation. Data were generated after (instantaneously) sampling a second order AR process as given in (2.1) with $n = 2$ and $\sigma^2 = 1$. Both a_1 and a_2 were equal to 2. The two parameters were estimated using $N = 10000$ data points. Each trial was repeated 100 times. The number of data points was chosen relatively large in order to emphasize the bias effect. The numerical findings are displayed in Figures 1 and 2 where the theoretical values of the estimates, which are obtained by solving the corresponding normal equations for a fixed h and $N \to \infty$, are also plotted. The standard deviations of the estimates are shown by vertical lines. Note that for $N = 10000$ the "end effects" are negligible. It is interesting, however, that the two new approaches of this paper still give smaller bias than the shift approach of [7].

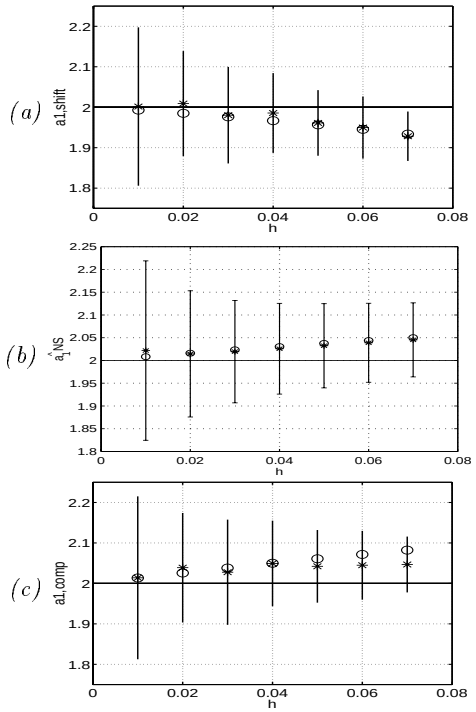

Figure 1: *Experimental(*) and theoretical(o) estimates for $a_1 = 2$ with various h. (a) With Shift; (b) Algorithm 1; (c) Algorithm 2.*

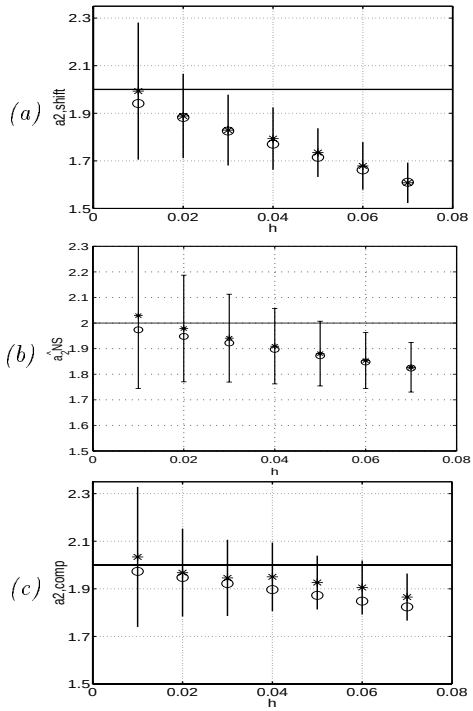

Figure 2: *Experimental(*) and theoretical(o) estimates for $a_2 = 2$ with various h. (a) With Shift; (b) Algorithm 1; (c) Algorithm 2.*

6. REFERENCE

[1] A. R. Bergström, *Continuous Time Econometric Modeling.* Oxford University Press, 1990.

[2] W. L. Luyben, *Process Modeling, Simulation, and Control for Chemical Engineers.* McGraw Hill, Tokyo, 1990.

[3] 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, 1974.

[4] R. Vijayan, H. V. Poor, J. B. Moore and G. C. Goodwin, "A Levinson-type algorithm for modeling fast-sampled data," *IEEE Trans. Autom. Contr.*, Vol. 36, No. 3, 314–321, 1991.

[5] H. Fan and X. Liu, "Delta Levinson and Schur-type RLS algorithms for adaptive signal processing," *IEEE Trans. on Signal Processing* Vol. SP-42, No. 7, 1629–1639, 1994.

[6] A. Feuer and G. C. Goodwin, *Sampling in Digital Signal Processing and Control.* Birkhäuser, Boston, 1996.

[7] T. Söderström, H. Fan, B. Carlsson and S. Bigi, "Least-squares parameter estimation of continuous-time ARX models from discrete-time data," *IEEE Trans. on Automatic Control*, Vol. 42, No. 5, pp. 659-673, May 1997.

[8] T. Söderström, H. Fan, M. Mossberg, and B. Carlsson, "Bias-compensation of least-squares estimates of continuous time AR process parameters," *Report UPTEC 97050R*, School of Engineering, Uppsala University, Sweden.