Computational Finance

Irreversibility of Financial Time Series: A Graph Theoretical Approach

Diptarka Saha(MB1516)

Anirban De(MB1428)

Introduction

A time series is said to be directional (or irreversible) if it possesses probabilistic properties which depend on the direction of time. A time series is said to be reversible if it has no such property. haviour. Appropriate plotting of time series can often reveal directionality; there is a lack of directional symmetry in certain characteristic behaviour sequences, such as rises and Falls. There has been relatively little written about directionality in time series. One reason for this may be that, until recently, time series modelling has been coterminus with standard Gaussian autoregressive moving average models (ARMA).However, Non-linear and non-Gaussian linear models typically have directionality as a property of their higher order depen

The basic idea is that directionality is an aspect of time series analysis which deserves wider recognition; for instance, it does not make sense to forecast with a time series model which is reversible, when past data are definitely irreversible. An example would help

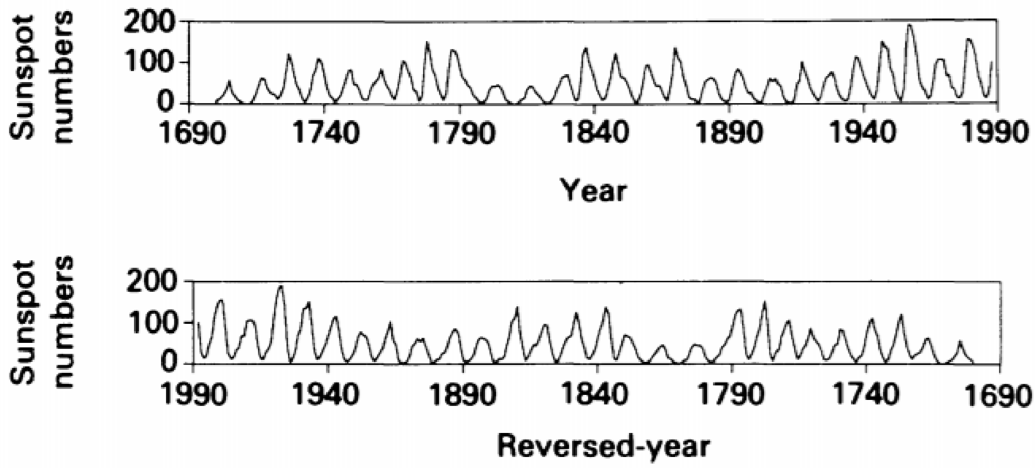


Figure 1. *The sunspot numbers* (1700–1988) *in forward and reverse time*; *scaling similar to Cleveland & McGill* (1987) *who note that 'the sunspot numbers rise more rapidly than they fall'*.

This data which is definitely irreversible should be modelled using irreversible models.

This concept of reversibility is also very interesting in the realm of financial time series. Basically a reversible financial time series would follow the efficient market hypothesis and hence attempts of predicting its behaviour would be futile. However, the more the irreversibility the more inefficient (and predictable in some sense) is the market.

It is important to stress at this point that financial time series are usually non-stationary. This is in principle a fundamental drawback, as to the best of our knowledge no rigorous theory has been advanced so far linking time series irreversibility and entropy production in the non-stationary case. As a matter of fact, according to the original definition, non-stationary series are infinitely irreversible, so the quantification of how irreversible a non-stationarytime series seems to be an ill-defined problem to begin with. Again, here we circumvent this problem by using the so-called visibility algorithms, a family of methods to make time series analysis in graph space that have been shown recently to be able to quantify different degrees of irreversibility in both stationary and non-stationary processes.

Accordingly, here we propose to apply the concept of graph-theoretical time series irreversibility in the context of financial time series. We first make use of the visibility algorithms to construct graph-theoretical representations of the stock prices of 15 companies from the NSE in the period 1997–2017. We then estimate time irreversibility in these representations through the Kullback–Leibler divergence of the *in* and *out* degree distributions. After checking that this measure is indeed genuine and not correlated to volatility, we show that all the companies under study are irreversible, and their degree of irreversibility varies across companies and fluctuates over time. The variance across companies allows us to rank companies, and the collective time fluctuations are finally used to provide a classification of financial periods.

Methodical Notes

Measure of Irreversibility

A dynamical process is said to be time reversible if any two time series $S = \{x_1, x_2, ..., x_n\}$ and $S^* = \{x_{-1}, x_{-2}, ..., x_{-n}\}$(where n denotes time) generated by this process asymptotically have the same joint distribution . In the concrete casewhere the process is stationary, the definition of time reversibility reduces to the equivalence of statistics between the forward and backward process: a stationary time series is thus time reversible if a series $\{x1, x2, ..., xn\}$ and its reverse $\{xn, ..., x2, x1\}$ are equally likely to occur. That is to say, the joint distributions $p(x1 , x2, ..., xn)$ and $q(xn, ..., x2, x1)$ coincide for reversible processes. If p ⍰= q we say that the process is (statistically) time irreversible.

First, we need a measure of irreversibility, Amongst other descriptors, we advocate that the Kullback–Leibler divergence (KLD) between the statistics (distributions) of the (appropriately symbolized) time series of the forward and backward process is indeed an interesting choice. We recall that if p and q are discrete distributions with domain X, then the Kullback–Leibler divergence $D\text{kld}(p|q)$ is defined as

$$D_{\text{kld}}(p|q) = \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})}$$

This is a semi-distance (i.e. non-symmetric) which is null if and only if $p = q$ and positive otherwise

Visibility Algorithms

Visibility algorithms are a family of methods to map time series into graphs, in order to explore the structure of the time series (and the dynamics underneath) using graph theory. Let $S = \{x(t)\}$ be a real-valued time series data.

We define and use the so called horizontal visibility graph (HVG) is defined by

(i) every datum $x(i)$ in the series is mapped to a node i in the graph (hence the graph nodes inherit a natural ordering),
(ii) two nodes i and j are connected by an edge if the associated data show mutual horizontal visibility if any other datum $x(k)$, where $i < k < j$, fulfil the following *ordering* criterion:
$x_k < \inf(x_i, x_j), \forall k: i < k < j$

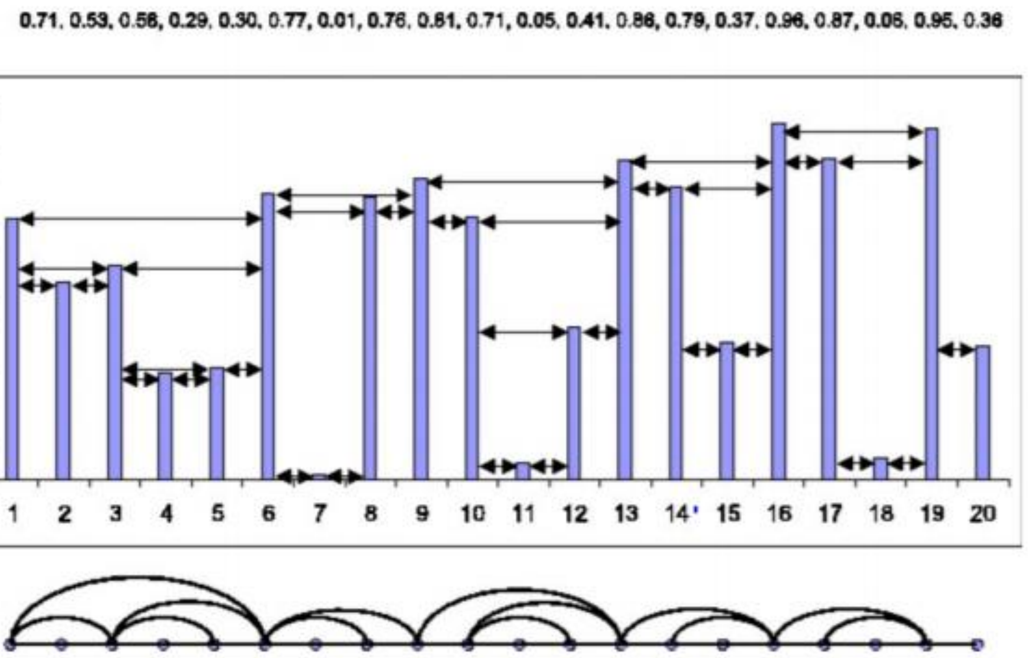


FIG. 2. (Color online) Illustrative example of the horizontal visibility algorithm. In the upper part we plot a time series and in the bottom part we represent the graph generated through the horizontal visibility algorithm. Each datum in the series corresponds to a node in the graph, such that two nodes are connected if their corresponding data heights are larger than all the data heights between them. The data values (heights) are made explicit in the top.

Note that previous definitions generate undirected graphs.
However, these can be made directed by assigning to the links the time arrow naturally induced by the node ordering. Accordingly, a link between *i* and *j* (where time ordering yields $i < j$), generates an *outgoing* link for *i* and an *ingoing* link for *j* in a directed version of a VG/HVG. The degree sequence of the VG/HVG
(which assigns to each node its degree or number of edges) thus splits into an ingoing degree sequence $\{kin(t)\}_{t=1}^{T}$, where $kin(t)$ is the ingoing degree of node $i = t$, and an outgoing degree sequence. An important property at this point is that the ingoing
and outgoing degree sequences are interchangeable under time series reversal. That is to say, if we define the time reversed series $S* = \{x_{T+1-t}\}_{t=1}^{T}$, then we have the following identities

$$\{k_{in}(t)\}[\mathcal{S}] = \{k_{out}(t)\}[\mathcal{S}^*]; \ \{k_{out}(t)\}[\mathcal{S}] = \{k_{in}(t)\}[\mathcal{S}^*],$$

Now, one can define, from the ingoing and outgoing degree sequences, an ingoing degree distribution $P(kin) \equiv Pin(k)$ and an outgoing degree distribution $P(k_{out}) \equiv P_{out}(k)$

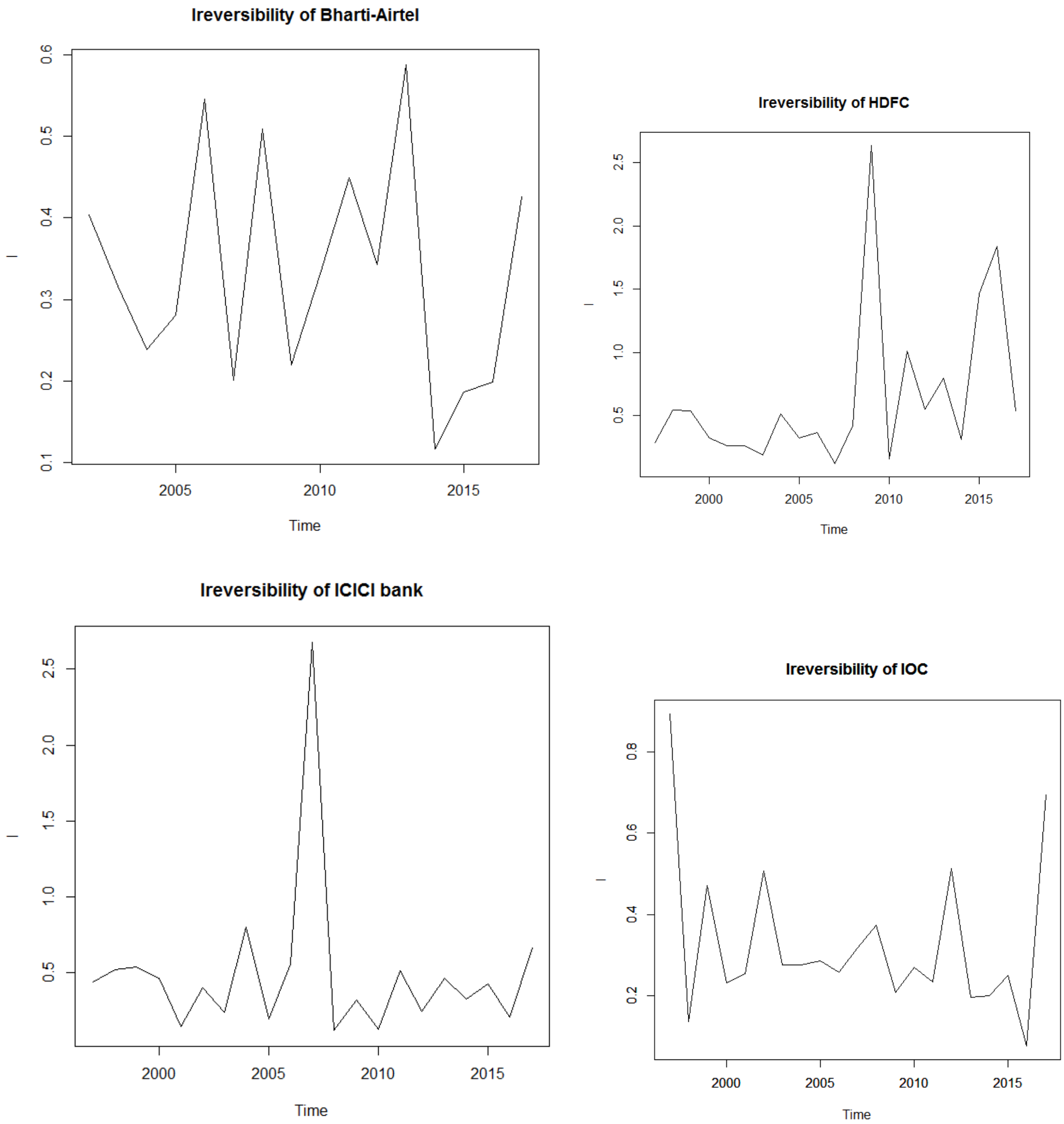
In other words, the statistics of the forward and backward process are encoded, in graph-space, in the *in* and *out* degree sequences. Time series irreversibility can then be estimated via the Kullback– Leibler divergence between the *in* and *out* degree distributions
This graph-theoretical measure reads

$$I_{HVG/VG}(\mathbf{x}) := D_{kld}(P_{in}||P_{out}).$$

Data and Results

We could only find data of resolution up to a day. So we gathered data for 15 companies from 1997-2017. For each of them we calculated the I_{HVG} for each year and then tried to find patterns.

Irreversibility plots for some of the companies follow



A clear pattern emerges, for banking companies such as HDFC and ICICI there is a huge peak at around 2008, which was presumably caused by 2008 financial meltdown. However, for non-banking companies such as Airtel and IOC this peak does not emerge. This pattern actually follows for other companies too. Mostly, there are few peaks of irreversibility followed by quasi-reversible areas. These irreversibility may have been caused due to different circumstances.

Ranking Companies:

In order to quantify the net amount of irreversibility of a certain company, we introduce Score[c]defined as the mean of the yearly I_{HVG}s, the score of a company c as
the average of the annualized irreversibility value. This quantity averages the degree of

irreversibility of a given company over large periods of time. According to the analogy between reversibility and entropy production, the larger the Score is, the more 'away from equilibrium' the signal generated by *c* is, thus producing larger amounts of entropy. This might be relevant from a financial perspective, as the larger the Score of a company, the less efficient it is and thus more interesting from an investment viewpoint. Here are all the scores

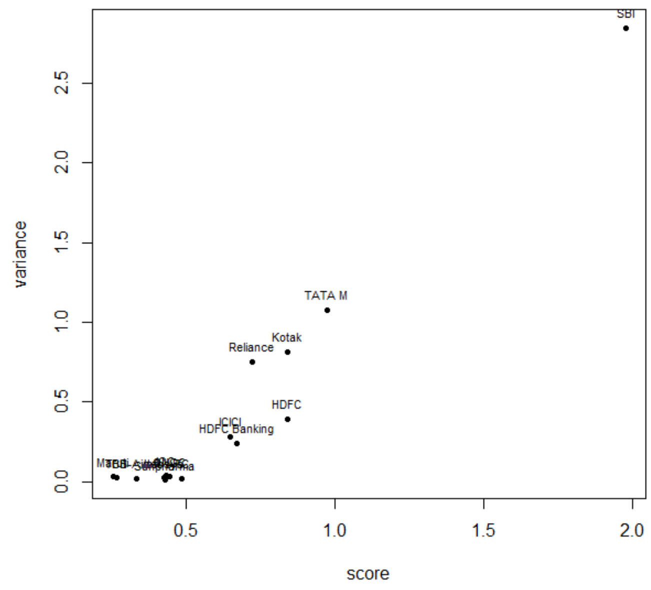
Company.1	score
BH-Airtel	0.3346248
HDFC Banking	0.6708538
ICICI	0.6500226
HDFC	0.8396314
Infosys	0.4273373
IOC	0.4333338
Reliance	0.7215806
TATA M	0.9759707
ONGC	0.4440302
Maruti	0.2559713
SBI	1.9748813
TCS	0.2674060
Sunpharma	0.4308516
Kotak	0.8423355
ITC	0.4867503

To further asses the the possibility that some companies may have suffered from large irreversibility only at sporadic occasions (which would yield a high irreversibility score even if the company were following a quasi-reversible evolution in most of the period), we also compute the irreversibility variance of a given company, defined as the variance of the I_{HVG} s

Here is the full table

Company.1	score	var_sc
BH-Airtel	0.3346248	0.01976251
HDFC Banking	0.6708538	0.24380498
ICICI	0.6500226	0.28400191
HDFC	0.8396314	0.39528441
Infosys	0.4273373	0.02385412
IOC	0.4333338	0.03674060
Reliance	0.7215806	0.75313449
TATA M	0.9759707	1.07879727
ONGC	0.4440302	0.03028263
Maruti	0.2559713	0.03181563
SBI	1.9748813	2.84708542
TCS	0.2674060	0.02330474
Sunpharma	0.4308516	0.01131465
Kotak	0.8423355	0.81352815
ITC	0.4867503	0.02058960

In general, the irreversibility Score will be a faithful static measure of a company's irreversibility as long as we have relatively small variance.



We see that in most of the cases the scores are pretty close and the variances are also low, however in some cases such as SBI both of them are high, TATA MOTORS shows a high variance in spite of low score. Interestingly, we find that the top six multinationals in the Score ranking, also have large variance. These are companies which have been dramatically affected by major external perturbations at certain specific times, perhaps acting as global sensors of the financial system's stability state.

Assessing periods of financial reversibility

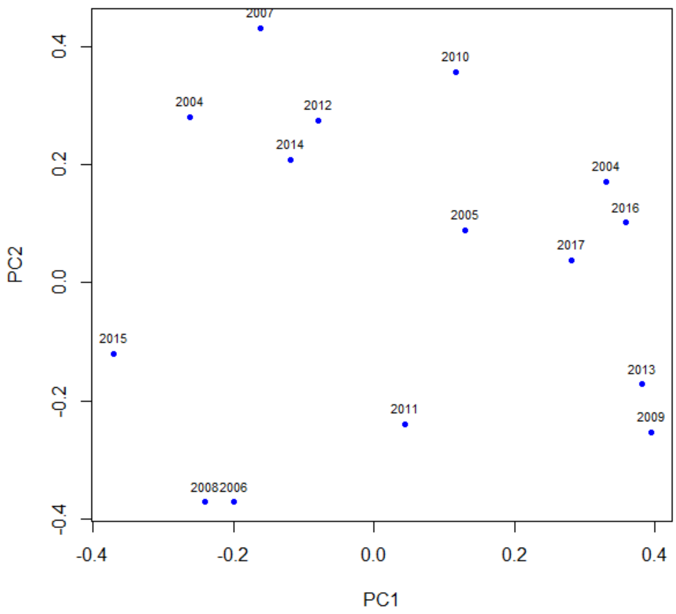
In the next section, we further explore this possibility, and investigate, in an unsupervised way, if the evolution of irreversibility features across companies over time reflects the stability of the whole financial system. This allows us to classify and cluster periods of time according to their level of systemic reversibility.

We have 14 data points, each belonging in a $\mathbf{R}^{15}$ space, that we would like to classify and find pattern amongst.

We try two methods

(i) Principal Component Analysis

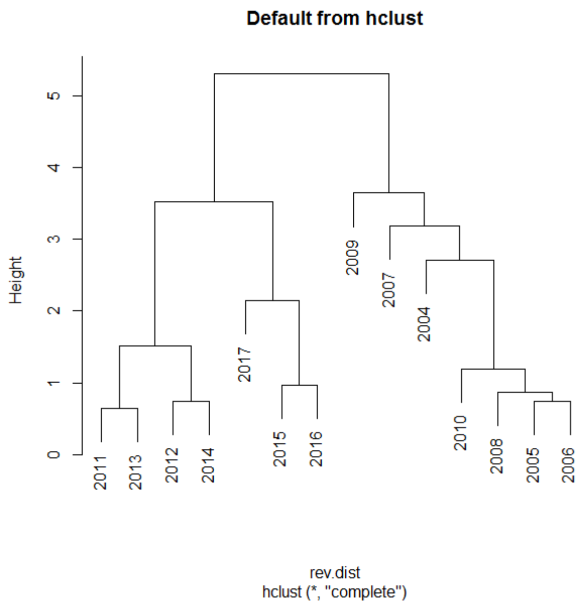
Principal Component Analysis (PCA) is a common statistical procedure to perform dimensionality reduction on data. We perform this on our dataset and find the following diagram



The PCA does not provide us with a clear classification at this stage, however since only 38% of the variability is expressed in the first two PC s we would not expect much, to have a finer perspective we use our next method

(ii) Hierarchical Clustering

We build a hierarchical cluster on the data using the hclust() function in R with complete linkage criterion to produce a hierarchical cluster tree. This can be viewed in the dendogram below



We can see, that the data is mainly divided into two parts, before 2011 and after 2011. It also seems the years closer to 2008 are also closer. Current years (2015,2016,2017) are also close which is possibly a sign of instability in the current market. However data with more resolution should be used to concur something definite .

Conclusion

From our computations with comparison to the original paper - one thing was obvious the 2008 meltdown had some say in the dynamics of the market, however it was not as strong a factor as it is in US market. Possibly pointing that the effect of the catastrophe was not as severe here.

Most important thing from here would be to find data with more resolution (minute wise, possibly seconds) to have some definite conclusions. But it is evident that each company has times ot irreversibility and times of quasi reversibility, however there is very much similarity in the behaviour of similar companies such as banking ones. Some further research with more companies would also be ideal

APPENDIX:

The codes

```
kld=function(p,q){

d=sum(p*log(p/q))

return(d)

}


data=read.csv("ge1.csv")


price=data[,2]


kld.ts=function(price){

adj.graph=matrix(0,ncol=length(price),nrow=length(price))


for (i in 1:length(price)){

for(j in i:length(price)){

if(j==(i+1))adj.graph[i,j]=1

if((j-i) >1 ){

ll=max(price[i+1]:price[j-1])

if(ll< min(price[i],price[j]))adj.graph[i,j]=1

}

}

}
```

```
graph=graph_from_adjacency_matrix(adj.graph,mode="directed")

d_in=degree(graph,1:length(price),mode="in")
d_out=degree(graph,1:length(price),mode="out")

h_in=as.numeric(names(table(d_in)))
h_out=as.numeric(names(table(d_out)))

k_in =numeric(max(h_in,h_out))
k_out =numeric(max(h_in,h_out))

for (i in 1:length(k_in)){

if(i %in% h_in )k_in[i]=table(d_in)[names(table(d_in))==i]

else k_in[i]=1/((length(price))^2)

}

k_in=k_in/sum(k_in)

for (i in 1:length(k_out)){

if(i %in% h_out )k_out[i]=table(d_out)[names(table(d_out))==i]

else k_out[i]=1/((length(price))^2)

}
```

```
k_out=k_out/sum(k_out)


return(kld(k_in, k_out))

}

#####
```