*Journal of Accounting, Finance and Economics*
*Vol. 1. No. 1. July 2011 Issue. Pp. 31-47*

# A Method of Detecting the Presence of Long Run Dependencies in Time Series

Dani AC Foo[1]

*It has been known for a while that the presence of long run dependencies in time series can confound analyses using statistical procedures that assume Gaussian or Poisson distributions. The implications are that unless this long memory effect is identified and factored into consideration, statistical conclusions assuming Gaussian and Poisson distributions can be overly biased. The report estimates the presence of long memory(LRD) in the stream of daily visitors, arriving from various sources to New Zealand from 1997 to 2010, using selected estimators of the Hurst-exponent. It is found that, after minimising the effects of short-term trends, periodicities, and cycles, there exist significant long memory embedded in data of all sources and in the aggregate. Further, existence of strong embedded "long memory" implies the existence of the "Joseph Effect" – that good times beget good times and bad times beget bad – in the underlying process and may have interesting implications for policy makers in the tourism industry in this particular case.*

**Field of Research:** Contemporary Issues in Economic and Financial Research, Development and Policy, Migration and Tourism, International Trade, Long Run Dependencies, Hurst Exponent.

## 1. Introduction

Tourism arrivals over time are of considerable interests to the tourism industry in New Zealand especially for formulation of long term policies towards enhancing various aspects of this segment of the economy. Such time series are usually analysed using conventional statistical procedures based on assumptions that do not take into account possible existence of LRD processes. This paper used a typical time series supplied by NZ Statistics Department to investigate whether LRD processes do exist in the said series and the implications to conventional statistical conclusions if LRD does exist.

The presence of long memory or "Long Run Dependecies" (LRD) and scaling behaviour (which is associated with "heavy tails") tend to taint conclusions based on assumptions of Gaussian distribution and on Poisson processes. The presence of LRD in a time-serial process affects the subsequent measurements far into the future, embedding a "memory" that tends to make measurements of the subsequent behaviour more volatile than expected. With the ready availability of reliable software such as SelQos[1] and SELFIS[2], it is now relatively easy to run preliminary tests for the presence of LRD. If LRD is detected, then one should be wary of drawing conclusions based purely on assumptions of Gaussian or Poisson distributions. It should be noted that LRD was commonly detected in fields as diverse as hydrology[3], economics[4], computer science[5], finance[6] and in the physical sciences[7]

[1] Department of Finance, Waikato Management School, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. Email: dfoo@waikato.ac.nz

The presence of LRD is characterized by a slow attenuation of power of the autocorrelation function, $Corr(X_{t,}X_{t+k})\sim Ck^{2H-2}\ as\ k\to\infty$ in a certain time series. $H$ is in fact the Hurst Exponent and the intensity of LRD can be estimated by the Hurst Exponent which varies between 0 and 1. At *H= 0.5* we have a series that does not have LRD, *H>0.5* implies the presence of LRD and the greater the intensity of LRD the higher the H estimate. Where *H<0.5* we have anti-dependence ("anti-persistence") where reversion towards the mean is higher that is normally expected.

The organization of this paper is as follows: In Section 2 Literature Review surveyed recent development in the applications of H-exponent estimators in various disciplines. Section 3 LRD Epistemology deals with theoretical aspects of H-exponent estimations. Section 4 Models of H-estimators Used deals with the respective constructions of each of the selected estimators. Section 5 Data, Methodologies Applied and Analyses, discuss how the data set is processed and analyzed by applying the selected estimators on various components of the data set. Section 6 Conclusions and Limitations were presented.

## 2. Literature Review

Much has been done in detecting LRD in computer science (Pacheco. Roman and Vargas, 2005), economics and finance (Taqqu, Teverovsky and Willinger, 1999) , and in the physical sciences using various *Hurst Index (H)* estimators, most of which have their own strength and weaknesses.

The presence of LRD in financial market has been a popular research topic starting with Mandelbrot (1967, 1971) who suggested that in the presence of LRD, the arrival of new market information cannot be fully arbitraged away and martingale models of asset prices cannot be justified. Therefore, pricing derivative securities with martingale methods is inappropriate if the underlying stochastic processes exhibit LRD and, consequently, statistical inferences concerning asset pricing models based on standard testing procedures may not be appropriate if LRD is presence.

Techniques such as the Rescaled Range (R/S) analysis, was first used by Hurst (1951). This popular methodology was subsequently used by Greene and Fielitz (1977), Wallis and Matalas (1970) and it was further popularized by Peters (1994). Aydogan and Booth (1988), began to find problems with the R/S estimator itself and Lo (1991) presented a refinement to the classical R/S estimator which was further found to be prone to Type II error bias by Teverovsky, Taqqu and Willinger (1999).

There are other methods of estimating the H-exponent and all these can be generally classified into two different domains: time domain based method and frequency domain based method. The R/S estimator discussed above is a time-domain method. So are the Absolute Moment Method and the Variance Method, developed later and used in this paper. Frequency domain based methods were later developed that include the Periodogram regression of Geweke and Porter-Hudak (1983), Local Whittle estimator (Kunsh, 1987) and Wavelet estimator (Veitch and Abry, 1998). All of which are used in this paper, but frequency domain methods are normally difficult to apply and to interpret.

It is well known that LRD estimators are themselves unstable and erratic (Karagiannis et al, 2006), mainly because H cannot be calculated in any definitive or direct way and has to be estimated as a by-product of some statistical estimating procedures. Depending on the estimator used, if sufficient care is not taken to ameliorate the underlying limitations of each of these respective estimators, the results can sometimes be mutually contradictory and conflicting (Karagiannis, Faloutsos and Reidi, 2002). For deeper methodological issues one should consult Beran (1992) or Allan (1996)

The more well known of problems in estimating LRD are the presence of short term trends, non-stationarity, discontinuity, periodicity and noise. Regardless, it is better to test whether LRD and other anomalies are present in a time series before deciding whether normal Gaussian analyses are appropriate. There is an increasing realisation that by just assuming that a time series is distributionally Gaussian or Poisson is unsatisfactory, simplistic and leads to estimation errors. The statistical properties of a series with LRD can be quite different from those of a series that are independent and identically distributed (iid). For instance, the variability properties of sample means of assumed iid observations are far from valid in the presence of LRD. We have selected the area of tourism arrivals for our study which we know has cycles, trends, seasonality and noise. We expected that the series is non-Gaussian time series, and decides to test for the presence of LRD. If the presence is confirmed, it is of practical significance; the least of which is that such presence may pose potentially significant problems in the statistical and substantive conclusions derived from traditional Gaussian-Poisson related techniques. A bonus is that the presence of LRD could be interpreted as "long memory" whose presence implies "persistence" which gave rise to the so called "Noah[8]" and "Joseph" effects.

## 3. LRD Epistemology

Theoretically, there are several ways of defining long memory process ("long range dependence" or LRD). An intuitively popular definition is couched in terms of the auto-covariance function, ρ(*k*), such that a long memory process is present if in the limit, $k \rightarrow \infty$:

$$\rho(k) \sim k^{-\alpha} L(k)$$

where $0 < \alpha < 1$and *L(x)* is a slow varying function. Hence a stationary process $X_t$ is long-range dependent, if there exists a real number $a \in (0,1)$ and a constant $c_1 > 0$ such that

$$\lim_{k \to \infty} \rho(k) / [c_1 k^{-a}] = 1$$

Where ρ(k) represents the sample correlation function and *k* is number of lags . The definition states that the autocorrelation function of long memory processes, decay to zero with rate approximately $k^{-a}$.The parameter that characterizes long-range dependence (LRD, "long memory") is the Hurst exponent (H), where $H = 1 - \alpha/2$. Long-memory occurs when ½ < H < 1 ("persistence") and 0 < H < ½ ("anti-persistent"). Long memory process can generate non-periodical cyclical patterns as ones observed by Hurst (1951) for the Nile River, where long periods of drought are followed by long periods of plenty. Mandelbrot and Wallis (1968) called this phenomenon as the "Joseph" or "Hurst" effect.

# 4. Models of H-Estimators Used

The following estimators were used in this study. These are chosen because they are the most common estimators and their efficacies and limitations are well documented above.

### 4.1 Time Domain Estimators

#### 4.1.1 Rescaled Range (R/S) Estimator[9]

The R/S estimator is a statistical estimator of H such that:

$$E\left[\frac{R(\tau)}{S(\tau)}\right] = Cn^H$$

where $R(\tau)$ is the amplitude range over a time window, τ, scaled to the standard deviation, $S(\tau)$, of the range. Here, $R(\tau) = \max\big(X(t,\tau)\big) - \min(X(t,\tau))\ for\ 1 \leq t \leq \tau$ and $S(\tau) = \sqrt{\frac{1}{\tau}\sum_{t=1}^{\tau}(\xi(t) - \langle\xi\rangle_\tau)^2}$ where $X(t,\tau) = \sum_{u=1}^{t}\xi(u) - \langle\xi\rangle_\tau$ and $\langle\xi\rangle_\tau = \frac{1}{\tau}\sum_{u=1}^{\tau}\xi(t)$. The R/S for any given τ is R/S(τ) = $\frac{R(\tau)}{S(\tau)}$, and H is the regression slope of log τ against log R/S(τ).

#### 4.1.2 Absolute Moment Estimator (A/M)

Here, a time series, $X_t$, is divided into blocks of size *m*, such that:

$_{(k)}X^m = \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i$ where *k = 1, 2,... N/m for a series* $X_i$, *i =1, 2,.... ,N* and *k* is the index that labels the block. The sum of the absolute values of the series is computed for various m, i.e.

$$A/M^{(m)} = \frac{1}{N/m}\sum_{k=1}^{N/m}\left|X_k^{(m)} - \bar{X}\right|,$$

where $\bar{X}$ denotes the original series' sample mean. Regressing the log of this statistic against the log of *m* should provide a line with a slope of *H-1,* where *H* is the Hurst exponent.

#### 4.1.3 Variance Method Estimator (VM)

VM exploits a characteristic property of the variance inherent in LRD processes that the variance of the sample mean converges to zero slower that the 1/N where N is the sample size. If LRD is present, then we have, for large N:

$$Var(\bar{X}_N) \sim cN^{2H-2}$$

where c > 0 and $\bar{X}_N$ is the sample mean. If we divide a time series, $X_t$, into block size of m, and within each block, m, aggregate the sub-series to produce a new sub-series, $X^{(m)}$ such that $_{(k)}X^{(m)} = \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i$ where k = 1, 2, .... and k is the label index of the block. The sample variance of $X^{(m)}$, is calculated as

$$s^2(m) = \frac{1}{\left(\frac{N}{m}\right) - 1} \sum_{k=1}^{N/m} \left(X_{(k)}{}^{(m)} - \bar{X}\right)^2$$

Where $\bar{X}$ denotes the global mean. Regressing log $s^2(m)$ against log(m) for each m, successively, we should have a line with a slope of *2H-2*, from which H can be inferred.

### 4.1.4 Variance Of Residuals Estimator

A time series, say, {$X_i$, i ∈ positive integers}, is divided into blocks *m*, defined as $\Psi_m$= $\{\Psi_1^m, \Psi_2^m, \ldots. \Psi_n^m, ..\}$, a derived new series from {$X_i$} with blocks size *m*. Each $\Psi_i^m \stackrel{\text{def}}{=} \{X_{(i-1)m+1}, X_{(i-1)m+2}, \ldots, X_{im}\}$ represents a block of size *m* of the original series. To every $\Psi_m$, the partial sum series $P(\Psi_i^m) = \{P_i^m(1), P_i^m(2), \ldots, P_i^m(m), \ldots\}$, where each of $P_j^m(j) = \sum_{j-1}^{m} X_{(i-1)m+j}$ and $P^m = \{P(\Psi_1^m), P(\Psi_2^m), \ldots.., P(\Psi_i^m), \ldots\}$, can be computed. Then a least square line is fitted to the partial sum series $P^m$ to give a new series $\mathcal{Z}^m = \{Z_1^m, Z_2^m, \ldots, Z_i^m, \ldots\}$ ("the least square series"). The variance of the residual is given as:

$$V_{res}^m(i) = \frac{1}{m}\sum_{j=1}^{m}(P_i^m(j) - Z_i)^2$$

If the process is known to be LRD, then the median of the residual series behaves as $Med(V_{res}^m) \sim m^{2H}$ for large m. A log-log plot of $Med(V_{res}^m)$ against varying *m* should give a straight line with slope 2H, where H is the Hurst exponent.

## 4.2 Frequency Domain Estimators

### 4.2.1 Periodogram Estimator

The periodogram may be defined as follow:

$$S(v) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X\,(j)e^{ijv}\right|^2$$

where *v* is the frequency and *X* is a given time series of length *N*. Given a series of finite variance, the *S(v)* is an estimator of the spectral density of X, then a series with LRD will have a spectral density *S(v)* proportional to $|v|^{1-2H}$at the lower frequencies close to zero. Hence, a log-log plot of the *S(v),* also known as the *periodogram*, against *v, frequencies*, should present a straight line with a slope of 1-2H. The frequencies used to estimate H are best taken at the lowest 10% of the spectrum in order to comply with the requirements stated above.

### 4.2.2 Local Whittle Estimator

It was stated in Taqqu and Teverovsky (1997) that the local Whittle estimator is semi-parametric and assumes the existence of LRD. :Local Whittle ("LWhittle") is preferred here because it makes less *a priori* assumptions that the standard Whittle. Since Local Whittle is based on the periodogram stated above, it's focus is also centred around low frequencies of the spectrum. LWhittle differs from the periodogram approach by its adding an extra parameter, M, which is an integer of less than N/2, and satisfying (1/M) + (M/N) → 0 as N→∞. Assuming only the functional form of spectral density we have:

$$f(v) \sim G(H)|v|^{1-2H} as\ v \ \rightarrow 0$$

Objective is to minimize:

$$R(H) = \log\left(\frac{1}{M}\sum_{j=1}^{M}\frac{S(v_j)}{v_j^{1-2H}}\right) - (2H-1)\frac{1}{M}\sum_{j=1}^{M} log v_j$$

Please note that unless the series is understood to be ideal, then M should be as small as possible, which is in effect using frequencies close to zero to minimize the effect of short range effects on the spectral density.

### 4.2.3 Abry-Veitch Estimator

Generally, the Hurst exponent is derived from a wavelet transform (in in case, Daubechies wavelets) of the time series X = ($x_1$, $x_2$, .... $x_n$). Given a series with long memory stochastic process, the variance at level *i* of the wavelet coefficients, $d_x(i,j)$ is given by:

$$Var\big(d_x(i,..)\big) = \frac{\sigma^2}{2} V_\psi(H)(2^j)^{2H+1}$$

where $V_\psi(H)$ depends on a particular wavelet chosen and the Hurst exponent. V is defined by:

$$V_\Psi(H) = -\int_{-\infty}^{\infty} \gamma_\psi(\tau)|\tau|^{2H} d\tau$$

Taking the log of the variance of wavelet coefficients above we have:

$$logVar\big(d_x(i,..)\big) = (2H+1)j + K$$

where K is a constant.

Specifically, a time average $\mu_i$, of $d_x(i.j)$ is computed at a given scale. $\mu_i = (n_i)^{-1}\sum_j^{n_i} d_x^2(i.j)$, where $n_i$ is the wavelet coefficient number at scale *i* and *n* the time series points. The Hurst exponent is estimated from the slope of a linear regression model stated as follows:

$$log_2(\mu_i) = \ log_2\ (\frac{1}{n_i}\sum_{j=1}^{n_i} d_x^2(i,j))$$

where *i* = 1, 2, ... ($log_2(n)$.

## 5. Data, Methodologies Applied and Analyses

In this paper, we will apply the most common H estimators to several series of daily visitor arrival data. These were procured from the New Zealand Statistics Department for the period spanning 1 September 1997 through 31 October 2010, totalling close 5000 data points. The series comprise of daily arrivals from Australia, China, Japan, UK, USA and the rest of the world and a derived series of Total Arrivals from these respective sources.

One of the main objectives is to assess whether there are pronounced LRD presence in these series, if so, there is profound implications for normal statistical analysis of these series if the presence of LRD is not allowed for because any statistical conclusions derived therefrom may be suspect. Further, if LRD is strong (H>>0.5), after accounting the presence of known "contaminants" such as short-term trends, non-stationarity, periodicity and noise, one might infer that there is LRD along with all the implications of the manifestations of persistence, self-similarity and the presence of an underlying fractal structure. All these are indications of the presence of "heavy tails" in the distributions of the time series concerned. These attributes violate the tenets of traditional Poisson modelling and the assumption of independence.

Evidence of a high H-exponent suggests "persistence" and the existence of the "Joseph Effect" where a good year tends to beget further good years and bad begets further bad. This could provide important signposts for long term policy making, especially in the strategic dimensions of tourism infra-structural development, market development and policy making.

### 5.1 Total Tourist Arrivals from All Sources

The plot, Figure 1, is quite typical of tourist arrival patterns and conforms to arrival patterns in other studies (Medeiros C, McAleer M, Slottje D, Ramos V and Rey-Maquieira J, 2008) The plot shows distinct peaks and troughs of a mix of periodicities, the bane of most LRD estimators. Further, the perceived existence of trend and noise do not help matter much either.

The Autocorrelation Function plot shown in Figure 2 clearly shows the effects of periodicities. All these are indications that the series are not amenable to traditional statistical analysis. For our purpose, we need to rid the series of these impediments as much as possible. Karagiannis, et al. 2006 suggests *Randomized Buckets* for ridding the series of most of these impediments. The procedure generally involves taking selected sub-series of a selected time series, randomizing these to "control the amount of correlation at different time scales." The procedure comprises External Randomization, Internal Randomization and Two Level Randomization. In this paper we will restrict ourselves to Internal Randomization.

Figure 1: Total Daily Arrivals: 1-9-1997 to 31-10-2010

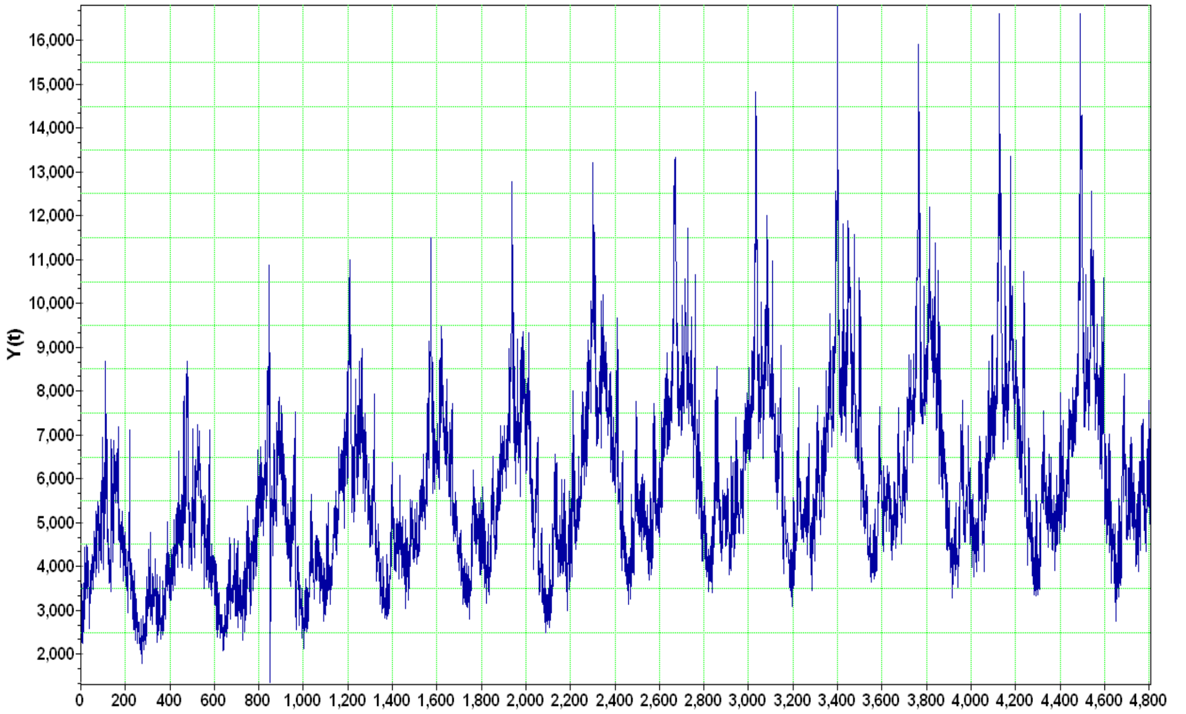
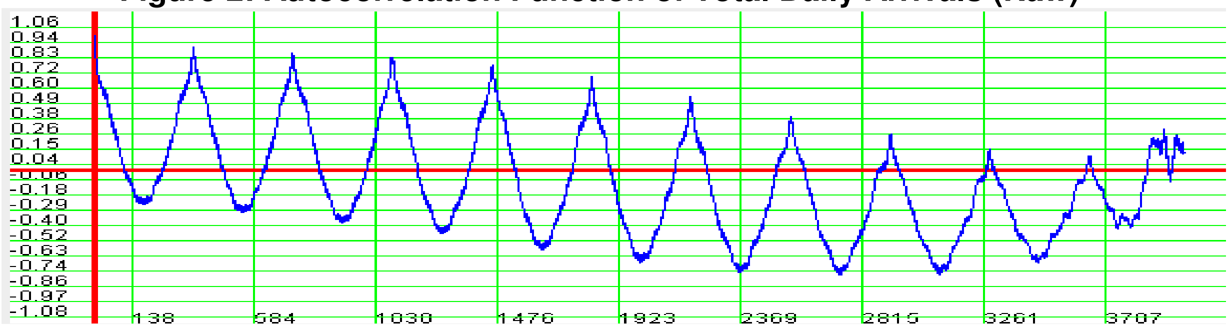



Figure 2: Autocorrelation Function of Total Daily Arrivals (Raw)




Briefly, the series is divided into segments ("buckets"), where intra-bucket data elements within the buckets are randomized, without the order of each of the contiguous buckets being changed. This has the effect of minimizing any correlations between intra-bucket data elements while correlations among the buckets themselves are maintained. If the original series contains LRD, then the ACF after this procedure should still manifest a "power-law" structure as the existence of LRD should be preserved.

The Internal Randomizing ("IR") procedure is applied to the Total Arrival data series and the result can be seen in Figure 3. Although the periodicity has been dampened, it is still discernible as is evident in the ACF presented in Figure 4. The IR procedure is now repeated, by increasing the bucket size, until all discernible evidence of periodicity or short-term trends disappear.

**Figure 3: Total Tourist Arrivals after Internal Randomizing (Bucket size=100)**

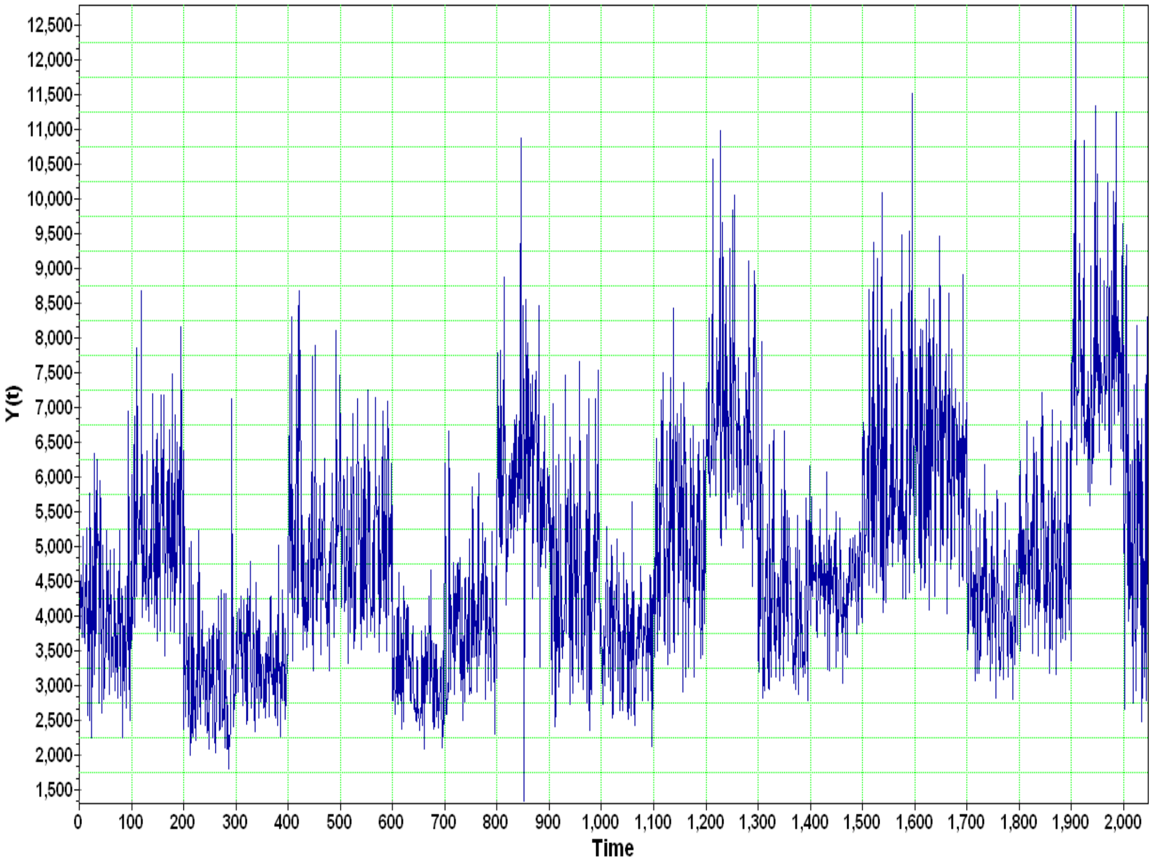
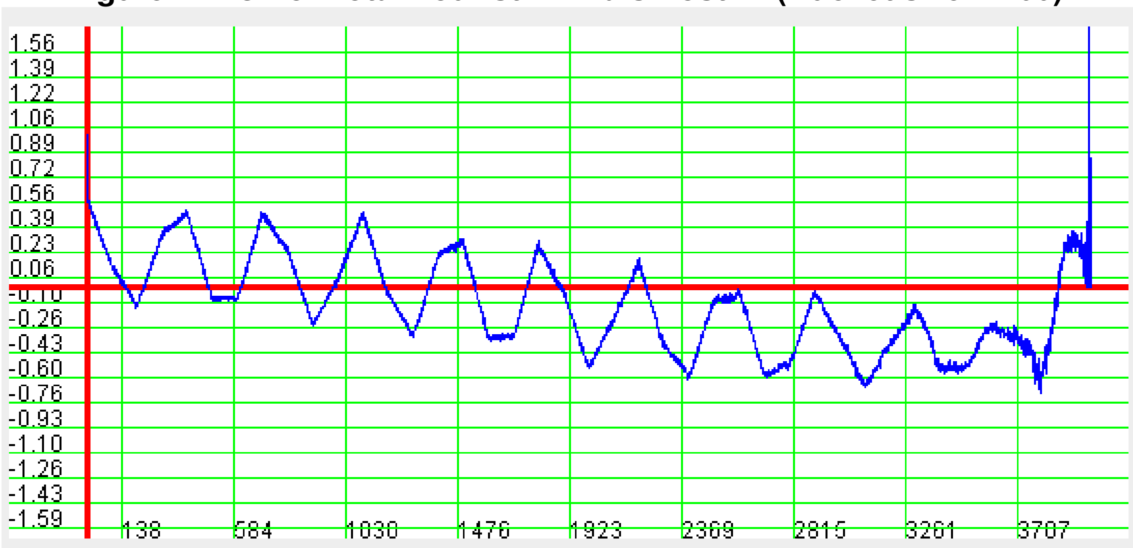



**Figure 4: ACF of Total Tourist Arrivals Post IR (Bucket size = 100)**




At bucket size of approximately 300 we have the following in Figure 5 and the corresponding ACF in Figure 6. Although the ACF here is not parabolic, it does show the characteristic slow attenuation of the function. Suffice to say we seem to have rid the data-set of most periodicity and short term trends.

**Figure 5: Total Tourist Arrival after Internal Randomization (Bucket size = 300)**

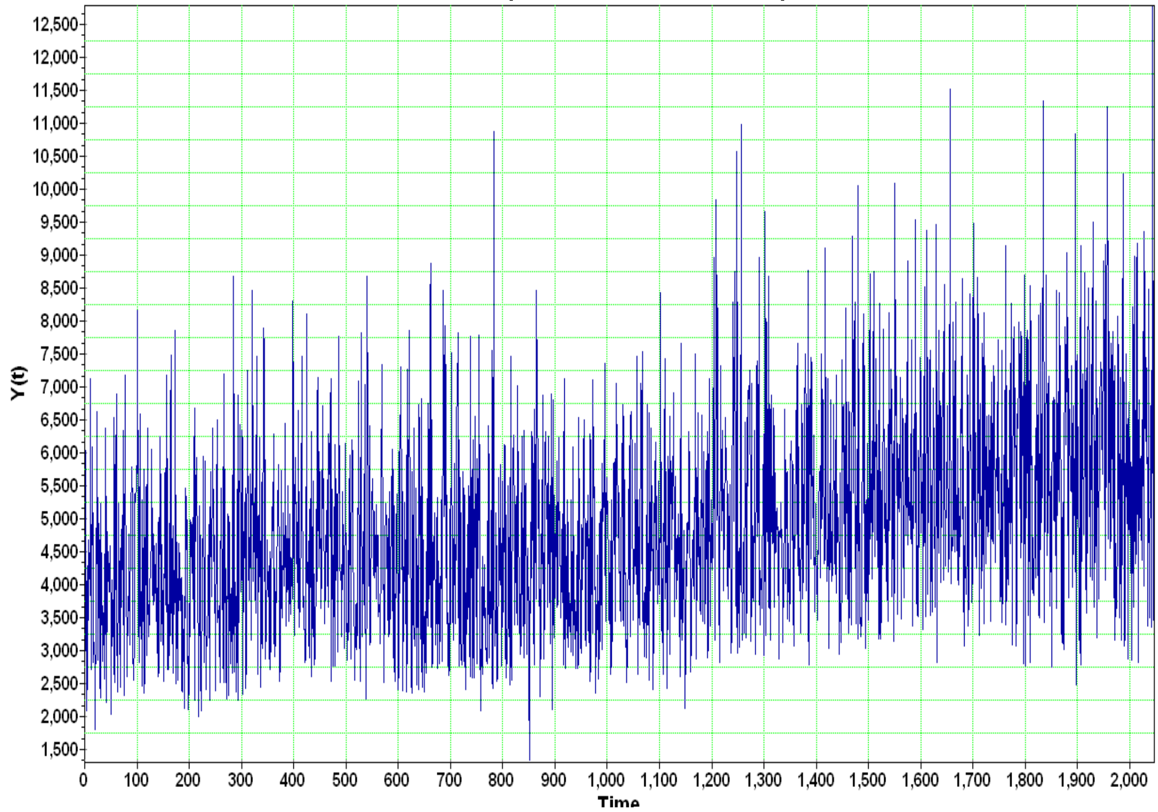
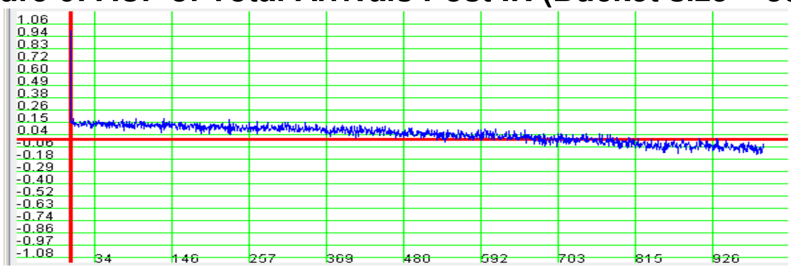



**Figure 6: ACF of Total Arrivals Post IR (Bucket size = 300).**




Now we estimate the Hurst Exponents (”H”) from this “cleansed” series using the various H estimators derived from two publicly available software, namely, SelQos[10] and SELFIS[11].Two common categories of estimators were used. The first category comprises estimators that use time-domain behaviour of the data set to estimate H. These are Rescaled Range (R/S), Absolute Momentum (A/M), Variance Method (VM), Variance of Residuals (VoR) and Modified Allan Variance (MAV). The second category comprises estimators that use frequency domain behaviour to estimate H. These are Periodogram (PDM), Local Whittle (Lwhittle) and Abry-Veitch Method (AV). Each of these methods have their own strength and weaknesses depending on the behaviour of the data and even after having rid the data series of short term “contaminants”. Further, note that both SelQoS and SELFIS implement these estimators differently as is evident in the outputs portrayed in Table 1. Combining the outputs from both the SelQoS and SELFIS implementations and also combining the all the H estimates (i.e. time and frequency domains) we have a Combined Median of 0.627 and Combined Mean of 0.683.

**Table 1: A summary of estimated H-Exponents using various estimators implemented in SelQoS and SELFIS**

| Estimators | SelQoS | SELFIS | Combined |
|---|---|---|---|
| **Time Domain** | | | |
| **R/S** | 0.8703 | 0.483 | |
| **A/M** | 0.8092 | 0.389 | |
| **VM** | 0.7737 | 0.893 | |
| **VOR** | 1.0759 | 0.841 | |
| **MAV** | 0.4632 | N/A | |
| **Median** | 0.8092 | 0.662 | 0.825 |
| **Mean** | 0.79846 | 0.652 | 0.733 |
| **Freq Domain** | | | |
| **PDM** | 0.5889 | 0.627 | |
| **Lwhittle** | 0.8236 | 0.607 | |
| **AV** | 0.4815 | 0.517 | |
| **Median** | 0.5889 | 0.607 | 0.598 |
| **Mean** | 0.6313 | 0.584 | 0.608 |
| | | | |
| **Comb Time-Freq** | | | |
| **Median** | 0.8038 | 0.627 | 0.627 |
| **Mean** | 0.7494 | 0.630 | 0.683 |
| **Max** | 1.0759 | 0.893 | |
| **Min** | 0.4632 | 0.389 | |

The SELFIS implementation of the various estimators was criticized in Ramirez-Pacheco et. al (2005) and Ramirez-Pacheco and Torres-Roman (2006) for consistently underestimating the H exponent for the various estimators under various controlled conditions. If one ignores the SELFIS results, and just taking SelQoS implementations we have the following extracted from Table 1, reconstituted in Table 2.

**Table 2: A summary of estimated H-Exponents using various estimators implemented in SelQoS**

| Estimators | SelQoS |
|---|---|
| **Time Domain** | |
| **Median** | 0.8092 |
| **Mean** | 0.79846 |
| **Freq Domain** | |
| **Median** | 0.5889 |
| **Mean** | 0.6313 |
| **Comb Time-Freq** | |
| **Median** | 0.8038 |
| **Mean** | 0.7494 |

Note that the median and the mean of all the H-exponent estimates in SeqQoS, the exception of the AV H-estimator, are, to a varying degree, greater than 0.5 where H = 0.5 implies "random walk" and the degree in which H>0.5 implies the degree of inherent Long Range Memory embedded in the data series. We may infer from Table 2 that there is strong suggestion of the presence of LRD in the Total Tourist Arrival data.

### 5.2 Tourist Arrivals from Australia, China, Japan, UK, USA, Rest of the world

Applying similar procedures outlined in the last section to various sub-components of the Total Arrivals, between 1/9/1997 and 30/10/2010, we have the estimated the following H-exponents after attempting to eliminate all signs of short-term trend, periodicities and cycles using the appropriate internal randomization bucket size ( i.e. "critical" bucket size). The "critical" bucket size is the approximate bucket size used in the IR process when all the short-term trends, periodicities and cycles are believed to have been eliminated in the IR process. In this case, all data sets seemed to reach this "critical" level at a bucket size of about 300. The transformed set of data is then used to estimate H using the various H-estimators provided by SelQoS. The results are summarized in Table 3.

**Table 3: A Summary of Time and Frequency Domain Estimates of H-exponents**

| Source | Time Domain | | | | | | Frequency Domain | | | | | Combined Time-Frequency Domains | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | R/S | A/M | VM | MAV | Median | Mean | PDM | Lwhittle | AV | Median | Mean | Median | Mean | Std-Dev |
| Australia | 0.8325 | 0.7645 | 0.7207 | 0.5307 | 0.7426 | 0.7121 | 0.6880 | 0.8483 | 0.5331 | 0.6880 | 0.6898 | 0.7207 | 0.7025 | 0.1296 |
| China | 1.0350 | 0.9654 | 0.9284 | 0.7522 | 0.9284 | 0.8820 | 0.7643 | 0.9963 | 0.4936 | 0.7643 | 0.7514 | 0.8464 | 0.8167 | 0.1888 |
| Japan | 0.6526 | 0.5581 | 0.5461 | 0.5970 | 0.5775 | 0.5884 | 0.6351 | 0.5954 | 0.5151 | 0.5954 | 0.5819 | 0.5954 | 0.5856 | 0.0491 |
| UK | 0.8182 | 0.7446 | 0.7221 | 0.5722 | 0.7334 | 0.7143 | 0.5641 | 0.7323 | 0.4678 | 0.5641 | 0.5881 | 0.7221 | 0.6602 | 0.1259 |
| USA | 0.7980 | 0.9408 | 0.9102 | 1.2884 | 0.9102 | 0.8830 | 0.8280 | 1.2968 | 0.8472 | 0.8376 | 0.8376 | 0.8472 | 0.8648 | 0.0591 |
| RoW | 0.8504 | 0.7778 | 0.7610 | 0.5599 | 0.7694 | 0.7373 | 0.5909 | 0.6504 | 0.5297 | 0.5909 | 0.5903 | 0.6504 | 0.6743 | 0.1230 |
| Aggregated | 0.8701 | 0.8153 | 0.7865 | 0.5681 | 0.8009 | 0.7600 | 0.6541 | 0.8146 | 0.5195 | 0.6541 | 0.6627 | 0.7865 | 0.6627 | 0.1477 |

### 5.3 Findings

From Table 3, the following results can be discerned.

#### 5.3.1 Australia Arrivals

Mean H-exponent, for Time Domain estimators, is 0.7121 and Median is 0.7426.
Mean H-exponent, for Frequency Domain estimators, is 0.6880 and Median is 0.6898. Combined Time-Frequency Domains Mean is 0.7025 and Median 0.7027.

Any future examination of Australian Arrival time series will have to contend with the strong presence of LRD. One may infer from the strong H-exponent (H>> 0.5) that there is strong "persistence" and "long memory". The "Joseph Effect" is highly likely to be operational and that good years beget good years and adverse years beget adverse years, although the length of the good-bad cycle is yet to be determined. Given the importance of Australian tourists to the New Zealand economy, the likely operations of the Joseph Effect can have policy implications.

#### 5.3.2 China Arrivals

Mean H-exponent, for Time Domain estimators, is 0.8820 and Median is 0.9284. Mean H-exponent, for Frequency Domain estimators, is 0.7514 and Median is 0.7643. Combined Time-Frequency Domains Mean is 0.8167 and Median 0.8464

Again any future examination of Chinese Arrival time series will have to contend with the presences of LRD. One may infer from the strong H-exponent (H>> 0.5) there is strong persistence – relatively stronger than the Australian case on the average. The "Joseph Effect" is highly likely to be operational and that good years beget good years and adverse years beget adverse years.

#### 5.3.3 Japan Arrivals

Mean H-exponent, for Time Domain estimators, is 0.5884 and Median is 0.5775.
Mean H-exponent, for Frequency Domain estimators, is 0.5819 and Median is 0.5954. Combined Time-Frequency Domains Mean is 0.5856 and Median 0.5954. Although the H-exponent estimates for Japanese arrivals are on the whole relatively smaller than both the Australian and Chinese estimates, they are still somewhat above 0.5. The important observation is that they are all consistently above the 0.5 threshold. Again, any future examination of Japanese Arrival time series will have to contend with the presences of LRD. The evidence of persistence implies the likely operation of the "Joseph Effect."

#### 5.3.4 UK Arrivals

Mean H-exponent, for Time Domain estimators, is 0.7143 and Median is 0.7334.
Mean H-exponent, for Frequency Domain estimators, is 0.5881 and Median is 0.5641. Combined Time-Frequency Domains Mean is 0.6602 and Median 0.7221.
Here, UK Arrivals again all exhibit H-exponent estimates of greater than 0.5. Again Joseph effect is highly likely to be operational with the same policy implications.

### 5.3.5 USA Arrivals

Mean H-exponent, for Time Domain estimators, is 0.8830 and Median is 0.9102. Mean H-exponent, for Frequency Domain estimators, is 0.8376 and Median is 0.8376. Combined Time-Frequency Domains Mean is 0.8648 and Median 0.8472. Arrivals from USA all exhibit H-exponent estimates of greater strength and consistency than all other arrival components. Joseph effect is highly likely to be operational with the same policy implications.

### 5.3.6 Rest of the World Arrivals

Mean H-exponent, for Time Domain estimators, is 0.7373 and Median is 0.7694 Mean H-exponent, for Frequency Domain estimators, is 0.5903 and Median is 0.5909. Combined Time-Frequency Domains Mean is 0.6743 and Median 0.6504 Arrivals from the Rest of the world, all exhibit H-exponent estimates of greater strength than 0.5. Joseph effect is likely to be operational with the same policy implications.

### 5.3.7 Aggregated Arrivals from all Sources

Mean H-exponent, for Time Domain estimators, is 0.7600 and Median is 0.8009 Mean H-exponent, for Frequency Domain estimators, is 0.6627 and Median is 0.6541. Combined Time-Frequency Domains Mean is 0.6627 and Median 0.7865. Arrivals from all Sources, exhibit H-exponent estimates of significantly greater strength than 0.5. Joseph effect is highly likely to be operational with the same policy implications.

## 6. Conclusions and Limitations

As is evident from the strengths of all the estimates of H-exponents, one may infer the following from the results presented:

1. In attempting to elicit information from this set of data, one has to be extremely careful when using statistics that rely on Gaussian and Poisson assumptions for the purpose. Pre-processing of the data set to determine the existence of LRD may be necessary in order to take into account the effects of LRD on statistical conclusions. The existence of IRD may render any statistical conclusions spurious.

2. One may infer that Joseph Effect may be in operation if one finds LRD in a reasonably “sanitized” data set, i.e. a data set that’s rid of short term trend, cycles and other periodicities. Internal Bucket shuffling is one of the methods of minimizing such impediments. One may conclude from this set of data that the Joseph Effect is quite prevalent.

3. At the moment although one may infer the existence of Joseph Effect, we are unable to determine the duration of such effect. This is subject to ongoing research. Each of the methods used suffers from its own respective weaknesses in estimation resulting from individual assumptions of the nature of the data. However, it is hoped that by using many approaches to estimating H after attempting to eliminate the

common impediments one could reasonably conclude the presence of LRD. The limitation here is that we could be overly conservative in detecting the presence of LRD.

## Endnotes

[1] SelQoS is the effort of JCR Pacheco, DT Roman, LE Vargas and OER Ferrera who have kindly provided me with the SelQoS software, a description of which can be found in Ramírez-Pacheco J C, Torres-Román D, Estrada-Vargas L, 2005. “Good and efficient hurst estimates with SELQOS” Retrieved August 2010 from: http://biblioteca.coqcyt.gob.mx/bvic/captura/resyaz.php?busqueda=GOOD%20AND%20EFFICIENT%20HURST%20ESTIMATES%20WITH%20SELQOS&where=meta&atributos=2&formato=2&tipo=normal&Zservers[0]=localhost:9999&Cantidad=5

[2] Detail of this publicly available software can be found in: Karagiannis T and Faloutsos M, 2002. “SELFIS: A Tool For SelfSimilarityand LongRange Dependence Analysis.” Retrieved from Microsoft Research on 21 July 2010 from: http://research.microsoft.com/apps/pubs/default.aspx?id=71485

[3]Hurst HE. 1951 “Long-term storge capacity of reservoirs.” *Trans. Am. Soc. Civil Engineers, 116, 770-799.*

[4] Granger CWJ and Newbold 1977, Forecasting EconomicTime Series, Academic Press, New York.
[5] Pacheco, J C R, Roman, D, Vargas L. 2005. Retrieved: 1 September 2010, from http://biblioteca.coqcyt.gob.mx/bvic/Captura/upload/Good-and-efficient-hurst-ANACON.pdf
[6] Taqqu, Teverovsky and Willinger (1999).

[7] Physical sciences and mathematics research in this area are too many to cite here, as a simple internet search using keywords such as “long run memory”, “persistence”, “Hurst Exponents”, etc. will testify.

[8] Mandelbrot B, and Hudson R. 2004, *The (Mis)behaviour of Markets,* Basic Books, New York. “*Persistent” time series (i.e. time series that have 0.50<H<1.00) tend to exhibit abrupt, and discontinuous changes whereas a normal distribution assumes continuous changes in a system. Thus, for a time series that exhibits Hurst statistics and abruptly change levels, i.e. skipping values either up or down, Mandelbrot coined the term "Noah effect," after the biblical story of the deluge.*

*For a time series that exhibit H between 0 and 0.5 then the series have fluctuations larger and more than what are expected to be normal random movements (i.e. “anti-persistence”). If H is 0.5, then the movements are said to be random movements. If H is between 0.5 and 1, the fluctuations are thought to be part of a long-term trend."Joseph Effect" alludes to an Old Testament story about Joseph’s Egypt experiencing seven years of feast followed by seven years of famine.*

[9] This section is adapted from: Coleman, R, “Research Article, Long Memory of Pathfinding Aesthetics”, *International Journal of Computer Games Technology, Vol. 2009, Article ID318505, page 3.*
[10] SelQoS is the effort of JCR Pacheco, DT Roman, LE Vargas and OER Ferrera who have kindly provided me with the SelQoS software, a description of which can be found in Ramírez-Pacheco J C, Torres-Román D, Estrada-Vargas L, 2005. “Good and efficient hurst estimates with SELQOS” Retrieved August 2010 from: http://biblioteca.coqcyt.gob.mx/bvic/captura/resyaz.php?busqueda=GOOD%20AND%20EFFICIENT%20HURST%20ESTIMATES%20WITH%20SELQOS&where=meta&atributos=2&formato=2&tipo=normal&Zservers[0]=localhost:9999&Cantidad=5

Pacheco JCR, Roman DT, Vargas LE and Ferrera OER. 2008. “A High Performance Tool for Long-Memory and Self-similarity Analysis” Retrieved 1 September 2010 from:http://biblioteca.coqcyt.gob.mx/bvic/Captura/upload/A-HIGH-PERFORMANCE-ANACON.pdf.

---

[11] Detail of this publicly available software can be found in: Karagiannis T and Faloutsos M, 2002. “SELFIS: A Tool For SelfSimilarityand LongRange Dependence Analysis.” Retrieved from Microsoft Research on 21 July 2010 from: http://research.microsoft.com/apps/pubs/default.aspx?id=71485

## References

Allan, DW, ‘Statistics of Atomic Frequency Clocks’, *Proceedings of the IEEE*, vol. 54, pp.221-30.

Aydogan, K and Booth, GG 1988, ‘Are There Long Cycles in Common Stock Returns’, *Southern Economic Journal*, 55(July), 141-149.

Beran, J 1992, ‘Statistical Methods for Data with Long –Range Dependence’, *Statistical Science*, vol. 7, pp 404-16.

Coleman, R ‘Research Article, Long Memory of Pathfinding Aesthetics’, *International Journal of Computer Games Technology*, Vol. 2009, Article ID318505, page 3.

Geweke, J and S. Porter-Hudak, 1983, ‘The estimation and application of long memory time series models’, *Journal of Time Series Analysis*, 4, 221–238

Granger CWJ and Newbold 1977, Forecasting EconomicTime Series, Academic Press, New York.

Green, M and Fielitz, B 1977, ‘Long term dependence in common stock returns’, *Journal of Financial Economics*, 4, 339-349

Hurst, H 1951, ‘Long term storage capacity of reservoirs’, *Transactions of the American Society of Civil Engineers*, 116: 770-799

Karagiannis T and Faloutsos M 2002, ‘SELFIS: A Tool For SelfSimilarityand LongRange Dependence Analysis’, Software retrieved from Microsoft Research on 21 July 2010 from: *http://research.microsoft.com/apps/pubs/default.aspx?id=71485*

Karagiannis, T, Molle, M and Faloutsos, M 2006, ‘Understanding the Limitations of Estimation Methods for Long-Range Dependence’, *UCR-CS-2006-10245*, November, 2006.

Karagiannis, T, Faloutsos, M and Reidi, R 2002, ‘Long Range Dependence: Now you seeit, now you don’t’, *IEEE GLOBECOM*, Global Internet Symposium, 2002.

Kunsh, HR 1987, *Statistical aspects of self-similar processes*, 1, 67–74, Utrecht, VNU Science Press.

Lo, AW 1991, ‘Long-Term Memory in Stock Market Prices’, *Econometrica*, 59, 5, 1279-1313.

Mandelbrot, B 1967, ‘The variation of some other speculative prices’, *Journal of Business*, 40, 393-413

Mandelbrot, B and Wallis, J 1968, ‘Noah, Joseph and operational hydrology’, *Water Resources Research*,4, 909-918.

Mandelbrot, B 1971, ‘When can a price be arbitraged efficiently? A limit to the validity of the random walk and martingale models’, *Review of Economics and Statistics*, 53, 225-236.

Medeiros, C, McAleer, M, Slottje, D, Ramos, V and Rey-Maquieira, J 2008, *Journal of Econometrics*, 147 (2008) 372-383.

Pacheco, JCR, Roman, D and Vargas, L 2005, Retrieved: 1 September 2010, from *http://biblioteca.coqcyt.gob.mx/bvic/Captura/upload/Good-and-efficient-hurst-ANACON.pdf*

Pacheco, JCR, Torres-Román, D and Estrada-Vargas, L 2005, 'Good and efficient hurst estimates with SELQOS', Software retrieved August 2010 from: *http://biblioteca.coqcyt.gob.mx/bvic/captura/resyaz.php?busqueda=GOOD%20AND%20EFFICIENT%20HURST%20ESTIMATES%20WITH%20SELQOS&where=meta&atributos=2&formato=2&tipo=normal&Zservers[0]=localhost:9999&Cantidad=5*

Pacheco, JCR and Torres-Roman, D 2006, 'A Tool for Long-range Dependent Analysis via the R/S statistic', *Proceedings of the 15th International Conference on Computing (CIC'06)*, 2006 IEEE.

Pacheco, JCR, Roman, DT, Vargas, LE and Ferrera, OER 2008, 'A High Performance Tool for Long-Memory and Self-similarity Analysis', Retrieved 1 September 2010 from:*http://biblioteca.coqcyt.gob.mx/bvic/Captura/upload/A-HIGH-PERFORMANCE-ANACON.pdf.*

Peters, EE 1994, *Fractal Market Analysis*, Wiley, New York

Taqqu, MS and Teverovsky, V 1996, 'Semi-parametric graphical estimation of long memory data', in P.M. Robinsion & M. Rosenblatt (Eds.), "*Athens conference on applied probability and time series*" Vol 2, Time series analysis in memory of E.J. Hannan. Springer.

Taqqu, Murad, S and Teverovsky, Vadim 1997, 'Robustness of Whittle-type estimators for time series with long-range dependence', *Stocahastic Models*, 13: 4, 723-757.

Taqqu, Teverovsky and Willinger 1999, 'Stock Market prices and long-range dependence', *Finance and Stochastics*, Springer-Verlag, 3, 1-13.

Veitch, D and Abry, P 1998, 'A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence', *IEEE Transactions on Information Theory*, 45

Wallis, JR and Matalas, NC 1970, 'Small Sample Properties of H and K Estimators of the Hurst Coefficient H', *Water Resources Research*, 6, 6, 1583-1594.