For Encyclopaedia of Environmetrics

LONG RANGE DEPENDENCE

by

P.M. Robinson[1]

Long range dependence, or *long memory*, usually refers to a strong correlation between distant observations in a time series. Evidence of long range dependence has been found in diverse fields, such as geophysics, agriculture, chemistry, economics and finance.

An early empirical investigation of river discharges provided the initial impetus for serious theoretical study; it has continued to influence the development of the subject. A long historical series of annual flood levels of the River Nile, recorded at the Roda Gorge at Cairo, suggested evidence of dependence over long intervals of time, with stretches when floods are high, and others when they are low; on the other hand, there was no regularity in their occurrence or duration, so that the series did not exhibit periodicity. For discrete, equally-spaced observations, letting X_t denote the level at time t and $\overline{X} = n^{-1}\sum_{t=1}^{n} X_t$ the sample mean based on a sample of size n, the *adjusted rescaled range statistic,*

$$R/S = \frac{\max\limits_{1\le s\le n} \sum\limits_{t=1}^{s}(X_t - \overline{X}) - \min\limits_{1\le s\le n} \sum\limits_{t=1}^{s}(X_t - \overline{X})}{\left\{\frac{1}{n}\sum\limits_{t=1}^{n}(X_t - \overline{X})^2\right\}^{\frac{1}{2}}}, \tag{1}$$

was found empirically to behave like n^H, $\frac{1}{2} < H < 1$; see Hurst (1951). However, if X_t is a sequence of independent (Gaussian) random variables, it can be shown theoretically that R/S behaves like $n^{\frac{1}{2}}$.

[1]Research supported by a Leverhulme Trust Personal Research Professorship.

The parameter H, known as the *Hurst coefficient*, arises in a time series model for X_t that can explain this behaviour. Let X_t, $t = 0, \pm 1, ...$, be a *stationary Gaussian process*, so that a complete description of its probabilistic structure is provided by specifying its mean μ and autocovariance function $\gamma_s = cov(X_t, X_{t+s})$, neither of which depend on t. If X_t is a *fractional noise*, having autocovariance

$$\gamma_s = \frac{\gamma_0}{2}\left\{|s+1|^{2H} - 2|s|^{2H} + |s-1|^{2H}\right\}, s = 0, \pm 1, ..., \tag{2}$$

then the R/S statistic (1) exhibits the n^H power law behaviour described, for large n. Moreover, for $H = \frac{1}{2}$, it follows from (2) that $\gamma_s = 0$, for all $s \neq 0$, so that X_t is an independent sequence, whereas for $\frac{1}{2} < H < 1$ we have

$$\gamma_s \sim \gamma_0 H(2H-1)|s|^{2H-2}, \text{ as } |s| \to \infty, \tag{3}$$

where "$\sim$" means that the ratio of left- and right-hand sides tends to one. The asymptotic behaviour in (3) indicates that the autocovariance decreases with long lags, but that it does so very slowly indeed, so that

$$\sum_{s=-\infty}^{\infty} \gamma_s = \infty. \tag{4}$$

As the earlier discussion indicates, one might estimate H by $\log(R/S)/\log n$.

The model (2) is connected with the interesting physical property of *self-similarity*. An underlying continuous-time process $Y(t)$ is called *self-similar* with parameter H if $Y(at)$ and $a^H Y(t)$ have identical finite-dimensional distributions for all $a > 0$; thus the distributions have the same shape irrespective of the frequency of sampling. If $Y(t)$ also has stationary increments, then $X_t = Y(t) - Y(t-1)$ has autocovariance function (2).

We can think of (4) as a *time domain* long range dependence property. An alternative, closely related, one is formulated in the *frequency domain*. Suppose that the stationary series X_t has a *spectral density*, denoted $f(\lambda)$, $\pi < \lambda \leq \pi$, so that we can write

$$\gamma_s = \int_{-\pi}^{\pi} f(\lambda) \cos s\lambda d\lambda, \ s = 0, \pm 1, ... \ .$$

Then the non-summability condition (4) is equivalent to an unbounded spectral density at zero frequency,

$$f(0) = \infty.$$

This is true if, for example, with $\frac{1}{2} < H < 1$,

$$f(\lambda) \sim C\lambda^{1-2H}, \text{ as } \lambda \to 0+, \tag{5}$$

for a positive, finite C.

A statistic that provides some indication of the magnitude of $f(\lambda)$ is the *periodogram*,

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} (X_t - \overline{X}) e^{it\lambda} \right|^2$$

In time series from diverse applications, plots of $I(\lambda)$ (or a smoothed version which can provide a more reliable estimate of a finite $f(\lambda)$) can appear consistent with the power law behaviour near frequency zero indicated in (5). Mathematically, the latter property often co-exists with (cf (3))

$$\gamma_s \sim c\,|s|^{2H-2}, \text{ as } s \to \infty, \tag{6}$$

for some finite, positive c. Both (5) and (6) indicate that the degree of dependence varies directly with H.

Recent research has focussed on models that are more naturally expressed in terms of the *fractional differencing* parameter d, which relates very simply to H,

$$d = H - \frac{1}{2}. \tag{7}$$

Letting L denote the lag operator, so that $LX_t = X_{t-1}$, we have formally the binomial expansion

$$(1 - L)^d = 1 - dL + \frac{d(d-1)}{2!} L^2 - \dots .$$

If the 'd-th fractional difference' of X_t is a sequence of uncorrelated random variables, V_t, with zero mean and common variance, so

$$(1 - L)^d X_t = V_t, \tag{8}$$

then, for $0 < d < \frac{1}{2}$, X_t has spectral density $f(\lambda)$ satisfying (5), while also (6) holds, with the identity (7); see Adenstedt (1974). Moreover, (5) and (6) are also satisfied if V_t is, more generally, a correlated, stationary sequence that is short range dependent, asymptotically having spectral density that is everywhere continuous, and thus bounded. This is the case when V_t is an *autoregressive moving average* sequence, so that

$$a(L)V_t = b(L)U_t, \tag{9}$$

where the U_t are uncorrelated with zero mean and common variance and $a(L) = 1 - \sum_{j=1}^{p} a_j L^j$ and $b(L) = 1 - \sum_{j=1}^{q} b_j L^j$ are polynomials of finite degrees, p and q, with all zeros outside the unit circle. The requirement on $a(L)$ entails stationarity, while that on $b(L)$ entails invertibility and identifiability. The resulting *fractionally integrated autoregressive moving average* model for X_t, obtained by combining (8) and (9),

$$(1 - L)^d a(L)X_t = b(L)U_t, \tag{10}$$

constitutes the most popular parameterization of long range dependence, though alternatives, besides (2), have been advanced.

In practice d and other parameters in (10) are unknown, but can be estimated by an approximation to Gaussian maximum likelihood, known as *Whittle estimation.* The spectral density of X_t given by (10) has the form

$$f(\lambda; \underset{\sim}{\theta}) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d} \left|\frac{b(e^{i\lambda})}{a(e^{i\lambda})}\right|^2, \pi < \lambda \leq \pi,$$

where $\underset{\sim}{\theta}$ denotes the vector of unknown parameters, $\underset{\sim}{\theta} = (d, a_1, ..., a_p, b_1, ..., b_q, \sigma^2)'$, where $\sigma^2 = V(U_t)$. A Whittle objective function is

$$L(\underset{\sim}{\theta}) = \sum_{j=1}^{n-1} \left\{ \log f(\lambda_j; \underset{\sim}{\theta}) + \frac{I(\lambda_j)}{f(\lambda_j; \underset{\sim}{\theta})} \right\}, \tag{11}$$

where $\lambda_j = 2\pi j/n$; note that the mean-correction in $I(\lambda_j)$ is redundant for $j = 1, ..., n - 1$. The periodogram $I(\lambda_j)$ can be rapidly computed by the *fast Fourier transform*, even when n is quite large. We estimate $\underset{\sim}{\theta}$ by the value $\underset{\sim}{\widehat{\theta}}$ minimizing $L(\underset{\sim}{\theta})$; in practice no closed-form solution exists, and numerical methods are needed.

For the purpose of statistical inference on $\underset{\sim}{\theta}$, it is known that, for Gaussian X_t, $\underset{\sim}{\widehat{\theta}}$ can be treated as approximately normally distributed with mean $\underset{\sim}{\theta}$ and covariance matrix

$$2\left[\sum_{j=1}^{n-1}\left\{\frac{\partial \log f(\lambda_j;\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right\}\left\{\frac{\partial \log f(\lambda_j;\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right\}'\right]^{-1}, \tag{12}$$

for large n; see Fox and Taqqu (1986). The same large-sample properties often hold even when X_t is non-Gaussian, though the approximate covariance matrix may involve an additional term besides (12), depending on fourth cumulants. It is not required that X_t have a known mean, the omission of the frequency for $j = 0$ (and equally, by periodicity, that for $j = n$) automatically corresponding to a mean-correction. Alternative methods of estimating $\underset{\sim}{\theta}$ are available but they will be less efficient than Whittle estimation when X_t is Gaussian, while possibly lacking some of the advantages it continues to enjoy even when X_t is non-Gaussian.

In practice the autoregressive and moving average orders p and q in (10) are likely to be unknown. It is possible to adapt methods for choosing p and q, based on the observed data, that have been derived in the short range dependent autoregressive moving average context (9). However, there is still a danger of under- or over-specifying p and q, which can lead to invalidation of the statistical properties described above. In studies of long range dependence, d is the parameter of greatest interest, but misspecification of $a(L)$ and $b(L)$ which essentially describe the short range dependent component of X_t, can seriously bias the estimation of d.

Semiparametric estimation of d is a way around this difficulty. Recalling the property (5), which holds over many models besides (10), under the identity (7), we consider a *local Whittle* estimate which rests only on the approximation of $f(\lambda)$ near frequency zero (cf. (11))

$$L(d, C) = \sum_{j=1}^{m}\left\{\log(C\lambda_j^{-2d}) + \frac{I(\lambda_j)}{C\lambda_j^{-2d}}\right\}, \tag{13}$$

where m is an integer which is much smaller than n; see Künsch (1987). We estimate C and d by (numerically) minimizing (13). Under mild regularity conditions, for large m and n we can treat the estimate of d as normal with mean d and variance $1/4m$;

see Robinson (1995). This argument requires m to be of smaller order than n, so that in view of (12) the parametric estimate described previously is the more precise. It is inadvisable to choose m too large as bias can then result, especially if the spectral density also contains peaks at non-zero frequencies. However, the longer the series length n, the larger we can choose m because (13) involves frequencies up to $2\pi m/n$, so that in very long series the extra robustness gained by the semiparametric approach may be worthwhile. Automatic, data-dependent, methods for choosing m, which balance the bias and imprecision that would be incurred by respectively choosing too large and too small a value, are available. An alternate method of estimating d that also uses only low frequencies, *log periodogram regression*, is longer established and more popular, but less efficient than the local Whittle estimate.

References

Adenstedt, R.K. (1974). On large sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 2, 1095-1107.

Fox, R. and Taqqu, M.S. (1986). Large sample properties of parameter estimates of strongly dependent stationary Gaussian time series. *Annals of Statistics* 14, 517-532.

Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 1, 519-543.

Künsch, H.R. (1987). Statistical aspects of self-similar processes. *Proceedings of the First World Congress of the Bernoulli Society* (Yu Prohorov and V. Sazonov, eds.) 1, 67-75, VNU Science Press, Utrecht.

Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 5, 1630-1661.