ACKNOWLEDGMENTS

This research was supported by ONR Contract N00014-91-J-1074. I am grateful to Chris Fraley for writing the Fracdiff software and for helpful comments, and to my long-time collaborator John Haslett for many useful discussions.

Comment

Richard L. Smith

Jan Beran has written an excellent and timely review of a topic that is gaining increasing attention in a whole variety of fields. As his review makes clear, the origins of the subject go back a long way and were rooted in practical problems in several fields. However, it is only in recent years, stimulated by the development of the fractional ARIMA model, that the subject has started to receive widespread attention among statisticians. Beran does a superb job of bringing together the extensive results that now exist on the effects of long-range dependence on a whole range of statistical inferences. Nevertheless, I suspect it is in the identification and estimation of long-range models themselves that readers will take the greatest interest, and it is here that I concentrate my comments.

A common feature of long-range models is that the spectral density $f(x)$ satisfies the relation

$$f(\omega) \sim b\omega^{1-2H}, \quad \omega \to 0. \tag{1}$$

Beran's equation (6) is a slight generalization of this, replacing the constant b by a slowly varying function, but for most purposes (1) suffices. One feature of many of the results about the effect of long-range dependence, such as Beran's equation (8), is that they depend on the spectral density only through the constants b and H. In fact, (8) itself depends only on H, but many related results depend also on the scaling constant b. For this reason, it is of interest to look for direct estimators of b and H, rather than assume some parametric model such as fractional ARIMA. I have been particularly interested in estimators based on the periodogram, which are among those reviewed in Section 2.4. If $I_n(\omega)$ denotes the periodogram at frequency ω based on n observations, then it is "well-known" that the sampling distribution of $I_n(\omega_j)$ at the Fourier frequencies $\omega_j = 2\pi j/n$ for $0 \le j < n/2$ is approximately that of independent exponential random variables with means $f(\omega_j)$. If we assume $f(\omega) = b\omega^{1-2H}$ then this suggests that $1 - 2H$ could be estimated as the slope of a linear regression of $\log I_n(\omega_j)$ on $\log \omega_j$. This idea has been suggested by a number of authors, in particular Geweke and Porter-Hudak (1983). Two refinements of Geweke and Porter-Hudak seem desirable:

a. Geweke and Porter-Hudak used least squares regression of log periodogram ordinate on log frequency. In contrast, since the asymptotic distribution of $I_n(\omega_j)$ is exponential, a regression of $\log I_n(\omega_j)$ based on errors from the Gumbel distribution function $1 - \exp(-e^x)$ would seem preferable. I call this the maximum likelihood (ML) approach, in contrast to Geweke and Porter-Hudak's least squares (LS) approach.
b. In addition, it is becoming increasingly clear that it is necessary to restrict the range of frequencies used in the regression, say to $n_0 \le j \le n_1$ where $1 < n_0 < n_1 << n/2$. At the lower end, the difficulty is that the above-mentioned "well-known" properties of the periodogram apparently break down for very low frequencies in the case of a long-range model (see, e.g., Künsch, 1987; Haslett and Raftery, 1989). At the upper end, the problem arises from the fact that (1) is only an asymptotic relation, not an identity, so attention must be restricted to small ω. A more formal argument along these lines was presented in my discussion of Haslett and Raftery (1989).

It seems to me that Graf's HUB00 and HUBINC estimators deal with problem (a), albeit in a quite different way from the ML approach being suggested here, but do not contain anything that corresponds directly to the selection of n_0 and n_1. In view of this, I am somewhat doubtful about the theoretical justification of these estimators.

The rest of this discussion concerns three examples, two of them taken from Beran's paper, which illustrate the importance of appropriate selection of n_0 and n_1 in this approach.

The first of these is the Nile data. Beran's Figure 3 plots the periodogram in log-log coordinates. It can be seen that the plot is decreasing at an approximately

Richard L. Smith is Professor, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27599-3260.