The Determination of the Order of an Autoregression
Author(s): E. J. Hannan and B. G. Quinn
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 41, No. 2 (1979), pp. 190-195
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2985032
Accessed: 18-02-2020 03:08 UTC

J. R. Statist. Soc. B (1979),
41, *No.* 2, *pp.* 190–195

The Determination of the Order of an Autoregression

By E. J. HANNAN and B. G. QUINN

The Australian National University

[Received September 1978. Revised December 1978]

SUMMARY

It is shown that a strongly consistent estimation procedure for the order of an autoregression can be based on the law of the iterated logarithm for the partial autocorrelations. As compared to other strongly consistent procedures this procedure will underestimate the order to a lesser degree.

Keywords: AUTOREGRESSION; AUTOREGRESSIVE ORDER; AIC; LAW OF ITERATED LOGARITHM; STRONG CONVERGENCE

1. INTRODUCTION

THERE is a considerable literature on the determination of the order, k, of an autoregression

$$\sum_0^k \alpha(j)\{x(n-j)-\mu\} = \varepsilon(n), \quad \alpha_0 = 1. \tag{1}$$

The context within which this is set is that of an ergodic stationary sequence $x(n)$ with mean μ and finite variance for which the $\varepsilon(n)$ are the linear innovations, so that

$$\sum_0^k \alpha(j) z^j \neq 0, \quad |z| \leqslant 1; \; \mathscr{E}\{\varepsilon(m)\,\varepsilon(n)\} = \delta_{mn}\,\sigma^2. \tag{2}$$

Often the $\varepsilon(n)$ are assumed to be Gaussian or at least independent but for almost all of the limit theory it is necessary only to assume that

$$\mathscr{E}\{\varepsilon(n)\,|\,\mathscr{F}_{n-1}\} = 0, \quad \mathscr{E}\{\varepsilon(n)^2\,|\,\mathscr{F}_{n-1}\} = \sigma^2, \quad \mathscr{E}\{\varepsilon(n)^4\} < \infty, \tag{3}$$

where $\mathscr{F}_n$ is the σ-algebra determined by $x(m), m \leqslant n$ (or equivalently by $\varepsilon(m), m \leqslant n$). The first of the conditions (3) is natural since it is equivalent to the requirement that the best linear predictor be the best predictor (in the least squares sense) but the second is not so natural. It could be eliminated at the expense of an increase in complexity. (See the end of Section 2.)

The best known rule for determining the true value of k, which we shall call k_0, is probably that due to Akaike (1969), namely the minimization of $\mathrm{AIC}(k) = \ln \hat{\sigma}_k^2 + 2k/N$ where N is the sample size and $\hat{\sigma}_k^2$ is the estimate of σ^2 obtained from the Yule–Walker relations for an autoregression of order k (see (4) below). Shibata (1976) has investigated the asymptotic properties of this estimate and his investigation shows that the estimate is not consistent but overestimates k_0, asymptotically, with a non-zero probability. Of course this need not be a defect in the method, apart altogether from the asymptotic nature of consistency, for the data will not be generated, precisely, by an autoregression and the estimated autoregression may be fitted only as an approximation. An alternative method of estimation has been proposed by Parzen (see Parzen, 1974) but again the estimate will not be consistent. The purpose of this paper is to provide an estimation procedure, of the same type as that due to Akaike, namely based on the minimization of $\ln \hat{\sigma}_k^2 + kC_N$, that is strongly consistent for k_0 and for which C_N decreases as fast as possible. Other strongly consistent procedures have been introduced (see Akaike, 1977; Rissanen, 1978; Shwarz, 1978). In these C_N is of order $N^{-1}\log N$ and as we show below this is not the slowest rate of increase possible. Since C_N

decreases faster than $N^{-1}\log N$, in the procedure introduced below, it is evident that it will underestimate the order, in large samples, less than these procedures.

The criterion suggested here is

$$\phi(k) = \ln \hat{\sigma}_k^2 + N^{-1} 2kc \ln\ln N, \quad c > 1.$$

It follows immediately from the strong consistency of the estimate, $\hat{k}$, minimizing $\phi(k)$ that, because k_0, $\hat{k}$ are integers, all of the established asymptotic theory for autoregressive estimation (under (3)) when k_0 is known applies also when $\hat{k}$ estimates k_0. One would not wish to make too much of all of this because of the asymptotic nature of the theorems and the fact, earlier referred to, that the true process may not be an autoregression. Nevertheless the result is of some interest. In Section 2 we shall prove the strong consistency of $\hat{k}$ when it is prescribed *a priori* that $k_0 \leqslant K < \infty$. It follows that K may be allowed to increase with N and the result will still hold but we do not know how fast that rate of increase may be. In Section 3 the result will be illustrated by some simulations and some further discussion of related results will be given. For those that do not wish to read the precise proof an outline is given below. We may write (Hannan, 1970, Chapter VI)

$$\phi(k) = \ln s^2 + \sum_1^k \ln\{1-\hat{\rho}^2(j\,|\,j-1)\} + N^{-1} 2kc \ln\ln N,$$

where s^2 is the sample variance. Thus $\phi(k)$ increases at k by $\ln\{1-\hat{\rho}^2(k\,|\,k-1)\}+2cN^{-1}\ln\ln N$. When $k = k_0$ then the true partial autocorrelation $\rho(k\,|\,k-1)$, is equal to $-\alpha(k_0)\neq 0$, since k_0 is the true order. Since $\hat{\rho}(k\,|\,k-1)$ estimates $\rho(k\,|\,k-1)$ consistently it is evident that an absolute minimum of $\phi(k)$ cannot be reached, asymptotically, for $k<k_0$. (The same, of course, is true for Akaike's and Parzen's estimates.) For $k>k_0$ we shall show that the law of the iterated logarithm (LIL) holds (see the statement of Theorem 1 below) so that there is an N_0, $P(N_0<\infty)=1$, such that for $N>N_0$, a.s., $\ln\{1-\hat{\rho}^2(k\,|\,k-1)\}+2cN^{-1}\ln\ln N>0$, $k_0<k\leqslant K$. Thus the strong consistency follows.

2. THE STRONG CONSISTENCY OF $\hat{k}$

Put

$$c(n) = N^{-1}\sum_{m=1}^{N-n}\{x(m)-\bar{x}\}\{x(m+n)-\bar{x}\} = c(-n).$$

We may assume $\mu = 0$. Put $\gamma(n) = \mathscr{E}\{x(m)\,x(m+n)\}$. It is now convenient to write $\hat{\alpha}_k(j)$ for the estimate of $\alpha(j)$ obtained from the Yule–Walker relations of order k, namely from

$$\sum_{j=0}^{k}\hat{\alpha}_k(j)\,c(n-j) = \delta_{0n}\,\hat{\sigma}_k^2; \quad n = 0,1,\ldots,k; \quad \hat{\alpha}_k(0) = 1. \tag{4}$$

It is well known (Hannan, 1970, Chapter VI) that $\hat{\rho}(k\,|\,k-1) = -\hat{\alpha}_k(k)$ and that

$$\hat{\alpha}_k(k) = \sum_0^{k-1}\hat{\alpha}_{k-1}(j)\,c(k-j)\Big/\sum_0^{k-1}\hat{\alpha}_{k-1}(j)\,c(j). \tag{5}$$

In a recursive calculation the denominator would be better computed as $\hat{\sigma}_{k-1}^2$ (see (4)) via

$$\hat{\sigma}_k^2 = \{1-\hat{\alpha}_k^2(k)\}\,\hat{\sigma}_{k-1}^2. \tag{6}$$

Theorem 1. Let $x(n)$ be generated by (1) for $k = k_0$ subject to (2), (3) and let $\hat{\rho}(k\,|\,k-1)$ be as defined above. Then

$$\hat{\rho}(k\,|\,k-1) = b_k(N)\,N^{-\frac{1}{2}}(2\ln\ln N)^{\frac{1}{2}}, \quad k>k_0, \tag{7}$$

where the sequence $b_k(N)$, $N = 1,2,\ldots$, has as its limit points the interval $[-1,1]$.

Theorem 2. Under the conditions of Theorem 1 if $\hat{k}$ is chosen to minimize $\phi(k)$ over $k \leqslant K$, for $k_0 \leqslant K$ then $\hat{k}$ converges a.s. to k_0.

Proof. We first point out that the mean correction may be eliminated. Indeed $\hat{\alpha}_k(k)$ is the ratio of two determinants whose elements are obtained from the $c(n)$ and the denominator determinant is that of the matrix having $c(m-n)$ in row m column n, $m, n = 1, \ldots, k$. The $c(n)$ are strongly consistent, by ergodicity, and the denominator determinant therefore converges a.s. to a non-zero constant. Hence it is sufficient to point out that

$$c(n) - N^{-1} \sum_1^{N-n} x(m)\, x(m+n) = \left(1 - \frac{n}{N}\right) \bar{x}^2 - \bar{x} N^{-1} \sum_{n+1}^{N} x(m) - \bar{x} N^{-1} \sum_1^{N-n} x(m)$$

and that the right side is $O(\ln \ln N/N)$ (where here and below such order relations hold a.s.). This follows from Heyde and Scott (1973) who show that $\bar{x}$ and hence also the remaining factors in the last two terms, obey the LIL.

We next point out that $\hat{\rho}(k \mid k-1)$ may be replaced by

$$\tilde{\rho}(k \mid k-1) = \frac{1}{\hat{\sigma}^2_{k-1}} N^{-1} \sum_{n=k+1}^{N} \left\{ \sum_{j=0}^{k-1} \hat{\alpha}_{k-1}(j)\, x(n-j) \right\} \left\{ \sum_{j=0}^{k-1} \hat{\alpha}_{k-1}(j)\, x(n-k+j) \right\}.$$

Indeed that numerator is

$$\sum_0^{k-1} \sum \hat{\alpha}_{k-1}(i)\, \hat{\alpha}_{k-1}(j) \{ c(j+i-k) + o(N^{-\frac{1}{2}}) \}$$

by the third part of (3). Since, by (4),

$$\sum \hat{\alpha}_{k-1}(i)\, c(j+i-k) = 0, \quad j \leqslant k-1$$

while $\hat{\alpha}_{k-1}(0) = 1$ this reduces to

$$\sum_0^{k-1} \hat{\alpha}_{k-1}(j)\, c(k-j) + o(N^{-\frac{1}{2}})$$

(observing that the $\hat{\alpha}_{k-1}(j)$ are uniformly bounded since they are known to satisfy the first part of (2)). It is sufficient therefore to examine $\tilde{\rho}(k \mid k-1)$. The denominator of this quantity converges almost surely to σ^2 since it is known that under (3) the Yule–Walker equations yield consistent estimates (see Hannan, 1973, for example). It will be sufficient therefore to establish the LIL for the numerator. Put

$$\xi_k(n) = \sum_{j=0}^{k_0} \alpha(j)\, x(n-k+j)$$

and observe that $\xi_k(n)$ has variance σ^2 since its variance when expressed in terms of the $\gamma(n)$ is precisely the same as that of

$$\varepsilon(n) = \sum_{j=0}^{k_0} \alpha(j)\, x(n-j).$$

Putting $\delta_{k-1}(j) = \hat{\alpha}_{k-1}(j) - \alpha(j)$ with $\alpha(j) = 0, j > k_0$, the numerator of $\tilde{\rho}(k \mid k-1)$ is

$$N^{-1} \sum_1^N \left\{ \varepsilon(n) + \sum_{j=1}^{k-1} \delta_{k-1}(j)\, x(n-j) \right\} \left\{ \xi_k(n) + \sum_1^{k-1} \delta_{k-1}(j)\, x(n-k+j) \right\} + O(N^{-1}).$$

It will follow from what is said below that

$$\sum_1^N \varepsilon(n)\, x(n-k+j), \quad j = 1, \ldots, k-1, \tag{8}$$

obeys the LIL. It will thus be sufficient to show that

$$c(n)-\gamma(n) = o(N^{-\frac{1}{2}+\eta}), \quad \eta>0. \tag{9}$$

Indeed if this is so then $\delta_{k-j}(j)$ is of this order also, by (4) for $n = 1, \ldots, k$. Also, then,

$$N^{-1}\sum_{1}^{N} x(n-j)\,\xi_k(n) = \sum_{0}^{k_0} \alpha(n)\,c(j-k+n)+O(N^{-1})$$

$$= \sum_{0}^{k_0}\alpha(n)\,\{c(j-k+n)-\gamma(j-k+n)\}+O(N^{-1}) = O(N^{-\frac{1}{2}+\eta}),$$

and consequently the numerator of $\tilde{\rho}(k\,|\,k-1)$ is, using the LIL for (8) and (9),

$$N^{-1}\sum_{1}^{N}\varepsilon(n)\,\xi_k(n)+O(N^{-1+\eta}), \quad \eta>0.$$

In fact the $c(n)$ also obey the LIL. This is shown in Heyde (1974) but since the proof given in detail covers only the case where the $\varepsilon(n)$ are independent we point out how the weaker result (9) may be established. The autocovariance $c(n)$ is just the mean of $x(m)\,x(m+n)$, $m = 1, 2, \ldots$, and this process has mean $\gamma(n)$ and absolutely continuous spectrum with spectral density that is continuous. Indeed the rth autocovariance of $x(m)\,x(m+n)$ is (Hannan, 1970, Chapter IV)

$$\gamma(r)^2+\gamma(r+n)\,\gamma(r-n)+\kappa_4\sum_{0}^{\infty} a(j)\,a(j+r)\,a(j+n)\,a(j+r+n),$$

where the $a(j)$ are generated by $\{\Sigma\,\alpha(j)\,z^j\}^{-1}$ and hence converge to zero at a geometric rate. The result now follows by standard theorems on the mean of a stationary process of this kind (see Hannan, 1978, for example). Thus we need consider the LIL only for

$$X(N) = \sum_{1}^{N}\varepsilon(n)\,\xi_k(n).$$

However, $X(N)$ is a martingale with stationary, ergodic, square integrable martingale differences and the result follows from Heyde and Scott (1973, Corollary 2).

The proof of Theorem 2 is now immediate along the lines given in Section 1.

It follows also from Theorem 1 that if $\phi(k)$ is replaced by $\log\hat{\sigma}_k^2+kC_N$, $\lim NC_N/\ln\ln N<2$ then strong consistency of the resulting estimate of k_0 cannot hold since $\ln\{1-\hat{\rho}^2(k\,|\,k-1)\}+C_N$ will be negative infinitely often, with respect to N, for any $k>k_0$.

Finally, we observe that the second part of (3) was used only to ensure that $\varepsilon(n)\,\xi_k(n)$ has variance σ^4 via the evaluation

$$\mathscr{E}\{\varepsilon(n)^2\,\xi_k(n)^2\} = \mathscr{E}[\mathscr{E}\{\varepsilon(n)^2\,|\,\mathscr{F}_{n-1}\}\,\xi_k(n)^2] = \sigma^2\,\mathscr{E}\{\xi_k(n)^2\} = \sigma^4.$$

If the second part of (3) is suppressed, Theorems 1 and 2 will still hold but we shall need to insert a further constant on the right of (7), namely $\nu = \sigma^{-2}[\mathscr{E}\{\varepsilon(n)^2\,\xi_k(n)^2\}]^{\frac{1}{2}}$ and correspondingly c will now have to be larger than ν. It is easy to see that ν can be strongly consistently estimated. However, we shall not go into details as it seems unlikely that this additional statistical computation would be used.

3. SIMULATIONS

All the simulations reported below were performed on a Univac U-1100 using pseudo-normal random numbers. The simulations are designed to show how $\phi(k)$ performs for "small" N. K was set equal to 15. Of course for small N the LIL can hardly be viewed as operating, as can be appreciated from the fact that $2\ln\ln 100\simeq 3$ while the 5 per cent point for a test of significance for $\hat{\rho}^2(k_0+1\,|\,k_0)$ would be about 4. For this reason also the value $c=1$ was used since it would seem pedantic, for the values of N used in Table 1, to choose some value of c such as 1·01. In Table 1 below the results of 100 replications of a number of simulations, for varying N and varying $\alpha(1)$ (k_0 being unity), are presented.

TABLE 1

Frequencies of estimated order. First-order autoregression

		N									
		50		100		200		500		1000	
α	$\hat{k}$	ϕ	AIC	ϕ	AIC	ϕ	AIC	ϕ	AIC	ϕ	AIC
0·1	0	85	72	79	59	55	41	30	17	9	2
	1	9	10	16	22	40	37	66	57	86	60
	2	2	5	5	10	2	9	3	14	4	17
	3	1	4	0	2	2	4	1	7	1	4
	>3	3	9	0	7	1	9	0	5	0	17
0·3	0	35	24	8	5	2	1	0	0	0	0
	1	56	51	87	72	95	79	90	70	94	74
	2	5	9	4	11	2	7	7	13	5	11
	3	1	6	1	4	0	7	3	7	1	7
	>3	3	10	0	8	1	6	0	10	0	8
0·5	0	1	1	0	0	0	0	0	0	0	0
	1	78	66	90	74	91	74	93	69	93	70
	2	14	17	7	11	6	14	5	11	6	12
	3	5	8	1	4	1	8	2	4	1	4
	>3	2	8	2	11	2	4	0	16	0	14
0·7	0	1	0	0	0	0	0	0	0	0	0
	1	78	67	91	72	93	78	94	78	95	70
	2	19	21	5	10	3	9	6	11	5	17
	3	1	6	2	7	3	5	0	5	0	2
	>3	1	6	2	11	1	8	0	6	0	11

The table shows what is to be expected, namely underestimation of the order relative to AIC for smaller N and smaller α but better results than for AIC for larger N or larger α.

In a second simulation we took $x(n)=\varepsilon(n)+\varepsilon(n-1)$ and again used $\phi(k)$ and AIC(k). The results of 100 replications for $N=100$ are shown in Table 2.

TABLE 2

Frequencies of estimated order. First-order moving average

	$\hat{k}$											
	0	1	2	3	4	5	6	7	8	9	10	>10
ϕ	0	0	10	19	30	12	15	5	7	1	1	0
AIC	0	0	2	3	19	9	22	13	15	4	7	6

Here $\rho(k\,|\,k-1)=(-1)^k/k$. Since $\{2N^{-1}\ln\ln N\}^{\frac{1}{2}}=0{\cdot}18$, $(2N^{-1})^{\frac{1}{2}}=0{\cdot}14$ it is evident that the results are much as might be expected.

It is not reasonable to expect any definite conclusions. If N is large and an autoregression is thought to be a good approximation then the use of $\phi(k)$ would have something to recommend it. This might not be true in other circumstances. The method provides some compromise, based on a rather precise analysis, between procedures such as those of Akaike (1977), with $C_N = N^{-1} \log N$ designed, in a Bayesian analysis, for a true autoregressive situation and those of the type of AIC designed for fitting an autoregression where the true structure may be more general.

REFERENCES

AKAIKE, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.*, **21**, 243–247.
—— (1977). On entropy maximisation principle. In *Applications of Statistics* (P. R. Krishnaiah, ed.), pp. 27–41. Amsterdam: North Holland.
HANNAN, E. J. (1970). *Multiple Time Series*. New York: Wiley.
—— (1973). The asymptotic theory of linear time series models. *J. Appl. Prob.*, **10**, 130–145.
—— (1978). Rates of convergence for time series regression. *Adv. Appl. Prob.*, **10**, 740–743.
HEYDE, C. C. (1974). An iterated logarithm result for autocorrelations of a stationary linear process. *Ann. Prob.*, **2**, 328–332.
HEYDE, C. C. and SCOTT, D. J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Prob.*, **1**, 428–436.
PARZEN, E. (1974). Some recent advances in time series modeling. *IEEE Trans.*, **AC-19**, 723–730.
RISSANEN, J. (1978). Modeling by shortest data description. *Automatica*, **14**, 465–471.
SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117–126.
SHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.