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.

Mehra (1971) showing how the Kalman filter may be employed to determine the Gaussian likelihood in the case of processes with state space formulations (in particular, ARMAs). In connection with this work it somehow seems appropriate to set down the classic engineer's line,

Every engineering problem has a solution.

(I suppose statistical consultants have to have the same attitude.)

In his paper Professor Hannan reviews a number of problems and techniques at the heart of contemporary time series research: state space representations and approximations, dead and real time algorithms for fitting state space representations and he also presents various asymptotic results. Throughout, his concern is with the case of vector-valued time series. Often the techniques for the vector case are seen to be far from elementary extensions of the usual techniques for real-valued series. It is interesting to see the development of a host of procedures for Hankel matrices. Procedures for Toeplitz matrices have long been central to work with stationary time series.

Material in the paper is presented concerning the approximation of "known" transfer functions and concerning model fitting to empirical data. I will separate my comments correspondingly.

In my own work with state space models, I strive to have the substantive matter suggest a specific form for the model. In particular this typically leads to a variety of structural zeros (in the matrices, A, B, C of (2.3)) and to scientifically interpretable parameters. This mechanistic modeling approach further tends to reduce the problems associated with estimating dimensionality substantially. This last material is meant as a comment on my experience, not a question. A question I do have is the following: Pade approximants have long been proposed to handle the construction of rational approximations. (A recent reference is Lii (1985).) I wonder, are there any clear connections between those approximations and the ones of this paper? I wonder too about connections with the e_n-transformation employed in what has been called G-spectral estimation for somehow this work seems related. (One reference is Gray, Houston and Morgan (1978).) A comment to end with here is a repetition of Professor Hannan's that a crucial issue in the utility of the approximations developed is—is the norm employed "a suitable measure of closeness with which to work"? Presumably, eventually experience will show its inadequacies and strengths.

Professor Hannan, in the paper, emphasizes the technique of "Gaussian estimation," with a penalty term for dimension. Recently, I (Brillinger, 1985) have found the use of a criterion involving third order moments as well as second order, to lead to improved estimates. It is meant for series that are not Gaussian. Specifically, let I_s^T denote the second order periodogram at frequency $2\pi s/T$ and let f_s denote the corresponding power spectral value. Then the estimating equations of Gaussian estimation are

$$\sum_s (I_s^T - f_s) \frac{\partial f_s}{\partial \theta} \Big/ f_s^2 = 0.$$

Suppose now that $I_{r,s}^T$ denotes the third order periodogram and $f_{r,s}$ the corresponding bispectral value. Then, the method of bispectral fitting looks for solutions of the equations

$$\sum_s (I_s^T - f_s) \frac{\partial f_s}{\partial \theta} \Big/ f_s^2 + \frac{2\pi}{T} \sum_r \sum_s (I_{r,s}^T - f_{r,s}) \frac{\partial f_{r,s}}{\partial \theta} \Big/ f_r f_s f_{r+s} = 0.$$

In the work cited, these equations were solved by iteratively reweighted least squares for an AR(2) model. They come from setting down a likelihood based on both second and third order periodograms. The standard errors of the coefficients were found to get smaller and a further parameter, the skewness of the errors, could be estimated. The asymptotic distribution of these estimates may be set down via the general lemma given in Brillinger (1975).

Consider next the problem of estimating dimensionality. In the fitting of $M(p, q)$, for example, -2 log likelihood is corrected by $\nu(\theta)/T$ at one place, with $\nu(\theta) = (p + q)n^2$, by $\nu(\theta)\log T/T$ at another and by $\nu(\theta)C_T/T$ at a third. Taking note of the current wave of research on penalized maximum likelihood, with the determination of the multiplier of the "smoothness" term by cross-validation, I am led to propose a correction term $\lambda p + \mu q$ instead, with λ and μ determined by some form of cross-validation. (For example by extensions of the procedures in Beltrão and Bloomfield (1987) or Wahba and Wold (1975).) Intuitively, I feel that p is the more important parameter here and that the fitting procedure should allow of this possibility.

I noticed that no mention was made of prewhitening in the sections on statistical aspects. (For the nontime series readers, prewhitening involves the fitting early on of a crude model to the data, to make the observations analyzed more nearly independent.) It seems to me that prewhitening substantially simplifies the analysis on many occasions. I wondered what Professor Hannan's current attitude to prewhitening was? I also wondered if Professor Hannan had any preferences as to how to assess the reasonableness of models fit? Presumably he would compute some measures, or plots, based on residuals.

I end by remarking what a pleasure it has been for me to read and learn from the many papers that Ted

Hannan has written. With his retirement, I expect to see papers being generated even more rapidly.

ADDITIONAL REFERENCES

BELTRÃO, K. I. and BLOOMFIELD, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* **8** 21–38.

BRILLINGER, D. R. (1975). Statistical inference for stationary point processes. In *Stochastic Processes and Related Topics* (M. L. Puri, ed.) 55–100. Academic, New York.

BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and non-Gaussian series data. *Water Res. Bull.* **21** 743–756.

GRAY, H. L., HOUSTON, A. G. and MORGAN, F. W. (1978). On G-spectral estimation. In *Applied Time Series Analysis* (D. F. Findley, ed.) 39–138. Academic, New York.

LII, K.-S. (1985). Transfer function model order and parameter estimation. *J. Time Ser. Anal.* **6** 153–170.

MEHRA, R. K. (1971). Identification of stochastic linear dynamic systems using Kalman filter representation. *AIAA J.* **9** 28–31.

SCHWEPPE, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Trans. Inform. Theory* IT-**11** 61–70.

WAHBA, G. and WOLD, S. (1975). Periodic splines for spectral density estimation: the use of cross validation for determining the degree of smoothing. *Comm. Statist.* **4** 125–141.

Comment

R. Dahlhaus

This paper by Hannan is an excellent review of an important topic in time series analysis: the approximation of a nonparametric time series by a parametric rational one. The paper gives insight into the problems which arise and offers a variety of methods to tackle these problems.

To regard a fitted parametric model as an approximation to a nonparametric time series is clearly the correct point of view when dealing with parametric time series analysis. Many interesting papers dealing with related problems have been published in recent years and there is great need for further results to develop the theory sufficiently. The present paper is an important contribution to this goal.

It was therefore a pleasure for me to read this stimulating paper and to have been asked to comment on it. I will restrict my comments to the problem of estimation and in particular to the case of a one-dimensional process which is approximated by an autoregressive process.

1. THE APPROXIMATION CRITERION

Since the goal of the paper is the approximation of the transfer function, it seems to be natural to take a criterion which measures the quality of the approximation directly. Suppose the original series has an infinite autoregressive representation

$$\sum_{s=0}^{\infty} a_s Y_{t-s} = \varepsilon_t \quad \text{with } \varepsilon_t \text{ i.i.d.},$$
$$E\varepsilon_t = 0, \quad \operatorname{var}(\varepsilon_t) = \sigma^2,$$

and Y_t is approximated by an AR(k)-process whose coefficients are estimated from the data by $\hat{a}_1(k), \cdots, \hat{a}_k(k), \hat{\sigma}_k^2$. An appropriate approximation criterion then would be, for example,

$$(1) \qquad \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \frac{\sigma \hat{A}_k(\lambda)}{\hat{\sigma}_k A(\lambda)} - 1 \right|^2 d\lambda,$$

where

$$A(\lambda) = \sum_{s=0}^{\infty} a_s \exp(-i\lambda s)$$

and

$$\hat{A}_k(\lambda) = \sum_{s=0}^{k} \hat{a}_s(k) \exp(-i\lambda s).$$

Considering the relative difference between $\hat{A}_k(\lambda)$ and $A(\lambda)$ is natural, since for Yule-Walker estimates (1) is approximately equal to $\sigma^{-2} T(\hat{a}(k) - a(k))$ $R(\hat{a}(k) - a(k))$ with $R = \{\operatorname{cov}(Y_i, Y_j)\}_{i,j}$, which tends weakly to a χ_k^2 distribution (if the true process Y_t is also an AR(k)-process), while the limit behavior of the absolute difference would depend on $A(\lambda)$. The choice of the $\mathscr{L}_2$ norm seems to be mainly for calculational convenience. However, by using the approximation $\log(\sigma/\hat{\sigma}_k) \approx (\sigma/\hat{\sigma}_k) - 1$ (or by adding the penalty term $2[(\sigma/\hat{\sigma}_k) - 1 - \log(\sigma/\hat{\sigma}_k)]$ for the innovation variance estimate to the criterion (1)) one

R. Dahlhaus is Privatdozent, Department of Mathematics, University of Essen, Postfach 103 764, D-4300 Essen 1, West Germany.