Long-Range Dependence

Peter. M. Robinson

Abstract

Abstract: Developments in research on stationary and nonstationary time series with long range dependence are reviewed.

London School of Economics, UK

Keywords: *long-range dependence, long memory, parametric inference, semiparametric inference, statistical properties*

The phenomenon of *long-range dependence,* or *long memory*, is a feature of statistical **time series.** It entails persistingly strong autocorrelation

---

Based in part on the article "Long-range dependence" by P. M. Robinson, which appeared in the *Encyclopedia of Environmetrics.*

This article was originally published online in 2013 in Encyclopedia of Environmetrics, © John Wiley & Sons, Ltd and republished in Wiley StatsRef: Statistics Reference Online, 2014.

between distant observations in a time series. Over the years, evidence of long-range dependence has been found in various substantive fields, including **geophysics**, **agriculture**, **chemistry**, **economics**, and **finance**.

The first of these provided the initial impetus for serious theoretical study, in particyular an empirical investigation of river discharges along the River Nile. A long historical series of annual flood levels, recorded at the Roda Gorge at Cairo, suggested the 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 $\hat{X} = n^{-1} \sum_{t=1}^{n} X_t$ the sample mean based on a sample of size n, the *adjusted rescaled range statistic*

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

was found empirically to behave like n^H, $1/2 < H < 1$; see Ref. [1]. However, if X_t is a sequence of independent (Gaussian) random variables (*See* **Normal Distribution**), it can be shown theoretically that R/S increases with n at the rate $n^{1/2}$.

The parameter H, known as the *Hurst coefficient*, arises in a time series model for X_t that can explain this behavior. Let $X_t, t = 0, \pm 1, \ldots$, be a ***stationary Gaussian process***, so that a complete description of its probabilistic structure is provided by specifying its mean μ and autocovariance function $\gamma_s = \mathrm{cov}(X_t, X_{t+s})$, neither of which depend on t. *Fractional noise* is the earliest form, and one of the simplest forms, of long-range dependent process. If X_t is fractional noise it has autocovariance

$$\begin{aligned} &\gamma_s = \tfrac{\gamma_0}{2} \left\{ |s+1|^{2H} - 2|s|^{2H} + |s-1|^{2H} \right\}, \\ &s = 0, \pm 1, \ldots \end{aligned} \tag{2}$$

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

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

where "$\sim$" means that the ratio of left- and right-hand sides tends to 1. The asymptotic behavior 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

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

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

$$f(0) = \infty$$

This is true if, for example, with $1/2 < H < 1$,

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

for a positive, finite constant C.

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

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

where i here denotes the imaginary unit, so $i^2 = -1$. 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 behavior near frequency zero indicated in (5). Mathematically, the latter property often coexists with [cf. (3)]

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

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

Recent research has focused 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 $LX_t = X_{t-1}$, formally we have the binomial expansion

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

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 $0 < d < 1/2, X_t$ has spectral density $f(\lambda)$ satisfying (5), while also (6) holds, with the identity (7); see Ref. [2]. 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;** this is known as *Whittle estimation.* The spectral density of X_t given by (10) has the form

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

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

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

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

For the purpose of statistical inference on θ, it has been shown that, for Gaussian X_t, $\hat{\theta}$ can be treated as approximately normally distributed with mean θ and covariance matrix

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

for large n; see Ref. [3]. 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 θ 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 order p and moving average order 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 overspecifying 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 Ref. [4]. 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 Ref. [5]. 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 nonzero 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.

More recently, there has been interest in *nonstationary* processes, where d in (8) satisfies $d \geq \frac{1}{2}$ (and a suitable initial condition on V_t for $t \leq 0$ is imposed). Also, *negatively dependent* processes have been considered, where $d < 0$. The class of such processes, indexed by real-valued d, is vast. Estimation of d, and other parameters, has been considered, using parametric and semiparametric methods that relate to and extend those described above. For example, Hualde and Robinson [6] show that estimates of d and other parameters that make no prior assumptions on the approximate location of d (such as the interval $(0, \frac{1}{2})$), have just as good statistical properties as ones that correctly impose such prior restrictions.

We have so far presumed that only a single time series is of interest, so that X_t is a scalar. But frequently we wish to consider a number of jointly dependent series X_t, and it is usually appropriate to model these jointly. Extensions of the modelling and estimation procedures described above have been developed, see again for example Hualde and Robinson [6], with a view also to examine the phenomenon of *cointegration*, where related series appear to move together and there exist one or more linear relations which have weaker dependence than the original series. An issue with vector valued time series, as often is the case with multivariate data, is the danger of *curse of dimensionality*, involving a proliferation of parameters and correspondingly loss in precision of estimation. This issue has been further highlighted in the recent '*big data'* literature, which does not necessarily require the dimensionality of X_t to remain fixed as n increases, but can even allow it to exceed n.

Some other areas of development are worthy of mention. One considers long-range dependence in *panel data* or *longitudinal data,* especially in the area of economics, see for example Robinson and Velasco (2015). Another applies notions of long-range dependence to *spatial data*, which arise in many fields, indeed this was considered very early on by Fairfield Smith (1938) in an agricultural context. Also there are definitions of long-range dependence that are not confined to properties of second moments, but refer to other distributional properties, see for example Robinson (2011), which also concerns spatial data.

References

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

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

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

[4] Künsch, H.R. (1987). Statistical aspects of self-similar processes, in Proceedings of the First World Congress of the Bernoulli Society, Vol. 1, Y. Prohorov & V. Sazonov, eds, VNU Science Press, Utrecht, pp. 67 – 75.

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

[6] Hualde, J. & Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models, *Annals of Statistics* **39**, 3152-3181..

[7] Robinson, P.M. & Velasco, C. (2015). Efficient inference on fractionally integrated panel data models with fixed effects, *Journal of Econometrics* **185**, 435-452.

[8] Fairfeld Smith, H. (1938). An empirical law describing heterogeneity in the yields of agricultural crops, *Journal of Agricultural Science.* **28,** 1-23.

[9] Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data, *Journal of Econometrics* **165,** 5-19.