1 Estimator

The ratio estimator is based on the Variance of Residuals method for estimating long-range dependence. The Variance of Residuals method, introduced by Peng *et. al.* [1] and further studied by Taqqu, Teverovsky and Willinger [3] and Taqqu and Teverovsky [2], works in the following way.

First, the original time series of length N is broken up into blocks of size m. Then, within each of the blocks, the partial sums of the series are calculated. The partial sums within a given block are called $Y(i), i = 1, 2, ..., m$. A least-squares line is fitted to the $Y(i)$ and the sample variance of the residuals is computed. This procedure is repeated for each of the blocks, resulting in approximately N/m estimates of the variance of the residuals. It was shown in [2] that, for finite variance long-range dependent series with scaling parameter H, the variance of the residuals scales like m^{2H} as m tends to infinity.

To obtain an estimate of the variance of the residuals, one can take the median of the N/m estimates. The median behaves like the original estimates themselves, at least in terms of scaling, so that the median of the N/m estimates of the variance of the residuals also scales like m^{2H}. Alternatively, one can take the average of the N/m estimates. Since these sample variances involve the squares of the original time series, the behavior of their average is somewhat counterintuitive when infinite variance series are involved. The square of a random variable with the heavy tail parameter α behaves like a random variable with the heavy tail parameter $\alpha/2 < 1$. Thus, taking the sample average over N/m realizations results in an additional scaling of $(m/N)^{1-2/\alpha}$. Since N does not change within a single series, we only concentrate on the scaling with respect to m. Since the variances themselves scale like $m^{2H} = m^{2d+2/\alpha}$ and since the averaging introduces an additional factor of $m^{1-2/\alpha}$, the sample average of the variance of residuals will scale like m^{2d+1}. A log-log plot versus m should follow a straight line with a slope of $2H$ or $2d + 1$ for the median and average respectively.

To obtain an estimate of α, we simply take the ratio of the median and the average for each block size m. This ratio should then scale like $m^{2/\alpha-1}$. It is this ratio which is plotted versus m on a log-log plot. A line is fitted to the plot between certain cut-off values of m, since all of the asymptotic results require both m and N/m to be large. For stable or Gaussian series, the lower cut-off can be quite small ($m = 5$) since these series scaled well even for small block sizes. For other series, such as Pareto, Lognormal, or Exponential, larger m's are required.

1.1 Modified ratio estimators

A problem, endemic to the estimation of parameters in the presence of infinite variance (heavy-tailed) processes, is that a few outliers in the time series can often greatly influence the estimate. This difficulty also exists for the ratio estimator.

Most points in the ratio estimator plot of a "problem" series follow a straight line, but there are a few which are considerably displaced. As an example, when the true α is 1.5, the estimate can equal 2.19. This is because the displaced points have been included in the least squares estimate of the slope and have skewed it considerably. To alleviate this problem, we can use a robustified least squares fit, formally a "least-trimmed squares regression" (*ltsreg* in S-Plus) which minimizes the sum of half of the smallest residuals. This method allows the outlying points of the plot to be ignored. Thus, for the series above, the robustified fit produces an estimate of 1.61. The price one has to pay for the increased robustness is that fewer points are used in the regression, and thus the standard deviation of the robustified estimator is greater. Ideally, one should only use the robustified version when necessary, and stay with the original version otherwise. In general, by looking at the plot, the practitioner should be able to notice when robustification is needed.

References

[1] C. K. Peng, S. V. Buldyrev, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E*, 49:1685–1689, 1994.

[2] 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.

[3] 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.