Information Entropic Measures Applied to Multivariate Fractional Brownian Motion

Vivek Anand

Thesis

Submitted Towards Partial Fulfillment

of

BS MS Dual Degree Program

Under the guidance of

Dr. M. S. Santhanam

Assistant Professor

Indian Institute of Science Education and Research

Pune, India

Department of Mathematical Sciences

Indian Institute of Science Education and Research Pune

April 2011

Certificate

This is to certify that this dissertation entitled "Information Entropic Measures Applied to Multivariate Fractional Brownian Motion" submitted towards the partial fulfillment of the BS MS Dual Degree Program at the Indian Institute of Science Education and Research, Pune, represents original research carried out by Vivek Anand at the Department of Mathematical Sciences, Indian Institute of Science Education and Research Pune under the supervision of Dr. M. S. Santhanam during the academic year 2010-2011.

Vivek Anand

Supervisor	Head, Mathematical Sciences
Date:	Date:
Place:	Place:

Acknowledgments

This study would not have been possible without the support and guidance of several people.

First, I would like to express my gratitude to Dr. M. S. Santhanam, who had not only been my guide and mentor for this final year project, but has also supported me during the period when I was preparing for the standerdised tests and my application process for my graduate studies. His expertise and insights in my field of study has been the foundation stones on which I carried out my investigations. Also, he always bore with my constant trespassing on his hectic schedule and came up with interesting ideas and suggestions regarding the project. In the past year, he helped me negotiate the complexities of eigenvector statistics while working on this project and never gave up on me.

Next, I sincerely thank Prof. L. S. Shashidhara, Dr. Farhat Habib and Dr. Neelesh Dahanukar for their constant encouragement in all my academic endeavors althrough my stay at IISER. Without their support, all this work would not have been possible.

Lastly, I thank my family for their constant endurance, my friends (Bedartha, Ashutosh, Smriti, Shambhavi and Parivesh) with whom I have had many insightful discussions during the course of this project and GOD almighty (Thakur, Maa and Swami Ji) who helped me to pick myself up even when the going got tough.

Abstract

Time series analysis gives us a window to look at the past events and make predictions about the future. It has been long since it was discovered that various natural process exhibit a long memory property, characterized by the Hurst parameter H. The main goal of this project is to extract significant information contained in large correlated multivariate time series in terms of information entropic measures. The data was projected onto principal components (using PCA) where maximum variance of the data was captured by information entropic measures. In this thesis we study the variation of the information entropy of the the top principal components (PCs) with a variation in H and find that as the value of H increases, the net information entropies of the top PCs decrease, indicating an increment in the amount of variation in top PCs as H increases.

Contents

List of Figures

Chapter 1

Introduction

Long memory processes are ubiquitous in natural and socio-economic processes [1, 2]. Use of long memory models started with Harold Hurst documenting the long term storage capacity of reservoirs [3]. Econometricians started using long memory models since around 1980s [4, 5]. Recently, long memory models found their applications in financial research, with these models efficiently capturing the essence of the various financial processes like inflation and exchange rate, properties of stock returns for long times etc. [6, 7]

In this thesis we study the properties related to the 'memory' of a time series using Principal Component Analysis (PCA) method. We use fractional Brownian motion (fBm) with varying values of Hurst exponent, denoted by H, as a prototype of time series with memory and then we explore the properties of the principal components for different values of H. We also verify the conjecture that the eigenvalue spectrum of PCA of fBm process is dependent on H and follows the power law

$$\lambda_n \sim n^{2H+1}. \tag{1.1}$$

i.e. the relation between the eigenvalues that we obtain upon performing the PCA and the index when plotted on log-log scale is a straight line with slope equal to

$-(2H+1)$. Figure 1.1 shows an example of the power law distribution of eigenvalues for $H = 0.50$ and figure 1.2 shows the power law distribution of figure 1.1 on log log scale.

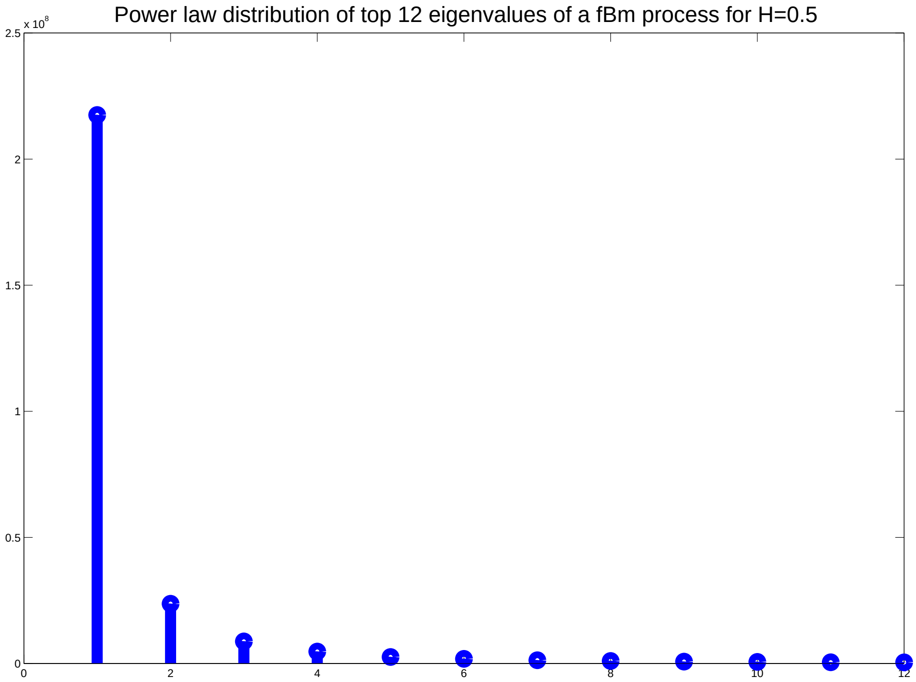


Figure 1.1: Power law distribution of top 12 eigenvalues of a fBm process for H=0.50. It is evident from the figure that the remaining eigenvalues are very small compared to the top eigenvalues.

The rest of the thesis has been organized as follows. In chapter 2 we walk through time series with memory, with an emphasis on long-memory. We also characterize the long memory with relation between the Hurst exponent and persistence and anti-persistence properties of time series data. We also discuss, compare and contrast the tools that we used to determine the Hurst exponent of the data. Chapter 3 discusses Principal Component Analysis (PCA), methodology of computing principal components (PCs) and the usefulness of PCA in data analysis. Chapter 4 deals with the concept of information entropy and we elaborate how information entropy can be used as proxy of the information content of the data. Chapter 6 explains the

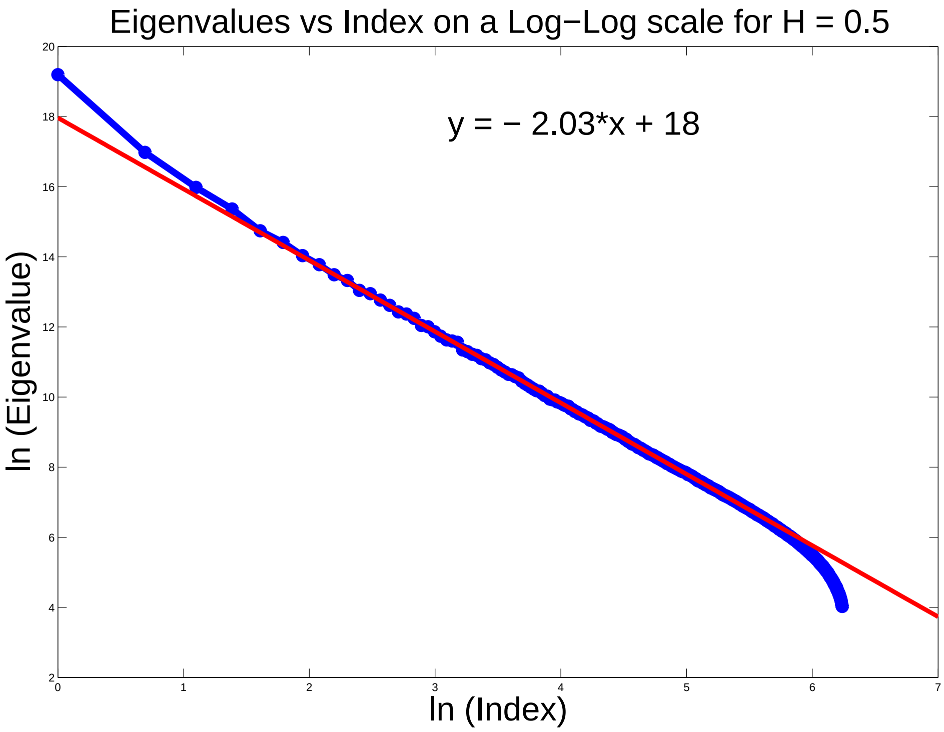


Figure 1.2: Power law distribution of the eigenvalues of a fBm process for H=0.50 on a log log scale. The red line is the regression line whose equation is given by $y = -2.03 * x + 18$

algorithm implemented to generate the simulated data. It also elaborates how the simulated data (here fBm) can be used as a prototype of a process with long memory. Here we perform PCA on the real time financial data and compare the results data vis-a-vis those obtained from the simulated data. The chapter also analyses the results obtained in terms of information entropic measures and concludes with the implication the result and the future course of action.

Chapter 2

Long Memory and Its Characterization

The odds and ends of long memory has long engaged the scientists since its first use by Harold Hurst in 1950s [3] that explained the effects of "Hurst Phenomenon" on the level of water in the Nile river. Today, diverse fields ranging from Hydrology, Climatology [2] to econometrics and finance [1] use long memory models.

Informally, autocorrelation of time series is the cross correlation with itself, *i.e.* the similarity between the observations as a function of time separation between them. A random stationary process whose integral of the auto-correlation function diverges is said to have long memory, *i.e.* the autocorrelation function $\rho(k) = Cov(X_i, X_{i+k})/Var(X_i)$ of a stationary process having long memory property holds: $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$. This implies that the autocorrelation functions decays asymptotically as a function of time lag k. For instance, a random process with an autocorrelation function assuming a power law of the form $\tau^{-\alpha}$ with $\alpha < 1$ will have long memory. This is because the autocorrelations decay to zero so slowly (at a hyperbolic rate) that their sum does not converge [2, 8].

Of several ways that describe long memory, the widespread definition is in the terms of autocovariance function $\gamma(k)$, which is the covariance of the time series with the time-shifted version of itself. A process is defined as 'long memory process' if in the limit $k \to \infty$

$$\gamma(k) \sim k^{-\alpha} L(k) \tag{2.1}$$

where $0 < \alpha < 1$ and $L(k)$ is a function[1] such that,

$$\lim_{x \to \infty} \frac{L(cx)}{L(x)} = 1 \tag{2.2}$$

The notation $x_n \sim y_n$ means that $x_n/y_n = 1$ as $n \to \infty$. The exponent α is the long memory exponent. The smaller the α, the longer the memory and vice versa. Often long memory is discussed in terms of Hurst exponent, H. The relation between H and α for a long memory process is given by

$$H = 1 - \frac{\alpha}{2} \Rightarrow \alpha = 2 - 2H \tag{2.3}$$

Processes that have short memory have H in the range $(0, 0.5)$ resulting in α being greater than 1 and thus, their autocorrelation function decays faster than k^{-1}. For a positively correlated time series, the Hurst exponent H lies in the interval (0.5, 1), resulting in long memory property.

2.1 Measuring Hurst Exponent

An attempt to empirically determine the long memory property of a time series is a daunting task. The main reason for this problem is the fact that testing for long memory requires large volume of data and quite often results in inconclusive or even conflicting results. However, several heuristic methods have been suggested to determine the long memory property of the investigated time series in the terms of its Hurst exponent. Here, in this thesis, we review the re-scaled (R/S) range analysis

[1] These type of functions are called slowly varying functions. Examples of slowly varying functions are $L(x) = log(x)$ or $L(x) = b$, where b is a constant.

and the detrended fluctuation analysis (DFA), the two most popular methods, which has been employed in this thesis to determine the Hurst exponent of the time series.

2.1.1 Estimating Hurst exponent from re-scaled range method

Suppose we have a stochastic process X_t at time points $t \in \tau = \{0, 1, \cdots, N\}$. We divide the time series of length N into A consecutive blocks each of length n such that endpoints of two neighboring blocks do not overlap; where $\{N, A, n\} \in I$. Now in every subinterval we modify the original datum X_t for location, using the slope of the series in the given subinterval. This is done by finding $X_t - (\frac{t}{n})(X_{an} - X_{(a-1)n})$ for all t with $(a-1)n \leq t \leq an$ for all $a = 1, 2, \ldots, A$.

Now, for any a^{th} subinterval, $I_a = [n(a-1), na]$, we construct the small possible box whose edges are parallel to the co-ordinate system and is such that it contains all fluctuations of $X_t - (\frac{t}{n})(X_{an} - X_{(a-1)n})$ that occur in the subinterval I_a. The length of the box is the length of the interval whereas the height of the box is given by

$$R_a = \max_{(a-1)n \leq t \leq an} \{X_t - (\frac{t}{n})(X_{an} - X_{(a-1)n})\} - \min_{(a-1)n \leq t \leq an} \{X_t - (\frac{t}{n})(X_{an} - X_{(a-1)n})\} \tag{2.4}$$

The construction of the boxes is depicted in the Figure **??**.

Let the empirical standard error of the variables between X_t and X_{t-1} be denoted by S_a, where S_a is the standard deviation of the interval. S_a is given by

$$S_a = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)} \tag{2.5}$$

where X_i's are the elements of time series contained in each box and μ is the mean of those elements. If S_a doesn't vary or varies slightly with a then the process is called stationary. If the process is not stationary, *i.e.*, there are large variations in S_a with changes in a, then dividing R_a by S_a rectifies the effects of scale inhomogeneity in

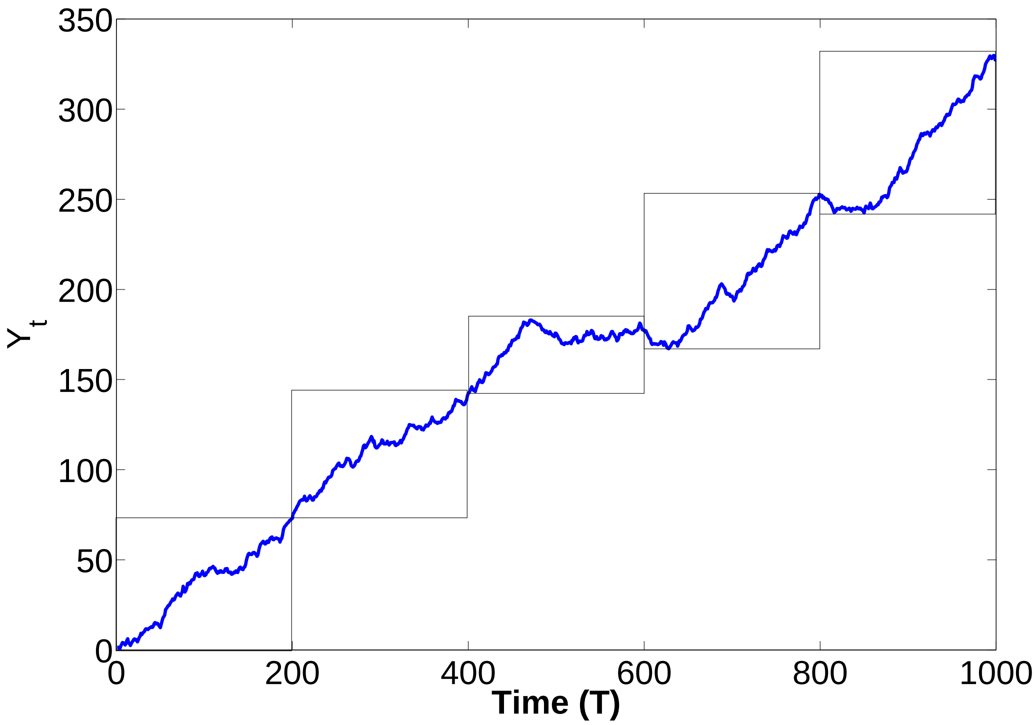


Figure 2.1: Construction of boxes for R/S analysis.

both temporal and spatial domains. In terms of n, the total area of the boxes after the correction for the scale is given by:

$$(\frac{R}{S})_n = A^{-1}\sum_{a=1}^{n}\frac{R_a}{S_a}. \tag{2.6}$$

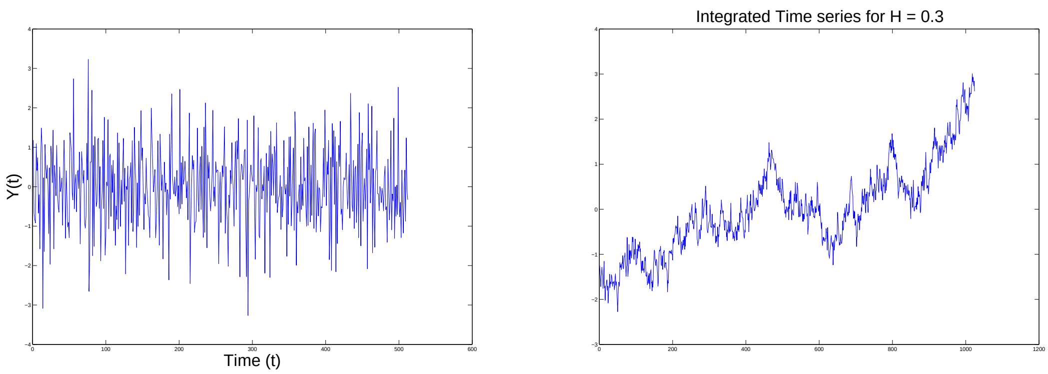
The estimator of the Hurst exponent, H is given by the slope of the regression of $\left(\frac{R}{S}\right)_n$ and n on a log-log scale for K values of n [2].

2.1.2 Estimating Hurst exponent from Detrended Fluctuation Analysis

Detrended Fluctuation Analysis, popularly known as DFA in scientific parlance, is a scaling analysis method that is used to estimate the self affinity of any signal and is useful in analysis of long memory property of time series and was introduced by

Peng. *et al.*[9]. The exponent obtained in DFA is very similar to Hurst exponent, except that DFA can also be used for non-stationary bounded time series. In this thesis we will be using bounded time series because for most physiological processes, their time series are bounded - it is rather unusual for them to have arbitrarily large amplitudes, irrespective of the length of the time series.

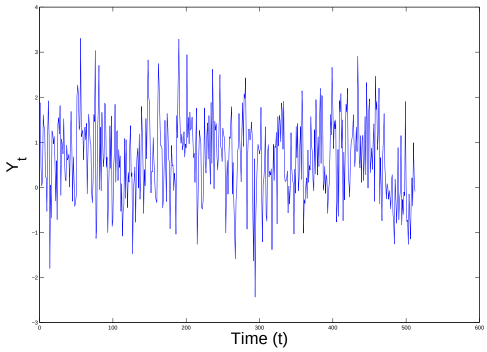
Though DFA works well for certain type of nonstationary time series, it is rather impossible for it to handle all possible nonstationarities in the real world data. Still, the property of DFA to detect intrinsic self similarity embedded in a seemingly nonstationary time series and avoiding the unintentional inclusion of apparent self-affinity (owing to some extrinsic trend) is something that makes makes DFA a robust method for analyzing the self similarity property of time series when compared to other conventional methods like R/S analysis. In order to determine the exponent the time series of length N is first integrated. This integration step maps the original time series into a self similar process. Figures 2.2 and 2.3 show the effects of integration.



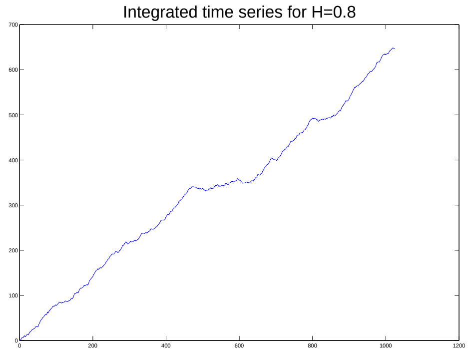
(a) Time series before integration (b) Time series after integration

Figure 2.2: Integration maps a time series into a self similar process

The integrated time series in then divided into K boxes of equal length i. A least squares line is fit to the data (which represents the trend in that box). Figure 2.4 explains the box construction and the line fitting in each box.



(a) Time series before integration



(b) Time series after integration

Figure 2.3: Integration maps a time series into a self similar process

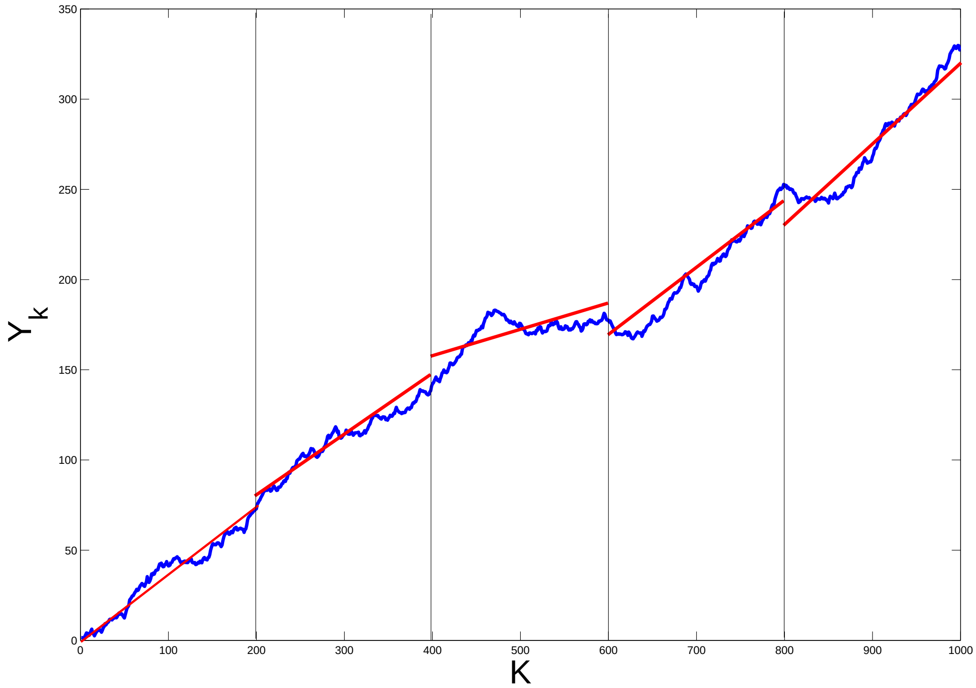


Figure 2.4: Box construction for DFA. The vertical lines indicate box of size $n = 200$ and the red line represents the'trend' estimated in each box by a linear least square fit.

The Y co-ordinate of the straight line segments is denoted by $Y_i(K)$. After line-fitting exercise, we detrend the integrated time series, Y_K, by subtracting the local trend $Y_i(K)$ in each box. The root-mean-square fluctuation or the characteristic size of fluctuation for this integrated and detrended time series is calculated by

$$F(i) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[Y(k) - Y_i(k)]^2}. \tag{2.7}$$

This computation is repeated over all box sizes (time scales) to characterize the relationship between $F(i)$ and the box size i. The regression plot of $F(i)$ and i on log-log scale is the scaling exponent α of the time series. The relationship between α and H is given as:

$$\alpha = \begin{cases} H \;\forall\; \alpha \;\in\; [0,1] \\ (H+1) \;\forall\; \alpha \;\in\; (1,2] \end{cases}$$

In this thesis we will mostly be using the DFA over R/S analysis because R/S analysis recently has been shown to overestimate H when compared to DFA, and thus considered less efficient [10, 11]. It has also been shown that in a random time series of length 2^9 to 2^{17}, the estimates of R/S analysis are significantly higher than 0.5 as compared to those of DFA which are very close to 0.5 [12]. Since the simulated data we have are of order of 2^{10} to 2^{12} whereas the real time data that we use are close to the order of 2^{17} to 2^{18}, we have preferred DFA.

Chapter 3

Principal Component Analysis

The central idea behind the Principal Component Analysis (PCA) is to reduce the dimension of a given data consisting of large number of intercorrelated variables, while preserving as much variation possible, as in the original data set. This is achieved by transforming the old interrelated variables to a new set of non-correlated variables, known as Principal Components (PCs). The PCs are ordered in a descending order where the top few PCs retain most of the variation present in the original data set.

3.1 Principal Components : Definition, Derivation and Properties

Suppose we have a vector Y with q time series $Y = \{y_i(t), y_2(t), \ldots, y_q(t)\}$. Unless q is small, it is not prudent to look at the q variances and $\frac{1}{2}q(q-1)$ covariances (or correlations) resulting from the covariance (or correlation) of each of the q time series with other $(q-1)$ time series. Alternatively, we can look for a few derived variables v, $v \ll p$, that preserve most of the information given by these variances and covariances (or correlation). Although PCA concentrates on variances, it doesn't ignore covariances and correlations. The first step to the PCA is to find a linear function $\alpha_1' y$ of the elements of y with maximum variance, where α_1 is a vector of q

constants $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1q}$ and denotes transpose. Thus we have

$$\alpha_1' y = \sum_{j=1}^{q} \alpha_{1q} y_j. \tag{3.1}$$

Next we look for another linear function $\alpha_2' y$ which is uncorrelated with $\alpha_1' y$ and has maximum variance. This continues to $\alpha_k' y$ where $\alpha_k' y$ again has a maximum variance is uncorrelated to $\alpha_1' y$, $\alpha_2' y$, $\ldots$, $\alpha_{k-1}' y$. We call $\alpha_i' y$ as the i^{th} PC. We can find up to q PCs. But in general, most of the variance in Y is accounted by the first m PCs, where $m << q$.

Having defined the principal components, we here elaborate the ways to find them. Let the previous case with q time series with known covariance matrix C. In our case, covariance matrix is a real symmetric matrix. In order to derive the form of principal components, let us consider first $\alpha_1' y$, the vector α_1 maximizes $Var[\alpha_1' y] = \alpha_1' C \alpha_1$. Since the elements of the eigenvectors are the weights associated with the variables, we have the condition $\alpha_1' \alpha_1 = 1$ *i.e.* the sum of the squares of the elements of α_1 equals 1. Also we have the condition $Max|\alpha_{1j}| = 1$. In order to maximize $\alpha_1' C \alpha_1$ (subject to the constraint $\alpha_1' \alpha_1 = 1$), we use the Lagrange multiplier method, *i.e.* maximize $\alpha_1' C \alpha_1 - \lambda(\alpha_1' \alpha_1 - 1)$ where λ is the Lagrange multiplier.

Differentiating with respect to λ_1, we have

$$C\alpha_1 - \lambda \alpha_1 = 0 \Rightarrow (C - \lambda I_q) = 0 \tag{3.2}$$

where I_q is the $q * q$ identity matrix. Hence λ is an eigenvalue of C and α_1 is the corresponding eigenvector. Thus the first principal components is the eigenvector of the covariance matrix with largest PC being eigenvector corresponding to the largest eigenvalue. Generalizing, the result for the other αs we have the other eigenvectors of the covariance matrix [13]. Some notable properties of the PCs that concern this thesis are:

i. PCs of real symmetric matrices are orthogonal.

ii. The largest PC accounts for the maximum variance.

iii. The sum of squares of the elements of PCs is equal to 1, *i.e.*, $\sum a_i^2 = 1$.

Other properties of the PCs are can be found in [13].

Chapter 4

Information Entropy and Information Content

Entropy has wide and multidisciplinary application and has been used in wide array of fields. Here we focus on its time series applications. Loosely speaking, entropy is a measure of the uncertainty that is associated with some random variable. In the context of information, it is regarded as the metric of information content or uncertainty of any stochastic event [14, 15].

Information content is the backbone of the information theory and mainly represents information measure and elimination of uncertainty in terms of information obtained. In this chapter we elaborate the concept of Shannon entropy and explain how Shannon entropy can be used as a proxy for information content of the data.

4.1 Shannon Entropy

Suppose some source independently emits a stream of n symbols $\{a_1, a_2, \ldots, a_n\}$, with respective probability being $\{p_1, p_2, \ldots, p_n\}$, where $\sum p_i = 1$. Shannon entropy tries to address the amount of information we get from each symbol in the stream. If we observe symbol a_i, we get $\log(1/p_i)$ information [1]. In a long run where N symbols are emitted, we will have $N * p_i$ occurances of a_i. Hence, with N observations, the total amount of information I that we will have is:

$$I = \sum_{i=1}^{n} (N * p_i) \log(\frac{1}{p_i}) \tag{4.1}$$

The average information per symbol, thus obtained is:

$$I/N = (1/N) \sum_{i=1}^{n} (N * p_i) \log(\frac{1}{p_i}) = \sum_{i=1}^{n} p_i \log(\frac{1}{p_i})$$

Hence the information entropy of a distribution is given by

$$-\sum_{i=1}^{n} p_i \log(p_i) \tag{4.2}$$

and is usually referred to as Shannon entropy.

4.1.1 Information Entropy as a Proxy of Information in Long Memory Process

Entropy increases with the degree of disorder and is maximum for absolutely random states. A time series with Hurst exponent $H = 1/2$ is considered as a random walk. For values of $H < 1/2$, the time series is anti-correlated whereas for values of $H > 1/2$, the time series is correlated. Thus, with the variation in the Hurst exponent, the information content of a time series varies. As the values of H increase, the information content in the time series, owing to strong correlations, increases. The principal components (PCs) or eigenvectors obtained after the PCA of the co-variance matrix obtained capture the variation in the data in a particular dimension.

[1] Shannon used the logarithm to provide additivity, characteristic for independent uncertainty.

For a given H, the eigenvector corresponding to the largest eigenvalue contains the maximum variance, and hence the maximum amount of information. As a result of this, its information entropy is expected to be the minimum. The information entropy is expected to increase as the eigenvalue decreases. This is because as we consider the lower rung eigenvalues, the variance captured by them decreases and hence the information content, and thereby increasing the information entropy.

Chapter 5

Fractional Brownian Motion

The fractional Brownian motion (fBm) is a continuous time is a stochastic process $\{X_t : t \in T\}$ for which any linear function applied to the sample function X_t will yield normally distributed results[1]. The fBm has zero mean and has stationary increments. Hurst exponent, H characterizes the correlation of the increments and has long range dependency property for values of H larger than 0.5. Unlike standard Brownian motion, this long range dependency property of fBm makes it an ideal model to study properties of processes with long memory. Even some empirical studies have modeled log return of financial time series using the fBm [16]. Fractional Brownian motion is said to have persistent correlations *i.e.* an upward jump is more likely to be followed by another upward jump or vice-versa when $1 > H > \frac{1}{2}$. For $0 < H < \frac{1}{2}$ the process is said to have anti-persistence property *i.e.* a jump up is more likely to be followed by jump down.

[1]This stochastic process is also known as Gaussian Process in the honor of German mathematician Carl Friedrich Gauss who pointed out the normal distribution, also known as Gaussian distribution

5.1 Generating a Fractional Brownian Motion

A fBm can be modeled as a process whose spectral density scales with its frequency f as a power law $f^{-\alpha}$. In order to generate such a process we follow the process suggested by Rangarajan and Ding [17]. We start with a discrete zero mean white Gaussian noise process[2] $\{\xi_K\}$, where $K = 0, 1, \ldots, (N-1)$, and variance σ^2. Upon performing Fourier transform, we obtain:

$$\Gamma_K = \sum_{n=0}^{N-1} \xi_n \exp\left(-i2\pi n \frac{K}{N}\right); K = 0, 1, \ldots, (N-1) \tag{5.1}$$

Γ_K is multiplied by the factor $f^{-\frac{\alpha}{2}} = (\frac{K}{N})^{-\frac{\alpha}{2}}$ to obtain a scaled quantity Γ'_K. The factor $f^{-\frac{\alpha}{2}}$ is chosen to ensure that the spectral density scales as $f^{-\alpha}$ with the frequency f. Finally an inverse Fourier transform is performed to obtain

$$x_n = \frac{1}{N} \sum_{K=0}^{N-1} \Gamma'_K \exp\left(2\pi n \frac{K}{N}\right); n = 0, 1, \ldots, (N-1). \tag{5.2}$$

The discrete process thus obtained has a mean power spectrum which scales with frequency as a power law $\frac{1}{f^\alpha}$.

5.2 Fractional Brownian Motion as a Prototype of a Process With Memory

In 2.1.2 we had seen the relationship between α and H. Gao *et al.* have shown that the power spectral density of a fBm should be $S(f) \sim f^{-(2H+1)}$, where H is the Hurst exponent [18].

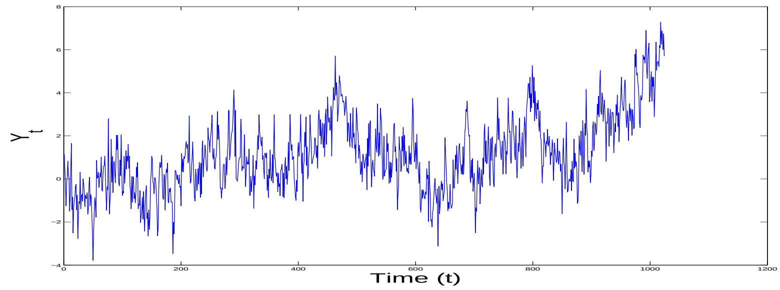
As we know, fractional Brownian motion is a zero mean Gaussian process with

[2] A normally distributed time series that has no correlation in time, and the Fourier transform of the the autocorrelation is flat.

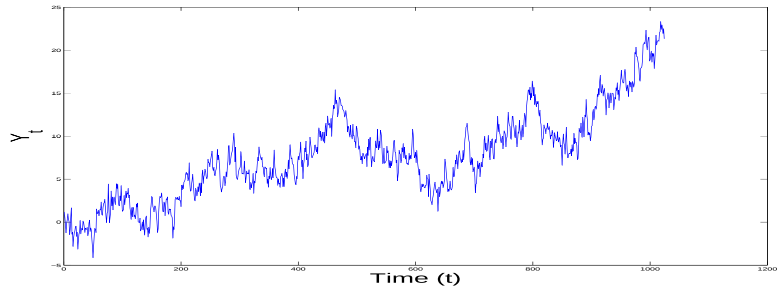
stationary increments whose covariance is given by

$$E[B_H(S)B_H(t)] = \frac{1}{2}\{S^{2H} + t^{2H} - |S - t|^{2H}\}. \tag{5.3}$$

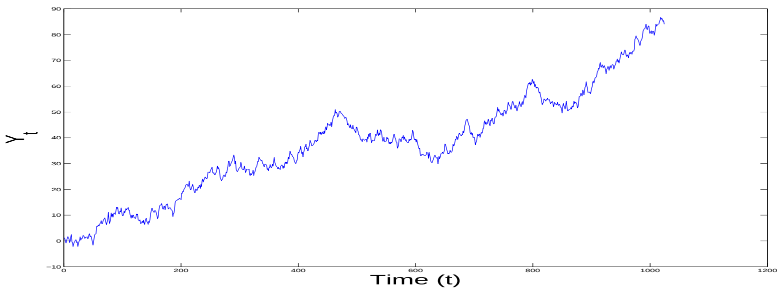
Now, when $H = \frac{1}{2}$, the fBm reduces gto standard Brownian motion. For values of $H < \frac{1}{2}$, the process has negatively correlated increments and is said to have anti-persistence property. For values of $H > \frac{1}{2}$ the process has persistence property and has positively correlated increments *i.e.* an upward jump is more likely to be followed by another upward jump and vice versa. It has been observed that with an increase in H the process becomes more trendy (smooth) and less irregular. Figure 5.1 shows various fBm processes with different values of H.



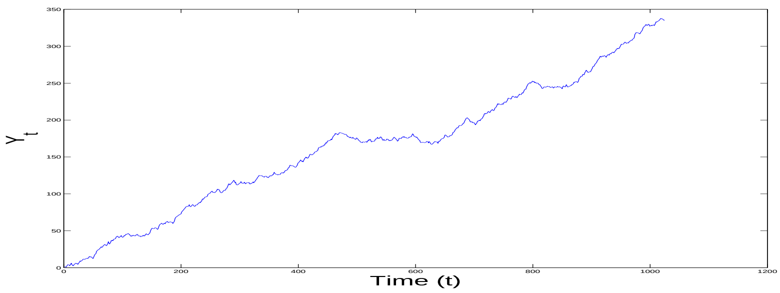
$H = 0.1$



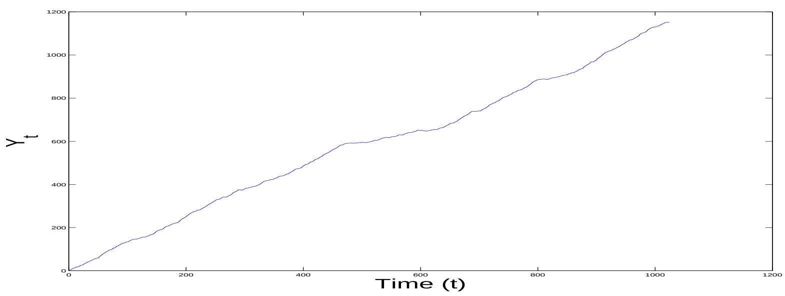
$H = 0.3$



$H = 0.5$



$H = 0.7$



$H = 0.9$

Figure 5.1: Several fBm process for different values of H

Chapter 6

Results and Discussions

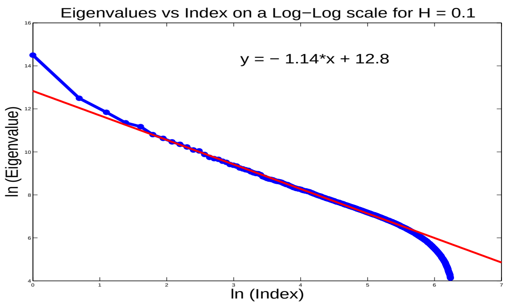
In this project we have tried to explore the variation in the information content of the principal components with the variation in H. To start with, we verified the conjecture by Gao *et al.* which states that since the power spectral density of a fBm process decays as $S(f) \sim \frac{1}{f^{2H+1}}$, the eigenvalue spectrum of the PCA decays as a power law [18]. We also investigated the properties of eigenvectors in terms of Shannon entropy and their dependence on H.

6.1 Eigenvalue Spectrum of the fBm Process

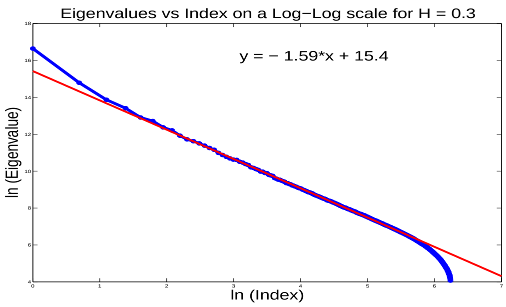
Conjecture: When n is large the eigenvalue spectrum from PCA of a fBm process with parameter H decays as a power law: $\lambda_n \sim n^{-(2H+1)}$.

Using the algorithm in 5.1, we generate fBm process with different H and computed the auto-covariance matrix of the simulated fBm process with different H. The eigen-analysis of these auto-covariance matrices do exhibit a power law decay in eigenvalue spectrum. Figure 6.1 shows the eigenvalue spectrum of various fBm processes with varying H.

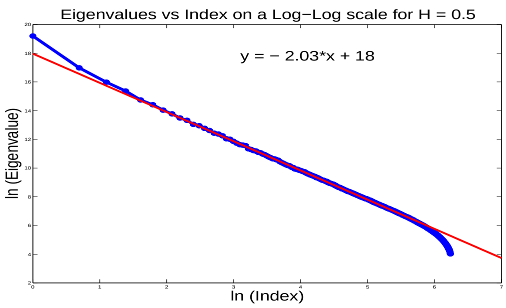
In every case, we fit a straight line to the eigenvalue spectrum on the log-log scale to find that the slope of the line is approximately equal to $[-(2H+1)]$. This verifies



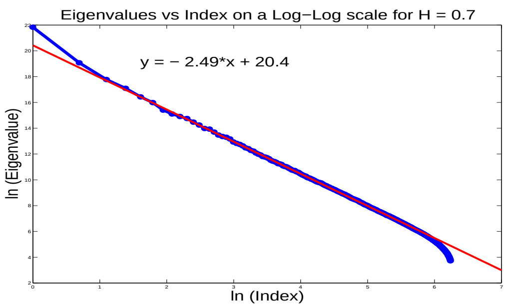
$H = 0.1$



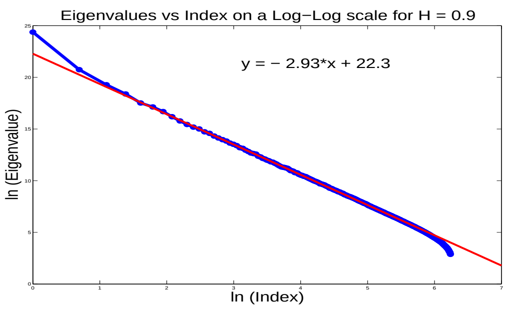
$H = 0.3$



$H = 0.5$



$H = 0.7$



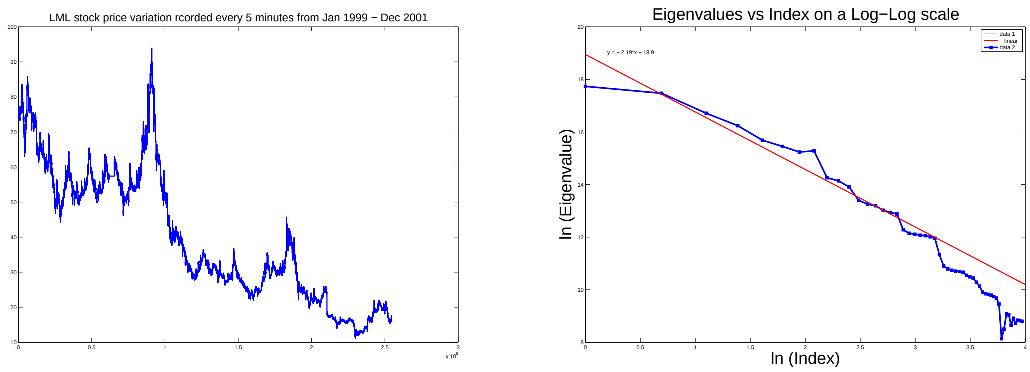
$H = 0.9$

Figure 6.1: Power law distribution for different values of H. The red line is the linear regression of the eigenvalues plotted on log-log scale.

the conjecture proposed by Gao *et al.*.

6.1.1 PCA of Stock Prices

It is assumed that the current market price contains all the information of the past and thus the time series of stock prices is assumed to have memory. Hence, we also tested the conjecture proposed by Gao *et al.* on real time data. We analyze the time series of the stock prices obtained from the Bombay Stock Exchange. We performed the DFA of over 150 time series of stock prices and grouped the stocks with similar Hurst exponents together. The stock prices were recorded from Jan 99 to Dec 01 at at 5 minute interval. We performed PCA on the autovariance matrix of the 53 stocks that had Hurst exponent close to 0.50 (actually in the range of $0.48 - 0.53$). One such stock is LML whose time series is shown in figure 6.2a. The eigenvalue spectrum of those stocks when plotted on log-log scale and fitted to a straight line had a slope of -2.19, indicating the power-law distribution and is shown in figure 6.2. Given the small number of stocks that we considered, the results can be considered to be in correspondence with the conjecture.



(a) (b)

Figure 6.2: Stock price variation and the eigenvalue spectrum. (a) Stock price variation of LML stock from Jan 99 to Dec 01 (b) The eigenvalue spectrum of stock prices for stocks with $H \approx 0.5$

6.2 Information Entropic Measures of Principal Components

Since for the values of $H < \frac{1}{2}$ time series is said to have negatively correlated increments. The extent of these negative correlations increase as the value of H decrease towards zero. For values of $H > \frac{1}{2}$, the time series has positively correlated increments. The extent of this correlation increases with increase in the value of H. This can also be observed from the Figure 5.1 that with the increase in the value of H, the fBm processes become smoother.

We know that the top few PCs obtained upon performing the PCA on the auto-correlation matrix of the fBm account for maximum variance in the data. Also, from the figure 6.3 we can see that with the increase in the values of H from 0 to 1, the persistence property of the time series increases *i.e.* with increase in the value of H, the time series starts to have positively correlated increments. Based on this observation we posit that as the values of H increases, the sum of information entropy of first few PCs (eigenvectors) decreases with increase in H.

To verify this we plot the mean of information entropy of eigenvectors of top 13 eigenvalues obtained from the auto-correlation matrix of simulated fBm process.The plot is shown in Figure 6.4.

We know that as the information entropy decreases with an increase in the information content. We also know that for low values of H, the data has negatively correlated increments which means the data has more noisy components compared to those with higher values of H. This means the few PCs of the fBm process with higher H have higher information content and thus their information entropy decreases with increase in H. For values of $H < \frac{1}{2}$, as H decreases, the long memory of data decreases and thus the information content of the data decreases, resulting in increase of information entropy.

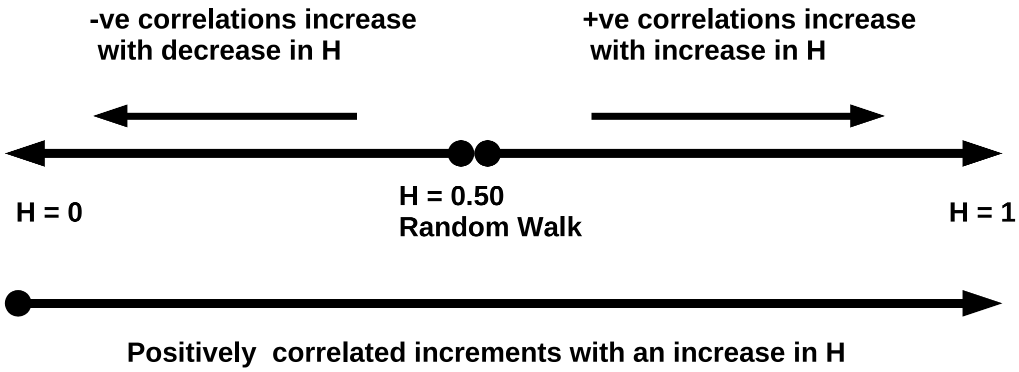


Figure 6.3: Variation of persistence with variation in H. As the value of H increases, the persistence increases.

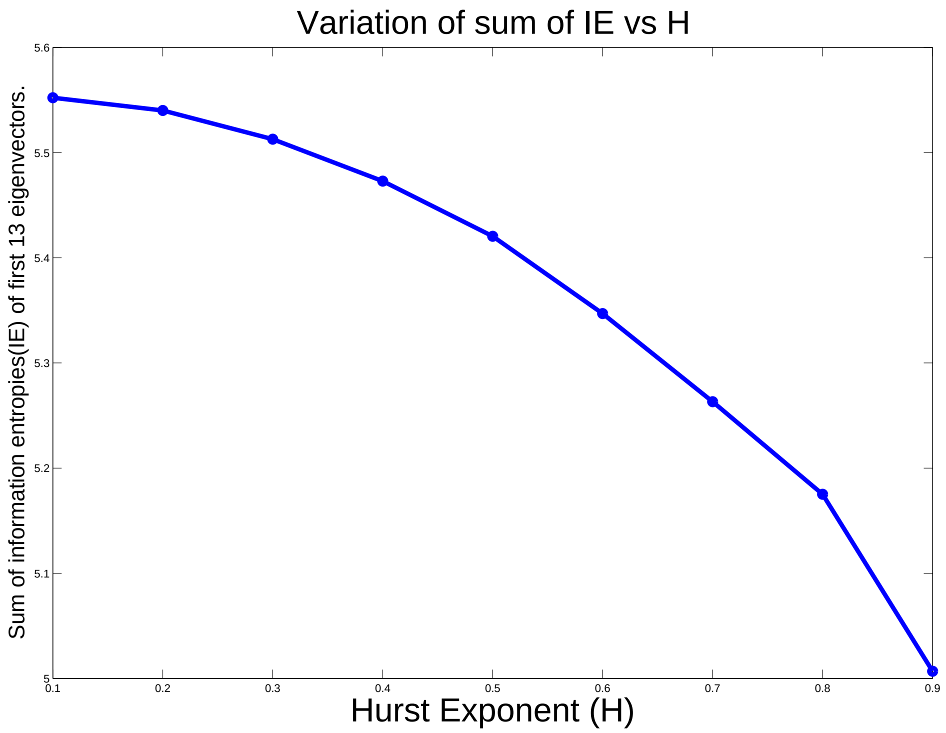


Figure 6.4: Variation of the sum of information entropies of top 13 PCs.

One point worth noting from figure 6.4 is that the sum of the information entropies of top 13 PCs shows significantly less variation for $H < \frac{1}{2}$. This is because, as the memory of the time series decreases significantly, the information content (variation) of the top eigenvectors is significantly low. For values of H between $0.1 - 0.2$, we find that the information entropy is almost constant, whereas it slowly starts decreasing for H between $0.3 - 0.4$ and decreases rapidly for $H > \frac{1}{2}$.
This verifies our claim that the information content in terms of 'weights' of the variables in the PCs decrease with a decrease in H and thus result in an increase in information content.

6.3 Discussions

In this thesis, we find that PCA can filter out the significant information contained large multivariate time series. We also found that, larger the correlations, more is the information contained in the top PCs. However, the cliche associated with the PCA is that it only finds orthogonal components that minimize the error in reconstructed data. The main limitation of PCA is that it only defines linear projections of the data and thus is not able to model the non-linear relationship amongst variables. And, given that we have used financial time series data, it is certain to have complex, non-linear relationship among the variables. However, in contrast to PCA, processes like independent component analysis (ICA) separate a multivariate signals into 'independent' additive subcomponents. Here 'independent' is used in the additive sense, *i.e.*, knowledge about one means does not give any information about the other. Thus, we believe that simultaneous use of PCA and ICA would yield better results as it has been shown that PCA enhances ICA performance by discarding the smaller eigenvalues before they whiten and thus reduce the computational complexity by minimizing the pair-wise dependencies [19].
Also, we can look how the total information entropy changes as a function of the

PCs considered, with the variation in H.

References

[1] R. T. Baillie, Long memory processand fractional integration in econometrics, Journal of Econometrics 73 (1996) 5–59.

[2] J. Beran, Statistics for Long-Memory Processes, Chapman and Hall, 1994, Ch. Stationary processes with long memory, pp. 41–65.

[3] H. Hurst, Long-term storage capacity of resorvoirs, Transactions of the American Society of Civil Engineers 116 (1951) 770–799.

[4] F. X. Diebold, G. D. Rudebusch, Long-memory and persistence in aggregate output., Journal of Monetary Economics 24 (1989) 180–209.

[5] C. W. J. Granger, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14 (1980) 222–238.

[6] R. T. Baillie, T. Bollerslev, Cointegration, fractional cointegration and exchange rate dynamics, Journal of Finance 49 (1994a) 737–745.

[7] R. T. Baillie, T. Bollerslev, Long memory in the forward premium, Journal of International Money and Finance 13 (1994b) 565–571.

[8] A. I. McLeod, A. W. Hippel, Preservation of the rescaled adjusted range, 1: A reassessment of the hurst phenomenon., Water Resour. Res. 14 (1978) 491–508.

[9] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, Mosaic organization of dna nucleotides, Physical Review E 49 (1994) 1685–1689.

[10] D. Grech, Z. Mazur, Can one make any crash prediction in finance using the local hurst exponent idea ?, Physica A 336 (2004) 133–145.

[11] L. Czarnecki, D. Grech, G. Pamula, Comparison study of global and local approaches describing critical phenomena on the polish stock exchange market, Physica A 387 (2008) 6801–6811.

[12] L. Kristoufek, R/s analysis and dfa: Finite sample properties and confidence intervals, ACTA PHYSICA POLONICA B 41 (2010) 1223–1236.

[13] I. T. Jolliffe, Principal Component Analysis, Springer, 2002, Ch. Chapter 1 - 4, pp. 1–77.

[14] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 379–423.

[15] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 623–656.

[16] B. Mandelbrot, Fractals and scaling in finance, discontinuity, concentration, risk, Springer, 1997.

[17] G. Rangarajan, M. Ding, Integrated approach to the assesment of long range correlation in time series data, Physical Review E 61 (2000) 4991–5001.

[18] J. B. Gao, Y. Cao, J. M. Lee, Principal component analysis of $\frac{1}{f_\alpha}$ noise, Physics Letters A 314 (2003) 392–400.

[19] C. Liu, H. Wechsler, Comparative analysis of independent component analysis (ica) for face recognition, International Conference on Audio and Visual Based Biometric Person Authentication.