Rejoinder

E. J. Hannan

A good deal of the discussion of the paper relates to the use of criteria such as AIC or BIC. The latter derives, via Gaussian assumptions, from (4.2). Of course the Gaussian assumptions are not necessary, as Jorma Rissanen points out. However it is not easy to prescribe a probability law for a stochastic process. That could be done by taking the innovations as independent but that also is a fiction. In addition, the statistical analysis often has a special purpose. The result of these two difficulties is that methods that are rather *ad hoc*, but still general and effective, will always be important. Fourier methods are an example of this, to some extent, as also is the technique, involving the use of the third order spectrum, in David Brillinger's comments. Raj Bhansali also deals with a special problem, namely s step ahead prediction, $s > 1$, which could not easily be treated via (4.2). Of course one cannot deal with very special techniques in a general survey, even if one had the wit to think of them.

Ritei Shibata and I are in agreement, I think, about the above and the relation of the purpose of the analysis to the method used. I cannot quite see his objection to Rissanen's encoding argument which, in a sense, treats data and parameters in the same way, since Ritei Shibata favors a Bayesian argument that does much the same. Rissanen's argument does not require finiteness of the true order. Indeed the notion of true order is rejected. Rissanen would use a prior for the autoregressive order, for example allotting $c2^{-\log^* h}$ to order h, $\log^* h = \log h + \log\log h + \cdots$, up to the last positive term. Of course this series converges but very slowly and it is not asserted that the truth lies in the model set. The results Ritei Shibata quotes about order of consistency relate to autoregressions. It is not clear to me that the boot cannot be on the other foot for ARMA model fitting. After all if there was a true finite order (or something very near to that) some overfitting that AIC might induce could result in false, nearly matching, poles and zeros. These could be troublesome if, say, pole placement was the end purpose of the statistical analysis.

It must be agreed that the structure theory in Section 2 of the paper has been little used in statistical practice. One reason for this may be a lack of familiarity with the theory and this the paper, partly, sought to redress. Another reason would be a lack of ready access to algorithms and programs. Once the dimension, n, of the output is increased the "curse of dimensionality" has its effects, even for an AR. Of course determining Kronecker indices allows the dimension of the parameter space to vary over a fine grid of integer values. For example for $n = 3$ all dimensions occur except 1, 2, 3, 7. Of course there is an arbitrary quality about Kronecker indices. One way to exorcise the curse of dimensionality is to use special knowledge about the elements of A, B, C in (2.3) so that only for $S = I$ will the change of basis, $x(t) \to Sx(t)$, in the state space leave A, B, C in the special form. Such prior knowledge may often be available as David Brillinger points out. However, there will be cases when prior knowledge is too vague for this. A related phenomenon to the use of prior constraints is that for $n = 1$ the systems are listed with $p \equiv q$. One may feel that $q \ll p$ will do. A way to handle this is to find an estimate, $\hat{d}$, of $d = \max(p, q)$ and then to examine, using AIC or BIC, pairs $(\hat{d}, q)$ $q < \hat{d}$ (or $(p, \hat{d})$, $p < \hat{d}$, for that matter). If T is large, when $\hat{d}$ will also be relatively big, this will be simpler than looking at (p, q), $p \le \hat{d}$, $q \le \hat{d}$. David Brillinger wants a heavier penalty on large q. One should perhaps be careful not to allow the investigator too much leeway to indulge his prejudices. (Referring to a related phenomenon, there is some evidence that careless use of rules for rejecting outliers has led to errors.) The same kind of objection can be raised to a proliferation of criteria, of which Rainer Dahlhaus introduces another in his (1). This criterion has appeal, as does also $m_h(T)$ in Section 5 of the paper. However, in relation to a special purpose both might do badly. (See the discussion below in $m_h(T)$ in Section 5.) AIC, BIC have the virtue that they (or their generalizations such as (4.2)) have a sound general principle behind them.

The idea of data reduction as a central statistical aim is an old one and underlies Rissanen's theory. It is not to be accepted uncritically but it also should not be rejected out of hand because it differs from received statistical theory. The consistency results in Section 5 are only of suggestive value, as are all theorems since reality is so complex. Such results are useful also in the development of further theory, albeit only of suggestive value. The same kind of theory, in any case, leads to (5.10) which is not constrained in its applications to a case of a true ARMA model.

I do not agree with Rainer Dahlhaus' statement about $\Phi(j)$ and $\Phi_h(j)$. Two situations can be contrasted. One is the fitting of an ARMA model of fixed order to nonARMA data. Then, θ being the parameter vector, one may show that $\hat{\theta} \to \Theta_0$, in the sense that any subsequence has a sub-subsequence converging to