CHAPTER 5

EMPIRICAL MODE DECOMPOSITION ANALYSIS OF FINANCIAL TIME SERIES

5.1. Introduction

In this chapter we present our work using EMD method to analyse and compare two financial time series using data from Indian and Hong Kong stock market. Through this work we establish the utility of Empirical Mode Decomposition as a tool for analysing financial time series. Most of this chapter draws from our published work (Guhathakurta et al. 2008)\(^{356}\).

Our study focuses on the question: “is there a similarity in the basic trading dynamics of different markets?” To analyse this we use the Hong Kong AOI and NIFTY indices of Hong Kong Stock Exchange, Hong Kong and National Stock Exchange, India, respectively for the study.

The reason for choosing these two indices is to compare the time series/characteristics of the stock returns of two different markets both qualitatively (because Indian market is very much an ‘emerging’ market while Hong Kong is reasonably efficient market), and culturally. The aspects under study are the nature of the probability distribution of the two indices, their phase distributions, amplitude distributions and also a comparison of the same.

\(^{356}\) Guhathakurta, Kousik, Mukherjee, Indranil and Roy Chowdhury, A. 2008, op cit
We first examine the graphical similarity of the probability distributions of the two financial return time series generated by the respective data. We examine the shape of the distributions as well as some other qualitative aspects and compare their similarity/dissimilarity.

We then apply a new technique for analysing the periodicity and properties of time series. Many “paradoxes” exist in standard decomposition of time signals. To avoid them, Huang et al. (1998)\textsuperscript{357} have developed a method, termed the Hilbert view, for studying nonstationary and nonlinear data in nonlinear dynamics. This method uses the empirical mode decomposition technique to generate finite number of Intrinsic Mode Functions (IMFs) that assume well behaved Huang–Hilbert Transform [Huang(1998)\textsuperscript{358}]. The results obtained from this technique can be further analysed to reveal more details about the time series. We generate the probability distributions of the phase and amplitude of the IMFs and then compare the same for the two different time series. Our results show a striking similarity between the two sets of results. This finding is significant, considering the geographical separation of the markets. These findings clearly indicate the possibility of examining the underlying characteristics of the financial market dynamics through this technique. Further, the results are encouraging enough to indicate that this technique may be utilised in building a platform for


\textsuperscript{358} \textit{ibid}
understanding asset price behaviour ultimately leading to appropriate asset price modelling.

5.2. Data

Our analyses are based on S & P CNX NIFTY (NIFTY) and Hong Kong AOI (AOI) from the NSE (National Stock Exchange, India) website database and Hong Kong Exchange website database respectively. S & P CNX Nifty is a well diversified 50 stock index accounting for 25 sectors of the Indian economy. AOI is ALL Ordinary Index, a value weighted index compiled by the Stock Exchange of Hong Kong based on all common stocks listed there. Thus the two indices represent diverse portfolios with diverse trading activities in two different countries. The fact that the indices are representative of two different countries makes the analysis all the more significant as we are able to show certain empirical universality on two different portfolios with distinguished characteristics.

Intra-day trading data, if analysed, would have certainly helped us in doing a scaling analysis which would have refined our study. However, due to non-availability of such data for free, we have restricted our study to inter-day or daily trading data. This constraint, however, does not pose any significant impediment to our endeavour to seek answer to the main question on underlying price behaviour similarity between two markets situated in two different countries. To the best of our knowledge such a study, comparing stock market return time
series of two different countries (one emerging market, the other a relatively developed market) employing the Hilbert technique, has not been done before. This makes this data set a unique combination under study.

Fig.5. 1a. and b. depict the time series paths of S & P CNX NIFTY and Hong Kong AOI, respectively.

Fig 5.1a. Daily close of NIFTY for the period July 1990 to January 2006
5.3. Analysing the basic return distribution

To begin with, let us define the fundamental variables governing the study and discuss the basic characteristics emerging from the data. We denote the Index values by \( I(t) \). The time series of logarithmic returns of the portfolio priced at \( I(t) \) over a time scale of \( s \) is defined as

\[
R_{\tau}(t) = \ln \left( \frac{I(t)}{I(t-\tau)} \right) \tag{5.3.1}
\]

where \( \tau \) is the primary unit, in this case 1 day. The time scale \( s \) being a parameter used to sample time series of returns, we can take different \( s \) for \( R_s(t) \) to analyse the behaviour of the return series with inter-day frequencies. We take \( s = 1-4 \) days for daily data and the results obtained for NIFTY and AOI are shown in Fig. 5.2a and b, respectively. A brief look at the figures tells us that the
amplitude of the time series is proportional to the sample time scales. The point to note is that for both the series representing two different markets, the proportionality is almost of the same order. Therefore, based on Eqn. (5.3.1) we define the normalised logarithmic returns as

\[
 r_t(t) = \frac{R_t(t) - \bar{R}_t(t)}{\sqrt{\frac{R_t^2(t) - \bar{R}_t^2(t)}}}
\]  

[5.3.2]

where the expectation values denoted by R are taken over the entire time period under consideration.

We further define the probability density function P as the normalised distribution of a measure ρ, which satisfies

\[
 \int_{-\infty}^{\infty} P(\rho) d\rho = 1
\]

[5.3.3]

where the measure ρ can be R, r, phase or amplitude. The probability distributions of the respective return series of NIFTY and AOI are depicted in Fig. 5.3a and b, respectively. The striking similarities in shape and behaviour of the distributions are quite noticeable. While most of the values for normalised returns on NIFTY are ranging between -3.45 and 3.45, those of AOI are ranging between -3.75 and 3.85. It was reported that the probability distributions of the normalised returns can be well described by the so-called double-exponential law also known as the Laplace distribution P(r) \sim \exp(-|r_t/\kappa|), where \kappa is a constant. The double-exponential distribution of return at not-too-long times t is a universal, ubiquitous feature of financial time series and was observed for different countries, stock-market indices, and individual stocks. Fig.5. 3a and b show the normalised returns of daily NIFTY and AOI, respectively. According to Silva
the central part of the curves shown in Fig. 5.3 can be fitted by the scaling form using a Bessel function, where 99% of probability resides and statistics is good, followed by power laws in the far tails, where data statistics is often poor. These features are well captured by the Heston stochastic process. For detailed discussions of the exponential-to-Gaussian crossover, see Silva (2004).  

Fig 5. 2. a. and b. Time series data of log returns $R(t)$ of inter-day NIFTY and AOI index, respectively, sampled by 1, 2, 3 and 4 days

---


*Ibid*
Fig. 5. 3. a. and b. Probability distribution of normalised returns of NIFTY and AOI index, respectively.
5.4. Empirical Mode Decomposition and the Intrinsic Mode Functions

5.4.1. Background

For a given real signal $s(t)$ we look for decomposition into simpler signals (modes)

$$s(t) = \sum_{j=1}^{m} a_j(t) \cos \phi_j(t)$$

where $a_j(t)$ is the amplitude and $\phi_j(t)$ is the phase of the $j$-th component. Each of the components needs to have a physical and mathematical meaning. Let $s(t)$ is mono component signal, i.e. we can find representation of the form

$$s(t) = a(t) \cos \theta(t)$$

that is both physically and mathematically meaningful. There are infinitely many ways to construct such representations but it is often advantageous to write the signal in complex form

$$S(t) = s(t) + is_i(t) = A(t) \exp i\phi(t)$$

and to take the actual signal to be the real part of the complex signal. The imaginary part $s_i(t)$ of $S(t)$ has to be chosen to achieve a sensible physical and mathematical description. If we can fix the imaginary parts we can then unambiguously define the amplitude and the phase by equations 5.4.1.4 and 5.4.1.5 respectively.
\[ A(t) = 2 \sqrt{s^2 + s_i^2} \]  

[5.4.1.4]

\[ \phi(t) = \arctan \left( \frac{s_i}{s} \right) \]  

[5.4.1.5]

The next important step is to define. Some of the methods are:

quadrature method

analytic signal method - \( s_i(t) \) is the Hilbert transform of \( s(t) \).

instantaneous frequency of \( s(t) \) is \( \theta(t) \)

\[ \left| \frac{a'(t)}{a(t)} \right| \]

instantaneous bandwidth is

narrow band instead of mono-component. The most popular example is

\[ s(t) = ACos \theta(t) \]

Some paradoxes exist. To avoid them, Huang, et al (1998)\(^{361}\) have developed a method, termed the Hilbert view, for studying nonstationary and nonlinear data in nonlinear mechanics.

The main tools of this methodology are:

Empirical Mode Decomposition Method (EMD)

Intrinsic Mode Functions (IMFs)

5.4.2 Main idea

The starting point of the Empirical Mode Decomposition (EMD) is to consider oscillations in signals at a very local level. In fact, if we look at the evolution of a signal $x(t)$ between two consecutive extrema (say, two minima occurring at times $t-$ and $t+$), we can heuristically define a (local) high-frequency part $d(t), t- \leq t \leq t+$ or local detail, which corresponds to the oscillation terminating at the two minima and passing through the maximum which necessarily exists in between them. For the picture to be complete, one still has to identify the corresponding (local) low-frequency part $m(t)$, or local trend, so that we have $x(t) = m(t) + d(t)$ for $t- \leq t \leq t+$. Assuming that this is done in some proper way for all the oscillations composing the entire signal, the procedure can then be applied on the residual consisting of all local trends, and constitutive components of a signal can therefore be iteratively extracted.

Given a signal $x(t)$, the effective algorithm of EMD can be summarized as follows:

- identify all extrema of $x(t)$
- interpolate between minima (resp. maxima), ending up with some envelope $e_{\text{min}}(t)$ (resp. $e_{\text{max}}(t)$)
- compute the mean $m(t) = (e_{\text{min}}(t) + e_{\text{max}}(t))/2$
- extract the detail $d(t) = x(t) - m(t)$
- iterate on the residual $m(t)$
5.4.3 Implementation

In practice, the above procedure has to be refined by a sifting process which amounts to first iterating steps 1 to 4 upon the detail signal $d(t)$, until this latter can be considered as zero-mean according to some stopping criterion. Once this is achieved, the detail is referred to as an Intrinsic Mode Function (IMF), the corresponding residual is computed and step 5 applies. By construction, the number of extrema is decreased when going from one residual to the next, and the whole decomposition is guaranteed to be completed with a finite number of modes.

Modes and residuals have been heuristically introduced on spectral arguments, but this must not be considered from a too narrow perspective. First, it is worth stressing the fact that, even in the case of harmonic oscillations, the high vs. low frequency discrimination mentioned above applies only locally and corresponds by no way to a pre-determined sub-band filtering (as, e.g., in a wavelet transform). Selection of modes rather corresponds to an automatic and adaptive (signal-dependent) time variant filtering.

As we have seen, the algorithm to create IMFs in EMD is rather elegant, and it mainly consists of two steps. First, the local extrema in the return time series data, $R_t(t)$ identified. Then, all the local maxima are connected by a cubic spline line $U(t)$, which forms the upper envelope of the time series. At the same time,
the same procedure is applied for the local minima to produce the lower envelope, L(t). Both envelopes will cover all the original time series. The mean of upper envelope and lower envelope, $m_1(t)$, given by

$$m_1(t) = \frac{U(t) + L(t)}{2} \quad [5.4.3.1]$$

is a running mean. We then subtract the running mean $m_1(t)$ from the original time series $R_\tau(t)$ and get the first component $h_1(t)$

$$R_\tau(t) - m_1(t) = h_1(t) \quad [5.4.3.2]$$

The resulting component $h_1(t)$ is an IMF if it satisfies the following conditions:

i. $h_1(t)$ is free of riding waves.

ii. It displays symmetry of the upper and lower envelopes with respect to zero.

iii. The numbers of zero crossing and extremes are the same or only differ by 1.

If $h_1(t)$ is not an IMF, the sifting process has to be repeated as many times as is required to reduce the extracted signal to an IMF. In the subsequent steps of sifting process, $h_1(t)$ is treated as the data

$$h_1(t) - m_1(t) = h_{1,1}(t) \quad [5.4.3.3]$$

If the resulting time series is the first IMF, then it is designated as

$$c_1(t) = h_{1,1}(t) \quad [5.4.3.4]$$
The first IMF component from the data contains the highest oscillatory frequency found in the original data $R_t(t)$.

Fig. 5.4. a. and b. The 1st 4 IMFs NIFTY and AOI index, respectively.
Subsequently, the first IMF is subtracted from the original data and the difference \( r_1(t) \), given by

\[
R_\tau(t) - c_1(t) = r_1(t)
\]  

[5.4.3.5]

is a residue. The residue \( r_1(t) \) is taken as if it were the original data, and we apply to it again the sifting process. Following the above procedures, the process of finding more intrinsic modes \( c_i \) continues until the last mode is found. The final residue will be a constant or a monotonic function, which represents the general trend of the time series data. Finally, we get

\[
R_\tau(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)
\]  

[5.4.3.6]

where \( r_n \) is a residue. To perform the EMD method on a financial time series, one may or may not impose intermittency as an additional condition in the sifting process, depending on the nature of the financial time series under consideration. The intermittency can be considered as a window used to eliminate the end effects and to facilitate computation. However, a characteristic intermittency in trading time of a stock market may be indefinite. In particular, for index return time series, there is no time scale to guide the choice of window size. Therefore, here we do not impose definite intermittencies in the sifting process. Hence, in the sifting process, the structures of the time series with primary time sampling intervals are closely preserved in the first mode.
We take returns $R(t)$ with time sampling interval of 1 day as the primary time series and then perform EMD to decompose $R(t)$ into 11 IMFs. The results are shown in Fig. 5.4a and b in which only the first four IMFs are shown, for NIFTY and AOI, respectively. The physical meanings of the decomposition are clear from the features of IMFs. Let us first compare time series $R(t)$ and IMFs $c_1$ and $c_2$ in Fig. 5.4. According to Eq. s (5.4.35) and (5.4.3.6), $R(t)$ consists of 11 IMFs and each IMF is independent from the others. The term 'independent' here is in some sense equivalent to the term 'orthogonal' in the theory of finite dimensional vector space. In other words, each IMF cannot be represented by other IMFs decomposed from the same primary time series. The main difference between IMFs $c_1$ and $c_2$ is the intermittencies they own. IMF $c_1$ is the first mode separated from $R(t)$ after the sifting process, and it has the highest frequency among 11 IMFs. Since no criterion is imposed on the intermittency, there is no specified relation between the intermittencies of $c_1$ and $c_2$. Furthermore, if one IMF dominantly catches characteristic features of $R(t)$, then its contribution is distinguishable in an observation like Fig.5. 4. It is obvious that $c_1$ catches the main structures of $R(t)$, since the time series of $R(t)$ is mainly characterized by its highest-frequency component. However, we should note that this is case by case and the conclusion may not be applicable to other time series. In our analysis, it is very important to note that IMF $c_1$ is not equal to time series $R(t)$. If we evaluate some quantities specifically defined for $R(t)$ from $c_1$s, the results may be quite different. Actually, it is not reasonable to impose all the fundamental statistics primarily performed on return $R(t)$ to IMFs. After IMFs have been
obtained from the EMD method, one can further calculate instantaneous phases of IMFs by applying the Hilbert transform to each IMF component, say the $r^{th}$ component. The procedures of the Hilbert transform consist of calculation of the conjugate pair of $c_r(t)$, i.e.

$$y_r(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_r(t')}{t-t'} dt'$$  \[5.4.3.7\]

where $P$ indicates the Cauchy principal value. With this definition, the two functions $c_r(t)$ and $y_r(t)$ forming a complex conjugate pair define an analytic signal $z_r(t)$:

$$z_r(t) = c_r(t) + i y_r(t)$$  \[5.4.3.8\]

which can also be expressed as

$$z_r(t) = A_r(t)e^{i \phi_r(t)}$$  \[5.4.3.9\]

with amplitude $A_r(t)$ and the phase $\phi_r(t)$ defined by

$$A_r(t) = [c_r^2(t) + y_r^2(t)]^{\frac{1}{2}}$$  \[5.4.3.10\]

$$\phi_r(t) = \arctan \left( \frac{y_r(t)}{c_r(t)} \right)$$  \[5.4.3.11\]

Then, we can calculate the instantaneous phase according to the equations. The amplitudes and phases of the IMFs calculated by the Hilbert transform are shown in Fig. 5.5a and b for NIFTY and AOI, respectively.
Fig.5. a. and b. The amplitude and phases of IMFs calculated by Hilbert transforms of NIFTY and AOI index, respectively.
5.4.4 The probability distribution of the 1st 4 IMF phases

First we a take a look at the phase distribution of the 1st 4 IMFs of the two time series as depicted in Fig. 5.6a and b.

We find that values of the phase distribution lie between -1.575 and 1.575 for both NIFTY and AOI. This indicates a striking similarity in the underlying periodicity of the two time series. Considering the separation of origin of the two series, this clearly indicates a mathematical universality of the behaviour of return time series in financial markets. A closer look at the amplitude distribution, as depicted in Fig. 5.7a and b, however, suggests that the range of values is much higher in case of AOI than NIFTY. This suggests a higher energy level for AOI signal, but the qualitative nature of distributions for both the cases is same.

Next we do a comparative analysis of the respective phase distributions of NIFTY and AOI for the first 2 IMFs. As one can see in Fig. 5.8, not only does the range of values match but also the polynomial regression lines are almost identical. This suggests a high degree of non-linear correlation. Again, this finding corroborates our initial findings.

As the amplitude value range is different for NIFTY and AOI, we use normalised values of the same and compare their probability distributions. Here, we also find a striking similarity between the two different time series as evident in Fig. 5.9.
Fig. 5. 6. a. and b. depict probability distributions of phases of 1st 4 IMFs of NIFTY and AOI index, respectively.
Fig. 5. a. and b. Probability distributions of amplitudes of 1st 4 IMFs of NIFTY and AOI index, respectively.
Fig. 5.8. Shows correlation graph of probability distribution of phases of 1st 2 IMFs of NIFTY and AOI, depicting polynomial regression curves as well.
Fig. 5.9. Shows correlation graph of probability distribution of amplitudes of 1st 4 IMFs of NIFTY and AOI, depicting polynomial regression curves as well.

5.5 Frequency domain characteristics of financial time series

The uniqueness of this project lay in the selection of the two financial time series under consideration. They represent the return on stock indices of two different countries. The trading volume, efficiency and structure of market are all different
in these two markets. However, our study showed that both time series exhibited strikingly similar patterns.

We further employed the Hilbert–Huang method of time signal analysis to define the instantaneous phase to catch characteristic features of index and return time series. The EMD method was used to decompose return time series into several IMFs, and the Hilbert transform was used to calculate the instantaneous phase of the first three IMFs accordingly.

We find that the phases of all IMFs are randomly distributed and have equal probabilities for all possible phases ranging between -1.575 and 1.575. This behaviour exists in both the time series. We expect that the same behaviour also exists in different time scale. The phase distributions corresponding to abruptly changing behaviours indicate unpredictable and stochastic features of the indexes. For a more detailed investigation, we needed access to intra-day data, which were not freely available. However, our initial findings strongly suggest the fact that the underlying return distributions of the two different stock markets essentially follow the same patterns. This establishes a common trading behavioural pattern in both Hong Kong Stock exchange and NSE, India.

While this is an important finding, it may be noted that we cannot comment on the reasons of this similarity based on this study alone. The econometrics of
financial markets depends on multiple variables and a separate causality study may be conducted to establish multivariate relationship. However, our endeavour was to use the EMD methodology to reveal characteristics of the time series which help us in comparing two or more time series effectively. In that direction, this work established the ability of EMD method to bring out characteristics of the time series graphically in the frequency domain that enables us to compare and comment on similarity/dissimilarity between two time series. This builds opportunities to revisit the theoretical implications of finance in terms of portfolio theory, asset pricing as well as international financial linkage. This justifies our choice of this method to test and compare the two models under test. Having established the utility of EMD as a test method we now present another work of ours in the next chapter where we use Recurrence Analysis to decipher patterns in the stock price movement.