International Statistical Review, (1985), **53**, 3, pp. 301–329. Printed in Great Britain

Methods for Determining the Order of an Autoregressive-Moving Average Process: A Survey

Jan G. de Gooijer[1], Bovas Abraham, Ann Gould and Lecily Robinson[2]

[1]*Department of Economic Statistics, University of Amsterdam, 1011 NH Amsterdam, Holland.* [2]*Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada N2L 3G1.*

Summary

Determining the order of an autoregressive-moving average process is an important and difficult part of time series analysis. Often time series analysts follow the Box–Jenkins approach to time series modelling. This approach relies somewhat on the subjective judgement of the analyst. Many other less heuristic methods have been proposed and used in the literature. In this survey the most important of these order determination methods are reviewed and their theoretical and practical relevance are discussed.

Key words: Akaike's information criterion; Bayesian methods; Corner method; Cross-validation; Extended autocorrelations; Final prediction error; Generalized partial correlations; Inverse autocorrelations; Lagrange multiplier test; S- and R-arrays.

1 Introduction

Autoregressive moving average models of order (p, q), which are often abbreviated to ARMA(p, q) models, are frequently used to describe and forecast time series observations. These models are commonly expressed as:

$$\phi_p(B)y_t = \theta_q(B)a_t, \tag{1.1}$$

where B is a backward shift operator such that $B^m y_t = y_{t-m}$ for any integer m,

$$\phi_p(B) = 1 - \sum_{j=1}^{p} \phi_j B^j, \quad \theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$$

are polynomial operators such that $\phi_p(B)\theta_q(B)$ has its roots outside the unit circle and $\{a_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of independent and identically distributed normal random variables with $E(a_t) = 0$ and $E(a_t^2) = \sigma_a^2$. Such a sequence is referred to as a white noise sequence. The y_t's may in general represent the dth difference or any other suitable transformation of some nonstationary time series.

In recent years the topic of order determination or specification has attracted considerable attention in the time series literature and in those research areas of control theory, econometrics and statistics which are closely related to time series analysis. Specifying the model means finding estimates of the order (p, q) of the process. If these estimates are available the parameters $\phi = (\phi_1, \ldots, \phi_p)$, $\theta = (\theta_1, \ldots, \theta_q)$ and σ_a^2 can then be estimated. We are not concerned with the latter here, although some of the methods do require estimation of the parameters of different models which can then be compared. The 'true'

order of a process is rarely, if ever, known and a difficult and delicate part of time series analysis then is the selection, based on a finite set of observations, of the order (p, q) of the process to be fitted. It often happens that the selected model is a simplification of the true model which usually is complicated. What is assumed and hoped is that it adequately describes the underlying process and that it may be potentially useful for some purpose.

Various methods have been proposed and explored in the literature, but still many practitioners usually follow the Box–Jenkins approach to time series modelling. To a very large extent this method is based on making inferences from the patterns of the sample autocorrelation and partial autocorrelation functions of the series. The characteristic properties of the corresponding theoretical functions for various types of ARMA models are used hereby as a reference; see, for example, Box & Jenkins (1976, Ch. 3). This approach relies heavily on the heuristic judgement of the time series analyst. Choosing the appropriate order of a process by making inferences from the patterns of the sample autocorrelation function is usually difficult and hence alternative methods have been proposed.

Although there exists no universal panacea to the question of determining the order of a time series model from empirical data, a large number of procedures has been put forth to choose the most appropriate model structure. In this survey we describe the main features of what seem to be the most useful ones. Our study does not have the pretention of being exhaustive. Neither is it an attempt to cover the whole of the literature on the technical details of the methods which, in some cases, can be very extensive. The aim is to give the broad framework of the most important order-determination techniques, used in time series analysis today. References will be given so that more detailed examination of particular techniques can be made from the original papers.

We acknowledge that there have been other studies with a similar objective. They differ from the present investigation in that they are restricted to linear regression models and/or focus on a more limited range of order determination procedures. Excellent surveys of the most important methods for the selection of the order of linear regression models are given by Amemiya (1980) and Geweke & Meese (1981). For a good fundamental treatment of the theory of econometric model selection, we refer to the unpublished Ph.D. thesis by Sawyer (1980). In time series analysis Rudra (1954) presents a snapshot picture of the state of the art up to 1954. Akaike (1974), van den Boom & van den Enden (1974) and Unbehauen & Göhring (1974) give more up-to-date, though rather limited, reviews on the literature of univariate time series model specification.

One way to classify the different methods is to assign to them one of the properties 'objective' or 'subjective'. For a subjective method there is always a decision to be made by man. This may be the choice of some significance level or the examination of graphs and tables to look for a characteristic pattern in the behaviour of a particular statistic. An objective method may be characterized by the fact that there is no necessary involvement of a human element in the modelling process.

This classification should not be taken too seriously since the existence of objectively defined methods does not eliminate the possibility of the use of judgement or critical use of the methods. However, it provides a convenient way to further group order determination methods in these two classes according to a relevant feature of the approach to the problem. In particular the class of subjective methods can be broadly divided into the following two subclasses: those methods which make use of the theory of statistical hypothesis testing and those which are based on deterministic or stochastic realization theory. In §§ 2 and 3 we describe the main features of some of the most relevant order-determination procedures in these subclasses.

Within the class of objective methods we can group order determination methods as

those based on (i) the one-step-ahead prediction error, (ii) information measures and (iii) Bayesian methods. These are discussed in §§ 4–6 and their relationships are considered in § 7. Section 8 contains some final remarks.

2 Methods based upon the theory of statistical hypothesis testing

One of the most frequently occurring problems in statistical inference is to decide, on the basis of a finite number of observations, whether a set of parameters β satisfies s independent restrictions $h_i(\beta) = 0\ (i = 1, \ldots, s)$, where $h_i(\beta)$ denotes some linear or nonlinear function of β for which the derivatives exist. This general problem can be formulated as one of hypothesis testing where the null hypothesis $h_i(\beta) = 0\ (i = 1, \ldots, s)$ is tested versus the alternative β is unrestricted. Several suitable statistics have evolved to test these hypotheses: in particular, the likelihood ratio (LR), Wald (W) and Lagrange multiplier (LM) test statistics.

One case of particular interest of the above testing problem is when the null hypothesis is that s of the parameters, involved in the specification of the model, take specified values. All the tests proposed in time series model building apply to this situation. Suppose that, for the sake of simplicity and without loss of generality, all these restricted parameters are equal to zero. Thus, if the parameter vector β is partitioned as $(\beta_1', \beta_2')'$, the null hypothesis H_0 may be taken as $\beta_2 = (0, \ldots, 0)'$. Then under H_0, each of the test statistics LR, W and LM has an asymptotically χ^2 distribution with s degrees of freedom.

It is clear from the above discussion that H_0 is nested within the alternative hypothesis by the restriction $\beta_2 = (0, \ldots, 0)'$. The hypotheses implicit in such a nest form a uniquely ordered set and this enables a sequential testing procedure to be carried out. T.W. Anderson (1963; 1971, pp. 34–43, 116–134, 270–276) presented a sequential testing procedure for determining the order of a Gaussian distributed AR process. A number L is specified such that the true order is known to be less than or equal to L and the following, mutually exclusive hypotheses are tested sequentially

$$\begin{aligned} &H_1\colon \phi_L = 0 \\ &H_2\colon \phi_L = \phi_{L-1} = 0 \\ &\quad\vdots \\ &H_L\colon \phi_L = \phi_{L-1} = \ldots = \phi_1 = 0. \end{aligned} \tag{2.1}$$

If any of these hypotheses is true, the preceding hypotheses must be true, and if any of these hypotheses is false the succeeding ones are false. In the latter case the testing procedure terminates.

It is worth noting that the probability of making a type I error (reject a given model structure when it is the true one) in such a sequential testing procedure differs from the significance level assigned to each hypothesis.

The tests employed in the sequential procedure (2.1) may be based on the likelihood ratio, Wald or Lagrange multiplier principles. As for the LR test, Whittle (1954) shows that for testing the two nested hypotheses H_0: the assumed model is an sth-order autoregression $(s = 0, 1, \ldots, L-1)$, against H_1: the given time series realization is generated by an AR(L) process, the LR test statistic becomes approximately

$$\text{LR} = c(n)\log(\hat{\sigma}_a^2/\hat{\sigma}_L^2), \tag{2.2}$$

where $\hat{\sigma}_a^2$ and $\hat{\sigma}_L^2$ are the maximum likelihood estimates of σ_a^2 under H_0 and H_1, respectively. The quantity $c(n)$ in (2.2) is a function of the sample size; one usually takes $c(n) = -(n-L)$; see Whittle (1952) for an alternative representation. A poor fit of the two

assumed models to the data corresponds to a large negative value of (2.2) and therefore rejection of H_0. Again the test statistic (2.2) follows an asymptotically χ^2 distribution with $L-s$ degrees of freedom under H_0, for Gaussian distributed data.

Anderson (1971, p. 216) suggested a statistic based on testing the null hypothesis $\phi_{s+1} = \ldots = \phi_L = 0$ $(s = 0, 1, \ldots, L-1)$ in an AR(L) model. Let R_L be an $L \times L$ matrix partitioned into s and $L-s$ rows and columns and the parameter vector $\hat{\phi}$ correspondingly

$$R_L = \left[\begin{array}{c:c} R_1 & R_{12} \\ \hdashline R_{12}' & R_2 \end{array}\right], \quad \hat{\phi} = \begin{bmatrix} \hat{\phi}^{(1)} \\ \hdashline \hat{\phi}^{(2)} \end{bmatrix}$$

where the (i, j)th element of R_L is the sample autocorrelation function r_{i-j}, defined by

$$r_k = \sum_{t=1}^{n-k} y_t y_{t+k} \Big/ \sum_{t=1}^{n} y_t^2 \quad (k = 1, \ldots, n-1),$$

and where $\hat{\phi}$ is the vector of estimates of the parameters in the AR(L) model. Then the statistic

$$n\hat{\phi}^{(2)\prime}[R_2 - R_{12}'R_1^{-1}R_{12}]\hat{\phi}^{(2)}/\hat{\sigma}_L^2 \tag{2.3}$$

is the Wald test statistic. The Lagrange multiplier test statistic for this case can be obtained in a similar way; see Hosking (1980) for several alternative representations of the LM test for general ARMA models.

Replacing the parameters ϕ_i in (2.1) by θ_i, the sequential testing procedure may also be used for testing hypotheses in pure MA models. The selection of the order can be based on either of the three previously discussed test statistics. Unfortunately, the sequential testing procedure has two serious drawbacks. The first is that it may take a relatively large order of the pure AR or pure MA model to approximate, in an appropriate way, a particular time series. This could lead to an overparametrized model and thus violate the principle of parsimony. The second drawback is the difficulty in choosing the significance level. Although the sequential testing procedure is designed to satisfy certain clearly defined optimality conditions, the choice of this significance level cannot be made, as it should be, on the basis of a suitable trade-off between probabilities of type I and type II errors (accept a false model structure) since the latter are unknown for the tests considered here.

The sequential testing procedure (2.1) starts from the most general, or least restrictive, hypothesis and tests successively more restricted hypotheses. Anderson (1971) shows that such a procedure is uniformly most powerful in the class of procedures that fix the probabilities of accepting a less restricted hypothesis than the true one. Clearly, an alternative way of sequentially testing hypotheses is to start off from the most restricted hypothesis and apply the procedure (2.1) in the reverse order. It is obvious that this approach has computational advantages, especially if the more general models are allowed to include highly nonlinear structures. However, it should be noted that, in general, the test statistics will not be independent and hence this approach may have a low power.

When the data are supposed to be generated by an ARMA process there may not be any obvious way in which the ARMA model is likely to be deficient nor any obvious alternative hypothesis to consider. Consequently, no unique ordering of hypotheses is possible and a sequential testing procedure is no longer applicable under such circumstances. One can overcome this problem by approximating an ARMA model by an AR or MA model of sufficiently high order, so that the true model can be assumed to be nested within this approximate model, or take $p = q$. One can also accept the lack of unique ordering in the hypotheses and proceed by applying sequential testing procedures within each ordered nest. In the latter case, it is necessary to apply one of the available tests for discrimination

between nonnested hypotheses, see MacKinnon (1983) for a review, in order to select the final model choice from the set consisting of the preferred models from each ordered nest.

Many applications of the above mentioned tests can be found in the literature. For instance, Zellner & Palm (1974) use the LR test to appraise alternative hypotheses in a small dynamic simultaneous equation model of the U.S. economy. Hsiao (1979a) uses the LR test to investigate the appropriateness of a bivariate AR model of Canadian money and income data. The small-sample properties of two Student t-tests are compared with the LR and W tests for an MA(1) model by Nelson & Shea (1979). Newbold (1983) uses the LM test to build an ARMA model for the money supply and gross national product of the U.S. Most other applications of these three tests have been confined to linear regression model specifications; see e.g. Breusch (1978) and Breusch & Pagan (1980). Finally, we shall mention that a test for nonnested ARMA models is discussed by Walker (1967).

3 Methods based on deterministic or stochastic realization theory

Many of the procedures for order selection require the estimation of the model parameters at each stage, the calculation of some measure of goodness of fit, and a comparison with previous steps over all possible choices of the order (p, q). Needless to say that such an approach may be quite involved for higher order processes and computationally rather expensive. We now consider some procedures which do not require prior model fitting. We will do this for each method separately. Recently, Piccolo & Tunnicliffe-Wilson (1984) have shown that several of the methods reviewed here have a common basis which allows for a unified approach to ARMA model identification. For more details about this approach we refer to their paper.

3.1 Inverse autocorrelations and partial autocorrelations

Chatfield (1979), see also Cleveland (1972), puts forward arguments in favour of using inverse autocorrelations (iac) instead of the autocorrelations in the Box–Jenkins procedure. The kth lag inverse autocovariance, denoted by $\gamma i(k)$, can be defined as the coefficient of B^k in the generating function $\Gamma i(B) = \sum \gamma i(k) B^k$, where the sum is over $k = -\infty, \ldots, \infty$, which satisfies $\Gamma i(B)\Gamma(B) = 1$, where $\Gamma(B)$ is the autocovariance generating function. The function $\Gamma i(B)$ is usually referred to as the inverse autocovariance generating function and $\Gamma i(B)/\gamma i(0)$ as the inverse autocorrelation generating function. It can be shown that for the ARMA(p, q) process (1.1),

$$\Gamma i(B) = \frac{\phi_p(B)\phi_p(B^{-1})}{\theta_q(B)\theta_q(B^{-1})} \frac{1}{\sigma_a^2}. \tag{3.1}$$

This implies that the kth lag inverse autocorrelation for an AR(p) process is given by

$$\rho i(k) = \begin{cases} \left(-\phi_k + \sum_{j=1}^{p-k} \phi_j \phi_{j+k}\right) \Big/ \left(1 + \sum_{j=1}^{p} \phi_j^2\right) & (k \leqslant p), \\ 0 & (k > p). \end{cases} \tag{3.2}$$

Thus in the autoregressive case $\rho i(k)$ has a cut-off after lag p. This property makes it a competitor for the partial autocorrelations. However, the estimation of $\rho i(k)$ is not easy. One method of estimating the iac of a series is to approximate the series by an AR process of 'sufficiently' high order, to estimate its parameters using the Yule–Walker equations, see Box & Jenkins (1976, Ch. 3), and then use the above equations (3.2) to estimate the

inverse autocorrelations. Bhansali (1980) has derived the asymptotic distributional properties of the sample inverse autocorrelations. In particular, he has shown that for a white noise process the sample inverse autocorrelations are asymptotically normally distributed with mean zero and variance $1/n$. Hence, significance of the iac at the 5% level can be checked by comparing them with the values $1{\cdot}96/n^{\frac{1}{2}}$.

An alternative function known as the inverse partial autocorrelation function (ipacf) has also been discussed in the literature, see Chatfield (1979), as a specification tool. This may be defined as the partial autocorrelation function for the inverse model $\theta_q(B)y_t = \phi_p(B)a_t$. These can be computed from the Yule–Walker equations by replacing the autocorrelations with the inverse autocorrelations. The behaviour of ipacf is similar to the autocorrelation function. However, its usefulness is limited because of estimation difficulties.

For a more complete discussion on this subject we refer to Bhansali (1980, 1983a), Hipel, McLeod & Lennox (1977) and Abraham & Ledolter (1984). McLeod, Hipel & Lennox (1977) give examples of the use of the inverse autocorrelation function in time series model identification.

3.2 *R- and S-arrays*

Gray, Kelley & McIntire (1978) have proposed another method of using the autocorrelations to identify the order (p, q) of an ARMA process. Their method primarily exploits the presence of certain patterns in two arrays of numbers, called the R- and S-arrays. The (i, j)th elements of these arrays are given by

$$R_j(\rho_i) = H_j(\rho_i)/H_j(1;\rho_i), \quad S_j(\rho_i) = H_{j+1}(1;\rho_i)/H_j(\rho_i) \tag{3.3}$$

respectively, where the jth and $(j+1)$th order determinants $H_j(\rho_i)$ and $H_j(1;\rho_i)$ are defined as

$$H_j(\rho_i) = \begin{vmatrix} \rho_i & \rho_{i+1} & \cdots & \rho_{i+j-1} \\ \rho_{i+1} & \rho_{i+2} & \cdots & \rho_{i+j} \\ \vdots & \vdots & & \vdots \\ \rho_{i+j-1} & \rho_{i+j} & \cdots & \rho_{i+2j-2} \end{vmatrix}, \quad H_{j+1}(1;\rho_i) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \rho_i & \rho_{i+1} & \cdots & \rho_{i+j} \\ \vdots & \vdots & & \vdots \\ \rho_{i+j-1} & \rho_{i+j} & \cdots & \rho_{i+2j-1} \end{vmatrix},$$

where $\rho_i \equiv \rho_{-i}$ is the autocorrelation at lag i, defined by $\mathrm{Cov}(y_t, y_{t-i})/\mathrm{Var}(y_t)$ and where $H_0(\rho_i) = 1$. Alternatively, one can replace ρ_i with $(-1)^i\rho_i$ in all the above formulae. Simple recursive relations to calculate the array values are given by Gray et al. (1978).

Table 1

Theoretical pattern of the R-array for an ARMA(p, q) *process*

i	$j=1$	$j=2$	...	$j=p$	$j=p+1$	...	$j=p+k$
$-l$	ρ_{-l}	$R_2(.)$	...	$R_p(.)$	0	...	0
⋮	⋮	⋮		⋮	⋮		⋮
$-q-p-1$	ρ_{-q-p-1}	$R_2(.)$	...	$R_p(.)$	0	...	0
$-q-p$	ρ_{-q-p}	$R_2(.)$	...	$R_p(.)$	nz	...	nz
⋮	⋮	⋮		⋮	⋮		⋮
$q-p$	ρ_{q-p}	$R_2(.)$	...	$R_p(.)$	nz	...	nz
$q-p+1$	ρ_{q-p+1}	$R_2(.)$	...	$R_p(.)$	0	...	0
⋮	⋮	⋮		⋮	⋮		⋮
g	ρ_g	$R_2(.)$	...	$R_p(.)$	0	...	0

nz, nonzero.

Table 2

Theoretical pattern of the S-array for an ARMA(p, q) *process*

i	$j=1$	...	$j=p$	$j=p+1$	...	$j=p+k$
$-l$	$S_1(.)$	...	C_2	u	...	u
.	.		.	.		.
.	.		.	.		.
.	.		.	.		.
$-q-p-1$	$S_1(.)$	...	C_2	u	...	u
$-q-p$	$S_1(.)$	...	C_2	$\pm\infty$	...	$\pm\infty$
$-q-p+1$	$S_1(.)$	...	nc	nc	...	nc
.	.		.	.		.
.	.		.	.		.
.	.		.	.		.
$q-p-1$	$S_1(.)$	...	nc	nc	...	nc
$q-p$	$S_1(.)$	...	C_1	$-C_1$	...	$(-1)^k C_1$
$q-p+1$	$S_1(.)$	...	C_1	u	...	u
.	.		.	.		.
.	.		.	.		.
.	.		.	.		.
g	$S_1(.)$	...	C_1	u	...	u

u, undefined; nc, nonconstant.

$$C_1=(-1)^p\left(1-\sum_{i=1}^{p}\phi_i\right), \quad C_2=-C_1/\phi_p.$$

If the observations are from a stationary ARMA(p, q) process, and the true autocorrelations are used to calculate the array elements, then the R-array will exhibit the behaviour shown in Table 1 and the S-array that shown in Table 2. In particular the pattern of the $(p+1)$ st column in the R-array and that of columns p and $(p+1)$ in the S-array are to be noted. When $p=0$ the S-array will have no fixed pattern and the Gray et al. (1978) method of order determination amounts to examining the first column of the R-array.

When the autocorrelations ρ_k are replaced by the mean corrected sample autocorrelations $r_k^{(n)}$, that is

$$r_k^{(n)}=\sum_{t=1}^{n-k}(y_t-\bar{y})(y_{t+k}-\bar{y})\Big/\sum_{t=1}^{n}(y_t-\bar{y})^2$$

is used instead of the true autocorrelations, the behaviour of the arrays will be an approximation of that described above; furthermore, the $(p+1)$st column of the S-array will consist of highly variable terms. One looks for the above behaviour in the arrays shown in Tables 1 and 2 in order to determine the values p and q. Gray et al. (1978) and Woodward & Gray (1978) suggest that using $r_k^{(n)}$ to calculate the arrays gives easier identification of high frequency data (in the spectral sense of the word) while $(-1)^k r_k^{(n)}$ works better with low frequency data, although this is not always the case. In addition these authors suggest that it is usually easier to identify p from the S-array and q from the R-array.

Clearly the Gray–Kelley–McIntire procedure to model selection is one of pattern recognition and in this respect, it is similar to the Box–Jenkins graphical approach to ARMA model identification. An obvious drawback of the R- and S-arrays approach is that the user may be confused by the large volume of numbers. To overcome this problem Gray et al. (1978) suggested a statistic which, using various properties of the R- and S-arrays, has been designed for testing the significance of the identical values of p and q. We do not discuss this statistic here since its sampling distribution has not yet been developed, and the patterns in the R- and S-arrays have to be quite distinct to give a significant value of the statistic. In the latter case, the statistic loses most of its usefulness.

Further comments to the Gray, Kelley & McIntire paper are given by Tukey et al. (1978). These reviewers make it clear that the R- and S-array procedure is quite complex

and its statistical properties, such as consistency of the selected order, are theoretically difficult to verify. Furthermore, for real data series the various arrays may not give such a clear-cut identification as the simulated time series presented by Gray et al. (1978) do.

3.3 *The corner method*

Yet another method which uses the autocorrelations to generate an array for identifying the order of a time series process is the so-called corner method of Beguin, Gourieroux & Monfort (1980). These authors base their order determination method on the $j \times j$ determinant $\Delta(i, j)$ which is defined as

$$\Delta(i,j) = \begin{vmatrix} \rho_i & \rho_{i-1} & \cdots & \rho_{i-j+1} \\ \rho_{i+1} & \rho_i & \cdots & \rho_{i-j+2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \rho_{i+j-1} & \rho_{i+j-2} & \cdots & \rho_i \end{vmatrix} \quad (i, j = 1, 2, \ldots, L). \tag{3.4}$$

Beguin et al. (1980) prove that a stationary process $\{y_t\}$ has a minimal ARMA(p, q) representation if and only if $\Delta(i, j) = 0$ for all values $i \geq q+1$ and $j \geq p+1$; while $\Delta(i, p) \neq 0$ for all values $i \geq q$ and $\Delta(q, j) \neq 0$ for all values $j \geq p$. Table 3 presents the tabulation of $\Delta(i, j)$ versus i and j in a so-called Δ-array. For stationary ARMA(p, q) processes, identifying p and q with this corner method is equivalent to identifying a corner of zero values in the Δ-array.

It is easy to see that the corner method is directly related to the Gray et al. (1978) order determination method, since $\Delta(i, j) = (-1)^{[j/2]} H_j(\rho_{i-j+1})$ where $[x]$ denotes the integer part of x. In practice, the autocorrelations ρ_k in $\Delta(i, j)$ are replaced by the mean corrected sample autocorrelations $r_k^{(n)}$, resulting in an array which will be somewhat different from the pattern in Table 3. Beguin et al. (1980) claim that visual inspection of this sample $\hat{\Delta}$-array is generally sufficient to select the appropriate order of a model. However, de Gooijer & Heuts (1981) seem to refute this claim. Also these authors point out that the application of a test suggested by Beguin et al. (1980) to detect whether a 'corner' is equal to zero, has severe limitations; see also Petruccelli & Davies (1984). An extension of the corner method to the determination of the order of nonstationary ARMA processes has been considered by Hamdi (1982). Liu & Hanssens (1982) have generalized the method to the identification of multiple-input transfer function models.

Table 3

Theoretical pattern of the Δ-array for an ARMA(p, q) *process; i, moving average order; j, autoregressive order*

i	$j=1$	$j=2$	...	$j=p$	$j=p+1$	...	$j=L$
1	$\Delta(1,1)$	$\Delta(1,2)$	...	$\Delta(1,p)$	$\Delta(1,p+1)$	...	$\Delta(1,L)$
.	.	.		.	.		.
.	.	.		.	.		.
.	.	.		.	.		.
q	$\Delta(q,1)$	$\Delta(q,2)$	...	$\Delta(q,p)$	$\Delta(q,p+1)$	...	$\Delta(q,L)$
$q+1$	$\Delta(q+1,1)$	$\Delta(q+1,2)$	...	$\Delta(q+1,p)$	0	...	0
.	.	.		.	.		.
.	.	.		.	.		.
.	.	.		.	.		.
L	$\Delta(L,1)$	$\Delta(L,2)$	...	$\Delta(L,p)$	0	...	0

Table 4

Theoretical pattern of the generalized partial autocorrelation array for an ARMA(p, q) *process; i, moving average order; j, autoregressive order*

i	$j=1$	...	$j=p-1$	$j=p$	$j=p+1$	...	$j=L$
0	$\phi_{11}^{(0)}$	...	$\phi_{p-1,p-1}^{(0)}$	$\phi_{pp}^{(0)}$	$\phi_{p+1,p+1}^{(0)}$	...	$\phi_{LL}^{(0)}$
⋮	⋮		⋮	⋮	⋮		⋮
$q-1$	$\phi_{11}^{(q-1)}$	...	$\phi_{p-1,p-1}^{(q-1)}$	$\phi_{pp}^{(q-1)}$	$\phi_{p+1,p+1}^{(q-1)}$	...	$\phi_{LL}^{(q-1)}$
q	$\phi_{11}^{(q)}$	...	$\phi_{p-1,p-1}^{(q)}$	ϕ_p	0	...	0
$q+1$	$\phi_{11}^{(q+1)}$	...	$\phi_{p-1,p-1}^{(q+1)}$	ϕ_p	u	...	u
⋮	⋮		⋮	⋮	⋮		⋮
L	$\phi_{11}^{(L)}$	...	$\phi_{p-1,p-1}^{(L)}$	ϕ_p	u	...	u

u, undefined.

3.4 *Generalized autocorrelations and partial autocorrelations*

If $\{y_t\}$ is an ARMA(p, q) process the theoretical autocorrelation function of such a process does not satisfy the first q Yule–Walker equations. However, it is well known that for $j > q$ the autocorrelations of such a process do satisfy the set of equations $\sum \phi_i \rho_{j-i} = 0$, where the sum is over $i = 0, \ldots, p$. Using this observation, Woodward & Gray (1981) define the generalized theoretical partial autocorrelation coefficient as

$$\phi_{jj}^{(i)} = \begin{cases} \rho_{i+1}/\rho_i & \text{if } j = 1, \\ \Delta^*(i, j)/\Delta(i, j) & \text{if } j > 1, \end{cases} \tag{3.5}$$

where $\Delta^*(i, j)$ is the determinant composed of the first $i-1$ columns of $\Delta(i, j)$ with the ith column given by the vector $(\rho_{i+1}, \ldots, \rho_{i+j})'$. Jenkins & Alavi (1981) refer to these as q-conditioned partial autocorrelations. It is interesting to note that $\phi_{jj}^{(0)}$ $(j = 1, 2, \ldots)$ are the ordinary partial autocorrelations.

Woodward & Gray (1981) show that the properties of $\phi_{jj}^{(i)}$ make it possible to uniquely identify p and q of a mixed process if the theoretical autocorrelations are known. They also establish the following relation between the S-arrays and the generalized theoretical partial autocorrelations:

$$\phi_{jj}^{(i)} = \begin{cases} -S_j(f_{-j+i-1})/S_j(f_{-j-i}) & \text{if } f_i = \rho_i, \\ (-1)^{j+1} S_j(f_{-j+i+1})/S_j(f_{-j-i}) & \text{if } f_i = (-1)^i \rho_i. \end{cases} \tag{3.6}$$

For an ARMA(p, q) process the $\phi_{jj}^{(i)}$ can be arranged as in Table 4. Thus a natural procedure for model identification is to find a column p of the generalized partial autocorrelation array with relatively constant entries and a row q in which the elements are zero for columns j $(j > p)$. Woodward & Gray (1981) illustrate this method, together with an application of the S-array procedure, by means of several examples.

In practice the theoretical autocorrelations ρ_k have to be replaced by the sample values $r_k^{(n)}$ and simulation results presented by Newbold & Bos (1983) indicate that the corresponding sample statistic is a poor estimator of (3.5). Further, it is shown by Davies & Petruccelli (1984) that the generalized sample partial autocorrelation function has an unstable behaviour when applied to time series of moderate length. This makes it less attractive for model specification; for additional discussion see also Glasbey (1982).

A generalized autocorrelation function useful for determining the order of an ARMA process has been recently proposed by Takemura (1984). Based on ideas given by Bartlett

& Diananda (1950) he defines this function, for $i \geq 0$ and $j \geq 0$, as

$$\rho(i+1, j+1) = \begin{cases} [\gamma_{i+j+1} - \gamma(j+1, i)'\Gamma(j, i)^{-1}\gamma(i+j, -i)] \\ \quad \times [\gamma_0 - 2\gamma(j, i)'\Gamma(j, i)'^{-1}\gamma(j+1, i) \\ \quad + \gamma(j+1, i)'\Gamma(j, i)^{-1}\Gamma(0, i)\Gamma(j, i)'^{-1}\gamma(j+1, i)]^{-1} & \text{if } |\Gamma(j, i) \neq 0, \\ 0 & \text{if } |\Gamma(j, i)| = 0, \end{cases} \tag{3.7}$$

where, for $i > 0$, $\gamma(j, i) = (\gamma_j, \gamma_{j+1}, \ldots, \gamma_{j+i-1})'$ and $\gamma(j, -i) = (\gamma_j, \gamma_{j-1}, \ldots, \gamma_{j-i+1})'$, with $\gamma_i = \text{Cov}(y_t, y_{t-i})$ the theoretical autocovariance at lag i, and where the $i \times i$ Toeplitz matrix $\Gamma(j, i)$ is given by

$$\Gamma(j, i) = \begin{bmatrix} \gamma_j & \gamma_{j+1} & \cdots & \gamma_{j+i-1} \\ \gamma_{j-1} & \gamma_j & \cdots & \gamma_{j+i-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{j-i+1} & \gamma_{j-i+2} & \cdots & \gamma_j \end{bmatrix}.$$

Takemura proves that $\rho(i, j)$ $(i > 0, j > 0)$ has the following properties: $\rho(1, j) = \rho_j$ for $j \geq 1$; $\rho(i, 1) = \phi_{ii}^{(0)}$ for $i \geq 1$; $-1 \leq \rho(i, j) \leq 1$; and, for an ARMA$(p, q)$ process, $\rho(i, j) = 0$ for $i > p$ and $j > q$. Clearly, this latter property can be used to identify the orders p and q of the process. Also he shows that the function (3.7) can be quickly calculated by recursive formulae which are a straightforward generalization of the well-known Durbin–Levinson recursive procedure for computing partial autocorrelations. In practical applications the theoretical autocovariances γ_i are, of course, replaced by the sample autocovariances. If $r(i+1, j+1)$ denotes the resulting generalized sample autocorrelation coefficient, then Takemura (1984) proves that for an ARMA(p, q) process with $\phi_p \neq 0$ and $\theta_q \neq 0$, the statistic

$$n^{\frac{1}{2}} r(p+1, q+1) \Big/ \left[1 + 2 \sum_{i=1}^{q} (r_i^x)^2\right]^{\frac{1}{2}}$$

has an asymptotically standard normal distribution. Here r_i^x denotes a consistent estimate of $\gamma_i^x = \text{Cov}(x_t, x_{t-i})$, where $x_t = \phi_p(B)y_t = \theta_q(B)a_t$. It is clear that this test statistic can be conveniently used to test for each pair (p, q) the null hypothesis $\rho(p+1, q+1) = 0$. Simulation results given by Takemura seem to confirm this conjecture.

3.5 *Iterated regression and the extended sample autocorrelation function* (ESACF)

Tiao & Tsay (1983a, b) and Tsay & Tiao (1984) used consistent estimates of the AR parameters of either stationary or nonstationary ARMA(p, q) processes to define the 'extended sample autocorrelation function' (ESACF) from which the values of p and q can be determined. If $\{y_t\}$ follows an ARMA(p, q) process then it can be shown that, in probability, $\hat{\phi}_{p(l)}^{(j)} \to \phi_l$ for $l = 1, \ldots, p$ if $j \geq q$ (regardless of whether or not the process is stationary), where $\hat{\phi}_{p(l)}^{(j)}$ is the AR parameter estimate at lag l of the jth iterated AR(p) regression. For ease of calculation, estimates of these parameters are obtained recursively for $j \geq 1$ using

$$\hat{\phi}_{p(l)}^{(j)} = \hat{\phi}_{p+1(l)}^{(j-1)} - \hat{\phi}_{p(l-1)}^{(j-1)} \hat{\phi}_{p+1(p+1)}^{(j-1)} \Big/ \hat{\phi}_{p(p)}^{(j-1)} \quad (l = 1, 2, \ldots, p), \tag{3.8}$$

where $\hat{\phi}_{p(0)}^{(j-1)} = -1$.

To obtain the ESACF, it is necessary to calculate the ordinary least squares estimates $\hat{\phi}_{p(l)}^{(0)}$, for $p = 1, 2, \ldots, p_0 + q_0$ and $l = 1, 2, \ldots, p$, by successive AR(1) through AR$(p_0 + q_0)$ fittings where p_0 and q_0 are predetermined values. Using the recursive formulae, $\hat{\phi}_{p(l)}^{(j)}$, for $p = 1, 2, \ldots, p_0$, $l = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q_0$, are then calculated. For any finite

Table 5

The extended sample autocorrelation table

Autoregressive order	Moving average order 0	1	2	3	...
0	$r_{1(0)}$	$r_{2(0)}$	$r_{3(0)}$	$r_{4(0)}$	...
1	$r_{1(1)}$	$r_{2(1)}$	$r_{3(1)}$	$r_{4(1)}$	
2	$r_{1(2)}$	$r_{2(2)}$	$r_{3(2)}$	$r_{4(2)}$	
.	.				
.	.				
.	.				

value m, the mth ESACF is defined as

$$r_{k(m)} = \text{lag } k \text{ sample autocorrelation of } \left(y_t - \sum_{l=1}^{m} \hat{\phi}^{(k)}_{m(l)} y_{t-l}\right). \tag{3.9}$$

Since $\hat{\phi}_{m(l)}$ is consistent for ϕ_l $(l = 1, 2, \ldots, m)$, the sample autocorrelations of $y_t - \sum \hat{\phi}^{(k)}_{m(l)} y_{t-l}$, where the sum is over $l = 1, \ldots, m$, are in large samples given by $r_{k(m)} = 0$ for $k > q$ and $p = m$.

Using the iterated regression estimates, the jth ESACF at lag k, $r_{k(j)}$, for $j = 1, 2, \ldots, p_0$ and $k = 1, 2, \ldots, q_0$, can be calculated from the residual processes. These values are arranged in a two-way table as shown in Table 5 in which the first row gives the ordinary sample autocorrelation function of the series y_t. Tsay & Tiao (1984) show that, for an ARMA(p, q) process, the ESACF has the following asymptotic property:

$$r_{k(j)} = \begin{cases} c(j-p, k-q) & (0 \leqslant k-q \leqslant j-p), \\ 0 & (k-q > j-p \geqslant 0), \end{cases} \tag{3.10}$$

where $c(j-p, k-q)$ is some nonzero constant or a continuous random variable bounded between -1 and 1.

The large-sample property (3.10) can be utilized to tentatively identify the order of an ARMA(p, q) process. In particular, the values of p and q can be determined by searching for the vertex of a triangle of asymptotic 'zero' values in Table 5, having boundary lines $j = d_1 \geqslant 0$ and $k - j = d_2 \geqslant 0$. The order of the process is then deduced from the row and column coordinates of the vertex. For finite values of n, the $r_{k(j)}$'s will not all be equal to

Table 6

The asymptotic extended sample autocorrelation table for an ARMA(p, q) *process*

Autoregressive order	Moving average order 0	1	...	q	$q+1$	$q+2$	$q+3$	...	L
0	c	X	...	X	X	X	X	...	X
1	c	X	...	X	X	X	X	...	X
.	.	.		.	.	.	.		.
.	.	.		.	.	.	.		.
.	.	.		.	.	.	.		.
$p-1$	c	X	...	X	X	X	X	...	X
p	c	X	...	X	0	0	0	...	0
$p+1$	c	X	...	X	X	0	0	...	0
$p+2$	c	X	...	X	X	X	0	...	0
.	.	.		.	.	.	.		.
.	.	.		.	.	.	.		.
.	.	.		.	.	.	.		.
L	c	X	...	X	X	X	X	...	0

X, a value $>(2/(n-j-k))$ or $<(-2/(n-j-k))$; 0, a value between $\pm 2(n-j-k)^{-1}$; c, a value between -1 and 1.

zero. As an additional tool to detect whether they are indeed equal to zero, Tsay & Tiao (1984) suggest to use Bartlett's (1946) formula $(n-j-k)^{-1}$ as an approximation to the asymptotic variance of the $r_{k(j)}$'s. The use of this formula is based on the hypothesis that the transformed series $\phi_k^{(j)}(B)y_t$ is a white noise process. As a simple informative guide to detect the patterns in Table 5, Tsay & Tiao (1984) advocate the determination of the order (p, q) by searching for the required 'cutting off' behaviour as shown in Table 6.

This method of order determination provides information on the 'maximum orders' of p and q since differencing of the data from nonstationary processes is not required. The identification technique is claimed to have performed quite well in practice, but a great deal of computation may be necessary to estimate the iterated regression coefficients and to calculate the residual processes and their sample autocorrelations. Tsay & Tiao (1983) have applied ESACF to the modelling of seasonal time series and Tiao & Tsay (1983b) use it to build multiple time series models.

4 Methods based on the one-step-ahead prediction error

The difficulties in determining the order of an ARMA model by using the Neyman–Pearson approach have introduced the important notion that one should not expect a finite number of observations on a time series process to give a clear-cut answer about the true order of that process. After all the existence of such an order is only a conceptual convenience. If one's primary goal is to select the best approximating model then it is much more reasonable to do this with respect to a particular purpose one has in mind for the model. The prediction of future values of a series is such a typical purpose. A number of objective order-determination procedures have been proposed to assess models on the basis of their predictive power, both within and outside the sample.

4.1 *Final prediction error* (FPE) *criterion*

Akaike (1969, 1970a) proposed a method which chooses the order p of the autoregression so that the expected one-step-ahead squared prediction error is minimized, considering the errors due to the inaccuracies when a low order is selected, and the error due to the increase of residual variance (caused by a decrease of the number of degrees of freedom) when a higher order is selected.

Let a stationary time series be generated by an AR(p) process, where p is finite and bounded by some integer L. If we assume that the parameters of this process are known, the minimum mean squared error forecast at time $t=n$, one step ahead, is

$$\hat{y}_n(1) = \phi_1 y_n + \ldots + \phi_p y_{n-p+1} \tag{4.1}$$

and the one-step-ahead prediction error $y_{n+1} - \hat{y}_n(1)$, has mean squared error given by $E[(y_{n+1} - \hat{y}_n(1))^2] = \sigma_a^2$; see, for example, Box & Jenkins (1976, § 5.1).

When the parameters ϕ_i are replaced by their least squares estimates and when n is large the expected mean squared error of the estimated one-step-ahead predictor becomes approximately

$$E[(y_{n+1} - \hat{y}_n(1))^2] = \sigma_a^2(1 + p/n); \tag{4.2}$$

see Yamamoto (1976). If we replace σ_a^2 by its 'unbiased' estimate $n\hat{\sigma}_a^2/(n-p)$, Akaike's (1970a) so-called final prediction error (FPE) statistic, is defined as

$$\text{FPE}(p) = \hat{\sigma}_a^2(n+p)/(n-p) \tag{4.3}$$

and the order is chosen for which FPE(p) ($p = 0, 1, \ldots, L$) attains its minimum value. In this method no subjective element is left in the definition of FPE(p), except for the determination of the upper limit L and the choice of the one-step-ahead prediction error

criterion. In practice, one usually fixes L at a rather large value ($<n$) to ensure that it exceeds the true order when the process is purely AR.

Generally, if the order of fit is increased, $\hat{\sigma}_a^2$ will tend to decrease to some level near the theoretical value σ_a^2, whereas the extra factor on the right-hand side of (4.3) penalizes an increase of p by 'pulling up' the FPE value. By seeking the minimum of FPE we select an order p which balances the risk between both terms optimally. It should be noted, however, that when $n \gg p$, $(n+p)/(n-p)$ is little affected by changes in p and it is, in general, difficult to choose the correct model order with a high degree of consistency (Gersch & Sharpe, 1973).

The properties of FPE for selecting the order of AR models were investigated by simulations, Bhansali (1973) concluded that FPE tends to overfit the true order of the generating AR process for series of length 50. Jones (1975) reanalysed Bhansali's results and pointed out that Bhansali had used an incorrect definition of the FPE criterion; see also Akaike (1974) for some comments on Bhansali's paper. In spite of this it was proved later, by Shibata (1976), that Bhansali's conclusion with respect to the properties of the FPE criterion is justified. We will discuss this in § 5 when we consider a refinement of the FPE statistic suggested by Akaike (1974).

Akaike (1971) extended the FPE method to the determination of the order of multivariate AR processes. Söderström (1977) showed that (4.3) can be used independently of the model structure. Thus, on replacing p in (4.3) by the number of estimable parameters it can be applied to all standard time series models including mixed ARMA models. An interesting generalization of the FPE method was obtained by Bhansali & Downham (1977). These authors have suggested that a choice of order could be based on the minimum of the function

$$\text{FPE}_\delta(p) = \hat{\sigma}_a^2(1+\delta p/n) \quad (p = 0, 1, \ldots, L), \tag{4.4}$$

where δ is a positive constant.

It is easy to see that a choice of $\delta = 2$ in (4.4) corresponds to using the FPE criterion. By choosing $\delta > 2$, one increases the penalty of overparametrization. In the simulation experiments of Bhansali & Downham (1977) values of δ in the range 1–4 were considered. Their results show that, if n is quite large ($n \geqslant 300$) the correct model is fitted more frequently for $\delta > 2$ than for $\delta = 2$, but if n is small ($n = 50$) this may not necessarily be the case. However, simulation results given by Akaike (1979) indicate that severe cases of underestimation of the true order can be expected with $\delta > 2$ for high-order models. This result is confirmed by Atkinson (1980) who claims that, on the basis of mean squared prediction error, the choice $\delta = 2$ should be adopted in (4.4).

For a further discussion of the FPE method see Akaike (1970b), Bhansali (1978) and McClave (1975, 1978). A similar criterion to the FPE_δ method has been developed by Bhansali (1983b) for determining the order of MA processes. Applications of the FPE method are, for example, reported by Caines, Keng & Sethi (1981). Hsiao (1979a, b; 1982) and Otomo, Nakagawa & Akaike (1972) use the concept of FPE to model vector AR processes. See also Lütkepohl (1985).

4.2 Cross-validatory criteria

It is apparent from the above discussion that the ability of predicting one-step-ahead values of $\{y_t\}$ is measured on the same data used to estimate the parameters of the model. To overcome this somewhat unrealistic situation several authors have advocated methods based on cross-validatory criteria. Though these methods were originally obtained in the context of linear regression models, they carry over to the selection of purely AR models.

The cross-validatory criteria are constructed as follows. Given a particular model

specification in the class of AR(p) ($p = 0, 1, \ldots, L$) models, the model is estimated n times each time deleting one observation from the sample and 'predicting' (in the minimum mean squared error sense) onto that observation. The cross-validatory criteria are then based on the sum of the weighted squared prediction errors

$$\sum_{t=1}^{n} w_{t(i)}\{y_t - \hat{y}_{t(i)}\}^2, \tag{4.5}$$

where $w_{t(i)}$ is some prescribed weighting function and $\hat{y}_{t(i)}$ is the tth 'predicted' observation of the autoregression formed from all the observations excluding the ith observation. Minimizing (4.5) for $p = 0, 1, \ldots, L$ gives the appropriate order of the model. This approach can be said to 'simulate prediction'. In contrast to FPE it does not use an observation to aid in the 'prediction' itself.

Different weighting functions $w_{t(i)}$ produce distinct order-determination methods. Allen (1971) proposed his so-called prediction sum of squares (PRESS) criterion by giving an equal weight to each deleted observation. The PRESS criterion has been applied to 25 U.S. economic time series by Bessler & Binkley (1980) who compared the obtained AR models with those suggested by the FPE criterion. Their results suggest that FPE tends to overfit the order of the AR models relative to those found by PRESS. Hjorth & Holmqvist (1981) apply PRESS to build a multivariate AR model for meteorological data. Other weighting functions have been advocated by Schmidt (1974), Stone (1978) and Hjorth (1982) but, as far as we know, only PRESS has received some attention in the literature.

4.3 *Criterion for autoregressive transfer functions* (CAT)

Another concept for determining the order of an AR process is given by Parzen (1974, 1977) by using a spectral theory approach. Since this procedure is closely associated with the FPE method, only its main characteristics will be presented here. Parzen (1974) assumes that an observed time series is generated by an AR process of infinite order described by $\phi_\infty(B)y_t = a_t$ with $\mathrm{Var}[a_t] = \sigma_\infty^2$, and that this process can be adequately approximated by a finite AR(p) process with variance $\sigma_a^2(p)$. Based on the spectral density functions of the two time series processes, Parzen (1974, p. 727) proposes the following criterion of 'distance' between the two AR processes

$$J_p = (2\pi)^{-1}\int_{-\pi}^{\pi} |\{\phi_p(e^{i\omega}) - \phi_\infty(e^{i\omega})\}/\phi_\infty(e^{i\omega})|^2\, d\omega.$$

Parzen shows that for a fixed p the minimum of J_p is $1 - \sigma_\infty^2/\sigma_a^2(p)$ and it is reached when the AR parameters are estimated by the Yule–Walker equations. Denote these estimates by $\hat{\phi}_i$ and let $\hat{\phi}_p(z) = \sum \hat{\phi}_i z^i$, it was proved by Kromer (1969) that approximately

$$E\left[(2\pi)^{-1}\int_{-\pi}^{\pi} |\{\hat{\phi}_p(e^{i\omega}) - \phi_\infty(e^{i\omega})\}/\phi_\infty(e^{i\omega})|^2\, d\omega\right] = 1 - \frac{\sigma_\infty^2}{\sigma_a^2(p)} + \frac{p}{n}. \tag{4.6}$$

The term $(1 - \sigma_\infty^2/\sigma_a^2(p))$ in (4.6) represents the bias due to approximation of $\phi_\infty(B)$ by $\phi_p(B)$, while p/n represents the overall variance of estimating the parameters ϕ_i in $\phi_p(B)y_t = a_t$.

Originally Parzen proposed to adopt that value of p as the most suitable order of the process for which (4.6) is minimized. Later, he recognized that this criterion is cumbersome to use in practice since it would require an estimate of σ_∞^2. Parzen (1975) modified his criterion to a version which he calls CAT ('the criterion for autoregressive transfer

functions'). It is given by the function

$$\text{CAT}^*(p) = \begin{cases} \left(n^{-1}\sum_{j=1}^{p}\hat{\sigma}_a^{-2}(j)\right) - \hat{\sigma}_a^{-2}(p) & (p = 1, 2, \ldots, L), \\ -(1+n^{-1}) & (p = 0), \end{cases} \tag{4.7}$$

where $\hat{\sigma}_a^2(j)$ is the residual variance estimate, adjusted for degrees of freedom, when an AR(j) model is fitted to the data.

Usually (4.7) is minimized with respect to $p = 0, 1, 2, \ldots, L$ to obtain the appropriate order of an AR time series model. However, Tong (1979) showed that the absence of $\hat{\sigma}_a^2(0)$ in (4.7) may lead to underestimation of the true order of the process. He suggested the following modification of CAT, namely

$$\text{CAT}(p) = \left(n^{-1}\sum_{j=0}^{p}\hat{\sigma}_a^{-2}(j)\right) - \hat{\sigma}_a^{-2}(p) \quad (p = 0, 1, \ldots, L). \tag{4.8}$$

The criteria (4.7) and (4.8) are further discussed and applied by Parzen (1975, 1978, 1979, 1980) and Parzen & Pagano (1979). Beamish & Priestley (1981) compare by simulation FPE, AIC and CAT* for the determination of the order of various AR processes. For most of the simulated series, the three order-determination methods led to the same order of fitted AR models. An extension of CAT for multiple AR processes is introduced by Parzen (1977).

5 Methods based on information measures

Fisher's well-known concept of information, through its elicitation of the maximum likelihood principle and ancillary statistics, has become the basis of a number of order determination procedures. The same cannot be said of another concept of information which bears a strong correspondence to the probabilistic interpretation of thermodynamic entropy. The use of the concept of entropy is relatively new in the area of time series analysis. But before concentrating on this area, we will deal with the 'standard' statistical situation of having n independent observations on a random variable Y.

Suppose Y is an absolutely continuous random variable characterized by a probability density function $f(Y \mid \beta)$ which is known apart from the k-dimensional parameter vector $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$, $\beta \in \mathbb{R}^k$. Assume that there exists a 'true' vector of values β^* of β and denote the true density $f(Y)$ by $f(Y \mid \beta^*)$. Within this framework it is required to select the β 'closest' to the true parameter vector β^*. The goodness of fit of $f(Y \mid \beta^*)$ with respect to $f(Y \mid \beta)$ can be measured by the entropy; see Akaike (1978b) for a formal justification. This is defined by

$$B(\beta^*, \beta) = \int_{\mathbb{R}} f(Y \mid \beta^*) \log f(Y \mid \beta)\, dy - \int_{\mathbb{R}} f(Y \mid \beta^*) \log f(Y \mid \beta^*)\, dy. \tag{5.1}$$

The second term on the right-hand side of (5.1) is a constant for given $f(Y \mid \beta^*)$. The first term determines the goodness of fit of $f(Y \mid \beta)$ to $f(Y \mid \beta^*)$.

Instead of maximizing the entropy criterion (5.1), the following information criterion

$$I(\beta^*, \beta) = -B(\beta^*, \beta) = \int_{\mathbb{R}} \{\log f(Y \mid \beta^*) - \log f(Y \mid \beta)\} f(Y \mid \beta^*)\, dy \tag{5.2}$$

may be minimized. This last quantity is known as the Kullback–Leibler mean information measure, see, for example, Kullback (1959) and is always greater than zero unless $f(Y \mid \beta^*) = f(Y \mid \beta)$ almost everywhere in the possible range of Y; namely when the model is essentially true.

An operational decision rule can be derived from (5.2) if it is assumed that β is sufficiently close to β^*. Let $\beta = \beta^* + \Delta\beta$ where $\Delta\beta = (\Delta\beta_1, \ldots, \Delta\beta_k)'$ is a $k \times 1$ vector, the norm of which is small, then the following Taylor series expansion of $I(\beta^*, \beta)$ is valid:

$$I(\beta^*, \beta^* + \Delta\beta) \doteq \int_{\mathbb{R}} \left\{ -\sum_i (\Delta\beta)_i \frac{\partial \log f(Y \mid \beta^*)}{\partial \beta_i^*} - \tfrac{1}{2} \sum_i \sum_j (\Delta\beta)_i (\Delta\beta)_j \frac{\partial^2 \log f(Y \mid \beta^*)}{\partial \beta_i^* \, \partial \beta_j^*} \right\} f(Y \mid \beta^*) \, dy.$$

If $f(Y \mid \beta^*)$ is a regular function, the first term on the right-hand side vanishes and we have $I(\beta^*, \beta^* + \Delta\beta) \doteq \frac{1}{2} \|\Delta\beta\|_I^2$, where $\|\Delta\beta\|_I^2 = \Delta\beta' I(\beta^*) \Delta\beta$ with $\|.\|_I^2$ the Euclidian norm and I the Fisher information matrix.

Next, assume that we restrict β to lie in an s-dimensional subspace ($s = 1, \ldots, k-1$), Θ_s, whilst the true parameter vector, β^*, lies in a k-dimensional space ($k > s$). Denoting by β_s^* the projection of β^* onto Θ_s, in the sense of the norm $\|.\|_I^2$, it can be shown that

$$2I(\beta^*, \beta_s) \doteq \|\beta_s^* - \beta^*\|_I^2 + \|\beta_s - \beta_s^*\|_I^2 \tag{5.3}$$

where $\beta_s \in \Theta_s$ and is close to β_s^*.

Replacing β_s in (5.3) by the vector of random variables $\hat{\beta}_s$ which is the restricted maximum likelihood estimates of β^* in Θ_s and using the fact that $n \|\beta_s^* - \hat{\beta}_s\|_I^2$ is asymptotically χ^2 distributed with s degrees of freedom, see, for example, Huber (1967), we obtain that for large n

$$2E[I(\beta^*, \hat{\beta}_s)] \doteq \|\beta_s^* - \beta^*\|_I^2 + s/n. \tag{5.4}$$

The equation in (5.4) above yields a measure of the extent to which $\hat{\beta}_s$ deviates from the true parameter vector β^*. Roughly, this equation implies that the 'expected' deviation of $\hat{\beta}_s$ from β^* has two components: one, representing the error made in selecting an approximate parameter space for β_s^*, the other being the error due to estimating the specified parameter vector.

In empirical work with finite samples, it is of course impossible to minimize (5.4) directly because the first term on the right-hand side of (5.4) is unknown and needs to be estimated. Akaike (1973) shows that under certain regularity conditions the likelihood ratio statistic

$$\text{LR}(Y) = -2 \sum_{i=1}^{n} \log [f(y_i \mid \hat{\beta}_s) / f(y_i \mid \hat{\beta})],$$

where $\hat{\beta}$ is the maximum likelihood estimate of β, is asymptotically distributed as a noncentral χ^2 variable, with $k - s$ degrees of freedom and noncentrality parameter $\|\beta_s^* - \beta^*\|_I^2$, so that

$$n^{-1}\{\text{LR}(Y) + 2s - k\} \tag{5.5}$$

is an unbiased estimate of (5.4). In practice (5.5) can be simplified by dropping terms common to every model since we are only concerned with choosing the minimum. Hence, we need only to minimize

$$-2L_n(Y; \hat{\beta}_s) + 2s \tag{5.6}$$

over $s = 1, \ldots, k-1$. This procedure is usually called Akaike's information criterion (AIC).

We may interpret AIC in the following way. The first term of (5.6) gives the penalty due to badness of fit whilst the $2s$ penalizes the selection of too high an order model and the resulting increased unreliability. As s increases, the first term decreases, and so AIC achieves its minimum, with the model giving the best compromise between fit and unreliability. In this respect the AIC may be regarded as a mathematical formulation of the principle of parsimony.

Akaike (1973, 1974) was the first who recognized the relevance of entropy for the determination of the order (p, q) of an ARMA model. When a time series is generated by a Gaussian distributed ARMA(p, q) process the approximate maximum value of the log likelihood function L_n is given by

$$\hat{L}_n = -\tfrac{1}{2}n \log \hat{\sigma}_a^2 - \tfrac{1}{2}n; \tag{5.7}$$

see, for example, Box & Jenkins (1976, Ch. 5). Ignoring the second term on the right-hand side of (5.7) which is a constant independent of the parameters of the process and the order (p, q), we may express (5.6) in the following equivalent form:

$$\text{AIC}(p, q) = n \log \hat{\sigma}_a^2 + 2(p+q). \tag{5.8}$$

Akaike's procedure suggests to choose that order (p, q) as the most appropriate order for which AIC(p, q) $(p, q = 0, 1, \ldots, L$; L is a preassigned upper limit to the order which needs not necessarily be the same for p and q) attains its minimum.

A number of authors have questioned the origin of the structure dependent term in (5.8). That term resulted from the observation made by Akaike (1977) that the bias caused by the asymptotically distributed estimate of AIC(p, q) is minus twice the number of parameters in the model and, hence, it could serve as the desired 'penalty term'. However, the curious thing about this is the coincidental appearance of the number of parameters in the bias and what seems to be needed to penalize overparametrization. After all, why should the penalty term be a linear function of the number of parameters? Certainly not because the bias happens to be such a function of them. This has motivated several authors, including Akaike himself, to continue the search for a more fundamental and natural method of selecting the order of a time series process.

A second motivation for many authors to consider other ways of determining the order of time series model stems from the fact that the AIC estimate is not consistent but asymptotically overestimates the true order with a nonzero probability. Shibata (1976) proved this result for general AR(p) models. The result also holds for FPE, the variant asymptotically related with AIC. A second and far less complete proof of this was given much later by Kashyap (1980).

Of course, inconsistency of AIC for AR models implies inconsistency for ARMA models. However, the proofs of the two cases are markedly distinct. If (p^*, q^*) is the true order of the ARMA(p, q) model and $p^* < P$, $q^* < Q$, where P, Q are the maximum values considered, then, under very general conditions, Hannan (1982a, Theorem 2) showed that the order obtained by AIC is sure to overestimate (p^*, q^*). Hannan indicates that this result needs to be taken cautiously because of its very asymptotic nature. Furthermore, it also assumes that there is a true ARMA model with order (p^*, q^*). In practice, however, time series will not be generated precisely by an ARMA process and the estimated model may be considered only as an approximation. For multivariate ARMA models Hannan (1982b, p. 462) essentially proved the same result as for the univariate ARMA models.

Asymptotic efficiency of AIC has been investigated by Shibata (1980). Based on a particular loss function for estimating the parameters in an AR(p) model, he shows that, when the true order is infinite, AIC results in an optimal sequence of order estimates as n approaches infinity. Unfortunately, Shibata's result is not above reproach because the loss function used is just the one to derive FPE. Hence, it is no surprise that FPE and methods asymptotically equivalent to it are efficient. An extension of Shibata's work to ARMA processes, which need not necessarily be Gaussian distributed, has been established by Taniguchi (1980).

Applications of AIC are, for example, reported by Akaike (1978b), Jones (1974) and

Tong (1977) for AR processes. The problem of choosing the order of ARMA models, using AIC, is considered by Kitagawa (1977) and Ozaki (1977). Neftçi (1982) uses AIC for the specification of models to economic time series. In the context of modelling nonstationary processes the use of AIC is discussed by Kitagawa (1981), Kitagawa & Akaike (1978) and Kozin & Nakajima (1980). Tong (1975) considers the determination of the order of a Markov chain. Other situations in which AIC has been applied include factor analysis and polynomial fitting (Akaike, 1977). A selection rule for selecting a subset of 'good' models using AIC has been proposed by Duong (1984).

6 Bayesian methods

Bayesian methods for selecting the order of a time series process, as classically proposed in Bayesian inference, should not be considered among the class of methods mentioned in the previous sections. They make use of prior knowledge about the parameter(s) of the model, which is of the form of a probability density function. Several Bayesian criteria have appeared in the literature. We will first briefly formulate the Bayesian methods within a general framework and then discuss these methods for determining the order of ARMA time series processes.

Consider again the situation where a sample of size n, $Y = (y_1, \ldots, y_n)'$ on a continuous random variable is available. Suppose that two models are entertained as having possibly generated or offered a satisfactory representation for this set of observations. Furthermore, assume that these models are characterized by the two hypotheses H_i $(i = 1, 2)$ represented by the probability density function $f(Y \mid \beta_i)$, where β_i $(i = 1, 2)$ is a k_i-dimensional vector of independent unknown parameters, $\beta_i \in \mathbb{R}^{k_i}$. In the Bayesian framework the choice between H_1 and H_2 may be made using the posterior odds ratio

$$K_{12} = \frac{P(H_1 \mid Y)}{P(H_2 \mid Y)} = \frac{P(H_1)}{P(H_2)} \frac{P(Y \mid H_1)}{P(Y \mid H_2)}, \tag{6.1}$$

where $P(H_i)$ denotes the prior probability that H_i is true; see, for example, Zellner (1971, Ch. 10).

Note that K_{12} is equal to the prior odds ratio times the ratio of averaged likelihoods, where the averaging is done using the prior probabilities for the parameters. Evidently this testing procedure differs from the likelihood ratio procedure where for discriminating between H_1 and H_2, maximum likelihood estimates are used as if they were true values of the unknown parameters. The posterior odds ratio is a measure of relative strength of belief in the competing hypotheses H_1 and H_2, given sample evidence. By choosing a model with the highest posterior probability one is acting to minimize the expected loss associated with the acceptance of one of the two hypotheses. That is, if $K_{12} > 1$, one chooses H_1 and if $K_{12} < 1$, one chooses H_2. When $K_{12} = 1$ one is indifferent in terms of expected loss. Of course, this rule is optimal only in case of symmetric loss functions of the models. Generalization of (6.1) to asymmetric loss structures which is often the case when choosing the order of an ARMA model, is straightforward.

It is clear from the above discussion that the model with the highest posterior probability $P(H_i \mid Y)$ must be chosen, given a symmetric loss function. Hence, maximize $P(H_i)P(Y \mid H_i)$. Unfortunately, in practice, it is usually difficult to specify completely the prior probability that a given model is the true one. In many cases $P(H_i)$ must be replaced by some probability that the model is the most adequate one. In the later case, and not entering into interpretation problems, the concept of 'vague' or 'diffuse' prior probability is often used in the literature. The difficulty with this concept can be avoided by following Jeffreys' (1961, p. 117) suggestion and taking equal prior probabilities $P(H_i)$ $(i = 1, 2)$ for each model. Hence the model with the highest $P(Y \mid H_i)$ must be chosen.

Using a theorem of Jeffreys (1961, pp. 139–), cited by Zellner (1971, pp. 31–33), Chow (1981) showed that, with a minimal data information prior, $\log P(Y \mid H_i)$ evaluated at $\hat{\beta}$ is given by

$$\log P(Y \mid H_i) = L_n(Y; \hat{\beta}_i) - \tfrac{1}{2}k_i \log n - \tfrac{1}{2}\log \det I(\hat{\beta}_i) + \tfrac{1}{2}k_i \log (2\pi) + \log f_{H_i}(\hat{\beta}) + O_p(n^{-\frac{1}{2}}), \quad (6.2)$$

where L_n denotes the log likelihood function, $f_{H_i}(\hat{\beta})$ is the prior density when model H_i is true, $\hat{\beta}$ is the maximum likelihood estimate of β_i, $I(\hat{\beta}_i)$ is the information matrix evaluated at $\beta_i = \hat{\beta}_i$, and $O_p(n^{-\frac{1}{2}})$ denotes order of probability.

6.1 *S criterion*

Unfortunately (6.2) involves a difficult problem which has been recognized by many Bayesian statisticians, namely the choice of the prior density $f_{H_i}(\beta)$ for each model H_i. Schwarz (1978) got around this problem by retaining only the first two terms on the right-hand side of (6.2). For large samples he proposed to choose the model for which

$$S = L_n(Y; \hat{\beta}_i) - \tfrac{1}{2}k_i \log n \quad (6.3)$$

is a maximum.

Qualitatively this criterion, like AIC, gives a mathematical formulation of the principle of parsimony. Quantitatively, the difference between (6.3) and AIC is only that in the former the dimension of the model is multiplied by $\frac{1}{2}\log n$. However, this difference is rather crucial since it can be shown that S is a strongly consistent method for AR models; for example, see Geweke & Meese (1981, Th. 5). For nested models AIC does not have this property; see also Bethel (1984).

A critical analysis of S has been given by Akaike (1981). He argued that the use of (6.3) is only possible when there is a clearly defined prior density of $f_{H_i}(\beta)$. This in fact would be a situation where straightforward application of results from Bayesian inference can be used to solve the problem of model selection. Chow (1981) shows that Schwarz's criterion will not necessarily be a good approximation to the log of the posterior probability for finite sample sizes, depending on the nature of the prior probabilities. In a comparative study of AIC and S, Stone (1979) also questions the viability of the latter procedure for finite values of n. He remarks that Schwarz's (1978) work includes as a special case an earlier effort on model discrimination by Jeffreys (1967).

For selecting the most appropriate model in a set of ARMA(p, q) models $(p, q = 0, 1, \ldots, L)$ equation (6.3) becomes asymptotically equivalent to minimization of the function

$$S(p, q) = n \log \hat{\sigma}_a^2 + (p+q) \log n \quad (6.4)$$

over all values of p and q $(p, q = 0, 1, \ldots, L)$.

Several similar forms of Bayesian criteria have appeared in the literature. For example, Atkinson (1980, 1981) suggests to use a part Y_1 of n_1 observations of the total number of n observations $Y = (Y_1, Y_2)$ to obtain the prior density $f_{H_i}(\beta \mid Y_1)$ $(i = 1, 2)$. Denote the residual sum of squares for the ith model $(i = 1, 2)$ from the initial sample Y_1 as $\hat{S}_i(Y_1)$ and from the complete sample as $\hat{S}_i(Y)$. Now the model specified by $f_{H_i}(\beta \mid Y_1)$ and $f(Y \mid \beta_i)$ yields the following log posterior probability:

$$\log P(H_i \mid Y) \propto \tfrac{1}{2}\sigma_a^{-2}\{\hat{S}_i(Y_1) - \hat{S}_i(Y)\} - \tfrac{1}{2}k_i \log (n/n_1) + \text{const.}, \quad (6.5)$$

where all models are assumed equally likely after the initial sample of observations and where $\propto$ denotes proportionality. If σ_a^2 is not known, the maximum likelihood estimate of σ_a^2 for the ith model can be used.

Atkinson suggests choosing the ith model for which (6.5) attains a minimum. For large

n the effect of the initial sample of n_1 observations becomes negligible and (6.5) reduces to

$$-2 \log P(H_i \mid Y) \propto \hat{S}_i(Y)/\sigma_a^2 + k_i \log n. \tag{6.6}$$

The Bayesian method (6.5) can be viewed as a special case of Schwarz's criterion (6.4). For choosing among regression models the use of S is equivalent to the use of (6.6). The distinction between (6.5) and S is that the latter is only valid for large samples. The passage from (6.5) to (6.6) clearly demonstrates that prior knowledge about $f_{H_i}(\beta)$ becomes of less consequence as n increases.

6.2 BIC *criterion*

On minimizing the neg-entropy of the distribution specified by the likelihoods with respect to the prior density of β, Akaike (1977, 1978a, 1979) determines a minimum value for the expected posterior loss function $E_{H_i}[\log f(Y \mid \beta)]$. This minimum value becomes the minimum attainable Bayesian information criterion (BIC) and it takes the form

$$\text{BIC} = (n-k_i) \log \{\hat{S}_i(Y)/(n-k_i)\} + \tfrac{1}{2} k_i \log \{(n\hat{\sigma}_y^2 - \hat{S}_i(Y))/k_i\}, \tag{6.7}$$

where $\hat{\sigma}_y^2$ is the raw sample variance of the observations. Akaike (1977) pointed out that the use of the method of maximum entropy avoids a gross misfit of the prior probability to the data under observation.

After some simple manipulations, it is easy to see that, for ARMA(p, q) models, (6.7) becomes

$$\begin{aligned} \text{BIC}(p, q) = {} & n \log \hat{\sigma}_a^2 - (n-p-q) \log (1-(p+q)/n) + (p+q) \log n \\ & + (p+q) \log \{(p+q)^{-1}(\hat{\sigma}_y^2/\hat{\sigma}_a^2 - 1)\}, \end{aligned} \tag{6.8}$$

where $\hat{\sigma}_a^2$, as before, represents the maximum likelihood estimate of σ_a^2. The set (p, q) $(p, q = 0, 1, \ldots, L)$ which produces the minimum BIC is chosen as the optimal order of the time series model.

When $n \gg (p+q)$ we may use the approximation, $-(n-p-q) \log \{1-(p+q)/n\} \doteq (p+q)$, so that BIC and AIC are connected in the following way:

$$\begin{aligned} \text{BIC}(p, q) &\doteq \text{AIC}(p, q) + (p+q)(\log n - 1) + (p+q) \log \{(p+q)^{-1}(\hat{\sigma}_y^2/\hat{\sigma}_a^2 - 1)\} \\ &\doteq \text{AIC}(p, q) + (p+q)(\log F_{p+q,n} - 1), \end{aligned} \tag{6.9}$$

where $F_{p+q,n} = n(\hat{\sigma}_y^2 - \hat{\sigma}_a^2)/\{(p+q)\hat{\sigma}_a^2\}$. It is clear from (6.9) that BIC penalizes overparametrization more strongly than AIC only if $F_{p+q,n} > e$. This is not automatically guaranteed. If the true model is white noise or if both the sample size and the AR and MA parameters are small then $F_{p+q,n}$ is small. In that case AIC penalizes overparameterization more than BIC.

Rissanen (1978) suggested to obtain the order of an ARMA(p, q) model by minimizing the function

$$\text{BIC}^*(p, q) = \log \hat{\sigma}_a^2 + (p+q) \log n/n \quad (p, q = 0, 1, \ldots, L). \tag{6.10}$$

His derivation is based upon the principle of minimizing the number of binary digits required to redescribe the observed data, when each observation is given with some precision, with the use of the best coding and the assumed model of the process.

We notice that (6.10) is essentially the criterion of Schwarz. The main difference, in addition to the approach, between BIC* and S is that the derivation of (6.10) does not need the Bayesian assumption that the parameters are obtained from a random experiment with a similar distribution. Rissanen (1980) proved that BIC* gives consistent estimates of the order of an AR(p) model. This was extended to ARMA models by Hannan (1980) and Hannan & Rissanen (1982). A slightly improved version of the criterion BIC* has recently been given by Rissanen (1983).

6.3 *Bayesian estimation criterion* (BEC) *and Hannan & Quinn* (HQ) *criterion*

Two other order-determination methods have emerged in the literature which are both Bayesian in inclination and we discuss them here. The first one is Geweke & Meese's (1981) so-called Bayesian estimation method (BEC), which originally was proposed to determine the order of Gaussian distributed linear regression models. The Bayesian estimation method evolves by considering minimization of the so-called estimation criterion (EC). For ARMA(p, q) models it becomes

$$\text{EC}(p, q) = \hat{\sigma}_a^2 + (p+q)g(n)/n \quad (p, q = 0, 1, \ldots, L), \tag{6.11}$$

where $g(n)$ is some function of the sample size n such that $g(n)/n = O(n^{-1})$.

If $g(n) = 0$, then (6.11) reduces to minimization of the residual variance estimate $\hat{\sigma}_a^2$ over all admissible values of p and q; that is (6.11) becomes an order-determination method which is usually employed in the traditional Box–Jenkins approach. If $g(n) > 0$ then EC ascribes a marginal penalty of $g(n)$ for each parameter added to the model, whose impact becomes negligible as n increases. Geweke & Meese (1981) show that (6.11) can be considered as a first-order approximation of Schwarz's criterion (6.4). Accordingly they choose $g(n)$ so that (6.11) is a first-order Taylor approximation to S. This leads to the criterion

$$\text{BEC}(p, q) = \hat{\sigma}_a^2 + (p+q)\hat{\sigma}_L^2 \log n/(n-L) \quad (p, q = 0, 1, \ldots, L), \tag{6.12}$$

where $\hat{\sigma}_L^2$ is the maximum likelihood estimate of the residual variance of the preassigned largest model order. For linear regression models Geweke & Meese (1981) demonstrate that BEC leads to consistent estimates of the model order. This result also holds for AR models.

A second order-determination method which can be considered as a variant of Schwarz's criterion, and is based on a mixture of the classical and Bayesian inference principles, has been proposed by Hannan & Quinn (1979) and Hannan (1980). These authors tried to determine a function $g(n)$ such that it decreases faster than $n^{-1} \log n$, and yet the criterion

$$\log \hat{\sigma}_a^2 + (p+q)g(n)/n$$

gives a consistent estimate of the order of an ARMA model. Obviously, $g(n)$ must decrease more slowly than $2/n$, the penalty term in AIC. Using special theorems about the law of iterated logarithms (Heyde, 1974; Heyde & Scott, 1973), Hannan & Quinn (1979) (HQ), for pure AR models, and Hannan (1980), for ARMA models, suggest to obtain the order of a time series model by minimizing the quantity

$$\text{HQ}(p, q) = \log \hat{\sigma}_a^2 + (p+q)c \log\log n/n \quad (p, q = 0, 1, \ldots, L), \tag{6.13}$$

where c is a constant to be specified.

If c is chosen so that $c > 2$, Hannan & Quinn have proven that for AR models minimization of (6.13) leads to a consistent estimate of the true order. Using (6.13) with $c = 2$ these authors simulated a number of AR(1) processes for various sample sizes and recorded the number of correctly fitted models. Based on a comparison between HQ and AIC, Hannan & Quinn conclude that their simulation results show that HQ leads to underestimation of the order relative to AIC for smaller n and smaller c but better results than AIC for larger n or larger c. They also indicate that for other types of processes this result might not be true. This modest assertion is supported by Hannan (1980), who shows that the result concerning $\log\log n$ rather draws a dividing line between consistent and nonconsistent order-determination methods than is a result to be used. In Hannan's

(1980) paper a proof is also given of the consistency of BIC* for ARMA models. For AR models the conditions upon which Hannan's proof is based are recently replaced by weaker conditions by Hannan & Kavalieris (1983).

Before we make some final comments about the general relevance of the order determination methods presented here, it is worth mentioning that we have not considered the problem of estimating unidentified models, i.e. models where both p and q exceed the corresponding true values. As Hannan (1982a, § 1) has pointed out, it will be very difficult to find the maximum of the Gaussian likelihood in such a case since there will be a myriad of close together local maxima with some zeros of $\theta_q(z)$ near the unit circle. Hannan (1980) and Hannan & Rissanen (1982) have considered this problem in more detail. Hannan (1980) proves theoretical results about the maximum likelihood estimation in unidentified models. Hannan & Rissanen (1982) produce a practical procedure for order selection which basically consists of the following three stages.

In the first stage an AR(L) polynomial is fitted to the data and the parameter estimates $\hat{\phi}_j$ ($j = 1, \ldots, L$) are used to obtain $\hat{a}_t = y_t - \sum \hat{\phi}_j y_{t-j}$ ($t \geq L+1$), where the sum is over $j = 1, \ldots, L$. The value of L is chosen such that it minimizes either AIC or BIC*. In the second stage, estimates of the true parameters of the model are computed by regressing y_t on y_{t-j} ($j = 1, \ldots, p$) and $\hat{a}_{t-j}$ ($j = 1, \ldots, q$). To do this, Hannan & Rissanen (1982) give a computationally efficient recursive estimation algorithm which may be considered as an extension of the Durbin–Levinson recursive algorithm (Durbin, 1960). If $\tilde{\sigma}^2_{p,q}$ denotes the residual variance of these regressions, then the estimate $(\tilde{p}, \tilde{q})$, say, of the true order is obtained by minimizing the function

$$\log \tilde{\sigma}^2_{p,q} + (p+q) \log n/n \quad (p \leq P, q \leq Q), \tag{6.14}$$

where P and Q are predetermined values, chosen sufficiently large. Since the coefficients $\tilde{\phi}_j$ and $\tilde{\theta}_j$ for which (6.14) attains its minimum value are not asymptotically efficient, Hannan & Rissanen (1982) suggest to estimate the parameters of the process using any of the available algorithms for the computation of the maximum of the Gaussian likelihood function, starting the computations with $(\tilde{p}, \tilde{q})$ for the order and $\tilde{\phi}_j$, $\tilde{\theta}_j$ and $\tilde{\sigma}^2_{p,q}$ as initial values for the parameters. This is considered to be the third stage of the procedure.

Hannan & Rissanen (1982) show that this method of order determination has huge computational advantages over straightforward computation of BIC* (or S). However, they also point out in a correction to their paper, that the procedure can lead to overestimation of the true orders at the second stage. Modifications of the second stage which correct this and which do not seriously affect the speed of convergence of the estimates, are recently given by Hannan & Kavalieris (1984). These authors also prove that their modifications of the original procedure yield consistent estimates of the true order of the process.

In closing this section we shall make some brief comments about the usefulness of 'Bayesian' inspired order-determination methods, in contrast to some of the previously discussed non-Bayesian ones. The methods based on the maximum likelihood principle use functions of the maximum likelihood estimates of the parameters and compare these to a prescribed fixed significance level. Often this significance level and the associated critical value are chosen with implicit consideration of the relative costs or losses associated with the probability of making a type I and type II error. However, in a Bayesian context explicit consideration is given to the loss structure, through the posterior odds ratio, and this leads to a natural choice of the critical value. For example, increasing the sample size will affect the posterior odds ratio and consequently will change a specified loss structure whereas with a method based on the likelihood ratio principle usually no changes will be made to the significance level in such a situation.

Another point which has been increasingly recognized in the literature is that use of a

fixed level of significance for choosing between models that differ appreciably in their parametric dimensionality is not entirely justified since it does not take account of the increase in 'variability' of the parameter estimates when new parameters are added to the model. In fact, the significance level of statistics based on the maximum likelihood principle (for example, LR, W and LM statistics) should be greater, in small samples, when the size of the competing models is large. Hence, order determination methods which incorporate some form of adjustment to prior information about the model structure and sample size, may be preferred to likelihood methods.

Adjustment of the order determination method to changes in the model structure is apparent in AIC. But, as stated in § 5, the heuristic argument introducing the structure dependent term in AIC is debatable. By relating prior information to the parameter structure of the model, we have seen in this section that Bayesian order-determination methods elicit, in a natural way, terms that penalize an overparametrization of the model. A practical method through which prior information can be incorporated in an easy way in the specification of ARMA models has been developed by Monahan (1983) for small samples.

Structure dependent terms penalizing overparameterization also appear in the 'Bayesian' inspired methods BEC and HQ, and in the Rissanen (1983) type of criteria. All these criteria will penalize overparameterization more strongly than AIC and its asymptotic equivalents. For this reason, AIC should not be used when the dimensions of the competing model are large relative to the sample size. Also when n is small one should not use such procedures like AIC. Instead one may have to take the risk involved in prior choice of a particular model but, as is clear from the order determination methods presented in the last part of this section, one does not have to be a Bayesian to do that.

7 Relationships between the various methods

Several relationships between the various order-determination methods have already been mentioned in the previous sections. To emphasize the close resemblance in form, a compendium of the most important linkages between certain methods will now be presented.

7.1 *The* AIC *and likelihood ratio methods*

An interesting parallel exists between AIC and the likelihood ratio testing procedure. For determining the best model structure within the set of AR models it follows from (2.3), with $c(n) = n$, and (5.8) that approximately

$$\text{AIC}(s, 0) - \text{AIC}(L, 0) = \text{LR}(s) - 2(L - s) \quad (s = 0, 1, \ldots, L-1). \tag{7.1}$$

Hence, this relation can be used to compute the AIC values if likelihood ratios are given. Also, for a given sample size n and for various values of s, there is a sequence of significance levels such that AIC and LR lead to the choice of the same AR model. For mixed ARMA models a similar relation to (7.1) exists when the order of the processes under H_0 and H_1 satisfies certain conditions.

7.2 *The* AIC *and* FPE *methods*

It is easily seen that, for values of $n \gg (p+q)$, we have the approximation

$$\frac{\{n+(p+q)\}}{\{n-(p+q)\}} = 1 + \frac{2(p+q)}{n} + O(n^{-2}).$$

Obviously, this implies that for an ARMA model

$$\begin{aligned}\log[\text{FPE}(p,q)] &= \log\left[\hat{\sigma}_a^2\frac{\{n+(p+q)\}}{\{n-(p+q)\}}\right]\\ &= \log\hat{\sigma}_a^2+\log\left\{1+\frac{2(p+q)}{n}+O(n^{-2})\right\}\\ &= \log\hat{\sigma}_a^2+\frac{2(p+q)}{n}+O(n^{-2}).\end{aligned} \tag{7.2}$$

Hence, from (5.8), we have

$$\text{AIC}(p,q)\doteq n\log\text{FPE}(p,q). \tag{7.3}$$

In other words, AIC and FPE are asymptotically equivalent methods for determining the order (p, q) of an ARMA model.

7.3 *Relation of* AIC, CAT *and some other criteria*

A slightly different approach was followed by Tong (1979) on relating AIC to CAT for AR processes. Denoting the unbiased residual variance $n\hat{\sigma}_a^2/(n-p)$ by $\tilde{\sigma}_a^2(p)$, p being the number of estimable parameters in the model, and ignoring terms of larger order than $O(n^{-1})$, he used the following asymptotically equivalent form of AIC:

$$\text{AIC}(p,0)\doteq n\log\tilde{\sigma}_a^2(p)+p. \tag{7.4}$$

Thus

$$\text{AIC}(p,0)-\text{AIC}(p-1,0)\doteq 1+n\log\{\tilde{\sigma}_a^2(p)/\tilde{\sigma}_a^2(p-1)\}=1+n\log\{1+\omega\},$$

where $\omega=(\tilde{\sigma}_a^2(p)-\tilde{\sigma}_a^2(p-1))/\tilde{\sigma}_a^2(p-1)$. Now it can be shown that

$$\text{AIC}(p,0)-\text{AIC}(p-1,0)\doteq 1+n\omega+nO_p(n^{-1}). \tag{7.5}$$

Furthermore, from (4.8),

$$\begin{aligned}\text{CAT}(p)-\text{CAT}(p-1) &= (n^{-1}-1)\tilde{\sigma}_a^{-2}(p)+\tilde{\sigma}_a^{-2}(p-1)\\ &= \tilde{\sigma}_a^{-2}(p-1)\{1+(n^{-1}-1)(1+\omega)^{-1}\}\\ &\doteq \tilde{\kappa}^2(p)\{\text{AIC}(p,0)-\text{AIC}(p-1,0)+O_p(n^{-\frac{1}{2}})\},\end{aligned} \tag{7.6}$$

where $\tilde{\kappa}^2(p)=n^{-1}\tilde{\sigma}_a^{-2}(p-1)$. Hence, modulo the term $\tilde{\kappa}^2(p)O_p(n^{-\frac{1}{2}})$, CAT and AIC are directly related by

$$\text{CAT}(p)-\text{CAT}(p-1)\doteq\tilde{\kappa}^2(p)\{\text{AIC}(p,0)-\text{AIC}(p-1,0)\}. \tag{7.7}$$

However, the fact that $\tilde{\kappa}^2(p)$ depends on $\tilde{\sigma}_a^{-2}(p-1)$ does not necessarily have to imply that the global minimum of CAT and that of AIC occur at the same autoregressive order p. In particular, Parzen (1977) shows that

$$\text{CAT}(p)\leqslant -\exp\{-\text{AIC}(p,0)\}. \tag{7.8}$$

The above results show that FPE, AIC and CAT exhibit similar asymptotic behaviour for AR processes. The connection between BIC and AIC was noted in § 6 for ARMA processes. Bednar & Roberts (1982) show that AIC can be calculated directly from the R- and S-arrays for ARMA processes. The asymptotic equivalence between AIC and cross-validation is established by Stone (1977). The relationships between the R- and S-arrays, the Δ-arrays and the extended (and regular) partial autocorrelations were cited in § 3.

8 Some concluding remarks

In this paper we briefly discussed several of the most important order-determination methods used in time series analysis today. It was not our intention to give a complete survey of all the methods presented in the literature. The picture of approaches, methods and results is too diverse and, moreover, scattered over several scientific disciplines. Neither have we made any recommendation of a single order-determination method as a definitely superior method. Such a recommendation can only be done on the basis of the final purpose one has for the selected time series model. For example, if one were to employ a time series model for forecasting alone, an order determination method which minimizes the one-step-ahead quadratic forecasting error is more preferable than methods which merely lead to fitted models that provide a good representation of the data in the sample.

It is clear from the discussion in this paper that there are many methods available, each with seemingly desirable characteristics. Though attempts have been made to link certain order-determination methods the whole selection process of an order may look, on the surface, somewhat esoteric. To an unskilled time series analyst it will not always be clear when to use a particular method. On the other hand, he or she may not be cognizant with the alternative methods available and their possible advantages. Hence, it seems highly desirable to order the order determination methods into a coherent and systematic framework. Such a framework is at present absent in the literature but we feel that the discussion given here may be useful as a stepping-stone for this. However, we shall emphasize that the formulation of a complete framework requires more theoretical as well as empirical work than is available at present.

Research should be directed to a theoretical investigation of the small-sample behaviour of the various order-determination methods. Many other theoretical as well as empirical problems should be investigated. For example, how sensitive are some of the order-determination methods to deviations from the familiar but restrictive assumption that the white noise process is Gaussian distributed? How robust are the various methods in the presence of outliers and bad data? It is our hope and expectation that these problems will be solved in the very near future. But, at present, they are beyond the scope of this paper.

Acknowledgements

The authors would like to thank Franz Palm, Pentti Saikkonen and two anonymous referees for valuable comments on a previous version of this paper.

References

Abraham, B. & Ledolter, J. (1984). A note on inverse autocorrelations. *Biometrika* **71,** 609–14.

Aitchison, J. & Silvey, S.D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29,** 813–828.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21,** 243–247.

Akaike, H. (1970a). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22,** 203–217.

Akaike, H. (1970b). A fundamental relation between predictor identification and power spectrum estimation. *Ann. Inst. Statist. Math.* **22,** 219–223.

Akaike, H. (1971). Autoregressive model fitting for control. *Ann. Inst. Statist. Math.* **23,** 163–180.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Ed. B.N. Petrov and F. Csáki, pp. 267–281. Budapest: Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19,** 716–723.

Akaike, H. (1977). On entropy maximization principle. In *Applications of Statistics*, Ed. P.R. Krisnaiah, pp. 27–41. Amsterdam: North-Holland.

Akaike, H. (1978a). A Bayesian analysis of the minimum AIC procedure. *Ann. Inst. Statist. Math.* **30,** 9–14.
Akaike, H. (1978b). On the likelihood of a time series model. *Statistician* **27,** 217–235.
Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure. *Biometrika* **66,** 237–242.
Akaike, H. (1981). Likelihood of a model and information criteria. *J. Econometrics* **16,** 3–14.
Allen, D.M. (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics* **13,** 469–475.
Amemiya, T. (1980). Selection of regressors. *Int. Econ. Rev.* **21,** 331–354.
Anderson, T.W. (1963). Determination of the order of dependence in normally distributed time series. In *Proceedings of the Symposium of Time Series Analysis*, Ed. M. Rosenblatt, pp. 425–446. New York: Wiley.
Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* New York: Wiley.
Atkinson, A.C. (1980). A note on the generalized information criterion of choice of a model. *Biometrika* **67,** 413–418.
Atkinson, A.C. (1981). Likelihood ratios, posterior odds and information criteria. *J. Econometrics* **16,** 15–20.
Bartlett, M.S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *J. R. Statist. Soc.* B **8,** 27–41.
Bartlett, M.S. & Diananda, P.H. (1950). Extensions of Quenouille's test for autoregressive schemes. *J. R. Statist. Soc.* B **12,** 108–115.
Beamish, N. & Priestley, M.B. (1981). A study of autoregressive and window spectral estimation. *Appl. Statist.* **30,** 41–58.
Bednar, J.B. & Roberts, B. (1982). A statistical interpretation of the *R* and *S* array approach to ARMA model identification. In *Applied Time Series Analysis*, Ed. O.D. Anderson, pp. 21–31, Amsterdam: North-Holland.
Beguin, J.M., Gourieroux, C. & Monfort, A. (1980). Identification of a mixed autoregressive-moving average process: The corner method. In *Time Series*, Ed. O.D. Anderson, pp. 423–436. Amsterdam: North-Holland.
Bessler, D.A., & Binkley, J.K. (1980). Autoregressive filtering of some economic data using PRESS and FPE. In *Proc. Bus. Econ. Statist. Sect., Am. Statist. Assoc.*, pp. 261–265.
Bethel, J.W. (1984). Asymptotic properties of information-theoretic methods of model selection. Ph.D. Thesis, University of California, Davis.
Bhansali, R.J. (1973). A Monte Carlo comparison of the regression method and the spectral methods of prediction. *J. Am. Statist. Assoc.* **68,** 621–625.
Bhansali, R.J. (1978). Linear prediction by autoregressive model fitting in the time domain. *Ann. Statist.* **6,** 224–231.
Bhansali, R.J. (1980). The autoregressive and the window estimate of the inverse correlation function. *Biometrika* **67,** 511–566.
Bhansali, R.J. (1983a). The inverse partial correlation function of a time series and its application. *J. Mult. Anal.* **13,** 310–327.
Bhansali, R.J. (1983b). Estimation of the order of a moving average model from autoregressive and window estimates of the inverse correlation function. *J. Time Series Anal.* **4,** 137–162.
Bhansali, R.J. & Downham, D.Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. *Biometrika* **64,** 547–551.
Boom, A.J.W. van den & van den Enden, A.W.M. (1974). The determination of the order of process and noise dynamics. *Automatica* **10,** 245–256.
Box, G.E.P. & Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden Day.
Breusch, T.S. (1978). Testing for autocorrelation in dynamic linear models. *Aust. Econ. Papers* **17,** 334–355.
Breusch, T.S. & Pagan, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Rev. Econ. Studies* **47,** 239–253.
Caines, P.E., Keng, C.W. & Sethi, S.P. (1981). Causality analysis and multivariate autoregressive modelling with an application to supermarket sales analysis. *J. Econ. Dynamics Control* **3,** 267–298.
Chatfield, C. (1979). Inverse autocorrelations. *J. R. Statist. Soc.* A **142,** 363–377.
Chow, G.C. (1981). A comparison of the information and posterior probability criteria for model selection. *J. Econometrics* **16,** 21–33.
Cleveland, W.S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics* **14,** 277–298.
Davies, N. & Petruccelli, J.D. (1984). On the use of the general partial autocorrelation function for order determination in ARMA(p, q) processes. *J. Am. Statist. Assoc.* **79,** 374–377.
Duong, Q.P. (1984). On the choice of the order of autoregressive models: A ranking and selection approach. *J. Time Series Anal.* **5,** 145–157.
Durbin, J. (1960). The fitting of time series models. *Rev. Int. Statist. Inst.* **28,** 233–244.
Gersch, W. & Sharpe, D.R. (1973). Estimation of power spectra with finite-order autoregressive models. *IEEE Trans. Autom. Control* **18,** 367–369.
Geweke, J.F. & Meese, R.A. (1981). Estimating regression models of finite but unknown order. *Int. Econ. Rev.* **22,** 55–70.
Glasbey, C.A. (1982). A generalization of partial autocorrelations useful in identifying ARMA models. *Technometrics* **24,** 223–228.
Gooijer, J.G. de & Heuts, R.M.J. (1981). The corner method: an investigation of an order determination procedure for general ARMA processes. *J. Oper. Res. Soc.* **32,** 1039–1042.
Gray, H.L., Kelley, G.D. & McIntire, D.D. (1978). A new approach to ARMA modelling. *Comm. Statist.* B **7,** 1–77.

Hamdi, A. (1982). Identification des processus ARMA à coéfficients dépendants du temps. Ph.D. Thesis, L'Université Paul-Sabatier de Toulouse.
Hannan, E.J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8,** 1071–1081.
Hannan, E.J. (1982a). Testing for autocorrelation and Akaike's criterion. In *Essays in Statistical Science*, Ed. J. Gani and E.J. Hannan, pp. 403–412. Applied Probability Trust, The University, Sheffield.
Hannan, E.J. (1982b). Estimating the dimension of a linear system. *J. Mult. Anal.* **11,** 459–473.
Hannan, E.J. & Kavalieris, L. (1983). The convergence of autocorrelations and autoregressions. *Aust. J. Statist.* **25,** 287–297.
Hannan, E.J. & Kavalieris, L. (1984). A method for autoregressive-moving average estimation. *Biometrika* **71,** 273–280.
Hannan, E.J. & Quinn, B.G. (1979). The determination of the order of an autoregression. *J. R. Statist. Soc.* B **41,** 190–195.
Hannan, E.J. & Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69,** 81–94. Correction (1983), **70,** 303.
Heyde, C.C. (1974). An iterated logarithm result for autocorrelations of a stationary linear process. *Ann. Prob.* **2,** 328–332.
Heyde, C.C. & Scott, D.J. (1973). Invariance principles for the laws of the iterated logarithm for martingales and processes with stationary increments. *Ann. Prob.* **1,** 428–436.
Hipel, K.W., McLeod, A.I. & Lennox, W.C. (1977). Advances in Box–Jenkins modelling 1. Model construction. *Water Resources Res.* **13,** 567–575.
Hjorth, U. & Holmqvist, L. (1981). A model selection based on validation with applications to pressure and temperature prognosis. *Appl. Statist.* **30,** 264–274.
Hjorth, U. (1982). Model selection and forward validation. *Scand. J. Statist.* **9,** 95–105.
Hosking, J.R.M. (1980). Lagrange-multiplier tests of time series models. *J. R. Statist. Soc.* B **42,** 170–181.
Hsiao, C. (1979a). Autoregressive modelling of Canadian money and income data. *J. Am. Statist. Assoc.* **74,** 553–560.
Hsiao, C. (1979b). Causality tests in econometrics. *J. Econ. Dynamics Control* **1,** 321–346.
Hsiao, C. (1982). Time series modelling and causal ordering of Canadian money, income and interest rates. In *Time Series Analysis: Theory and Practice*, **1,** Ed. O.D. Anderson, pp. 671–699. Amsterdam: North-Holland.
Huber, P.J. (1967). The behaviour of maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symp.* **1,** 221–233.
Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford: Clarendon Press.
Jeffreys, H. (1967). *Theory of Probability*, 4th ed. Oxford University Press.
Jenkins, G.M. & Alavi, A.S. (1981). Some aspects of modelling and forecasting multivariate time series. *J. Time Series Anal.* **2,** 1–47.
Jones, R.H. (1974). Identification and autoregressive spectrum estimation. *IEEE Trans. Autom. Control* **19,** 894–897.
Jones, R.H. (1975). Fitting autoregressions. *J. Am. Statist. Assoc.* **70,** 590–592.
Kashyap, R.L. (1980). Inconsistency of the AIC rule for estimating the order of autoregressive models. *IEEE Trans. Autom. Control* **25,** 996–998.
Kitagawa, G. (1977). On a search procedure for the optimal AR-MA order. *Ann. Inst. Statist. Math.* **29,** 319–332.
Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *J. Time Series Anal.* **2,** 103–116.
Kitagawa, G. & Akaike, H. (1978). A procedure for the modelling of nonstationary time series. *Ann. Inst. Statist. Math.* **30,** 351–363.
Kozin, F. & Nakajima, F. (1980). The order determination problem for linear time varying AR models. *IEEE Trans. Autom. Control* **25,** 250–257.
Kromer, R.E. (1969). Asymptotic properties of the autoregressive spectral estimator. Ph.D. Thesis, Stanford University.
Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley.
Liu, L.-M. & Hanssens, D.M. (1982). Identification of multiple-input transfer function models. *Comm. Statist.* B **11,** 297–314.
Lütkepohl, H. (1985). Comparison of criteria for estimating the order of a vector autoregressive processes. *J. Time Series Anal.* **6,** 35–52.
MacKinnon, J.G. (1983). Model specification tests against non-nested alternatives. *Econometric Rev.* **2,** 85–110.
McClave, J.T. (1975). Subset autoregression. *Technometrics* **17,** 213–220.
McClave, J.T. (1978). Estimating the order of autoregressive models: The max χ^2 method. *J. Am. Statist. Assoc.* **73,** 122–128.
McLeod, A.I., Hipel, K.W. & Lennox, W.C. (1977). Advances in Box–Jenkins modelling 2. Applications. *Water Resources Res.* **13,** 577–586.
Monahan, J.F. (1983). Fully Bayesian analysis of ARMA time series models. *J. Econometrics* **21,** 307–331.
Neftçi, S.N. (1982). Specification of economic time series models using Akaike's criterion. *J. Am. Statist. Assoc.* **77,** 537–540.
Nelson, C.R. & Shea, G.S. (1979). Hypothesis testing based on goodness-of-fit in the moving average time series model. *J. Econometrics* **10,** 221–226.

Newbold, P. (1983). Model checking in time series analysis. In *Applied Time Series Analysis of Economic Data*, Ed. A. Zellner, pp. 133–143. Washington D.C.: Government Printing Office.
Newbold, P. & Bos, T. (1983). On q-conditioned partial correlations. *J. Time Series Anal.* **4,** 53–55.
Otomo, T., Nakagawa, T. & Akaike, H. (1972). Statistical approach to computer control of rotary kilns. *Automatica* **8,** 35–48.
Ozaki, T. (1977). On the order determination of ARIMA models. *Appl. Statist.* **26,** 290–301.
Parzen, E. (1974). Some recent advances in time series modelling. *IEEE Trans. Autom. Control* **19,** 723–730.
Parzen, E. (1975). Some solutions to the time series modelling and prediction problem. In *The Search for Oil*, Ed. D.B. Owen, pp. 1–16. New York: Marcel Dekker.
Parzen, E. (1977). Multiple time series: Determining the order of approximating autoregressive schemes. In *Multivariate Analysis-IV*, Ed. P.R. Krishnaiah, pp. 283–295. Amsterdam: North-Holland.
Parzen, E. (1978). Comments on 'Analysis and modelling of seasonal time series' by George E.P. Box, Steven C. Hillmer, and George C. Tiao. In *Seasonal Analysis of Economic Time Series*, Ed. A. Zellner, pp. 338–341. Washington D.C.: U.S. Government Printing Office.
Parzen, E. (1979). Forecasting and whitening filter estimation. In *Forecasting Studies in Management Sciences*, Ed. S. Makridakis and S.C. Wheelwright, pp. 149–165. Amsterdam: North-Holland.
Parzen, E. (1980). Time series modelling, spectral analysis, and forecasting. In *Directions in Time Series*, Ed. D.R. Brillinger and G.C. Tiao, pp. 80–111. Hayward: California, Institute of Mathematical Statistics.
Parzen, E. & Pagano, M. (1979). An approach to modelling seasonally stationary time series. *J. Econometrics* **9,** 137–153.
Petruccelli, J.D. & Davies, N. (1984). Some restrictions on the use of corner method hypothesis tests. *Comm. Statist.* **13** (5), 543–551.
Piccolo, D. & Tunnicliffe-Wilson, G. (1984). A unified approach to ARMA model identification and preliminary estimation. *J. Time Series Anal.* **5,** 183–204.
Rissanen, J. (1978). Modelling by shortest data description. *Automatica* **14,** 465–471.
Rissanen, J. (1980). Consistent order-estimates of autoregressive processes by shortest description of data. In *Analysis and Optimization of Stochastic Systems*, Ed. O. Jacobs, M. Davis, M. Dempster, C. Harris and P. Parks, pp. 451–461. New York: Academic Press.
Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* **11,** 416–431.
Rudra, A. (1954). A critical survey of some test methods in time series analysis. *Cal. Statist. Bull.* **5,** 165–177.
Sawyer, K. (1980). The theory of econometric model selection. Ph.D. Thesis, Australian National University.
Schmidt, P. (1974). Choosing among alternative linear regression models. *Atl. Econ. J.* **1,** 7–13.
Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6,** 461–464.
Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63,** 117–126.
Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8,** 147–164.
Söderström, T. (1977). On model structure testing in systems identification. *Int. J. Control* **26,** 1–18.
Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Statist. Soc.* B **39,** 44–47.
Stone, M. (1978). Cross-validation: A review. *Math. Oper. Statist.* **9,** 127–139.
Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *J. R. Statist. Soc.* B **41,** 267–278.
Takemura, A. (1984). A generalization of autocorrelation and partial autocorrelation functions useful for identification of ARMA(p, q) processes. Technical Report no. 84–16, Department of Statistics, Purdue University.
Taniguchi, M. (1980). On selection of the order of the spectral density model for a stationary process. *Ann. Inst. Statist. Math.* **32,** 401–419.
Tiao, G.C. & Tsay, R.S. (1983a). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11,** 856–871.
Tiao, G.C. & Tsay, R.S. (1983b). Multiple time series modelling and extended sample cross-correlations. *J. Business Econ. Statist.* **1,** 43–56.
Tong, H. (1975). Determination of the order of a Markov chain by Akaike's information criterion. *IEEE Trans. Info. Theory* **21,** 476–480.
Tong, H. (1977). Some comments on the Canadian lynx data. *J. R. Statist. Soc.* A **140,** 432–436.
Tong, H. (1979). A note on local equivalence of two recent approaches to autoregressive order determination. *Int. J. Control* **29,** 441–446.
Tsay, R.S. & Tiao, G.C. (1983). Identification of multiplicative ARMA models for seasonal time series. Technical Report No. 7, Statistics Research Center, Graduate School of Business, University of Chicago.
Tsay, R.S. & Tiao, G.C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *J. Am. Statist. Assoc.* **79,** 84–96.
Tukey, J.W. Akaike, H., Robinson, E.A. & Granger, C.W.J. (1978). Comments on Gray, Kelley & McIntire. *Comm. Statist.* B **7,** 79–92.
Unbehauen, H. & Göhring, B. (1974). Tests for determining the model order in parameter estimation. *Automatica* **10,** 233–244.
Walker, A.M. (1967). Some tests of separate families of hypothesis in time series analysis. *Biometrika* **54,** 39–68.

Whittle, P. (1952). Tests of fit in time series. *Biometrika* **39,** 309–318.
Whittle, P. (1954). Some recent contributions to the theory of stationary time series. In *A Study in the Analysis of Stationary Time Series*, 2nd ed., by H. Wold, pp. 196–228. Stockholm: Almquist and Wicksell.
Woodward, W.A. & Gray, H.L. (1978). New ARMA models for Wölfer's sunspot data. *Comm. Statist.* B **7,** 97–115.
Woodward, W.A. & Gray, H.L. (1981). On the relationship between the *S* array and the Box–Jenkins method of ARMA model identification. *J. Am. Statist. Assoc.* **76,** 579–587.
Yamamoto, Y. (1976). Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. *Appl. Statist.* **25,** 123–127.
Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.
Zellner, A. & Palm, F.C. (1974). Time series analysis and simultaneous equation econometric models. *J. Econometrics* **2,** 17–54.

Résumé

Le problème du choix de l'ordre d'un modèle ARMA a été étudié depuis longtemps. Souvent la construction de ce modèle est basée sur les fonctions empiriques d'autocorrélation et d'autocorrélation partielle de Box et Jenkins. Plusieurs autres procédures sont disponsibles. Dans cet article nous passons en revue les plus importantes de ces procédures et nous examinons leurs propriétés statistiques.

[*Received September 1983, revised November 1984*]