Estimators for Long Range Dependence: An Empirical Study

William Rea, Marco Reale and Jennifer Brown
Department of Mathematics and Statistics
University of Canterbury
Christchurch
New Zealand

December 13, 2007

Abstract

We present the results of a simulation study into the properties of 11 different estimators of the Hurst parameter, H, or the fractional integration parameter, d, in long memory time series. We compare and contrast their performance on simulated Fractional Gaussian Noises and fractionally integrated series with lengths between 100 and 10,000 data points and H values between 0.55 and 0.90 or d values between 0.05 and 0.40. We apply all 11 estimators to the Campito Mountain data and estimate the accuracy of their estimates using the Beran goodness of fit test for long memory time series.

Keywords: Strong dependence, global dependence, long range dependence, Hurst parameter estimators

1 Introduction

The subject of long-memory time series was brought to prominence by Hurst (1951) and has subsequently received extensive attention in the literature. See the volumes of Beran (1994), Embrechts and Maejima (2002), and Palma (2007) and the collections of Doukhan et al. (2003) and Robinson (2003) and the references therein.

Of critical importance in analysing and modeling long memory time series is estimating the strength of the long-range dependence. Two measures are commonly used. The parameter H, known as the Hurst or self-similarity parameter, was introduced to applied statistics by Mandelbrot and van Ness (1968) and arises naturally from the study of self-similar processes. The other measure, the fractional integration parameter, d, arises from the generalization of the Box-Jenkins ARIMA(p,d,q) models from integer to non-integer values of the integration parameter d. This generalization was accomplished independently by Granger and Joyeux (1980) and by Hosking (1981). The fractional integration parameter d is also the discrete time counterpart to the self-similarity parameter H and the two are related by the simple formula $H = d + 1/2$.

A number of estimators of H and d have been developed. These are usually validated by an appeal to some aspect of self-similarity, or by an asymptotic analysis of the distributional properties of the estimator as the length of the time series converges to infinity. Nine of these estimators were discussed in some detail by Taqqu et al. (1995). They also carried out an empirical study of these estimators for a single series length, five values of both H and d, and 50 replications. Since then available computer power has increased considerably. We have extended their study to a larger number of parameters, higher number of replications and two additional estimators as detailed in Section (2) below.

The remainder of the paper is organized as follows. Section (2) gives details of the method. Section (3) presents the results. Section (4) applies the methods to the Campito Mountain data which is regarded as a standard example of a long memory time series. Section (5) contains the discussion and Section (6) gives our conclusions and suggests avenues of future research.

2 Method

Ten estimators are implemented in the contributed package `fSeries` of Wuertz (2005) for the popular statistical software `R` (R Development Core Team (2005)). They are the absolute value, aggregated variance, boxed periodogram, differenced variance, Higuchi, Peng, periodogram, rescaled range, wavelet, and the Whittle. The wavelet estimator is discussed in some detail by Abry and Veitch (1998), Abry et al. (1998), and Veitch and Abry (1999) and the other nine are discussed briefly by Taqqu et al. (1995). Further, the estimator of Haslett and Raftery (1989) for d is implemented in the contributed package `fracdiff` of Fraley et al. (2006).

Taqqu et al. (1995) simulated both fractional Gaussian noises (FGNs) and the corresponding discrete time fractionally integrated (FI(d)) series and found that each estimator perfomed similarly whether estimating H in simulated FGNs or d in simulated FI(d)s. For example, if an estimator was biased when estimating H it was also biased in a very similar manner when estimating d. Thus, with the exception of the Haslett-Raftery estimator, we only investigated each estimator's performance in estimating H for simulated FGNs. FGNs were generated using the function `fgnSim` in `fSeries`. We ran 1000 replications of simulated FGNs with 100 different lengths and eight different H values. The lengths were between 100 and 10,000 data points in steps of 100. The H values were between 0.55 and 0.90 in steps of 0.05. For each series H was estimated by each of these ten estimators. For each H value and series length we estimated the median, 75% and 95% confidence intervals empirically from the simulated data. The H or d estimates were sorted into ascending order and the median obtained by averaging the 500th and 501st values. Similar calculations were done for the upper and lower values of the 75% and 95% confidence intervals.

For the Haslett-Raftery estimator of d we generated FI(d) series with the function `farimaSim` in `fSeries` over the range 0.05 to 0.40 in steps of 0.05. The other details are the same as above. In the presentation of the results we converted the Haslett-Raftery d estimates to H equivalents to facilitate comparisons among the estimators.

The simulations and estimations were performed on a SunBlade 1000 with 750Mhz UltraSPARC-III CPU with 2Gb of memory and a Sun Ultra 10 with

440Mhz UltraSPARC-IIi CPU and 1Gb of memory.

3 Results

To present the results in tabular form would require a very large amount of space. Thus we present them in graphical form. Figures (1) through (7) present some of the results. The remainder are available from the authors' website. Figures (1) through (6) are presented with a vertical axis covering 1.2 H units to facilitate comparisons of the estimators' standard deviation of their estimates. It should be noted that stationary long memory occurs in the range $0.5 < H < 1.0$. Baillie (1996) states that for $1.0 \leq H < 1.5$ the series is non-stationary but mean reverting while for $0 \leq H \leq 0.5$ the series in anti-persistent. Figure (7) presents the mean squared error (MSE) as a function of series length. We report MSE for series lengths greater than or equal to 500 data points. Again the vertical axes all have the same range to facilitate comparisions.

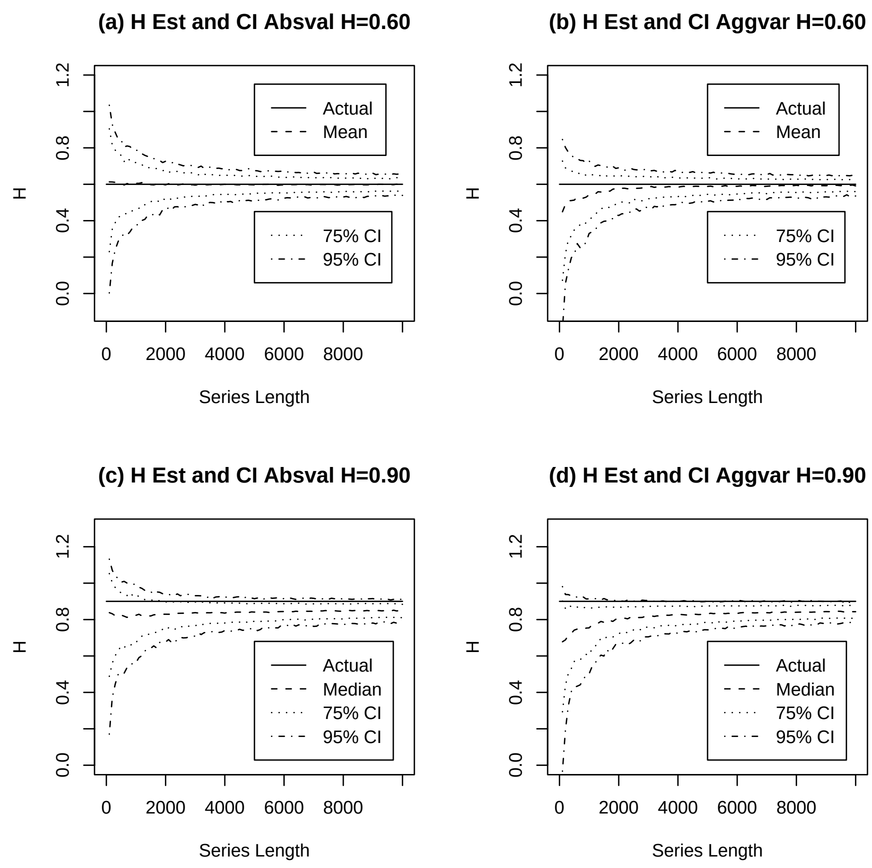

Figure 1: Empirical confidence intervals for the H estimates with $H = 0.60$ and $H = 0.90$; a and c absolute value method, b and d aggregated variance estimator.

The absolute value of the variance method was unbiased at all lengths when H was low (0.55 or 0.60) but became progressively biased and underestimated H as H increased (Figure 1 (a) and (c)).

The aggregated variance method (Figure 1 (b) and (d)) exhibited bias and underestimated H in short series when H was low. As H increased the estimator became increasingly biased at all series lengths examined. With $H = 0.90$ the true value of H lay above the upper 95% empirical confidence interval for all but the shortest series lengths.

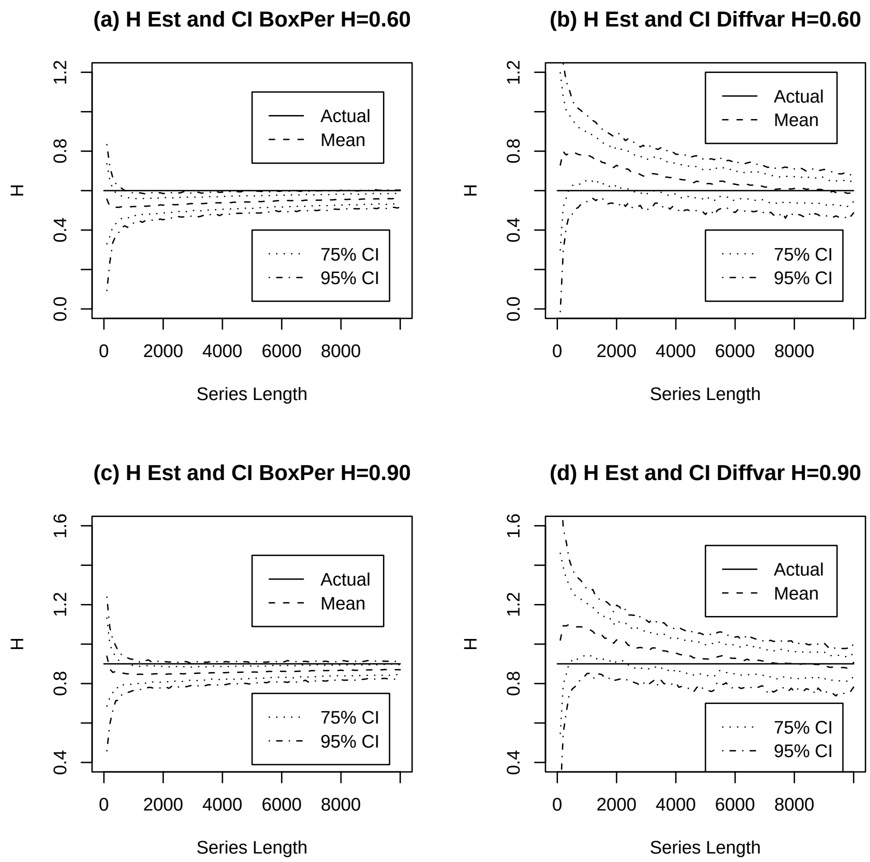

Figure 2: Empirical confidence intervals for the H estimates with $H = 0.60$ and $H = 0.90$; a and c boxed periodogram method, b and d differenced variance estimators

The boxed periodogram method was developed specifically to deal with perceived problems with the periodogram estimator (see Section (5) below for details). Comparing the boxed periodogram (Figure 2 (a) and (c)) with the unmodified periodogram method (Figure 4 (a) and (c)) we can see that for FGNs where the series were short and H was high that the periodogram method was biased towards over estimating H. The boxed periodogram was biased towards underestimating H for almost all values of H and series lengths examined.

The differenced variance method (Figure 2 (b) and (d)) had one of largest confidence intervals of the estimators when the series are short but this slowly decreased as sample size increased. Only the periodogram and wavelet methods had a similarly wide confidence interval for short series. The difference variance estimator exhibited bias towards over estimating H for any series with less than 7,000 observations. This bias was very serious in the short series. For series longer than about 9,000 observations the estimator exhibited a small amount of bias towards underestimating H.

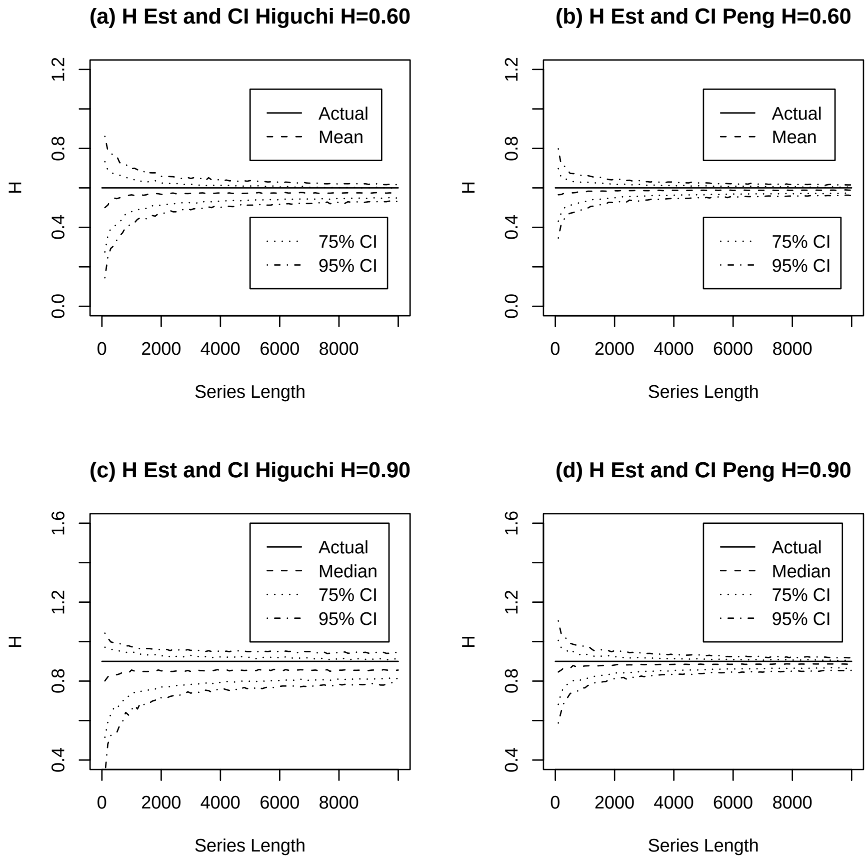

Figure 3: Empirical confidence intervals for the H estimates with $H = 0.60$ and $H = 0.90$; a and c Higuchi estimator, b and d Peng estimator.

The Higuchi (1988) estimator (Figure 3 (a) and (c)) was biased towards underestimating H but the magnitude of the bias appears relatively independent of H. The confidence interval of the estimate increased with increasing H.

The Peng et al. (1994) estimator (Figure 3 (b) and (d)) was biased toward under estimating H in the series lengths we investigated. This bias appeared to be independent of H but greater in short series.

The R/S estimator is of considerable historical interest because it was first proposed by Hurst and used extensively in early studies of long-memory processes. However, as can be seen from Figure (4b) and (4d) the R/S estimator

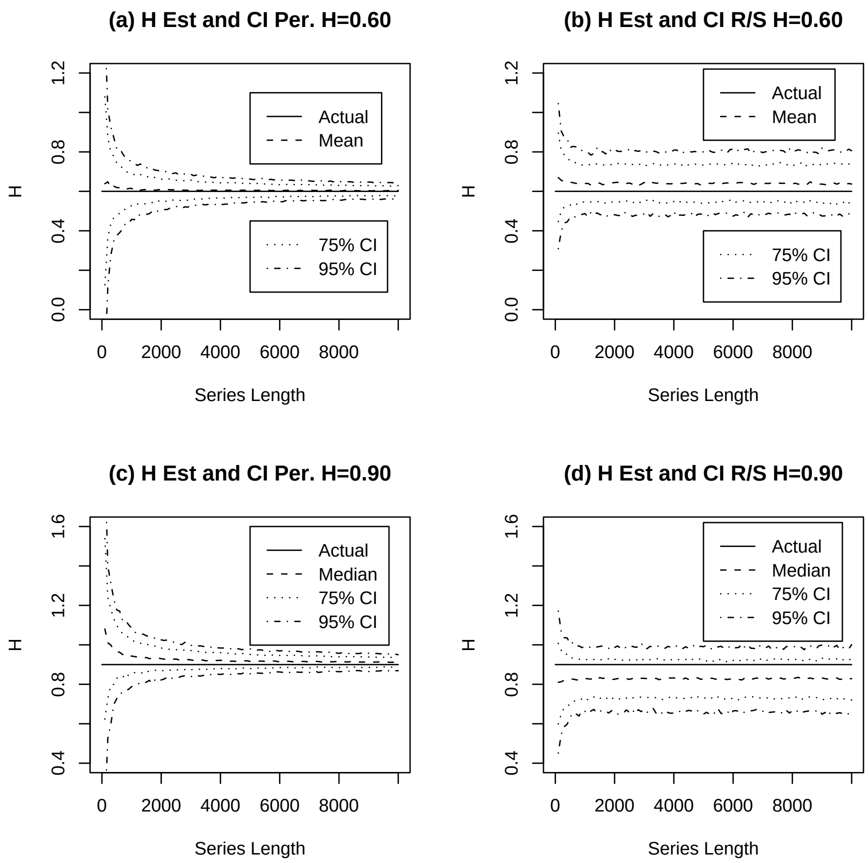

Figure 4: Empirical confidence intervals for the H estimates with $H = 0.60$ and $H = 0.90$; a and c periodogram estimator, b and d R/S estimator.

exhibited three problems; it was biased upwards when H was low, it was biased downwards when H was high, and the confidence interval of the estimate did not decrease with increasing series length once the series reached about 1000 observations.

Compared to the other nine estimators implemented in `fSeries` the Whittle estimator (Figure 5 (a) and (c)) was remarkable for its narrow confidence interval. It only displayed a small amount of downwards bias when the series was short and H was high. There was an implementation issue in the software we used. The Whittle estimator would terminate with an error when H was low and the series contained only a few hundred observations. Thus in Figure (5) there is no data for series with less than 300 observations in the $H = 0.65$ results.

The wavelet estimator was unbiased for all H values at series lengths over 4,100 data points. The bias present in series shorter than 4,100 data points was very small. The availability of a new *octave* can be seen in Figures (5b) and (5d) with each doubling of the series length. New *octaves* result in a series of

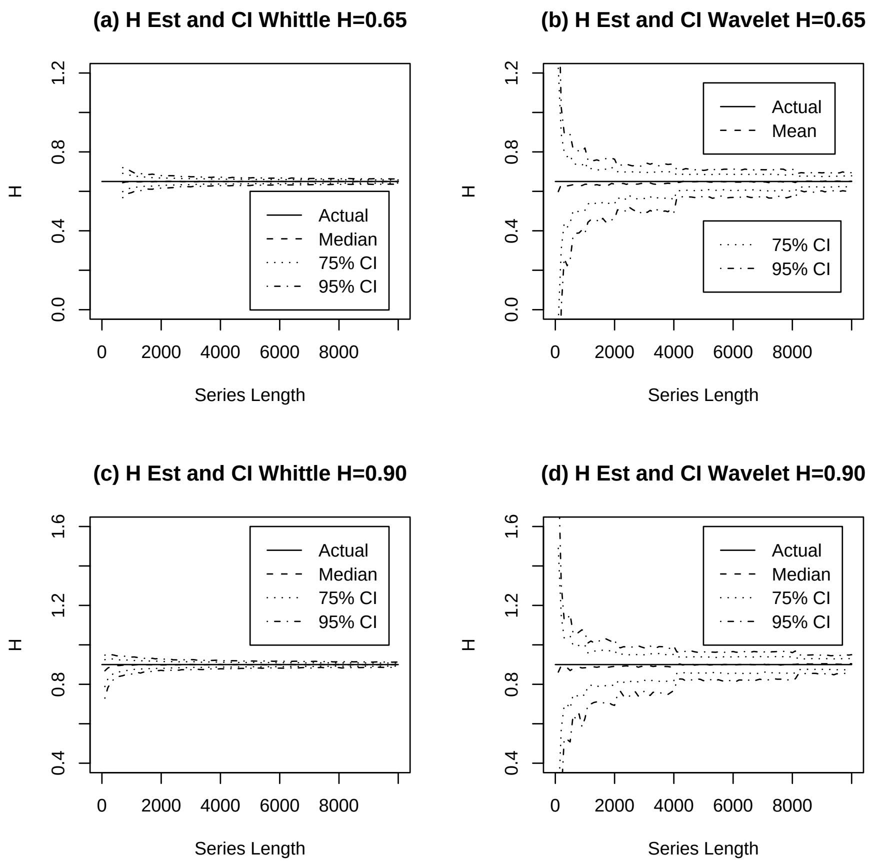

Figure 5: Empirical confidence intervals for the H estimates with $H = 0.65$ and $H = 0.90$; a and c Whittle estimator, b and d wavelet estimator.

steps in the reduction of the confidence interval of the estimate with increasing series length. The estimator has constant variance when the number of *octaves* is fixed.

The Haslett-Raftery estimator, Figures (6a) and (6b), does not return estimates of d less than zero ($H < 0.5$). Hence for low d and short series the distribution is truncated on the low side at $d = 0$ or $H = 0.5$ in Figure (6a). The Haslett-Raftery estimator was an excellent estimator with only small amounts of bias in the short series and had a narrow confidence interval.

Figure (7) presents the MSEs for the estimators for $H = 0.9$ or $d = 0.4$ as appropriate. This is an alternative way to look at the data from the simulations. We only report MSEs for series of 500 data points and longer because of the high MSEs for some estimators in the short series. The Whittle and Haslett-Raftery both had low MSEs in all series greater than 500 data points in length.

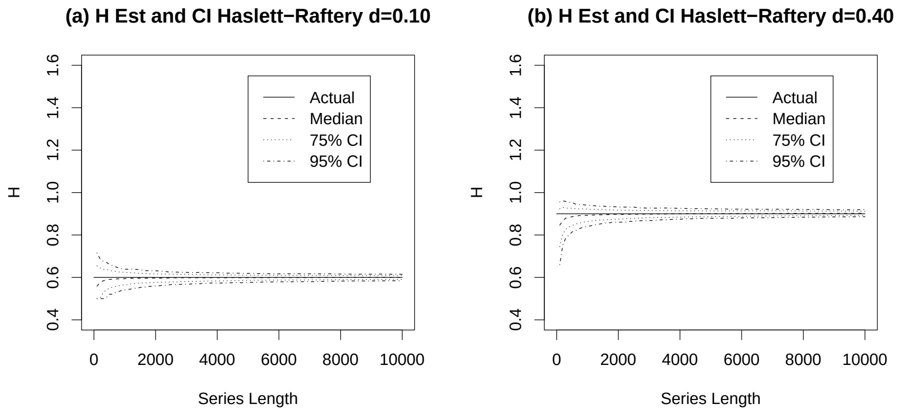

Figure 6: Empirical confidence intervals for the H estimates with $d = 0.10$ ($H = 0.60$) and $d = 0.40$ ($H = 0.90$); a and b Haslett-Raftery estimator.

4 Application: Campito Mountain Data

The Campito Mountain bristlecone pine data is regarded as a standard example of a long memory time series. It is a 5405 year series of annual tree ring widths of bristlecone pines on Campito Mountain, Califonia. It was studied by Baillie and Chung (2002) who determined that an ARFIMA(0,0.44,0) model fitted the data best. The lack of additional short term correlation in the data means it is a good candidate for modelling with an FGN.

We used the `camp` data as supplied in the `R` (R Development Core Team (2005)) package `tseries`. We applied the 11 estimators to this series and estimated the goodness of fit to an FGN for all estimators, except the Haslett-Raftery, using the test of Beran (1992) as implemented in the `R` package `longmemo` of Beran et al. (2006). The Beran test is more powerful against under estimation of H than over estimation. The Beran test was unable to be used for H values exceeding unity.

The results are presented in Table (1). The Beran test indicated an H value close to 0.89 fitted this data best. The maximum p-value was 0.577 for values of H estimated by the aggregated variance and rescaled range estimators. Nine of the 11 estimators returned H or d values which lie in an acceptible range on the basis of the Beran (1992) test assuming we set our level of statistical signicance at 0.05 to reject the null hypothesis of an FGN. The remaining two could not be tested.

Given the results from our simulated FGNs there were some unexpected H estimates for the Campito data. On the basis of the simulations we expected the aggregated variance, absolute value, boxed periodogram, Higuchi, and rescaled range to return a low estimate for H. None of these estimators did so. As the Beran test reported that $H = 0.89$ yielded the best fit we used the median value from the simulations with series length 5400 and $H = 0.90$, and adjusted for the difference of 0.01 H units, to estimate the value of H which would be reported by each estimator if the data was from an FGN. This value is reported in Table (1) as "Expected H". The sixth column reports the empirically

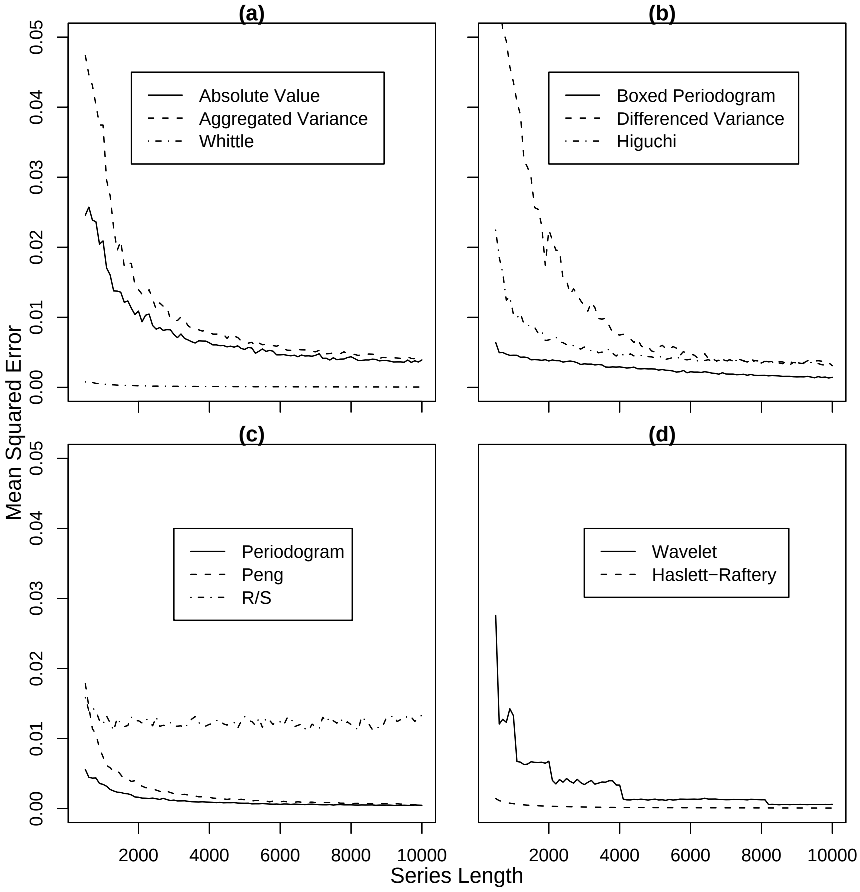

Figure 7: Mean squared errors (MSE) as a function of series length for all 11 estimators with d=0.4 for the Haslett-Raftery and H=0.9 for the other ten. MSEs are reported starting at a series of 500 data points.

determined p-value for the actual estimate again using the simulated data. We do not report values for the Haslett-Raftery estimator as it estimates d not H. It is interesting that six of the ten estimators returned H estimates which are statistically significantly higher than their expected values.

The estimator which had the least bias and narrowest confidence interval in the simulations with series length 5400 and $H = 0.90$, namely the Whittle, was marginally out performed by the aggregated variance and rescaled range judged on the basis of the Beran test.

The fourth column of Table (1) reports CPU times in seconds. With present day computer speeds estimation times on the Campito series are not an issue. Only three estimators required more than one second of CPU time on this 5405 observation series. It is evident that some estimators which require a long compute time, such as the Higuchi and Peng, did not yield a more accurate estimation of H for this data.

Method	H Est	Beran p-value	CPU Seconds	Expected H	Empirical p-value
Absolute Value	0.862	0.435	0.19	0.831	0.70
Aggregated Variance	0.889	0.577	0.34	0.821	0.04*
Boxed Periodogram	0.914	0.509	0.09	0.849	0.01**
Differenced Variance	1.089	-	0.21	0.925	0.01**
Haslett-Raftery	0.947	0.241	0.17	-	-
Higuchi	0.966	0.102	19.65	0.845	< 0.001***
Peng	0.936	0.344	18.46	0.875	< 0.001***
Periodogram	1.007	-	0.06	0.908	< 0.001***
Rescaled Range	0.892	0.577	0.04	0.816	0.36
Wavelet	0.927	0.421	0.07	0.889	0.25
Whittle	0.876	0.540	1.05	0.890	0.15

Table 1: The first two columns of results presents the H estimates and p-values returned by the Beran (1992) test for the Campito Mountain data for each of the 10 estimators of H and the Haslett-Raftery estimator of d converted to H equivalent. CPU times are in seconds on the SunBlade described in the text. The expected H column is the expected value that the estimator would return if the Campito data was an FGN with $H = 0.89$. The empirical p-value column is estimated empirically from the simulated data.

5 Discussion

It is clear from the simulations that not all estimators are created equal. Long memory occurs in the range $0.5 < H < 1$. Thus any estimator used to estimate the strength of the long memory needs to be both accurate and have a low variance.

The absolute value of the variance method is similar to both the aggregated variance method and the Higuchi method. The theoretical properties of the absolute value method have been investigated by Giraitis et al. (1999) who showed that this estimator had a bias of not less than $(\log N)^{-1}$ where N is the series length.

The boxed periodogram method was developed specifically to deal with the problem of having most of the points used to estimate H on the right-hand side of the graph. This was believed to, possibly, cause bias in the periodogram estimator. Beran (1994) p133 and Taqqu et al. (1995) outline some of the reasons such a method could be expected to be biased. Theoretical properties of periodogram estimators have been investigated by Robinson (1994), Robinson (1995), and Lobato and Robinson (1996).

The differenced variance method was developed to be robust to trends which are known to cause spurious long memory in the R/S estimator, see Bhattacharya et al. (1983). We did not test its robustness.

Teverovsky and Taqqu (1999) established that the differenced variance method had a higher variance than the aggregated variance method, a result supported by our simulation study. In fairness to the method it must be pointed out that Teverovsky and Taqqu (1999) did not intend for it to be used alone but rather in conjunction with the aggregated variance method to test for the presence of shifting means or deterministic trends.

The Higuchi (1988) estimator only indirectly estimates H. It estimates the fractal dimension, D, of a series by estimating its path length. As implemented it then converts the estimate of D to H by the simple relationship $H = 2 - D$. This should be borne in mind if a researcher wishes to estimate D rather than H as it is a simple matter to recover D from the H estimate.

Taqqu et al. (1995) give a detailed proof that the method of Peng et al. (1994) is asymptotically unbiased. In the simulations the bias was never large but even at a sample size of 10,000 observations the estimator cannot be considered unbiased. However, its MSE approaches that of the Periodogram method as the series length increases which, in turn, is better than several others.

Some theoretical properties of the R/S estimator have been examined by Mandelbrot (1975) and Mandelbrot and Taqqu (1979). Mandelbrot (1975) proved that the R/S statistic is robust to the increment process having a long-tailed distribution in the sense that $E[X_i^2] = \infty$.

However, Bhattacharya et al. (1983) proved that the R/S statistic was not robust to departures from stationarity. Thus for a short memory process with slowly decaying deterministic trend the R/S statistic will return an estimate of H which implies the presence of long-memory.

The use of wavelets to estimate H is a natural use of wavelets as pointed out by Abry and Veitch (1998), Abry et al. (1998), and Veitch and Abry (1999). The wavelet estimator is asymptotically unbiased. In the simulations the bias was always small but disappeared for series with more than 4,000 observations.

Some of the theoretical properties of the Whittle estimator have been investigated by Fox and Taqqu (1986), Dalhaus (1989) and Horvath and Shao (1999).

6 Conclusions and Future Research

Of the eleven estimators examined here the Whittle and Haslett-Raftery estimators performed the best on simulated series. They only exhibited bias in short series with high H or d values and had the lowest and second lowest MSEs respectively of the 11 estimators at all series lengths. These empirical results are supported for the Whittle estimator by the theoretical work of Fox and Taqqu (1986) and Dalhaus (1989).

For series with 4,000 or more data points, the Peng, periodogram and wavelet estimators look to be good choices based on their MSEs.

The Higuchi estimator is useful if the researcher wishes to recover the fractal dimension of the time series. In contrast to the other estimators it provides useful information on a time series if the series is not an FGN (or FI(d)) series. As an estimator of H it is inferior to several others.

The boxed periodogram method is clearly inferior to the periodogram method it was intended to improve upon for FGNs. Further research would be needed to test if it is more robust than the periodogram method in series with departures from a pure FGN. This could be accomplished, for example, by simulating ARFIMA series with non-zero AR and MA components or series with structural breaks.

The R/S estimator is of considerable historical interest but had a major deficiency in that its MSE plateaued while all other estimators' MSEs decreased with increasing series length. Against this we must note that it was one of the two best performing estimators when applied to the Campito data when judged by the Beran test.

The differenced variance estimator was the worst of the eleven estimators in short series. For series longer than 6,000 data points its MSE was better than the R/S and on a par with the absolute value, aggregated variance and Higuchi methods. As noted above, Teverovsky and Taqqu (1999) do not recommend its use in isolation as it is part of a test for shifting means or deterministic trends. Teverovsky and Taqqu (1999) also recommend the aggregated and differenced variance plots always be examined visually. We agree with these recommendations. We did not test its robustness to shifting means or deterministic trends. Some numerical results on its performance in these two situations can be found in Teverovsky and Taqqu (1999).

The application to the Campito data six of the estimators returned value statistically significantly different H estimates than expected based on the evidence from the simulated series. Although the fit of the Campito data to an FGN is good (p=0.577), these six estimators do not seem to be robust to whatever specific departures from an FGN that are present in the data. This suggests that a researcher should not rely on a single estimator when estimating H and that the Beran (1992) test should always be applied to test the goodness of fit of the data to an FGN.

Because of the apparent lack of robustness, a useful avenue of future research would be to quantify the sensitivity of these estimators to various types of departures from an FGN, e.g. FGN series with a small number of shifts in mean or a small number of outlier data points.

References

Abry, P. and Veitch, D. (1998). Wavelet Analysis of Long-Range-Dependent Traffic. *IEEE Transactions on Information Theory*, 44(1):2–15.

Abry, P., Veitch, D., and Flandrin, P. (1998). Long-Range Dependence: Revisiting Aggregation with Wavelets. *Journal of Time Series Analysis*, 19(3):253–266.

Baillie, R. T. (1996). Long memory processes and fractional integration in economics. *Journal of Economics*, 73:5–59.

Baillie, R. T. and Chung, S.-K. (2002). Modeling and forecasting from trend-stationary long memory models with applications to climatology. *International Journal of Forecasting*, 18:215–226.

Beran, J. (1992). A Goodness-of-Fit Test for Time Series with Long Range Dependence. *Journal of the Royal Statistical Society B*, 54(3):749–760.

Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman & Hall/CRC Press.

Beran, J., Whitcher, B., and Maechler, M. (2006). *longmemo: Statistics for Long-Memory Processes (Jan Beran) – Data and Functions*. R package version 0.9-3.

Bhattacharya, R., Gupta, V., and Waymire, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, 20:649–662.

Dalhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17:1749–1766.

Doukhan, P., Oppenheim, G., and Taqqu, M. (2003). *Theory and Applications of Long-Range Dependence*. Birkhaüser.

Embrechts, P. and Maejima, M. (2002). *Selfsimilar Processes*. Princeton University Press.

Fox, R. and Taqqu, M. S. (1986). Large-Sample Properties of Parameter estimates for Strongly Dependent Stationary Gaussian Time Series. *The Annals of Statistics*, 14:517–532.

Fraley, C., Leisch, F., Maechler, M., Reisen, V., and Lemonte, A. (2006). *fracdiff: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models*. R package version 1.3-0.

Giraitis, L., Robinson, P. M., and Surgailis, D. (1999). Variance-type estimation of long memory. *Stochastic Processes and their Applications*, 80:1–24.

Granger, C. W. J. and Joyeux, R. (1980). An Introduction to Long-range Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1:15–30.

Haslett, J. and Raftery, A. E. (1989). Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resources (with Discussion). *Applied Statistics*, 38:1–50.

Higuchi, T. (1988). Approach to an irregular time series on the basis of fractal theory. *Physica D*, 31:277–283.

Horvath, L. and Shao, Q.-M. (1999). Limit Theorems for Quadratic Forms with Applications to Whittle's Estimate. *The Annals of Statistics*, 9(1):146–187.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika*, 68(1):165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808.

Lobato, I. and Robinson, P. M. (1996). Averaged periodogram estimation of long memory. *Journal of Econometrics*, 73:303–324.

Mandelbrot, B. B. (1975). Limit Theorems of the self-normalized range for weakly and strongly dependant processes. *Z. Wahr. verw. Geb.*, 31(1):271–285.

Mandelbrot, B. B. and Taqqu, M. S. (1979). Robust r/s analysis of long-run serial correlation. In *Proceedings of the 42nd Session of the International Statistical Institute*, volume 48 of *Bulletin of the International Statistical Institute*, pages 69–104.

Mandelbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437.

Palma, W. (2007). *Long-Memory Time Series Theory and Methods*. Wiley-Interscience.

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

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robinson, P. M. (1994). Semiparametric Analysis of Long-Memory Time Series. *The Annals of Statistics*, 22(1):515–539.

Robinson, P. M. (1995). Log-Periodogram Regression of Time Series with Long Range Dependence. *The Annals of Statistics*, 23(3):1048–1072.

Robinson, P. M. (2003). *Time Series with Long Memory*. Oxford University Press.

Taqqu, M., Teverovsky, V., and Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals*, 4:785–798.

Teverovsky, V. and Taqqu, M. (1999). Testing for Long-Range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-Type Estimator. *Journal of Time Series Analysis*, 18:279–304.

Veitch, D. and Abry, P. (1999). A Wavelet-Based Joint Estimator of the Parameters of Long-Range Dependence. *IEEE Transactions on Information Theory*, 45(3):878–897.

Wuertz, D. (2005). *fSeries: Financial Software Collection - fSeries*. R package version 220.10063.