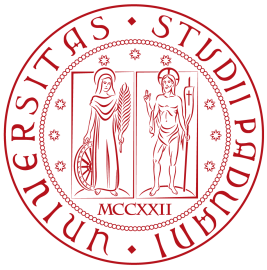

Università Degli Studi Di Padova

Dipartimento di Fisica e Astronomia "Galileo Galieli"

Master Degree in Physics of Data

Final Dissertation

MEASURING THE INFORMATION FLOW BETWEEN THE WEB AND STOCK MARKET VOLUMES: A MULTIVARIATE TRANSFER ENTROPY ANALYSIS

Supervisor
Prof. Elisa Omodei

Co-supervisors
Prof. Guido Caldarelli
Prof. Samir Suweis

Candidate
Giulio Vicentini

Academic Year

2021-2022

Ai miei genitori
e ai miei compagni di laboratorio

Abstract

The recent discovery of strong analogies between speculative markets and some well known physical phenomena and concepts, such as spin systems, universality, criticality and complexity has led to a growing interest of physicists in the dynamics of financial markets. Moreover the immense amount of data that nowadays is provided by the internet represents both a challenge and an opportunity to find appropriate models that are able to describe emergent statistical laws.
Following the procedure of recent studies, in this thesis we investigate the interplay between finance-related news and tweets and financial markets. In particular, we consider, in a period of 9 years, the Twitter-and-news volume of the 30 stock companies that form the Dow Jones Industrial Average (DJIA) index and, as a first attempt, we explore results of Granger-causality test. However, the non-stationary and non-gaussian nature of financial data requires a different tool that can overcome the limits of linear statistics. We found this tool in information theory; allowing us to propose a novel approach based on a multivariate transfer entropy analysis.

Contents

Listing of figures

Listing of tables

1 Introduction

In recent years there has been a fast growth of data-production and data-accessibility. This availability has led to the emergence of a new field of study, as well introduced by Easley and Kleinberg [4]: ”Over the past decades there has been a growing public fascination with the complex “connectedness” of modern society. This connectedness is found in many incarnations: in the rapid growth of the Internet and the Web, in the ease with which global communication now takes place, and in the ability of news and information as well as epidemics and financial crises to spread around the world with surprising speed and intensity. These are phenomena that involve networks, incentives, and the aggregate behavior of groups of people; they are based on the links that connect us and the ways in which each of our decisions can have subtle consequences for the outcomes of everyone else.

Motivated by these developments in the world, there has been a coming-together of multiple scientific disciplines in an effort to understand how highly connected systems operate. Each discipline has contributed techniques and perspectives that are characteristically its own, and the resulting research effort exhibits an intriguing blend of these different flavors. From computer science and applied mathematics has come a framework for reasoning about how complexity arises, often unexpectedly, in systems that we design; from economics has come a perspective on how people’s behavior is affected by incentives and by their expectations about the behavior of others; and from sociology and the social sciences have come insights into the characteristic structures and interactions that arise within groups and populations. The resulting synthesis of ideas suggests the beginnings of a new area of study, focusing on the phenomena that take

place within complex social, economic, and technological systems".

1.1 Predicting the behaviour of techno-socio systems

The recent technological revolution has created an unprecedented situation of data availability, changing the way in which we look at social and economic sciences. The constantly increasing use of the Internet as a source of information, such online news and social media, started an analogous increasing online activity. The interaction with technological systems is generating large data-sets that illustrate collective behavior in a previously unimaginable way [5, 6]. In this vast repository of Internet activity we can find the interests, concerns, and intentions of the global population with respect to various economic, political, and cultural phenomena.
Modern techno-social systems are made of large scale physical infrastructures embedded in a variety of communications and computing networks that evolves and develops mirroring human behaviors. To predict how these systems work, we must start formally describing patterns found in the real world. The models that we can use to anticipate future events, risks and trends are based on these formalizations. Computational models, when provided with appropriate data, can return high levels of anticipation power in very complex framework, such as weather forecasting. As a matter of fact, in this context, thanks to recent development of computational systems and a wide availability of historical meteorological data, we managed to reach a great level of accuracy in predicting daily weather.
Despite these promising results in weather forecasting, we can not reach the same accuracy in the quantitative prediction of phenomena in techno-social systems. Indeed our little knowledge of social behavior limits the possibility to predict emergent human actions. This represents the main difference in prediction power between physical systems (for which we have a vast knowledge of underlining laws) and social systems.
The level of information flow regarding social systems is not just due to the development of modern super-computers. Understanding the links between people and technology and the attenuation of borders between the real world and the online one are changing our accessibility to data. A great instance of the people/technology inter-linkage can be found in the analysis of human mobility. In the past, this field was based on often limited and incomplete data; such as census and survey, which were often incomplete and/or limited to a specific context. Despite advances in the study of human transport, this lack of data has impeded the construction of a general framework of human mobility. However, in pioneering work, Brockmann et al. [7] opened the path to the general exploitation of proxy data for human interaction and mobility.

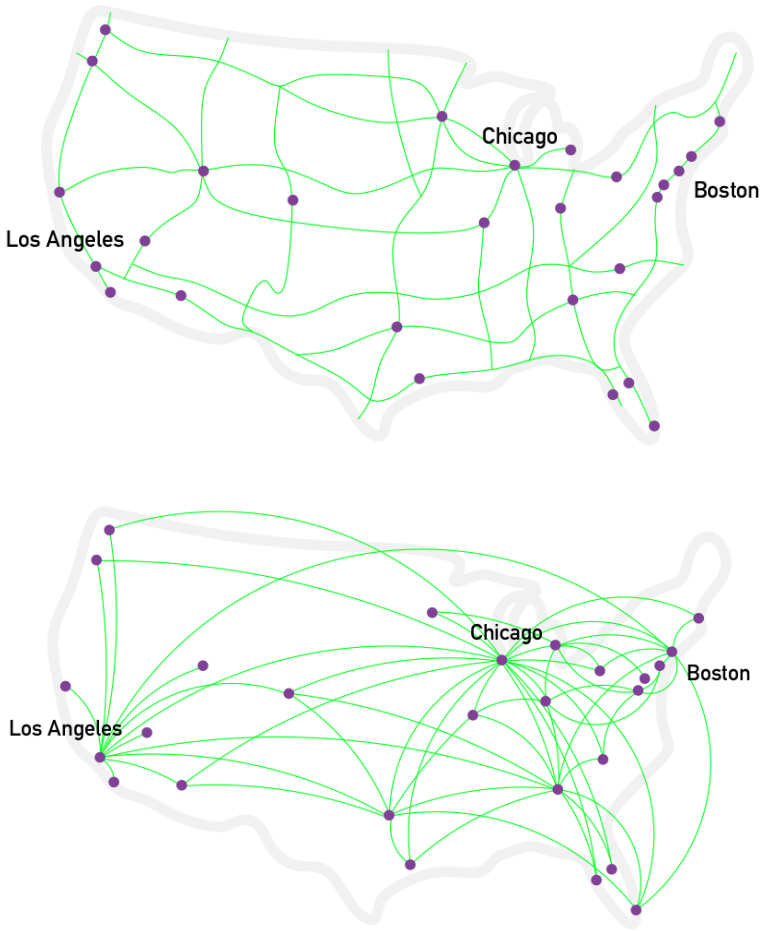

Figure 1.1: U.S. highway network (up) and U.S. flights network (down) [1].

Complex systems, network science, non-equilibrium statistical physics, and computer science all play a key role to face these challenging aspects of predicting and managing events in techno-social networks. Although these approaches are not completely mature yet, it now seems possible to imagine computational predicting systems that will help us design cities, supply-chains, connection infrastructures and resources distributions.

1.1.1 Financial markets

Financial markets such as the New York Stock Exchange (NYSE) or the NASDAQ stock market are a cornerstone of modern financial economics and offer a great example of techno-socio system thanks to the immense amount of electronically recorded financial data available and to

the tight link with human behaviour. Thanks to stock exchanges, companies are able to grow their capitals purchasing shares in change of investors funding. In practice, there are two ways through which an investor can receive a reward:

- if a company perform well, it can earn a profit trough dividends (a percentage of the company profits that are divided among the shareholders);
- otherwise it can sell the original shares at an higher price.

Market indices are one of the most important measure of market performances. For example the Standard and Poor's 500 (S&P500), the NASDAQ Composite Index and the Dow Jones Industrial Average (DJIA) summarise, respectively, the market's performance of the 500 largest publicly traded equities on the New York Stock Exchange (by market capitalisation), of all the stocks listed on the Nasdaq stock exchanges and of the 30 major US manufacturing publicly traded equities. These aggregate measures of the overall market performance of a subset of the equities traded on that market. The importance of these indices it is not related only with the summary of equity values but it extends as a measure of economic performance of a market sector or even of a country as a whole.
We have understood that, through financial markets, companies can raise capital, while an investor can expect a reward in dividends, and high expected dividend can push the price of the stocks. So, good performances of a company can lead to a grow of a financial index.
Nevertheless, there are many factors, not directly related to the performance of a company, that can influence the stocks price. A prime example is he terrorist attacks on the World Trade Center and the US Pentagon on September 11, 2001, in which nothing changed regarding the underlying performance of the companies; while the DJIA lost the 7.14%. In this case it was the behaviour of investors that expected lower returns in the future. From this point of view, markets are a reflection of, not only the underlying performance of companies, but also of overall market's expectations. While markets themselves are exceptionally hard to predict with any level of assurance, there may be very broad drivers in the wider economy that can guide our expectations of what might happen in the financial markets. With this in mind there is considerable interest in finding out what the underlying drivers of our markets and economies really are. So an interesting area to explore is the relationship between equities and indices, indices and indices, and indices and the economy as a whole in order to understand the extent to which changes in one financial or economic measure act as a precursor or driver to changes in the other. The changes in prices of equities show some unusual behaviours that have made the study of their statistics a non-trivial matter. So in practice what is most likely being observed

3M (MMM)	American Express (AXP)	Amgen (AMGN)
Apple (AAPL)	Boeing (BA)	Caterpillar (CAT)
Chevron (CVX)	Cisco (CSCO)	Coca-Cola (KO)
Disney (DIS)	Dow (DOW)	Goldman Sachs (GS)
Home Depot (HD)	Honeywell (HON)	IBM (IBM)
Intel (INTC)	Johnson&Johnson (JNJ)	JPMorgan (JPM)
McDonald's (MCD)	Merck (MRK)	Microsoft (MSFT)
Nike (NKE)	Procter&Gamble (PG)	Salesforce (CRM)
Travelers (TRV)	UnitedHealt (UNH)	Verizon (VZ)
Visa (V)	Walgreens (WBA)	Walmart (WMT)

Table 1.1: List of the Dow Jones components.

in the price variations is the very rapid diffusion of both relevant and irrelevant information through a financial market and its influence on how traders perceive the future value of individual equities.

The goal of this thesis is to measure the information flow from the web (financial news and tweets) to the stock market volumes of the 30 companies that compose the DJIA index (Table 1.1). Indeed, financial turnovers, financial contagion and, ultimately, crises, are often originated by collective phenomena such as herding among investors (or, in extreme cases, panic) which signal the intrinsic complexity of the financial system [8]. Therefore, the possibility to anticipate anomalous collective behavior of investors is of great interest to policy makers [9, 10, 11] because it may allow for a more prompt intervention, when this is appropriate.

1.2 Looking for cause-effect relationships

The advent of the scientific method and its application have led to rapid progress and constant technological development. The mathematical modeling of phenomena allows the formulation of accurate quantitative predictions, and the rigorous verification of the latter with reproducible experiments.

Scientific models still not falsified by experimental measures are characterized by one predictive capacity much higher than that obtained from empirical knowledge, and allow practical applications that cannot otherwise be achieved. Scientific theories often have the property of generalizing and unifying the description of apparently disconnected phenomena, such as the fall of an apple from the tree and the revolution motion of the earth around the sun. In this sense, the resulting knowledge can be said to be of a higher level.

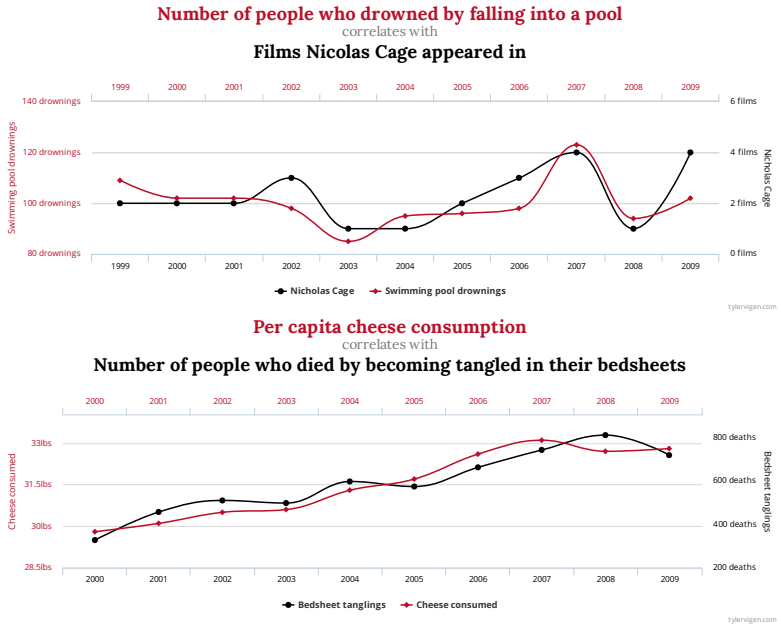

Figure 1.2: Examples of spurious correlations [2].

This approach is characteristic of physics in general and is applied in very different contexts, from the dynamics of the universe to the study of biological systems. This gives excellent results when the systems studied are ”simple”, while it is more difficult to frame ”complex” phenomena, those in which the interacting bodies are very many and the dynamics critically depend from the surrounding conditions: for example disciplines such as Medicine or Biology are very far from developing a set of mathematical laws that explain in detail the functioning of living organisms. However, a quantitative model that makes falsifiable predictions remains the main objective of the scientific approach to problems.

The alternative to the theoretical development of a mathematical model is the search for correlations: to try to understand the mechanisms that regulate a given phenomenon, one relies on events that seem correlated, assuming that this correlation is the result of a cause and effect relationship. If it is true, however, that a cause-effect relationship implies a correlation, the opposite is not always true; that is, correlation does not imply causation. Examples of spurious (i.e. non-causal) correlations are everywhere; Tyler Vigen collects various paradoxical examples of such correlations in a site that has become famous [2], a couple of them are reported in Figure 1.2.

The empirical knowledge that allows us to interact with the environment is based on the observation of correlations, learned over time thanks to our experiences.

To understand the reasons, let's consider the Bayesian approach to statistics. Formally, thinking to all the possible, mutually exclusive, hypotheses H_i (our model) which could condition the event E (the data). What is the probability of H_i under the hypothesis that E has occurred? The answer is given by Bayes Theorem [12]:

$$p(H_i|E) = \frac{p(E|H_i)p(H_i)}{p(E)} \tag{1.1}$$

where

$$p(E) = \sum_j p(E|H_j)p(H_j) \tag{1.2}$$

is the total probability.
From a data-model point of view:

$$p(model|data) \quad \alpha \quad p(data|model)p(model)$$

$$posterior \quad \alpha \quad likelihood \quad prior$$

it is highlighted how our degree of confidence regarding the reliability of a certain model (*posterior probability*) depends on how well the data are in agreement with the model's predictions (*likelihood*), but also on the subjective preconception that we have regarding the model itself (*prior probability*). The examples in Figure 1.2 are paradoxical, because, considering for example the second of the two, the probability that we assign 'a priori' to a model in which US per capita cheese consumption, is causally related to the number of people died tangled in their bedsheets is essentially zero; therefore even after observing the data, this probability, proportional to the product of prior and likelihood, remains very low. The problem arises when with regard to a certain phenomenon our prior is "flat", i.e. we have no ab-origin information or preconceptions: in those cases a likelihood that exhibits a correlation leads us to believe that this correlation is not accidental. To take shelter it is therefore necessary to have correct preconceptions, that is a solid knowledge of the phenomena, which is impossible for all the phenomena we may have to deal with.
What has just been discussed makes it clear how difficult it is to extract information from data without a theoretical model to guide. The approach that is therefore more efficient in the development of knowledge consists in seeking a cause-effect link only where this link is predicted by a well-justified theoretical model, essentially where the Bayesian prior has values not too far from unity. Under these conditions, the theory suggests which measures to take, which data

to analyze and under which hypotheses, triggering the virtuous circle of the scientific method. The central aim of many studies in the physical, behavioral, social, and biological sciences is the clarification of cause–effect relationships among variables or events. However, the appropriate methodology for extracting such relationships from data – or even from theories – has been debated a lot. When talking about causality, the two fundamental questions are [13]:

- 'What empirical evidence is required for legitimate inference of cause–effect relationships?'
- 'given that we are willing to accept causal information about a phenomenon, what inferences can we draw from such information, and how?'

These questions have been without clear answers in part because we have not had an hard formalization for causality and in part because we have not had effective mathematical tools for answering causal questions.
Fortunately, in the last decade, causality has been revised into a mathematical object with a defined meaning. Practical problems based on causal information can now be solved using logic and mathematics.
In this thesis, we consider a statistical form of causality, which can be observed in codependent time series where a response in the dependent series is more likely to follow after some change in the driving series. The direction of information transfer is forced by requiring the cause to precede the effect.
This concept takes shape into the context of Granger causality [14]. In this work we exploit this formalization together with its natural generalization: the transfer entropy, that allows the multivariate analysis needed for financial studies.

Previous Works and Novelty of the Thesis

The idea of this thesis starts from the works of Caldarelli [15, 16] and Novak [17, 18, 19], where different kinds of sources and methods have been proposed to predict the behaviour of financial markets. In particular, in [15] the authors show how 'Web Search Queries can predict Stock Market Volumes' through an analysis based on time-lagged cross-correlation and Granger Causality test; while in [16], using the same methods, 'the effects of Twitter sentiment on Stock Price Returns' is showed. In this framework we aim, first, to reproduce some previous results and, second, to step forward from these researches thanks to the use of contemporary data. More precisely, the novelties that this thesis propose are the followings:

- the analysis and comparing of different web sources updated to contemporary days;
- the use of Transfer Entropy as measure of information flow and as natural generalization of Granger Causality;
- a novel approach based on the Multivariate Transfer Entropy to study the combination and the effect of multiple sources.

In order to do this in chapter 2 a formal definition of Transfer Entropy, together with the comparison with Granger Causality and its Multivariate form is given. Chapter 3 provides the detailed description of our data-set. Lastly, the analysis with the results and the numerical methods are presented, followed by the conclusions.

2 Methods

2.1 Information Theory

In the early decades of the 20th century, Bell Labs laid the foundations for information theory. The major contribution was given by Claude Shannon, who built a mathematical theory of communication, the results of which still stand today [20].
Shannon's interest was how to transmit information over a channel in the most efficient way possible. The analysis introduced the idea of entropy of signals and channels [21].
In its classical and original interpretation, production of entropy quantifies the irreversibility of a process. More precisely, it is a function of state, depending on macroscopic observables of an equilibrium system. The difference in entropy between two equilibrium states A and B is defined as [22]:

$$H(B) - H(A) = \int_A^B \left(\frac{dQ}{T}\right)_R \tag{2.1}$$

where A and B are connected by a reversible transformation R. The system is at equilibrium at every point on the path $A \rightarrow B$, possessing a definite temperature T_{sys}, and exchanging an infinitesimal amount of heat dQ with a thermal bath at the same temperature $T = T_{sys}$.
The second law of thermodynamics states that [22], for a reversible process, the total entropy - i.e. that of the system and anything it has interacted with - remains the same. For any irreversible transformation, however, it increases.

It was thanks to Boltzmann that it has been possible to move from an axiomatic definition of entropy to a formal definition in terms of microscopic states [23]. Formally, Boltzamm proved a connection between the thermodynamical H and the amount of states in phase-space available to a system.

$$H_B(\mathcal{E}, V, N) = k_B \log \Omega(\mathcal{E}, V, N) \tag{2.2}$$

where $\Omega(\mathcal{E}, V, N)$ is the number of microscopic states that correspond to a macroscopic state with energy $\mathcal{E}$, volume V and N number of particles; while k_B is the Boltzmann constant. In this way, as more microscopic states are present, the larger the entropy will be; leading to the interpretation of entropy as some sort of "disorder" [23].
An even more general interpretation of entropy comes from information theory, where H is a measure of the experimenter's ignorance about the system [23].

2.1.1 Information Entropy

In information theory there is an additional way of looking at entropy: it is as a measure of our ignorance about a system [23]. To understand this we need to go back to the definition of information itself. Consider a discrete event space $E = \{x_i\}_{i=1,\dots,N}$, with probabilities $p_i \geq 0$, such that $\sum_i p_i = 1$. We want to quantify the amount of information $\eta(x_i)$ acquired by the observation of event x_i occurring. Shannon wanted an information measure which satisfied a number of conditions, notably:

- it should be additive for independent pieces of information:

  $$\eta(P[x_1 \wedge x_2]) = \eta(x_1) + \eta(x_2) \tag{2.3}$$

  where x_1 and x_2 are two independent events with, respectively, probability p_1 and p_2;

- it should reflect likelihood of events, in particular capturing increasing uncertainty associated with an increasing number of (equally likely) events:

  $\eta(x)$ is a decreasing function of p,

  the probability that x occurs;

- it should be continuous with respect to changes in these likelihoods.

He was interested in how much information a message conveyed. If something is very likely to happen, the information gleaned from it happening is not very great, a bit like the sun rising in the morning does not actually tell us very much. On the other hand, rare events (such

as the sun shining while it is raining) convey a great deal of information because they are relatively surprising. Thus, his measure of the information, $\eta(x)$, of an event x, was the log of the probability, $p(x)$, of x happening, being observed:

$$\eta(x) = -\log_2 p(x) \tag{2.4}$$

Shannon used natural logs, giving information in *nats*. When we consider Gaussian variables, natural logs appear directly, but in most cases we shall use logs to base 2, denoted $\log_2$, giving information in *bits*, the more common unit today. One can interpret the values of $\eta(x)$, in bits, as the optimal number of yes/no questions that one needs to ask (on average) to determine the value of x.

Given this definition of information, the entropy is now the average information over sets of events, which can be measured as repeated observations over time, or over sets of different realisations of a system.

If we average or take the expectation value of the information according to the probability of each event occurring, we end up with the Shannon entropy:

$$H(X) = E[\eta(x)] = -\sum p(x) \log_2 p(x) \tag{2.5}$$

We also need the idea of conditional entropy, the uncertainty left after we have taken into consideration some context:

$$H(X|Y) = \sum_y p(y) H(X|y) \tag{2.6}$$

where

$$H(X|y) = -\sum_x p(x|y) \log_2 p(x|y) \tag{2.7}$$

and $p(x|y)$ is the conditional probability.

2.1.2 Mutual Information

The mutual information is the amount of shared information between X and Y. It is a measure of their statistical dependence. Thus, we should be able to take the entropy of X and subtract from it the entropy of X given Y, since this chunk of the entropy has, by definition, nothing to

do with Y. This is exactly the case as in Eqn. 2.6:

$$I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{2.8}$$

which is clearly symmetric in X and Y.
The mutual information can be thought of as a non-linear form of correlation. The corollary of this is that:

$$I(X:Y) = 0 \iff X \text{ is independent of } Y$$

An important generalisation of mutual information, which is crucial to the development of transfer entropy is the idea of mutual information between two processes, X and Y, conditioned on a third process, Z. In fact the expression for the conditional mutual information, $I(X:Y|Z)$, is very straightforward. We simply condition each of the entropy terms in Eqn. 2.4:

$$I(X:Y|Z) = H(X|Z) - H(X|Y,Z) \tag{2.9}$$

or with the following conditional independence criterion:

$$I(X:Y|Z) = 0 \iff X, \text{ conditional on } Z, \text{ is independent of } Y.$$

2.1.3 Transfer Entropy

Given jointly distributed random variables X, Y -discrete or continuous, and possibly multivariate -we have seen that the mutual information $I(X:Y)$ furnishes a principled and intuitive answer to the questions:

- How much uncertainty about the state of Y is resolved by knowing the state of X (and vice versa)?
- How much information is shared between X and Y?
- How may we quantify the degree of statistical dependence between X and Y?

Suppose now that, rather than static variables, we have jointly distributed sequences of random variables X_t, Y_t labelled by a sequentially enumerable index $t = \ldots, 1, 2, 3, \ldots$. Intuitively the processes X_t, Y_t may be thought of as an evolution in time t of some unpredictable

variables X, Y, that is, random time-series processes. Such joint or multivariate stochastic processes are natural models for a huge variety of real-world phenomena, from stock market prices to neural signals, which may be viewed as non-deterministic dynamic processes.
How, then, might we want to frame, interpret and answer comparable questions to the above for dynamic stochastic processes rather than static variables? We may, of course, consider the mutual information $I(X_t : Y_t)$ between variables at a given fixed time t. But note that, by jointly distributed for stochastic processes, we mean that there may be dependencies within any subset $X_t, Y_s : t \in T, s \in S$ of the individual variables. Thus, for instance, X_t, the variable X as observed at time t, may have a statistical dependency on its value X_{t-s} at the earlier time $t - s$, or indeed on its entire history $X_{t-1}, X_{t-2}, \ldots$, or the history $Y_{t-1}, Y_{t-2}, \ldots$ of the variable Y. A particularly attractive notion is that of quantifying a time-directed transfer or flow of information between variables. Thus we might seek to answer the question:

- How much information is transferred (at time step t) from the past of Y to the current state of X (and vice versa)?

This information transfer, which we would expect - unlike the contemporaneous mutual information $I(X_t : Y_t)$ - to be asymmetric in X and Y, is precisely the notion that transfer entropy aspires to quantify [24].
The notion of transfer entropy (TE) was formalised by Thomas Schreiber [25]. Schreiber realised that an obvious candidate for a time-asymmetric measure of information transfer from X to Y, namely the lagged mutual information $I(X_t : Y_{t-s})$ [25, 26], is unsatisfactory for the reason that it fails to take into account shared history (as well as common external driving influences) between the processes X and Y, and that this is likely to lead to spurious inferences of directed information transfer.
Information theory supplies just the tool to effect this accounting: we must condition on the past of Y as a conditional mutual information. Such conditioning removes any redundant or shared information between current Y and its own past, but also includes any synergistic information about current Y in the source X that can only be revealed in the context of the past of Y.

$$T_{X \to Y}(t) = I(Y_t : X_{t-1} | Y_{t-1}) = H(Y_t | Y_{t-1}) - H(Y_t | Y_{t-1}, X_{t-1}) \tag{2.10}$$

Nevertheless it is possible that the shared information between the target and its past extends to a longer history length and that the earlier values of the source contain additional information

about the target. For these reasons it is needed to define also the general form of the (k, l)-history transfer entropy:

$$T^{(k,l)}_{X\to Y}(t) = I(Y_t : X^{(l)}_{t-1} | Y^{(k)}_{t-1}) = H(Y_t | Y^{(k)}_{t-1}) - H(Y_t | Y^{(k)}_{t-1}, X^{(l)}_{t-1}) \tag{2.11}$$

where l and k are respectively the history lenght of the source X and the target Y. The key idea is that $T_{X\to Y}(t)$ may be interpreted intuitively as the degree of uncertainty about current Y resolved by the past states X and Y, over and above the degree of uncertainty about current Y already resolved by its own past state alone.

2.2 Relationship with Granger Causality

As mentioned in the introduction to this thesis, transfer entropy is closely related to and shares a common history with Wiener–Granger causality (Granger causality for short).
Granger causality is based on the premise that cause precedes effect, and a cause contains information about the effect that is unique, and is in no other variable. In its purest form, the essence of the idea is surprisingly close to that of transfer entropy. Let $F(y_t | y^{(k)}_{t-1}, x^{(l)}_{t-1})$ denote the distribution function of the target variable Y conditional on the joint (k, l)-history $Y^{(k)}_{t-1}, X^{(l)}_{t-1}$ of both itself and the source variable X, and let $F(y_t | y^{(k)}_{t-1})$ denote the distribution function of Y_t conditional on just its own k-history. Then [27, 14] variable X is said to Granger-cause variable Y (with lags k, l) if and only if

$$F(y_t | y^{(k)}_{t-1}, x^{(l)}_{t-1}) \neq F(y_t | y^{(k)}_{t-1}) \tag{2.12}$$

In other words:

> X Granger-causes $Y \iff$ Y, conditional on its own history, is not independent of the history of X

The connection with transfer entropy is clear: in fact (2.12) holds precisely when $T^{(k,l)}_{X\to Y} \neq 0$. Thus Transfer Entropy might be construed as a non-parametric test statistic for pure Granger causality.
Granger's parametric formulation was, specifically, based on linear vector auto-regressive (VAR) modelling [28]. X_t, Y_t are assumed to be multivariate real-valued, zero-mean, jointly stationary

stochastic processes. Following Geweke [29] , we consider the nested VAR models:

$$X_t = A_1 X_{t-1} + \cdots + A_k X_{t-k} + B_1 Y_{t-1} + \cdots + B_l Y_{t-l} + \varepsilon_t \tag{2.13}$$

$$X_t = A_1^{'} X_{t-1} + \cdots + A_k^{'} X_{t-k} + \varepsilon_t^{'} \tag{2.14}$$

The parameters of the models are the VAR coefficient matrices A_i, B_j, $A_i^{'}$ and the covariance matrices $\Sigma = c(\varepsilon_t)$, $\Sigma^{'} = c(\varepsilon_t^{'})$ where $\varepsilon_t, \varepsilon_t^{'}$ are the residuals, assumed to be serially uncorrelated; (2.13) and (2.14) are referred to, respectively, as the full and reduced models.
The $X \to Y$ Granger Causality statistic stands to quantify the degree to which the full model yields a better prediction of the target variable than the reduced model. The most convenient form for the Granger causality statistic is given by

$$F_{X\to Y}^{(k,l)} = \log \frac{|\Sigma^{'}|}{|\Sigma|} \tag{2.15}$$

where $|\cdot|$ denotes the matrix determinant.
Adopting an approach based on a maximum-likelihood (ML) framework, we note that $F_{X\to Y}$ is precisely the log-likelihood ratio statistic for the model (2.13) under the null hypothesis

$$H_0 : B_1 = B_2 = \cdots = B_l = 0 \tag{2.16}$$

Note that, given that X_t, Y_t is described by the model (2.13), the null hypothesis (2.16) is precisely the negation of condition (2.12) for non-causality. An immediate payoff of the ML approach is that we have an (asymptotic) expression for the sample distribution of the statistic $F_{X\to Y}$ as a χ^2 with degrees of freedom equal to the difference in number of free parameters between the full and reduced models.
Finally, Barnett et al. [30] proved the following theorem, stating that Granger Causality and Transfer Entropy are equivalent for gaussian variables:

> If the joint process X_t, Y_t is Gaussian (more precisely, if any finite subset $\{X_{t_1}, Y_{t_2} : (t_1, t_2) \in S\}$ of the variables is distributed as a multivariate Gaussian) then there is an exact equivalence between the Granger causality and transfer entropy statistics:
>
> $$T_{X\to Y}^{(k,l)} = \frac{1}{2} F_{X\to Y}^{(k,l)} \tag{2.17}$$

For some aspects, Granger causality offers some obvious advantages over non-parametric

transfer entropy as a data-driven, time-directed, functional analysis technique; in particular the ease and efficiency of VAR model parameter estimation as compared with the difficulties of entropy/mutual information estimation, as well as the existence of known theoretical sampling distributions for statistical inference. We might ask, then, why we should bother with (non-parametric) transfer entropy at all. The answer depends largely on the nature of the data and the stochastic generative processes underlying it. In particular, for this thesis, a multivariate analysis is investigate; pushing for a transfer entropy approach.

2.3 Multivariate Transfer Entropy

With many systems there are many interacting variables, so we need to be able to handle additional influences on the pairwise interaction we have discussed so far. When a third (possibly multivariate) process, Z_t, say, is jointly distributed with the processes X_t ,Y_t then the pairwise, bivariate or apparent transfer entropy $T_{X\to Y}$ may report a spurious information flow from X to Y, due to (possibly lagged) joint influences of Z on X and Y (i.e. $Z \to X$ and $Z \to Y$). This is known as a common driver effect. Similarly, $T_{X\to Y}$ may report a spurious information flow from X to Y due to cascade effects, e.g. where we actually have $X \to Z \to Y$. Further, $T_{X\to Y}$ will not detect any synergistic transfer from X and Z together in these scenarios. It is, however, a simple matter to discount redundant joint influences and include synergies by conditioning on the past of Z. We thus define conditional transfer entropy [31, 32]:

$$T^{(k,l,m)}_{X\to Y|Z}(t) = I(Y_t : Y^{(l)}_{t-1}|Y^{(k)}_{t-1}, Z^{(m)}_{t-1}) = H(Y_t|Y^{(k)}_{t-1}, Z^{(m)}_{t-1}) - H(Y_t|X^{(k)}_{t-1}, X^{(l)}_{t-1}, Z^{(m)}_{t-1}) \quad (2.18)$$

$T_{X\to Y|Z}(t)$ may be interpreted intuitively as the degree of uncertainty about current Y resolved by the past state of X, Y and Z together, over and above the degree of uncertainty about current Y already resolved by its own past state and the past state of Z.
A case of particular practical importance is where we have a system of n jointly distributed processes $X_t = (X_{1,t}, \ldots, X_{n,t})$. Then since, as we have seen, the pair-wise transfer entropies $T_{X_j\to X_i}(t)$, $i, j = 1, \ldots, n$ are susceptible to confounds due to common influences of the remaining X_k, an alternative measure of pairwise information flows in the full system X is given by the pairwise- or bivariate-conditional or complete transfer entropies [31]:

$$T_{X_j\to X_i|X_{[ij]}}(t) = I(X_{i,t} : X_{j,t-1}|X_{[ij],t-1}) = H(X_{i,t}|X_{[j],t-1}) - H(X_{i,t}|X_{t-1}) \quad (2.19)$$

where the notation $[\dots]$ indicates omission of the corresponding indices.
Similarly, we may define collective transfer entropy [32] as the transfer from some multivariate set of n jointly distributed processes $X_t = (X_{1,t}, \dots, X_{n,t})$ to a specific univariate process, Y:

$$T^{(k,l)}_{X\to Y}(t) = I(Y_t : X^{(l)}_{t-1} | Y^{(k)}_{t-1}) \tag{2.20}$$

In particular, we focus on the case with $n = 2$, for which (2.17) becomes

$$T^{(k,l)}_{X\to Y}(t) = I(Y_t : X^{(l)}_{1,t-1}, X^{(l)}_{2,t-1} | Y^{(k)}_{t-1}) \tag{2.21}$$

that is, as we will see, exactly the quantity of our interest.

2.4 Kernel Estimation

Before to proceed with our analysis, we give a look to the computational methods exploited for this work. The Java Information Dynamics Toolkit (JIDT) [3] is a Google code project which provides a standalone, (GNU GPL v3 licensed) open-source code implementation for empirical estimation of information-theoretic measures from time-series data. While the toolkit provides classic information-theoretic measures (e.g. entropy, mutual information, conditional mutual information), it ultimately focuses on implementing higher-level measures for information dynamics. It provides implementations for both discrete and continuous-valued data for each measure, including various types of estimator for continuous data.
For continuous variables one could simply discretise or bin the data and apply discrete estimators. This is a simple and fast approach, though it is likely to sacrifice accuracy. Alternatively, we can use an estimator that harnesses the continuous nature of the variables, dealing with the differential entropy and probability density functions. The latter is more complicated but yields a more accurate result. For this work we chose a *kernel estimator*.
With this method the relevant joint PDFs are estimated with a kernel function Θ, which measures "similarity" between pairs of samples x_n, y_n and $x_{n'}, y_{n'}$ using a resolution or kernel width r:

$$p_r(x_n, y_n) = \frac{1}{N} \sum_{n'=1}^{N} \Theta \left(\left| \begin{pmatrix} x_n - x_{n'} \\ y_n - y_{n'} \end{pmatrix} \right| - r \right) \tag{2.22}$$

By default Θ is the step kernel:

$$\Theta(x) = \begin{cases} 0 & x > 0 \\ 1 & x \leq 0 \end{cases} \tag{2.23}$$

and the norm $|\cdot|$ is the maximum distance. This combination – a box kernel – is what is implemented in JIDT. It results in $p_r(x_n, y_n)$ being the proportion of the N values which fall within r of x_n, y_n in both dimensions X and Y.
Kernel estimation can measure non-linear relationships and is model-free, though is sensitive to the parameter choice for resolution r. Selecting a value for r can be difficult, with a too small value yielding under-sampling effects whilst a too large value ignores subtleties in the data. In our analysis the default value $r = 0.5$ remained unchanged.

3 Data

The importance of high-quality data as a proxy for a social system and for studying cause-effect relationships has already been discussed. In this framework, a main part of this work concerns data collection. First of all it is needed to define what we mean with *The Web*. To the best of our knowledge, three distinct classes of online data sources have been investigated for financial predictions, namely: news media, web search query data and social media feeds [33]. For this thesis we collected the volume of websites news and the volume of tweets related to the 30 companies that form the Dow Jones Index in a period of 9 years (from October 2011 to December 2020), on a daily basis. Let's see the details.

3.1 Web News

Access to structured information regarding the financial market with its various instruments and indicators is available for several decades, but the systematic quantification of unstructured information hidden in news from diverse Web sources is of relatively recent origin.
We base our analyses on a newly developed text processing pipeline, New-Stream, which was designed and implemented within the scope of the EU FP7 projects FIRST[1] and FOC[2]. New-Stream continuously downloads articles from more than 200 worldwide news sources, such as yahoo.com, reuters.com, nytimes.com and bbc.co.uk. It extracts the content, stores complete

[1] http://project-first.eu/
[2] http://www.focproject.eu/

texts of articles and extracts finance-related entities. It is a domain-independent data acquisition pipeline but is biased towards finance by the selection of news sources and the taxonomy of entities that are relevant to finance. More details can be found in [17].
Thanks to this new portal, we have been provided – directly by one of the authors – with a data-set of over 10 million news regarding the companies that form the S&P500 index; including news that satisfy the followings search criteria:

- full-text search: ticker symbol or "company name" in article titles;
- constrain: the document content needs to be more then 1000 characters long.

This data-set, divided for each company of interest, will compose one of our sources that we call **News**.
In order to understand better and to investigate how stock markets are influenced by online information, we selected a sub-set of news that are directly related to finance; for which the content talks about financial topics, and not general arguments about the company. Specifically, these news have been filtered using an ontology of financial terms built in collaboration with economy experts [17]. This sub-set is our second source, named **Financial News**.

Since they were already provided, the temporal range of this work is mainly due to these data.

3.2 Twitter

Social media are increasingly reflecting and influencing behavior of other complex systems, in particular social media feeds are becoming an important source of data to support the measurement of investor and social mood extraction.
Because of its willingness to share data with academia and industry, Twitter has been the primary social media platform for scientific research as well as for the consulting of businesses and governments in the last decade. In recent years, a series of publications have studied and criticized Twitter's APIs and Twitter has partially adapted its existing data streams. The newest Twitter API for Academic Research allows to "access Twitter's real-time and historical public data with additional features and functionality that support collecting more precise, complete, and unbiased data-sets." The main new feature of this API is the possibility of accessing the full archive of all historic Tweets.

The second source of our data is exactly from Twitter and consists of relevant tweets. We collected this data using the Twitter API[3] for Academic Research, which is freely made available by Twitter for research purposes upon request. For each stock of the Dow Jones Index, we collected the corresponding daily time-series by exploiting the function 'Client.get_all_tweets_count()' available through the Python Tweepy library[4] and for academic research only. As parameters of this function, we used the initial and final date (23 Oct 2011 and 17 Dec 2020, respectively) and the corresponding stock cash-tag (e.g. "$APPL" for Apple). Cash-tags are a Twitter feature used in the financial sector instead of hashtags to tag conversations in order to allow users to see all the other Tweets that include it.
To the best of our knowledge, all the available tweets with cash-tags are acquired. These will compose our third source **Tweets**.

3.3 Stock Market Volumes

The last part of our data-set is, of course, the financial data. Various financial terms have been used in previous research as proxy for market performances, from log prices return to volumes[16, 15]. The choice of the variables does not affect the outcome of the present work, as a matter of fact it has been shown how volume shifts can be correlated with price movements [34]. For this thesis the trading volume has been chose.
The daily financial data for all of stocks are publicly available from Yahoo! Finance [5] and we collected them trough the yfinance library[6] available in Python. We focused our attention on the daily trading volumes, forming the target of our information-flow analysis: the **Trading Volume**.

The whole data-set has been filtered for the days in which the stock market is open. An example (Apple) of the trends of our time series is shown in Figure 3.1.

[3] https://developer.twitter.com/en/docs/twitter-api
[4] https://docs.tweepy.org/en/stable/client.html
[5] http://finance.yahoo.com/
[6] https://pypi.org/project/yfinance/

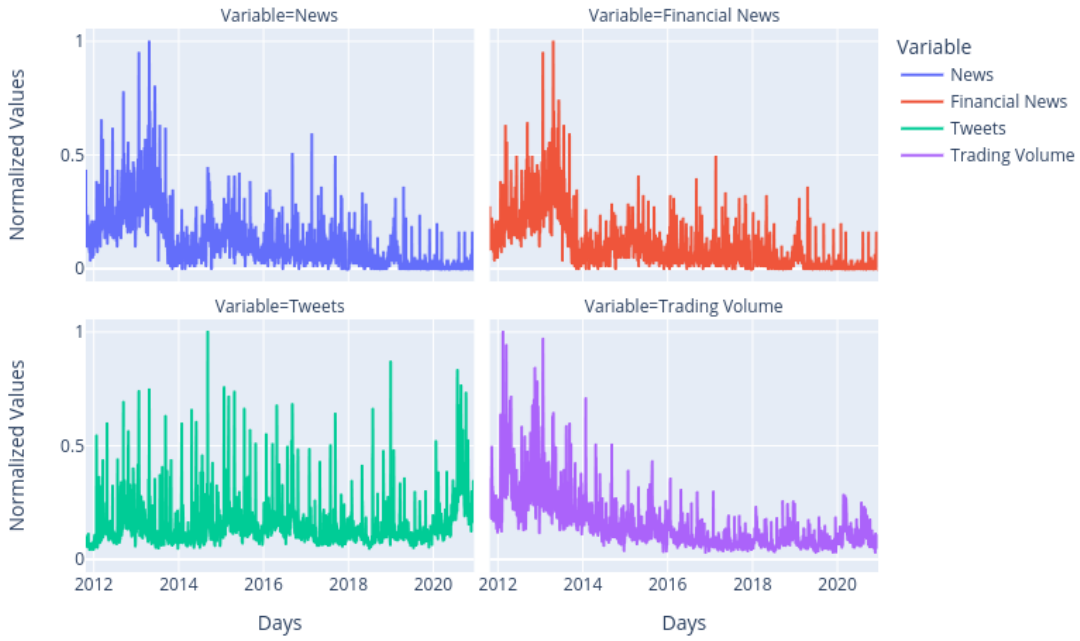

Figure 3.1: Apple example data-set.

4 Results

4.1 Granger Causality

In practice, we have three different available data sources and we want to understand if there is information flow between these and the trading volume for each company in consideration. The first attempt is done with Granger Causality; probably the main definition of causality in econometrics that aims to state if a time-series X 'causes' another time-series Y. The method follows the formalization of a statistical test:

- a null hypothesis H_0 for which X does not 'granger-cause' Y is built;
- the test returns a p-value;
- if the p-value is greater than a predefined threshold H_0 is not rejected, otherwise we can reject the null hypothesis and we can estabilish the causality relation between X and Y.

Hence the results come in the form of a p-value, for this reason it is needed to define a statistical threshold for which we can reject the H_0 hypothesis. In this case we choose two values, namely 0.01 and 0.05.
Nevertheless, before to start the analysis, a study on the-time lag has to be performed. The data are daily collected and we need to be sure that the minimum p-value corresponds to a lag of exactly one day; as expected given previous studies [15]. In order to do this the time-lag effect has been studied. In particular, it has been done for each company, and then represented by the

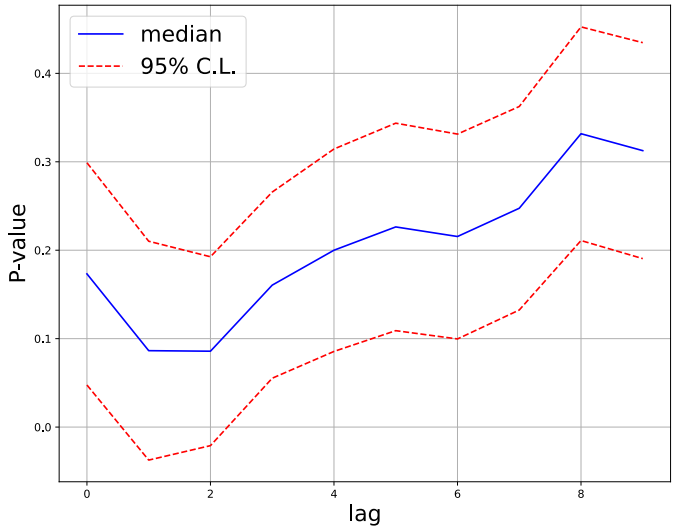

Figure 4.1: Twitter Granger Causality time-lag study.

median and its 95% C.L. among all the companies, as shown in Figure 4.1 in case of Twitter as source. From Figure 4.1 it is clear that the number of tweets can better anticipate the trading volume one day in advance. Therefore the following analysis is based on this assumption.
Going back to our goal, we want to see for how many companies we can reject the null hypothesis of the Granger Causality test between the sources and the target on a one day lag, given the two chosen threshold. From now on this is the main metrics with which we evaluate the results of the methods, expressed as percentage over the total number of companies.
Before to perform the test it is important to keep in mind the hypothesis done on the data. As a matter of fact Granger Causality can be applied only for stationary data [27]. In order to respect this constraint we apply a difference transform on our data, subtracting the present step to the next step:

$$X_{diff}(t) = X(t+1) - X(t) \tag{4.1}$$

computing the daily shifts and removing the trends.
The results for the Granger Causality are shown in Table 4.1

Source	p-value = 0.01	p-value = 0.05
News	27%	37%
Financial News	33%	40%
Tweets	27%	33%

Table 4.1: Granger Causality results.

4.2 Transfer Entropy

The main character of this thesis is, without doubts, the Transfer Entropy. The advantages in moving to a non-parametric information theory measure are many. In particular, in this context, we care about three major points that make us prefer it instead of the Granger Causality:

- it is able to capture non-linear relations;
- it works better with non-gaussian data;
- it provides an absolute value, allowing to identify a 'strenght' of the information flow and not only a yes/no test.

Strong with these arguments, we perform our analysis. It starts in the same way as before, with a time-lag study that find the optimal lag day for which the Transfer Entropy is maximized. In Figure 4.2 the Twitter results are showed, following the same median and C.L. procedure already discussed.
In this case we find the maximum value of TE in correspondence of a time lag $\Delta L = 0$ days. This could seem to contradict previous results but it is explained by the definition of Transfer Entropy and by its computational implementation. We recall eq. 2.8 and we observe how the mutual information is computed between the target X_t and a step back of the source Y_{t-1}, leading to a pre-embedded one step (one day in our case) lag. This concept is made even clearer in one illustration (Figure 4.3) provided by JIDT, the online library exploited for this work, introduced in Chapter 2.

Thanks to this observations it is possible to perform the same kind of analysis done with Granger Causality and to compare with it.

4.2.1 Statistical Significance

The first goal when passing to a Transfer Entropy analysis is to understand how to compare this measure with Granger Causality or, more precisely, how to match an absolute value with

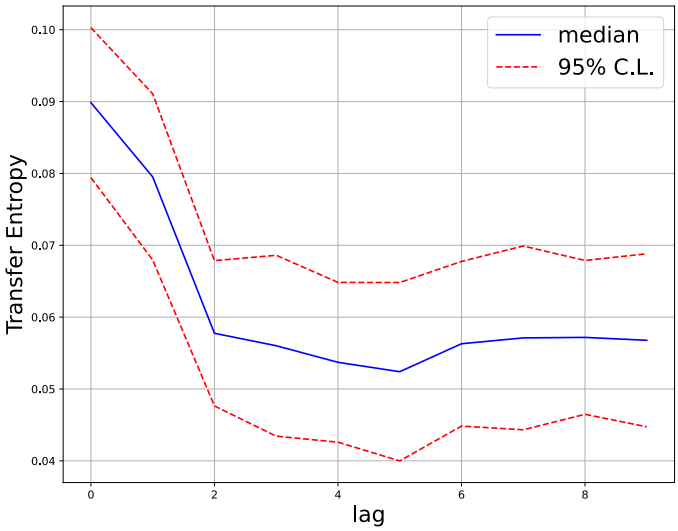

Figure 4.2: Twitter Transfer Entropy time-lag study.

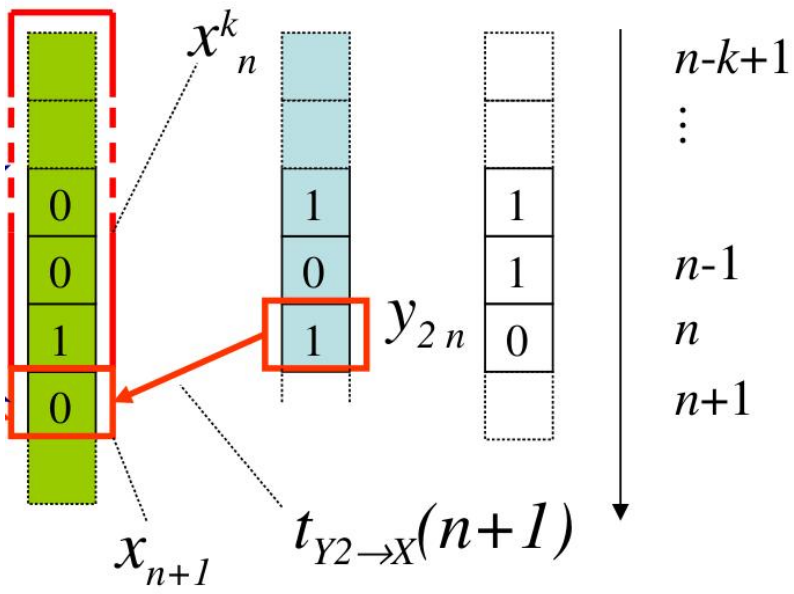

Figure 4.3: Transfer Entropy computational scheme [3].

Source	C.L. = 99%	C.L. = 95%
News	60%	73%
Financial News	67%	83%
Tweets	90%	97%

Table 4.2: Transfer Entropy results.

a statistical test. To address this, standard sub-sampling techniques such as permutation testing and bootstrapping may be employed for significance testing and estimation of confidence intervals for the transfer entropy [24]. This is done by forming a null hypothesis H_0 that there is no such relationship, and making a test of evidence (our original measurement) in support of that hypothesis.
In practice we adopt a shuffling approach producing null-hypothesis transfer entropy values from independently shuffled time-series over the same domain, containing no causal relationships. By calculating the mean and standard deviation of the shuffled transfer entropy, we estimate a Z-score that identify the significance of a causal result as the distance between the result and the average shuffled result, standardizing by the shuffled standard deviation:

$$Z = \frac{T_{X\to Y} - E[T_{X^s\to Y}]}{\sigma[T_{X^s\to Y}]} \tag{4.2}$$

where $T_{X\to Y}$ is the transfer entropy of the temporally ordered sample, $E[T_{X^s\to Y}]$ is the average transfer entropy over all shuffled realizations and $\sigma[T_{X^s\to Y}]$ is the standard deviation of the sample of shuffled realizations. This quantity corresponds to the degree to which the result lies in the right tail of the distribution of the zero-causality shuffled samples, and hence how unlikely the result is due to chance. Therefore, the Z-score represents the significance of the excess transfer entropy in the un-shuffled case.
At the end it is important to manage to compare the Granger Causality results. In order to do this we build a confidence level for which the Z-score test can fail and hence accept the alternative hypothesis for which we have a direct relationships. Starting from the p-values 0.05 and 0.01 we chose C.L. at 95% and 99% that, for a one-tail distribution, correspond to $Z = 1.645$ and $Z = 2.32$. In this framework we can count all the companies that present a Z-score larger than the two thresholds. The results are presented in Table 4.2.

A deeper analysis could be conducted observing the distributions of the computed Z-scores. In fact, focusing on the 'causal flow strenght', one could individuate patterns among different

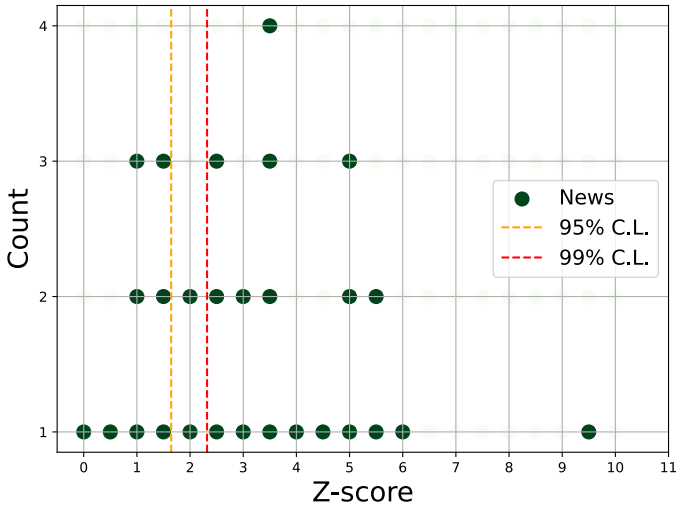

Figure 4.4: News Z-score distribution.

kind of companies and conduct an *ad hoc* analysis on a specific stock. In Figure 4.4, Figure 4.5 and Figure 4.6 we present the number of financial entities that are individuate by a specific Z-score. To conduct a specific study on particular firms is out of the scope of this thesis. Nevertheless we can point out how the evident News outlier correspond with one of the member of the Financial News couple (namely Microsoft); while the Twitter outliers are completely different companies (namely Cisco, Disney and Nike).

4.2.2 Data Distribution

It is instantly clear how the Transfer Entropy manage to state a relation of information flow between the sources and the trading volume for many more companies with respect to the Granger Causality. If this results are confirmed, they are of great interest since they show two aspects:

- first, the sub-set of the Financial News is way more informative than the complete Web News set;
- second, Twitter is revealed as the most important source that can be used as proxy of the Trading Volumes for almost all the companies in consideration, and so it can have taken the role of news spreading all over the world.

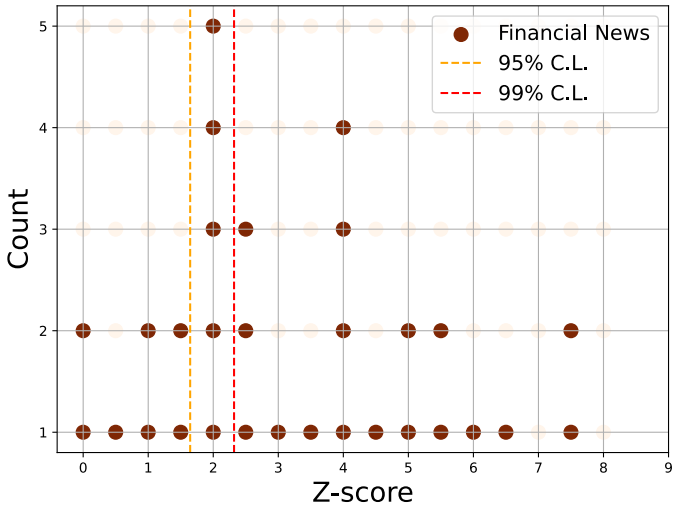

Figure 4.5: Financial News Z-score distribution.

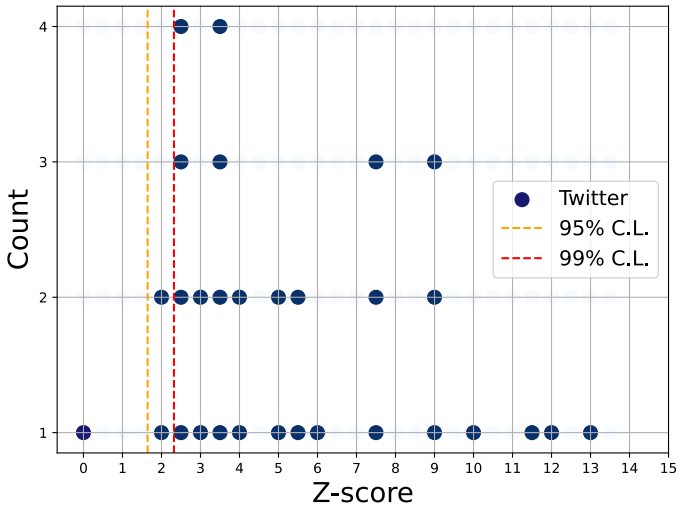

Figure 4.6: Twitter Z-score distribution.

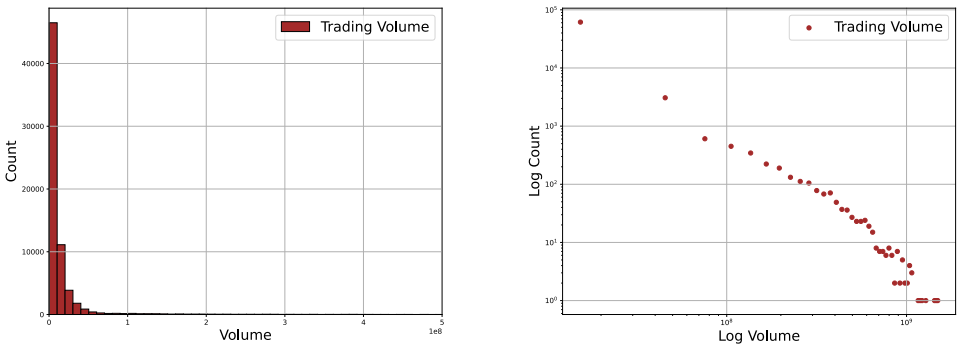

Figure 4.7: Trading Volume distribution (log-scale on the right).

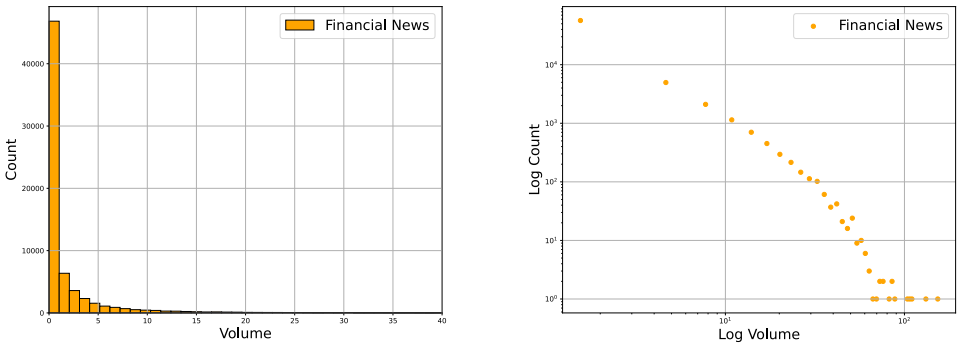

Figure 4.8: Financial News distribution (log-scale on the right).

From this point of view, if we take these result as true, it is important to understand why we have to neglect the Granger Causality outcomes. At first glance, the most obvious assumption is that the hypothesis made on the data do not match with the ones required by Granger. As a matter of fact, as already stated in [15], financial data are non-Gaussian distributed. They instead present fat-tailed distributions. It is easy to give a look the total data distributions, and this is exactly the case of our data-set, as showed in Figure 4.7 Figure 4.8 and Figure 4.9; where the data are plotted in simple histograms, as well with a log-scale for the y-axes.
Given these observations, the differences between the results of Granger Causality and Transfer Entropy are more obvious, and we can continue with our analysis based on the information theory tool.

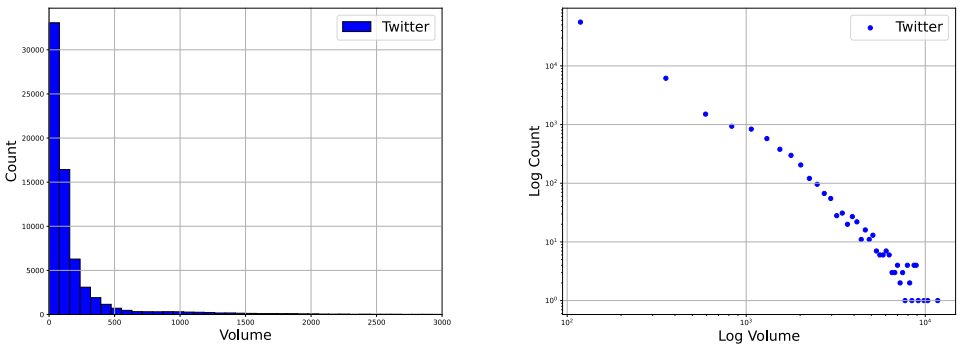

Figure 4.9: Twitter distribution (log-scale on the right).

4.3 Multivariate Analysis

Once that the Transfer Entropy results are confirmed it is possible to exploit the power of this measure and to take advantage of its main feature: the ability to be multivariate.
Until now we have understood how the sub-set of Financial News is more informative of the complete set of News. Starting from this point we want to see if combining the signal present in our most informative sources (Financial News and Twitter), one can get better outcomes. In order to do this we recall the definition of *Collective Transfer Entropy* defined in section 2.3. In particular we focus on the bivariate case, that allows to estimate the information flow from a system of $n = 2$ sources to our target of Trading Volumes.
From an operating perspective, comparing absolute values of Transfer Entropy should be immediate. Nevertheless, it is possible that different pairs of variables in the same system experience different types of dynamics, and one should correct somehow for these differences before making comparisons. One key method here is bias correction, since bias could be higher or lower under different dynamics. This may be performed for other estimator by computing the null distribution $T_{X^s \to Y}$ as introduced before and then subtracting out the mean $E[T_{X^s \to Y}]$ of this distribution. Another step is to consider Transfer Entropy as a fraction of the maximum value that it could potentially take under the given dynamics. Marschinski and Kantz [35] proposed the Relative Explanation Added (REA):

$$REA = \frac{T_{X \to Y} - E[T_{X^s \to Y}]}{H'_Y(t)} \tag{4.3}$$

Source	Count
Financial News, Twitter > Financial News	93%
Financial News, Twitter > Twitter	80%

Table 4.3: Multivariate Transfer Entropy results.

which first removes the bias and then normalises by the entropy rate $H'_Y(t)$, that represents the fraction of information in the target Y not explained by its own past that is explained by X in conjunction with that past.

Thanks to this normalization, we can make a comparison between the Transfer Entropy results in the two cases: univariate and bivariate. In particular we want to investigate for how many companies the multivariate analysis presents an information flow larger than the simple one; and to understand the differences adding one source with respect to the other. Table 4.3 presents the results.

In practice we are comparing the differences in the TE values between measures performed with both sources with respect to a measure computed with just one of the two. In this way it is possible to analyze if the Multivariate returns better results and which one of the two sources provides more information. These numbers are not surprising, given the high values already recorded in the univariate case. However they reveal the consistency of this analysis, confirming the higher predicting power of Twitter with respect the Web News. Indeed, adding the information contained in Twitter, we get more predicting power than the other case of adding the Financial News source.

5 Conclusion

This thesis takes part in a context in which physicist are more often interested in the study of economics and financial topics, and in which the recent digital revolution has led to the emergence of new data-driven methods to investigate social systems. In this framework financial markets are specifically interesting since they provided a lot of data and they are highly linked with users behaviours.
In particular this work starts from a series of scientific papers that study the predictions of financial markets starting form web sources. Here, the Granger Causality analysis between Web News, Tweets and Stock Market Volumes has been re-proposed, supported by new and contemporary data. Nevertheless, the non-stationary and non-gaussian nature of financial data required a different approach that can overcomes the limits of linear statistics. This tool has been found in information theory, particularly with the use of Transfer Entropy. Thanks to this non parametric measure, the following results have been observed:

- the sub-set of Financial News contains more information than the whole set of News, letting think that a more sever filtering can produce a more precise market's proxy;
- Twitter is revealed as the most important source of information, able to anticipate market performances for almost all the companies (except one).

Given these outcomes, it is possible to note how Twitter is playing the role of main information spreader around the world, substituting news job.
In this direction, as component of novelty, the use of Transfer Entropy allows a step forward

in the study of different sources to forecast financial markets; making accessible a multivariate analysis. Thanks to this quantity, it has been possible to investigate the combination of multiple sources in information flow detection. In particular, Financial News and Twitter have been taken into consideration. Consistently with previous results, the leading role of Twitter has been confirmed. Moreover, with respect the univariate case, a significant increase in performance has been detected; making the multivariate analysis the preferred version.
These considerations lead to the following conclusions and future perspectives:

- the power and flexibility of Transfer Entropy in this field of application are clear, future works can try to optimize the computational efficiency trough different methods; such as the Kraskov-Stögbauer-Grassberger (KSG) technique or the Symbolic Transfer Entropy;
- Twitter has been confirmed as main vector of information world-wide; as a future step one can investigate more limiting constraints on search queries; as well as a sentiment analysis on News and Twitter corpora;
- as best of our knowledge, this thesis open the path for multivariate analysis in the financial field; in the future different kind of sources can be tried, such as Google Trends, Reddit or Telegram;
- ultimately, this work is of course affected by limitations, being restricted to a small number of companies, as well to a defined time range. An immediate improvement can be tried focusing on bigger indices, such as the S&P500; or on a different time-scale; going under the daily base. Moreover, a deeper analysis can be made on specific stocks, going to study the outliers with very small or big Z-score values and their relation with the company history.

References

[1] A.-L. Barabási and M. Pósfai, *Network science*. Cambridge: Cambridge University Press, 2016. [Online]. Available: http://barabasi.com/networksciencebook/

[2] V. Tyler, "Spurious correlation," 2011. [Online]. Available: https://www.tylervigen.com/spurious-correlations

[3] J. T. Lizier, "JIDT: an information-theoretic toolkit for studying the dynamics of complex systems," *CoRR*, vol. abs/1408.3270, 2014. [Online]. Available: http://arxiv.org/abs/1408.3270

[4] D. A. Easley and J. M. Kleinberg, *Networks, Crowds, and Markets - Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[5] G. King, "Ensuring the data-rich future of the social sciences," *Science*, vol. 331, no. 6018, pp. 719–721, 2011. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1197872

[6] A. Vespignani, "Predicting the behavior of techno-social systems," *Science (New York, N.Y.)*, vol. 325, pp. 425–8, 08 2009.

[7] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, jan 2006. [Online]. Available: https://doi.org/10.1038%2Fnature04292

[8] J.-P. Bouchaud, "The (unfortunate) complexity of the economy," 2009. [Online]. Available: https://arxiv.org/abs/0904.0805

[9] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. White, "Economic networks: The new challenges," *Science (New York, N.Y.)*, vol. 325, pp. 422–5, 08 2009.

[10] J.-P. Bouchaud, "Economics needs a scientific revolution," *Nature*, vol. 455, no. 7217, pp. 1181–1181, oct 2008. [Online]. Available: https://doi.org/10.1038%2F4551181a

[11] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli, "The physics of financial networks," *Nature Reviews Physics*, vol. 3, no. 7, pp. 490–507, jun 2021. [Online]. Available: https://doi.org/10.1038%2Fs42254-021-00322-5

[12] G. D'Agostini, *Bayesian Reasoning in Data Analysis*. WORLD SCIENTIFIC, 2003. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/5262

[13] J. Pearl, *Causality*, 2nd ed. Cambridge, UK: Cambridge University Press, 2009.

[14] C. Granger and P. Newbold, *Forecasting Economic Time Series*, 2nd ed. Elsevier, 1986. [Online]. Available: https://EconPapers.repec.org/RePEc:eee:monogr:9780122951831

[15] I. Bordino, S. Battiston, G. Caldarelli, M. Cristelli, A. Ukkonen, and I. Weber, "Web search queries can predict stock market volumes," *PLOS ONE*, vol. 7, 07 2012. [Online]. Available: https://doi.org/10.1371/journal.pone.0040014

[16] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The effects of twitter sentiment on stock price returns," *PLOS ONE*, vol. 10, 09 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0138441

[17] P. Kralj Novak, M. Grcar, and I. Mozetic, "Analysis of financial news with newsstream, technical report ijs-dp-11892," 07 2015.

[18] P. K. Novak, L. D. Amicis, and I. Mozetič, "Impact investing market on twitter: influential users and communities," *Applied Network Science*, vol. 3, no. 1, sep 2018. [Online]. Available: https://doi.org/10.1007%2Fs41109-018-0097-9

[19] P. M., A.-F. N., and N. P. et al., "Cohesiveness in financial news and its relation to market volatility," *Scientif Reports*, 2014. [Online]. Available: https://doi.org/10.1038/srep05038

[20] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[21] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Copyright Cambridge University Press, 2003.

[22] K. Huang, *Statistical Mechanics*, 2nd ed. John Wiley & Sons, 1987.

[23] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters and Complexity*, first edition ed. Great Clarendon Street, Oxford OX2 6DP: Oxford University Press, 2006.

[24] T. Bossomaier, L. Barnett, M. Harré, and J. Lizier, *An Introduction to Transfer Entropy: Information Flow in Complex Systems*. Springer International Publishing, 2016. [Online]. Available: https://books.google.it/books?id=p8eADQAAQBAJ

[25] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, Jul 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.85.461

[26] A. Kaiser, "Information transfer in continuous processes," *Physica.*, vol. 166, no. 1, 2002.

[27] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–38, 1969. [Online]. Available: https://EconPapers.repec.org/RePEc:ecm:emetrp:v:37:y:1969:i:3:p:424-38

[28] J. D. Hamilton, *Time Series Analysis*. Princeton University Press, 1994. [Online]. Available: https://www.worldcat.org/title/time-series-analysis/oclc/1194970663&referer=brief_results

[29] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 304–313, 1982. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1982.10477803

[30] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for gaussian variables," *Phys. Rev. Lett.*, vol. 103, p. 238701, Dec 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701

[31] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, "Local information transfer as a spatiotemporal filter for complex systems," *Phys. Rev. E*, vol. 77, p. 026110, Feb 2008. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.77.026110

[32] ——, "Information modification and particle collisions in distributed computation," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 3, p. 037109, 2010. [Online]. Available: https://doi.org/10.1063/1.3486801

[33] H. Mao, S. Counts, and J. Bollen, "Predicting financial markets: Comparing survey, news, twitter and search engine data," 2011. [Online]. Available: https://arxiv.org/abs/1112.1051

[34] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Amaral, and H. Stanley, "Econophysics: financial time series from a statistical physics point of view," *Physica A: Statistical Mechanics and its Applications*, vol. 279, no. 1, pp. 443–456, 2000. [Online]. Available: https://ideas.repec.org/a/eee/phsmap/v279y2000i1p443-456.html

[35] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series. an improved estimator for transfer entropy," *European Physical Journal B*, vol. 30, pp. 275–281, 11 2002.