

# ON SOME AMBIGUITIES ASSOCIATED WITH THE FITTING OF ARMA MODELS TO TIME SERIES

BY DAVID F. FINDLEY

*U.S. Bureau of the Census*

**Abstract.** Examples are presented illustrating some ambiguities associated with the application of ARMA models to problems of signal extraction, multistep-ahead forecasting, spectrum approximation and linear quadratic control. Except in the signal extraction example, the ambiguities arise either from lack of sufficient autocovariance data to completely determine the process, or, often relatedly, from the approximate nature of the models used.

**Keywords.** Time series; ARMA models; Forecasting; Linear quadratic control; Signal extraction; Spectrum estimation.

## I. INTRODUCTION

The ambiguities we discuss are of various kinds: In section 2, we illustrate some consequences for signal extraction of inappropriately using 'parsimony' to remove the ambiguity inherent in the selection of an ARMA model for an unobserved component of a stationary ARMA  $(p, p)$  process when only the spectrum of the observed process is known. In the remaining sections, we mainly look at consequences for various applications of only knowing a finite segment  $\Gamma(0), \dots, \Gamma(m)$  of the autocovariance sequence of the observed stationary process. The examples make it clear that which model is 'best' can depend not only on  $\Gamma(0), \dots, \Gamma(m)$  but also on the purpose for which the model is fit, i.e., on criteria appropriate to the intended application.

It is shown in section 3 that there are always AR, MA and ARMA processes compatible with  $\Gamma(0), \dots, \Gamma(m)$ , provided that  $[\Gamma(k-j)]_{0 \leq j, k \leq m}$  is nonsingular. In section 4, we demonstrate that modelling, in the usual sense of fitting a model with optimal one-step-ahead forecasting properties, may not be the most desirable procedure if the goal is multi-step-ahead forecasting or spectrum estimation. It is pointed out in section 5 that, as a consequence of the somewhat restricted type of linear control law specified in the usual formulation of the optimal linear quadratic controller design problem, see Åström (1970, pp. 258–260), the resulting optimal controllers enjoy a certain possibly valuable robustness property against model misspecifications.

In the final section 6, additional references are given to recent literature dealing with ARMA models as approximations. Shibata's results, showing that model selection procedures giving consistent estimates of model order can perform badly in this context in comparison with an inconsistent procedure, are mentioned.

## 2. SIGNAL EXTRACTION AND PARSIMONY

If the purely nondeterministic, stationary, univariate, zero mean series  $x(t)$ , with *known* spectral density  $f(\lambda)$ , is the sum of two uncorrelated, unobservable, component series,  $s(t)$  and  $n(t)$ , with spectral densities  $f_s(\lambda)$  and  $f_n(\lambda)$ , then it is a classical result that the frequency response function of the linear filter which provides the least mean square estimate of  $s(t)$  based on  $x(t+k)$  ( $k=0, \pm 1, \dots$ ) is given by

$$h(e^{i\lambda}) = f_s(\lambda)/f(\lambda) = \{1 + f_n(\lambda)/f_s(\lambda)\}^{-1}, \quad (2.1)$$

see, for example, Hannan (1970, p. 169). Often,  $f_s(\lambda)$  and  $f_n(\lambda)$  must be determined by assuming models for  $s(t)$  and  $n(t)$ . If the assumed model for  $s(t)$  is incorrect, it is to be expected that the expression attempting to specify  $h(e^{i\lambda})$  as in (2.1) will produce a degraded estimate of  $s(t)$ . In this section we illustrate degradations which result from inappropriately using the most parsimonious candidate model for  $s(t)$  (compatible with  $f(\lambda)$ ).

In what follows, it is assumed that  $n(t)$  is a white noise process with variance  $\sigma_n^2$ . Thus  $f(\lambda) = f_s(\lambda) + \sigma_n^2$  and  $f(\lambda) \geq \sigma_n^2$  for all  $\lambda$  in  $[-\pi, \pi]$ . For any  $\tilde{\sigma}^2$  satisfying

$$0 < \tilde{\sigma}^2 \leq \inf_{\lambda} f(\lambda), \quad (2.2)$$

if we set  $\tilde{f}_s(\lambda) = f(\lambda) - \tilde{\sigma}^2$ , then the decomposition

$$f(\lambda) = \tilde{f}_s(\lambda) + \tilde{\sigma}^2 \quad (2.3)$$

is *admissible* in the sense that  $\tilde{f}_s(\lambda)$  is a spectral density function and  $\tilde{f}_n(\lambda) = \tilde{\sigma}^2$  is the spectral density function of a white noise process. We shall assume that  $x(t)$  is an ARMA( $p, p$ ) process for some  $p > 0$ , so that

$$f(\lambda) = \frac{\sigma^2 |1 - \beta_1 e^{i\lambda} - \dots - \beta_p e^{ip\lambda}|^2}{|1 - \phi_1 e^{i\lambda} - \dots - \phi_p e^{ip\lambda}|^2}$$

with  $\phi_p \neq 0 \neq \beta_p$ . Then  $\tilde{f}_s(\lambda) = f(\lambda) - \tilde{\sigma}^2$  is easily seen to be the spectral density function of an ARMA( $p, q$ ) process for some  $q \leq p$ . Box, Hillmer and Tiao (1978) describe and illustrate consequences of the variety of choices for  $\tilde{\sigma}^2$ , without focusing on the case  $q < p$ . If  $q < p$ , we say that the decomposition (2.3) has *reduced order*. Since  $f(\lambda)$  is known, the coefficients of the backshift operator polynomial  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  can be determined, as can the autocovariance at lag  $p$ ,  $\Gamma_z(p)$ , of the MA( $p$ ) process  $z(t) = \phi(B)x(t)$ . With this information, it is easy to determine whether or not  $f(\lambda)$  has a reduced order decomposition: If we multiply (2.3) by  $|\phi(e^{i\lambda})|^2$  and then equate coefficients of  $e^{ip\lambda}$  on both sides, we obtain the following result from (2.2).

**PROPOSITION 2.1.** *With  $\Gamma_z(p)$  as defined above, set  $\sigma_{n,0}^2 = -\phi_p^{-1}\Gamma_z(p)$  and  $f_{s,0}(\lambda) = f(\lambda) - \sigma_{n,0}^2$ . Then  $f(\lambda)$  has a reduced order decomposition if and only if*

$$0 < \sigma_{n,0}^2 \leq \min_{\lambda} f(\lambda). \quad (2.4)$$

The right-hand inequality in (2.4) must be strict if  $p = 1$ . If these inequalities are satisfied, then there is exactly one reduced order decomposition and it is given by

$$f(\lambda) = f_{s,0}(\lambda) + \sigma_{n,0}^2.$$

If (2.4) is satisfied and no additional information is available for determining  $f_s(\lambda)$ , then considerations of parsimony and simplicity might make it appealing to choose the signal extraction filter determined by  $h_0(e^{i\lambda}) = f_{s,0}(\lambda)/f(\lambda)$ , corresponding to the assumption that  $f_s(\lambda) = f_{s,0}(\lambda)$ , i.e., that  $s(t)$  is an ARMA ( $p, q$ ) process with  $q < p$ . If this assumption is incorrect, it follows from the identity

$$\{h_0^2(e^{i\lambda})/h^2(e^{i\lambda})\}h^2(e^{i\lambda})f(\lambda) = h_0^2(e^{i\lambda})f(\lambda)$$

that the factor  $h_0^2(e^{i\lambda})/h^2(e^{i\lambda})$  measures the distortion to the spectrum of the optimally estimated signal resulting from the use of the sub-optimal filter defined by  $h_0(e^{i\lambda})$ . Similarly,

$$\text{RELERR} = \left\{ \frac{\int_{-\pi}^{\pi} |h(e^{i\lambda}) - h_0(e^{i\lambda})|^2 f(\lambda) d\lambda}{\int_{-\pi}^{\pi} h^2(e^{i\lambda}) f(\lambda) d\lambda} \right\}^{1/2} \tag{2.5}$$

measures the root mean square difference between the optimally and suboptimally extracted signal estimates, relative to the standard deviation of the optimal estimate.

We shall now evaluate these quantities for the case  $p = 1$ , assuming  $s(t)$  is an ARMA (1, 1) process satisfying

$$s(t) = \phi s(t-1) + e(t) - \theta e(t-1),$$

where  $e(t)$  is a white noise process with variance  $\sigma_s^2$ , such that  $e(t)$  is uncorrelated with  $s(t-k)$  for  $k \geq 1$ . If  $\Gamma(k)$  and  $\Gamma_s(k)$  denote the autocovariance at lag  $k$  of  $x(t)$  and  $s(t)$ , respectively, for  $k = 0, 1, 2, \dots$ , then  $\Gamma(0) = \Gamma_s(0) + \sigma_n^2$  and  $\Gamma(k) = \Gamma_s(k)$  for  $k \geq 1$ . Since  $f(\lambda)$  is known, so are  $\Gamma(0)$ ,  $\Gamma(1)$ ,  $\Gamma(2)$  and  $\phi = \Gamma(2)/\Gamma(1)$ . By setting  $\theta = 0$ , temporarily, we calculate

$$\sigma_{n,0}^2 = \Gamma(0) - \phi^{-1}\Gamma(1)$$

and  $f_{s,0}(\lambda) = \sigma_{s,0}^2 |1 - \phi e^{i\lambda}|^{-2}$ , where

$$\sigma_{s,0}^2 = \phi^{-1}(1 - \phi^2)\Gamma(1).$$

Thus,  $h_0(e^{i\lambda}) = \{1 + (\sigma_{n,0}^2/\sigma_{s,0}^2)|1 - \phi e^{i\lambda}|^2\}^{-1}$  and

$$h_0^2(e^{i\lambda})/h^2(e^{i\lambda}) = (\sigma_{s,0}^2/\sigma_s^2)^2 |1 - \theta e^{i\lambda}|^{-4}. \tag{2.6}$$

From the formulas for  $\Gamma_s(0)$  and  $\Gamma_s(1)$  given in (3.4.6-3.4.7) of Box and Jenkins (1976, p. 76), we obtain the alternative formulas  $\sigma_{n,0}^2 = \sigma_n^2 + \phi^{-1}\theta\sigma_s^2$  and  $\sigma_{s,0}^2 = (1 - \phi\theta)(1 - \phi^{-1}\theta)\sigma_s^2$ , which can be used to evaluate (2.5) and (2.6) for various choices of  $\sigma_n^2$ ,  $\sigma_s^2$ ,  $\phi$  and  $\theta$ . (An algebraic formula for RELERR in (2.5) is given in the Appendix.) The inequality  $\sigma_{s,0}^2 > 0$ , or  $\phi^{-1}\theta < 1$ , is equivalent to the strict

form of the right hand inequality in (2.4). The tabled values of (2.5) and (2.6) given below demonstrate two of the possibilities. The tabled quantity  $\Gamma_s(0)/\sigma_n^2$  gives the total power signal-to-noise ratio for each example.

TABLE 2.1  
VALUES OF  $h_0^2(e^{i\lambda})/h^2(e^{i\lambda})$ ,  $\lambda = 0, 0.1\pi, \dots, \pi$  AND RELERR FOR TWO  
CHOICES OF  $(\phi, \theta, \sigma_s^2)$  WITH  $\sigma_n^2 = 1$

$(\phi, \theta, \sigma_s^2)$ :	(-0.8, 0.4, 1.0)	(0.5, 0.3, 10.0)
0	30.25	0.48
0.1 $\pi$	24.60	0.43
0.2 $\pi$	14.91	0.32
0.3 $\pi$	8.24	0.21
0.4 $\pi$	4.71	0.14
$\lambda$ 0.5 $\pi$	2.91	0.10
0.6 $\pi$	1.98	0.07
0.7 $\pi$	1.46	0.06
0.8 $\pi$	1.20	0.05
0.9 $\pi$	1.06	0.04
$\pi$	1.02	0.04
RELERR:	0.17	0.59
$\Gamma_s(0)/\sigma_n^2$ :	5.00	10.53

A signal extraction approach to seasonal adjustment due to W. P. Cleveland (see Cleveland, Dempster and Stith (1980)) involves specifying fairly simple uncorrelated ARMA processes  $\tilde{s}(t)$  and  $\tilde{n}(t)$  for the components,  $s(t)$  and  $n(t)$ , of the observed series  $x(t)$  and using the process spectral densities  $\tilde{f}_s(\lambda)$  and  $\tilde{f}_n(\lambda)$  in (2.1) to obtain the signal extraction filter. Cleveland's procedure is to choose the component models' parameters so that a segment of the autocovariance sequence of  $\tilde{s}(t) + \tilde{n}(t)$  approximates the corresponding segment of the autocovariance sequence of  $x(t)$ . An economic time series seldom has a 'true' seasonal component, so that for seasonal adjustment there is latitude in the choice of models for the seasonal and nonseasonal components. Except in such circumstances, however, procedures analogous to Cleveland's cannot be recommended: In the examples of table 2.1, the approximation is perfect for all segments, but the distortion due to the use of too simple a model for  $s(t)$  is severe.

### 3. MODELS COMPATIBLE WITH FINITELY MANY AUTOCOVARIANCES

The most widely used stationary autoregressive-moving average model 'identification' procedures for  $d$ -dimensional vector time series ( $d \geq 1$ ) are based on the analysis of a finite number of autocovariance or sample autocovariance matrices at lags 0 through  $m$ ,  $\Gamma(k)$  ( $0 \leq k \leq m$ ) ( $m$  is variously determined), or on the analysis of quantities derivable from these: see Akaike (1974), Box and Jenkins (1976), Gray, Kelly and McIntire (1978), Woodward and Gray (1981), and Box and Tiao (1981). Such models are usually fit in order to obtain forecasts or

spectrum estimates, to design controllers, or to do signal extraction, rather than to identify the 'true' model generating the observed series. In fact, consider the symmetric block Toeplitz matrix  $[\Gamma(k-j)]_{0 \leq j, k \leq m}$ , in which the  $(j, k)$ -block,  $\Gamma(k-j)$ , is set equal to the transpose  $\Gamma'(j-k)$  of  $\Gamma(j-k)$  when  $j > k$ . If this matrix is nonsingular, then these model identification procedures cannot, by themselves, uniquely identify the true model when it exists, as the following result shows.

PROPOSITION 3.1. *If  $[\Gamma(k-j)]_{0 \leq j, k \leq m}$  is positive definite, then there exist processes of the following sorts, all having  $\Gamma(0), \dots, \Gamma(m)$  as their autocovariance matrices at lags  $0, \dots, m$ , respectively: (i) (Robinson and Wold (1963)) an AR( $p$ ) process with  $p \leq m$ ; (ii) an MA( $q$ ) process; (iii) for any  $r$  satisfying  $1 \leq r \leq p-1$ , an ARMA( $r, n$ ) process, with  $n \leq r + \max\{m, q\}$ , having any preassigned autoregressive polynomial  $B(z) = I + b(1)z + \dots + b(r)z^r$  such that the zeros of  $\det B(z)$  lie in  $\{|z| > 1\}$ ; and, finally, in the univariate case ( $d = 1$ ), (iv) (Georgiu (1983)) for any  $s$  satisfying  $1 \leq s \leq p$ , an ARMA( $n, s$ ) process, with  $n \leq p$ , whose moving average polynomial can be an arbitrarily prespecified polynomial of degree  $s$ .*

PROOF. The result of Robinson and Wold (1963) shows that a mean zero, stationary,  $d$ -dimensional, AR( $p$ ) process  $y(t)$  of order  $p \leq m$  exists whose autocovariance matrix at lag  $k$  coincides with  $\Gamma(k)$  for  $0 \leq k \leq m$ . We wish to establish the existence of an MA( $q$ ) process with this property. We can assume  $p \neq 0$ . If we use  $\Gamma(k)$  to denote the autocovariance matrix of  $y(t)$  at lag  $k$  for every  $k > m$ , then the spectral density matrix of this autoregressive process is, by definition,

$$f_y(\lambda) = \Gamma(0) + \sum_{k=1}^{\infty} \{ \Gamma(k)e^{ik\lambda} + \Gamma'(k)e^{-ik\lambda} \}. \tag{3.1}$$

On the other hand, if the autoregression equation for  $y(t)$  is

$$y(t) + a(1)y(t-1) + \dots + a(p)y(t-p) = e(t),$$

where  $e(t)$  has covariance matrix  $G$  and is uncorrelated with  $y(t-1), y(t-2), \dots$ , then Whittle (1963) shows that the zeros of the determinant of  $A(z) = I + a(1)z + \dots + a(p)z^p$ ,  $z$  a complex variable, have modulus greater than 1. It therefore follows from the alternative formula,  $f_y(\lambda) = A^{-1}(e^{i\lambda})GA'^{-1}(e^{i\lambda})$ , that  $f_y(\lambda)$  is positive definite for all  $\lambda$  and that the convergence of the series in (3.1) is uniform on  $[-\pi, \pi]$  (even exponentially rapid). Consequently, there exists a smallest  $N \geq m$  such that

$$f^N(\lambda) = \Gamma(0) + \sum_{k=1}^N \{ \Gamma(k)e^{ik\lambda} + \Gamma'(k)e^{-k\lambda} \} \tag{3.2}$$

is positive definite for all  $\lambda$ . If  $q$  denotes the smallest positive integer (possibly less than  $m$  if  $\Gamma(m) = 0$ ) such that  $f^q(\lambda) = f^N(\lambda)$ , then it follows from the matrix version of the theorem of Fejér and Riesz given in Hannan (1970, p. 66) that  $f^q(\lambda)$  is the spectral density matrix of a  $d$ -dimensional moving average process

of order  $q$ , which, by (3.2), has  $\Gamma(0), \dots, \Gamma(m)$  as its first  $m+1$  autocovariance matrices. This verifies the assertion (ii).

Given  $B(z)$  as in (iii), an ARMA( $r, n$ ) process having this property, with  $n \leq r + \max\{m, q\}$ , can be obtained by defining its spectral density matrix to be

$$\tilde{f}(\lambda) = f^q(\lambda) + \sum_{k=s+1}^{\infty} \{\tilde{\Gamma}(k)e^{ik\lambda} + \tilde{\Gamma}'(k)e^{-ik\lambda}\}$$

where  $s = \max\{m, q\}$  and where  $\tilde{\Gamma}(k)$ ,  $k = 0, 1, 2, \dots$ , is an autocovariance matrix sequence satisfying

$$\tilde{\Gamma}(k) + b(1)\tilde{\Gamma}(k-1) + \dots + b(r)\tilde{\Gamma}(k-r) = 0 \quad (3.3)$$

for  $k \geq 1$ . If necessary, we rescale the  $\tilde{\Gamma}(k)$ -sequence by multiplication with a positive constant to insure that

$$-\delta I < \sum_{k=s+1}^{\infty} \{\tilde{\Gamma}(k)e^{ik\lambda} + \tilde{\Gamma}'(k)e^{-ik\lambda}\} < \delta I \quad (-\pi \leq \lambda \leq \pi), \quad (3.4)$$

for some  $\delta > 0$  for which  $f^q(\lambda) \geq \delta I$  ( $-\pi \leq \lambda \leq \pi$ ) holds. Then  $\tilde{f}(\lambda)$  is positive definite for all  $\lambda$ , by (3.4), and (3.3) and (3.4) imply that  $B(e^{i\lambda})\tilde{f}(\lambda)B'(e^{-i\lambda})$  is a positive definite matrix polynomial of degree  $n \leq r + s$ . Consequently,  $\tilde{f}(\lambda)$  is the spectral density matrix of an ARMA( $r, n$ ) process whose first  $m+1$  autocovariance matrices are  $\Gamma(0), \dots, \Gamma(m)$ . This establishes (iii).

The final result (iv) appears as Theorem 7.1 in the Ph.D. dissertation of Georgiu (1983).

We noted in (i) that the autoregressive order  $p$  of the process  $y_t$  satisfies  $p \leq m$ . However, for a given value of  $m$ , there is no such generally valid upper limit for the smallest moving average order  $q$  in (ii). This follows from the case  $d = 1$ , where, as C. W. J. Granger showed, for a fixed  $q$ , the lag 1 autocorrelation  $\rho_q(1)$  of a moving average process of order  $q$  satisfies  $|\rho_q(1)| \leq \cos\{\pi/(q+2)\}$  (see Davies, Pate and Frost (1974)), so that  $q$  can be forced to be larger than any prespecified integer by choosing  $|\Gamma(1)/\Gamma(0)|$  close enough to 1.

The models of the AR( $p$ ) and MA( $q$ ) processes described in (i) and (ii) of Proposition 3.1 may have more parameters than the most parsimonious ARMA model compatible with  $\Gamma(0), \dots, \Gamma(m)$ . Unfortunately, even in the univariate case, deciding when a compatible ARMA( $p_0, q_0$ ) model exists with  $p_0 + q_0 < \min\{p, q\}$  is difficult at present. Kalman (1982) discusses the case  $m = 3$  under the restriction  $q_0 < p_0$ .

#### 4. SOME CONSEQUENCES OF THE USE OF APPROXIMATE MODELS

In the two subsections of this section, we present simple theoretical examples from the applications areas of multistep-ahead prediction and spectrum estimation. In each case we assume that the correct values are available for whatever autocovariances of the observed series  $x(t)$  are required to determine the parameters of the model being fit. By using correct autocovariance data, we avoid

confounding the consequences of modeling error or model inexactness with those due to the sampling variability of parameter estimates. In each example,  $x(t)$  is assumed to be a covariance stationary time series with mean 0, whose autocovariance sequence is  $\Gamma(k)$ ,  $k=0, 1, \dots$ . We set  $\rho(k) = \Gamma(k)/\Gamma(0)$ ,  $k=0, 1, \dots$ .

4a. *m-Step-Ahead Prediction*

Suppose that for  $m \geq 1$  we seek to approximate  $x(t+m)$  by an expression of the form  $\psi x(t)$ , where the optimal value  $\psi(m)$  of  $\psi$  minimizes the expected mean square error  $E\{x(t+m) - \psi x(t)\}^2$ . From the easily verified fact that  $x(t+m) - \rho(m)x(t)$  is uncorrelated with  $x(t)$ , it follows that

$$E\{x(t+m) - \psi x(t)\}^2 = E\{x(t+m) - \rho(m)x(t)\}^2 + (\psi - \rho(m))^2 E x^2(t), \quad (4.1)$$

and hence that the optimal choice of  $\psi$  is

$$\psi(m) = \rho(m), \quad m = 1, 2, \dots$$

If  $x(t)$  is a first order autoregressive process, then  $\psi(m) = \psi^m(1) = \rho^m(1)$  for any  $m \geq 1$ , but, in general,  $\psi(m)$  will not be a function of  $\psi(k)$ ,  $k < m$ . *In particular, the optimal m-step-ahead forecast function of a given form is not necessarily determined by the optimal one-step-ahead prediction-error-filter, i.e., the optimal whitening filter (or 'model') of that form.* If  $x(t)$  is modeled as a first-order autoregression and, as a consequence, an  $m$ -step-ahead predictor for  $x(t+m)$  having the form  $\rho^m(1)x_t$  is used, then, by (4.1), an increase over the optimal mean square prediction error in the amount

$$(\rho(m) - \rho^m(1))^2 \Gamma(0) \quad (4.2)$$

will result if this model specification is incorrect.

To illustrate (4.2), we suppose that  $m > 2$  and that  $x(t)$  is an  $AR(m)$  process whose partial autocorrelations at lags 1 to  $m$  are given by  $\rho(1), 0, \dots, 0, \phi_{mm}$ , respectively, where  $|\rho(1)| < 1$  and  $|\phi_{mm}| < 1$ . (If  $m$  is not too small, or if  $|\phi_{mm}|$  is not too large, a realization from such a process could quite conceivably be mismodeled as an  $AR(1)$  process.) It follows from the Levinson-Durbin algorithm, Box and Jenkins (1976, p. 83), that

$$\rho(k) = \begin{cases} \rho^k(1), & 1 \leq k \leq m-1 \\ \rho^m(1) + (1 - \rho^2(1))\phi_{mm}, & k = m, \end{cases}$$

and that  $x(t)$  satisfies

$$x(t) = \rho(1)x(t-1) - \rho(1)\phi_{mm}x(t-m+1) + \phi_{mm}x(t-m) + e(t)$$

where  $e(t)$  is a white noise process (with variance  $(1 - \rho^2(1))(1 - \phi_{mm}^2)\Gamma(0)$ ). For this process  $x(t)$ , the quantity in (4.2) is equal to  $(1 - \rho^2(1))^2 \phi_{mm}^2 \Gamma(0)$ , which can be quite substantial (relative to  $\Gamma(0)$ ) if  $\rho(1)$  is fairly small and  $\phi_{mm}$  is fairly large (e.g.,  $|\rho(1)| \leq 1/2$ ,  $|\phi_{mm}| \geq 1/\sqrt{2}$ ).

Since observed series can rarely be expected to be realizations of ARMA processes or other simple parametric models, the above analysis suggests that if the goal is to do  $m$ -step-ahead forecasting for  $m$  somewhat larger than 1, then, rather than using a fitted (whitening) model to produce the forecasts, a direct approach might be better: For example, if  $x(0), \dots, x(N)$  are observed and tentatively identified as being fairly well modeled as an autoregressive process, then a direct predictor could be calculated by choosing  $\psi_0, \dots, \psi_{p-1}$  to minimize

$$\sum_{t=p}^{N-m} \left\{ x(t+m) - \sum_{k=0}^{p-1} \psi_k x(t-k) \right\}^2. \quad (4.3)$$

Shibata (1980) describes a strategy which is asymptotically optimal, according to a rather natural criterion (equivalent to MAIC when  $m=1$ ), for choosing the length  $p$  of the predicting autoregression in (4.3) for a given prediction lead  $m$  in a situation in which  $x(t)$  is not an autoregressive process. Findley (1983b) applies this procedure to two well known series. Findley (1983a) includes a theoretical analysis of Shibata's procedure for a simple example. The analysis suggests that the problem of selecting  $p$  increases in difficulty with increasing forecast lead  $m$ . Gersch and Kitagawa (1983) describe and illustrate an alternative procedure for doing  $m$ -step-ahead forecasting directly for series with trend and seasonality.

#### 4b. Spectrum Estimation

We now assume that the covariance structure of  $x(t)$  is determined by a spectral density function  $f(\lambda)$  and that an approximating AR ( $p$ ) or ARMA ( $p, q$ ) spectral density  $g(\lambda)$  is sought which minimizes the relative mean square error,

$$\int_{-\pi}^{\pi} \left( \frac{f(\lambda) - g(\lambda)}{f(\lambda)} \right)^2 d\lambda. \quad (4.4)$$

When  $f(\lambda)$  is unknown, Shibata (1981) shows that his procedure for determining the order  $p$  of the best one-step-ahead autoregressive predictor for  $x(t)$  also leads to asymptotically optimal AR ( $p$ ) spectral estimators of  $f(\lambda)$  under (4.4) in a situation in which  $x(t)$  is not an autoregressive process. Shibata's results (which have been generalized by Taniguchi (1980) to the case of approximating ARMA models) require that  $p$  be allowed to increase without limit as the number of observations becomes infinite. We shall demonstrate that if  $f(\lambda)$  is known and  $p$  is held fixed, then the AR ( $p$ )-spectral density  $g(\lambda)$  minimizing (4.4) need not be the spectral density associated with the AR ( $p$ ) model for  $x(t)$  minimizing the mean square one-step-ahead prediction error.

If  $g(\lambda)$  has the form  $\sigma^2 |1 - \phi e^{i\lambda} - \dots - \phi_p e^{ip\lambda}|^{-2}$ , then by setting equal to zero the derivative of (4.4) with respect to  $\sigma^2$ , one obtains an expression for  $\sigma^2$  as a function of the coefficients  $\phi_1, \dots, \phi_p$ . Using this expression in (4.4), the minimization of (4.4) is seen to reduce to choosing these coefficients to minimize the



scale invariant quantity

$$\int_{-\pi}^{\pi} \frac{g^2(\lambda)}{f^2(\lambda)} d\lambda / \left( \int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda \right)^2.$$

Applying Bessel's equality to the numerator, we see that this quantity is equal to

$$1 + 2 \sum_{k=1}^{\infty} \rho_{g/f}^2(k), \tag{4.5}$$

where, for  $k = 1, 2, \dots$ ,  $\rho_{g/f}(k)$  denotes the autocorrelation at lag  $k$  of a time series whose spectral density is  $g(\lambda)/f(\lambda)$ . In the simple case in which  $x(t)$  is an MA (1) process and an AR (1) spectral density is being fit, we have  $f(\lambda) \sim |1 - \theta e^{i\lambda}|^2$  and  $g(\lambda) \sim |1 - \phi e^{i\lambda}|^{-2}$ , so that  $g(\lambda)/f(\lambda)$  is the spectral density of an AR (2) process whose characteristic polynomial has the roots  $\theta^{-1}$  and  $\phi^{-1}$ . In this case, by (3.2.20) of Box and Jenkins (1976),

$$\rho_{f/g}(k) = C(\theta, \phi)\theta^k + C(\phi, \theta)\phi^k, \quad k = 0, 1, 2, \dots$$

with  $C(\theta, \phi) = \theta(1 - \phi^2)/(\theta - \phi)(1 + \theta\phi)$ . Hence, (4.5) is easily shown to equal  $1 + 2H(\theta, \phi)$ , with

$$H(\theta, \phi) = \theta^2 C^2(\theta, \phi)(1 - \theta^2)^{-1} + 2\theta\phi C(\phi, \theta)C(\theta, \phi)(1 - \phi\theta)^{-1} + \phi^2 C^2(\phi, \theta)(1 - \phi^2)^{-1} \quad (\phi \neq \theta).$$

For this special case, the autocorrelation at lag one of  $x(t)$  is given by  $\rho(1) = -\theta(1 + \theta^2)^{-1}$ . In particular,  $\rho(1) = 0.4$  if  $\theta = -0.5$ , so that the characteristic polynomial of the AR (1) model for  $x(t)$  having optimal one-step-ahead forecasts is  $1 - 0.4z$ . However,  $H(-0.5, 0.4) = 0.15476$ , and the value of  $\phi$  minimizing  $H(-0.5, \phi)$  is 0.45, with  $H(-0.5, 0.45) = 0.09931$ . Thus the characteristic polynomial of the AR (1) model providing the best spectrum estimate according to the criterion (4.4) is  $1 - 0.45z$ . These two AR (1) models are different, and this illustrates the point mentioned in the Introduction: When approximate models are used, which model is best is determined by the intended application.

5. DESIGN OF LINEAR QUADRATIC CONTROLLERS

Let  $x(t)$  be an observed vector variable which is to be controlled by manipulation of the observed vector variable  $u(t)$ . Suppose that the joint process  $[x'(t)u'(t)]'$  is stationary with mean zero and with known autocovariance matrices at lags 0 and 1, and that the joint process has been tentatively identified as a first order autoregression. From the given autocovariance data, we calculate the coefficient matrices  $\phi$  and  $\psi$  of the least mean square approximation equation,

$$x(t) = \phi x(t-1) + \psi u(t-1) + e(t), \tag{5.1}$$

in which the components of  $e(t)$  are uncorrelated with those of  $x(t-1)$  and  $u(t-1)$ , and, if the joint process is actually first order autoregressive, also with

the components of  $x(t-k)$  and  $u(t-k)$  for all  $k \geq 2$ . Suppose a mean quadratic cost function is given of the form

$$E \left\{ \sum_{t=2}^N x'(t)Ax(t) + u'(t-1)Bu(t-1) \right\}, \quad (5.2)$$

where  $A$  is positive semidefinite and  $B$  is positive definite. From the uncorrelatedness of  $e(t)$  with  $x(t-1)$  and  $u(t-1)$ , it follows by way of the derivation of the optimal controllers based on the identity (6.5) of Åström (1970, pp. 278–279), that the controller gain functions  $L(t)$  ( $1 \leq t \leq N-1$ ) for the linear control laws of the form

$$u(t) = -L(t)x(t) \quad (1 \leq t \leq N-1) \quad (5.3)$$

which minimize (5.2) are the usual ones given by the Dynamic Programming algorithm as described, for example, by Åström (1970, p. 267). The resulting optimal control laws of the form (5.3) can be shown to also be the optimal laws defining  $u(t)$  as a linear function of  $x(t)$ ,  $x(t-1)$ , ... when  $e(t)$  is uncorrelated with  $x(t-k)$  and  $u(t-k)$  for all  $t \geq 2$ , but they need not be optimal in this extended sense when  $e(t)$  is uncorrelated only with  $x(t-1)$  and  $u(t-1)$ : Consider, for example, the simple situation of scalar variables  $x(t)$  and  $u(t)$  and the cost function (5.2) with  $N=2$  and  $A \neq 0$ . Suppose that  $\gamma = Ee(t)x(t-2)$  is nonzero. If we substitute  $x(2) = \phi x(1) + \psi u(1) + e(2)$  and  $u(1) = -\tilde{L}(1)x(1) - \tilde{M}(1)x(0)$  into this cost function, it follows, by setting equal to zero the partial derivatives of the resulting quadratic in  $\tilde{L}(1)$  and  $\tilde{M}(1)$ , that the optimal (cost minimizing) values of  $\tilde{L}(1)$  and  $\tilde{M}(1)$  are given by

$$\tilde{L}(1) = \psi\phi(B/A + \psi^2)^{-1} + \rho(1)\tilde{M}(1)$$

and

$$\tilde{M}(1) = \gamma\psi(B/A + \psi^2)^{-1}\{1 - \rho^2(1)\}^{-1},$$

where  $\rho(1)$  denotes the lag 1 autocorrelation of  $x(t)$ . Since  $\tilde{M}(1) \neq 0$ ; because  $\gamma \neq 0$ , we have shown that the usual 'optimal' control law for this problem, given by  $u(1) = -L(1)x(1)$  with  $L(1) = \psi\phi(B/A + \psi^2)^{-1}$ , can be improved upon by using a law which also includes a lagged value of the controlled variable.

An alternative method for obtaining this improved control law is to define the vector  $x_{\text{new}}(t) = [x(t) \ x(t-1)]'$  and find the optimal control law of the form  $u(t) = L_{\text{new}}(t)x_{\text{new}}(t)$ , with  $L_{\text{new}}(t) = [\tilde{L}(t) \ \tilde{M}(t)]$ , using analogues of (5.1) and (5.2) in which  $x_{\text{new}}(t)$  replaces  $x(t)$ . Thus, in this example, the ambiguity associated with the choice of the form of the control law can be viewed as arising from the ambiguity associated with the choice of the components of  $x(t)$ ,  $u(t)$ ,  $A$  and  $B$ .

The observation made above, that the optimal control law of the form (5.3) is completely determined by the least mean square approximation equation (5.1) rather than by the 'true model' for  $x(t)$ , may help to explain the success of controllers obtained using least-squares-fitted statistical models, with carefully chosen observed vectors  $x(t)$  and  $u(t)$ , in situations in which the physical model of the process makes it clear that the statistical model used can only be approximate, see Otomo, Nakagawa and Akaike (1972) and Nakamura and Akaike (1981).

6. FINAL COMMENTS CONCERNING APPROXIMATING OR MISSPECIFIED MODELS

In addition to the references given above, there are a number of interesting recent papers dealing with properties of approximating or misspecified models and the estimates of their parameters: See, for example, An, Chen and Hannan (1982), Bhansali (1981), DuFour (1982), Hosoya and Taniguchi (1982), Lewis and Reinsel (1982), Ljung and Caines (1979) and Ogata (1980).

The derivation of the AIC model selection criterion by Akaike (1973) explicitly includes the possibility that the models being fit are, at best, only approximate. The impact of Akaike's work (see, for example, This Week's Citation Classic (1981)) stimulated the investigations of Shibata and Taniguchi referred to above, which, along with Akaike (1970a, 1970b, 1971), are the only substantial theoretical papers known to the author which discuss application-specific model selection criteria. This would seem to be a worthwhile focus for future research, particularly in the area of statistical modeling for controller design: for example, the question of the optimality of the procedure of Akaike (1971) does not seem to have been investigated. Forecasting and spectrum estimation performance criteria different from those used by Shibata and Taniguchi could also be examined.

The papers by Shibata and Taniguchi also show that modifications of Akaike's AIC to obtain a procedure which consistently estimates the correct model orders  $p$  and  $q$  when the observed series comes from an ARMA ( $p, q$ ) process (see, for example, Hannan (1980)) can lead to unboundedly large relative loss, compared with the efficient performance of minimum AIC models, under loss functions appropriate for prediction or spectrum estimation, when ARMA ( $p, q$ ) models are only approximate. Thus, an important consequence of the usually approximate nature of models is that, for model selection, consistency can be an *undesirable* property. This suggests that the approximating model viewpoint is not without some rather profound implications for statistics.

ACKNOWLEDGEMENTS

The author is indebted to the referee for the Whittle (1963) reference and for a number of suggestions, and to William Bell, David Dickey, Will Gersch, Myron Katzoff and Peter Zadrozny for comments on preliminary versions of this paper.

APPENDIX

In this appendix, a derivation is sketched of an algebraic formula for RELERR in (2.5) for the case  $p = 1$ . To begin, one verifies that, in the notation of section 2, the function  $|h(e^{i\lambda}) - h_0(e^{i\lambda})|^2 f(\lambda)$  is proportional to

$$\frac{|-\theta e^{-i\lambda} + (1 + \theta^2 - \sigma_{\varepsilon_0}^2 / \sigma_{\varepsilon}^2) - \theta e^{i\lambda}|^2}{|1 - \phi e^{i\lambda}|^2 |1 - \beta e^{i\lambda}|^2}, \tag{A.1}$$

where

$$\beta = \begin{cases} \gamma - (\gamma^2 - 1)^{1/2}, & \gamma > 1 \\ \gamma + (\gamma^2 - 1)^{1/2}, & \gamma < -1, \end{cases}$$

with

$$\gamma = (1/2)\{1 + \theta^2 + (\sigma_n^2/\sigma_s^2)(1 + \phi^2)\}/\{\theta + \phi\sigma_n^2/\sigma_s^2\}.$$

Also,

$$\frac{-\theta e^{-i\lambda} + (1 + \theta^2 - \sigma_{s,0}^2/\sigma_s^2) - \theta e^{i\lambda}}{(1 - \phi e^{i\lambda})(1 - \beta e^{i\lambda})} = -\theta e^{-i\lambda} + \{1 + \theta^2 - \sigma_{s,0}^2/\sigma_s^2 - \theta(\gamma_1\phi + \gamma_2\beta)\} \\ + \sum_{k=1}^{\infty} \{\gamma_1\tilde{\gamma}_1\phi^{k-1} + \gamma_2\tilde{\gamma}_2\beta^{k-1}\}e^{ik\lambda}, \quad (\text{A.2})$$

where  $\gamma_1 = \phi/(\phi - \beta)$ ,  $\gamma_2 = \beta/(\beta - \phi)$ ,

$$\tilde{\gamma}_1 = \phi(1 + \theta^2 - \sigma_{s,0}^2/\sigma_s^2) - \theta(1 + \phi^2),$$

and

$$\tilde{\gamma}_2 = \beta(1 + \theta^2 - \sigma_{s,0}^2/\sigma_s^2) - \theta(1 + \beta^2).$$

Now apply Bessel's equality to evaluate the integral of (A.1) over  $[-\pi, \pi]$  as  $2\pi$  times the sum of squares of the coefficients in (A.2). Using the geometric sum formulas  $\sum_{k=0}^{\infty} \phi^k = (1 - \phi)^{-1}$ , etc. to simplify the sum of squares, one obtains that this integral is  $2\pi$  times

$$F(\sigma_{s,0}^2/\sigma_s^2, \sigma_n^2/\sigma_s^2, \phi, \theta) = \theta^2 + \{1 + \theta^2 - \sigma_{s,0}^2/\sigma_s^2 - \theta(\gamma_1\phi + \gamma_2\beta)\}^2 \\ + \frac{(\gamma_1\tilde{\gamma}_1)^2}{1 - \phi^2} + \frac{2\gamma_1\tilde{\gamma}_1\gamma_2\tilde{\gamma}_2}{1 - \phi\beta} + \frac{(\gamma_2\tilde{\gamma}_2)^2}{1 - \beta^2}.$$

With this notation, RELERR is given by

$$\text{RELERR} = F(\sigma_{s,0}^2/\sigma_s^2, \sigma_n^2/\sigma_s^2, \phi, \theta)/F(0, \sigma_n^2/\sigma_s^2, \phi, \theta).$$

#### REFERENCES

- AKAIKE, H. (1970a) Statistical Predictor Identification. *Ann. Inst. Stat. Math.* 22, 203-217.
- AKAIKE, H. (1970b) A Fundamental Relation between Predictor Identification and Power Spectrum Estimation. *Ann. Inst. Stat. Math.* 22, 219-223.
- AKAIKE, H. (1971) Autoregressive Model Fitting for Control. *Ann. Inst. Stat. Math.* 23, 163-180.
- AKAIKE, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium on Information Theory*, (edited by B. W. Petrov and F. Csaki), Akademiai-Kiado; Budapest, pp. 267-281.
- AKAIKE, H. (1974) Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. In *System Identification: Advances and Case Studies* (edited by R. K. Mehra and D. G. Lainiotis) New York: Academic Press, 27-96.
- AN, J.-Z., Z.-G. CHEN AND E. J. HANNAN (1982) Autocorrelation, Autoregression and Autoregressive Approximation. *Annals of Statistics* 10, 926-936.
- ÅSTRÖM, K. J. (1970) *Introduction to Stochastic Control Theory*. Academic Press: New York.
- BHANSALI, R. (1981) Effects of Not Knowing the Order of An Autoregressive Process. *J. Am. Statist. Assoc.* 76, 588-597.
- BOX, G. E. P. AND G. M. JENKINS (1976) *Time Series Analysis—Forecasting and Control*, 2nd ed. Holden-Day: San Francisco.
- BOX, G. E. P., S. C. HILLMER AND G. C. TIAO (1978) Analysis and Modeling of Seasonal Time Series. In *Seasonal Analysis of Economic Time Series* (edited by A. Zellner). U.S. Department of Commerce, Bureau of the Census: Washington, pp. 309-334.
- BOX, G. E. P. AND G. C. TIAO (1981) Modelling Multiple Time Series with Applications. *J. Am. Statist. Assoc.* 76, 802-816.
- CLEVELAND, W. P., A. P. DEMPSTER AND J. STITH (1980) Advances in Model-Based Seasonal Adjustment. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* 30-37.
- DAVIES, N., M. B. PATE AND M. G. FROST (1974) Maximum Autocorrelations for Moving Average Processes. *Biometrika* 61, 199-200.

- DUFOR, J.-M. (1982) A Specification Error Theorem for Predictions from Estimated Autoregressions. *IMS Bulletin*, 11, 197 and Département de science économique et Centre de recherche en développement économique, Université de Montréal, Cahier 8132.
- FINDLEY, D. F. (1983a) On the Unbiasedness Property of AIC for Exact or Approximating Multivariate ARMA Models. U.S. Bureau of the Census. SRD Research Report/RR-83/01. (Submitted for publication).
- FINDLEY, D. F. (1983b) On Using a Different Time Series Model for Each Forecasting Lead. Proceedings of the Business and Economics Section of the American Statistical Association, 528-531.
- GEORGIU, T. T. (1983) Partial Realization of Covariance Sequences. Center for Mathematical System Theory. University of Florida, Gainesville.
- GERSCH, W. AND G. KITAGAWA, (1982) The Prediction of Time Series with Trends and Seasonalities. *J. Bus. Econ. Statist.* 1, 253-264.
- GRAY, H. L., G. D. KELLY AND D. D. MCINTIRE (1978) A New Approach to ARMA Modeling. *Communications in Statistics* B7, 1-77.
- HANNAN, E. J. (1970) *Multiple Time Series*. Wiley: New York.
- HANNAN, E. J. (1980) The Estimation of the Order of an ARMA Process. *Annals of Statistics* 8, 1071-1081.
- HOSOYA, Y. AND M. TANIGUCHI (1982) A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes. *Annals of Statistics* 10, 132-153.
- KALMAN, E. (1982) Realization of Covariance Sequences. In *Toeplitz Centennial* (edited by I. G. Gohberg). Birkhäuser: Basel, pp. 135-164.
- LEWIS, R. AND G. REINSEL (1982) Prediction of Multivariate Time Series by Autoregressive Model Fitting. Technical Report No. 664. Department of Statistics, University of Wisconsin, Madison.
- LIUNG, L. AND P. CAINES (1979) Asymptotic Normality of Prediction Error Estimates for Approximate System Models. *Stochastics* 3, 29-45.
- NAKAMURA, H. AND H. AKAIKE (1981) Statistical Identification of Optimal Control of Supercritical Thermal Power Plants. *Automatica* 17, 143-155.
- OGATA, Y. (1980) Maximum Likelihood Estimates of Incorrect Markov Models and the Derivation of AIC. *Journal of Applied Probability* 17, 59-72.
- OTOMO, T., T. NAKAGAWA AND H. AKAIKE (1972) Statistical Approach to Computer Control of Cement Rotary Kilns. *Automatica* 8, 35-49.
- ROBINSON, E. A. AND H. WOLD (1963) Minimum Delay Structure of Least-Squares and Eo-Ipso Predicting Systems for Stationary Stochastic Processes. In *Time Series Analysis* (edited by M. Rosenblatt). Wiley: New York, pp. 192-196.
- SHIBATA, R. (1980) Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process. *Annals of Statistics* 8, 147-164.
- SHIBATA, R. (1981) An Optimal Autoregressive Spectral Estimate. *Annals of Statistics* 9, 300-306.
- TANIGUCHI, M. (1980) On the Selection of the Order of the Spectral Density Model for a Stationary Process. *Ann. Inst. Stat. Math.* 32A, 401-409.
- THIS WEEK'S CITATION CLASSIC (1981) *Current Contents* 51, 22.
- WHITTLE, P. (1963) On the Fitting of Multivariate Autoregressions and the Canonical Factorization of a Spectral Density Matrix. *Biometrika* 50, 129-134.
- WOODWARD, W. AND H. L. GRAY (1981) On the Relationship between the S-array and the Box-Jenkins Method of ARMA Model Identification. *J. Am. Statist. Assoc.* 76, 579-587.