PERFORMANCE ANALYSIS OF SOME METHODS FOF IDENTIFYING CONTINUOUS-TIME AUTOREGRESSIVE PROCESSES

Torsten Söderström and Magnus Mossberg
Systems and Control Group, Uppsala University
PO Box 27, SE-751 03 Uppsala, Sweden
e-mail: `ts@syscon.uu.se` and `mm@syscon.uu.se`

ABSTRACT

Identification of continuous-time AR processes by least squares and instrumental variables methods using discrete-time data in a 'direct approach' is considered. The derivatives are substituted by discrete-time differences, for example by replacing differentiation by a delta operator. In this fashion the model is casted into a (discrete-time) linear regression. In earlier work we gave sufficient conditions for the estimates to be close to their true values for large data sets and small sampling intervals. The purpose of this paper is to further analyse the statistical properties of the parameter estimates. We give expressions for the dominating bias term of the estimates, for a general linear estimator applied to the continuous-time autoregressive process. Further, we consider the asymptotic distribution of the estimates. It turns out to be Gaussian, and we characterise its covariance matrix, which has a simple form.

1 INTRODUCTION

Parameter estimation of continuous-time systems is an important subject which has numerous applications ranging from control, signal processing, to astrophysics and economics. Examples of publications about identification of continuous-time systems and sampling are [3], [9]–[10]. The book [3] deals with sampling techniques in digital signal processing and control. In [9] a tutorial on identification of continuous-time systems is given and in [10] a bias-compensating least squares method for identification of continuous-time systems is presented.

2 BASIC SETUP

Consider a continuous-time AR model which describes a process with the spectral density

$$\phi(\omega) = \frac{\sigma^2}{A(i\omega)A(-i\omega)} \qquad (1)$$

where

$$A(p) = p^n + a_1 p^{n-1} + \ldots + a_n \qquad (2)$$

In the time domain such a process is represented by

$$\begin{aligned} A(p)y(t) &= e(t) \\ Ee(t)e(s) &= \sigma^2 \delta(t-s) \end{aligned} \qquad ($$

where p denotes the differentiation operator, $y(t)$ is tl output, and $e(t)$ a (continuous-time) white noise sour(The time series is observed at $t = h, 2h, 3h, \ldots, N$ The model order n is supposed to be known. It is interest to estimate the parameter vector

$$\theta = (a_1 \ldots a_n)^T \qquad ($$

from the available discrete-time data.

To rewrite the model (2), we substitute the kth or(differentiation operator p^k by the operator

$$D^k = \frac{1}{h^k} \sum_j \beta_{k,j} q^j$$

where q is the shift operator ($qy(t) = y(t+h)$) and h the sampling interval. We impose the so called 'natu. conditions', [5],

$$\sum_j \beta_{k,j} j^m = \begin{cases} 0 & m < k \\ k! & m = k \end{cases}$$

Then for any differentiable function $f(t)$ it holds tha

$$D^k f(t) = p^k f(t) + O(h)$$

Hence, the approximation error introduced by us: the operator D^k instead of p^k should be small if sampling is fast and the underlying signals are smo(enough.

After substituting derivatives in (2) by the appr(mation (5), the following linear regression model can constructed

$$\begin{aligned} w(t) &= \varphi^T(t)\theta + \varepsilon(t) \\ w(t) &= D^n y(t) \\ \varphi^T(t) &= \left[-D^{n-1}y(t) \ldots - y(t)\right] \end{aligned}$$

where t is now a discrete-time index, and $\varepsilon(t)$ is an equation error.

The identification problem is now to estimate the parameter vector θ in the linear regression model (8) from the available data. It is of interest to examine if and how a least-squares approach

$$\hat{\theta}_N = \left[\Sigma_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\left[\Sigma_{t=1}^{N}\varphi(t)w(t)\right] \qquad (9)$$

can be used. It turns out that the least squares estimate (9) is severly biased even as $h \to 0$, unless the differential operators $\{\beta_{n,j}\}$ and $\{\beta_{n-1,k}\}$ are selected to fulfil certain conditions, [5]. In [6] we studied an alternative approach based on a bias-compensation, while an instrumental variable alternative was considered in [2]. As a matter of fact, several approaches can be described into the same framework. The general estimator can, in the asymptotic case of large data sets, be written as

$$\hat{\theta} = \left[Ez(t)\varphi^T(t)\right]^{-1} F\left[Ez(t)w(t)\right] \qquad (10)$$

where $z(t)$ and F have different interpretations for different approaches. Three different estimators are obtained from the description (10) as follows:

1. The least squares (LS) estimator is obtained with the choices

$$z(t) = \varphi(t), \quad F \equiv I \qquad (11)$$

For the estimate (10) to have a bias term of order $O(h)$, we must further require that the weights fulfil

$$\sum_j\sum_k \beta_{n,j}\beta_{n-1,k}\left[|j-k|^{2n-1} - (j-k)^{2n-1}\right] = 0 \qquad (12)$$

2. In the instrumental variable (IV) case, we nominally take

$$F \equiv I, \quad z(t) = \left(y(t) \; \dots \; y(t-nh+h)\right)^T \qquad (13)$$

3. For the bias-compensated least squares estimator (BCLS), see [6], we have

$$z(t) = \varphi(t), \quad F = \begin{pmatrix} 1/\xi & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \qquad (14)$$

where

$$\xi = \frac{(-1)^{n-1}}{(2n-1)!}\sum_{\ell=0}^{n}\sum_{m=0}^{n-1}\beta_{n,\ell}\beta_{n-1,m}|\ell-m|^{2n-1} \qquad (15)$$

Note that the condition (12) is no longer required.

If the sampling interval h is small, the estimator will give a small bias of order $O(h)$, and therefore the estimate can be written as

$$\hat{\theta} = \theta_0 + \tilde{\theta}h + O(h^2) \qquad (16)$$

3 ANALYSING THE BIAS

We will now present an expression for the bias coefficient $\tilde{\theta}$ in (16) for the general estimator.

Lemma 1 *It holds that*

$$\tilde{\theta} = R_c^{-1}x \qquad (17)$$

where the matrix

$$R_c \triangleq \operatorname{cov}\begin{pmatrix} p^{n-1}y(t) \\ \vdots \\ y(t) \end{pmatrix} \qquad (18)$$

satisfies

$$(R_c)_{i,j} = \begin{cases} 0 & i+j \text{ odd} \\ (-1)^{n-i}p^{2n-i-j}r(0^+) & i+j \text{ even} \end{cases} \qquad (19)$$

for $i,j = 1, \dots, n$ and the vector x is given by

$$x_k = \sum_{l=0}^{n} a_l \sum_i \gamma_{n-k,i}\sum_j \beta_{n-l,j}\frac{|i-j|^{2n-k-l+1}}{(2n-k-l+1)!} \times p^{2n-k-l+1}r(0^+) \qquad (20)$$

□

Proof See [8]. □

As the bias coefficient $\tilde{\theta}$ is a vector, it is tricky how to compare it for different estimators. A pertinent *scalar* measure would highly facilitate such comparisons. One way to proceed, [8], is to start with the scalar measure

$$V = V(\tilde{\theta}) = \frac{1}{4\pi}\int_{-\infty}^{\infty}\left[\frac{\hat{\Phi}(\omega) - \Phi(\omega)}{\Phi(\omega)}\right]^2 d\omega \qquad (21)$$

and approximate it as a quadratic form in $\tilde{\theta}$ for small biases. This turns out to lead to

$$V = \tilde{\theta}^T W \tilde{\theta}, \qquad W = \frac{R_c}{\sigma^2} \qquad (22)$$

see [8] for details.

Example 1 Figure 1 shows which of the three methods LS, IV and BCLS that gives the smallest value of the quadratic form (22), used as a scalar-valued measure of the bias, for different values of a_1 and a_2 of a CAR(2) process. It is clear that the IV method is preferable for processes with parameters in one of two parts of the a_1, a_2-plane, while the BCLS method is preferable for processes with parameters in the other part of the a_1, a_2-plane. Apparently, when identifying a second order process, the LS method never gives the lowest amount of bias, measured as the quadratic form (22). □

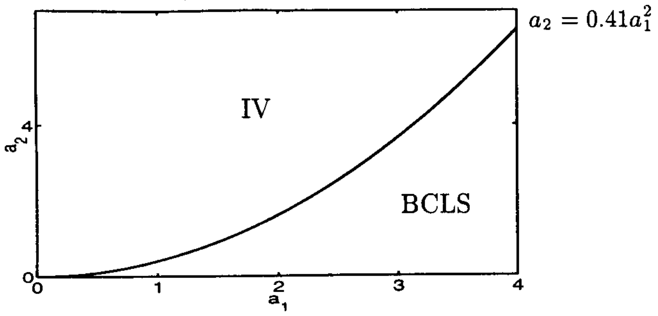

Figure 1: The figure shows which of the three methods LS, IV and BCLS that gives the smallest value of the quadratic form (22), used as a scalar-valued measure of the bias, in different areas of the a_1, a_2-plane.

4 THE COVARIANCE MATRIX OF THE ASYMPTOTIC ESTIMATES

In this section, we will briefly present a general expression for the estimation error covariance in the case of a finite but large N. More details can be found in [7], [8].

Introduce the notation

$$\varepsilon(t) = w(t) - \varphi^T(t)\hat{\theta} \tag{23}$$

and the covariance functions

$$\begin{aligned} R_z(\tau) &= Ez(t+\tau)z^T(t) \qquad & R_{z\varphi}(\tau) &= Ez(t+\tau)\varphi^T(t) \\ R_{z\varepsilon}(\tau) &= Ez(t+\tau)\varepsilon(t) \qquad & r_\varepsilon(\tau) &= E\varepsilon(t+\tau)\varepsilon(t) \end{aligned}$$

The estimation error can, for large N, be written as

$$\begin{aligned} \hat{\theta}_N - \hat{\theta} &= \left[\frac{1}{N}\sum_{k=1}^{N} z(kh)\varphi^T(kh)\right]^{-1} \\ &\times \left[\frac{1}{N}\sum_{k=1}^{N} z(kh)w(kh) - \frac{1}{N}\sum_{k=1}^{N} z(kh)\varphi^T(kh)\hat{\theta}\right] \\ &\approx R_{z\varphi}^{-1}(0)\frac{1}{N}\sum_{k=1}^{N} z(kh)\varepsilon(kh) \end{aligned} \tag{24}$$

The covariance matrix of the estimation error can be approximated as

$$E(\hat{\theta}_N - \hat{\theta})(\hat{\theta}_N - \hat{\theta})^T \approx R_{z\varphi}^{-1}(0)P_o R_{z\varphi}^{-T}(0) \tag{25}$$

where

$$P_o = \operatorname{cov}\left(\frac{1}{N}\sum_k z(kh)\varepsilon(kh)\right) \approx \frac{1}{N}\overline{P} \tag{26}$$

and

$$\overline{P} = \sum_{k=-\infty}^{\infty} \left[R_z(kh)r_\varepsilon(kh) + R_{z\varepsilon}(kh)R_{z\varepsilon}^T(-kh)\right] \tag{27}$$

It turns out that a detailed and more strict analysis of the covariance matrix is quite involved. We will therefore merely state the result here, and refer to the report [7] for the remaining details.

Theorem 1 *Consider the continuous-time autoregres sive process (2), identified with the least squares metho (9), subject to the 'natural conditions' (6) and the con sistency condition, [5]. Then the parameter estimate. are asymptotically Gaussian distributed,*

$$\sqrt{N}(\hat{\theta}_N - \hat{\theta}) \xrightarrow{\text{dist}} \mathrm{N}(0, C) \tag{28}$$

with the covariance matrix

$$C = R_{z\varphi}^{-1}(0)\overline{P}R_{z\varphi}^{-T}(0) \tag{29}$$

Further, the covariance matrix C satisfies

$$\lim_{h\to 0} hC = \sigma^2 R_c^{-1} \tag{30}$$

where R_c is given by (18). □

The consequence of the theorem is that we may ap proximate the covariance matrix of the estimate as

$$\operatorname{cov}(\hat{\theta}_N) \approx \frac{\sigma^2}{Nh}R_c^{-1} \tag{31}$$

Some comments on the above result are appropriate.

1. Note that the expression (31) is not influenced b the specific choice of the approximate derivatives that is the weights $\{\beta_{k,j}\}$.
2. The same result holds for the bias-compensate LS method, [6], and for the instrumental variabl method, [2]. In fact, even the limit (30) takes th same value for these estimators. Hence, all th methods can be said to have the same accuracy fo fast sampling.
3. The matrix R_c is proportional to the noise intensit σ^2. Hence the right hand side does not depend o σ^2.
4. The product Nh is precisely the length of the identi fication experiment (measured in continuous-time)
5. The expression (31) is remarkably simple and ha strong similarities with the corresponding expres sion for a least squares fit to a discrete-time linea regression, in which case one has for consistent es timates

$$\operatorname{cov}(\hat{\theta}) \approx \frac{\lambda^2}{N}[E\varphi(t)\varphi^T(t)]^{-1} \tag{32}$$

   where λ^2 is the innovation variance. It is strikin to note the appearance of h in (31) but not in (32)
6. For the expression (31) to be valid, it is require that N is large and h is small, so that both type of asymptotic evaluations apply. It is most reason able to further require that the total experimen time Nh should be significant compared to the tim constants of the process. This requirement leads t quite large values of N for the expression (31) to b accurate.

5 NUMERICAL STUDIES

In this section a numerical study of the least squares methods [5], the bias-compensated least squares method [6] and the instrumental variables method [2] is made. Also a method based on transfering the poles from an estimated discrete-time ARMA model into continuous-time is studied. The ARMA model parameters were estimated using a prediction error method. Denote the estimated ARMA poles by $\{e^{-\hat{x}_i h}\}_{i=1}^{n}$. Then the estimated continuous-time model is

$$\left[\prod_{i=1}^{n}(p-\hat{x}_i)\right] y(t) = e(t) \tag{33}$$

which can be reformulated as (2).

Data were generated after instantaneous and *exact* sampling a second order AR process

$$(p^2 + a_1 p + a_2)y(t) = e(t) \tag{34}$$

where $e(t)$ is a continuous-time white noise process as given in (2) with $\sigma^2 = 1$, [1], [4]. Note that the continuous-time white noise does not exist, but that the spectrum of y is modelled well. Both a_1 and a_2 were chosen equal to 2 and the two parameters were estimated using $N = 10\,000$ data points. Each trial was repeated 100 times. The variances for the methods considered in the paper are shown in Figure 2, together with the variance given by (31). The theoretical results for the

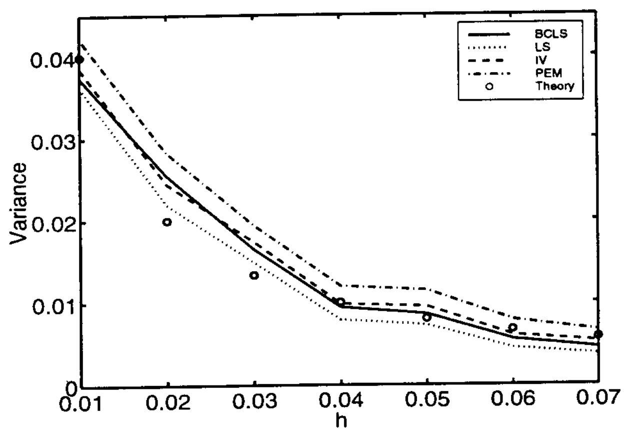

Figure 2: The variance of $\hat{a}_1$.

variance expression derived in Section 4 are supported by the numerical studies. In Figure 2 it is also seen that the variances for the methods are almost the same and that all methods perform as a prediction error method for fast sampling.

REFERENCES

[1] K. J. Åström. *Introduction to Stochastic Control Theory.* Academic Press, New York, NY, 1970.

[2] 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. System Identification*, volume 3, pages 645–650, Copenhagen, Denmark, July 4–6 1994.

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

[4] T. Söderström. *Discrete-Time Stochastic Systems: Estimation and Control.* Prentice–Hall International, Hemel Hempstead, United Kingdom, 1994.

[5] T. Söderström, H. Fan, B. Carlsson, and S. Bigi. Least squares parameter esimation of continuous-time ARX models from discrete-time data. *IEEE Trans. on Automatic Control*, AC-42:659–673, May 1997.

[6] T. Söderström, H. Fan, M. Mossberg, and B. Carlsson. Bias-compensation schemes for estimating continuous-time AR process parameters. In *Proc. 11th IFAC Symp. System Identification*, volume 3, pages 1337–1342, Kitakyushu, Japan, July 8–11 1997.

[7] T. Söderström and M. Mossberg. The accuracy of least squares estimates of continuous-time autoregressive parameters. Technical Report UPTEC 98001R, Systems and Control Group, Uppsala Univ., Uppsala, Sweden, April 1998.

[8] T. Söderström and M. Mossberg. Performance evaluation of methods for identifying continuous-time autoregressive processes. *In preparation*, 1998.

[9] H. Unbehauen and G. P. Rao. Identification of continuous-time systems - a tutorial. In *Proc. 11th IFAC Symp. System Identification,* volume 3, pages 1023–1049, Kitakyushu, Japan, July 8–11 1997.

[10] 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. Control* 56(6):1441–1452, 1992.