Correspondence

The Prediction Error of Autoregressive Small Sample Models

PIET M. T. BROERSEN

Abstract—A fundamental problem in order selection is that one single realization of a stochastic process is used twice, for the estimation of parameters for different model orders and for the selection of the best model order. Parameters are estimated by the minimization of the residual variance; higher model orders with more estimated parameters will always give a smaller residual variance. The purpose of order selection is to find the model order that gives the best fit to other realizations of the same stochastic process. This fit is expressed by the squared prediction error and it will increase if too many parameters are used. The weak parameter criterion (WPC) is an estimate for the squared prediction error, with as special feature that it is computed from the same observations that are used for the estimation of the parameters.

I. INTRODUCTION

The model order in autoregressive estimation is often selected with Akaike's FPE [1] or AIC criterion [2]. This gives satisfactory results in many areas of application, especially if a sufficient number of observations is available. Both criteria are based on the asymptotic properties of estimation. The FPE uses the asymptotic equivalence [3] between ordinary least squares regression and autoregression. However, Jones [4] has shown that the small sample behavior of autoregressive estimates depends on the method of estimation, in contrast with the asymptotic performance. This dependence influences the relation between the residual variance and the squared prediction error. Therefore, order selection should be adapted to the method of parameter estimation. To achieve this, Broersen [5]-[7] introduced the weak parameter criterion (WPC). Berryman [8] evaded the order selection problem by the choice of a model order that depends only on the number of observations. Fougere, who advocates the use of reflection coefficients for stable models [9], used a variance ratio as threshold [10] that was derived with asymptotically based χ^2 distributions. Unlike WPC, that ratio is not adapted to the estimation method.

This paper presents a new justification for the principle of the WPC. In a simulation run, it is possible to compute the squared prediction error from an extra second realization of the stochastic process. This can be compared with the residual variance and the WPC that are obtained from the single realization that would be available in practice. It shows the correspondence between WPC and the squared prediction error; both have the same average over many repeated simulation runs. So, looking for a model order with a small WPC is equivalent to the minimization of the best available estimate for the squared prediction error, as was the original intention of Akaike [1]. Calibration formulae are presented that describe the averages over many simulation runs of WPC, the squared prediction error and the residual variance, all as a function of the order of the estimation model.

Manuscript received August 29, 1987; revised June 26, 1989.

The author is with Delft University of Technology, 2600 GA Delft, The Netherlands. IEEE Log Number 9034422.

II. RESIDUAL VARIANCE

The residual variance S_M^2 for a model of order *M* is defined for Yule–Walker and for Burg estimation as:

$$S_M^2 = \frac{R(0)}{N} \prod_{i=1}^{M} (1 - k_i^2)$$
(1)

where k_i is an estimated reflection coefficient and R(0) is the sum of squares of N observations. For least squares methods, S_M^2 is defined as the sum of all squared residuals divided by the their number. Two different least squares methods exist: the covariance method LSF with only forward or only backward residuals and the modified covariance method LSFB where both forward and backward residuals are minimized simultaneously. The methods are described by Marple [11]. LSF has N - M residuals; LSFB has twice as many residuals, so S_M^2 becomes the sum of all squared residuals divided by (N - M) and 2(N - M), respectively.

The behavior of the residual variance as a function of the order of the estimation model will depend on the true parameters of the autoregressive process. On theoretical grounds, however, the statistical properties of estimated reflection coefficients are independent of the true process for model orders higher than the order of the true process that generates the observations [12, p. 65]. In other words, all estimated models of orders higher than the true order will have some common properties. The simplest process that can be studied is white noise with true order zero and this will be typical for all other processes above their true order. Therefore, the white noise process is discussed first. Section IV gives a simple correction term for the description of the experimental results for nonwhite processes. The average value of the residual variance as a function of the model order follows from (1) as:

...

$$\tilde{S}_M^2 = \sigma_{\epsilon}^2 \prod_{i=1}^{M} (1 - v_i)$$
⁽²⁾

if *M* coefficients are estimated from a white noise process with variance σ_i^2 . The ~ sign denotes an empirical approximation that may serve as a calibration value; its accuracy can be determined in simulations where everything about the true process is known. The WPC coefficients v_i in (2) depend on the method of parameter estimation. They have been defined originally as empirical approximations for the average of the squared reflection coefficients for Burg and Yule-Walker estimates. They have been determined for increasing model orders *i* in white noise simulations [5] and are given by:

$$v_{i,B} = 1/(N+1-i)$$
(3)

$$v_{i,YW} = (N - i) / \{ N(N + 2) \}.$$
(4)

Least squares methods estimate all parameters simultaneously. The variance of estimated parameters from white noise can be approximated [7] with:

$$v_{i,\text{LSFB}} = 1/(N + 2 - 1.5i)$$
 (5)

$$v_{i\,\text{LSE}} = 1/(N+2-2i). \tag{6}$$

GA Delit, The The lower drawn lines in Figs. 1-4 represent the empirical relation (2), for $\sigma_{\epsilon}^2 = 1$. The crosses are measured averages $\overline{S_M^2}$, obtained

0096-3518/90/0500-0858\$01.00 © 1990 IEEE



Fig. 1. Residual measures for LSFB in white noise with $\sigma_e^2 = 1$, N = 20.



Fig. 2. Residual measures for LSF in white noise with $\sigma_c^2 = 1$, N = 20.



Fig. 3. Residual measures for Burg in white noise with $\sigma_t^2 = 1$, N = 20.

in Monte Carlo simulations by averaging the residual variance as a function of the model order for 500 realizations of 20 white noise observations each. The figures show that the residual variance depends on the method of estimation. However, by using the different WPC coefficients for each method, all measured averages are accurately described by the same formula (2).



Fig. 4. Residual measures for Yule-Walker in white noise with $\sigma_r^2 = 1$, N = 20.

III. WPC AND SQUARED PREDICTION ERROR The WPC has been defined [5] as:

WPC(M) =
$$S_M^2 / \prod_{i=0}^{M} (1 - 2v_i)$$
 (7)

with $v_o = 0$. The WPC(M) is calculated for each order from zero to a prefixed maximum order L. In order selection, that order M is selected for which WPC(M) attains a minimum. In this paper we calibrate the behavior of the average of WPC(M) as a function of the order of the fitted model, $M = 0, 1, \dots, 15$, without any selection. An empirical formula for WPC(M) in white noise follows with (2) and (7) as:

$$W\tilde{P}C(M) = \sigma_{\epsilon}^{2} \prod_{i=0}^{M} (1 - v_{i}) / \prod_{i=0}^{M} (1 - 2v_{i}).$$
(8)

The squared prediction error PE(M) for order M is defined as

$$\operatorname{PE}(M) = \frac{1}{N - M} \sum_{n=M+1}^{N} \left\{ y_n - \sum_{i=1}^{M} \hat{a}_i y_{n-i} \right\}^2$$
(9)

where the estimates \hat{a}_i , computed from one realization x_n , are substituted into a different realization y_n of the same stochastic process. The right-hand side of the same formula (9) denotes the LSF residual variance S_M^2 if the \hat{a}_i are obtained from the same realization y_n .

The upper drawn lines in Figs. 1-4 represent the formula for WPC(M). Also WPC(M) and PE(M) are indicated; these are measured averages over 500 simulation runs, evaluated for $M = 0, 1, \dots, 15$. It turns out that WPC(M) is a good estimate for PE(M), for the calculation of which two independent realizations are required in each simulation run. Moreover, WPC(M) of (8) gives an accurate calibration formula for both WPC(M) and PE(M). The extension of these results to general autoregressive processes is given in the next section.

IV. AUTOREGRESSIVE PROCESSES

Suppose that an autoregressive process is characterized by *K* reflection coefficients with true values κ_i , $i = 1, \dots, K$. Although a one-to-one correspondence exists between reflection coefficients and parameters, the parametrization with reflection coefficients has an advantage here, because the variance of the observations is given by:

$$E\{S_0^2\} = E\left\{\frac{R(0)}{N}\right\} = \sigma_\epsilon^2 / \prod_{i=1}^K (1 - \kappa_i^2)$$
(10)

for an autoregressive process with K reflection coefficients driven by a white noise sequence ϵ_n with variance σ_i^2 . An empirical ap-



Fig. 5. Residual measures for LSFB in AR process generated with $\sigma_t^2 = 1$, N = 20, $\kappa_1 = \kappa_3 = -0.5$, $\kappa_2 = \kappa_4 = 0.5$.

proximation for S_M^2 can be found by making use of the fact that, for model order M less than the true order K, the fraction of the variance (10) with index between M and K cannot be explained. Using that fraction as a correction for (2) and (8) gives:

$$\bar{S}_{M}^{2} = \sigma_{\epsilon}^{2} \prod_{i=1}^{M} (1 - v_{i}) / \prod_{j=M+1}^{K} (1 - \kappa_{j}^{2})$$
(11)

$$\widetilde{WPC}(M) = \sigma_{\epsilon}^{2} \prod_{i=0}^{M} (1 - v_{i}) / \left\{ \prod_{i=0}^{K} (1 - 2v_{i}) \prod_{j=M+1}^{K} (1 - \kappa_{j}^{2}) \right\}$$
(12)

with the final product replaced by 1 for $M \ge K$. Both σ_{ϵ}^2 , K and κ_i are known exactly in our simulations. The empirical approximations (11) and (12) are a mixture of asymptotic values for κ_j^2 and small sample values for v_i . Fig. 5 shows simulation results for an LSFB example of true order K = 4. For model orders $M \ge 4$, the drawn lines are equal to those of the white noise process in Fig. 1, as follows from (6) and (11), (8) and (12). The measured averages

 S_{M}^2 , WPC (M) and $\overline{PE}(M)$ are very close to the empirical approximations. This gives two conclusions. First, (11) is a good choice for a calibration formula for the average of S_{M}^2 and (12) for WPC (M) and PE (M). Second, the white noise behavior given in Fig. 1 is representative for the behavior of other models above the true process order. Of course, numerous other simulations with different N, other parameters and other estimation methods support these conclusions; see also [13].

The difference between the small sample behavior of the WPC on the one hand and asymptotically based selection criteria on the other hand is indicated in Fig. 5 by the measured average of the final prediction error: $FPE(M) = S_M^2(N+M)/(N-M)$ of Akaike [1]. This final prediction error doesn't give an accurate approximation of the squared prediction error in small samples. More over, a selected FPE order will depend heavily on the maximum order that is considered for selection [13].

For M > K, (2) and (11) become identical, and (2) can be used to compute an estimate for σ_i^2 from a measured value for S_M^2 . A comparison of the quality of selected models shows [13] that order selection with WPC yields models with a significantly lower prediction error than with other selection methods over a wide range of circumstances. The reason is the close correspondence between WPC and PE in small samples.

V. CONCLUSION

Simulations have shown that the average WPC is about equal to the average squared prediction error PE. In other words: WPC, that can be computed from one single realization of a stochastic process, is an estimate for the important squared prediction error, that can only be found if two realizations are available. Empirical calibration formulae are given for the WPC and PE and for the residual variance. They are adapted to the various estimation methods by using different WPC coefficients v_i . All given formulae are multiplicative with $1 - v_i$ or $1 - 2v_i$; in this sense, the small sample behavior of autoregressive processes is multiplicative. The presented formulae describe the numerous simulation results and hence may be considered as a useful tool in modeling practice.

REFERENCES

- H. Akaike, "Fitting autoregressive models for prediction," Ann. Inst. Statist. Math., vol. 21, pp. 243-247, 1969.
- [2] —, "A new look at the statistical model identification," IEEE Trans. Automat. Contr., vol. AC-19, pp. 716–723, 1974.
- [3] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, vol. 11, pp. 173–220, 1943.
- [4] R. H. Jones, "Autoregression order selection," *Geophysics*, vol. 41, pp. 771-773, 1976.
- [5] P. M. T. Broersen, "Selecting the order of autoregressive models from small samples," *IEEE Trans. Acoust.*, Speech, Signal Processing, vol. ASSP-33, pp. 874-879, 1985.
- [6] —, "Subsets of reflection coefficients," in Proc. ICASSP 86 (Tokyo, Japan), 1986, pp. 1373–1376.
- [7] —, "Subsets of autoregressive parameters." in *Proc. EUSIPCO 86* (The Hague, The Netherlands), 1986, pp. 961–964.
- [8] J. G. Berryman, "Choice of operator length for maximum entropy spectral analysis," *Geophysics*, vol. 43, pp. 1384–1391, 1978.
 [9] P. F. Fougere, "A solution to the problem of spontaneous line split-
- [9] P. F. Fougere, "A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis," J. Geophys. Res., vol. 82, pp. 1051-1054, 1977.
- [10] —, "Spectrum model-order determination via significant reflection coefficients," in *Proc. ICASSP 85* (Tampa, FL), 1985, pp. 1345– 1347.
- [11] S. L. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs,: Prentice-Hall, 1987.
- [12] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed. San Francisco: Holden-Day, 1976.
- [13] P. M. T. Broersen, "A comparison of criteria for autoregressive order selection," in *Preprints 8th Symp. Identification Syst. Parameter Estimation* (Beijing, China), 1988, pp. 1054-1059.

Optimal Hydrophone Placements Under Random Perturbations

NILANJAN CHANDRA AND CHARLES KNAPP

Abstract—Hydrophone location uncertainty can deteriorate the performance of an array in the estimation of the location of an acoustic source. An approach is proposed for hydrophone placement which minimizes a conventional bearing or range variance bound averaged over the random deviation of sensor positions. That is, we use bounds on localization error covariances called Cramer-Rao lower bounds (CRLB's) to distribute the sensors in an effective manner for line and towed arrays.

Manuscript received December 2, 1987; revised May 9, 1989,

- N. Chandra is with the Department of Engineering Technology, Central Michigan University, Mt. Pleasant, MI 48859.
- C. Knapp is with the Department of Electrical Engineering, University of Connecticut, Storrs, CT 06269. IEEE Log Number 9034430.

-

0096-3518/90/0500-0860\$01.00 © 1990 IEEE