1 R/S Method

This is one of the better known methods. It is discussed in detail in Mandelbrot and Wallis [4], Mandelbrot [2] and Mandelbrot and Taqqu [3]. For a time series, $X = \{X_i, i \geq 1\}$, with partial sum $Y(n) = \sum_{i=1}^{n} X_i$, and sample variance

$$S^2(n) := (1/n) \sum_{i=1}^{n} {X_i}^2 - (1/n)^2 Y(n)^2,$$

the R/S statistic, or the *rescaled adjusted range*, is given by :

$$\frac{R}{S}(n) := \frac{1}{S(n)} \left[\max_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) - \min_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) \right]. \tag{1.1}$$

For fractional Gaussian noise or fractional ARIMA,

$$E[R/S(n)] \sim C_H n^H, \tag{1.2}$$

as $n \to \infty$, where C_H is another positive, finite constant not dependent on n.

To determine H using the R/S statistic, proceed as follows. For a time series of length N, subdivide the series into K blocks, each of size N/K. Then, for each lag n, compute $R(k_i, n)/S(k_i, n)$, starting at points $k_i = iN/K + 1$, $i = 1, 2, \ldots$, such that $k_i + n \leq N$. For values of n smaller than N/K, one gets K different estimates of $R(n)/S(n)$. For values of n approaching N, one gets fewer values, as few as 1 when $n \geq N - N/K$.

Choosing logarithmically spaced values of n, plot $\log[R(k_i, n)/S(k_i, n)]$ versus $\log n$ and get, for each n, several points on the plot. This plot is sometimes called the pox plot for the R/S statistic. The parameter H can be estimated by fitting a line to the points in the pox plot. Since any short-range dependence in the series typically results in a transient zone at the low end of the plot, set a cut-off point, and do not use the low end of the plot for the purposes of estimating H. Usually, the very high end of the plot is not used as well, because there are too few points on the plot at the high end to make reliable estimates. The values of n that lie between the lower and higher cut-off points are used to estimate H.

The R/S statistic has recently been studied in detailed simulations studies by Taqqu, Teverovsky and Willinger [6] and Taqqu and Teverovsky [5].

2 Modified Lo's R/S

Define R as above, but now do not consider multiple lags, but only the lag $n = N$, the length of the series. However, instead of simply using the sample standard deviation, S, to normalize R, Lo suggests using a weighted sum of autocovariances, namely,

$$S_q(N) := \sqrt{\frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X}_N)^2 + \frac{2}{N} \sum_{j=1}^{q} \omega_j(q) \left[\sum_{i=j+1}^{N} (X_i - \bar{X}_N)(X_{i-j} - \bar{X}_N) \right]} \tag{2.1}$$

where $\bar{X}_N$ denotes the sample mean of the time series, and

$$\omega_j(q) := 1 - \frac{j}{q+1}, \quad q < N.$$

Observe that $S_0(N)$ is the sample variance $S(N)$ defined above. Another way of representing $S_q(N)$ is:

$$S_q(N) = \sqrt{S^2 + 2\sum_{j=1}^{q} \omega_j(q)\hat{\gamma}_j}, \tag{2.2}$$

where $\hat{\gamma}_j$ are the sample autocovariances. DThe weights ω are chosen so that S_q is the sample variance of an aggregated or averaged series. We now compute the statistic

$$V_q(N) = N^{-1/2}R(N)/S_q(N). \tag{2.3}$$

If a series has no long-range dependence, Lo shows that given the right choice of q, the distribution of $V_q(N)$ is asymptotic to

$$W_1 = \max_{0 \le t \le 1} W_0(t) - \min_{0 \le t \le 1} W_0(t) \tag{2.4}$$

where W_0 is the standard Brownian bridge. Thus, confidence intervals can be computed,

$$P\{W_1 \in [.809, 1.862]\} = 0.95. \tag{2.5}$$

Thus, Lo uses the interval $[0.809, 1.862]$ as the 95% (asymptotic) acceptance region for the test

$$H_0 = \{\text{no long-range dependence}\}$$

versus the alternative $H_1 = \{\text{there is long-range dependence, i.e. } 1/2 < H < 1\}$.

The main innovation here is, of course, using $S_q(N)$ in the denominator, and Lo's motivation for using that statistic instead of S is as follows. If $q = 0$, the "range" of the time series, R, is normalized by the sample standard deviation, S. If the series is short-range dependent, e.g. if the process is actually AR(1) (no long-range component), then as $N \to \infty$, the statistic $V_0(N)$ will converge to a constant times W_1, where the constant depends on the short-range dependence structure. This can cause a problem if one is not using the graphical method. Because of the presence of this constant, the value of the statistic could easily fall outside the confidence interval (2.5), and then a false conclusion of long-range dependence would be drawn. On the other hand, given an appropriate choice of q, $S_q(N)$ will compensate for the "extra" short-range dependencies and $V_q(N)$ will then converge to W_1 for large N. If a time series has both short- and long-range dependence, Lo's statistic V_q should typically fall outside the confidence interval (2.5), implying correctly that long-range dependence is present. Lo's results are asymptotic, in that they assume $N \to \infty$ and $q = q(N) \to \infty$. In practice, the sample size N is finite and the question naturally arises as to whether there is a right choice of $q(N)$. The choice of $q(N)$ influences both the actual size of the test ($P(\text{reject } H_0 \mid H_0)$) and its power ($P(\text{accept } H_0 \mid H_1)$)

In addition, Lo's method can only indicate whether long-range dependence is present or not, without yielding an estimate of H. While the graphical method is not always very accurate in its estimates, it does usually provide a rough value for H.

We use the following modification of Lo's method: use V_q at a wide range of values of q, and then plot the estimates. Lo's conclusions seem to indicate that there will be some sort of short-range behavior for small q's and then a period of stability for larger q's, where only the long-range effects would be evident.

In practice, the plots do not actually behave in this fashion, but, in general, simply decrease with increasing q.

For reference see [1] and [7].

References

[1] A. W. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.

[2] B. B. Mandelbrot. Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31:271–285, 1975.

[3] B. B. Mandelbrot and M. S. Taqqu. Robust R/S analysis of long-run serial correlation. In *Proceedings of the 42nd Session of the International Statistical Institute*, Manila, 1979. Bulletin of the International Statistical Institute. Vol.48, Book 2, pp. 69-104.

[4] B .B. Mandelbrot and J. R. Wallis. Computer experiments with fractional Gaussian noises, Parts 1,2,3˙ *Water Resources Research*, 5:228–267, 1969.

[5] M. S. Taqqu and V. Teverovsky. Estimating long-range dependence in finite and infinite variance series. In R. Adler, R. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy-Tailed Distributions*, Boston, 1996. Birkhäuser. To appear in 1997.

[6] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(4):785–798, 1995. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

[7] V. Teverovsky and M. S. Taqqu. Analysis of lo's r/s method for detecting long-range dependence. Preprint, 1996.