SHIBATA, R. (1981). An optimal autoregressive spectral estimate. *Ann. Statist.* **9** 300–306.

SOLO, V. (1984). The exact likelihood for a multivariate ARMA model. *J. Multivariate Anal.* **15** 164–173.

TJØSTHEIM, D. (1986). Some doubly stochastic time series models. *J. Time Ser. Anal.* **7** 51–72.

TJØSTHEIM, D. and PAULSEN, J. (1983). Bias of some commonly-used time series estimates. *Biometrika,* **70** 389–399.

WHITTLE, P. (1963a). On the fitting of multivariate autoregressions and the approximate canonical factorisation of a spectral density matrix. *Biometrika* **50** 129–134.

Comment

R. J. Bhansali

I would like to congratulate Ted Hannan on a masterly survey of the current state of the art for fitting multivariate autoregressive moving average models, ARMA(p, q). Hannan is quite correct in emphasizing that there may not be a true (p, q) and thus a fitted ARMA model is at best thought of as an approximation to the generating structure of the observed time series. The question then arises: what are the properties of the order selected by minimizing AIC when viewed in this light rather than as an estimator of an underlying "true" order? The work of Shibata (1980) would suggest that the order selected by AIC is such that the one-step mean square error of prediction is minimized within the class of all order selection procedures. However, the more-than-one-step mean square error of prediction may not be minimized (see also Whittle, 1963b, page 36). Indeed, for autoregressive model fitting, Findley (1983) has advocated that a different order should be selected for each forecast lead, and he has suggested that a criterion introduced by Shibata (1980, page 163) may be used for this purpose. However, a justification for introducing this criterion has not been given. A related but different criterion is suggested by the work of Hannan and Rissanen (1982).

As has already been noted by Franke (1985a) and Chen (1985), at the second stage of the Hannan-Rissanen procedure for ARMA model selection, "autoregressive" estimates of the coefficients $b(u)$, say, in the moving average representation of a univariate stationary nondeterministic process $\{x_t\}$ are obtained as

$$\hat{b}_h(u) = \hat{c}_h(u)/\hat{c}_h(0), \quad u = 0, 1, \cdots,$$

where

$$\hat{c}_h(u) = \sum_{j=0}^{h} \hat{a}_h(j) R^{(T)}(u + j)$$

provides the corresponding estimator of the cross-covariance $c(u)$ say, between x_{t+u} and the linear innovations ε_t; the $\hat{a}_h(j)$ are the hth order "Yule-Walker" estimates of the autoregressive coefficients;

$$R^{(T)}(u) = T^{-1} \sum_{t=1}^{T-u} x_t x_{t+u}, \quad u = 0, 1, \cdots,$$

is a "positive definite" estimator of the covariance function of $\{x_t\}$; and $x_1, \cdots, x_T$ denotes an observed realization of length T of $\{x_t\}$.

Now, if the complete past $\{x_t, t \leq 0\}$, say of $\{x_t\}$ is known, the s step mean square error of prediction is given by

$$V(s) = \sigma^2 \sum_{j=0}^{s-1} b^2(j), \quad s = 1, 2, \cdots,$$

where $\sigma^2 = c(0)$ is the variance of ε_t.

Bhansali (1978) and Lewis and Reinsel (1985) consider the effect on the mean square error of prediction of estimating the prediction constants by fitting an autoregressive model of order h, when h is a function of T and tends to infinity simultaneously with it. It is clear from their work that if $o_p(h^{1/2}/T^{1/2})$ terms are ignored and certain additional regularity conditions are satisfied, the resulting mean square error of prediction may be approximated by

$$L(s) = V(s)\left(1 + \frac{h}{T}\right).$$

On adopting an argument similar to that used by Akaike (1970) for deriving his FPE criterion, which as discussed by Bhansali (1986) is closely related to the argument used for deriving AIC, one may consider

R. J. Bhansali is Senior Lecturer, Department of Statistics and Computational Mathematics, University of Liverpool, Liverpool L69 3BX, England.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science.* www.jstor.org

selecting a different h for each s by minimizing $L(s)$. An estimator of $V(s)$ is given by

$$\hat{V}_h(s) = \sum_{j=0}^{s-1} \hat{c}_h(j)\hat{b}_h(j).$$

However, as in the case $s = 1$, $\hat{V}_h(s)$ for any $s > 1$ provides a biased estimator of $V(s)$. Indeed, under regularity conditions similar to, but slightly stronger than, those specified in Theorem 4.2 of Bhansali (1986), an asymptotically unbiased estimator of $V(s)$, to terms of order h/T, is given by

$$\tilde{V}_h(s) = \hat{V}_h(s)\left(1 - \frac{s}{T}\right)^{-1}.$$

We may, therefore, estimate $L(s)$ by

$$\tilde{L}_h(s) = \hat{V}_h(s)\left(1 + \frac{s}{T}\right)\left(1 - \frac{s}{T}\right)^{-1},$$

and for each s, choose h so that $\tilde{L}_h(s)$ is minimized. Observe that for $s = 1$, $\tilde{L}_h(s)$ reduces to the FPE criterion of Akaike (1970).

We note that in the context of fitting ARMA models nonparametrically, i.e. without requiring that there is a "true" order (p_0, q_0), the question of how to estimate $V(s)$ is closely connected with the discussion in Section 3. My question to Hannan is: what, if any, are the *statistical* properties of the procedure due to Adamyan, Arov and Krein (1971), which has been described after equation (3.4). Of course, as discussed by Franke (1985b), in the Hannan-Rissanen procedure, an approximation to the transfer function $B(w)$, say of the $b(u)$, is constructed by solving the "Box-Jenkins" equations. This approximation is realizable if, e.g., $f(w)$ is known, and it is optimal from the point of view of entropy maximization. When should the approximation to $B(w)$ proposed by Adamyan, Arov and Krein be preferred to this approximation? Note that for $n = 1$, an "autoregressive" estimator of the function $g(w) = B(w)B^{-1}(w)$ is given by $\hat{g}_h(w) = \hat{A}_h(w)\{\hat{A}_h(w)\}^{-1}$, where $\hat{A}_h(w)$ is the transfer function of the $\hat{a}_h(j)$.

ADDITIONAL REFERENCES

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

BHANSALI, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain. *Ann. Statist.* **6** 224–231.

BHANSALI, R. J. (1986). A derivation of the information criteria for selecting autoregressive models. *Adv. Appl. Probab.* **18** 360–387.

CHEN, ZHAO-GUO. (1985). The asymptotic efficiency of a linear procedure of estimation for ARMA models. *J. Time Ser. Anal.* **6** 53–62.

FINDLEY, D. F. (1983). On using a different time series forecasting model for each forecast lead. Research Report CENSUS/SRD/RR-83/06, Statistical Res. Div., Bureau of the Census, Washington, D.C.

FRANKE, J. (1985b). ARMA processes have maximal entropy among time series with prescribed autocovariances and impulse responses. *Adv. Appl. Probab.* **17** 810–840.

LEWIS, R. and REINSEL, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.* **16** 393–411.

WHITTLE, P. (1963b). *Prediction and Regulation by Linear Least-Squares Methods.* Van Nostrand, Princeton, N. J.

Comment

David R. Brillinger

Throughout his whole career Ted Hannan has invariably put a finger on directions in which the field of time series later moved. We can anticipate that being the case with this present paper as well. State space representations and corresponding ARMA models seem destined to be in the forefront of time series research for many future years in much the same way that linear regression is so pervasive in traditional statistics research.

On a surprising number of occasions, techniques developed to handle time series problems have gone on to become central to statistics generally, so all statisticians may gain from paying some attention to the problems studied here. As examples of techniques going on to broader use, one may mention the work by Parzen and Rosenblatt on spectral density estimation that led to later work on probability density estimation and the work by Akaike on dimension estimation for autoregressive processes that led to techniques for dimension estimation in general parametric problems.

One thing this paper does is to make apparent the debt time series researchers owe to engineers. The engineers recognized basic problems and often developed effective solutions. Engineering contributions abound in the book by Kailath (1980). A particularly important one is the work by Schweppe (1965) and

David R. Brillinger is Professor of Statistics, Department of Statistics, University of California, Berkeley, California 94720.