SCUOLA
NORMALE
SUPERIORE

Efficiency and instabilities of financial markets

Lucio Maria Calcagnile

relatori
Prof. Stefano Marmi
Prof. Fulvio Corsi

Tesi di Perfezionamento
in Matematica per le tecnologie industriali

anno 2016

Efficiency and instabilities of financial markets

Lucio Maria Calcagnile

relatori
Prof. Stefano Marmi
Prof. Fulvio Corsi

Tesi di Perfezionamento
in Matematica per le tecnologie industriali

Scuola Normale Superiore
anno 2016

Preface

This thesis tracks my research path, which started right after my master's degree in mathematics, then developed in the 3-year period in which I attended the *Perfezionamento in Matematica per le tecnologie industriali* at the Scuola Normale Superiore (January 2009–December 2011) and in the subsequent years (2012–present), within the stimulating and original research context created by the collaboration between the mathematical finance group of the SNS and LIST[1].[2] This collaboration was agreed in November 2011 with the signing of a convention between the two parts, for a 36-month period (later extended) starting from January 2012. The collaboration was signed with the stated primary goals of cooperating in doing research work in common interest areas, which were mainly—but not limited to—quantitative finance, and the exchange of knowledge, both of technical and scientific nature, and of academic and industrial interest. The SNS-LIST collaboration developed within a quantitative finance research laboratory, named QuantLab[3], created by LIST within its Advanced Technology Direction. It is in this context that I continued my research path, initially with a 12-month research associate position[4] at SNS funded by LIST within the signed convention (January 2012–January 2013), then being directly employed by LIST as quantitative analyst to continue working on the QuantLab activities (May 2013–present).

Retracing my research activity from the beginning, in 2010 my first work—made with my two master's thesis advisors Stefano Galatolo and Giulia Menconi—was published on the *Journal of Nonlinear Science*. It was an

[1] LIST S.p.A. (member of List Group) is a company developing innovative software solutions for the financial industry, with its headquarters situated in Pisa. Web site: `www.list-group.com`.

[2] In May 2011, the SNS organised the workshop *L'instabilità nei mercati finanziari: il Flash Crash un anno dopo* for a review of the scientific literature inspired by 6$^{\text{th}}$ May 2010 Flash Crash. It was during this workshop that LIST and the mathematical finance group of the SNS found out the mutual interest in the subject of market instabilities and talked about a possible collaboration to do research on this topic together.

[3] Web site: `www.quantlab.it`.

[4] *Assegno di ricerca*.

evolution of my master's thesis work and it was finally printed during the second year of my *Perfezionamento* course. In this work we studied the entropy—viewed as an ergodic theory concept—of different classes of dynamical systems, using an information theory approach.

During my *Perfezionamento* course, first under the guidance of my advisor Stefano Marmi and then with the joint mentoring of S. M. and Fulvio Corsi, the idea was born to employ the Shannon entropy as a tool to quantify randomness in high frequency financial time series, i.e., as a means to study market informational efficiency analysing high resolution market data. There was some literature about using information theory tools to study market efficiency, but the majority dealt with low frequency (daily) data and those few works actually using high frequency data suffered—from our viewpoint—from serious shortcomings. We took on this problem with a combined theoretical and empirical approach, which allowed me to learn a lot of quantitative finance topics and to study in depth the pecularities of high frequency time series, from the stylised facts of intraday price data to many different mathematical approaches used to model them, and to the deep and subtle market microstructure issues. This study was particularly important for my scientific education, because it represented the bridge from genuine mathematics and the fascinating world of quantitative finance.

Thanks to this study on market efficiency with high frequency financial data, I was able to start building a valuable expertise made up both of theoretical knowledge and practical competences, which was very useful in my subsequent work within the SNS-LIST collaboration. Doing research in QuantLab provided me with the double opportunity, on one side, to make the most of the acquired knowledge and, on the other side, to widely deepen it.

The main topic of the research developed in QuantLab was that of financial market systemic instabilities, which revealed to be a fertile research area and—to me personally—a very interesting and stimulating one. The second part of this thesis is made up of two articles written in this context.

The development of the research activity in QuantLab needed a considerable effort by the involved partners, requiring the engagement of people, the transmission of knowledge between people with different backgrounds, the building of new competences, the designing and realisation of the necessary technological structures. All this was at its very beginning as I was finishing my 3-year *Perfezionamento* course, so that the research in the QuantLab context could be carried out only in the subsequent years.

I acknowledge LIST for the funding of the research associate position I had at the Scuola Normale Superiore to collaborate in the research project

Metriche di instabilità sistemica dei mercati. I also acknowledge the support of the research grant GR12CALCAG *Entropia di Shannon ed efficienza informativa dei mercati finanziari* at the Scuola Normale Superiore.

Throughout my long research path, I have had the privilege to work with people who have taught me a lot from a scientific point of view and also with whom I have spent many enjoyable moments. I am much obliged to my advisor Stefano Marmi, for his wise guidance throughout my long and composite research path. I am also grateful to my advisor Fulvio Corsi, for his precise and valuable help. I am thankful to Fabrizio Lillo, for his valuable guidance in a research area where I had a lot to learn. I warmly thank Giacomo Bormetti, for the considerable amount of time we spent working together, for his kind encouragement and for the many enjoyable daily moments we shared.

Finally, I want to thank my parents, who, in their own way, kept encouraging me in difficult moments of my life. This thesis is dedicated to them.

Lucio Maria Calcagnile

Pisa, July 2016

Abstract

In this thesis, we consider two topics in the field of high frequency financial econometrics. The first one is the measurement of market efficiency from high frequency data, within an information theory framework. The study of this topic is performed with an analytical and empirical combined approach. The second topic is that of financial market systemic instabilities at high frequency level and is analysed mainly with an empirical and modelling approach.

In the first part of the thesis, we deal with the question of how market efficiency can be measured from high frequency data. Our interest is motivated on one hand by the availability of high resolution data, that are an incredibly rich information source about financial markets, and on the other hand by some literature on the problem of efficiency measurement that suffers, in our opinion, from serious methodological shortcomings. The process of incorporating the available information into prices does not occur instantaneously in real markets. This gives rise to small inefficiencies, that are more present at the intraday level (high frequency) than at the daily level (low frequency).

Analysing 1-minute and 5-minute price time series of 55 *exchange traded funds* (ETFs) traded at the New York Stock Exchange (NYSE), we develop a methodology to isolate true inefficiencies from other sources of regularities, such as the intraday pattern, the volatility clustering and the microstructure effects. Our tool to measure efficiency is the Shannon entropy. Since it applies to finite-alphabet symbolic sources, we work with 2-symbol and 3-symbol discretisations of the data.

In the 2-symbol discretisation, one symbol stands for the positive returns, the other for the negative returns. The zero returns, corresponding to price stationarity between consecutive records, are simply discarded. In the 2-symbol discretisation, the intraday pattern and the volatility have no effect, since they are modelled as multiplicative factors. Only the microstructure effects affect the symbolisation. In a first analysis, we test the null hypoth-

esis of the symbolised data being indistinguishable from white noise, i.e., uncorrelated data with maximum uncertainty and Shannon entropy equal to 1. Results show, as expected, that the hypothesis is not realistic and that perfect efficiency is rejected in almost all cases. The entropy measurements, however, provide us with a quantification of inefficiency. Data sampled at the frequency of 5 minutes show a higher level of efficiency than those at the 1 minute frequency, confirming that the microstructure effects are greater at higher frequencies.

Market microstructure gives rise to linear autocorrelation that have been modelled in the literature as autoregressive moving average (ARMA) processes. This motivates us to conduct an analytical study of the Shannon entropy of these processes. We find analytical expressions for the Shannon entropy of order two and three of the AR(1) and MA(1) cases. We also prove a number of results about general properties of the probabilities of finite binary sequences derived from the AR(1) and MA(1) processes, that are applicable and useful to obtain (at least partial) results about the entropies of order greater than three. All this analytical work and the results about the AR(1) and MA(1) Shannon entropy are original and we dedicate a separate chapter to this subject. We believe that these results and the whole approach have their own interest and can be useful in other contexts as well.

Relying on the ARMA modelling of asset returns, we follow two ways. For the series that are best fitted with an AR(1), an MA(1) or a simple white noise process, we define a measure of inefficiency as the (suitably normalised) difference between the Shannon entropy measured on the data and the theoretical value of the Shannon entropy of the corresponding process. For the other series, which are best fitted with ARMA(p, q) models with $p + q > 1$, we define a measure of inefficiency as the (suitably normalised) difference between 1 (the entropy of a white noise process) and the Shannon entropy of the ARMA residuals. Results show that in some cases a large part of the apparent inefficiency is explained by the linear dependence structure, while in many other cases the ARMA residuals continue to contain some (nonlinear) inefficiencies. We also rank the ETFs according to the measured inefficiency and notice that the rankings are not very sensitive to the choice of the entropy order. By rigorously testing the ARMA residuals for efficiency, we reject the hypothesis of efficiency for the large majority of the ETFs.

We find a strong relationship between low entropy and high relative tick size. This is explained by noting that an asset's price with a large relative tick size is subject to more predictable changes. We also notice an interesting relation between the inefficiency of ETFs tracking the indices of countries, with the ETFs of the Asian/Oceanic countries being very inefficient, the ETFs tracking the indices of European countries showing a moderate amount

of inefficiency and ETFs relative to American countries being among the most efficient. The levels of detected inefficiency for the country ETFs can be related to the opening time overlap between the country markets and the NYSE, where the ETFs are exchanged with a trading dynamics that can or cannot rely on a simultaneous evolution of the corresponding indices. We hypothesise that those ETFs that track indices of markets that are closed during the ETFs' trading time are deliberately given a low price, since their dynamics is only coarsely determined.

In a 3-symbol discretisation, one symbol represents a stability basin, encoding all returns in a neighbourhood of zero. Negative and positive returns lying outside of this basin are encoded with the other two symbols. Previous works working with a 3-symbol discretisation of data fix absolute thresholds to define this basin. We argue that there are numerous problems in doing so and, as a major enhancement with respect to such literature, we propose a very flexible approach to the 3-symbol discretisation, which is also rather general. We define the thresholds for the symbolisation to be the two tertiles of the distribution of values taken by the time series. Such a definition has many advantages, since, unlike a fixed-threshold symbolisation scheme, it adapts to the distribution of the time series. This is important because different assets have different distributions of returns and fixed thresholds could introduce discrepancies in treating the different time series. Moreover, the distribution of returns also varies with the sampling frequency, so that a fixed symbolisation scheme for different frequencies appears inappropriate. Finally, our flexible tertile symbolisation can be applied not only to the raw return series, but also to series of processed returns whose values range on different scales. Such are the cases of the returns filtered for the intraday pattern, of the volatility-standardised returns and of the ARMA residuals.

Using the tertile symbolisation, we investigate to what degree the intraday pattern, the volatility and the microstructure contribute to create regularities in the return time series. We follow a whitening procedure, starting from the price returns, removing the intraday pattern, thus getting the deseasonalised returns, then standardising them by the volatility, finally filtering the standardised returns for the ARMA structure and getting the ARMA residuals. We symbolise all these series with the dynamic tertile thresholds and estimate the Shannon entropy of the symbolis series to measure their degree of randomness.

Results show that, in the vast majority of cases, the raw return symbolised series are the most predictable, with the removal of the intraday component making the series more efficient. The most noteworthy results come however from the standardisation step. The removal of the volatility is responsible for the largest increase in the entropy, meaning that volatility gives the return

series a huge amount of regularity. On average, this effect is larger (62%) than the combined effect of the intraday (18%) and microstructure (20%) regularity. This result convincingly demonstrates that, when studying the randomness of a three-symbol discretised time series, the volatility must be filtered out. Omitting this operation would give results that tell more on the predictable character of the volatility than on, for example, market efficiency. The last whitening step, consisting in removing the ARMA structure, further contributes to move return series towards perfect efficiency. Overall, our results show that much of the assets' apparent inefficiency (that is, the predictability of the raw return series) is due to three factors: the daily seasonality, the volatility and the microstructure effects.

In the second part of the thesis we deal with the topic of abnormal high frequency returns of asset prices[5] and systemic instabilities of financial markets. We also develop two different modelling approaches. Modelling the dynamics of large movements in security prices is of paramount importance for risk control, derivative pricing, trading and to understand the behaviour of markets. Our interest for the topic is motivated by the importance of understanding the consequences that the increase of algorithmic and high frequency trading has on market stability, in a context of more and more interconnected markets, where technological progress has deeply changed both the way that trading is performed and the velocity of the spread of information.

The purpose of this research is a deep understanding of the dynamics of market instabilities, from many different viewpoints, with particular interest in systemic events. We devote our attention to the study of their statistical properties, their historical evolution, their relation with macroeconomic news, their autocorrelation and dependence properties, their modelling.

We identify the intraday times of abnormal returns by means of a jump identification procedure based on the standardisation of one minute returns by the intraday pattern and the local volatility. In a first work (Chapter 8), we show that the dynamics of jumps of a portfolio of stocks deviates significantly from a collection of independent Poisson processes. The deviation that we observe is twofold. On one side, by considering individual assets, we find evidence of time clustering of jumps, clearly inconsistent with a Poisson process. This means that the intensity of the point process describing jumps depends on the past history of jumps and that a recent jump increases the probability that another jump occurs. The second deviation from the Pois-

[5]We refer to abnormal returns of asset prices also as *jumps* or *instabilities*.

son model is even more relevant, especially in a systemic context. We find a strong evidence of a high level of synchronisation between the jumping times of a portfolio of stocks. In other words, we find a large number of instances where several stocks jump within the same one minute interval (which is our maximum resolution). This evidence is absolutely incompatible with the hypothesis of independence of the jump processes across assets.

In order to model the time clustering of jumps for individual assets we propose the use of Hawkes processes, a class of self-exciting point processes, which reveal to be very effective in describing the single asset jumping properties. However, the natural extension of the application of Hawkes processes to describe the dynamics of jumps in a multi-asset framework, i.e., considering multivariate Hawkes processes, is highly problematic due to the exponential growth of the parameter number and, more importantly, it is inconsistent with data. In fact, even considering just two stocks, we find that a bivariate Hawkes model is unable to describe the high number of *synchronous* jumps that we empirically observe in the data. This is due to the fact that the kernel structure of Hawkes is more suited to model lagged jumps rather than synchronous jumps. Our main contribution is the introduction of a Hawkes factor model to describe systemic events. We postulate the presence of an unobservable Hawkes process describing a market jump factor. When this factor jump, each asset jumps with a given probability, which is different for each stock. An asset can jump also by following an idiosyncratic Hawkes process. We show how to estimate this model and how to discriminate between systemic and idiosyncratic jumps. Results from simulations of the model show that it is able to reproduce the main features that characterise the departure from a random behaviour of jumps, namely, the time clustering of jumps on individual stocks, the large number of simultaneous systemic jumps and the time lagged cross-excitation between different stocks.

In a second work (Chapter 9), we analyse one-minute data of 140 stocks for each year from 2001 to 2013, addressing other questions. Concerning the evolution of the dynamics of market instabilities in the last years, our research provides the empirical evidence that, while the total number of extreme price movements has decreased along the years, the occurrence of systemic events has significantly increased. This trend is more and more pronounced when considering events of higher and higher level of systemicity. To identify the possible causes of such events we compare their time occurrences with a database of pre-scheduled macroeconomic announcements, which can be expected to have a market-level influence and to possibly explain market-wide events. Unexpectedly, only a minor fraction (less than 40%) of events involving a large fraction of assets has been preceded by the release of a macro news. This evidence leaves room for the hypothesis that financial markets

exhibit systemic instabilities due to some endogenous dynamics, resulting from unstable market conditions such as a temporary lack of liquidity.

Finally, we provide the evidence that highly systemic instabilities have the double effect of (i) increasing the probability that another systemic event takes place in the near future and (ii) increasing the degree of systemicity of short-term instabilities. In order to reproduce such empirical evidences, we propose a model within the class of multivariate Hawkes processes, with mutually exciting properties. We present a multidimensional—yet parsimonious in the number of parameters—Hawkes process which provides a realistic description of the market behaviour around systemic instability events.

Chapter 2 has a mathematical character and introduces the concepts of information source and entropy in the context of dynamical systems and random processes.

Chapter 3 presents and compares some methods to estimate the entropy of dynamical systems, with a combined ergodic theory and information theory approach. It is constituted by the publication (P1).

In Chapter 4 we develop the analytical study of the entropy of the AR(1) and MA(1) processes, motivated by the study on market efficiency of Chapter 7. This is original content and constitutes part of the preprint (P4).

In Chapter 5 we review the application of information theory to gambling and portfolio theory developed by Kelly. The presentation follows [1]. We also review the *Efficient Market Hypothesis* and some literature on the measuring of market efficiency from real data.

In Chapter 6 we present the stylised facts of high frequency financial data and the data cleaning procedures that we use to filter them out in the analyses of Chapters 7, 8 and 9.

In Chapter 7 we develop our study of market efficiency measurement using the Shannon entropy. We use both an analytical approach, using the theoretical results obtained in Chapter 4, and a more empirical one. The content of this chapter is original and constitutes—together with Chapter 4—the preprint (P4).

Chapter 8 contains our first work on market instabilities. It focuses on methods for their detection and on the development of the Hawkes models

for their modelling. This chapter also constitutes the publication (P2).

Chapter 9 is our second work on market instabilities, treating the topics of the historical evolution of instabilities, their exogenous/endogenous origin, the analyses about the systemicity levels and their modelling. It constitutes the preprint (P3).

List of publications

(P1) L. M. Calcagnile, S. Galatolo, G. Menconi, *Non-sequential Recursive Pair Substitutions and Numerical Entropy Estimates in Symbolic Dynamical Systems*, Journal of Nonlinear Science, **20** (2010), pp. 723–745

(P2) G. Bormetti, L. M. Calcagnile, M. Treccani, F. Corsi, S. Marmi, F. Lillo, *Modelling systemic price cojumps with Hawkes factor models*, Quantitative Finance 2015, DOI: 10.1080/14697688.2014.996586

(P3) L. M. Calcagnile, G. Bormetti, M. Treccani, S. Marmi, F. Lillo, *Collective synchronization and high frequency systemic instabilities in financial markets*; preprint at `http://arxiv.org/abs/1505.00704`

(P4) L. M. Calcagnile, F. Corsi, S. Marmi, *Entropy and efficiency of the ETF market*; preprint at `http://arxiv.org/abs/1609.04199`

Contents

Chapter 1

Introduction

Central theme of this thesis is the analysis of high frequency financial data. In recent years, the wide availability of high resolution data has renewed and stimulated the interest of researchers with a mathematical and physical background towards the complex world of financial markets.

The mathematical study of financial markets has a long history and a multitude of approaching methods. One that we find particularly fascinating is the attempt to transport to financial markets some concepts from the theory of dynamical systems. For example, in the late 1980s and especially during the 1990s, some literature focused on the concept of chaos in financial markets, i.e., the apparently random behaviour of markets as the result of a nonlinear dynamics. See [2] for a literature review on the topic of chaos in financial markets and [3, 4] for notable works on the subject. There is also a sizeable amount of literature on the concept of entropy in finance, which has been used as an important tool in portfolio selection [5, 6, 7], asset pricing [8, 9, 10, 11], Efficient Market Hypothesis testing and market efficiency measuring [12, 13, 14]. We investigate how entropy can be used to quantify inefficiency, with the awareness that high frequency data, alongside with the advantage of providing a much higher amount of information than low frequency—daily, at most—data, also present serious challenges.

Another topic that, especially in recent years, has attracted the attention of researchers is that of market instabilities. Recent crashes—the biggest and most famous one being the 6^{th} May 2010 Flash Crash—have posed several questions about their nature, e.g., how frequently do they happen?, are they occurring more and more frequently?, are their causes exogenous or endogenous to the markets? We study detection techniques, correlation properties, historical evolution, exogenous/endogenous origin, modelling possibilities.

In this thesis, we deal with the two aforementioned topics, namely the problem of measuring the informational efficiency of markets and the study

and modelling of market instabilities, in both cases with high frequency financial data as our starting point.

1.1 Measuring the informational efficiency of markets with the Shannon entropy

Financial markets are places where information plays a major role. The agents in the market trade based on all the available information—public and private—and it can also be said that the collective trading activity itself *generates* new information. It is no surprise that, in this context of hidden and manifest information, information theory was applied. Information theory has a precise beginning, identifiable in Claude Shannon's groundbreaking 1948 paper *A Mathematical Theory of Communications*, published on the *Bell System Technical Journal* [15]. Though—as acknowledged by Shannon himself in the beginning of his paper—a basis for a general theory of communication was contained in the work of Nyquist and Hartley at Bell Laboratories in the 1920s, it was Shannon's unifying vision that revolutionised communication, developing the first key concepts and inspiring a multitude of research topics that we now define as the field of information theory.

In his paper Shannon introduced four major concepts: the *channel capacity*, i.e., the concept that every communication channel has a speed limit, measured in binary digits per second; the digital representation of data (the term *bit* is used in his article for the first time), be it text, sound, image or video; the concept of *source coding*, which deals with efficient representation of data and is now synonymous with data compression, based on the removal of redundancy in the information to make the message smaller; the quantification of the amount of information in a signal, called *entropy*. The entropy of a source is essentially an average measure of the randomness of its outputs. It represents the average amount of information that the source transmits with each symbol, or, equivalently, the average amount of uncertainty removed with the transmission of each symbol.

Soon after Shannon developed the entropy theory of information, many attempts were made to give information a more intuitive meaning and practical applications. In his 1956 paper *A New Interpretation of Information Rate* [16], Kelly proved that a gambler with side information could use some of Shannon's equations to achieve the highest possible return on his capital. Presenting his idea as a system for betting on horse races, Kelly showed that the gambler's optimal policy is to maximise the expected logarithm of wealth

and that this goal is achieved with the gambling scheme where bets on each horse are proportional to the probability that each horse will win.[1] Kelly also showed that the financial value of side information (i.e., the increase in wealth that results from that information) is equal to the mutual information between the horse race and the side information. The strong connection between information theory and gambling on horse races was also extended to the financial markets in the broader concept of portfolio selection, highlighting how striking the parallels between the theory of optimal investment in the stock market and information theory are. In fact, gambling on horse races can be viewed as a special case of portfolio selection, since placing bets in a race with m horses corresponds to investing on m special stocks in the market, where the final value of the stock for horse i is either 0 or the value of the odds for horse i. Both in the context of horse races and of a stationary stock market, it has been shown that the growth rate of the wealth is a dual of the entropy rate. The theory of growth optimal portfolios was introduced by Kelly [16] and Latané [17] and generalized by Breiman [18]. Barron and Cover [19] proved that the mutual information between some side information and the stock market is an upper bound on the increase in the growth rate.

A market is said to be *informationally efficient* if it is efficient in processing the information. In an efficient market observed prices express a correct evaluation of all available information. The concept of market efficiency was first anticipated by Bachelier in 1900, who writes in his PhD dissertation that "past, present and even discounted future events are reflected in market price, but often show no apparent relation to price changes." [20]. Modern literature on market informational efficiency begins however in the 1960s, with the independent works of Samuelson and Fama. The idea that asset prices fully reflect all available information (the so-called *Efficient Market Hypothesis* (EMH) was introduced in modern economics by Samuelson, with his article [21]: *Proof that properly anticipated prices fluctuate randomly.* Samuelson observed that "[i]f one could be sure that a price would rise, it would have already risen." It was Fama, though, the first to introduce the term 'efficient market' [22]. In his seminal papers [23, 24, 22, 25] he was concerned with the old debate between *technical analysis* (the use of patterns in historical data to forecast future prices) and *fundamental analysis* (the use of accounting and economic data to establish the assets' fair prices).

Fama proposed three forms of the efficient market hypothesis, corresponding to different information sets available to market agents. The *weak form*

[1] This strategy is called *Kelly criterion*, *Kelly betting*, *Kelly formula*, *proportional betting* or with similar expressions.

says that prices fully reflect the information contained in past prices. The *semi-strong form* maintains that prices fully reflect all publicly available information. The *strong form* asserts that market prices fully reflect any information, public and private.

Market efficiency implies the random character of prices (the more efficient the market, the more random the sequence of price changes). A perfectly efficient market is one in which price changes are completely random and unpredictable. This ideal situation is the result of many market participants attempting to profit from their information. When investors try to exploit even the smallest informational advantages at their disposal, their actions have the effect of incorporating the information into market prices, so that in the end the profit opportunities that first motivated the trades are quickly eliminated.

A statistical description of the unforecastability of price changes is provided by the so-called random walk hypothesis (RWH), suggesting a model in which prices evolve in a purely random manner. Much of the literature about market efficiency revolved around the RWH and testing for the presence of patterns in historical market data. A number of works testing data for the RWH were published from the beginning of the 1960s to the beginning of the 1990s, with general results supporting the RWH for daily data and only occasional results of departure from randomness and rejection of the RWH [26].

Fama [25] summarises the early random walk literature, his own contributions and other studies about the information contained in the historical price time series and concludes that the results strongly support the weak form of the EMH. The evidence accumulated in the 1960s and 1970s shows that, while markets cannot be completely efficient in the strong form, there is convincing support for the weak and semi-strong forms. Departures from the EMH are commonly explained as consequences of over- or underreaction to new information by the investors, when prices are temporarily pushed beyond the fair or rational market value and subsequently adjusted, with the formation of patterns.

As pointed out by Campbell, Lo, MacKinlay [27], perfect informational efficiency is an idealised condition which is never met in real markets. Inefficiencies are always present in real markets and there is no point in investigating the efficiency of markets in absolute terms. Rather than the absolute view taken by the traditional literature on market efficiency, it is much more interesting to study the notion of *relative* efficiency, that is, the degree of efficiency of a market measured against the benchmark of an idealised perfectly efficient market.

We stress that this is the point of view we take in this work. We shall

investigate to what degree assets depart from the idealised perfect efficiency, ranking them according to relative efficiency.

Relative efficiency was studied in [28, 29, 30, 13, 31, 12], where different concepts and tools were used to measure the amount of randomness in the historical return time series, namely, the algorithmic complexity, a variable-order Markov model, the Shannon entropy, the Hurst exponent, the approximate entropy. All these works have in common the procedure of translating a data time series into a symbolic one, by assigning a symbol to a particular behaviour of the price (for example, symbols 0, 1, 2 are given the meaning of 'the price goes down', 'the price goes up', 'the price remains stable' and the return time series is translated into a sequence of 0s, 1s and 2s). This is also what we do in Chapters 4 and 7, investigating symbolisations of stochastic processes and historical time series into 2 or 3 symbols. Like [13], we use the Shannon entropy as a tool for measuring the randomness of the series.

In [28, 29, 30] the authors measure the relative efficiency of the return time series by means of a 3-symbol discretisation, with one symbol representing the price stability behaviour, defined as the absolute value of price returns—or of price differences, in one case—being less than a threshold. A serious shortcoming of their procedure, however, is that they do not normalise returns for the volatility. In our opinion, much of the apparent inefficiency they find in the series is essentially due to the persistency of volatility, which translates into the three symbols forming patterns (in periods of high volatility the two symbols not representing price stability will be very frequent, while the opposite will hold in periods of low volatility).

All the cited papers dealing with relative efficiency, with the exception of [30], deal with daily data. Working instead with high frequency data (i.e., intraday price records) is both very interesting and very challenging. As long as the amount of information is concerned, intraday data obviously provide a much richer source than daily data. Moreover, it is more natural to investigate the presence of recurring patterns and inefficiencies at a high frequency level, since at the daily level they tend to compensate. On the other hand, we have to confront ourselves with new problems. The fact that a security has different buy and sell prices and the fact that prices move on a discrete grid are known as *microstructure* effects, since they depend on the structure of the market and on the price formation mechanism. They are sources of regularity in recorded price series—they add autocorrelation—and can play a major role especially for low liquidity assets. These effects add spurious regularity to the series, which must be properly discounted if one wants to quantify the true efficiency.

The scientific literature has proposed a variety of approaches to the modelling of microstructure. An interesting one is the modelling by means of

the autoregressive–moving-average (ARMA) models. These models are able to capture the return series autocorrelation due to the microstructure features. The benchmark provided by the ARMA models can be used in two different ways. One way is estimating the best model on the data and taking the time series of *residuals*[2], then analysing the residual time series to quantify the remaining regularities. Another way to use the benchmark of the autoregressive moving-average models is to compare the randomness of the data time series with the theoretically calculated value of the randomness of the ARMA models. We follow both approaches, developing an interesting theoretical study about the Shannon entropy of ARMA models.

Our method to measure the assets' efficiency is designed to discount all known sources of regularity of high frequency data: the intraday pattern, the long-term volatility and the microstructure effect. Only the inefficiency that remains in the processed series, after these effects have been filtered out, can be regarded as the true inefficiency.

1.2 Studying and modelling the properties of market instabilities

Recent years have seen an unprecedented rise of the role that technology plays in all aspects of human activities. Unavoidably, technology has heavily entered the financial markets, to the extent that all major exchanges are now trading exclusively using electronic platforms. This has profoundly changed the way the trading activity is conducted in the market, from the old phone conversation or click and trade on a screen to software programming. In algorithmic trading, computer programs process a large amount of information—from market data to political news and economic announcements—and output trading instructions with no human intervention being required. Trading algorithms include the search for arbitrage opportunities (for example, small discrepancies in the exchange rates between three currencies), the search for optimal splitting of large orders at the minimum cost, the implementation of long-term strategies to maximise profits. An important class of algorithms is devoted to the automatic parsing and interpretation of news releases, resulting in trading orders being generated before humans have even read the news title. Electronic platforms and new technologies have facilitated the flourishing of the so-called *high frequency trading*, which is nowadays a significant fraction of all the trading activity in US and Europe electronic exchanges

[2]Residuals can be thought of as the term-by-term difference between the data and the fitted model.

[32].

It has been argued that high frequency trading is beneficial to the discovery of the efficient price [33]. Arbitrage opportunities that may originate under some circumstances are promptly exploited—and thus eliminated—by algorithms that act as high frequency traders, resulting in a more efficient price. On the other side, the ultra fast speed of information processing, order placement and cancelling potentially allows large price movements to propagate very rapidly through different assets and exchanges [34], and is believed to be a means of contagion and the cause of an increased synchronisation among different assets and different markets.

On 6$^{\text{th}}$ May 2010, this synchronisation effect had its most spectacular appearance in what is known as the *Flash Crash*. The crash started from a rapid decline in the E-Mini S&P 500 market and in a very short time the anomaly became systemic and the shock propagated towards ETFs, stock indices and their components, and derivatives [35, 36]. The price of the Dow Jones Industrial Average plunged by 9% in less than 5 minutes but recovered the pre-shock level in the next 15 minutes of trading. The SEC reported that such a swing was sparked by an algorithm executing a sell order placed by a large mutual fund. Then high frequency traders, even though did not ignited the event, caused a "hot potato" effect amplifying the crash. In the aftermath of the crash, several studies have focused on events, evocatively named *mini flash crashes*, concerned with the emergence of large price movements of an asset in a very limited fraction of time and attributing their origin to the interaction between several automatic algorithms [37] or to the unexpected product of regulation framework and market fragmentation [38].

In recent years, it has been observed a growing tendency of the financial markets to exhibit systemic instabilities, i.e., sudden large price movements involving a great part of the market. Researchers have tried to answer the following questions. Are these extreme price fluctuations driven by the release of price sensitive information? Or, on the contrary, are these events an effect of any unrevealed endogenous dynamics? Many such events cannot be associated with exogenous events such as news releases and it has been suggested in literature that this susceptibility may be generated in an endogenous way by the trading process itself. The Flash Crash has dramatically shown how strongly interconnected different markets and asset classes can become, especially during extreme events.

Scientific literature about discontinuities in asset prices has concentrated mostly on the identification problem at the daily time scale, i.e., on trying to answer the question of whether an extreme return occurred on a given day. This topic, referred to as the jump identification problem in the econometric literature, was extensively studied in theoretical works as well as in empirical

analysis and applications to asset pricing. Significantly less research—and mainly of empirical nature—has been devoted to the study of *cojumps*, i.e., large price movements occurring synchronously in two or more assets. In literature, the modelling of discontinuities in the dynamics of asset prices is commonly performed by means of a compound Poisson process, see e.g. [39]. While having the great advantage of being analytically tractable, Poisson processes (as well as the more general Lévy processes) have independent increments and thus cannot account for any kind of serial or cross-sectional correlation. It is instead reasonable to presume that the jump component of the price process has some clustering properties and that a better choice for its modelling would be one that allows some dependence between the jumping times. Moreover, considering the increased interconnection of electronic markets, it is reasonable to expect some synchronisation between the jumping times of different assets. This is particularly important when aiming to describe and model events of systemic instabilities such as the Flash Crash, for which using independent jump processes would clearly be inadequate.

A class of self-exciting point processes that are good candidates to well describe the correlation properties of the jump time series is that of Hawkes processes. These processes were introduced in seismology more than forty years ago [40] and have been widely employed to model earthquake data [41, 42]. In recent years, Hawkes processes have become popular also in mathematical finance and econometrics [43, 44, 45]. They have been effectively used to address many different problems in high-frequency finance, such as estimating the volatility from transaction data, devising optimal execution strategies or modelling the dynamics of the order book.

A univariate Hawkes process is a point process whose intensity is the sum of two components: the baseline intensity, responsible for the happening of events unrelated to the history of the process up to the present time, and a backward-looking kernel, which adds a further intensity for every past event (more recent events give a higher contribution) and is responsible for the generation of *child* events, i.e., events triggered by previous ones. A univariate Hawkes process is thus a natural choice for trying to model a clustered time series like the one of an asset price jumps. While the baseline intensity could explain the exogenously generated jumps, the kernel could account for the endogenous mechanisms of self-excitement and auto-triggering.

In a global perspective, considering market instability properties means looking also at cross-asset dependencies and mutual excitement properties. Since in a highly connected market large price movements in one asset can propagate and thus be responsible of extreme returns in other assets, a model describing global instabilities must necessarily possess inter-asset correlations or some kind of multi-asset features.

It is natural to consider multivariate Hawkes processes, multidimensional point processes which possess, along with the self-exciting properties of standard univariate Hawkes processes, also cross-exciting properties for every possible pair of processes. However, as we show in Chapter 8, they present two notable downsides. Firstly, the number of parameters of a multivariate Hawkes process grows quadratically with the dimension, i.e., an N-dimensional Hawkes process has $O(N^2)$ parameters, and thus its estimation becomes unfeasible for high dimensions. Secondly, but more importantly, we find that even a 2-dimensional Hawkes model is not able to reproduce the high number of *synchronous* jumps observed in empirical data. These considerations led us to search for different approaches.

In a first work, Chapter 8, we investigate the one-asset and multiple-asset jump clustering properties. After studying the performance of our jump identification method and analysing some statistical properties of jumps of individual stocks and of systemic jumps, we first compare the empirical data with a Poisson benchmark and then develop the Hawkes modellisation of the jump dynamics. Since, as said before, bivariate Hawkes processes fail at describing the cojump properties of asset pairs, we develop a Hawkes factor model, which postulates the presence of an unobservable point process describing a market factor. When this factor jumps, each asset jumps with a given probability. Each asset can jump also according to an idiosyncratic point process. Both the factor process and the idiosyncratic processes are modelled as Hawkes processes, in order to capture the time clustering of jumps.

In a second work, Chapter 9, we propose a further deep study on systemic instabilities, from different perspectives. First of all, we investigate the historical evolution of systemic jumps in the years from 2001 to 2013, answering the questions "How have systemic instability properties changed in the last years?" and "Have markets become systemically more unstable?". Technology and market structure have evolved and continue evolving, thus it is reasonable to hypothesise that also market systemic instabilities have changed in some of their properties.

We then study to what extent systemic instabilities are due to macroeconomic news. Announcements about new interest rates, employment data, industrial production and other economical/financial relevant news usually cause an adjustment of markets to the new available information. This adjustment could be violent if the released information differs significantly from the experts' opinions or if it is unanticipated to some degree. However, not all systemic instabilities can be associated with the release of macroeconomic news and some of them can also have an endogenous origin, like the Flash Crash.

Finally, we study the relation of systemic instabilities with other systemic instabilities. We do this in two ways. The first one consists in looking at how the occurrence of a systemic event increases the *probability* that another systemic event happens in a few minutes, based on the level of systemicity[3] of the conditioning event. The second way consists in studying how systemic events with a higher level of systemicity anticipate new events with high level of systemicity.

In order to model the dipendencies among cojump multiplicities, we design a multivariate Hawkes process where each component is meant to account for the point process of a different cojump multiplicity. In other words, we avoid modelling the single-asset jump processes and then looking at their aggregated properties, instead we directly model the point processes of multiplicity 1, multiplicity 2, ..., multiplicity N, where N is the maximum multiplicity, i.e., the total number of assets. The idea supporting this approach is what we observe on data, namely, the fact that a cojump event of a certain multiplicity is generally followed by other events with a similar multiplicity (multiplicity is highly autocorrelated). This accounts for both effects of (i) highly systemic events triggering *more* new events and (ii) higher-multiplicity events triggering other higher-multiplicity events. The parametrisation of the multivariate Hawkes kernel is chosen in such a way that the exciting effect of multiplicity i on multiplicity j is highest for $j = i$ and decreases as $|i - j|$ increases, i.e., the closest the two multiplicities, the highest the mutual exciting effect, and the more distant the two multiplicities, the lowest the mutual exciting effect.

[3]We call *level of systemicity* of a cojump the fraction of the market exhibiting a price jump. Similarly, we talk of the *multiplicity* of a cojump event as the number of assets involved in the collective jump. The two expressions differ in that the first one expresses a relative magnitude, while the second one is an absolute number. However, unless their relative/absolute character is crucial for what we say, we use them interchangeably.

Part I

Entropic measures of market efficiency

Chapter 2

Information sources and entropy

2.1 Introduction

An *information source* or *source* is a mathematical model for a physical entity that produces a succession of symbols called "outputs" in a random manner. The space containing all the possible output symbols is called the *alphabet* of the source and a source is essentially an assignment of a probability measure to events consisting of sets of sequences of symbols from the alphabet. A natural definition for a source is then as a *discrete-time finite-alphabet stochastic process*. However, it is often useful to explicitly treat the notion of time as a transformation of sequences produced by the source. Thus in addiction to the common random process model we shall also consider modelling sources by dynamical systems as considered in ergodic theory.

A source is *ergodic* if, intuitively, its statistical features can be deduced from a single typical realisation. In other words, if the source is ergodic, observing many samples of a process is statistically equivalent to observing a single sample, provided that it is *typical*, that is, it reflects the process' properties. Ergodic theory is the mathematical study of the long-term average behaviour of systems. Historically, the idea of ergodicity came from Boltzmann's *ergodic hypothesis* in statistical mechanics, which states that the average of a physical observable over all possible initial conditions (phase space mean) should equal the long-term time average along a single orbit (time mean). For an observable f on a system (X, μ) with evolution law T, the ergodic hypothesis can be expressed by

$$\int_X f \, \mathrm{d}\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) \quad \text{a. e.}$$

The systems for which the hypothesis holds are called *ergodic*.

The *entropy* of a source is essentially an average measure of the randomness of its outputs. It represents the average amount of information that the source transmits with each symbol, or, equivalently, the average amount of uncertainty removed with the transmission of each symbol.

2.2 Information sources

We can describe an information source using two mathematical models: A random process and a dynamical system. The first is just a sequence of random variables, the second a probability space together with a transformation on the space. The two models are connected by considering a time shift to be a transformation.

2.2.1 Random processes and sequence spaces

A discrete-time stochastic process is a sequence of random variables $X_1, X_2, X_3, \ldots$ defined on a probability space (X, Σ, μ). In many cases it is useful to consider bi-infinite sequences $\ldots, X_{-1}, X_0, X_1, \ldots$, and what we present for processes with indices in $\mathbb{N}$ holds, with the straightforward generalisations, for processes with indices in $\mathbb{Z}$. For more details see Shields [46].

The process has *alphabet* A if every X_i takes values in A. We shall only be interested in finite-alphabet processes, so, from now on, "process" will mean a discrete-time finite-alphabet process.

The finite sequence $a_m, a_{m+1}, \ldots, a_n$, where $a_i \in A$ for $m \leq i \leq n$, will be denoted with a_m^n. The set of all strings a_m^n will be denoted with A_m^n, except for $m = 1$, when A^n will be used.

The *k-th order joint distribution* of the process $\{X_k\}$ is the measure μ_k on A^k defined by

$$\mu_k(a_1^k) = \text{Prob}(X_1^k = a_1^k), \quad \text{for every } a_1^k \in A^k.$$

The *distribution* of the process is the set of joint distributions $\{\mu_k : k \geq 1\}$. The distribution of a process can also be defined by specifying the start distribution μ_1 and, for each $k \geq 2$, the conditional distributions

$$\mu(a_k|a_1^{k-1}) = \text{Prob}(X_k = a_k|X_1^{k-1} = a_1^{k-1}) = \frac{\mu_k(a_1^k)}{\mu_{k-1}(a_1^{k-1})}.$$

The distribution of a process is thus a family of probability distributions, one for each k. The sequence, however, cannot be completely arbitrary, for

implicit in the definition of process is that the following *consistency condition* must hold for every $k \geq 1$,

$$\mu_k(a_1^k) = \sum_{a_{k+1} \in A} \mu_{k+1}(a_1^{k+1}), \quad \text{for every } a_1^k \in A^k. \tag{2.1}$$

A process is defined by its joint distributions and this means that the particular space on which the measurable functions X_n are defined is not important. Given a process, there is a canonical way to define a probability space and a sequence of random variables on it which realise the process. This is the so-called *Kolmogorov representation* and it is based on the symbolic sequences spaces. We briefly describe it here, referring the reader to Shields' book [46] for a more detailed and formal presentation.

Let A be a finite set. We denote by $A^{\mathbb{N}}$ the set of infinite sequences

$$x = (x_i), \quad x_i \in A, \ i \geq 1.$$

The *cylinder* determined by a_m^n, with $a_i \in A$ for $m \leq i \leq n$, denoted by $[a_m^n]$, is the subset of $A^{\mathbb{N}}$ defined by

$$[a_m^n] = \{x \in A^{\mathbb{N}} : x_i = a_i, \, m \leq i \leq n\}.$$

Let Σ be the σ-algebra generated by the cylinders. The elements of Σ are called the Borel sets of $A^{\mathbb{N}}$. For every $n \geq 1$ the coordinate functions $\hat{X}_n : A^{\mathbb{N}} \to A$ is defined by $\hat{X}_n(x) = x_n$. The next theorem states that every process with alphabet A can be thought of as the coordinate function process $\{\hat{X}_n\}$ on the space $A^{\mathbb{N}}$ endowed with the Borel *sigma*-algebra and a probability measure.

Theorem 2.2.1 (Kolmogorov representation theorem)**.** *If $\{\mu_k\}$ is a sequence of measures for which the consistency conditions (2.1) hold, then there is a unique Borel probability measure μ on $A^{\mathbb{N}}$ such that $\mu([a_1^k]) = \mu_k(a_1^k)$ for each k and each a_1^k.*

In other words, if $\{X_n\}$ is a process with finite alphabet A, there is a unique Borel measure μ on $A^{\mathbb{N}}$ for which the sequence of coordinate functions $\{\hat{X}_n\}$ has the same distribution as $\{X_n\}$.

For the proof, see Shields [46], p. 2.

A process is *stationary* if the joint distributions do not depend on the choice of time origin, that is, if

$$\text{Prob}(X_i = a_i, \ m \leq i \leq n) = \text{Prob}(X_{i+1} = a_i, \ m \leq i \leq n), \tag{2.2}$$

for all m, n and a_m^n. The intuitive idea of stationarity is that the mechanism which generates the random variables does not change with time.

When the space of bi-infinite sequences $A^{\mathbb{Z}}$ is used, the definition of cylinder extends straightforwardly to this setting and an analogue of Theorem 2.2.1 holds for $A^{\mathbb{Z}}$. That is, if $\{\mu_k\}$ is a family of measures for which the conditions of consistency (2.1) and stationarity (2.2) hold, then there is a unique Borel measure μ on $A^{\mathbb{Z}}$ such that $\mu([a_1^k]) = \mu_k(a_1^k)$, for each $k \geq 1$ and each a_1^k. The projection of μ on $A^{\mathbb{N}}$ is obviously the Kolmogorov measure on $A^{\mathbb{N}}$ for the process defined by $\{\mu_k\}$.

2.2.2 Shift transformations

The shift is the transformation that rules the dynamics of sequence spaces.

The (left) shift is the map $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$ defined by

$$(\sigma x)_n = x_{n+1}, \quad x \in A^{\mathbb{N}}, n \geq 1,$$

or, in symbolic form,

$$x = (x_1, x_2, x_3, \ldots) \Rightarrow \sigma x = (x_2, x_3, x_4, \ldots).$$

Note that

$$\sigma^{-1}[a_m^n] = [b_{m+1}^{n+1}], \quad b_{i+1} = a_i, m \leq i \leq n,$$

so that the set map σ^{-1} takes a cylinder set onto a cylinder set. It follows that $\sigma^{-1}B$ is a Borel set for each Borel set B, which means that σ is a Borel measurable mapping.

The stationarity of the process translates into the condition $\mu([a_m^n]) = \mu([b_{m+1}^{n+1}])$ for all m, n and a_m^n, that is, $\mu(B) = \mu(\sigma^{-1}B)$ for every cylinder set B, which is equivalent to the condition that $\mu(B) = \mu(\sigma^{-1}B)$ for every Borel set B. This latter condition is usually summarised by saying that σ *preserves the measure* μ or, alternatively, that μ is σ-invariant.

For stationary processes there are two canonical representations. One as the Kolmogorov measure on $A^{\mathbb{N}}$, the other as the σ-invariant extension to the space $A^{\mathbb{Z}}$. The latter model is called the *bi-infinite Kolmogorov model* of a stationary process. Note that the shift σ in the bi-infinite model is invertible.

2.2.3 Dynamical systems

Let $\mathcal{X} = (X, \mathcal{B}, \mu)$ be a probability space and T a measure-preserving transformation, that is, a measurable $T : \mathcal{X} \to \mathcal{X}$ such that

$$\forall B \in \mathcal{B} \quad T^{-1}(B) \in \mathcal{B} \text{ e } \mu(T^{-1}(B)) = \mu(B).$$

Equivalently, the condition that T preserves the measure μ is expressed by saying that μ is T-invariant.

Since the composition of measure-preserving transformations is still measure-preserving, we can consider the iterates $T^n : \mathcal{X} \to \mathcal{X}$.

Using the given definition, it is difficult to verify whether a given transformation is measure-preserving, for in general we have not explicit knowledge of all the elements of $\mathcal{B}$. It is sufficient, though, to explicitly know a semi-algebra that generates $\mathcal{B}$. A *semi-algebra* is a collection $\mathcal{S}$ of subsets of X for which the following conditions hold:

(i) $\varnothing \in \mathcal{S}$;

(ii) if $A, B \in \mathcal{S}$ then $A \cap B \in \mathcal{S}$;

(iii) if $A \in \mathcal{S}$ then $X \setminus A = \cup_{i=1}^{n} E_i$ with $E_i \in \mathcal{S}$ for each i and the E_i's pairwise disjoint.

For example, if the space is the unit interval $[0, 1]$, the collection of subintervals is a semi-algebra. This follows from the following theorem (Walters [47], p. 20).

Theorem 2.2.2. *Let $\mathcal{X}_1 = (X_1, \mathcal{B}_1, \mu_1)$ and $\mathcal{X}_2 = (X_2, \mathcal{B}_2, \mu_2)$ two probability spaces and $T : \mathcal{X}_1 \to \mathcal{X}_2$ a transformation. Let $\mathcal{S}_2$ be a semi-algebra that generates $\mathcal{B}_2$. If for each $A_2 \in \mathcal{S}_2$ it holds $T^{-1}(A_2) \in \mathcal{B}_1$ and $\mu_1(T^{-1}(A_2)) = \mu_2(A_2)$, then T is measure-preserving.*

With some notation abuse, in the following we shall write $T : X \to X$ (and similar) instead of the more correct $T : \mathcal{X} \to \mathcal{X}$.

Isomorphism of measure-preserving transformations

In the study of measure-preserving transformations, it is very useful to introduce a notion of isomorphism between them. Thanks to this notion, in many cases it is possible to translate the study of a measure-preserving transformation of a given space to the shift transformation of a sequence space (see Section 2.2.4).

Two isomorphic transformations both possess or do not possess some properties (for instance, *ergodicity*, see Section 2.2.5) and for them we shall find some invariant quantities (for instance, *entropy*, see Section 2.3.3).

Definition 2.2.1. *Let $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ be two probability spaces with respective measure-preserving transformations*

$$T_1 : X_1 \to X_1, \quad T_2 : X_2 \to X_2.$$

We say that T_1 *and* T_2 *are* isomorphic *(or, more precisely, that the dynamical systems* $(X_1, \mathcal{B}_1, \mu_1, T_1)$ *and* $(X_2, \mathcal{B}_2, \mu_2, T_2)$ *are* isomorphic*), and we shall write* $T_1 \sim T_2$*, if there exist two sets* $M_1 \in \mathcal{B}_1$*,* $M_2 \in \mathcal{B}_2$ *with* $\mu_1(M_1) = 1$*,* $\mu_2(M_2) = 1$ *such that*

(i) $T_1 M_1 \subseteq M_1$*,* $T_2 M_2 \subseteq M_2$*,*

(ii) there exists an invertible measure-preserving transformation

$$\phi : M_1 \to M_2 \quad \textit{with } \phi T_1(x) = T_2 \phi(x) \ \forall x \in M_1,$$

that is, such that the following diagram is commutative.

$$\begin{array}{ccc} (M_1, \mu_1) & \xrightarrow{T_1} & (M_1, \mu_1) \\ \phi \big\downarrow & & \phi \big\downarrow \\ (M_2, \mu_2) & \xrightarrow{T_2} & (M_2, \mu_2) \end{array}$$

(In the condition (ii), the set M_i*,* $i = 1, 2$*, is meant to be endowed with the* σ*-algebra* $M_i \cap \mathcal{B}_i = \{M_i \cap B \mid B \in \mathcal{B}_i\}$ *and the restriction of the measure* μ_i *to such* σ*-algebra.)*

We say that T_2 *is a* factor *of* T_1 *if there exist* M_1 *and* M_2 *as before such that (i) and (ii') hold, where*

(ii') there exists a (not necessarily invertible) measure-preserving transformation

$$\phi : M_1 \to M_2 \quad \textit{con } \phi T_1(x) = T_2 \phi(x) \ \forall x \in M_1.$$

In such a case, T_1 *is said to be an* extension *of* T_2*.*

Observe that the relation of isomorphism between measure-preserving transformations is an equivalence relation. Moreover, if $T_1 \sim T_2$, then $T_1^n \sim T_2^n$ for all $n > 0$.

Example. Let T be the transformation $Tz = z^2$ on the unit circle S^1 with the Borel σ-algebra and the Haar measure, and let U be defined by $Ux = 2x \pmod 1$ on $[0, 1)$ with the Borel σ-algebra and the Lebesgue measure. Let us consider the map $\phi : [0, 1) \to S^1$ defined by $\phi(x) = e^{2\pi i x}$. The map ϕ is bijective and measure-preserving (it is sufficient to verify it on the intervals and use Theorem 2.2.2). Furthermore it holds $\phi U = T\phi$ and so $U \sim T$.

2.2.4 From dynamical systems to symbolic dynamics

We describe in these section a procedure to relate the study of a dynamical system on a generic space to the study of a sequence space with the shift transformation.

Let T be an invertible measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$ and let $\alpha = \{A_1, \ldots, A_k\}$ be a finite measurable partition of X. We denote by Ω the product space $\{1, 2, \ldots, k\}^{\mathbb{Z}}$, so that a point of Ω is a bi-infinite sequence $\omega = (\omega_n)_{-\infty}^{\infty}$, where $\omega_n \in \{1, 2, \ldots, k\}$ for all n. On Ω the shift transformation σ is defined as in Section 2.2.2.

We now show that there exist a σ-invariant probability measure ν on the Borel σ-algebra $\mathcal{B}(\Omega)$ of Ω and a measure-preserving map $\phi_\alpha : (X, \mathcal{B}, \mu) \to (\Omega, \mathcal{B}(\Omega), \nu)$ such that $\phi_\alpha T = \sigma \phi_\alpha$. For $x \in X$ define

$$\phi_\alpha x = (\omega_n)_{-\infty}^{\infty} \quad \text{if } T^n x \in A_{\omega_n}.$$

The n-th coordinate of $\phi_\alpha x$ is the index of the element of α which contains $T^n x$.

It is thus defined an application $\phi_\alpha : X \to \Omega$ and it obviously holds $\sigma(\phi_\alpha x) = \phi_\alpha(Tx)$, for all $x \in X$. To see that ϕ is measurable, note that for every cylinder set $[\omega_m^n]$, $m < n$, it holds

$$\phi_\alpha^{-1}([\omega_m^n]) = \bigcap_{j=m}^{n} T^{-j} A_{\omega_j}.$$

Then for each cylinder set C it holds $\phi_\alpha^{-1}(C) \in \mathcal{B}$, and since cylinders generate $\mathcal{B}(\Omega)$ it holds $\phi_\alpha^{-1}(\mathcal{B}(\Omega)) \subseteq \mathcal{B}$.

The map ϕ_α transports in a natural way the measure μ on $(\Omega, \mathcal{B}(\Omega))$:

$$\nu(E) \stackrel{\text{def}}{=} \mu(\phi_\alpha^{-1} E) \quad \text{for each measurable set } E \subseteq \Omega.$$

Clearly $\phi_\alpha : X \to \Omega$ is measure-preserving and from the map equality $\sigma \phi_\alpha = \phi_\alpha T$ it follows that σ preserves ν.

When the map T is not invertible, an analogue construction is possible taking into consideration the space of symbolic sequences $\Omega^+ = \{1, 2, \ldots, k\}^{\mathbb{N}}$.

The explained construction is very important and we base on it the computer simulations we present in Section 3.6. Let us now establish some terminology.

Definition 2.2.2. *The sequence $\phi_\alpha(x)$ is called the* (α, T)-name of x. *A point $x \in X$ is called* α-typical for T *(or, simply,* typical*) if every sequence $\omega_1 \omega_2 \ldots \omega_n$ appears in the (α, T)-name of x with frequency $\mu(A_{\omega_1} \cap T^{-1} A_{\omega_2} \cap \ldots \cap T^{-n+1} A_{\omega_n})$.*

Note that, by Birkhoff's Ergodic Theorem 2.2.5, almost every point of X is α-typical for T.

The map $\phi_\alpha : X \to \Omega$ is not in general invertible, so it does not always give an isomorphism between X and Ω. In general, $(\Omega, \mathcal{B}(\Omega), \nu, \sigma)$ is only a factor or $(X, \mathcal{B}, \mu, T)$ (see Definition 2.2.1). Still, if the partition α is generating (see Definition 2.2.3 below), then it is indeed an isomorphism, as stated by Proposition 2.2.3, whose proof the reader can find in Petersen [48], p. 274.

Definition 2.2.3. *Let T be a transformation of the probability space $(X, \mathcal{B}, \mu)$ that preserves the measure μ. A finite partition $\alpha = \{A_1, \ldots, A_n\}$ of X is said to be* generating *if the sets of the form*

$$A_{i_1} \cap T^{-1}A_{i_2} \cap \ldots \cap T^{-(k-1)}A_{i_k}$$

generate the σ-algebra $\mathcal{B}$.

Proposition 2.2.3. *The map ϕ_α is an isomorphism if and only if α is a generating partition for T.*

2.2.5 Ergodicity

Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving transformation. If $T^{-1}B = B$ for some $B \in \mathcal{B}$ (in this case B is said to be *completely invariant*), then also $T^{-1}(X \setminus B) = X \setminus B$, and we could study T considering the two transformations $T|_B$ e $T|_{X\setminus B}$. If $0 < \mu(B) < 1$ this simplifies the study of T. If, on the contrary, $\mu(B) = 0$ or $\mu(X \setminus B) = 0$, then considering the restrictions of T to B and to its complement does not imply a significant simplification. From these considerations it emerges the importance of studying those transformations that cannot be decomposed as above. Non-decomposable transformations are called *ergodic*.

Definition 2.2.4. *Let $(X, \mathcal{B}, \mu)$ be a probability space. A measure-preserving transformation T on $(X, \mathcal{B}, \mu)$ is said to be* ergodic *if the only measurable sets that are completely invariant have either zero or full measure, that is, if*

$$T^{-1}B = B,\ B \in \mathcal{B} \quad \Rightarrow \quad \mu(B) = 0 \text{ or } \mu(B) = 1.$$

Ergodicity may be defined in many different though equivalent ways, some of which we list in the next theorem. For its proof, see for example Walters [47] or any textbook of ergodic theory.

Theorem 2.2.4. *If $T : X \to X$ is a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$, then the following properties are equivalent.*

(i) T *is ergodic.*

(ii) For every $A \in \mathcal{B}$ *with* $\mu(A) > 0$ *it holds* $\mu(\cup_{n=1}^{\infty} T^{-n}A) = 1$.

(iii) For every A, $B \in \mathcal{B}$ *with* $\mu(A) > 0$, $\mu(B) > 0$ *there exist* $n > 0$ *such that* $\mu(T^{-n}A \cap B) > 0$.

(iv) If $f \in L^1(X, \mathcal{B}, \mu)$ *is* T*-invariant (that is,* $f \circ T = f$ *a. e.) then* f *is constant almost everywhere.*

(v) For every A, $B \in \mathcal{B}$ *it holds* $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \mu(A)\mu(B)$.

One of the most important results in ergodic theory was proved by Birkhoff in 1931 [49]. We state it in the setting of probability spaces, but it holds more generally for σ-finite measure spaces.

Theorem 2.2.5 (Ergodic Theorem, Birkhoff 1931)**.** *Let* $(X, \mathcal{B}, \mu)$ *be a probability space and* $T : X \to X$ *a measure-preserving transformation. Given* $f \in L^1(X, \mathcal{B}, \mu)$*, there exist* $f^* \in L^1(X, \mathcal{B}, \mu)$ *such that*

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \quad a.\ e.$$

Moreover f^* *is* T*-invariant (that is,* $f^* \circ T = f^*$ *a. e.) and* $\int_X f^* \mathrm{d}\mu = \int_X f \mathrm{d}\mu$.

We remark that, if T is ergodic, from statement (iv) of Theorem 2.2.4 and from Birkhoff's Ergodic Theorem it follows that $f^*(x) = \int_X f^* \mathrm{d}\mu$ a. e. From this remark the following corollary follows.

Corollary 2.2.6. *Let* $(X, \mathcal{B}, \mu)$ *be a probability measure and* $T : X \to X$ *a measure-preserving transformation. Furthermore, let* $E \in \mathcal{B}$*. If* T *is ergodic then*

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_E\big(T^i(x)\big) \to \mu(E) \quad a.\ e.$$

In other words, if T *is ergodic then the orbit of almost every point of* X *enters the set* E *with asymptotic frequency equal to* $\mu(E)$.

The following theorem asserts that ergodicity is invariant by isomorphism (for the proof, see for example Walters [47]).

Theorem 2.2.7. *Let* $(X_1, \mathcal{B}_1, \mu_1)$ *and* $(X_2, \mathcal{B}_2, \mu_2)$ *two probability spaces with measure-preserving transformations* $T_i : X_i \to X_i$*,* $(i = 1, 2)$*. If* T_1 *is isomorphic to* T_2 *and* T_2 *is ergodic, then also* T_1 *is ergodic.*

2.3 Entropy

2.3.1 Introduction

The concept of entropy was first introduced in thermodynamics by Clausius in 1854. Shannon introduced it in information theory in 1948 and Kolmogorov in ergodic theory in 1958. In all these cases entropy is a measure of randomness and disorder.

Claude Shannon based the mathematical study of information transmission on the two fundamental ideas: (1) the amount of information transmitted is measured by the amount of uncertainty removed and (2) the uncertainty of a transmission is proportional to the expected value of the logarithm of the probability of the received symbol.

To be more clear, let us consider a source emitting sequences $\ldots x_0x_1x_2\ldots$, where the a_i's are elements of a finite alphabet $A = \{a_1, a_2, \ldots, a_k\}$. Suppose to receive symbol a_i with probability p_i, for each $i = 1, \ldots, k$. If each symbol is transmitted independently of what has been previously transmitted, the average amount of information transmitted per symbol can be measured by the quantity

$$H = -\sum_{i=1}^{k} p_i \log_2 p_i$$

(with the convention $0 \log_2 0 = 0$). H is the expected information content per symbol, when the information content of an event E is measured by $-\log_2 P(E)$. That this is a reasonable choice comes from the fact that the only measures of information of the form $f(P(E))$, with $f : \mathbb{R} \to [0, \infty)$ continuous and additive on independent events ($f(P(E_1 \cap E_2)) = f(P(E_1)P(E_2)) = f(P(E_1)) + f(P(E_2))$), are $-\log_r P(E)$ for any $r > 0$. For another justification of the definition of H see Khinchin [50]. It is common to use the logarithm to base 2 to measure information in bits. Any other base is of course allowed, as long as one changes the unit of measurement by a constant factor. If some $p_{i_0} = 1$ (and all others $p_i = 0$), then there is no uncertainty on what the source will transmit and in this case $H = 0$. On the other hand, when all the symbols are equally likely to be transmitted the uncertainty on the next symbol is maximum and in fact H attains its maximum value when all the p_i's are equal to $1/k$.

For a generic source, the probability of receiving a given symbol may depend on what has been previously received. For example, if the source's output is English text, the probability of receiving a 'u' rises enormously immediately after a 'q' has been transmitted. These dependencies can be taken into account when measuring information considering blocks of symbols.

For each $k = 1, 2, \ldots$ let $\mathcal{C}_k$ be the set of blocks of length k made up by symbols of the alphabet A. If a block C has a probability $P(C)$ of being received, the average information per symbol in a transmission of length k is

$$H_k = -\frac{1}{k} \sum_{C \in \mathcal{C}_k} P(C) \log_2 P(C). \tag{2.3}$$

The *entropy of the source* is defined as the limit

$$h = \lim_{k \to \infty} -\frac{1}{k} \sum_{C \in \mathcal{C}_k} P(C) \log_2 P(C) \tag{2.4}$$

and it represents the average amount of information the source transmits with each symbol. Equivalently, h measures the average amount of uncertainty that a reader of the source has about the next symbol.

2.3.2 Entropy of a random process

Definition 2.3.1 (Entropy of a random variable)**.** *Let X be a finite-valued random variable with distribution defined by $p(x) = Prob(X = x)$, $x \in A$. The* entropy *of X is the expected value of the variable $-\log p(X)$, that is,*

$$H(X) = -\sum_{x \in A} p(x) \log p(x).$$

The logarithm base is 2 and the convention $0 \log 0 = 0$ is used. If p is the distribution of X, then $H(p)$ may be used in place of $H(X)$. For a random vector $(X_1, \ldots, X_n)$ with distribution $p(x_1, \ldots, x_n) = \text{Prob}(X_1 = x_1, \ldots, X_n = x_n)$, we will use the notation

$$H(X_1, \ldots, X_n) = -\sum_{x_1, \ldots, x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n).$$

Proposition 2.3.1. *(a)* **Positivity.** *$H(X) \geq 0$, with equality if and only if X is constant.*

(b) **Boundedness.** *If X has k values then $H(X) \leq \log k$, with equality if and only if each $p(x) = 1/k$.*

(c) **Subadditivity.** *$H(X, Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent.*

Positivity is obvious since $0 \leq p(x) \leq 1$, for all x.

To prove boundedness, consider the function f defined on $[0,1]$ by

$$f(t) = \begin{cases} 0 & \text{if } t = 0 \\ -t \log t & \text{if } 0 < t \leq 1. \end{cases}$$

The function f is continuous, non-negative and convex. From Jensen's inequality it follows that for all positive $\lambda_1, \lambda_2, \ldots, \lambda_k \in [0,1]$,

$$f\left(\frac{1}{k}\sum_{i=1}^{k} \lambda_i\right) \geq \frac{1}{k}\sum_{i=1}^{k} f(\lambda_i).$$

Thus for every probability vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$

$$\frac{1}{k}\sum_{i=1}^{k} f(\lambda_i) \leq f\left(\frac{1}{k}\sum_{i=1}^{k} \lambda_i\right) = f\left(\frac{1}{k}\right) = -\frac{1}{k}\log\frac{1}{k} = \frac{1}{k}\log k,$$

and so $H(\lambda) \leq \log k$. Clearly $H(1/k, 1/k, \ldots, 1/k) = \log k$.

For the proof of subadditivity, see Shields [46], Theorem I.6.3.

We now define conditional entropy, which is of great importance.

Definition 2.3.2 (Conditional entropy of random variables)**.** *Let X and Y be two random variables with joint distribution $p(x,y)$ and conditional distributions $p(x|y) = p(x,y)/p(y)$ along with its symmetrical. The* conditional entropy *of X given Y is defined by*

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y).$$

It is very easy to prove the addition law

$$H(X,Y) = H(Y) + H(X|Y).$$

The previous unconditional inequalities extend to conditional entropy as follows.

Proposition 2.3.2. *(a)* **Positivity.** *$H(X|Y) \geq 0$, with equality if and only if X is a function of Y.*

(b) **Boundedness.** *$H(X|Y) \leq H(X)$, with equality if and only if X and Y are independent.*

(c) **Subadditivity.** *$H((X,Y)|Z) \leq H(X|Z) + H(Y|Z)$, with equality if and only if X and Y are conditionally independent given Z.*

We can now define entropy for random processes. We state the definition only for stationary processes, because throughout this thesis we shall never deal with non-stationary ones.

Definition 2.3.3 (Entropy of a random process)**.** *Let $X = \{X_1, X_2, \ldots\}$ be a stationary random process with finite alphabet A and measure μ. The* n-th order entropy *of X is*

$$H_n(\mu) = H(X_1^n) = -\sum_{x_1^n \in A^n} \mu(x_1^n) \log \mu(x_1^n).$$

The n-th order conditional entropy *of X is*

$$h_n(\mu) = H_n(\mu) - H_{n-1}(\mu) = -\sum_{x_1^n \in A^n} \mu(x_1^n) \log \mu(x_n | x_1^{n-1}) = H(X_n | X_1^{n-1}).$$

The entropy rate *or* process entropy *of X is*

$$h(\mu) = H(\{X_n\}) = \lim_{n\to\infty} \frac{H_n(\mu)}{n}. \tag{2.5}$$

The limit in Equation (2.5) exist (see for example Shields [46], pp. 59–60).

The entropy rate of a process can be defined in an alternative way as limit of conditional entropies, as stated in the following proposition. For its proof we refer the reader to any textbook treating process entropy.

Proposition 2.3.3. *For a random process as in Definition 2.3.3, it holds*

$$h(\mu) = H(\{X_n\}) = \lim_{n\to\infty} h_n(\mu) = \lim_{n\to\infty} H(X_n | X_1^{n-1}). \tag{2.6}$$

2.3.3 Entropy of a dynamical system

Let us consider a measure-preserving dynamical system $(X, \mathcal{B}, \mu, T)$. A finite-state stationary random process can be defined by specifying a finite measurable partition $\alpha = \{A_1, A_2, \ldots, A_n\}$ of X. Each point $x \in X$ determines an infinite symbolic sequence which at time j has the symbol A_i if $T^j x \in A_i$ (see Section 2.2.4).

Definition 2.3.4. *The* entropy of the partition $\alpha = \{A_1, A_2, \ldots, A_n\}$ *is the quantity*

$$H(\alpha) = -\sum_{i=1}^{n} \mu(A_i) \log_2 \mu(A_i). \tag{2.7}$$

A block of symbols $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ starting from place of index j corresponds to the entry of point $T^j x$ into the set $A_{i_1} \cap T^{-1}A_{i_2} \cap \ldots \cap T^{-k+1}A_{i_k}$. This set is an element of the partition $\alpha_0^{k-1} = \alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-k+1}\alpha$, which is the minimum common refinement of partitions $\alpha, T^{-1}\alpha, \ldots, T^{-k+1}\alpha$.

Definition 2.3.5. *If $\alpha = \{A_1, \ldots, A_n\}$ and $\beta = \{B_1, \ldots, B_m\}$ are two partitions, then*

- *$T^{-1}\alpha$ is the partition $\{T^{-1}(A_1), \ldots, T^{-1}(A_n)\}$,*
- *$\alpha \vee \beta$ is the partition $\{A_i \cap B_j : i = 1, \ldots, n; j = 1, \ldots, m\}$,*
- *β is a* refinement *of α (and we write $\beta \geq \alpha$), if each B_j is, except possibly for a zero-measure set, a subset of some A_i.*

With these definitions, the quantity H_n in Definition 2.3.3 is

$$\frac{1}{n}H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha)$$

and the entropy rate in Equation (2.5) is

$$h(\alpha, T) = h_\mu(\alpha, T) = \lim_{n\to\infty} \frac{1}{n}H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha).$$

Definition 2.3.6. *The quantity $h(\alpha, T)$ is called the* entropy of the transformation T with respect to the partition α.

The *entropy of the transformation* T is defined as the maximum uncertainty over all the possible finite-state processes associated with T:

$$h(T) = h_\mu(T) = \sup_\alpha h(\alpha, T). \tag{2.8}$$

The number $h(T)$ measures how much the transformation T 'mixes' the space.

Clearly, $h(T)$ is invariant in the class of isomorphism of T. More generally, if S is a factor of T then $h(S) \leq h(T)$ (because each partition relative to the system S determines a partition of the same entropy relative to T, but not necessarily viceversa).

A natural question, given a transformation T, is whether there exist a partition α such that

$$h(\alpha, T) = h(T).$$

In 1958 Kolmogorov ([51]) defined the entropy of T as $h(\alpha, T)$ when T has a generating partition α and ∞ otherwise. The definition given by Equation 2.8 was given by Sinai ([52]) and the equivalence between the two definitions is known as the Kolmogorov-Sinai theorem.

Theorem 2.3.4 (Kolmogorov-Sinai)**.** *If α is a generating partition for T, then*

$$h(T) = h(\alpha, T)$$

For the proof of the theorem we refer the reader to Petersen [48].

The existence of a generating partition for a dynamical system is assured by the following theorem [53].

Theorem 2.3.5 (Krieger, 1970)**.** *For an ergodic dynamical system $(X, \mathcal{B}, \mu, T)$ such that $h_\mu(T) < \infty$, there exists a finite generating partition α.*

The identification of a generating partition is generally a challenging task. In the following, we shall provide some examples of generating partitions in specific cases.

Chapter 3

Estimating entropy from samples

When a process' invariant measure is explicitly known, we could in principle estimate the entropy by applying the definition. On the other hand, when we are not given explicit knowledge of the measure, we are often not able to know exactly the entropy of the process and the problem of entropy estimation arises. A usual approach to this problem is considering long sample sequences, which are looked at as parts of infinite typical sequences and thus representing the statistical features of the system. To such samples several entropy estimating algorithms can be applied.

In particular, we investigate a symbolic substitution method denominated *non-sequential recursive pair substitution* (briefly, NSRPS) as a tool to estimate entropy of an ergodic source. The entropy we deal with is the Shannon entropy of finite-alphabet stationary stochastic processes, in particular those that can be obtained as a symbolic model of a dynamical system.

The use of symbolic models of dynamical systems as a benchmark for this kind of study is motivated by the following two important features.

- Dynamical systems can produce strings with many kinds of nontrivial statistical features (slow decay of correlations, no Markov structure, and so on).
- The dynamical/geometrical properties of the system under consideration often allow the entropy of the system to be estimated (sometimes rigorously calculated) by some other method (e.g., Lyapunov exponents and geometrical properties of the invariant measure) whose results can be compared with the estimation done by symbolic methods.

We shall compare the estimating algorithm suggested by the NSRPS method with two others, which we shall call the *empirical frequencies* (briefly,

EF) method and the *return times* (briefly, RT) method. We remark that these three methods can be applied directly to the symbolic sequence without having any other information on the source. For the ergodic transformations of the unit interval we shall also use another estimating algorithm which does not apply to symbolic processes. This is the approximation of the *Lyapunov exponent*, which converges very fast and will be considered as a reference value for the entropy. We remark that the estimation of entropy by this method uses some additional information on the system (the derivative of the map, which is calculated at each step of the dynamics, and the dimension of the invariant measure). Each estimation algorithm is supported by rigorous mathematical results, as it is shown in the following sections.

In Section 3.1 we formally present the NSRPS method and state the main theorem about it. In Sections 3.2, 3.3 and 3.4 we give a review of rigorous results supporting the estimation of entropy by the other methods we chose: empirical frequencies, return times and Lyapunov exponent. In Section 3.5 we discuss the details of the implementation of the above methods and the reasons of some arbitrary choice we could not avoid. In Section 3.6 we present the experimental results, with some tables and figures.

3.1 Non-sequential recursive pair substitution

3.1.1 Introduction

The idea of applying recursive pair substitutions to symbolic sequences was first proposed by Jiménez-Montaño, Ebeling and others (see [54], [55], [56]), but it was put into the formal context of probability theory and studied more deeply by Grassberger [57] in 2002 and Benedetto, Caglioti, Gabrielli [58] in 2006.

We now briefly explain how the NSRPS method works.

Let us suppose to have a finite-state stationary source, that is a device providing infinite sequences of symbols $x_0x_1x_2\ldots$ where each x_i is an element of a finite alphabet A, in such a way that the probability of receiving a given finite string does not vary with time. Given a sequence from such a source, the NSRPS method prescribes to individuate the pair (or one of the pairs) of symbols of maximal frequency and to substitute all its non-overlapping occurrences with a new symbol $\alpha \notin A$. For example, given the sequence

$$01101011101100011101101001 1\ldots,$$

taken from a source μ for which $\mu(01)$ is the highest among the probabilities of symbol pairs, we substitute the pair 01 with the new symbol 2, thus

obtaining

$$2122112100211212021\ldots.$$

In the case the pair to substitute is made up of two equal symbols, not *all* the occurrences are to be substituted, but only the non-overlapping ones. For example, given the sequence

$$00110100001010001000001100001\ldots,$$

we substitute the pair 00, obtaining

$$2110122101201220112 21\ldots.$$

Starting from a source μ with alphabet $A = \{0, 1\}$, after the first substitution we shall have a new source with alphabet $A_1 = \{0, 1, 2\}$ and a measure μ_1 on the finite strings inherited from μ. We can then go on repeating the steps, introducing new symbols $3, 4, \ldots$ and obtaining new sources $\mu_2, \mu_3, \ldots$.

The main theorem about the NSRPS method (Theorem 3.1.3) says that the entropy h of an ergodic source μ, which is given by

$$h(\mu) = \lim_{k\to\infty} -\frac{1}{k} \sum_{\text{length}(\underline{x})=k} \mu(\underline{x}) \log_2 \mu(\underline{x}),$$

can be calculated, in the limit for the number of substitutions N approaching infinity, knowing only the probabilities according to μ_N of the individual symbols and of the pairs in the new sources, after many substitutions. We remark that the hypotheses of substituting at each step one of the pairs with the maximum probability is a sufficient but not necessary condition for the conclusion of the main theorem 3.1.3 to hold (see [58]).

Numerical results about the use of this method for the estimation of the entropy of the English language were sketched by Grassberger in [57]. We show a first systematic comparison of this method with other classical ones, by performing several experiments on artificial symbolic sequences constructed from dynamical systems.

3.1.2 Definitions and results on the NSRPS method

In this section we briefly recall from [58] definitions and main results on the NSRPS method. We introduce the terms and the notations which are fundamental to state the main theorem 3.1.3. We omit all the technical details and the proofs, for which we refer to [58].

We recall from the introduction to the NSRPS method that the method we study is applied to symbolic sequences which are supposed to come from a finite-state stationary source.

Let us call our finite alphabet A and denote with $A^* = \cup_{k=1}^{\infty} A^k$ the collection of all finite words in the alphabet A. A word $\underline{w} \in A^*$ has length $|\underline{w}|$ and, if $|\underline{w}| = k$, it will also be indicated with $w_1^k = w_1 \dots w_k$.

Let $x, y \in A$, $\alpha \notin A$ and $A_1 = A \cup \{\alpha\}$.

Definition 3.1.1.

A pair substitution is a function $G = G_{xy}^{\alpha} : A^* \to A_1^*$ which is defined by recursively substituting all the non-overlapping occurrences of the pair xy. More precisely, $G\underline{w}$ is defined substituting in $\underline{w}$ the first occurrence from left of xy with α and repeating this procedure to the end of the sequence.

We consider again the example sketched in the introduction and show some general notation. Given the sequence

$$\underline{w} = 01101011101100011101101001 1 \in \{0,1\}^*,$$

performing the substitution $01 \mapsto 2$ leads to

$$G_{01}^2(\underline{w}) = 2122112100211212021 \in \{0,1,2\}^*.$$

We indicate with $\mathcal{E}(A)$ the set of all the stationary ergodic measures on $A^{\mathbb{Z}}$, the only ones we shall deal with. If $\mu \in \mathcal{E}(A)$ and $\underline{w} \in A^*$, we shall use the notation $\mu(\underline{w})$ to indicate the μ-measure of the cylinder set $[w_1, \dots, w_k] = \cap_{i=1}^{k} \{X_i = w_i\}$, where the X_i's are the random variables which describe the stochastic process.

The map $G = G_{xy}^{\alpha}$ naturally induces a map $\mathcal{G} = \mathcal{G}_{xy}^{\alpha} : \mathcal{E}(A) \to \mathcal{E}(A_1)$, as the following theorem shows. We indicate with $\#\{\underline{s} \subseteq \underline{r}\}$ the number of occurrences of a subword $\underline{s}$ in a word $\underline{r}$.

Theorem 3.1.1. *If $\mu \in \mathcal{E}(A)$ and $\underline{s} \in A_1^*$, then the limit*

$$\mathcal{G}\mu(\underline{s}) = \lim_{n \to \infty} \frac{\#\{\underline{s} \subseteq G(w_1^n)\}}{|G(w_1^n)|}$$

exists and is constant μ almost everywhere in $\underline{w}$. Furthermore, the values $\{\mathcal{G}\mu(\underline{s})\}_{\underline{s} \in A_1^*}$ are the marginals of an ergodic measure on $A_1^{\mathbb{Z}}$.

It is obvious that a pair substitution shortens the sequence it is applied to. The following proposition gives an average quantification of this shortening.

Proposition 3.1.2. *If $x \neq y$ then*

$$Z_{xy}^{\mu} \stackrel{def}{=} \lim_{n \to \infty} \frac{n}{|G(w_1^n)|} = \frac{1}{1 - \mu(xy)} \quad (\mu \text{ a. e. in } \underline{w}).$$

If $x = y$ then

$$Z^{\mu}_{xx} \stackrel{def}{=} \lim_{n\to\infty} \frac{n}{|G(w_1^n)|} = \frac{1}{1 - \sum_{k=2}^{\infty}(-1)^k \mu(\underline{x}^k)} \quad (\mu \text{ a. e. in } \underline{w}),$$

where $\underline{x}^k$ is the string made up of k symbols x.

Recall from Definition 2.3.3 that, given $\mu \in \mathcal{E}(A)$ and $n \geq 1$, the *n-th order entropy* is

$$H_n(\mu) = -\sum_{|\underline{w}|=n} \mu(\underline{w}) \log_2 \mu(\underline{w}),$$

the *n-th order conditional entropy* is

$$h_n(\mu) = H_{n+1}(\mu) - H_n(\mu),$$

and the *Shannon entropy of the process μ* is

$$h(\mu) = \lim_{n\to\infty} h_n(\mu) = \lim_{n\to\infty} H_n(\mu)/n.$$

3.1.3 The main theorem

Intuitively, after a pair substitution the information is more concentrated, with respect to the original sequence. After several substitutions, the most important blocks (the most frequent ones) are concentrated into symbols and the value of the entropy can be well estimated by the entropy of short blocks (H_k with small k). This can be formulated in precise terms (see [58], Theorem 3.2 and Corollary 3.1.5) and suggests that a sequence of substitutions might asymptotically transfer *all* the information to the distribution of the pairs and individual symbols. This is precisely the content of the main theorem.

To state it, we define the following objects:

- the alphabets $A_N = A_{N-1} \cup \{\alpha_N\}$ where $\alpha_N \notin A_{N-1}$ and $A_0 = A$;

- the maps $G_N = G^{\alpha_N}_{x_N y_N} : A^*_{N-1} \to A^*_N$, where $x_N, y_N \in A_{N-1}$;

- the maps between measures $\mathcal{G}_N = \mathcal{G}^{\alpha_N}_{x_N y_N}$;

- the measures $\mu_N = \mathcal{G}_N \mu_{N-1}$, with $\mu_0 = \mu$;

- the quantities $Z_N = Z^{\mu_{N-1}}_{x_N y_N}$ and $\overline{Z}_N = Z_N \dots Z_1$.

Theorem 3.1.3 ([58], Theorem 3.2)**.** *If*

$$\lim_{N\to\infty} \overline{Z}_N = +\infty$$

then

$$h(\mu) = \lim_{N\to\infty} \frac{h_1(\mu_N)}{\overline{Z}_N}.$$

Theorem 3.1.4. *If at each step N the pair $x_N y_N$ is a pair with the maximum frequency among all the pairs of symbols of A_{N-1}, then*

$$\lim_{N\to\infty} \overline{Z}_N = +\infty.$$

Theorems 3.1.3 and 3.1.4 combined guarantee that, by performing at each step the substitution of a pair with maximum probability, the entropy of the original ergodic process is approximated by the 1-st order conditional entropy, which takes into consideration only the distribution of the single symbols and of the pairs of symbols. In this sense, through this method "the ergodic process becomes 1-Markov in the limit".

In practical utilisations of the above theorem we have access to the statistical properties of the source by measuring the empirical frequency of digit sequences in the experimental data we have. Given a sequence $x_1 x_2 \dots x_n$, the empirical distribution of the (overlapping) k-blocks a_1^k is defined naturally by

$$p_k(a_1^k | x_1^n) = \frac{\#\{i \in [1, n-k+1] : x_i^{i+k-1} = a_1^k\}}{n-k+1} \tag{3.1}$$

and its empirical k-entropy is defined by

$$\tilde{H}_k(x_1^n) = -\sum_{|\underline{w}|=k} p_k(\underline{w}|x_1^n) \log_2 p_k(\underline{w}|x_1^n).$$

Let us call G the substitution operation on the maximal frequency pair (if there are more than one string of maximal frequency, the lexicographic order is used). By ergodicity, it is possible to rephrase the above theorem into a statement which is more similar to what can be pratically done on long strings coming from the source.

Corollary 3.1.5. *If μ is ergodic, for almost every $\omega \in A^{\mathbb{N}}$*

$$h(\mu) = \lim_{n\to\infty} \lim_{l\to\infty} \frac{\tilde{H}_2(G^n(\omega_1^l)) - \tilde{H}_1(G^n(\omega_1^l))}{\tilde{Z}_n(\omega_1^l)},$$

where $\tilde{Z}_n(\omega_1^l) = \frac{l}{|G^n(\omega_1^l)|}$ is the shortening rate after n substitutions.

Proof. Let ω be a typical realisation of the system. Since the system is ergodic, $\lim_{l\to\infty} \tilde{H}_k(G^n(\omega_1^l)) = H_k(\mu_n)$, hence $\lim_{l\to\infty} \tilde{H}_2(G^n(\omega_1^l)) - \tilde{H}_1(G^n(\omega_1^l)) = h_1(\mu_n)$. Moreover, in the same way, when n is fixed and $l \to \infty$, $\tilde{Z}_n(\omega_1^l) \to \overline{Z}_n$ and the corollary follows from the above Theorem 3.1.3. □

3.2 Empirical frequencies (EF)

To estimate entropy directly by the definition, a simple procedure consists in determining the empirical distribution p_k of the overlapping k-blocks and taking $\frac{H_k(p_k)}{k}$ as an estimate for h. If k is fixed and the length of the sample sequence n tends to infinity, then $\frac{H_k(p_k)}{k}$ almost surely converges to $\frac{H_k(\mu_k)}{k}$, which tends to h as $k \to \infty$. Theorem 3.2.1 below guarantees that these two limits can be performed together with $k(n) \sim \log_2 n$.

Given the sequence $x_1 x_2 \dots x_n$, the empirical distribution $p_k(\cdot|x_1^n)$ of the overlapping k-blocks is defined as in (3.1). The following holds.

Theorem 3.2.1. *If μ is an ergodic measure of entropy $h > 0$, if $k(n) \to \infty$ as $n \to \infty$ and if $k(n) \leq \frac{\log_2 n}{h}$, then*

$$\lim_{n\to\infty} \frac{1}{k(n)} \tilde{H}_{k(n)}(x_1^n) = h, \quad a.\ s.$$

The above theorem gives an indication about how in practice it is possible to deal with the double limit in the estimation of entropy (length of the sample, length of the blocks) reducing it to a single limit. For the proof and further details see [46], Theorem II.3.5 and Remark II.3.6. We remark that the same result holds for non-overlapping distributions. The reason why we chose to consider the overlapping one is to enrich the statistic as much as possible, as it will be explained in Section 3.5.

In practical applications it turns out that the empirical entropy is not the best possible statistical estimator for the entropy. In cases when the length of the sample is not large enough compared to the length of the blocks (see Theorem 3.2.1 for the orders of magnitude), the empirical entropy tends to underestimate the entropy. To correct this bias, some better estimators have been defined in the literature, first by Miller [59] and then by Grassberger with a furtherly improved estimator.

In our experiments we use the Grassberger estimator $\hat{H}_{\mathrm{G}}$, defined as follows (see [60] for details). Let us consider the sequence $G_1, G_2, \dots$ defined

by

$$\begin{aligned} G_1 &= -\gamma - \ln 2 \\ G_2 &= 2 - \gamma - \ln 2 \\ G_{2n+1} &= G_{2n} \\ G_{2n+2} &= G_{2n} + \tfrac{2}{2n+1} \quad (n \geq 1), \end{aligned}$$

where $\gamma = 0.577215\ldots$ is Euler's constant. For N points distributed randomly and indipendently into M boxes, according to probabilities $p_1, \ldots, p_M$, the entropy $H = -\sum_{i=1}^{M} p_i \ln p_i$, which is naively estimated by $\hat{H}_{\text{naive}} = \ln N - \frac{1}{N}\sum_{i=1}^{M} n_i \ln n_i$ (the n_i's being the observed occurrences; notice that in the case of strings this corresponds to the empirical entropy), is better estimated by

$$\hat{H}_{\text{G}} = \ln N - \frac{1}{N}\sum_{i=1}^{M} n_i G_{n_i}.$$

In our context, when estimating the entropy H_k, the set of boxes is the set of possible strings of length k. This set has $M = 2^k$ elements in the case of a binary alphabet. The occurrences of the k-blocks in the sample give the various n_i's.

We end remarking that, while experimental examples contained in this study are long artificial trajectories mostly coming from dynamical systems, when working on short sequences (for instance, finite realisation of some biophysical processes or experiments), surrogate analysis can be useful in order to take into account fluctuations of entropy (see e.g. [61], [62]).

3.3 Return times (RT)

Ornstein and Weiss proved an interesting result which links entropy and the so-called return times for ergodic processes. They showed in [63] that the logarithm of the waiting time until the first n terms of a sequence x occur again in x is almost surely asymptotic to nh.

Definition 3.3.1. *Given a sequence x taken from an ergodic process, we define the n-th return time as*

$$R_n(x) = \min\{m \geq 1 : x_{m+1}^{m+n} = x_1^n\}.$$

Theorem 3.3.1. *If μ is an ergodic process with entropy h, then*

$$\lim_{n\to\infty} \frac{1}{n} \log_2 R_n(x) = h, \quad a.\ s.$$

For the original proof see [63], for an alternative one see [46], Theorem II.5.1.

3.4 Lyapunov exponent

If we are interested in the estimation of the entropy of a one dimensional system a powerful tool is the Lyapunov exponent.

Let us consider a map $T : [0,1] \to [0,1]$ having an ergodic invariant measure μ. We define its Lyapunov exponent by

$$\lambda_\mu = \int_0^1 \log_2 T' \mathrm{d}\mu.$$

Under some assumptions (see below) this quantity is related to the fractal dimension $HD(\mu)$ of μ and the entropy h_μ of the system by the formula $HD(\mu) = \frac{h_\mu}{\lambda_\mu}$. Hence, if we know $HD(\mu)$ and estimate λ_μ numerically, we obtain an estimate for h_μ.

Let us give a precise statement for one dimensional systems (see [64] for a generalisation to multidimensional systems). A map $T : [0,1] \to [0,1]$ is called piecewise monotonic if there is a sequence $\{Z_i\}_{i\in\mathbb{N}}$ of disjoint open subintervals of $[0,1]$ such that $T|_{Z_i}$ is strictly monotone and continuous for each i.

Let us consider the set $E_Z = \cap_{i\in\mathbb{N}} T^{-i}(\cup_{j\in\mathbb{N}} Z_j)$, where all iterates of T are in the open intervals. Let μ be an invariant ergodic measure such that $\mu(E_Z) = 1$. Let us consider its Lyapunov exponent λ_μ and its Kolmogorov-Sinai entropy h_μ. Let us denote by $HD(X)$ the Hausdorff dimension of a subset $X \subset [0,1]$. The Hausdorff dimension $HD(\mu)$ of a measure μ is defined as the infimum $HD(\mu) = \inf_{\mu(X)=1}(HD(X))$ of the dimension of full measure sets.

Let us consider the p-variation of a function $f : [0,1] \to \mathbb{R}$ on a subinterval $[a,b]$ defined by:

$$\mathrm{var}^p_{[a,b]}(f) = \sup\left\{ \sum_{i=1}^m |f(x_{i-1}) - f(x_i)|^p \;\middle|\; m \in \mathbb{N}, a \le x_0 < \ldots < x_m \le b \right\}.$$

We say that the derivative of a piecewise monotonic map has bounded p-variation if there is a function g such that $g(x) = 0$ on $[0,1] \setminus E_Z$, $g = T'$ on each Z_i and $\mathrm{var}^p_{[0,1]}(g) < \infty$.

Theorem 3.4.1 ([65])**.** *Let T be a map on $[0,1]$ with finitely many monotonic pieces and a derivative of bounded p-variation for some $p \ge 0$. If μ is an ergodic invariant measure with Lyapunov exponent $\lambda_\mu > 0$, then*

$$HD(\mu) = \frac{h_\mu}{\lambda_\mu}.$$

In many of the systems we will study we have that the invariant measure μ we consider is absolutely continuous with respect to the Lebesgue measure with a regular (bounded variation or continuous) density, hence $HD(\mu) = 1$. The Lyapunov exponent will be then numerically estimated with a Birkhoff average along a typical orbit of the system, hence giving

$$h_\mu = \int_0^1 \log_2 T' \mathrm{d}\mu = \lim_{n\to\infty} \frac{\sum_{i=1}^n \log_2(T'(T^i(x_0)))}{n}$$

for μ-a.e. x_0, by the Ergodic Theorem. Experimental results indicate that this limit converges very fast and gives a very good estimation for h_μ.

3.5 Computer simulations

Concerning the results of the computer simulations, some comments are due on the way we implemented the entropy estimating algorithms.

- About empirical frequency estimation, in our simulations we could not consider blocks much longer than 23 bits. This is because the algorithm takes a time which grows exponentially in the length of the blocks considered. The empirical distribution of blocks of various lengths was calculated on the entire symbolic sequence. The values taken as entropy estimates are those calculated by means of the estimator $\hat{H}_\mathrm{G}$ (divided by $\ln 2$ to have the entropy expressed in bits). In all but three cases they do not differ significantly (the differences being at most 10^{-3}) from the ones calculated in the naive way and in the other three cases they are actually better (the expanding map E, the logistic map Λ_4 and the Manneville map with exponent $\frac{3}{2}$).

- The return times method was performed by calculating the return times of strings long up to $\log_2 n$, where n is the length of the symbolic sequences. Moreover, in order to have more reliable results, for every binary sequence we considered not only the return times of the initial strings x_1^k, but also of x_2^{k+1}, x_3^{k+2}, $\ldots$, x_{1000}^{k+999}, and took the average of their logarithms, hence what we measure is an average return time indicator.

- In the implementation of the NSRPS method, at every step the substitution with a new symbol of a pair with maximum probability was performed, then we calculated the conditional entropy of order 1 and the inverse of the mean shortening Z_N estimating the entropy according to Corollary 3.1.5.

The implementation of the substitutions method did not show meaningful computational constraints, since performing a pair substitution requires a very short time. Nevertheless, there is one algorithmic question to be answered: the identification of a stop condition.

For the estimation of the entropy with NSRPS, at the moment we have not an analogous of Theorem 3.2.1, hence we have to find how many substitutions it is convenient to made on a finite sample string. We had to understand when to stop the substitutions before the sequence becomes too short and consequently the statistics becomes too poor. We chose to stop when the following condition has occurred:

StopCond: the substituted pair has frequency < 0.02.

The stop condition above is somewhat artificial and has no intrinsic relation with the symbolic process. In all the cases we studied we knew the true entropy or estimated it quite precisely by means of the Lyapunov exponent, so that we could understand when the approximation through the pair substitutions method was good. In all our processes, for which we took symbolic samples long 15 millions bits, it seems that few tens of pair substitutions are enough for the estimate to become more or less constant when considering the first three decimal digits. Obviously, when the process is independent or 1-Markov at most one pair substitution is needed in order to have a very precise estimate of the entropy. On the contrary, processes which have long memory properties need many pair substitutions. The stop condition we used does not take into account the memory properties of the process, so that it lets the algorithm performing unnecessary pair substitutions in low-Markov cases and stops it before useful substitutions in long-memory processes. Although a threshold lower than 0.02 in *StopCond* could improve the estimates, the goal is to find some criterion, both user-independent and sequence-dependent, which determines for each case the most appropriate number of substitutions to perform.

3.6 Experiments

We now describe the transformations of the unit interval generating the symbolic sequences to which we applied the entropy estimating algorithms.

3.6.1 Maps

We considered a few maps of the interval, to which we applied the construction explained in Section 2.2.4 to obtain symbolic sequences.

Piecewise expanding maps

We considered a piecewise expanding map E, defined by

$$Ex = \begin{cases} \frac{4x}{3-2x} & \text{if } x \in [0, \frac{1}{2}[\\ \frac{2x-1}{2-x} & \text{if } x \in [\frac{1}{2}, 1] \end{cases},$$

which is discontinuous in $\frac{1}{2}$ and has two surjective branches (see Figure 3.1). It holds $E'(x) > k$ for all x, where $k > 1$ is a constant. As it is well known (see e.g. [66]), a map of this kind has a unique absolutely continuous invariant measure with dimension 1. Moreover, Theorem 3.4.1 applies and we can estimate the entropy by the Lyapunov exponent. A generating partition for E is $\{[0, \frac{1}{2}[, [\frac{1}{2}, 1]\}$ (see [67], Exercise 3.4).

We show the results of the entropy estimates in Table 3.1 and Figure 3.1.

map	h_{Lyap}	h_{EF}	h_{RT}	h_{NSRPS} (N_{sub})
E	0.8673	0.864	0.838	0.867 (17)

Table 3.1: *Entropy estimates for the piecewise expanding map E. The values h_{Lyap}, h_{EF}, h_{RT} and h_{NSRPS} are the entropy estimates as Lyapunov exponent or by empirical frequencies, return times, NSRPS, respectively. N_{sub} is the number of pair substitutions executed when the stop condition* StopCond *occurs.*

The NSRPS method gives the best estimate. Though, the substitutions themselves have no particular role, since the map seems to be 1-Markov (the first value calculated with the substitutions algorithm is already very close to the true entropy).

Lorenz-like maps

Another example of map with two non-surjective branches is a Lorenz-like map (similar maps are involved in the study of the famous Lorenz system) defined by

$$Lx = \begin{cases} 1 - \left(\frac{-6x+3}{4}\right)^{\frac{3}{4}} & \text{if } x \in [0, \frac{1}{2}[\\ \left(\frac{6x-3}{4}\right)^{\frac{3}{4}} & \text{if } x \in [\frac{1}{2}, 1] \end{cases}.$$

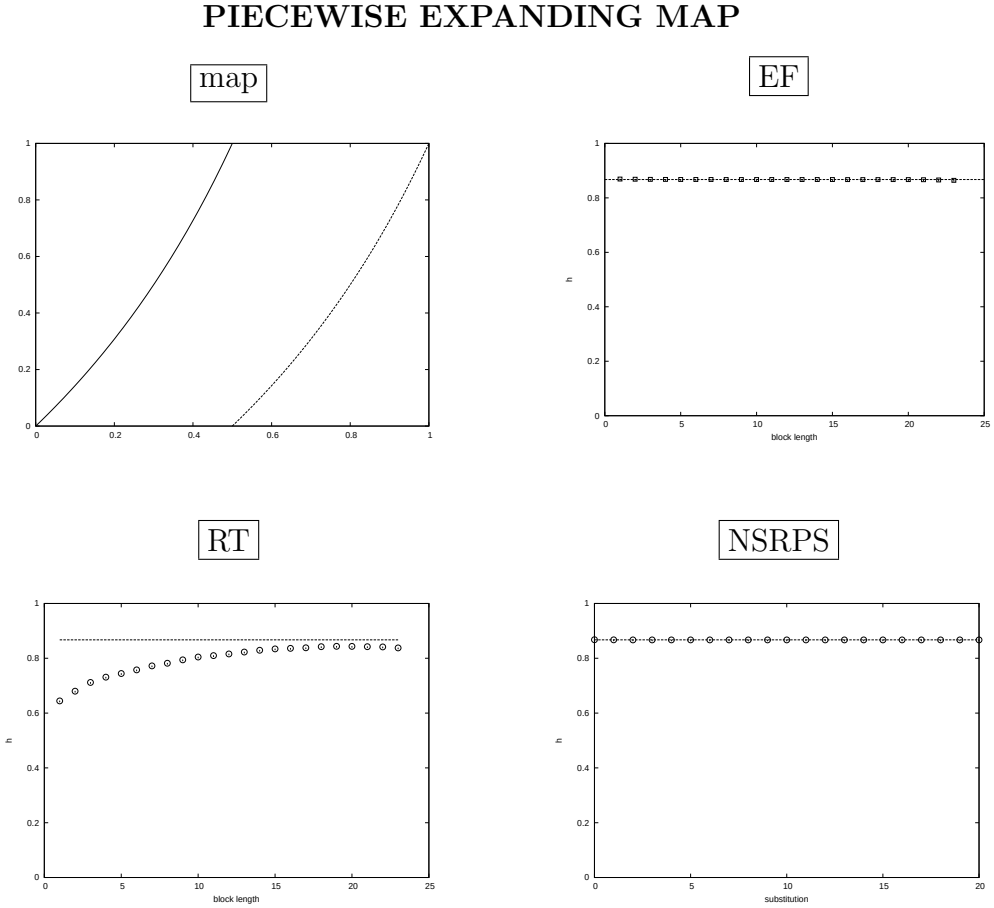

Figure 3.1: *Piecewise expanding map E and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.*

The derivative of L is uniformly greater than 1 for all $x \in [0,1] \setminus \{\frac{1}{2}\}$ and $L'\left(\frac{1}{2}^{\pm}\right) = +\infty$ (see Figure 3.2).

As for the previous piecewise expanding map E, the Lorenz-like map L has a unique absolutely continuous invariant measure with dimension 1 (see [66]). Theorem 3.4.1 does not apply in this case because the derivative is not bounded and hence has not p-bounded variation. However the usual relation between entropy and Lyapunov exponent holds and can be recovered by [68]. Moreover, the natural partition $\{[0, \frac{1}{2}[, [\frac{1}{2}, 1]\}$ is generating (see again [67]).

In Table 3.2 and Figure 3.2 the results obtained for the map L are shown.

The Lorenz-like map L appears not to be 1-Markov. In fact, from the plot relative to NSRPS in Figure 3.2 it can be noticed that the best value is not the

map	h_{Lyap}	h_{EF}	h_{RT}	h_{NSRPS} (N_{sub})
L	0.7419	0.765	0.723	0.748 (17)

Table 3.2: *Entropy estimates for the Lorenz-like map L. N_{sub} is the number of pair substitutions executed when the stop condition* StopCond *occurs.*

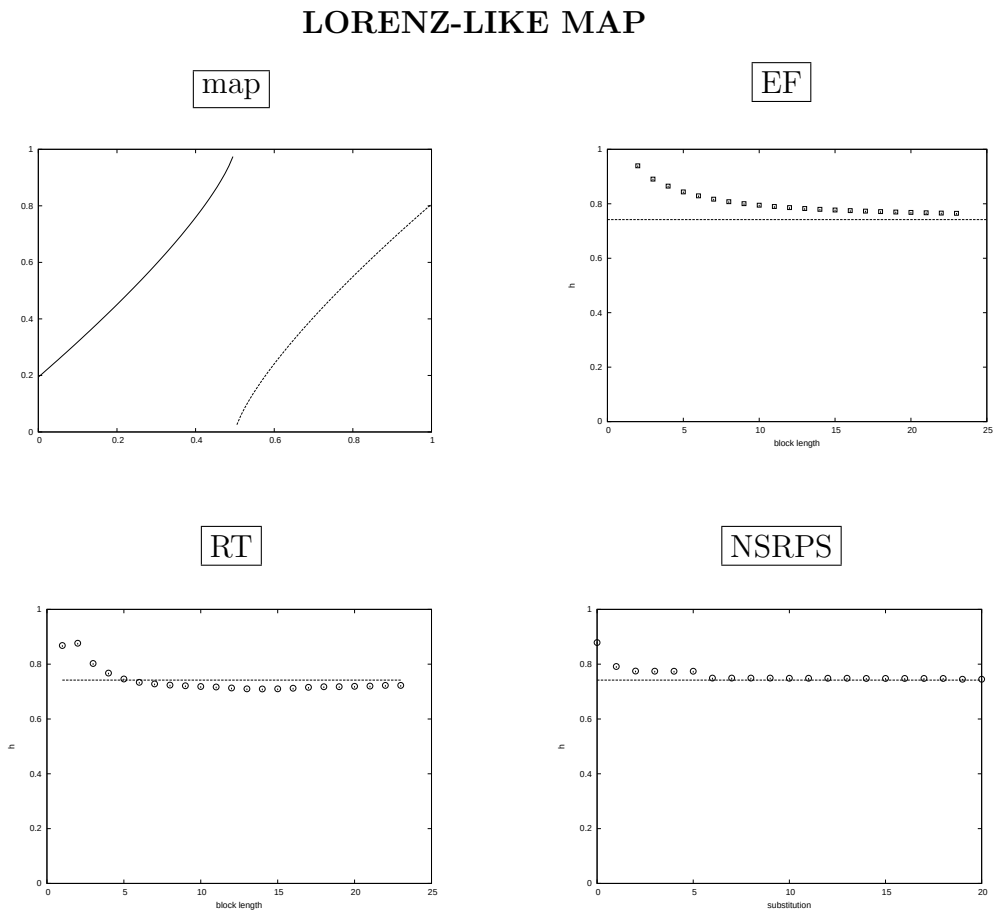

Figure 3.2: *Lorenz-like map L and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.*

first estimated, that is simply the 1-st order conditional entropy h_1. Instead, there are pair substitutions that significantly improve the approximation of the entropy. These substitutions are those which condense more information than others. Furthermore, this is one of the cases in which a few more pair

substitutions after condition *StopCond* occurs give a better estimate.

Logistic maps

The logistic maps are of the form

$$\Lambda_\lambda x = \lambda x(1-x), \quad 1 \leq \lambda \leq 4$$

We took $\lambda = 4$, 3.8 and 3.6 (the graph of $\Lambda_{3.8}$ is shown in Figure 3.3 (map)). For all these three maps, the partition $\{[0, \frac{1}{2}[, [\frac{1}{2}, 1]\}$ is generating (see [67]).

For $\lambda = 4$ there is a unique invariant measure, which is ergodic and absolutely continuous with respect to Lebesgue and whose density is $\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}$. Furthermore, the dynamical system $([0,1], \mathcal{B}([0,1]), \rho(x)\mathrm{d}x, \Lambda_4)$ is isomorphic to the shift on the Bernoulli process with alphabet $\{0,1\}$ and parameter $\frac{1}{2}$. Thus, for the entropy it holds $h(\Lambda_4) = 1$.

About the maps $\Lambda_{3.8}$ and $\Lambda_{3.6}$ we remark that the assumptions of Theorem 3.4.1 still hold and the dimension of the invariant measure is estimated to be very close to 1 (see [69]). Hence we assume to be reasonable to estimate the entropy by the Lyapunov exponent.

In Table 3.3 we summarise the final entropy estimates obtained with the four methods for the three logistic maps, while in Figure 3.3 we show in graphical form the complete results for the map $\Lambda_{3.8}$.

map	h_{Lyap}	h_{EF}	h_{RT}	h_{NSRPS} (N_{sub})
Λ_4	1.0000	1.000	0.959	1.000 (17)
$\Lambda_{3.8}$	0.6234	0.652	0.610	0.628 (18)
$\Lambda_{3.6}$	0.2646	0.348	0.314	0.269 (18)

Table 3.3: *Entropy estimates for the logistic maps Λ_λ. N_{sub} is the number of pair substitutions executed when the stop condition* StopCond *occurs.*

For the map Λ_4 the NSRPS method does not require any substitution to correctly estimate the entropy up to the sixth decimal digit. This is no surprise, since the symbolic process associated with Λ_4 is independent.

Instead, for the map $\Lambda_{3.8}$ it happens that, similarly to the NSRPS case of the map L (see Figure 3.2), there are pair substitutions which are more important than others in approximating the value of the entropy, as it can be noticed in Figure 3.3.

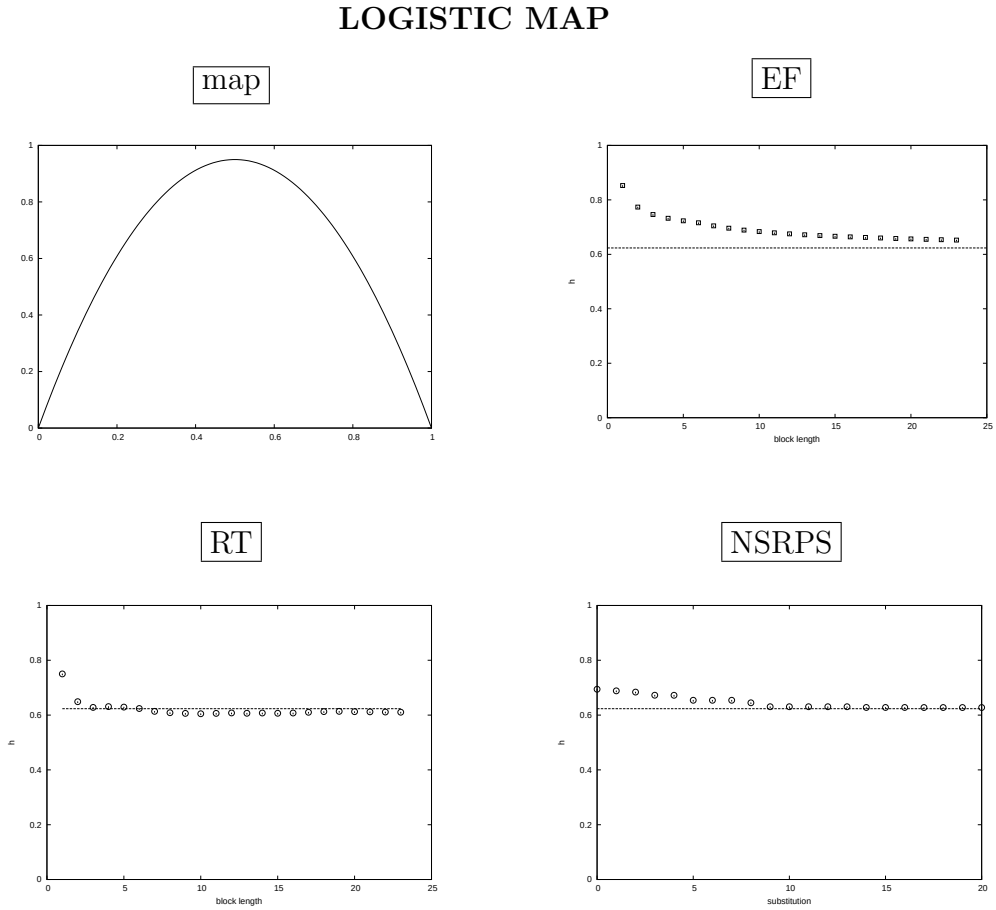

Figure 3.3: *Logistic map for $\lambda = 3.8$ and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.*

The entropy estimating algorithms give for the map $\Lambda_{3.6}$ results that are qualitatively similar to those of $\Lambda_{3.8}$.

Manneville-Pomeau maps

Manneville maps exhibit dynamics with long range correlations. They are defined by

$$M_z x = x + x^z \pmod 1, \quad z \in \mathbb{R}^+.$$

Such maps have great interest in physics and possess different characteristics as the exponent z varies. We focused our attention on the values $1 < z < 2$, for which the maps admit a unique absolutely continuous invariant probability measure (with unbounded density). For these parameters,

the system has power law decay of correlations, and the rate is slower and slower as z approaches 2 (see e.g. [66], section 3). In this case the system has "long memory" and to estimate entropy by the empirical frequencies we would need long blocks. For $z \geq 2$ the absolutely continuous invariant measure is no longer finite. We also remark that since those maps have bounded variation derivative, in the cases where the absolutely continuous invariant measure is finite we can again estimate the entropy by the Lyapunov exponent. We took values of z which go very close to 2: $z_1 = \frac{3}{2}$, $z_2 = \frac{7}{4}$, $z_3 = \frac{15}{8}$, $z_4 = \frac{31}{16}$, $z_5 = \frac{63}{32}$, $z_6 = \frac{127}{64}$ (see the plot of M_{z_4} in Figure 3.4). For all $1 \leq i \leq 6$ it holds $M'_{z_i}(x) > 1$ for all $x \in]0, 1]$ and $M'_{z_i}(0^+) = 1$. For these maps the natural partitions $\{[0, c_i[, [c_i, 1]\}$, where $c_i \in]0, 1[$ is that value such that $M_{z_i}(c_i^-) = 1$ and $M_{z_i}(c_i^+) = 0$, are obviously generating.

The presence in 0 of an indifferent fixed point is the main responsible for the peculiar behaviour of the Manneville maps. When, starting from a random point x_0, after a certain number $\overline{n}$ of iterations the point $M_z^{\overline{n}} x_0$ happens to be very close to 0, the subsequent iterations remain very close to 0 for a long time. This fact translates in having many consecutive zeros in the binary symbolic sequence associated with the orbit of x_0. These strings of zeros can be long even hundreds of thousands of bits or more. The closer to 2 is the exponent z, the longer and more frequent these strings.

In carrying out the simulations for the Manneville maps and commenting their results, one cannot ignore the peculiarities of these maps. It turns out that the symbolic sequences we generated are too short to reflect the general characteristics of the maps. If in a sequence of 15 millions bits there happen to be groups of consecutive zeros that are hundreds of thousands of bits long, then the results obtained from such a sequence cannot be completely reliable. The usual approach to this problem is considering many sequences, generated from different initial random points, and taking the averages of the estimates. For the map M_{z_i} we considered $2i$ sequences, with $1 \leq i \leq 6$. Still, the values obtained from the various sequences are quite different, so that we cannot consider completely reliable the averages as well.

Bearing in mind these considerations, we report in graphic form the results for the Manneville map $M_{z_4} = z + z^{\frac{31}{16}} \pmod 1$ (see Figure 3.4), while the results for all the six Manneville maps considered are shown in Table 3.4.

For each Manneville map that we studied (except for $M_{\frac{7}{4}}$), the entropy estimates obtained through the NSRPS method were clearly the closest to the true entropy (which we assumed to be equal to the average Lyapunov exponent), although they were not as close as for the other maps or processes (see Section 3.6.2).

map	h_{Lyap}	h_{EF}	h_{RT}	h_{NSRPS} (N_{sub})
$M_{\frac{3}{2}}$	0.811	0.816	0.821	0.813 (18)
$M_{\frac{7}{4}}$	0.519	0.529	0.558	0.511 (20)
$M_{\frac{15}{8}}$	0.314	0.344	0.442	0.322 (19)
$M_{\frac{31}{16}}$	0.228	0.247	0.444	0.226 (21)
$M_{\frac{63}{32}}$	0.175	0.237	0.400	0.216 (21)
$M_{\frac{127}{64}}$	0.168	0.217	0.358	0.196 (21)

Table 3.4: *Entropy estimates for the Manneville maps M_{z_i}. N_{sub} is the average number of pair substitutions executed when the stop condition* StopCond *occurs.*

A skew product

We consider an example of a two dimensional system having long range correlations which is quite different from the Manneville map. Let us consider the following map $\mathcal{S} : [0,1]^2 \to [0,1]^2$ defined by

$$\mathcal{S}(x,y) = (Ex, y + \alpha\phi(x) \bmod 1)$$

where: $\phi(x) = \begin{cases} 1 \; if \; x \geq \frac{1}{2} \\ 0 \; if \; x < \frac{1}{2} \end{cases}$, α is a diophantine irrational and E is the one dimensional piecewise expanding map considered in Section 3.6.1. In the system the x coordinate is subjected to a chaotic transformation, while the y is rotated according to the value of x. Such systems preserve an absolutely continuous invariant measure and are mixing. Some estimations for the decay of correlations are given in [70].

We partitioned the unit square in four equal squares $Q_1, \ldots, Q_4$ having a common vertex at $(\frac{1}{2}, \frac{1}{2})$.

The entropy of $\mathcal{S}$ with respect to the partition $\{Q_1, \ldots, Q_4\}$ is the same as the entropy of E. Indeed, the rotation has zero entropy and a symbolic orbit for the two dimensional system can be constructed by the information given by its symbolic orbit for the one dimensional map E and the information relative to the rotation part. Although the entropy is the same, its estimation is much more complicated, as the experiments show.

In Figure 3.5 we consider the case where $\alpha = \frac{1+\sqrt{5}}{2}$ is the golden ratio. The empirical frequencies and the substitutions seem to converge to a value which is slightly greater than the true entropy. The return time instead seems

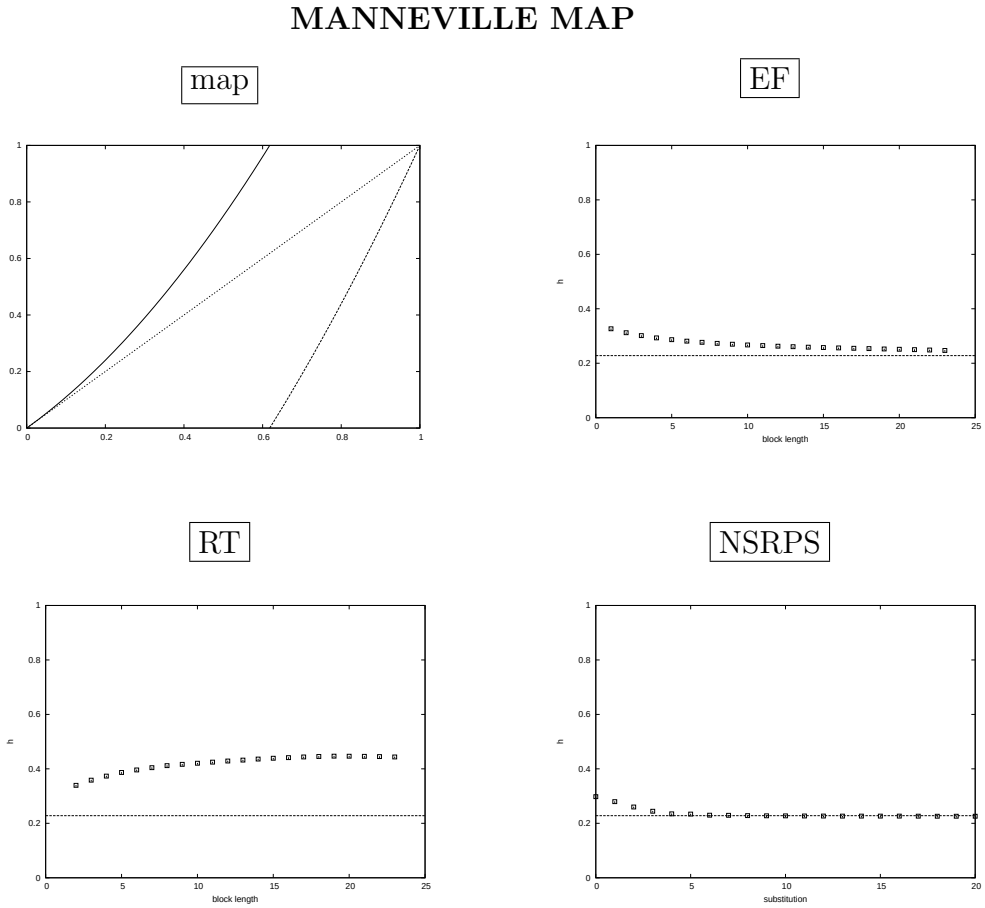

Figure 3.4: *Manneville map with $z = \frac{31}{16}$ and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.*

to better approximate the entropy in this case.

3.6.2 Renewal processes

Apart from the symbolic sequences obtained from ergodic transformations of the unit interval, we considered sequences taken from the so-called renewal processes.

A renewal process is a stationary process with alphabet $\{0, 1\}$ for which the distances between consecutive ones are independent and identically distributed random variables. When a symbol '1' occurs, the sequence forgets all its past and the probability of having the next '1' after j bits is p_j, where

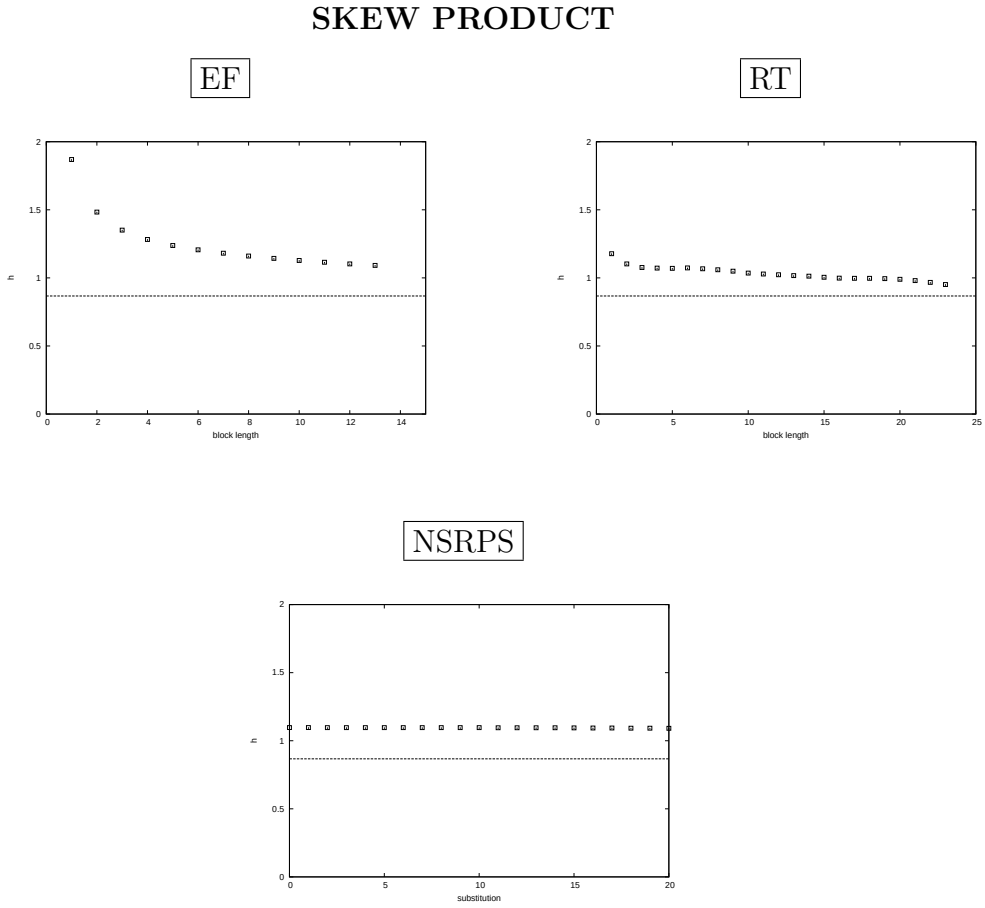

Figure 3.5: *Results for the skew product $\mathcal{S}$: entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the entropy value.*

$0 \leq p_j \leq 1$ and $\sum_{j=1}^{\infty} p_j = 1$.

We considered such renewal processes, with $p_1 = p_2 = \ldots = p_{2^k} = \frac{1}{2^k}$, $5 \leq k \leq 9$, which we shall indicate with RP_{2^k}.

For these renewal processes, the value of their entropy can be calculated exactly. We recall in fact that the entropy of a process is the number of bits per symbol that are necessary to describe the process itself. The quantity $C = -\sum_j p_j \log_2 p_j$ represents the number of bits that one needs to describe the process of the jumps between consecutive ones. In other words, C is the entropy of a random variable which describes the length of the jumps. If n is large, with n jumps (nC bits) we describe a sequence long about $n\overline{L}$ symbols,

where $\overline{L}$ is the average length of the jumps. Thus,

$$h(RP_{2^k}) \overset{n\to\infty}{\approx} \frac{nC}{n\overline{L}} = \frac{-\sum_{j\geq 1} p_j \log_2 p_j}{\sum_{j\geq 1} j p_j}.$$

In our cases, where $p_1 = \ldots = p_{2^k} = \frac{1}{2^k}$ and $p_j = 0$ for $j > 2^k$, we have

$$h(RP_{2^k}) = \frac{2k}{2^k + 1}.$$

In Table 3.5 we show the results of the entropy estimates for the renewal processes RP_{2^k} and those of RP_{32} are also plotted in Figure 3.6.

map	h	h_{EF}	h_{RT}	h_{NSRPS} (N_{sub})
RP_{32}	0.303030	0.321	0.272	0.303067 (11)
RP_{64}	0.184615	0.196	0.153	0.184793 (22)
RP_{128}	0.108527	0.115	0.110	0.108498 (25)
RP_{256}	0.062257	0.066	0.055	0.062239 (18)
RP_{512}	0.035088	0.037	0.039	0.035112 (16)

Table 3.5: *Entropy estimates for the renewal processes RP_{2^k}. N_{sub} is the number of pair substitutions executed when the stop condition* StopCond *occurs.*

For this process, the substitutions method gives an excellent approximation of the entropy already after five pair substitutions. After these substitutions all the memory of the process has been transferred to the distribution of the pairs, so that the sequence has become 1-Markov.

3.7 Conclusions

The performance of the three symbolic methods is summarised in Figure 3.7. NSRPS results to be the method that best approximates the entropy value. To this aim, it is a fast and light computational tool that may be used also for systems having low entropy or long range correlations where other statistical methods fail.

We showed for the first time a comparison in entropy estimation among NSRPS and other well-known methods. The results also open some further questions about NSRPS:

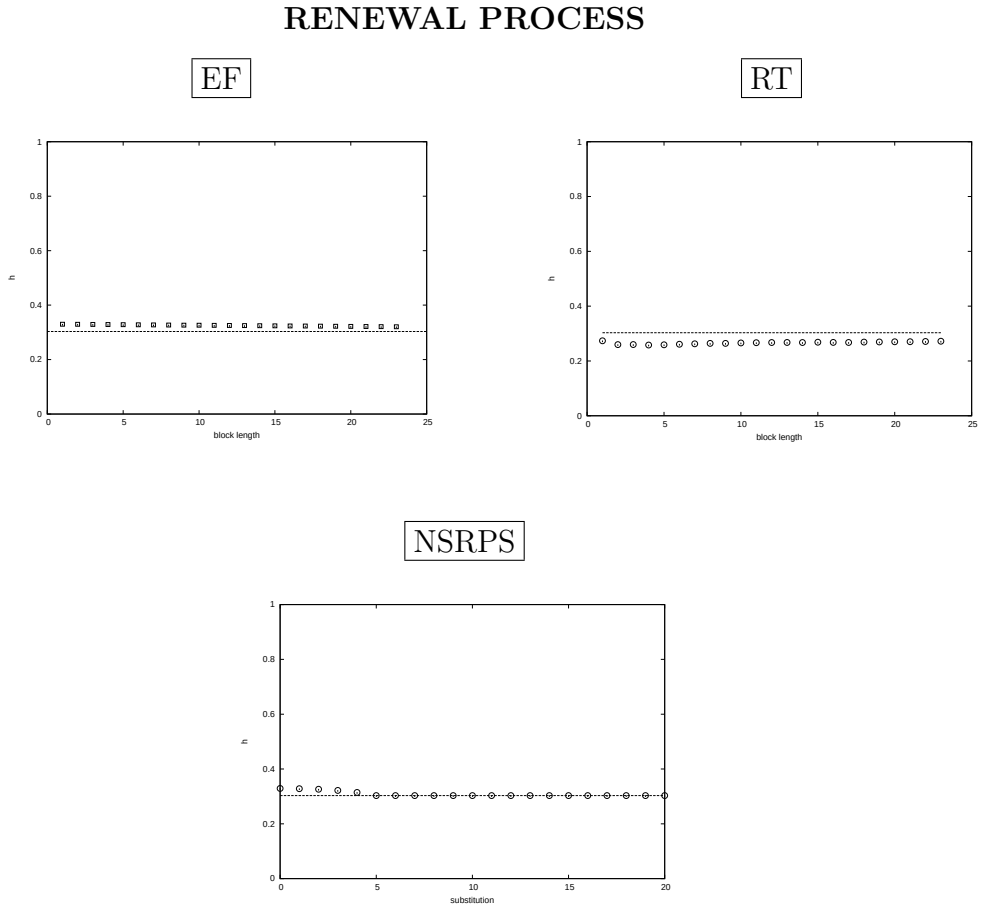

Figure 3.6: *Results for the renewal process RP_{32}: entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the entropy value.*

- How to prove an analogous of Theorem 3.2.1 for NSRPS giving a sufficient number of substitutions in function of the length of the string?
- Are there other meaningful substitution methods (different from the recipe given in Theorem 3.1.4) that may be proved to be (at least) sufficient for Theorem 3.1.3 to hold?
- Can the joint use of NSRPS and Lyapunov exponent (which are both fast converging and fastly computable) together with Theorem 3.4.1 give a particularly good method to numerically estimate the Hausdorff dimension of an attractor?
- Concerning the applications of NSRPS to non-artificial processes, such

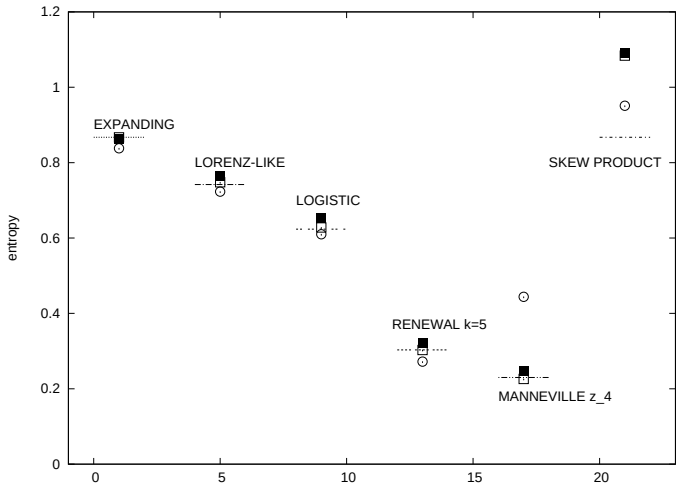

Figure 3.7: *Entropy estimates for the maps E, L, $\Lambda_{3.8}$, $M_{\frac{31}{16}}$, $\mathcal{S}$ and the renewal process RP_{32}: symbol $\square$ refers to the NSRPS value under* StopCond *condition; $\blacksquare$ refers to the EF value and $\circ$ refers to the RT method. Straight lines show the entropy values (for the maps they are the estimated Lyapunov exponents).*

as literary texts, biological sequences (DNA, proteins) and time series in general, what interesting features of the driving dynamics may be extracted?

- NSRPS method might be the core of some data compression algorithm (see [71]). This should pave the way to some investigations towards its compression capabilities in comparison with other well-known algorithms. We remark that data compression procedures have also been successfully used as entropy estimators (see e.g. [72] and [73]).

Chapter 4

Entropy of the AR(1) and MA(1) processes

In Chapter 3 we have studied the entropy of some nonlinear deterministic dynamical systems, while now we devote our attention to two linear stochastic dynamical systems.

In this chapter we study from an analytical point of view the Shannon entropy of suitably discretised versions of the AR(1) and MA(1) processes. We start with the usual definitions of these processes and then translate them from continuous to discrete *symbolic* phase state, by using two symbols to play the roles of negative real values and positive real values. This is a rather coarse discretisation, but it has the advantage of being by far the most simple one to treat analytically.

We eventually find analytical expressions for the entropies H_2, H_3, h_2, h_3. The techniques used to derive these results cannot be immediately extended to obtain analytical forms of Shannon entropies H_k and h_k of higher order, that is, with $k > 3$. Other tools, in addition to the ones we use, seem to be needed in order to derive further results.

The theoretical study reported in this chapter is motivated by the study of high-frequency financial time series of Chapter 7. In that context, the AR(1) and the MA(1) processes are often used to describe and model the time series of price returns of financial assets. The theoretical approach adopted in this chapter to finding the Shannon entropy of these processes and the results we find can help in empirical analyses where such processes are employed.

4.1 The AR(1) and MA(1) processes

The building blocks of the AR(1) and MA(1) processes (or, in general, of all ARMA(p, q) models) are random variables $\{\epsilon_t\}_{t\in\mathbb{Z}}$ constituting a *white noise process*, that is, a sequence of random variables such that the following three conditions hold:

(i) $\mathbb{E}[\epsilon_t] = 0$, for each t;

(ii) $\mathrm{Var}[\epsilon_t] = \sigma^2$;

(iii) $\mathrm{Cov}[\epsilon_s, \epsilon_t] = 0$, for all $s \neq t$.

This is the standard definition of white noise process and much theory on linear stochastic models is developed making no further assumption. However, to obtain analytical results we crucially use properties of a *Gaussian* white noise, which do not hold for general white noise processes. From now on, in this chapter, we indicate with $\{\epsilon_t\}_t$ a Gaussian white noise process, that is, we assume

$$\epsilon_t \text{ are i.i.d. } \mathcal{N}(0, \sigma^2). \tag{4.1}$$

4.1.1 The AR(1) process

Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a Gaussian white noise process as defined by (4.1). For $\phi \in \mathbb{R}$, the *autoregressive* process of order 1 is defined by[1]

$$X_t = \phi X_{t-1} + \epsilon_t. \tag{4.2}$$

In what follows we shall consider only values of the autoregressive parameter ϕ for which the AR(1) process defined by (4.2) is stationary, which happens for $|\phi| < 1$.

By recursive substitution, the random variable X_t may be expressed as a linear combination of infinite past white noise terms, that is,

$$X_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \ldots = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}. \tag{4.3}$$

This definition is referred to as the MA(∞) form of the AR(1) process.

[1] A more general definition would include a mean constant term to be added to the right-hand side of Equation (4.2). However, we do not intend to make a presentation of the most general AR(1) process and just define the process in the form we treat it.

From the stationarity of $\{X_t\}_t$ and the expressions (4.2) and (4.3) it follows that

$$\begin{cases} \mathbb{E}[X_t] & = 0 \\ \mathrm{Var}[X_t] & = \sum_{i=0}^{\infty} \mathbb{E}[\phi^{2i}\epsilon_{t-i}^2] = \sum_{i=0}^{\infty} \phi^{2i}\sigma^2 = \frac{\sigma^2}{1-\phi^2} \end{cases},$$

so that the unconditional distribution of X_t is $\mathcal{N}(0, \frac{\sigma^2}{1-\phi^2})$. If we were told the value of X_{t-1}, for some t, we would have

$$\begin{cases} \mathbb{E}_{t-1}[X_t] & = \phi X_{t-1} \\ \mathrm{Var}_{t-1}[X_t] & = \sigma^2 \end{cases},$$

where with $\mathbb{E}_{t-1}$ and Var_{t-1} we denote the conditional expectation and conditional variance $\mathbb{E}[\cdot|X_{t-1}]$ and $\mathrm{Var}[\cdot|X_{t-1}]$. The conditional distribution of X_t, given the knowledge of X_{t-1}, is thus $\mathcal{N}(\phi X_{t-1}, \sigma^2)$.

It is convenient to find also the conditional distribution of X_t, given the knowledge of X_{t-i}. This will be useful when calculating entropies in Section 4.2.2. By partial substitution we have $X_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \ldots + \phi^{i-1}\epsilon_{t-i+1} + \phi^i X_{t-i}$ and thus the conditional distribution is $\mathcal{N}\big(\phi^i X_{t-i}, \frac{1-\phi^{2i}}{1-\phi^2}\sigma^2\big)$.

4.1.2 The MA(1) process

Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a Gaussian white noise process as before. Given $\theta \in \mathbb{R}$, the *moving average* process of order 1 with parameter θ is defined by[2]

$$X_t = \epsilon_t + \theta\epsilon_{t-1}. \tag{4.4}$$

The MA(1) process is stationary for all values of the parameter $\theta \in \mathbb{R}$, and what we will say and the results we will obtain about this process hold for all possible values of θ. However, in the applications we are concerned only with values $|\theta| < 1$. For these values the MA(1) process is invertible and it can be expressed in an AR(∞) form, so that Equation (4.4) is equivalent to

$$X_t = \theta X_{t-1} - \theta^2 X_{t-2} + \theta^3 X_{t-3} - \ldots + \epsilon_t = \sum_{i=1}^{\infty} (-1)^{i+1}\theta^i X_{t-i} + \epsilon_t. \tag{4.5}$$

The random variable X_t has unconditional distribution $\mathcal{N}(0, (1+\theta^2)\sigma^2)$ and conditional distribution, given the knowledge of X_{t-1}, $\mathcal{N}(\frac{\theta}{1+\theta^2}X_{t-1}, (\frac{1+\theta^2+\theta^4}{1+\theta^2}\sigma^2))$.

[2]As in the case of the AR(1) process, this is not the most general definition of the MA(1) process, but it is the one we shall work with.

4.2 The Shannon entropy of the processes AR(1) and MA(1)

We recall from Definition 2.3.3 that, for a finite-state random process $\{S_t\}$ with measure μ (we shall refer to this object also as the *source* μ), the k-th order Shannon entropy is

$$H_k(\mu) = H(S_1^k) = -\sum_{s_1^k \in A^k} \mu(s_1^k) \log \mu(s_1^k), \tag{4.6}$$

where A is the finite alphabet of random variables S_t, and the k-th order conditional entropy is

$$h_k(\mu) = H_k(\mu) - H_{k-1}(\mu) = -\sum_{s_1^k \in A^k} \mu(s_1^k) \log \mu(s_k | s_1^{k-1}).$$

Moreover, we recall that the quantities $\frac{H_k(\mu)}{k}$ and $h_k(\mu)$ converge, as $k \to \infty$, to a common value $h(\mu)$, which is called the *entropy rate* of μ.

In order to talk about the Shannon entropy of the processes AR(1) and MA(1), whose phase space is continuous, we need some kind of discretisation. Among the infinitely many possible discretisations, we choose the simplest one which is not trivial. If $\{X_t\}_t$ is an AR(1) or an MA(1) process as defined in Sections 4.1.1 and 4.1.2, we define the binary symbolisation

$$s_t = B(X_t) = \begin{cases} 0 & \text{if } X_t < 0 \\ 1 & \text{if } X_t > 0 \end{cases} . \tag{4.7}$$

The symbolisation (4.7) is almost always defined, since the case $X_t = 0$ has obviously measure zero. In probabilistic terms, there would not be any difference between the given definition and one where the equality to zero is assigned to either symbol 0 or 1. The symbolisation (4.7) thus defines a binary process $\{s_t\}_t$, which will be the object we shall be studying throughout this section. This finite-state process has a measure μ inherited from and depending on the original process X_t. When we want to specify to which process we are referring to, we shall use the notations $H_k^{AR(1)}$, $H_k^{MA(1)}$, $h_k^{AR(1)}$, $h_k^{MA(1)}$, $h^{AR(1)}$, $h^{MA(1)}$.

To calculate the Shannon entropies of the discretised AR(1) and MA(1) processes, we exploit some properties of symmetry that they possess. We start with proving a result about the parity of the entropies as functions of the parameters ϕ and θ.

Proposition 4.2.1. *The entropy H_k is an even function of the parameter ϕ or θ, for all $k = 1, 2, \ldots$. Moreover, also h_k, for all $k = 1, 2, \ldots$, and h are even functions. In formulas, it holds*

(i) $H_k^{AR(1)}(\phi) = H_k^{AR(1)}(-\phi)$,

(ii) $H_k^{MA(1)}(\theta) = H_k^{MA(1)}(-\theta)$,

(iii) $h_k^{AR(1)}(\phi) = h_k^{AR(1)}(-\phi)$,

(iv) $h_k^{MA(1)}(\theta) = h_k^{MA(1)}(-\theta)$,

for all $k = 1, 2, \ldots$, *and*

(v) $h^{AR(1)}(\phi) = h^{AR(1)}(-\phi)$,

(vi) $h^{MA(1)}(\theta) = h^{MA(1)}(-\theta)$.

Proof. First note that, if $\{\epsilon_t\}_t$ is a Gaussian white noise, then also $\{\epsilon'_t\}_t = \{(-1)^t\epsilon_t\}_t$ is a Gaussian white noise and it is indeed the same process as $\{\epsilon_t\}_t$ since a Gaussian random variable ϵ_t has the same distribution as its opposite $-\epsilon_t$. The AR(1) process defined by $X'_t = -\phi X'_{t-1} + \epsilon'_t$ has the MA(∞) form (see Equation (4.3))

$$X'_t = \sum_{i=0}^{\infty}(-\phi)^i\epsilon'_{t-i} = \sum_{i=0}^{\infty}(-1)^i\phi^i(-1)^{t-i}\epsilon_{t-i} = \sum_{i=0}^{\infty}(-1)^t\phi^i\epsilon_{t-i}.$$

Thus we have $X'_t = (-1)^t X_t$, for all t. This relation between the two continuous-state processes $\{X'_t\}$ and $\{X_t\}$ translates into an analogous one for the binary processes $S' = \{s'_t\}_t$ and $S = \{s_t\}_t$ defined by discretisation as in (4.7). This means that a single realisation of the process $\{X_t\}_t$ (or, equivalently, of the process $\{\epsilon_t\}_t$) produces two binary sequences s and s' for which it holds $s_t = s'_t$ for even t and $s_t = -s'_t$ for odd t. We therefore have a bijective correspondence between realisations of $\{s_t\}$ and of $\{s'_t\}$ which also preserves the measure, that is, it holds

$$\mu_S(s_{t_1}^{t_k}) = \mu_{S'}(s_{t_1}^{t_k\prime}), \quad \text{for all } k \text{ and all } t_1 \leq \ldots \leq t_k. \tag{4.8}$$

The following diagram provides a picture of the process isomorphism:

$$\begin{array}{ccc} \{X_t\} & \xrightarrow{\ \prime\ } & \{X'_t\} \\ {\scriptstyle B}\big\downarrow & & {\scriptstyle B}\big\downarrow \\ \{s_t\} & \xrightarrow{\ \prime\ } & \{s'_t\} \end{array} .$$

From Equation (4.8) it follows that $H_k^{AR(1)}(\phi) = H_k^{AR(1)}(-\phi)$, which means that (i) is proved.

Equality (ii) is proved in the very same way as for (i), by noting that the MA(1) process defined by $X'_t = \epsilon'_t - \theta\epsilon'_{t-1}$, with $\epsilon'_t = (-1)^t\epsilon_t$ for all t, is isomorphic to that defined by $X_t = \epsilon_t - \theta\epsilon_{t-1}$.

Finally, (iii) and (v) follow immediately from (i), and (iv) and (vi) follow immediately from (ii). □

Proposition 4.2.1 tells us that it suffices to calculate the entropies of the AR(1) and the MA(1) processes only for $\phi \geq 0$ and for $\theta \geq 0$. We now state and prove a result which derives from the symmetry of the normal distribution.

Proposition 4.2.2. *Let μ be the measure of an AR(1) or an MA(1) process discretised as in (4.7). Let $s_1^k \in \{0,1\}^k$ be a binary string of length k and $\bar{s}_1^k$ its* complementary *string defined by $\bar{s}_i = 1 - s_i$, for each $i = 1, \ldots, k$. Then it holds $\mu(s) = \mu(\bar{s})$.*

Proof. If $\{\epsilon_t\}_t$ is a Gaussian white noise process defining the process $\{X_t\}_t$ (either AR(1) or MA(1)), the white noise $\bar{\epsilon} = \{-\epsilon_t\}_t$ defines the process $\bar{X} = \{-X_t\}_t$. This is actually isomorphic to the process X itself, since the random variables ϵ_t have the same distributions as their opposites. The processes S and $\bar{S}$, discretised versions of the processes X and $\bar{X}$, are therefore isomorphic and the thesis follows. □

Finally, the last property that we need is the time-reversibility of stationary Gaussian linear models, which is the content of the next theorem. We first give the formal definition of time-reversibility.

Definition 4.2.1. *A stationary process is time-reversible if, for every n and every $t_1, \ldots, t_n$, the vectors $\{X_{t_1}, \ldots, X_{t_n}\}$ and $\{X_{t_n}, \ldots, X_{t_1}\}$ have the same joint probability distribution.*

Theorem 4.2.3. *Stationary ARMA processes built from a Gaussian white noise are time-reversible.*

For the proof see [74]. What we are interested in is the following specification to the AR(1) and MA(1) cases.

Corollary 4.2.4. *Let μ be the measure of an AR(1) or an MA(1) process discretised as in (4.7). Then for every binary string $s_1 \ldots s_k \in \{0,1\}^k$ it holds $\mu(s_1 \ldots s_k) = \mu(s_k \ldots s_1)$.*

4.2.1 A geometric characterisation of the Shannon entropies $H_k^{AR(1)}$

In this section we give a general characterisation of the Shannon entropies $H_k^{AR(1)}$, for all $k = 1, 2, \ldots$, in terms of the entropy of some partition of the unit sphere $\mathbb{S}^{k-1} = \{\mathbf{x} \in \mathbb{R}^k \,|\, ||\mathbf{x}|| = 1\}$. In principle, the same path can be followed to obtain an analogous general characterisation for the entropies $H_k^{MA(1)}$. However, this does not seem to be feasible, since for the process MA(1) the general formulas for the conditional distributions of X_k, given $X_1, X_2, \ldots, X_{k-1}$, are not as simple as the ones for the process AR(1), which is Markov.

Let $s_1^k \in \{0,1\}^k$ be one of the 2^k binary strings of length k. According to the symbolisation (4.7), it corresponds to the event $\{X_1 \in I_1, \ldots, X_k \in I_k\}$, where $I_i = (-\infty, 0)$ if $s_i = 0$ and $I_i = (0, \infty)$ if $s_i = 1$. For the process AR(1) we have

$$\begin{cases} X_1 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\phi^2}\right) \\ X_2 \sim \mathcal{N}(\phi X_1, \sigma^2) \\ \quad \vdots \\ X_k \sim \mathcal{N}(\phi X_{k-1}, \sigma^2) \end{cases}$$

and therefore

$$\mu(s_1^k) = \int_{I_1} \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{1-\phi^2}}} e^{-\frac{1}{2}\left(\frac{X_1}{\frac{\sigma}{\sqrt{1-\phi^2}}}\right)^2} \int_{I_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{X_2-\phi X_1}{\sigma}\right)^2} \ldots$$
$$\ldots \int_{I_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{X_k-\phi X_{k-1}}{\sigma}\right)^2} \mathrm{d}X_k \ldots \mathrm{d}X_2 \, \mathrm{d}X_1. \quad (4.9)$$

Let us now consider the normalising linear transformation

$$\begin{cases} Y_1 = \frac{1}{\frac{\sigma}{\sqrt{1-\phi^2}}} X_1 \\ Y_2 = \frac{X_2-\phi X_1}{\sigma} \\ \quad \vdots \\ Y_k = \frac{X_k-\phi X_{k-1}}{\sigma} \end{cases}, \quad (4.10)$$

described in matrix form by $Y = A_\phi X$, with

$$A_\phi = \frac{1}{\sigma}\begin{pmatrix} \sqrt{1-\phi^2} & 0 & 0 & \dots & 0 & 0 \\ -\phi & 1 & 0 & \dots & 0 & 0 \\ 0 & -\phi & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\phi & 1 \end{pmatrix}.$$

The random variables Y_t are $\mathcal{N}(0,1)$ and Equation (4.9) can be written

$$\mu(s_1^k) = \int_{I'} \frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{1}{2}\left(Y_1^2+Y_2^2+\ldots+Y_k^2\right)} \, \mathrm{d}Y_1 \, \ldots \, \mathrm{d}Y_k, \tag{4.11}$$

where $I' = A_\phi(I_1 \times I_2 \times \ldots \times I_k)$. The integral in Equation (4.11) is equal to the fraction of k-dimensional solid angle determined by the cone I', or, equivalently, to the fraction of hypersphere $\frac{\lambda(I' \cap \mathbb{S}^{k-1})}{\lambda(\mathbb{S}^{k-1})}$, being λ the Lebesgue measure.

The 2^k solid angles of the form I', corresponding to the strings of k binary symbols, are those that result from sectioning the k-dimensional Euclidean space with the hyperplanes π_1, π_2, π_3, …, π_k of equations

$$\begin{aligned} \frac{\phi^{k-1}}{\sqrt{1-\phi^2}} x_1 &= 0 \\ \frac{\phi^{k-1}}{\sqrt{1-\phi^2}} x_1 + \phi^{k-2} x_2 &= 0 \\ \frac{\phi^{k-1}}{\sqrt{1-\phi^2}} x_1 + \phi^{k-2} x_2 + \phi^{k-3} x_3 &= 0 \\ \vdots & \\ \frac{\phi^{k-1}}{\sqrt{1-\phi^2}} x_1 + \phi^{k-2} x_2 + \phi^{k-3} x_3 + \ldots + \phi x_{k-1} + x_k &= 0. \end{aligned}$$

The problem of calculating the measures $\mu(s_1^k)$ of Equation (4.6) has thus been translated into a purely geometric problem: calculating the solid angles in $\mathbb{R}^k$ cut by the hyperplanes π_i, $i = 1, \ldots, k$. The entropy of Equation (4.6) is thus nothing else that the entropy of the partition of $\mathbb{S}^{k-1}$ determined by the hyperplanes π_i.

4.2.2 Calculating the entropies H_k

At least for the AR(1) process, the characterisation given in Section 4.2.1 is very general. Though it can be exploited to calculate the entropies through the calculation of the solid angles for $k = 2, 3$, there seem to exist no general formula for calculating the solid angles in $\mathbb{R}^k$ determined by k hyperplanes, for $k \geq 4$. However, the symmetries of the problem, presented in Proposition 4.2.2 and Corollary 4.2.4, allow to considerably reduce the number of degrees of freedom of the problem even when $k \geq 4$. We now find the entropies H_k, for $k = 1, 2, 3$, both for the AR(1) and the MA(1) processes. Thanks to Proposition 4.2.1, we can restrict our attention to values of the parameters $\phi \geq 0$ and $\theta \geq 0$.

k = 1 When $k = 1$ we must find the measures $\mu(0)$ and $\mu(1)$. Since the random variables of the process (either the AR(1) or the MA(1)) have a symmetrical distribution we simply have $\mu(0) = \mu(1) = \frac{1}{2}$. Therefore, for the 1st order entropy we have $H_1 = -\mu(0)\log_2\mu(0) - \mu(1)\log_2\mu(1) = 1$.

k = 2 We deal with the case $k = 2$ by proving a general result, which is very useful also for the case $k = 3$. To establish some notation, if $\mathbf{a} = a_1^l$ and $\mathbf{b} = b_1^m$ are two finite binary strings, let us denote by $\mathbf{a} \cdot^i \mathbf{b}$ the cylinder set defined by $\{S_1^l = a_1^l\} \cap \{S_{l+i+1}^{l+i+m} = b_1^m\}$. We prove the following two propositions for the processes AR(1) and MA(1).

Proposition 4.2.5. *Let μ be the measure of the discretised AR(1) process. Then it holds*

$$\mu(0 \cdot^i 0) = \frac{1}{2\pi}\arccos(-\phi^{i+1}), \tag{4.12}$$

$$\mu(0 \cdot^i 1) = \frac{1}{2\pi}\arccos(\phi^{i+1}). \tag{4.13}$$

Proposition 4.2.6. *Let μ be the measure of the discretised MA(1) process. Then it holds*

$$\mu(00) = \frac{1}{2\pi}\arccos\left(-\frac{\theta}{1+\theta^2}\right), \tag{4.14}$$

$$\mu(01) = \frac{1}{2\pi}\arccos\left(\frac{\theta}{1+\theta^2}\right), \tag{4.15}$$

and

$$\mu(\mathbf{s_1} \cdot^i \mathbf{s_2}) = \mu(\mathbf{s_1})\mu(\mathbf{s_2}), \tag{4.16}$$

for $i \geq 1$.

Note that these results suffice for calculating the entropies H_2, since we just take $i = 0$ in Proposition 4.2.5 and, furthermore, by Proposition 4.2.2, for both processes AR(1) and MA(1) we have $\mu(10) = \mu(01)$ and $\mu(11) = \mu(00)$.

Proof of Proposition 4.2.5. The quantities $\mu(0 \cdot^i 0)$ and $\mu(0 \cdot^i 1)$ are the probabilities of the events $\{X_1 < 0\} \cap \{X_{i+2} < 0\}$ and $\{X_1 < 0\} \cap \{X_{i+2} > 0\}$, respectively. Recall from Section 4.1.1 that $X_{i+2}|X_1 \sim \mathcal{N}\big(\phi^{i+1}X_1, \frac{1-\phi^{2(i+1)}}{1-\phi^2}\sigma^2\big)$. Thus, proceeding as in Section 4.2.1, we are left with calculating the measures of the subsets of $\mathbb{S}^1$ cut by the lines in $\mathbb{R}^2$ given by equations $x_1 = 0$ and $\frac{\phi^{i+1}}{\sqrt{1-\phi^{2(i+1)}}}x_1 + x_2 = 0$. Equalities (4.12) and (4.13) follow immediately. □

Proof of Proposition 4.2.6. Just as in Proposition 4.2.5, $\mu(00)$ and $\mu(01)$ are the probabilities of the events $\{X_1 < 0\} \cap \{X_2 < 0\}$ and $\{X_1 < 0\} \cap \{X_2 > 0\}$, respectively. Since the conditional distribution of $X_2|X_1$ is $\mathcal{N}(\frac{\theta}{1+\theta^2}X_1, (\frac{1+\theta^2+\theta^4}{1+\theta^2}\sigma^2))$, we have that $\mu(00)$ and $\mu(01)$ are the relative measures of the subsets of $\mathbb{S}^1$ cut by the lines of equations $x_1 = 0$ and $\frac{\theta}{\sqrt{1+\theta^2+\theta^4}}x_1 + x_2 = 0$. Expressions (4.14) and (4.15) follow straightforwardly.

Finally, equality (4.16) is easily proved by noting that the conditional distribution of a random variable X_t, given X_{t-i} with $i \geq 2$, is the same as its unconditional distribution because X_t and X_{t-i} $(i \geq 2)$ are independent. □

For the AR(1) process we thus have

$$H_2^{AR(1)}(\mu) = -2\mu(00)\log_2\mu(00) - 2\mu(01)\log_2\mu(01),$$

where

$$\mu(00) = \frac{1}{2\pi}\arccos(-\phi),$$
$$\mu(01) = \frac{1}{2\pi}\arccos(\phi).$$

For the MA(1) process we have

$$H_2^{MA(1)}(\mu) = -2\mu(00)\log_2\mu(00) - 2\mu(01)\log_2\mu(01),$$

with $\mu(00)$ and $\mu(01)$ given by Equations (4.14) and (4.15).

k = 3 We could find the quantities $\mu(s_1s_2s_3)$, with $s_i \in \{0, 1\}$ for $i = 1, 2, 3$, by using the formula to calculate the solid angles in $\mathbb{R}^3$ cut by the hyperplanes of which we have the equations. However, it is much simpler and much more instructive to exploit the symmetry properties affirmed by Proposition 4.2.2 and Corollary 4.2.4. This method can be applied also to

strings of length greater than 3, and though in this cases it does not seem sufficient to completely determine all the measures $\mu(\mathbf{s})$, still it allows to obtain partial results that we do not detail. Since no general formula is known for the solid angle in the k-dimensional Euclidean space with $k > 3$, using the symmetry properties of the measures of the strings is the only way that we have found to partially explore higher-dimensional settings.

Initially, we let μ indicate the measure of either process, AR(1) or MA(1). By Proposition 4.2.2 we have $\mu(100) = \mu(011)$, $\mu(101) = \mu(010)$, $\mu(110) = \mu(001)$, $\mu(111) = \mu(000)$. Furthermore, by Corollary 4.2.4 we also have $\mu(001) = \mu(100)$. The symmetries thus reduce the number of unknown quantities from $2^3 = 8$ to three. Now note that we have the following three independent relations:

$$\begin{aligned} &\mu(000) + \mu(001) = \mu(00), \\ &\mu(010) + \mu(011) = \mu(01), \\ &\mu(000) + \mu(010) = \mu(0 \cdot 0). \end{aligned} \tag{4.17}$$

Since $\mu(00)$, $\mu(01)$, $\mu(0{\cdot}0)$ are determined for the AR(1) and MA(1) processes in Propositions 4.2.5 and 4.2.6, we can solve the system (4.17).

For the AR(1) process we finally have

$$H_3^{AR(1)}(\mu) = -2\mu(000)\log_2\mu(000) - 4\mu(001)\log_2\mu(001) - 2\mu(010)\log_2\mu(010),$$

where

$$\begin{aligned} \mu(000) &= \frac{1}{2\pi}\arccos(-\phi) - \frac{1}{4\pi}\arccos(\phi^2), \\ \mu(001) &= \frac{1}{4\pi}\arccos(\phi^2), \\ \mu(010) &= \frac{1}{2\pi}\arccos(\phi) - \frac{1}{4\pi}\arccos(\phi^2). \end{aligned}$$

For the MA(1) process we have

$$H_3^{MA(1)}(\mu) = -2\mu(000)\log_2\mu(000) - 4\mu(001)\log_2\mu(001) - 2\mu(010)\log_2\mu(010),$$

where

$$\begin{aligned} \mu(000) &= \frac{1}{2\pi}\arccos\left(-\frac{\theta}{1+\theta^2}\right) - \frac{1}{8}, \\ \mu(001) &= \frac{1}{8}, \\ \mu(010) &= \frac{1}{2\pi}\arccos\left(\frac{\theta}{1+\theta^2}\right) - \frac{1}{8}. \end{aligned}$$

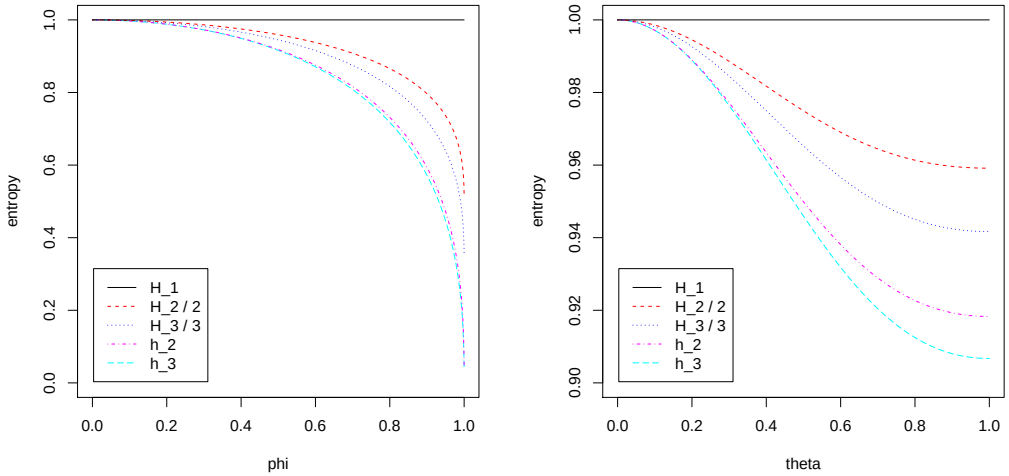

Figure 4.1: Theoretical Shannon entropies H_1, $\frac{H_2}{2}$, $\frac{H_3}{3}$, together with conditional entropies $h_2 = H_2 - H_1$, $h_3 = H_3 - H_2$, of the AR(1) (left) and MA(1) (right) processes, as functions of the autoregressive parameter $0 \leq \phi < 1$ and of the moving average parameter $0 \leq \theta < 1$, respectively.

In Figure 4.1 we graph the theoretical entropies H_1, $\frac{H_2}{2}$, $\frac{H_3}{3}$ and the conditional entropies $h_2 = H_2 - H_1$, $h_3 = H_3 - H_2$ as calculated above, for positive values of the autoregressive parameter ϕ and the moving average parameter θ. It is interesting to note that in the AR(1) case the conditional entropies h_2 and h_3 go to 0 as ϕ approaches 1, while in the MA(1) process they converge to values of about 0.918 and 0.907, respectively.

Chapter 5

Gambling, portfolio theory and market efficiency

At first sight, information theory seems to be unrelated to gambling and portfolio theory. But in fact there is strong duality between the growth rate of investment in a horse race and the entropy rate of the horse race. Indeed, the sum of the growth rate and the entropy rate is a constant. The horse race can be seen as a special case of investment in the stock market and the connection extends to a duality between the growth rate of wealth in the stock market and the entropy rate of the market.

5.1 Gambling

Assume that m horses run in a race, with horse i having probability p_i of winning the race. If horse i wins, the payoff is o_i for 1 (that is, an investment of 1 dollar on horse i results in o_i dollars if horse i wins and 0 dollars if horse i loses). An equivalent way of expressing an a-for-1 odd is saying that it is b-to-1 with $b = a - 1$. For example, fair odds on a coin flip are 2-for-1 or 1-to-1 (*even* odds). Assume that a gambler distributes his entire wealth across the horses, with b_i denoting the fraction of his wealth invested in horse i. Thus of course $b_i \geq 0$ for all i and $\sum b_i = 1$. If the race is won by horse i, the gambler will have multiplied his wealth by a factor $b_i o_i$.

The wealth at the end of the race is a random variable, and the gambler wishes to "maximise" its value. Assume that the gambler repeatedly reinvest his money in a sequence of races. Then, after n races, the gambler's wealth is

$$S_n = \prod_{i=1}^{n} S(X_i), \tag{5.1}$$

where $S(X) = b(X)o(X)$ is the factor by which the gambler's wealth is multiplied when horse X wins.

Definition 5.1.1. *The* doubling rate *of a horse race is*

$$W(\mathbf{b}, \mathbf{p}) = \mathbb{E}(\log S(X)) = \sum_{k=1}^{m} p_k \log(b_k o_k).$$

Theorem 5.1.1. *Let the race winners $X_1, \ldots, X_n$ be i.i.d. $p(x)$. Then the wealth of the gambler using betting strategy $\mathbf{b}$ grows exponentially at rate $W(\mathbf{b}, \mathbf{p})$, that is,*

$$\frac{1}{n} \log S_n \to W(\mathbf{b}, \mathbf{p}).$$

The proof is an easy application of the weak law of large numbers to the i.i.d. random variables $\log S(X_i)$, $i = 1, 2, \ldots$.

Definition 5.1.2. *The* optimum doubling rate $W^*(\mathbf{p})$ *is the maximum doubling rate over all choices of the portfolio* $\mathbf{b}$*:*

$$W^*(\mathbf{p}) = \max_{\mathbf{b}} W(\mathbf{b}, \mathbf{p}) = \max_{\mathbf{b}:\, b_i \geq 0, \sum b_i = 1} \sum_{i=1}^{m} p_i \log(b_i o_i).$$

The maximisation of $W(\mathbf{b}, \mathbf{p})$ as a function of $\mathbf{b}$, subject to the constraint $\sum b_i = 1$, is performed by studying the functional

$$J(\mathrm{b}) = \sum p_i \log(b_i o_i) + \lambda \sum b_i,$$

where λ is a Lagrange multiplier. One finds that $\mathbf{b} = \mathbf{p}$ is a stationary point of the function $J(\mathbf{b})$ with $\lambda = -1$. That it is actually a maximum is the content of the next theorem.

Theorem 5.1.2 (Proportional gambling is log-optimal)**.** *The optimum doubling rate is given by*

$$W^*(\mathbf{p}) = \sum p_i \log o_i - H(\mathbf{p})$$

and is achieved by the proportional gambling scheme $\mathbf{b}^* = \mathbf{p}$.

Proof. Rewriting the function $W(\mathbf{b}, \mathbf{p})$ in a suitable way, it holds

$$\begin{aligned} W(\mathbf{b}, \mathbf{p}) &= \sum p_i \log \left(\frac{b_i}{p_i} p_i o_i \right) \\ &= \sum p_i \log o_i - H(\mathbf{p}) + \sum p_i \log(\frac{b_i}{p_i}) \\ &\leq \sum p_i \log o_i - H(\mathbf{p}), \end{aligned}$$

with equality if and only if $\mathbf{b} = \mathbf{p}$. □

Proportional gambling (that is, betting on each horse in proportion to its probability of winning) is known as *Kelly gambling* ([16]).

A special case is when the odds are m-for-1 on each horse, in which case the odds are fair with respect to the uniform distribution and the optimum doubling rate is given by the following corollary.

Corollary 5.1.3 (Conservation theorem)**.** *For uniform fair odds, the sum of the doubling rate and the entropy rate is a constant. In particular, it holds*

$$W^*(\mathbf{p}) + H(\mathbf{p}) = \log m.$$

Corollary 5.1.3 expresses the duality between the doubling rate and the entropy rate of a horse race. Every bit of entropy decrease doubles the gambler's wealth. Low entropy races are the most profitable.

5.1.1 Dependent horse races and entropy rate

Assume now that many races take place and that there is some dependence among them, so that the gambler can use the results of previous races to determine his strategy for the next one. The sequence $\{X_k\}$ of horse race winners forms a stochastic process and the optimal doubling rate for uniform fair odds is

$$\begin{aligned} W^*(X_k|X_{k-1},\ldots,X_1) &= \mathbb{E}\left[\max_{\mathbf{b}(\cdot|X_{k-1},\ldots,X_1)} \mathbb{E}[\log S(X_k)|X_{k-1},\ldots,X_1]\right] \\ &= \log m - H(X_k|X_{k-1},\ldots,X_1), \end{aligned}$$

which is achieved by $b^*(x_k|x_{k-1},\ldots,x_1) = p(x_k|x_{k-1},\ldots,x_1)$.

After n races, the gambler's wealth is as in (5.1) and the exponent in the growth rate (assuming m-for-1 odds) is

$$\begin{aligned} \frac{1}{n}\mathbb{E}[\log S_n] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\log S(X_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} (\log m - H(X_i|X_{i-1},\ldots,X_1)) \\ &= \log m - \frac{H(X_1,\ldots,X_n)}{n}. \end{aligned} \tag{5.2}$$

For a sequence of races forming a stationary process $\{X_i\}$ with entropy rate $h(\{X_i\})$, taking the limit for $n \to \infty$ in (5.2) yields $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\log(S_n)] + h(\{X\}) = \log m$, which is again the result that the sum of the doubling rate and the entropy rate is a constant.

5.2 Information theory and portfolio theory

A stock market may be represented as a vector of stocks $\mathbf{X} = (X_1, \ldots, X_m)$, with $X_i \geq 0$, $i = 1, \ldots, m$, where X_i is the ratio between the price at the end of the day and the price at the beginning of the day. For example, $X_i = 1.03$ means that the price of stock i went up 3 percent that day. A *portfolio* $\mathbf{b} = (b_1, \ldots, b_m)$, with $b_i \geq 0$, $\sum b_i = 1$, is an allocation of wealth across the stocks, where b_i is the fraction of the total wealth invested in stock i. The ratio of the wealth at the end of the day to the wealth at the beginning of the day is $S = \mathbf{b}^T\mathbf{X} = \sum_{i=1}^{m} b_i X_i$. The goal is to maximise S in some sense.

Definition 5.2.1. *The* growth rate *of a stock market portfolio* $\mathbf{b}$ *with respect to a stock market* $\mathbf{X} \sim F(\mathbf{x})$ *is defined as*

$$W(\mathbf{b}, F) = \int \log(\mathbf{b}^T\mathbf{x})\, \mathrm{d}F(\mathbf{x}) = \mathbb{E}[\log(\mathbf{b}^T\mathbf{X})].$$

If the logarithm is to base 2, the growth rate is also called the doubling rate.

Definition 5.2.2. *The* optimal growth rate $W^*(F)$ *is*

$$W^*(F) = \max_{\mathbf{b}} W(\mathbf{b}, F), \tag{5.3}$$

where the maximum is over all possible portfolios $\mathbf{b}$*, with* $b_i \geq 0$*,* $\sum_i b_i = 1$*.*

A portfolio that achieves the maximum in Equation 5.3 is called a log-optimal portfolio *or* growth optimal portfolio*.*

Theorem 5.2.1. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be i.i.d.* $\sim F(\mathbf{x})$*. Let*

$$S_n^* = \prod_{i=1}^{n} \mathbf{b}^{*T}\mathbf{X}_i$$

be the wealth after n *days using the constant rebalanced portfolio* $\mathbf{b}^*$*. Then*

$$\frac{1}{n}\log S_n^* \to W^* \quad \textit{with probability } 1.$$

The proof easily follows from the strong law of large numbers.

As a counterpart to Kelly proportional gambling, where one invests in proportions that remain unchanged in expected value after the investment period, in the context of portfolio theory it can be proved (see [1], Theorem 16.2.1 and consequences) that the expected proportion of wealth in each stock under the log-optimal portfolio is unchanged from day to day. In formulas, being $\mathbf{b}^*$ the allocation of wealth at the beginning of the day, the proportion of wealth at the end of the day is $\frac{b_i^* X_i}{\mathbf{b}^{*T}\mathbf{X}}$, whose expected value is b_i^*.

5.2.1 Duality between the growth rate and the entropy rate of a market

We now want to show a strong formal analogy between the growth rate and the entropy rate of a stochastic market process. Let $\{\mathbf{X}_n\} = (\mathbf{X}_1, \mathbf{X}_2, \ldots)$ be a vector-valued stochastic process, with $\mathbf{X}_i \geq 0$, representing the most general time-dependent market process. We also consider investment strategies that depend on the past values of the market in a causal fashion.

Definition 5.2.3. *A* causal *portfolio strategy is a sequence of mappings* $b_i : \mathbb{R}^{m(i-1)} \to \mathcal{B} = \{\mathbf{b} \in \mathbb{R}^m \,|\, \mathbf{b}_i \geq 0, \sum_{i=1}^m \mathbf{b}_i = 1\}$*, with the interpretation that portfolio* $b_i(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$ *is used on day* i*.*

Let

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i^T(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}\mathbf{X}_i).$$

The objective is to maximise $\mathbb{E}[\log S_n]$ over all causal portfolio strategies $\{\mathbf{b}_i(\cdot)\}$. It holds

$$\begin{aligned}\max_{\mathbf{b}_1,\ldots,\mathbf{b}_n} \mathbb{E}[\log S_n] &= \sum_{i=1}^{n} \max_{\mathbf{b}_i(\mathbf{X}_1,\ldots,\mathbf{X}_{i-1})} \mathbb{E}[\log(\mathbf{b}_i^T\mathbf{X}_i)] \\ &= \sum_{i=1}^{n} \mathbb{E}[\log(\mathbf{b}_i^{*T}\mathbf{X}_i)],\end{aligned}$$

where $\mathbf{b}_i^*$ is the log-optimal portfolio for the conditional distribution of $\mathbf{X}_i$ given the past values of the stock market. In other words, $\mathbf{b}_i^*(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$ is the portfolio that achieves the conditional maximum, which is denoted by

$$\begin{aligned}\max_{\mathbf{b}} \mathbb{E}[\log(\mathbf{b}^T\mathbf{X}_i)|(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}) &= (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})] \\ &= W^*(\mathbf{X}_i|\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}).\end{aligned}$$

Taking the expectation over the past, we get the conditional optimal growth rate

$$W^*(\mathbf{X}_i|\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}) = \mathbb{E}\big[\max_{\mathbf{b}} \mathbb{E}[\log(\mathbf{b}^T\mathbf{X}_i)|\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}]\big],$$

where the maximum is over all portfolio-valued functions $\mathbf{b}$ defined on $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$. Thus, the highest expected log return is achieved by using the conditional log-optimal portfolio at each step. Let

$$W^*(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \max_{\mathbf{b}_1,\ldots,\mathbf{b}_n} \mathbb{E}[\log S_n],$$

where the maximum is over all causal portfolio strategies. Then, since $\log S_n^* = \sum_{i=1}^n \log \mathbf{b}_i^{*T} \mathbf{X}_i$, we have the following chain rule for W^*:

$$W^*(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \sum_{i=1}^{n} W^*(\mathbf{X}_i | \mathbf{X}_1, \ldots, \mathbf{X}_{i-1}).$$

This chain rule is formally the same as the chain rule for the entropy H (see Section 2.3.2). In some way, W can be seen as the dual of H. In particular, conditioning reduces H but increases W. We now extend the formal analogy between W and H by defining the counterpart of the entropy rate for time-dependent stochastic processes.

Definition 5.2.4. *The* growth rate W_∞^* *is defined as*

$$W_\infty^* = \lim_{n \to \infty} \frac{W^*(\mathbf{X}_1, \ldots, \mathbf{X}_n)}{n} \tag{5.4}$$

if the limit exists.

Theorem 5.2.2. *For a stationary market* $(\mathbf{X}_1, \mathbf{X}_2, \ldots)$*, the growth rate exists and is equal to*

$$w_\infty^* = \lim_{n \to \infty} W^*(\mathbf{X}_n | \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}). \tag{5.5}$$

The existence of the limit in Equation (5.4) and the equality (5.5), expressing the growth rate in the conditional form, are proved in the same way as for the entropy rate.

We end this section by noting that the horse race treated in Section 5.1 is a special case of a stock market, in which there are m stocks corresponding to the m horses in the race. At the end of the race, the value of the stock for horse i is either 0 or o_i, the value of the odds for horse i. Thus, X is nonzero only in the component corresponding to the winning horse. In this case, the log-optimal portfolio is proportional betting (known as Kelly gambling, see Section 5.1), and in the case of uniform fair odds (that is, $o_i = m$, for all i) it holds

$$W^* = \log m - H(X).$$

In the case of a sequence of correlated horse races, the optimal portfolio is conditional proportional betting and the asymptotic growth rate is

$$W_\infty^* = \log m - h(\{X\}),$$

where $h(\{X\}) = \lim \frac{1}{n} H(X_1, \ldots, X_n)$ if the limit exists.

5.3 Market efficiency

A market is said to be *informationally efficient* if it is efficient in processing the information. In an efficient market observed prices express a correct evaluation of all available information.

5.3.1 A brief history

The concept of market efficiency was first anticipated by Bachelier in 1900, who writes in his PhD dissertation that "past, present and even discounted future events are reflected in market price, but often show no apparent relation to price changes." [20]. Modern literature on market informational efficiency begins however in the 1960s, with the independent works of Samuelson and Fama. The idea that asset prices fully reflect all available information (the so-called *Efficient Market Hypothesis* (EMH) was introduced in modern economics by Samuelson, with his article [21]: *Proof that properly anticipated prices fluctuate randomly.* Samuelson observed that "[i]f one could be sure that a price would rise, it would have already risen." It was Fama, though, the first to introduce the term 'efficient market' [22]. In his seminal papers [23, 24, 22, 25] he was concerned with the old debate between *technical analysis* (the use of patterns in historical data to forecast future prices) and *fundamental analysis* (the use of accounting and economic data to establish the assets' fair prices).

5.3.2 The random walk hypothesis

Market efficiency implies the random character of prices (the more efficient the market, the more random the sequence of price changes). A perfectly efficient market is one in which price changes are completely random and unpredictable. This ideal situation is the result of many market participants attempting to profit from their information. When investors try to exploit even the smallest informational advantages at their disposal, their actions have the effect of incorporating the information into market prices, so that in the end the profit opportunities that first motivated the trades are quickly eliminated. In an idealised frictionless market with costless trading this occurs instantaneously and prices always fully reflect all available information. In mathematical terms, prices follow martingales.

A statistical description of the unforecastability of price changes is provided by the so-called random walk hypothesis (RWH), suggesting a model in which prices evolve in a purely random manner according to some specifications. Much of the literature about market efficiency revolved around

the RWH and many tests were developed in order to see if it could provide a satisfying model for market data. One of the first tests of the RWH was developed in [75], where the frequencies of same-sign and opposite-sign consecutive returns were compared. A number of works testing data for the RWH were published from the beginning of the 1960s to the beginning of the 1990s, with general results supporting the RWH for daily data and only occasional results of departure from randomness and rejection of the RWH. See [26] for a survey on the topic.

5.3.3 Efficient Market Hypothesis: the *weak*, the *semi-strong* and the *strong* form

Fama proposed three forms of the efficient market hypothesis, corresponding to different information sets available to market agents. The *weak form* says that prices fully reflect the information contained in past prices. The *semi-strong form* maintains that prices fully reflect all publicly available information. The *strong form* asserts that market prices fully reflect any information, public and private.

Fama [25] summarises the early random walk literature, his own contributions and other studies about the information contained in the historical price time series and concludes that the results strongly support the weak form of the EMH. Studies of the semi-strong form of the EMH are tests about the speed of adjustment of prices to new information, such as stock splits and earnings announcements. When this announcements happen, the market appears to anticipate the information with most of the price adjustment occurring before the news is released and a remaining rapid and accurate adjustment happening shortly after. The evidence accumulated in the 1960s and 1970s shows that, while markets cannot be completely efficient in the strong form, there is convincing support for the weak and semi-strong forms. Departures from the EMH are commonly explained as consequences of over- or underreaction to new information by the investors, when prices are temporarily pushed beyond the fair or rational market value and subsequently adjusted with the formation of patterns.

5.3.4 Refutability of the EMH

The possibility of testing data for efficiency remains however somewhat elusive. Fama also demonstrated that the notion of market efficiency could not be rejected without an accompanying rejection of the model assumed for normal returns (e.g., the price formation mechanism). He emphasises that

the hypothesis of market efficiency must be tested in the context of expected returns and it thus depends on the benchmark choice. The so-called *joint hypothesis problem* states that when a model yields a return significantly different from the actual return, one can not be certain if the market is inefficient or if there exists an imperfection in the model. Researchers can only modify their models by adding different factors to eliminate the anomalies, in the hope of fully explaining the return within the model. Therefore the notion of market efficiency is not a well-posed and empirically refutable hypothesis. To make it operational, one must specify additional structure, e.g., investors' preferences, information structure, etc. But then a test of market efficiency becomes a test of several auxiliary hypotheses as well, and a rejection of such a joint hypothesis tells us little about which aspect of the joint hypothesis is inconsistent with the data.

5.3.5 Relative efficiency

As pointed out by Campbell, Lo, MacKinlay [27], perfect informational efficiency is an idealised condition which is never met in real markets. Inefficiencies are always present in real markets and therefore there is little point in investigating the efficiency of markets in absolute terms. Rather than the all-or-nothing view taken by much of the traditional literature on market efficiency, it would be much more interesting to study the notion of *relative* efficiency, that is, the degree of efficiency of one market measured against another. The idealised perfectly efficient market may be regarded as a useful benchmark for measuring relative efficiency.

We stress that this is the point of view we take in this work. We shall investigate to what degree assets depart from the idealised perfect efficiency, ranking them according to relative efficiency.

5.4 Literature review on measuring market efficiency

In this section we review some recent studies that deal with *measuring* the informational efficiency of financial markets. Market efficiency is not a matter that is to be studied only in absolute terms. Markets which are not efficient may display different levels of inefficiency. The presented studies provide different instruments to quantify the efficiency level of a market.

5.4.1 Measuring market efficiency with the Lempel-Ziv compression index

Relative efficiency of financial markets was investigated by Giglio and coauthors in the work [28]. The authors particularly stress the fact that the efficient market is an idealised concept that is unattainable, but that serves as a benchmark for measuring relative efficiency. They make use of the algorithmic complexity theory, which provides a connection between the efficient market hypothesis and the unpredictable character of asset returns, because a time series that has a large amount of non-redundant information (such as that of the idealised efficient market) exhibits statistical features that are almost indistinguishable from those observed in a random time series. Measurements of the deviation from randomness determined in terms of algorithmic complexity provide a tool to assess the relative efficiency of a financial market. The more algorithmically complex is a time series from a given market, the more efficient that market.

The algorithmic complexity of a string is defined (in Kolmogorov's formulation) as the length of the shortest computer program that can produce the string. A practical problem of this definition is that determining the shortest program is not computable. In 1976, Lempel and Ziv propose in [76] a measure of the complexity of a string that does not rely on the shortest algorithm. In 1987, Kaspar and Schuster (see [77]) provide an easily calculable measure of the Lempel-Ziv index, which we now briefly explain.

The goal is to build a given string, by making a series of elementary operations consisting in either inserting a new digit or copying the digit from a word contained in the part of the string that has been already constructed. Assume that a given string $S = s_1 \dots s_n$ has been reconstructed by the program up to the digit s_r and that s_r has been newly inserted. In order to check whether the remaining substring $s_{r+1} \dots s_n$ can be reconstructed by simple copying operations or whether new digits must be inserted, one proceeds as follows. One first checks whether s_{r+1} is in the *vocabulary* of $s_1 \dots s_r$, that is, the set of *words* of the substring $s_1 \dots s_r$ in which it is divided by the points where newly inserted digits are. If this is the case, one checks whether $s_{r+1}s_{r+2}$ is contained in the vocabulary of $s_1 \dots s_{r+1}$ and so on until $s_{r+1} \dots s_{r+k}$ becomes so large that it can no longer be obtained by copying a word from the vocabulary of $s_1 \dots s_{r+k-1}$ and one has to insert a new digit. The Lempel-Ziv complexity index of the string S is defined as the number c of the steps required to construct S, that is, the number of the newly inserted digits (plus one, if the last copy step is not followed by inserting a digit).

Let us show how the Lempel-Ziv index is calculated for some simple

strings. For the string made up of only zeros, $S = 0000000000$, one inserts the first digit and then proceeds with copying previously occurred words. Thus, $c = 2$. For the string $T = 0101010101$, one inserts the first digit, then inserts the second digit, and then proceeds with the copying operation until the end of the strings. In this case, $c = 3$. For the string $U = 0110001001$, one inserts the first digit (0), then inserts the second digit (1), then further inserts 10 and 001, and finally copies 001. Thus, in this case, $c = 5$. We note that these complexity indices reflect the intuition that the string S is the less complex of the three and the string U the most complex one.

It has been shown in [76] that for almost all random strings the complexity index of the initial n-character string, $c(n)$, asymptotically grows when $n \to \infty$ as $\rho(n) = \frac{n}{\log_2 n}$, that is, it holds

$$\lim_{n\to\infty} \frac{c(n)}{\rho(n)} = 1.$$

One may thus compute a finite normalised complexity index $LZ = \frac{c(n)}{\rho(n)}$ to get the complexity of a string relative to that of a random one. Note however that the LZ index is not an absolute measure of complexity, nor is it ranged between zero and one.

The dataset studied by Giglio and coauthors is made up of seven years of daily data from July 2000 to July 2007, for a total amount of 2000 observations, for 36 stock exchange indices and 20 US Dollar exchange rates. Analyses are performed with the returns of the raw series, which are coded as ternary strings as follows. Assuming a stability basin b around zero, the return time series r_t is coded into the symbolic sequence s_t defined by

$$s_t = \begin{cases} 0 & \text{if } r_t \leq -b \\ 2 & \text{if } -b < r_t < b \\ 1 & \text{if } r_t \geq b \end{cases} . \tag{5.6}$$

Following Shmilovici *et al.* [29], the authors take $b = 0.0025$ and argue that some experiments on changing the value of b seem to show that results do not alter too much. Nonetheless, they point out that more research is needed to consider a more sophisticated analysis in the choice of b.

Given a symbolic sequence of length 2000, the authors consider sliding time windows, each of width 1000, calculate the LZ index for every window and get the average value. If a time series is autocorrelated, then it will present more patterns and will have lower complexity than an independent one. Their measure of relative efficiency is the proportion of windows on which the LZ index is above 1. For a pseudo-random sequence they find

that the $LZ = 1$ threshold is surpassed 98.8% of the time. This leads to saying that the sequence is 98.8% efficient.

Results of this analysis on the 36 stock exchange indices lead to ranking them according to efficiency. For instance, the S&P 500 index (USA) is found to be the most efficient, with an average score of 99.1%, followed by the DAX 30 (Germany) and the Nikkei 225 (Japan) with scores of 98.4% and 98.2%, respectively. The other indices with an efficiency score above 90% are the All Ordinaries (Australia, 97.8%), the ATX (Austria, 97.4%), the Dow Jones (USA, 95.4%), the Korea Composite (South Korea, 94.9%), the Tel Aviv 100 (Israel, 92.9%), the Hang Seng (Hong Kong, 91.5%), the Straits Times (Singapore, 90.3%). In the last five positions the following indices are ranked: Shanghai Composite (China, 49.5%), Philippines (Philippines, 43.1%), Lima General (Peru, 37.9%), Karachi 100 (Pakistan, 23.7%), Colombo SE (Sri Lanka, 10.5%).

Among the US Dollar exchange rates, the first six in the ranking all show an efficiency score above 99%. They are the exchange rates with the UK Pound Sterling, the Swedish Krona, the Norwegian Krone, the Euro, the New Zealand Dollar, the Swiss Franc. On the other end of the ranking, the five exchange rates showing the minimum efficiency scores are the Indian Rupee (43.54%), the Colombian Peso (21.98%), the Taiwan New Dollar (21.17%), the Chinese Yuan (17.94%), the Sri Lanka Rupee (11.84%).

A general consideration on the efficiency scores and the rankings is that the indices and the exchange rates of developed markets are top ranked, showing a great amount of informational efficiency. On the other hand, the last positions of the rankings are occupied by indices and exchange rates of emerging markets, which show the greatest amount of inefficiency.

We remark that we share Giglio and coauthors' idea of measuring relative efficiency, instead of studying efficiency in absolute terms. This is precisely the point of view we take in the study presented in Chapter 7. We also share with them the approach of coding return series into symbolic ones, in order to study the presence of patterns and inefficiencies. However, we use a different tool to measure efficiency and, what is more important, we investigate high-frequency data rather than daily data, so that the amount of information at our disposal is much larger. We also argue that the redundancy of symbolised return series may be greatly affected by the long-memory properties of the volatility. A ternary symbolisation of returns like the one performed by Giglio and coauthors, which has a stability basin for the central symbol fixed in an absolute way, with no adjustment to take volatility dynamics into account, can give rise to spurious inefficiencies. Searching for true inefficiencies, all factors that are known to possess autocorrelation and memory properties must be filtered out from the time series one works with.

5.4.2 Measuring market efficiency with a Variable Order Markov model

Interesting studies on the weak form of the Efficient Market Hypothesis have been conducted by Shmilovici with other authors ([29], [30]), who make use of a Variable Order Markov model to detect the recurring patterns in return time series. The return series are symbolised and then checked for compressibility above random. The authors also study the predictability of the symbolic series and the profitability of some trading strategies suggested by high predictability. The starting point is the consideration that in an efficient market compression of the time series is not possible, because there are no patters and the complexity of the series is high.

In [29] the authors use Rissanen's context tree algorithm [78], a universal coding method, for measuring the complexity of a time series. In [30] a slightly modified version of the algorithm is used, to smooth observed empirical probabilities in order to account also for events of zero frequency. However, the two versions of the algorithm are conceptually the same and we briefly explain how it works in its original formulation.

The VOM model represents a collection of statistically significant patterns in symbolic sequences by a parsed tree. Each node in the tree contains the conditional distribution of symbols, given a pattern that is represented by the branch (path) from the root to the node. The branches are not necessarily equal in length, what makes the algorithm particularly effective for predicting short series. Let A be a finite alphabet of symbols. For example, $A = \{0, 1, 2\}$, representing time series behaviour of *decrease*, *stability* and *increase*. Given a finite sequence $\sigma_1^n = \sigma_1 \dots \sigma_n$, with $\sigma_i \in A$, the VOM algorithm generates a conditional probability distribution $\hat{P}(\sigma|s)$ for a symbol $\sigma \in A$ given a context $s \in A^*$, where A^* is the set of finite A-sequences of any length. VOM models attempt to estimate conditional distributions of the form $\hat{P}(\sigma|s)$, where the context length $|s|$ depends on the available statistics.

The VOM learning algorithm includes three phases: counting, smoothing, context selection. In the counting phase, an initial context tree is built with a maximal depth (that is, the contexts' length) D. The tree has a root node, from which the branches are developed. A branch from the root to a node represents a context that appears in the training set in a reversed order. Each node has at most $|A|$ children and contains $|A|$ counters of symbols $N_\sigma(s)$, $\sigma \in A$, given the context s. The tree is constructed as follows. The symbolic sequence is parsed one symbol at a time and, for every symbol σ_i, its D-sized context σ_{i-D}^{i-1} defines a path in the tree, which is constructed if it does not yet exist. The counts of σ_i at each node of the path are incremented by one. The second phase of the VOM construction is to use the counts $N_\sigma(s)$ for

generating the predictor $\hat{P}(\sigma|s)$. The following equation is used to smooth the probability to account also for events of zero frequency:

$$\hat{P}(\sigma|s) = \frac{\frac{1}{2} + N_\sigma(s)}{\frac{|A|}{2} + \sum_{\sigma' \in A} N_{\sigma'}(s)}.$$

In the third phase, the size of the tree is reduced by recursively eliminating the leaves that do not contribute significant additional information with respect to their parent node. We refer the reader to the cited articles for the details.

Once the VOM tree is constructed, it is used to predict the next symbol in a series, looking for the one that maximises the likelihood $\hat{P}(\sigma|s)$, given the context s.

The time series data used in study [29] are daily returns for 13 national stock market indices and for the 25 stocks of the TA25 index of Tel Aviv Stock Exchange. Returns are discretised to binary (positive or negative) or ternary (positive, stable, negative) series, where *stable* means a change smaller than 0.25%. In study [30], the authors use high-frequency data: tick-by-tick bid prices of 12 currency pairs for the year 2002 are sampled at frequencies of 1, 5, 10, 15, 20, 25, 30 minutes. The difference series are quantised to ternary-symbols series, such that an increase (respectively, decrease) of 3 pips (i.e. 0.0003) or more is coded by symbol '1' (respectively, '3') and any change smaller than 2 pips is coded by the stability symbol '2'.

Since temporal patterns can be of short duration, a VOM model is constructed for each temporal (sliding) window of 50, 75, 100 symbols. For each fixed window length, final results are averages over all the sliding windows. As long as the compressibility of the series is concerned, a compression coefficient is computed for each sequence, by dividing the stochastic complexity (as measured by the context-tree algorithm) by the number of bits needed for the original representation of the sequence (e.g. 75 bits for a binary sequence of length 75). Then a compression threshold γ is fixed and a null hypothesis that the sequence be random is rejected if the number of windows showing compressibility higher than γ is greater than the number that is expected for a random sequence, at 90% confidence levels. Values of γ considered are 0.86, 0.88, 0.9.

Results of the analysis in [29] reveal that significant compressibility is detected in ten out of the thirteen international stock index series. The authors stress that it is mostly for small volume markets, like the Brazilian, the Swiss and the Belgian, that the Efficient Market Hypothesis is rejected. By contrast, indices such as the AEX (Netherlands) and the HSI (Hong Kong) demonstrate no statistically significant compression. The authors also emphasise that, since stock indices are a weighing of stocks, inefficiencies in

single stocks may be masked in the process. This is why they test each individual stock in the TA25 index. It is found that 8% of the stocks are incompressible for all the sliding windows lengths and all the compression threshold checked; 36% of the stocks are compressible only for some sliding windows lengths and compression thresholds; 56% of the stocks are compressible for all the sliding windows lengths and all the compression thresholds. In the overall picture, the Tel Aviv Stock Exchange is found to be efficient most of the time (about 83%-96% of the trading days), but with periods of inefficiency for some stocks.

We want to attract the attention of the reader on two further results that the authors find and report without much commenting. The first one is that windows of lengths 50 and 75 demonstrate higher compressibility than those of length 100, indicating a short-term effect of 2-3 months in length. The second one is that the percentage of windows that are compressible beyond random is higher for the ternary series representation. We argue that both these effects can be due to the memory properties of the volatility, which the authors do not take into account in handling data. A ternary discretisation with fixed thresholds incorporates some predictability in periods of high or low volatility, even under the null assumption of no correlation among the returns, because during such periods returns have different distributions. Any discretisation of returns into more than two symbols must take volatility issues into account.

In [30], the analysis of the compressibility of the series shows that the Efficient Market Hypothesis is rejected in 248 experiments out of 252 (a different experiment is conducted for each combination of the 12 currency pairs, the 3 window lengths and the 7 sampling frequencies). An analysis of predictability of the series similarly shows that higher frequency series are more predictable that lower frequency series, confirming that the presence of patterns is higher at higher frequencies. Finally, the Efficient Market Hypothesis is tested as the absence of profitable opportunities, according to the definition of efficiency given by Jensen. Taking this point of view, predictable patterns invalidate the Efficient Market Hypothesis only if they produce excess returns that are consistently large enough to cover for transaction costs that are associated with trading operations. The authors develop and simulate some trading strategies to investigate this issue. Results show that the strategies generate too many transactions, leading to net losses in almost all experiments. Only in very few cases the strategies result in a net profit, that is however very small. In conclusion, theoretical market's inefficiencies are largely present, but they are not *practical* inefficiencies, since they are not exploitable to make profit. In other words, statistically significant anomalies are not economically significant.

We remark that the analyses of series compressibility conducted in [29] and [30] would fit well in the study of relative efficiency. Although the authors are only interested in rejecting or not rejecting the null hypothesis of randomness, their procedure leads to a concrete quantification of the randomness of the series, which could be used to rank assets by efficiency. We also emphasise that the 2009 study is the first, to the best of our knowledge, to approach the relative efficiency problem using high-frequency data. However, we believe that the results may be significantly affected by the fact that the ternary symbolisations employed in the two studies ignore the predictable character of seasonalities (for high-frequency data) and of volatility. Especially when searching for redundancies in time series, they must be taken into account and filtered out in order to avoid spurious predictability. When we deal with ternary discretisations of high-frequency returns in Chapter 7, we will pay great attention to the removal of intraday patterns and long-term volatility.

5.4.3 Measuring market efficiency with the Shannon entropy

Relative efficiency is studied by Risso in [13], where the quantification of informational efficiency is a means to compare the emerging markets with the developed markets. Efficiency is once again measured as closeness of symbolised time series to perfect randomness.

The evaluation of randomness is pursued by means of the Shannon entropy of binarised time series. In particular, given a sequence of returns $\{r_t\}$, a symbolic sequence $\{s_t\}$ is defined by

$$s_t = \begin{cases} 0 & \text{if } r_t < \bar{r} \\ 1 & \text{if } r_t > \bar{r} \end{cases},$$

where $\bar{r}$ is the average return. The entropy of words of length k is calculated as

$$H_k = -\frac{1}{N_0} \sum_{|\mathbf{s}|=k} \mu(\mathbf{s}) \log_2 \mu(\mathbf{s}), \tag{5.7}$$

where $\mu(\mathbf{s})$ is the observed frequency of word $\mathbf{s}$ and N_0 is the number of observed sequences with non-zero frequency. The latter is a punishment in the entropy for frequencies equal to zero. It turns out that the values of entropies H_k, for $k = 1, 2, \ldots$, calculated according to definition 5.7, reach a minimum value in correspondence of some $\bar{k}$. In the empirical analyses, typical values of $\bar{k}$ are around 10. The value $H_{\bar{k}}$ is taken as measure of the efficiency of the series.

The dataset is made up of daily data of stock market indices for 20 different countries, covering the period from 1st July 1997 to 14th December 2007. The twenty countries are ranked by value of relative efficiency. Results show that three Asian markets take the first positions as the most efficient. The stock markets of Taiwan (market index: TSEC), Japan (NIKKEI) and Singapore (Straits Times) are indeed important financial centres in the world. In the last four positions as the most inefficient markets there are three ex-socialist countries (Czech Republic, Russia and Slovenia) and the only African stock market (Egypt). The last results seem to support the hypothesis that the emerging stock markets are more inefficient than the developed markets. In particular, the ex-socialist countries have not achieved levels of efficiency similar to the more developed European markets. A clear difference in the efficiency level is found between Western Europe markets (UK, Germany, Austria, Holland, Switzerland) and Eastern Europe markets (Czech Republic, Russia, Slovenia).

We remark once more that ranking different assets by informational efficiency will also be the goal of our approach. In our study of market efficiency in Chapter 7, we shall use precisely the Shannon entropy as a means for measuring departure of symbolised return series from randomness, even though our entropy estimators are different. These similarities make the study [13] the closest to ours, as long as the goal and the basic means for measuring the presence of patterns in symbolic time series are concerned. However, we will deal with high-frequency data and this will be a major enhancement for the study of market efficiency.

5.4.4 Measuring market efficiency with the Hurst exponent and the Approximate Entropy

We briefly mention two other ways that have been followed in literature to measure relative efficiency. References are [31] and [12].

In [31] the authors stress that the analyses that focus on single measures of efficiency should be looked with care. They employ a rolling windows approach to show that efficiency seems to evolve over time. They use the median Hurst exponent (we refer to the paper for the details) as measure of efficiency to compare the relative efficiency of different markets. The empirical analysis is conducted on the indices of 13 equity markets (11 emerging markets, plus the US and Japan, which are included for comparison purposes), for which daily closing prices from January 1992 through December 2002 are used. Logarithmic returns $\ln \frac{p_t}{p_{t-1}}$ are calculated and corrected to eliminate short-range dependencies. This is performed by estimating the

Hurst exponent on AR(1) and GARCH(1,1) adjusted time series instead of using the true observed time series. Countries are ranked using median Hurst exponent values measures for efficiency. Results show that inefficiencies are more significant for Asian countries than for Latin American countries, while US and Japan rank as the most efficient ones. The authors also find a positive relationship between market capitalisation and efficiency and an inverse relationship between trading costs and their measures of market efficiency, which suggest that market microstructure variables may play a role in efficiency results.

In [12] the authors use the *Approximate Entropy* (ApEn) to quantify the randomness in the time series and to study the market efficiency of the global foreign exchange markets. The data used are volatility-normalised return series of daily foreign exchange rates for 17 countries from 1984 to 1998 and from 1999 to 2004, before and after the Asian crisis, respectively. The Approximate Entropy of a time series $\mathbf{x} = (x_1, \ldots, x_N)$ is defined by

$$\mathrm{ApEn}(\mathbf{x}, m, r) = \Phi^m(\mathbf{x}, r) - \Phi^{m+1}(\mathbf{x}, r),$$

where m is the embedding dimension and r the tolerance in similarity. The function $\Phi^m(\mathbf{x}, r)$ is given by

$$\Phi^m(\mathbf{x}, r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln \frac{B_i(\mathbf{x}, r)}{N-m+1},$$

where $B_i(\mathbf{x}, r)$ is the number of pairs (i, j) for which $\max_{k=1,\ldots,m}(|x_{i+k-1} - x_{j+k-1}) \leq r$. The ApEn value compares the relative magnitude between repeated pattern occurrences for the embedding dimensions m and $m+1$. The ApEn is large when the time series data have a high degree of randomness and is small for the time series with a low degree of randomness, which makes it a suitable tool to measure market efficiency. Parameter values considered in [12] are $m = 2$ and r equal to the 20% of the standard deviation of the time series. Values of Approximate Entropy on different time series are used to rank the corresponding markets by efficiency. Results show higher market efficiency for European and North American foreign exchange market than for African and Asian ones except Japan.

Chapter 6

High frequency data

When dealing with financial time series, high frequency data indicate datasets of asset prices recorded more often than daily. The more prices a day, the higher is the frequency of the observations. In this chapter we describe the data used for our analyses, as well as all the preliminary procedures conducted in order to have clean data to work with. Procedures and techniques explained in Section 6.2 constitute an essential part in the handling of high-frequency data, which are somewhat more difficult to treat than low-frequency data, for they carry, along with more information, more errors and more features of seasonal behaviour.

High-frequency data have many advantages but they also present some challenges. On the plus side, the additional price observations allow us to learn more about how prices react to information. Moreover, it is more natural to investigate the presence of inefficiencies, recurring patterns and extreme events at a high frequency level, since they tend to compensate at a daily level. More observations also enable us to estimate and forecast volatility more accurately. On the other hand, they are much noisier than low-frequency data, since microstructure effects (such as the spread between buying and selling prices) become more important and in fact play a major role. They also present typical intraday patterns in trading behaviour that have to be taken into account and properly modelled. Finally, there is the problem that high-frequency data are not completely regular, since they may lack some observations in periods of low liquidity.

6.1 Stylised facts for intraday returns

Intraday returns have stylised facts that are similar to but distinct from those of daily returns. We briefly present the most important ones, which

are about (i) the distribution of intraday returns, (ii) the autocorrelations of intraday returns, (iii) the autocorrelations of intraday absolute returns, (iv) the intraday volatility pattern. We refer the reader to [79] for a thorough presentation.

(i) The distribution of intraday returns is leptokurtic and more so than for daily returns. Intraday returns have a fat-tailed distribution, whose kurtosis increases as the frequency of price observations increases. The distributions of high-frequency returns often show a sharp spike at zero, which reflects the feasible set of discrete prices.

(ii) More dependence might be anticipated in intraday returns than in daily returns for two reasons. First, bidask bounce in transaction prices will show most clearly at higher frequencies. Second, the exploitation of any dependence is more difficult when expected profits per trade decline as data frequency increases, but costs do not. The magnitude of observed dependence is, however, often very small. It is a stylised fact that intraday returns from traded assets are almost uncorrelated, with any important dependence usually restricted to a negative correlation between consecutive returns.

(iii) There is substantial positive dependence among intraday absolute returns, which occurs at many low lags and also among returns separated by an integer number of days. There is a periodic behaviour of the intraday volatility, visible in the autocorrelations of absolute returns.

(iv) The volatility is usually highest around the opening and the closing of the market. The average level of volatility depends on the time of day, with a significant intraday variation. There are short bursts of high volatility in intraday prices that follow major macroeconomic announcements.

6.2 Cleaning and whitening

6.2.1 Outliers

Anomalous values that may be present in recorded data, for example because of errors of transmission, are often referred to as outliers. To detect and remove them from high-frequency data, we use the cleaning algorithm proposed by Brownlees and Gallo in [80]. The algorithm was developed to identify the price records at tick-by-tick level which are too distant from a mean value

calculated in their neighbourhood. More precisely, a price observation p_i is regarded as anomalous and removed if

$$|p_i - \bar{p}_i(k)| \geq c\, s_i(k) + \gamma, \tag{6.1}$$

where $\bar{p}_i(k)$ and $s_i(k)$ denote respectively the δ-trimmed sample mean and sample standard deviation of the closest k observations around i, c is a constant which amplifies the standard deviation and γ is a granularity parameter useful in the case of k equal prices producing a zero variance. Although the algorithm was originally developed for tick-by-tick prices, we use it also with 1-minute data.

6.2.2 Stock splits

A *forward stock split* (or, simply, a *stock split*) is an operation consisting in an increase of the number of shares of a company and in a simultaneous adjustment of the price so that the market capitalisation of the company remains the same. A *reverse stock split* (also called a *stock merge*) is the opposite operation, leading to a decrease of the number of shares and to a corresponding increase of the price. A split is *m-for-n* if m new shares are released for every n old ones, with a simultaneous price change from p to $\frac{n}{m}p$. Splits must be identified and removed before any analysis on price returns, since they appear as huge returns and leaving them would greatly compromise the results.

In handling the data for the analyses of Chapters 7 and 8, we identify stock splits in correspondence of price returns for which it holds

$$|r| > 0.2.$$

This would detect, for example, a 3-for-2 split or a 4-for-5 merge.

6.2.3 Intraday patterns

We filter out the intraday volatility pattern from returns by means of a simple model with intraday volatility factors. Returns at intraday time t are rescaled by a factor ζ_t, which is calculated as the average, over all days, of adjusted absolute returns at time t. More precisely, if $R_{d,t}$ is the raw return of day d and intraday time t, we define the rescaled return

$$\tilde{R}_{d,t} = \frac{R_{d,t}}{\zeta_t}, \tag{6.2}$$

where

$$\zeta_t = \frac{1}{N_{\text{days}}} \sum_{d'} \frac{|R_{d',t}|}{s_{d'}}, \tag{6.3}$$

with N_{days} indicating the number of days in the sample and $s_{d'}$ the standard deviation of absolute intraday returns of day d'.

Throughout this thesis we shall refer to the rescaled returns $\tilde{R}$ defined by Equation (6.2) as *deseasonalised returns*.

6.2.4 Heteroskedasticity

To remove the heteroskedasticity from the (deseasonalised) return series, we estimate the time series of local volatility σ_t and define the *standardised* returns by

$$r_t = \frac{\tilde{R}_t}{\sigma_t}. \tag{6.4}$$

The series of standardised returns are the arrival point of the preliminary data-cleaning procedures and represent the main object that we shall study.

For estimating the local volatility, we use the *realised absolute variation* and the *realised bipower variation* (see [81, 82]). Let the logarithmic prices $p(t)$ be generated by a process

$$\mathrm{d}p(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \tag{6.5}$$

with $\mu(t)$ a finite variation process, $\sigma(t)$ a càdlàg volatility process, $W(t)$ a standard Brownian motion. Let the interval $[0, t]$ be divided into subintervals of the same length δ and denote by r_i the return $p(i\delta) - p((i-1)\delta)$. Then the following probability limits hold for the *realised absolute variation* and the *realised bipower variation*:

$$\begin{aligned} \mathrm{p}-\lim_{\delta \searrow 0} \delta^{\frac{1}{2}} \sum_{i=1}^{\lfloor t/\delta \rfloor} |r_i| &= \mu_1 \int_0^t \sigma(s)\,\mathrm{d}s, \\ \mathrm{p}-\lim_{\delta \searrow 0} \sum_{i=1}^{\lfloor t/\delta \rfloor - 1} |r_i||r_{i+1}| &= \mu_1^2 \int_0^t \sigma^2(s)\,\mathrm{d}s, \end{aligned} \tag{6.6}$$

where $\mu_1 = \mathbb{E}(|u|) = \sqrt{\frac{2}{\pi}} \simeq 0.797885$, $u \sim \mathcal{N}(0, 1)$.

Our estimators of local volatility based on these quantities are defined by the exponentially weighted moving averages

$$\hat{\sigma}_{\text{abs},t} = \mu_1^{-1} \alpha \sum_{i>0} (1-\alpha)^{i-1} |r_{t-i}| \tag{6.7}$$

and

$$\hat{\sigma}^2_{\mathrm{bv},t} = \mu_1^{-2}\alpha \sum_{i>0}(1-\alpha)^{i-1}|r_{t-i}||r_{t-i-1}|, \tag{6.8}$$

where α is the parameter of the exponential averages to be specified.

Chapter 7

The informational efficiency of financial markets

Throughout this chapter we estimate entropies by using the empirical distributions of finite strings in symbolic samples, as defined by Equation (3.1). Following Section 3.2, we shall be concerned with the estimation of the entropies H_k and h_k, with $k = 1, 2, \ldots, \log_2 n$, where n is the length of the series. For values of k greater than $\log_2 n$ the statistics provided by the series is too poor and the entropies H_k are underestimated. For the sake of uniformity in presenting the results, we shall actually choose few values for the order k, typically 2, 3, 6, 10.

As we shall see in Section 7.2 (Table 7.3), there will be cases where the different symbols of the alphabet appear in the series with significantly different frequencies. Since what we want to measure are the correlations among consecutive symbols, we want to filter out the difference in the frequencies of single symbols. Put another way, we want to measure the degree of randomness of the series, given that symbols appear in the series with the observed frequencies. To this aim, the entropies we shall calculate are

$$\tilde{H}_k = \frac{H_k}{H_1} \quad \text{and} \quad \tilde{h}_k = \frac{h_k}{H_1}. \tag{7.1}$$

Since the Shannon entropy deals with finite-alphabet information sources, a symbolisation of the returns is needed before being able to perform any analysis of entropy estimation. To be precise, we must say that of course return values are already discrete, since prices move on a discrete grid. However, what we intend to study by means of the Shannon entropy is the degree of randomness in the time sequence of *few* coarsely identified behaviours of the price. Indeed, we shall be interested only in symbolisations into 2 or 3

symbols, each representing a notable behaviour, such as “the price goes up”, “the price is stationary”, “the price goes down”.

7.1 ETF data

For our study we took high-frequency historical data of 55 Exchange Traded Funds traded at the New York Stock Exchange. In Section 7.1.1 we briefly explain what Exchange Traded Funds are, while in Section 7.1.2 we describe the data used in this study.

7.1.1 Exchange Traded Funds

An Exchange Traded Fund (ETF) is an investment whose performance is based on an index (or, less commonly, on other underlying assets). For example, the OIH is an oil ETF that follows the OSX oil index. It is a mini-portfolio of oil industry equities. Therefore, if the OSX goes up, generally so does the OIH. Purchasing shares of an ETF is like purchasing shares of all the equities that constitute the corresponding index, but with the advantage of purchasing just one product instead of many, each in the right proportion. An ETF is in fact a mini-portfolio, not a basket of stocks. Like equities, ETFs are traded on an exchange and during market hours. They can also be sold short or on margin. The goal of an ETF is to mimic its corresponding index and to yield the same return on investment. Exceptions to this rule are inverse ETFs and leveraged ETFs.

An *inverse* exchange traded fund is created by using various assets and derivatives (like options) in order to create profits when the underlying index declines in value. It is basically an index ETF that gains value when the correlating index falls. For example, the Short DOW 30 ETF profits when the Dow Jones Industrial Average index goes down.

Leveraged ETFs aim at outperforming the index or commodity they track. A leveraged ETF wants to provide 2-3 times the return of the correlating asset. So if the tracked index rises 1%, a 2x leveraged ETF wants to create a 2% return on investment. Leveraged ETFs are designed to include the securities in the underlying index, but also include derivatives of the securities and the index itself. These derivatives include, but are not limited to, options, forward contracts, swaps and futures.

There are also inverse leveraged ETFs, which offer multiple positive return if an index declines in value. They work the same as normal inverse ETFs, they are just designed for multiple returns.

7.1.2 Dataset

In Table 7.1 we present the list of the ETFs studied, along with the asset tracked by each one.

The ETFs studied include market ETFs, country ETFs, commodity ETFs and industry ETFs. The Select Sector SPDRs are ETFs that divide the S&P 500 into nine sectors.

The data used in this study cover a period of about forty months, from the 13th July 2006 to the 1st December 2009. We use closing prices at the sampling frequency of 1 minute, and a resampling of the same data at a 5-minute frequency. In the former case we have the advantage of using the greatest possible amount of price data that is available to us, but this goes to the detriment of the regularity of the price series, since the 1-minute series present a number of missing observations depending on the level of liquidity. In the latter case, instead, we use less information but the series are more regular. So, the choice of the frequency to work with is a tradeoff between amount of information and regularity of the data. From the point of view of information theory, the best is of course to use as much data as possible, and this is what we shall predominantly do in the analyses. In order to assess to what extent results depend on the chosen frequency or on the regularity of the series, in some cases we shall perform the analyses on the 5-minute data as well. In Table 7.2 we report the number of available price observations for the 55 ETFs. The number of days in the data sample is 854 for the vast majority of the ETFs, with six exceptions: 853 for EWU and RTH, 851 for EWW and MDY, 850 for RKH, 849 for EFA.

7.1.3 Data cleaning

We now outline the steps of the data cleaning procedure and establish some terminology. We then detail below on the single steps.

Starting from 1-minute closing prices, we first remove outliers (see Section 6.2.1) and then calculate logarithmic returns

$$R_t = \ln \frac{p_t}{p_{t-1}}.$$

We then search for possible stock splits (see Section 6.2.2), in order to remove the huge returns in correspondence of them. Throughout this chapter, we shall be referring to returns cleaned for possible splits, but not yet processed in any other way, as *raw returns*. Then, the procedure of filtering out the daily seasonalities, by the removal of the intraday pattern as explained in Section 6.2.3, leads to what we refer to as *deseasonalised returns*. Finally,

ticker	name ETF	provider[a]	asset tracked	lev./inv.
DIA	DIAMONDS Trust Series 1	St. Str. Gl. Adv.	Dow Jones Industrial Average Index	
DXD	ProShares UltraShort Dow30	ProShares	Dow Jones Industrial Average Index	-2x
EEM	iShares MSCI Emerging Markets Index Fund	iShares	MSCI Emerging Markets Index	
EFA	iShares MSCI EAFE Index	iShares	MSCI EAFE Index	
EWA	iShares MSCI Australia Index	iShares	MSCI Australia Index	
EWG	iShares MSCI Germany Index	iShares	MSCI Germany Index	
EWH	iShares MSCI Hong Kong Index	iShares	MSCI Hong Kong Index	
EWJ	iShares MSCI Japan Index	iShares	MSCI Japan Index	
EWM	iShares MSCI Malaysia Index	iShares	MSCI Malaysia Index	
EWS	iShares MSCI Singapore Index	iShares	MSCI Singapore Index	
EWT	iShares MSCI Taiwan Index	iShares	MSCI Taiwan Index	
EWU	iShares MSCI United Kingdom Index	iShares	MSCI United Kingdom Index	
EWW	iShares MSCI Mexico Investable Market Index	iShares	MSCI Mexico Investable Market Index	
EWY	iShares MSCI South Korea Index	iShares	MSCI South Korea Index	
EWZ	iShares MSCI Brazil Index	iShares	MSCI Brazil Index	
FXI	iShares FTSE China 25 Index	iShares	FTSE China 25 Index	
GDX	Market Vectors Gold Miners	Van Eck	AMEX Gold Miners Index	
GLD	SPDR Gold Shares	St. Str. Gl. Mkts.	Gold Bullion	
IBB	iShares Nasdaq Biotechnology Index	iShares	NASDAQ Biotechnology Index	
ICF	iShares Cohen & Steers Realty Majors Index	iShares	Cohen & Steers Realty Majors Index	
IJH	iShares S&P MidCap 400 Index	iShares	S&P MidCap 400 Index	
IJR	iShares S&P SmallCap 600 Index	iShares	S&P SmallCap 600 Index	
IVE	iShares S&P 500 Value Index	iShares	S&P 500 Value Index	
IVV	iShares S&P 500 Index	iShares	S&P 500 Index	
IVW	iShares S&P 500 Growth Index	iShares	S&P 500 Growth Index	
IWD	iShares Russell 1000 Value Index	iShares	Russell 1000 Value Index	
IWF	iShares Russell 1000 Growth Index	iShares	Russell 1000 Growth Index	
IWM	iShares Russell 2000 Index	iShares	Russell 2000 Index	
IWN	iShares Russell 2000 Value Index	iShares	Russell 2000 Value Index	
IWO	iShares Russell 2000 Growth Index	iShares	Russell 2000 Growth Index	
IYR	iShares Dow Jones U.S. Real Estate Index	iShares	Dow Jones U.S. Real Estate Index	
MDY	SPDR S&P MidCap 400	St. Str. Gl. Adv.	S&P MidCap 400 Index	
MZZ	ProShares UltraShort MidCap400	ProShares	S&P MidCap 400 Index	-2x
PHO	PowerShares Water Resources	PowerShares	NASDAQ OMX US Water Index	
QID	ProShares UltraShort QQQ	ProShares	NASDAQ-100 Index	-2x
QLD	ProShares Ultra QQQ	ProShares	NASDAQ-100 Index	2x
QQQQ	PowerShares QQQ	PowerShares	NASDAQ-100 Index	
RKH	Market Vectors Bank and Brokerage	Van Eck	Market Vectors US Listed Bank and Brokerage 25 Index	
RTH	Market Vectors Retail	Van Eck	Market Vectors US Listed Retail 25 Index	
SDS	ProShares UltraShort S&P500	ProShares	S&P 500 Index	-2x
SLV	iShares Silver Trust	iShares	price of Silver	
SPY	SPDR S&P 500	St. Str. Gl. Adv.	S&P 500 Index	
SSO	ProShares Ultra S&P500	ProShares	S&P 500 Index	2x
TIP	iShares Barclays TIPS Bond	iShares	Barclays U.S. Treasury Inflation Protected Securities Index (Series-L)	
USO	United States Oil	United States Commodity Funds LLP	price of West Texas Intermediate light, sweet crude oil	
VWO	Vanguard MSCI Emerging Markets	Vanguard	MSCI Emerging Markets Index	
XHB	SPDR S&P Home Builders	St. Str. Gl. Adv.	S&P Homebuilders Select Industry Index	
XLB	Materials Select Sector SPDR	St. Str. Gl. Adv.	Materials Select Sector Index	
XLE	Energy Select Sector SPDR	St. Str. Gl. Adv.	Energy Select Sector Index	
XLF	Financial Select Sector SPDR	St. Str. Gl. Adv.	Financial Select Sector Index	
XLI	Industrial Select Sector SPDR	St. Str. Gl. Adv.	Industrial Select Sector Index	
XLK	Technology Select Sector SPDR	St. Str. Gl. Adv.	Technology Select Sector Index	
XLP	Consumer Staples Select Sector SPDR	St. Str. Gl. Adv.	Consumer Staples Select Sector Index	
XLU	Utilities Select Sector SPDR	St. Str. Gl. Adv.	Utilities Select Sector Index	
XLV	Health Care Select Sector SPDR	St. Str. Gl. Adv.	Health Care Select Sector Index	
XLY	Consumer Discretionary Select Sector SPDR	St. Str. Gl. Adv.	Consumer Discretionary Select Sector Index	

Table 7.1: List of ETFs, with provider, tracked asset and possible leverage or inverse feature. A "2x" leveraged ETF is one which seeks to provide 2 times the daily performance of the tracked index, with "-2x" standing for 2 times the inverse of the daily performance. [a]St. Str. Gl. Adv. = State Street Global Advisors, St. Str. Gl. Mkts. = State Street Global Markets.

ETF	1-minute observations	5-minute observations
DIA	328243 (98.81%)	65497 (99.60%)
DXD	276498 (83.23%)	63160 (96.05%)
EEM	329574 (99.21%)	65414 (99.48%)
EFA	327608 (99.20%)	65035 (99.48%)
EWA	250225 (75.32%)	64266 (97.73%)
EWG	192693 (58.00%)	61246 (93.14%)
EWH	273999 (82.48%)	64911 (98.71%)
EWJ	307276 (92.50%)	65401 (99.46%)
EWM	213888 (64.38%)	61752 (93.91%)
EWS	251315 (75.65%)	64113 (97.50%)
EWT	296362 (89.21%)	65143 (99.06%)
EWU	125763 (37.90%)	53109 (80.86%)
EWW	309977 (93.64%)	64921 (99.08%)
EWY	288868 (86.95%)	64922 (98.73%)
EWZ	328035 (98.74%)	65389 (99.44%)
FXI	317549 (95.59%)	65258 (99.24%)
GDX	280744 (84.51%)	64605 (98.25%)
GLD	325293 (97.92%)	65483 (99.58%)
IBB	218702 (65.83%)	62490 (95.03%)
ICF	278878 (83.95%)	64651 (98.32%)
IJH	208636 (62.80%)	62956 (95.74%)
IJR	277711 (83.60%)	65016 (98.87%)
IVE	224241 (67.50%)	63684 (96.85%)
IVV	302419 (91.03%)	65330 (99.35%)
IVW	239155 (71.99%)	64142 (97.54%)
IWD	293429 (88.33%)	65285 (99.28%)
IWF	300553 (90.47%)	65310 (99.32%)
IWM	330785 (99.57%)	65508 (99.62%)
IWN	291552 (87.76%)	65127 (99.04%)
IWO	290830 (87.55%)	65111 (99.02%)
IYR	313139 (94.26%)	65114 (99.02%)
MDY	321856 (97.23%)	65087 (99.33%)
MZZ	172022 (51.78%)	58381 (88.78%)
PHO	164836 (49.62%)	59896 (91.09%)
QID	324741 (97.75%)	65335 (99.36%)
QLD	299218 (90.07%)	64862 (98.64%)
QQQQ	330694 (99.54%)	65483 (99.58%)
RKH	246159 (74.45%)	61953 (94.66%)
RTH	283523 (85.45%)	64630 (98.40%)
SDS	302133 (90.95%)	64599 (98.24%)
SLV	251230 (75.62%)	64052 (97.41%)
SPY	330781 (99.57%)	65502 (99.61%)
SSO	255237 (76.83%)	61252 (93.15%)
TIP	190454 (57.33%)	61374 (93.33%)
USO	312469 (94.06%)	65243 (99.22%)
XHB	255791 (77.00%)	63412 (96.43%)
XLB	306421 (92.24%)	65040 (98.91%)
XLE	330192 (99.39%)	65404 (99.46%)
XLF	325164 (97.88%)	65276 (99.27%)
XLI	287595 (86.57%)	64975 (98.81%)
XLK	282942 (85.17%)	65095 (98.99%)
XLP	253028 (76.17%)	64589 (98.22%)
XLU	298217 (89.77%)	65011 (98.86%)
XLV	265267 (79.85%)	64688 (98.37%)
XLY	276167 (83.13%)	64726 (98.43%)

Table 7.2: Number of observations in the price series at frequencies of 1 minute and 5 minutes, both in absolute and relative terms.

we normalise the deseasonalised returns by the volatility (see Section 6.2.4), thus obtaining the *standardised returns*.

Outliers

We search for anomalous values in recorded price data by means of the algorithm of Brownlees and Gallo presented in Section 6.2.1. Parameter values are set equal to $k = 20$, $\delta = 10\%$, $c = 5$, $\gamma = 0.05$. Results of this outlier detection procedure lead to the removal of a number of 1-minute observations ranging from 9 to 310 across the 55 ETFs. Their distribution with respect to the time of the day shows that these observations occur for the great majority at the very beginning and at the very end of the trading day, suggesting that the algorithm spuriously identifies as outliers some genuine observations where high variability is physiological. However, the number of 1-minute observations detected as outliers is so limited (about one every three days in the worst case) that even spurious removal has negligible impact on the results.

Splits

The search for possible stock splits is performed as outlined in Section 6.2.2. We detect four splits in our data: two 3-for-1 splits in the ETFs EEM and FXI, and one 10-for-1 split in SLV, all occurring at minute 11:10 of 17th July 2008; the fourth one is a stock merge with no clear ratio in the RTH ETF, occurring at 11:00 of 18th June 2009, corresponding to a price change from about 6 dollars to about 77 dollars.

Intraday pattern

We filter out the daily seasonalities from the raw returns by means of the model presented in Section 6.2.3. As an example, we report in Figure 7.1 the intraday volatility profile of the DIA 1-minute return series.

Heteroskedasticity

The distributions of high-frequency returns often show a sharp spike at zero, due to the discreteness of the set of prices. We find these two facts in our data. Filtering out the heteroskedasticity by means of Equation (6.4), with the volatility estimated by Equation (6.7), considerably reduces the excess kurtosis of the returns distribution for all the ETFs, thus proving to be an effective method. For instance, for the FXI ETF the excess kurtosis of 1-minute returns equals 11.87 before removing the heteroskedasticity and 0.88 after doing it. Figure 7.2 shows the histograms of FXI intraday 1-minute

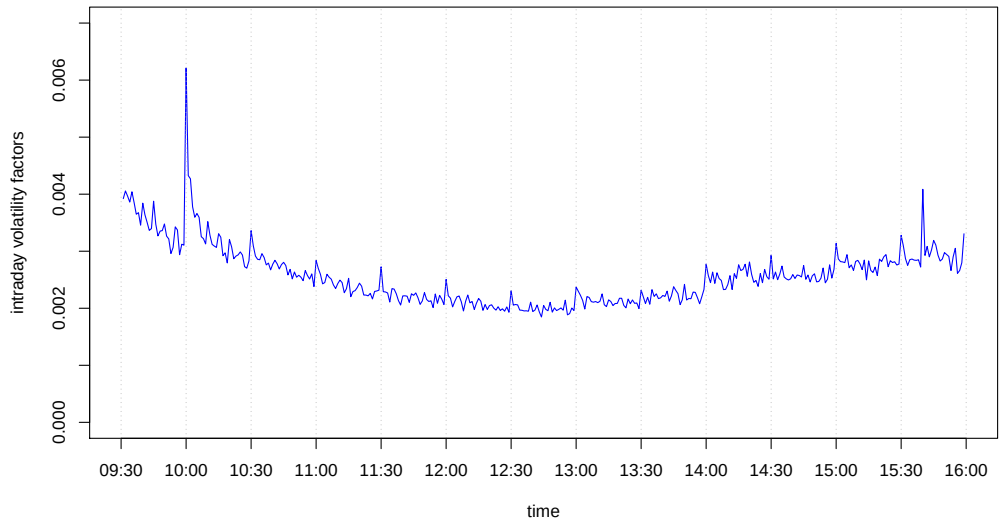

Figure 7.1: Intraday volatility profile of 1-minute returns of the DIA ETF.

returns, the intraday pattern being already removed, before and after the removal of heteroskedasticity by means of Equation (6.4). As can be seen in the figures, there is a spike at 0 representing the great number of zero returns.

7.2 Binary alphabet

7.2.1 Discretising returns

The simplest symbolisation of price returns is the one which distinguishes only the two cases of positive and negative return, corresponding to the two behaviours of price moving up and moving down, respectively. Stationarity of price can not be included in either of the two behaviours as long as the symbolisation is defined in a symmetrical way. This point highlights the fact that returns are distributed on a discrete set of values. If they were instead distributed on a continuous set of values, taken according to an absolutely continuous distribution, the probability of taking a precise value would be zero, thus negligible in practical cases.

If $\{r_t\}_t$ is the time series of non-zero returns, we define the 2-symbol

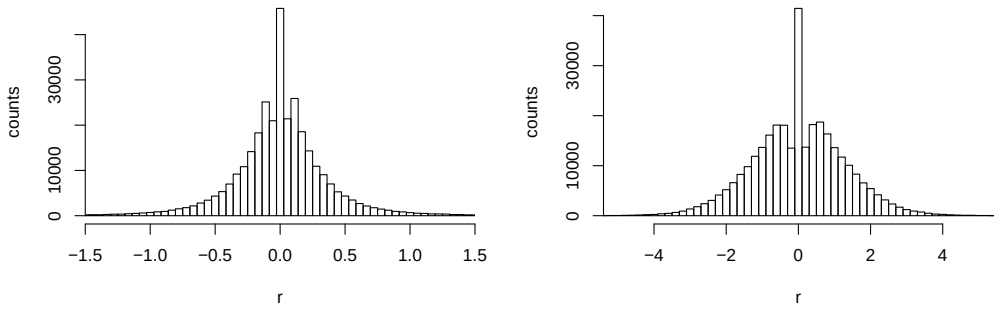

Figure 7.2: Histograms of 1-minute returns of the FXI ETF, before (left) and after (right) removing the heteroskedasticity. The intraday pattern has already been filtered out in both series.

sequence $\{s_t\}_t$

$$s_t = \begin{cases} 0 & \text{if } r_t < 0 \\ 1 & \text{if } r_t > 0 \end{cases} . \tag{7.2}$$

In Table 7.3 we report the number of the two symbols '0' and '1' in the ETF return series symbolised according to symbolisation (7.2). Ten ETFs exhibit a difference in the number of the two symbols, which is statistically significant at the 1% significance level when a null model is assumed in which at each time instant prices have the same probability of moving up and moving down. Under such assumptions, the number of either upward and downward movements of the price has a binomial distribution.

The simplest null hypothesis to check consists in the returns being independent, that is, indistinguishable from strong white noise. For such a model, the Shannon entropy of the process symbolised according to (7.2) equals one, since all the strings s_1^k have the same probability $\mu(s_1^k) = \prod_{i=1}^{k} \mu(s_i) = \frac{1}{2^k}$ and the uncertainty is maximum. We stress that the assumption of independence is not realistic, since intraday returns are known to possess some features of correlation, mainly due to microstructure effects. However, this first analysis allows to measure to what degree the ETF series of symbolised non-zero returns depart from complete randomness. To put it another way, it provides an average quantification of how much the sign of the returns is predictable.

We test this hypothesis by comparing the entropy of the ETF return series and the entropy of a white noise process with independent Gaussian innovations. The former is of course the entropy of a single realisation (the

ETF	negative 1-min r	positive 1-min ret	negative 5-min ret	positive 5-min ret
DIA	147229 (49.84%)	148163 (50.16%)	31193 (49.79%)	31457 (50.21%)
DXD	128403 (49.99%)	128454 (50.01%)	30530 (50.06%)	30462 (49.94%)
EEM	**143965** (49.61%)	**146212** (50.39%)	**30645** (49.29%)	**31525** (50.71%)
EFA	**136687** (49.75%)	**138083** (50.25%)	29936 (49.60%)	30419 (50.40%)
EWA	91641 (49.91%)	91962 (50.09%)	26932 (49.59%)	27376 (50.41%)
EWG	**68745** (49.61%)	**69819** (50.39%)	**25099** (49.16%)	**25956** (50.84%)
EWH	88689 (49.97%)	88792 (50.03%)	25504 (49.79%)	25720 (50.21%)
EWJ	89201 (50.07%)	88943 (49.93%)	23583 (49.86%)	23715 (50.14%)
EWM	59347 (49.76%)	59918 (50.24%)	21354 (49.43%)	21845 (50.57%)
EWS	73803 (49.92%)	74028 (50.08%)	23376 (49.60%)	23749 (50.40%)
EWT	89682 (49.97%)	89796 (50.03%)	24732 (49.77%)	24965 (50.23%)
EWU	46786 (49.93%)	46915 (50.07%)	22403 (49.88%)	22510 (50.12%)
EWW	127767 (50.07%)	127398 (49.93%)	29984 (49.98%)	30006 (50.02%)
EWY	117983 (49.87%)	118594 (50.13%)	29416 (49.57%)	29928 (50.43%)
EWZ	**146076** (49.69%)	**147908** (50.31%)	**30861** (49.44%)	**31562** (50.56%)
FXI	142538 (49.78%)	143791 (50.22%)	**30939** (49.46%)	**31609** (50.54%)
GDX	125416 (50.05%)	125179 (49.95%)	30850 (50.13%)	30687 (49.87%)
GLD	**143987** (49.75%)	**145424** (50.25%)	30773 (49.69%)	31152 (50.31%)
IBB	96749 (50.08%)	96446 (49.92%)	29466 (49.79%)	29713 (50.21%)
ICF	126453 (49.96%)	126646 (50.04%)	30910 (49.97%)	30949 (50.03%)
IJH	95337 (49.74%)	96324 (50.26%)	30047 (49.65%)	30465 (50.35%)
IJR	121864 (49.85%)	122588 (50.15%)	30438 (49.60%)	30928 (50.40%)
IVE	98649 (49.99%)	98680 (50.01%)	29731 (49.89%)	29867 (50.11%)
IVV	137548 (49.92%)	138016 (50.08%)	31272 (49.84%)	31478 (50.16%)
IVW	102972 (49.83%)	103664 (50.17%)	29584 (49.80%)	29821 (50.20%)
IWD	127387 (50.00%)	127382 (50.00%)	30500 (49.90%)	30623 (50.10%)
IWF	123566 (49.78%)	124655 (50.22%)	29797 (49.70%)	30158 (50.30%)
IWM	145131 (49.79%)	146334 (50.21%)	30908 (49.69%)	31291 (50.31%)
IWN	129162 (49.88%)	129777 (50.12%)	30716 (49.61%)	31193 (50.39%)
IWO	**128803** (49.74%)	**130171** (50.26%)	30909 (49.86%)	31079 (50.14%)
IYR	137814 (50.03%)	137646 (49.97%)	30887 (49.92%)	30990 (50.08%)
MDY	147600 (49.77%)	148983 (50.23%)	**31168** (49.48%)	**31819** (50.52%)
MZZ	81551 (50.11%)	81179 (49.89%)	28481 (50.35%)	28080 (49.65%)
PHO	67273 (49.72%)	68019 (50.28%)	26724 (49.47%)	27297 (50.53%)
QID	150071 (50.09%)	149533 (49.91%)	31609 (50.10%)	31487 (49.90%)
QLD	**139639** (49.63%)	**141716** (50.37%)	31519 (49.94%)	31598 (50.06%)
QQQQ	140048 (49.81%)	141141 (50.19%)	30368 (49.87%)	30523 (50.13%)
RKH	115708 (50.23%)	114637 (49.77%)	30159 (50.33%)	29769 (49.67%)
RTH	94231 (49.88%)	94683 (50.12%)	26214 (50.28%)	25921 (49.72%)
SDS	140510 (49.96%)	140740 (50.04%)	31313 (50.07%)	31230 (49.93%)
SLV	**110158** (49.61%)	**111891** (50.39%)	**29885** (49.43%)	**30569** (50.57%)
SPY	150017 (49.89%)	150698 (50.11%)	31376 (49.78%)	31649 (50.22%)
SSO	118571 (49.85%)	119300 (50.15%)	29536 (49.85%)	29718 (50.15%)
TIP	**81407** (49.36%)	**83517** (50.64%)	**27780** (49.26%)	**28612** (50.74%)
USO	139722 (49.91%)	140246 (50.09%)	30851 (49.60%)	31346 (50.40%)
XHB	102953 (50.25%)	101927 (49.75%)	**29155** (50.59%)	**28475** (49.41%)
XLB	121761 (49.79%)	122788 (50.21%)	29421 (49.67%)	29815 (50.33%)
XLE	**146881** (49.66%)	**148876** (50.34%)	**30955** (49.48%)	**31607** (50.52%)
XLF	124317 (50.21%)	123291 (49.79%)	28938 (50.25%)	28646 (49.75%)
XLI	108042 (50.02%)	107957 (49.98%)	28508 (49.95%)	28566 (50.05%)
XLK	103627 (49.87%)	104160 (50.13%)	27995 (49.81%)	28208 (50.19%)
XLP	89120 (49.97%)	89211 (50.03%)	26559 (49.62%)	26961 (50.38%)
XLU	111554 (49.92%)	111901 (50.08%)	28330 (49.80%)	28553 (50.20%)
XLV	96419 (49.91%)	96754 (50.09%)	27511 (49.65%)	27897 (50.35%)
XLY	102484 (50.03%)	102359 (49.97%)	28201 (49.83%)	28391 (50.17%)

Table 7.3: Number of symbols '0' and '1' in the ETF return series symbolised according to (7.2). Bold values indicate return series where statistically significant asymmetry is found in the number of the two symbols (at 99% confidence level).

time series), while the latter is the entropy estimated on a Monte Carlo simulation of 1000 realisations, each of which having the same length of the ETF time series to be compared with. Indicating with $\hat{h}_k$ the estimators of the rescaled entropies $\tilde{h}_k$ of Equation (7.1), we check whether it holds or not that

$$\hat{h}_k^{\mathrm{ETF}} \in [\hat{h}_k^{\mathrm{WN},0.5\%}, \hat{h}_k^{\mathrm{WN},99.5\%}], \tag{7.3}$$

where $\hat{h}_k^{\mathrm{WN},x\%}$ denotes the x-th percentile of the white noise Monte Carlo simulation. It turns out that the efficiency hypothesis of the ETF return series being indistinguishable from white noise is rejected for the great majority of the ETFs, as expected. There are however few cases for which this basic test fails to reject the efficiency hypothesis. We note that there is some dependence of the results on the order k of the considered entropies. For $k = 2$, property (7.3) does not hold for ETFs IYR, XHB, XLB; for $k = 3$ it is violated by ETFs MZZ, RKH, XLB; for $k = 6$ it does not hold only for RKH; for $k = 10$ exceptions to (7.3) are EWZ, IBB, IYR, MZZ, XLE. In Figure 7.3 we show the order 10 entropies $\hat{h}_{10}^{\mathrm{ETF}}$ with confidence bands of 99%. Looking at the top picture, we see that there is a number of ETFs quite close to the efficiency level of 1, although just five of them are not statistically distinguishable from perfectly efficient white noise. There is however a remarkable number of ETFs which are very far from the condition of efficiency. In what follows we shall analyse to what extent this is attributable to microstructure effects that can be modelled and consequently how much of this inefficiency remains after filtering out the predictability due to microstructural dependence.

By repeating the basic test (7.3) with the 5-minute return series (bottom graph in Figure 7.3), we see how a lower sampling frequency makes the return series more efficient. Of course this is expected, since many inefficiencies must be imputable to microstructure effects, which are greater at higher frequencies. However, we still see that there are a number of ETFs (namely, EWJ, EWM, EWS, EWT, EWU, TIP) that, compared with others, maintain a large degree of inefficiency. It is also notable that in two cases (PHO and RTH) a relatively low entropy in the 1-minute return series corresponds to a relatively high entropy in the 5-minute return series. In these two cases it appears that the sole change in the frequency of the observations removes almost all the inefficiency. This contrasts with other cases of comparable low entropy for the 1-minute return series, such as EWU, which exhibit the same feature of having a relatively low entropy in the 5-minute return series as well. We point out that this should be considered as evidence of the fact that equally inefficient series may have different causes at the origins of their inefficiencies.

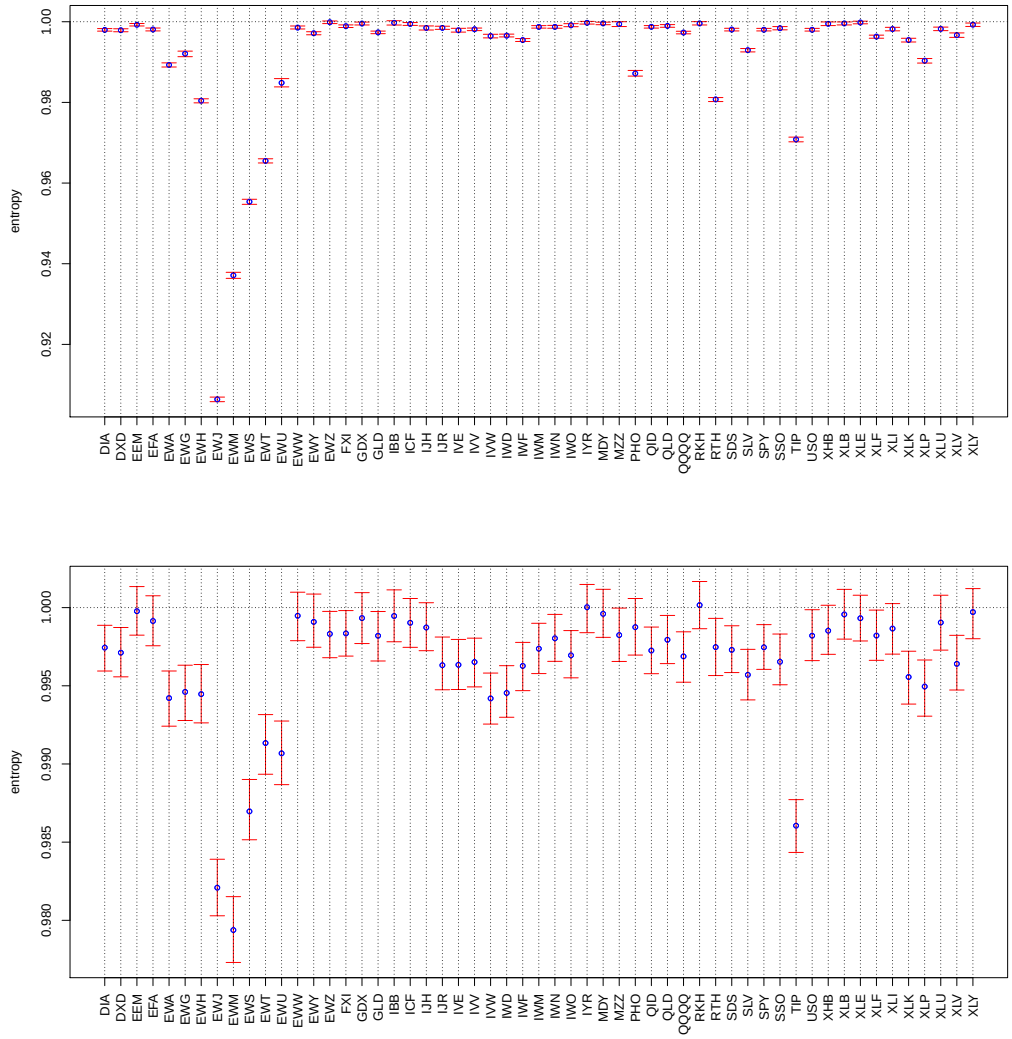

Figure 7.3: Estimated $\hat{h}_{10}^{\mathrm{ETF}}$ values (circles) of the 55 ETFs, with 99% confidence bands obtained from Monte Carlo simulations of a white noise process of the same length of the ETFs 1-minute (top) and 5-minute (bottom) return series.

7.2.2 Modelling return series as AR(1) or MA(1) processes

As already anticipated in Section 7.2.1, the fact that intraday return series generally show significant departure from perfect randomness is expected, since it is a stylised fact (see, for example, [79]) that intraday returns possess

some significant correlation, at least at the first lag. One source of this correlation is the bid-ask bounce in transaction prices, which shows most clearly at higher frequencies. The bouncing prices are responsible for a negative autocorrelation at the first lag.

A possible model to theoretically explain this stylised fact is the following. Market (logarithmic) prices p_t are supposed to differ from the latent efficient (or rational or fundamental) prices p_t^* by temporary pricing errors u_t, so that it holds

$$p_t = p_t^* + u_t. \tag{7.4}$$

The market returns

$$r_t = r_t^* + u_t - u_{t-1} \tag{7.5}$$

are therefore the sum of two terms. The first is represented by the rational returns r_t^*, which are the rational response to fundamental information and are assumed to be white noise from the theory of efficient markets. Let σ^2 indicate their variance, that is, assume that $r_t^* \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$. The second component is $u_t - u_{t-1}$. Depending on the structure imposed on the pricing error component u_t, many structural models for microstructure effects can be recovered. In the simplest case, $\{u_t\}$ is an i.i.d. noise process independent of the price process. Let η^2 indicate the variance of random variables u_t. The observed returns process is then MA(1) with $\mathbb{E}[r_t] = 0$ and autocovariance function given by

$$\mathbb{E}[r_t r_{t-\tau}] = \begin{cases} \sigma^2 + 2\eta^2 & \text{for } \tau = 0 \\ -\eta^2 & \text{for } \tau = 1 \\ 0 & \text{for } \tau \geq 2 \end{cases}.$$

If the pricing errors u_t are assumed to follow instead an AR(1) process, then the returns process is ARMA$(1, 1)$, with more complex autocorrelation structure.

Some empirical high-frequency data indeed show a typical MA(1) structure in the autocorrelation of returns, although others do not. In many cases, returns exhibit significant autocorrelation also at lags greater than 1 (see Figure 7.5). A typical picture is one where the autocorrelation function shows an alternating sign, decreasing in absolute size as the lag gets larger (see the top right and bottom left panels of Figure 7.5; see also [83]). In [83], the authors propose a simple model to capture this alternating-sign higher order dependence, which resembles an AR(1) model and is slightly more complicated. Motivated by this similarity and by the fact that the procedure in Section 7.2.3 identifies the AR(1) model as the best ARMA(p, q) model for some return series, we regard the AR(1) model as a good compromise between effectiveness and simplicity.

For these two simple return models (the AR(1) and the MA(1)), we developed in Chapter 4 an analytical approach to determine the theoretical values of their Shannon entropies. In this section, we take the ETFs whose non-zero return series are indeed well described by an AR(1) or an MA(1) process and we provide a quantification of their inefficiency as the degree to which the entropy measured on the empirical series differs from the theoretical one.

More precisely, our procedure is the following. We first estimate the best ARMA(p, q) model for each one of the non-zero return series (see Section 7.2.3 for the details of the estimate procedure and the results on all the ETF series). Then, for those series whose estimated parameters p and q are such that $p + q \leq 1$—that is, for the series whose best ARMA estimate is an AR(1), an MA(1) or just a white noise process—we can easily calculate the theoretical values h_2^{th} and h_3^{th} of the entropies h_2 and h_3. They just equal 1 in the case of the white noise process, while for the AR(1) and the MA(1) processes can be obtained by applying the formulas derived in Section 4.2.2. Finally, for each of these series we compute the inefficiency scores I_2 and I_3 defined by

$$I_k = \frac{h_k^{\text{th}} - \hat{h}_k^{\text{ETF}}}{\sigma_k}, \qquad k = 2, 3, \tag{7.6}$$

where $\hat{h}_k^{\text{ETF}}$ is the conditional entropy of order k measured on the binarised return series and σ_k is the standard deviation of the estimator $\hat{h}_k$ on Monte Carlo realisations of the estimated process with the same length as the ETF's return series. The scores I_k measure how much the empirical series depart from being a pure AR(1), MA(1) or white noise process. Making the basic assumption that a perfectly efficient ETF should perfectly follow one of these processes (the linear ARMA dependencies being due only to microstructure features), the scores I_k indeed provide a measure of the amount of inefficiency present in the empirical return series.

In Figure 7.4 we show the comparison between the entropy $\tilde{h}_3$ of the ETF return series and the theoretical value of h_3—given by the formulas obtained in Chapter 4—of the processes AR(1), MA(1) or white noise, which best fit the return series. We notice that, as can be expected, the theoretical entropy values are always higher than the return series entropy, meaning that the return series are less efficient than the corresponding AR(1), MA(1) or white noise processes. The difference between the two values provides us with a quantification of the inefficiency of the return series against their benchmark.

In Table 7.4 we report the inefficiency rankings as determined by the inefficiency scores (7.6). As the 17 ETFs treated here are a subset of the 55 ETFs for which we determine in Table 7.6 other rankings, according to

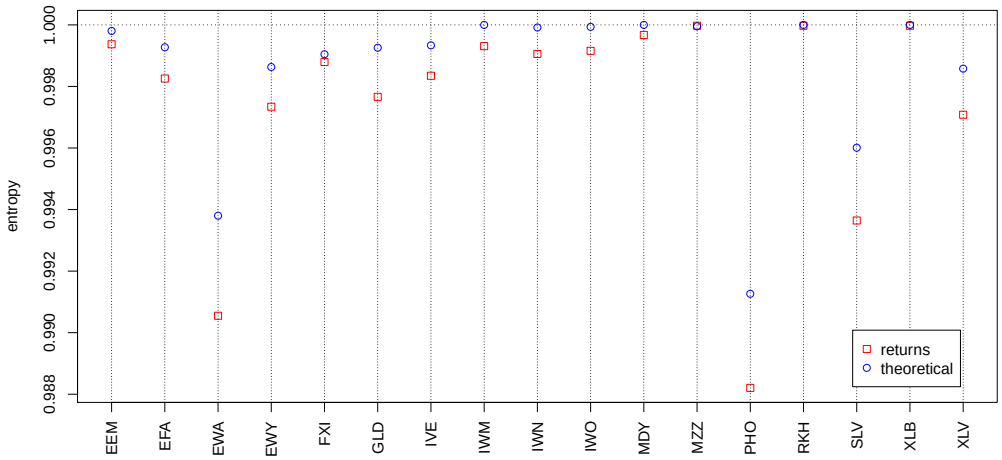

Figure 7.4: The entropy $\tilde{h}_3$ of the non-zero 1-minute return series and the theoretical values of the h_3 entropy of the corresponding best fitting AR(1), MA(1) or white noise models, for the 17 ETFs well described by one of these models.

a different score of inefficiency, we report in parentheses this latter ranking for comparison purposes. We notice that, although the two approaches differ to a certain degree for the most inefficient ETFs (top and central rankings), they give quite stable results for the most efficient ones (bottom rankings).

7.2.3 Modelling return series as ARMA(p, q) processes

In Section 7.2.2 we showed how a simple model for the efficient and the observed price implies that returns follow an MA(1) process. As already anticipated, the same model can explain more complex autocorrelation structures, by changing the structure of the pricing error component u_t in Equations (7.4) and (7.5). We now change perspective and adopt an approach which is data-driven. In this section we make no a priori assumption and start from the empirical autocorrelation functions of the ETF data, which show different scenarios.

In Figure 7.5 we show the sample autocorrelation of the series made up of non-zero 1-minute returns. At the top left we see a case (the ETF SLV) where significant autocorrelation is present at the lag 1 and almost at no other lag, which is the typical structure of a MA(1) process. At the top right the ETF EWA shows a situation where the autocorrelation is significant at the first

ETF	I_2	rank		I_3	rank	
IWM	81.035	1	(9)	65.330	1	(8)
IWO	43.994	2	(4)	38.037	2	(9)
MDY	41.642	4	(12)	30.622	3	(12)
EWA	43.527	3	(1)	29.653	4	(1)
PHO	40.688	5	(10)	26.298	5	(11)
IWN	27.514	8	(8)	20.956	6	(6)
GLD	27.258	9	(5)	19.355	7	(4)
XLV	28.054	7	(7)	19.274	8	(7)
SLV	28.914	6	(14)	19.192	9	(14)
IVE	23.009	11	(6)	17.939	10	(5)
EWY	25.015	10	(2)	17.012	11	(2)
EEM	18.109	13	(11)	13.342	12	(10)
EFA	19.589	12	(3)	13.131	13	(3)
FXI	11.726	14	(13)	7.517	14	(13)
XLB	2.133	17	(16)	2.285	15	(16)
RKH	5.917	15	(15)	2.144	16	(15)
MZZ	2.716	16	(17)	0.360	17	(17)

Table 7.4: Inefficiency scores I_k, for $k = 2, 3$, and corresponding rankings of the 17 ETFs well described by AR(1), MA(1) or white noise model (first means most inefficient). Rankings in parentheses refer to relative positions in Table 7.6.

lag, which is the dominant part, but also at lags 2 and 3, with an alternating sign and a decreasing absolute value which are typical of an AR(1) process with negative parameter. The ETF EWM at the bottom left of the figure has some features which recall those of EWA, such as the alternating sign and the decreasing absolute value, yet the situation is more complicated. The autocorrelation function is in fact statistically significant for many lags (from 1 to 11). Furthermore, since the decay does not look exponential, one single autoregressive parameter should not suffice to describe the dynamics of returns. Finally, at the bottom right we report a case (the ETF SPY) where some negative autocorrelation is significant at many lags, but there is no clear structure in the autocorrelation function. The four scenarios depicted here do not cover all the behaviours of the autocorrelation function of the 55 ETFs. However they provide a picture of how different the autocorrelation functions of different assets can be.

For these reasons, we do not assume any a priori model and instead we look for the best linear ARMA(p, q) model that describes the data. Of course there does not exist a model that is the best in fitting the data, since by augmenting the number of parameters of the model we can describe them better and better, increasing the likelihood. However this may result in overfitting and therefore what one seeks is a compromise between good description of the data and the complexity of the model, that is, the number of parameters. There is no unique way to do this, and two possibilities are using the *Akaike Information Criterion* (AIC), developed by Akaike ([84]),

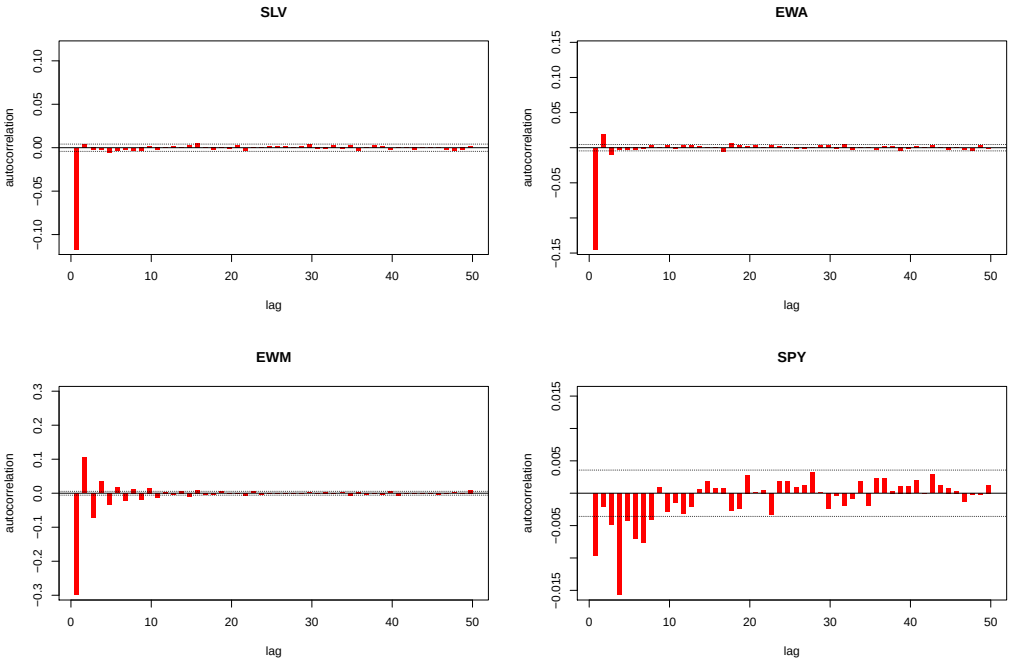

Figure 7.5: Autocorrelation functions of non-zero 1-minute returns for the ETFs SLV, EWA, EWM, SPY. Dotted lines indicate significance levels given by $\pm\frac{1.96}{\sqrt{L}}$, where L is the length of the series.

and the *Bayesian Information Criterion* (BIC), developed by Schwarz ([85]). They are both based on the likelihood and have a penalty term to weigh the complexity of the model in terms of the number of parameters to be estimated. Given an observed time series $x_1, \ldots, x_n$ of length n and a model with m parameters $\alpha_1, \ldots, \alpha_m$, they are defined by

$$\text{AIC} = -2\log(L) + 2m, \tag{7.7}$$

$$\text{BIC} = -2\log(L) + m\log(n), \tag{7.8}$$

where $L = \mathcal{L}(\alpha_1, \ldots, \alpha_m | x_1, \ldots, x_n) = \mathbb{P}(x_1, \ldots, x_n | \alpha_1, \ldots, \alpha_m)$ is the likelihood of the parameters, given the data x_i. The *best* model, according to either of the two information criteria, is the one for which the quantity (7.7) or (7.8) attains its minimum value. The penalty term in the BIC is larger than that in the AIC. In selecting the best ARMA(p, q) model to fit our data we shall use the BIC, in order to strongly penalise complex models with a large number of parameters.

In Table 7.5 we report, for every ETF, the AR and MA orders p and q, such that the series of non-zero 1-minute returns is best fitted with an

ARMA(p, q) model, where *best* means that the model is the one that minimises the quantity (7.8). As we can see, the return series of the ETFs whose autocorrelations are represented in Figure 7.5 are best fitted with different models: SLV with an MA(1) model, EWA with an AR(1) model, EWM with an ARMA$(3, 2)$ model, SPY with an ARMA$(2, 4)$ model. There are 13 out

ETF	p	q	ETF	p	q	ETF	p	q	ETF	p	q	ETF	p	q
DIA	0	4	EWU	2	1	IVE	0	1	PHO	1	0	USO	3	2
DXD	0	4	EWW	1	2	IVV	1	3	QID	2	2	XHB	2	0
EEM	1	0	EWY	1	0	IVW	2	1	QLD	2	4	XLB	0	0
EFA	1	0	EWZ	2	1	IWD	1	3	QQQQ	0	4	XLE	0	2
EWA	1	0	FXI	1	0	IWF	2	2	RKH	0	0	XLF	1	1
EWG	2	1	GDX	1	2	IWM	0	0	RTH	3	1	XLI	2	2
EWH	1	2	GLD	1	0	IWN	1	0	SDS	1	3	XLK	2	2
EWJ	3	2	IBB	0	2	IWO	1	0	SLV	0	1	XLP	2	1
EWM	3	2	ICF	0	3	IYR	0	2	SPY	2	4	XLU	0	2
EWS	3	2	IJH	0	4	MDY	0	0	SSO	1	1	XLV	1	0
EWT	3	2	IJR	5	1	MZZ	0	1	TIP	1	1	XLY	0	2

Table 7.5: AR and MA orders p and q resulting from the minimisation of the BIC (7.8) among all models ARMA(p, q) with $p + q \leq 8$.

of the 55 series that are best fitted either with an AR(1) model or an MA(1) model, plus other 4 that are best fitted with a simple white noise process. We stress that for these 17 cases we know what the corresponding entropies H_2, H_3, h_2, h_3 of the binary symbolisation should be, since for a white noise process they equal 1 and for the AR(1) and the MA(1) models they are given by the analytical results obtained in Chapter 4. However, we now want to perform a hypothesis test to assess whether there are further dependencies other than the linear ARMA structure. If all the amount of predictability of the series is due to their linear ARMA structure, once it is filtered out there would remain no other dependence and the series of residuals should not be distinguishable from white noise. Figure 7.6 shows how the ARMA residuals no longer contain the significant autocorrelation detected in returns. In order to measure how much inefficiency is there in the return series that is not due to the ARMA structure, we analyse the residuals of the ARMA estimates of the series, symbolise them according to (7.2) and then compare the entropy of the symbolised series with the values obtained by a Monte Carlo simulation of a white noise process. Before showing the results of this test, we first show in Figure 7.7 the comparison between the entropy $\tilde{h}_2$, defined by Equations (7.1), of the return series and of their ARMA residuals. We stress that the same qualitative picture that we are about to describe is valid also for the entropies $\tilde{h}_k$, with $k \neq 2$. It is interesting to note how different the behaviour of the various ETFs is. Some ETFs already have a high entropy in the return series, so that taking the ARMA residuals does not lead to a noticeable increase in the entropy value. Others, that have a relatively low

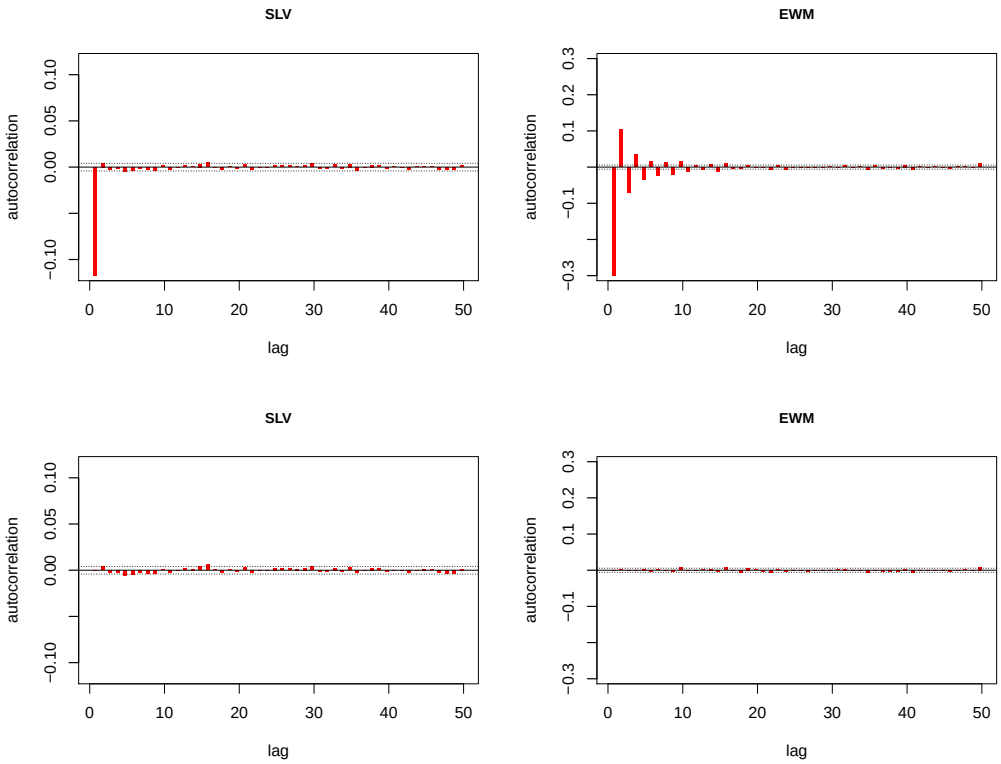

Figure 7.6: Autocorrelation of the series of non-zero 1-minute returns (top) and of their ARMA(p, q) residuals (bottom) for the ETFs SLV (left) and EWM (right). The values of (p, q) are $(0, 1)$ for SLV and $(3, 2)$ for EWM.

entropy, show a significant increase in the value of the entropy when ARMA residuals are considered in place of returns. However, there are cases where the inefficiency of the return series vanishes almost completely when taking the ARMA residuals, such as for PHO, SLV, TIP. For these ETFs it seems that a large part of the apparent inefficiency is embodied and explained by the linear dependence structure of ARMA models. For many other cases, instead, the ARMA residuals continue to contain some (nonlinear) inefficiencies.

We now go back to the hypothesis test we mentioned earlier, to see whether the residuals of ARMA estimates are distinguishable from white noise. Analogously to what we did in Section 7.2.1, we check the condition

$$\hat{h}_k^{\text{res}} \in [\hat{h}_k^{\text{WN},0.5\%}, \hat{h}_k^{\text{WN},99.5\%}], \tag{7.9}$$

where $\hat{h}_k^{\text{res}}$ denotes the estimated entropy of the series of ARMA residuals. It turns out that even considering the ARMA residuals of return series we reject

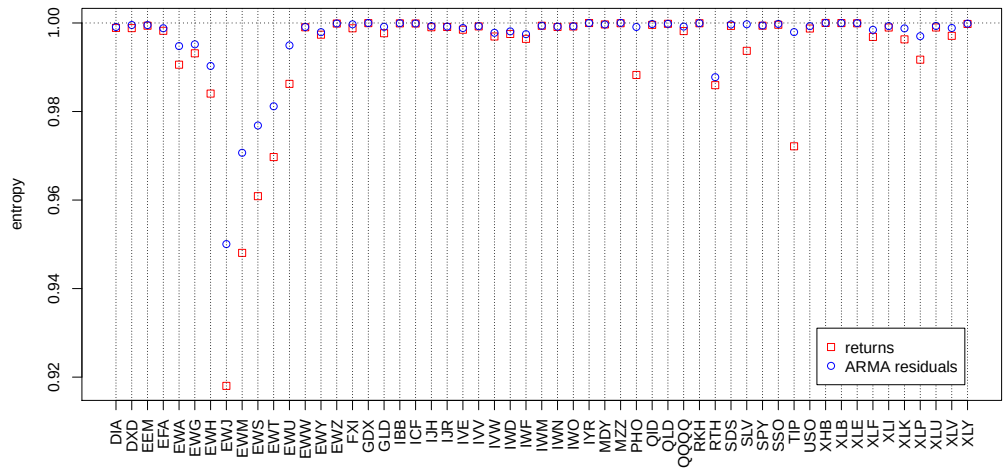

Figure 7.7: The entropy $\tilde{h}_2$ of the non-zero 1-minute return series and of the corresponding series of ARMA(p, q) residuals, where p and q are those of Table 7.5, for the 55 ETFs.

the hypothesis of efficiency for the large majority of the ETFs. The cases where the test does not reject the efficiency hypothesis are the following. For $k = 2$, GDX, MZZ, XHB, XLB; for $k = 3$, IBB, MZZ, XHB, XLB; for $k = 6$, IBB, XHB, XLE; for $k = 10$, EWZ, FXI, GDX, IBB, ICF, IYR, XHB, XLE, XLY. Referring to the results found in Section 7.2.1, we note that the cases for which the efficiency was rejected already analysing the entropy of the return series are cases where it is also rejected after filtering out the linear part, as one would expect. There are however a few exceptions (IYR for $k = 2$, RKH for $k = 3, 6$, MZZ for $k = 10$).

We argue that the predictability that remains in a series after removing the amount due to the linear component, that is, after taking the residuals of ARMA estimates, constitutes nonlinear inefficiency proper to that asset. It is interesting to quantify these inefficiencies and rank the ETFs according to their measures. We perform this by measuring how faraway the entropy of the residuals is from the value 1 of pure white noise. In particular, for each ETF we define the following measure of inefficiency:

$$I_k^{\text{res}} = \frac{1 - \hat{h}_k^{\text{res}}}{\sigma_k^{\text{WN}}}, \tag{7.10}$$

where σ_k^{WN} is the standard deviation of the estimator $\hat{h}_k$ on a white noise process of the same length as the ETF's return series. A perfectly efficient series

has values of I_k equal to 0. The farther from 0 the value of I_k, the greater the inefficiency. Results are reported in Table 7.6, for $k = 2, 3, 6, 10$. Note that varying k the ranking positions are quite stable for the most inefficient ETFs (those in the first nine positions). Many other ETFs display different degrees of variability, ranging from differences across the four rankings of at most one position to differences of eleven places.

We do not know what the reasons of the inefficiencies and the mechanisms generating them are. They may have origin in technical details of how the trading of that particular asset is regulated by the rules of the market, or in the particular strategies adopted by market makers, or also in microstructure details (such as the relative tick size) that may have different impact on different assets. Inefficiencies may also have a more *fundamental* origin, that is, they may be due to the economics of the asset and of other financial assets related to it.

Investigating the relationship between entropy and the relative tick size we find interesting results. Note that for our ETF data looking at the relative tick size (that is, the ratio between the minimum possible price variation and the price) is equivalent to looking at the price, since the absolute tick size is equal for all the 55 ETFs. It clearly emerges from the scatter plots in Figure 7.8 that the five most inefficient ETFs are those with the lowest price. A possible interpretation of this lies in the fact that an asset's price with a large relative tick size is subject to more predictable price changes. If we think of the rational price moving on a continuous scale, a change in the observed price means that the rational price has passed a tick level. Now, for an asset with a low price (or, equivalently, with a large relative tick size) it would be much more difficult to cross another tick level in the same direction than it is for an asset with a large price (that is, with a small relative tick size). Thus, most probably the observed price either moves backward or remains constant. Recall that, in the symbolisation we are considering, the stationarity of observed price (corresponding to a zero return) is ignored and simply discarded. With this considerations in mind, it is reasonable to expect the price of an asset with a large relative tick size to show some oscillating behaviour producing a more predictable symbolic sequence.

Another consideration on the rankings of Table 7.6 is one which concentrates on the ETFs which track a country index, in particular those from EWA to EWZ in Table 7.1. We notice that the most inefficient ETFs are those relative to the Asian countries (such as Japan, Malaysia, Singapore, Taiwan, Hong Kong) and to Australia. The ETFs tracking the indices of Germany and United Kingdom, though quite inefficient as well, are not in the very first places, lying behind the group of Asian countries. Finally, ETFs relative to Mexican and Brazilian indices are much behind in the inefficiency

ETF	I_2^{res}	rank	I_3^{res}	rank	I_6^{res}	rank	I_{10}^{res}	rank
EWJ	4933.980	1	2807.921	1	1177.880	1	318.380	1
EWM	1505.712	3	1132.796	2	510.979	2	147.319	2
EWS	1495.015	4	1105.510	3	445.117	3	119.970	3
EWT	1911.672	2	1053.470	4	427.720	4	110.355	4
RTH	1307.800	5	790.725	5	328.179	5	94.070	5
EWH	737.441	6	546.053	6	228.858	6	70.411	6
EWA	417.549	7	292.831	7	118.729	7	36.094	7
IWF	271.103	9	207.461	9	83.385	8	23.151	8
EWG	371.408	8	212.230	8	81.883	9	23.147	9
IWD	214.192	13	166.222	11	69.487	10	20.591	10
EWU	208.818	15	152.508	14	55.773	14	16.896	11
IVW	256.876	10	155.098	13	56.317	13	16.794	12
XLP	232.995	11	172.684	10	66.368	11	16.750	13
EWY	213.702	14	156.279	12	57.306	12	15.309	14
TIP	187.414	16	108.947	16	48.645	16	13.820	15
QQQQ	98.620	27	81.404	22	39.632	20	13.484	16
XLF	214.455	12	122.775	15	48.263	17	13.475	17
DIA	156.796	18	104.936	17	49.220	15	13.401	18
IVV	113.026	20	79.714	23	42.456	19	13.142	19
SPY	95.455	28	75.835	25	45.118	18	12.925	20
USO	106.876	23	70.114	26	34.931	24	11.419	21
GLD	106.747	24	81.771	21	34.481	25	10.763	22
IVE	100.316	26	76.238	24	35.123	23	10.693	23
IWM	81.035	31	65.330	29	30.048	27	10.258	24
EFA	186.516	17	104.336	18	38.174	21	9.839	25
XLK	113.568	19	89.810	19	36.301	22	9.646	26
SDS	52.276	37	47.967	35	29.521	28	8.911	27
IJR	113.007	21	63.456	30	28.002	29	8.881	28
EWW	102.933	25	85.422	20	33.454	26	8.665	29
XLU	67.310	33	48.481	34	26.354	32	7.975	30
DXD	52.030	38	48.874	33	24.541	34	7.344	31
IWO	107.853	22	62.741	31	25.647	33	7.327	32
XLV	95.166	29	68.339	28	27.939	30	7.320	33
IWN	93.928	30	68.919	27	27.043	31	7.275	34
XLI	66.079	34	51.481	32	21.970	35	7.262	35
QLD	23.058	43	23.942	43	15.057	40	7.062	36
SSO	21.837	44	23.412	44	16.504	38	6.430	37
IJH	57.385	36	43.514	37	18.616	36	6.090	38
SLV	28.499	42	19.818	45	10.301	44	5.525	39
QID	38.927	40	32.219	39	17.209	37	5.203	40
PHO	67.431	32	40.896	38	16.414	39	5.196	41
MDY	42.881	39	28.769	40	14.460	42	4.976	42
EEM	60.268	35	45.020	36	14.625	41	4.530	43
MZZ	-1.888	54	0.130	54	3.503	50	3.358	44
XLB	2.133	52	2.285	52	5.464	47	2.800	45
RKH	4.666	49	3.995	51	2.763	52	2.543	46
XHB	-2.395	55	0.067	55	-0.129	55	2.320	47
XLY	12.893	46	11.997	47	10.748	43	1.943	48
IYR	2.885	51	8.558	48	4.722	48	1.715	49
FXI	36.978	41	26.709	41	9.662	46	1.558	50
ICF	11.986	47	26.368	42	10.106	45	1.364	51
GDX	1.047	53	4.637	49	3.174	51	0.614	52
IBB	4.424	50	1.568	53	1.220	54	-0.372	53
EWZ	17.015	45	14.064	46	4.359	49	-0.747	54
XLE	4.929	48	4.279	50	1.440	53	-1.457	55

Table 7.6: Inefficiency scores I_k^{res}, for $k = 2, 3, 6, 10$, and corresponding inefficiency rankings of the 55 ETFs (first means most inefficient).

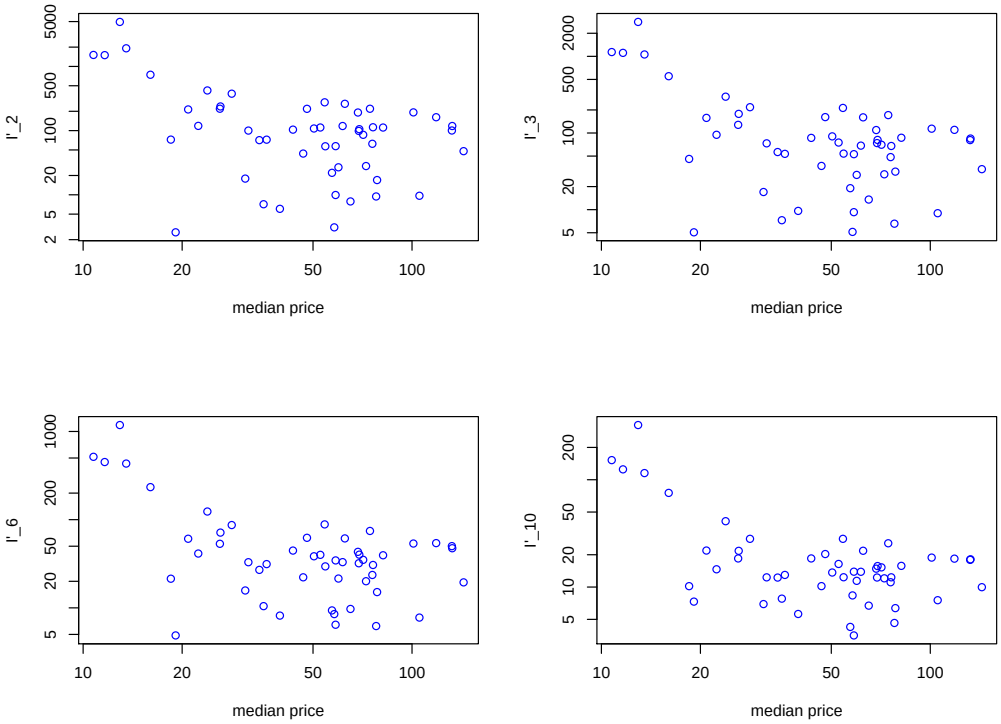

Figure 7.8: Scatter plots in log-log scale showing the relation between inefficiency (scores I_2^{res}, I_3^{res}, I_6^{res}, I_{10}^{res}) and median price of 51 ETFs (the four ETFs with a detected split/merge are omitted). The graphed scores $I'_k = I_k^{\text{res}} + 5$ are shifted versions of the inefficiency scores I_k^{res} in order to be positive for the logarithmic plot.

rankings, with the latter being among the most efficient ETFs. An exception to this classification is the ETF tracking the index of South Korea, which is not ranked in the group of ETFs of the other Asian countries. Apart from this exception, it can be argued that the levels of detected inefficiency follow the time distances from the New York time. In Figure 7.9 we show the relationship between inefficiency and opening time overlap of the NYSE with the country markets, for the mentioned country ETFs. The markets of the Asian countries of Japan, Malaysia, Singapore, Taiwan, Hong Kong, as well as the Australian market, are closed when the corresponding ETFs are being traded at the New York Stock Exchange. Therefore these assets are traded while the tracked index has no dynamics. There is instead some time overlap in the opening times of the New York Stock Exchange and of European mar-

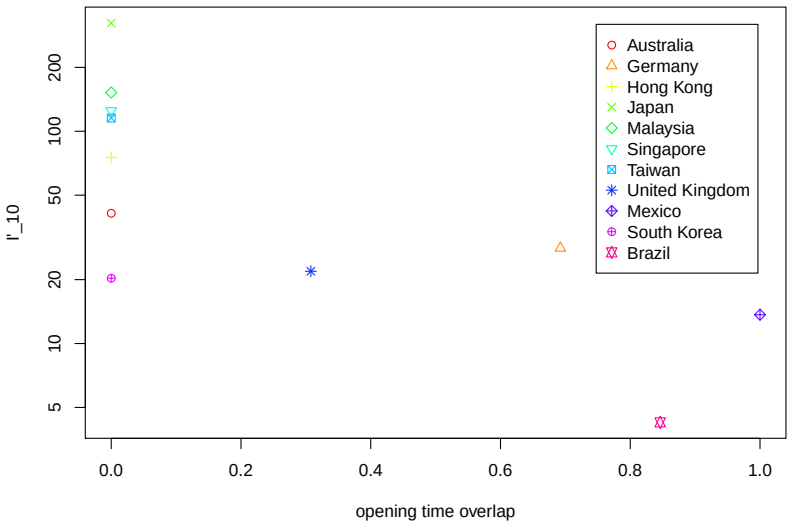

Figure 7.9: Relationship between inefficiency and opening time overlap of the NYSE with the country markets, for the country ETFs from EWA to EWZ in Table 7.1. The graphed score $I'_{10} = I^{\text{res}}_{10} + 5$ is a shifted version of the inefficiency score I^{res}_{10}, in order to be positive for the logarithmic plot. The opening time overlap on the x axis is the time when both the NYSE and the country market are simultaneously open, divided by the NYSE opening time (6.5 hours).

kets, while there is great overlap of the former with the markets of Mexico and Brazil. The trading dynamics of the ETFs on these last two countries can therefore rely on a simultaneous evolution of the corresponding indices.

We remark that, if the two mechanisms that we propose as possible explanations to what we observe in the inefficiency rankings are legitimate, it may well be that the two things are related. It may be that those ETFs that track indices of markets that are closed during the ETFs' trading time are deliberately given a low price, since their dynamics is only coarsely determined. However, we could not find any founded indication in this respect.

7.3 Ternary alphabet

As we point out in Section 7.2, a symmetrical binary discretisation of returns that takes into account all the information does not exist. The most natural one is defined by (7.2), but it ignores all the zero returns, which correspond

to intervals of price stationarity. This waste of information is larger as one moves to higher frequencies, since the probability of observing a price change in a fixed interval decreases with the decreasing of the interval length. In our data, at frequencies of 1 and 5 minutes the amount of zero returns is huge. In Section 7.2 we showed how much information on market efficiency can be extracted even ignoring the zero returns. In this section we instead use all the returns.

The idea of ternary-alphabet discretisations of returns is that a symbol represents a stability basin, encoding all returns in a neighbourhood of zero. Negative and positive returns lying outside of this basin are encoded with the two other symbols. In the papers [29] and [28], the three-symbol discretisation of returns is performed as defined in (5.6), with a threshold $b = 0.0025$. In [30], high-frequency exchange rates differences are discretised in a similar fashion with a threshold equal to three pips (i.e. $b = 0.0003$). We argue that there are numerous problems in fixing an absolute threshold in these discretisations. The main objection is that a fixed symbolisation scheme does not take into account the heteroskedasticity of the series. Time series of returns are known to display periods of different volatility, that is, periods with different average absolute size. We believe that, in such contexts, a ternary discretisation of returns should possess a character of variability, in order to consistently adapt to the volatility dynamics. The risk in not doing so is that the memory properties of the volatility are encoded in the symbolised series and spurious dependencies are introduced.

There are also other reasons that make fixed-threshold ternary symbolisations inadequate. First of all, different assets have different distributions of returns (or rates differences in the case of currency exchanges), so that a fixed neighbourhood of zero includes portions of the return distributions which are different across the assets. This introduces discrepancies in treating the time series that can potentially affect the results of the analyses. Secondly, the distribution of returns also varies as the sampling frequency varies, so that choosing a fixed symbolisation scheme for different frequencies (as in [30]) appears inappropriate. Finally, the three-symbol discretisation can be applied not only to the raw returns, but also to returns filtered for the intraday pattern as defined by (6.2) or to standardised returns as defined by (6.4). These latter two series, as well as other possible series obtained by processing the returns in some other way, range on different scales and therefore the ternary symbolisations with fixed thresholds are not the proper way to deal with their discretisation.

Concerning the three-symbol discretisations of time series, we propose a more flexible approach, which is also rather general. We define the thresholds for the symbolisation to be the two tertiles of the distribution of values taken

by the time series. More formally, if $\{r_1, \ldots, r_N\}$ is the time series, we define its tertile-symbolised series by

$$s_t = \begin{cases} 0 & \text{if } r_t < \vartheta_1 \\ 1 & \text{if } \vartheta_1 \leq r_t \leq \vartheta_2 \\ 2 & \text{if } r_t > \vartheta_2 \end{cases}, \tag{7.11}$$

where ϑ_1 and ϑ_2 denote the two tertiles of the empirical distribution of the time series $\{r_t\}$.

7.3.1 The impact of intraday patterns and volatility on market efficiency measures

As already pointed out, a three-symbol discretisation of returns, with one symbol encoding returns in a stability basin around zero, embeds to some degree the intraday pattern and the volatility into the symbolic series. Thus, when these components are not properly filtered out, the symbolic series will possess a certain amount of predictability due to the memory and regularity properties of these factors. We now proceed with a quantitative study in this respect.

As outlined in Section 7.1.3, the whitening procedure that we apply to the series of logarithmic returns R_t starts with removing the intraday pattern, getting the deseasonalised returns $\tilde{R}_t$, and continues with removing the volatility, getting the standardised returns r_t. We further treat the standardised returns to remove any ARMA component that may be due to microstructure factors, thus getting also a series of ARMA residuals ϵ_t. We symbolise all these series with tertile thresholds as in (7.11) and estimate the Shannon entropy of the symbolic series to measure their degree of randomness. By doing so, we can assess to what degree the intraday pattern, the volatility and the microstructure contribute to create regularities in the return time series.

We show in Figure 7.10 the values of the Shannon entropy $\tilde{h}_8$ for the 1-minute and 5-minute series of raw returns R_t, deseasonalised returns $\tilde{R}_t$, standardised returns r_t and ARMA$(\hat{p}, \hat{q})$ residuals of standardised returns, where $\hat{p}$ and $\hat{q}$ are the ones that minimise the Bayesian Information Criterion value, given by Equation (7.8), among all the models ARMA(p, q) with $p+q \leq 5$. Similar features to those discussed below also hold for results obtained with the entropies $\tilde{h}_k$, with $k \leq 10$ and $k \neq 8$.

For five 1-minute series (ETFs EWH, EWJ, EWM, EWS, EWT) it happens that the proportion of zero returns is so large that the two tertile thresholds ϑ_1 and ϑ_2 are equal to zero. The corresponding symbolic series thus have

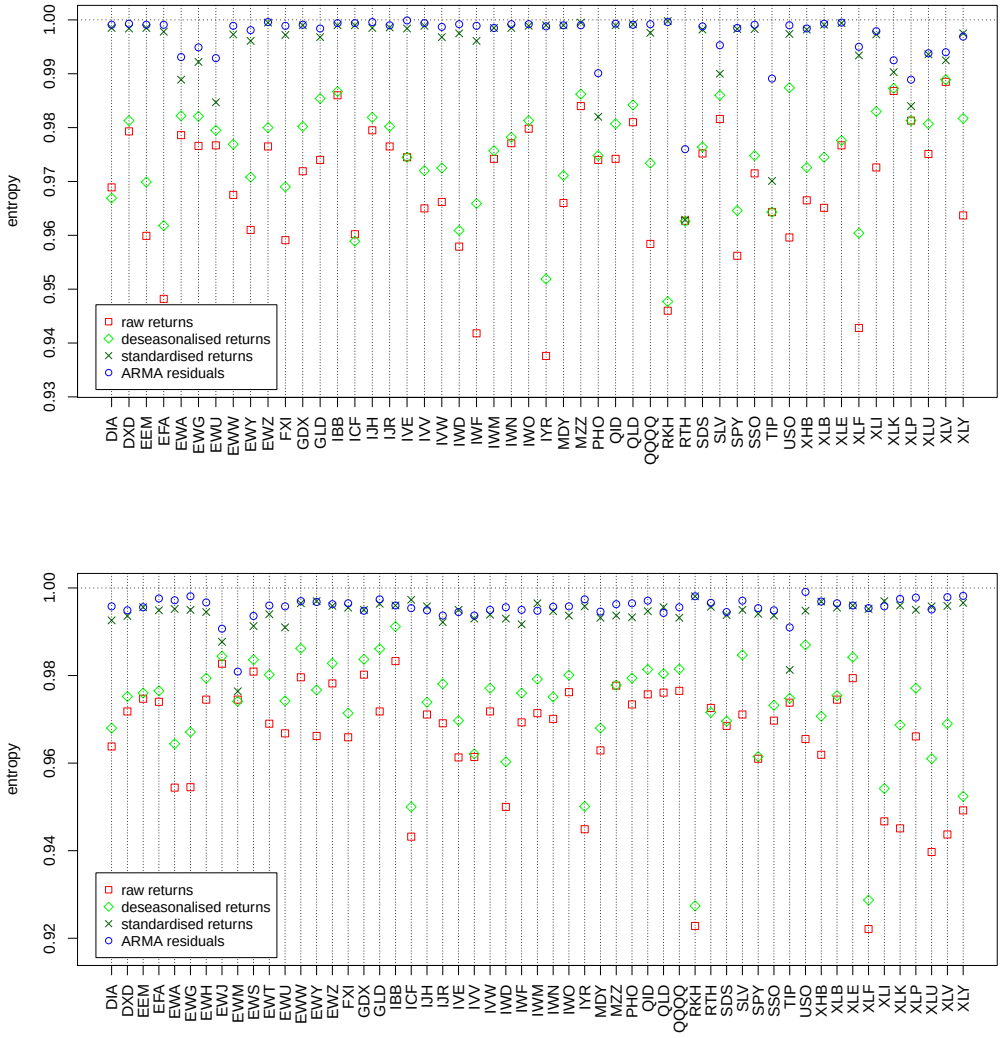

Figure 7.10: The entropy $\tilde{h}_8$ of the 1-minute (top) and 5-minute (bottom) series of raw returns, deseasonalised returns, standardised returns, ARMA residuals, for the ETFs. Results for the ETFs EWH, EWJ, EWM, EWS, EWT are considered for the 5-minute series only.

an unbalanced number of the three symbols. In order to avoid comparing series with balanced distributions of symbols with series with unbalanced distributions, we ignore the results on the latter. We instead report results for all 5-minute series, since at this frequency the tertiles of the return distributions never happen to equal zero.

Concerning the results reported in Figure 7.10, we notice that in the vast majority of cases the symbolised series of the raw returns are the most predictable, as is expected since they still carry all the regularities of the intraday pattern and the correlation due to the volatility and the microstructure. However, there are cases (DIA, ICF, RTH for the 1-minute series; EWM, RTH for the 5-minute series) in which removing the intraday pattern from the raw returns leads to series of lower entropy. This may seem to be not possible, but note that the tertile values ϑ_1 and ϑ_2 defining the symbolisation of the series change in an unpredictable manner when passing from the raw return series R_t to the deseasonalised $\tilde{R}_t$.

The most noteworthy results of the entropy estimates reported in the two graphs of Figure 7.10, however, are those regarding the standardisation of the returns by the volatility. In almost all the cases, it can be seen that the passage from the deseasonalised returns to the standardised ones is responsible for the largest increase in the entropy. Averaging across all the ETFs, a percentage of around 18% of the entropy increase obtained with the three whitening procedures is attributable to the removal of the intraday pattern, while about 62% is due to the standardisation for the volatility and 20% is the entropy gain given by the removal of correlations due to microstructure effects. This means that the removal of the volatility from the return series increases their randomness more than the other two procedures taken alone. Put another way, the volatility gives the return series a huge amount of predictability and it does so much more, on average, than the daily seasonality and the microstructure effects. This result should be regarded as a convincing demonstration of the fact that, when studying the randomness of a three-symbol discretised time series, the volatility must be filtered out. Omitting this operation would give results that tell more on the predictable character of the volatility than on, for example, market efficiency.

The last refinement that we do on standardised return series is the removal of dependence structures due to market microstructure effects. As we did in Section 7.2 when dealing with two-symbol discretisations, we perform it by taking the residuals of the ARMA(p, q) model that best describes the series of the standardised returns r_t. This last procedure has the general effect of removing some remaining predictability, further contributing to move return series towards perfect efficiency. We note however that this is not always the case: the entropy of the ARMA residuals is lower than that of standardised returns for the 1-minute series of IYR, MZZ, QLD, RKH, XLY and for the 5-minute series of EWY, GDX, ICF, IJH, IVE, IWM, QLD, XLI, XLU. We think that this counterintuitive behaviour might be caused by the large amount of zero returns present in the data, in correspondence of which some spurious randomness is *introduced* by taking the ARMA residuals.

Overall, the results shown in Figure 7.10 clearly indicate that much of the apparent inefficiency (that is, the predictability of the raw return series) is due to three factors: the daily seasonality, the volatility and the microstructure effects. For example, note in the 1-minute picture of Figure 7.10 how four of the most apparently inefficient ETFs (namely, EFA, IWF, IYR, RKH) see their predictability almost vanish after their series are filtered for these three factors.

We remark that, although the daily seasonality and the microstructure effects are characteristic of high-frequency time series, the memory properties of the volatility play a major role also in low-frequency (for example, daily) data. Therefore, we conclude that studies on measuring relative efficiency as randomness of symbolised return series should carefully deal with the issue of removing the volatility, for failing to do so would heavily affect the results, in fact invalidating them.

Our three-step procedure aims at removing from the return series all the predictability that is imputable to factors having their own dynamics that can be modelled. What remains in the filtered series that separates them from being purely random is what we assess as the true inefficiency of the assets. It may be due to other features of the market that we do not take into account or to more *fundamental* aspects.

7.4 Conclusions

In this chapter we study how relative market efficiency can be measured from high frequency data, filtering out all the known sources of regularity, such as the intraday pattern, the persistence of volatility and the microstructure effects. To this aim we employ the Shannon entropy as tool to measure the randomness degree of binary and ternary symbolised time series of 55 ETFs.

Using the analytical study of the entropy of the AR(1) and MA(1) processes conducted in Chapter 4, we develop an original theoretical approach to discount microstructure effects from the measuring of efficiency of return time series which exhibit simple autocorrelation structures. A very interesting topic for future work is the extension of the analytical results found in Chapter 4 to higher entropy orders and to more complex ARMA processes.

A more empirical approach, in which we choose the ARMA(p, q) process that best describes each time series, allows us to filter out the linear microstructure effects for all the ETFs and to measure residual regularities. Results show that in some cases a large part of the regularities is explained by the linear dependence structure, while in other cases the ARMA residuals still contain some (nonlinear) regularities. By rigorously testing the ARMA

residuals for efficiency, we reject the hypothesis of efficiency for the large majority of the ETFs. We also rank the ETFs according to an inefficiency score and find that the rankings are not very sensitive to the choice of the entropy order.

We find a strong relationship between low entropy and high relative tick size. This is explained by noting that an asset's price with a large relative tick size is subject to more predictable changes. We also notice that the inefficiency scores for the country ETFs can be related to the opening time overlap between the country markets and the NYSE, where the ETFs are exchanged. We hypothesise that those ETFs that track indices of markets that are closed during the ETFs' trading time are deliberately given a low price, since their dynamics cannot rely on a simultaneous evolution of the corresponding indices.

With the 3-symbol discretisation, we find that the removal of the volatility is responsible for the largest amount of regularity in the return series. This effect amounts to the 62% of the total entropy gain and thus it is larger, on average, than the combined effect of the intraday (18%) and microstructure (20%) regularity. This result convincingly demonstrates that, when studying the randomness of a three-symbol discretised time series, the volatility must be filtered out. Omitting this operation would give results that tell more on the predictable character of the volatility than on, for example, market efficiency.

Part II

Market instabilities

Chapter 8

Modelling systemic price cojumps with Hawkes factor models

8.1 Introduction

Modelling the dynamics of large movements in security prices is of paramount importance for risk control, derivative pricing, trading, and to understand the behaviour of markets. The importance of understanding the dynamics of abnormal returns, together with the recent wide availability of ultra-high frequency data for a broad range of assets, has spurred the growth of a large body of literature on their estimation and modelling. The mainstream contribution to this field of research typically deals with the identification problem at the daily time scale, i.e. answering the question whether on a given day an extreme return occurred. An incomplete list of recent studies on this topic, which is usually addressed in the econometric literature as jump identification problem, includes theoretical work with a nonparametric approach [86, 87, 88, 89, 90, 91, 92, 93], as well as empirical analysis [94, 95, 96, 97, 98, 99], and applications to asset pricing [100, 101, 102, 103, 104]. General reviews on jumps are [39] and [105]. Significantly less research has been devoted to the study of events happening synchronously in two or more stock prices. The investigation of such events, usually referred to as cojumps, is mainly of empirical nature, see e.g. [106, 107, 108, 109]; relevant exceptions dealing with cojump estimation and modelling are the papers of [110, 111] and [112].

Taking a complementary, yet different, perspective, in this study we want to explicitly identify the intraday times of abnormal returns. Thus, we resort

on an identification approach based on the standardisation of high frequency returns, in a similar spirit to [96, 113, 114]. We continue to generically refer to such extreme events as jumps and cojumps. Contrary to other types of jump tests (such as those based on multipower variations or multiscale power variations), intraday tests enable to identify the exact intraday time of a jump, as well as the number of jumps within a given trading day.

The modelling of a discontinuous component in asset price dynamics is commonly performed in the literature by a compound Poisson process, see e.g. [39]. Poisson jump processes have the great advantage of being analytically tractable. However, Poisson jumps (as well as the more general Lévy processes) have independent increments and thus they do not allow for any type of serial and cross-sectional dependence. Instead, one might argue that the jump component of the price process could be better described by a process where jumps are clustered in time. The same objection emerges when one considers many assets. Markets are nowadays more and more interconnected and it is a priori reasonable to expect that some sort of synchronisation between the jumping times of different assets is present.

This synchronisation effect had its most spectacular appearance during the May 6, 2010 Flash Crash. According to the SEC-CFTC, the crash started from a rapid price decline in the E-Mini S&P 500 market [35, 36]. However in a very short time the price drop propagated towards ETFs, stock indices and their components, and derivatives. For example, the Dow Jones Industrial Average plunged about nine percent, only to recover those losses within minutes. The contagion effect can be extremely rapid in liquid markets [115, 116] and leads to a strongly synchronised discontinuous movement of the price of many assets. This type of systemic events can not be described by a model where price jumps follow independent processes.

In this work we show that indeed the dynamics of jumps of a portfolio of stocks deviates significantly from a collection of independent Poisson processes. The deviation that we observe is twofold. On one side, by considering individual assets, we find evidence of time clustering of jumps, clearly inconsistent with a Poisson process. This means that the intensity of the point process describing jumps depends on the past history of jumps, and a recent jump increases the probability that another jump occurs.

The second deviation from the Poisson model is probably more important, especially in a systemic context. We find a strong evidence of a high level of synchronisation between the jumping times of a portfolio of stocks. In other words, we find a large number of instances where several stocks (up to 20, the size of our set) jump within the same one minute interval (which is our maximum resolution). This evidence is absolutely incompatible with the hypothesis of independence of the jump processes across assets.

Taking together these empirical deviations from the independent Poisson model, there is a need of a suitable modelling of the multi-asset jump process and this is the main methodological contribution of this work. In order to model the time clustering of jumps for individual assets we propose the use of a class of self-exciting point processes, termed Hawkes processes. These processes were introduced more than forty years ago [40], and have been widely employed to model earthquake data [117, 41, 42]. In the last years, Hawkes processes have experienced an increasing popularity in mathematical finance and econometrics. One of the first applications to financial time series is due to [43], and a wide literature review in this context is collected in [44]. Among more recent developments not covered by the latter reference we mention [118] where these processes are applied to the order flow in a continuous double auction market, [119] for the modelling of trades-through orders in a limit order book, [120] where Hawkes processes are used to introduce a new stochastic model for the variation of tick-by-tick asset price both in one and two dimension able to reproduce the strong microscopic mean reversion and the Epps effect, and [121], which introduces a measure of the market activity providing a direct access to the level of its endogeneity and as a potential predictor of market microinstabilities (a critical review of this paper is given in [122]). Finally, [123] discusses an innovative approach to foreign exchange market in which the high frequency price dynamics is affected by a self exciting mechanism and an exogenous component, generated by the pre-announced arrival of macroeconomic news.

In this study we use Hawkes processes for modelling the dynamics of jumps of individual assets and we show that they describe well the time clustering of jumps. However the direct extension of the application of Hawkes processes to describe the dynamics of jumps in a multi-asset framework is highly problematic and inconsistent with data. In fact, from a methodological point of view, even by using a simple two parameter kernel (for example, exponential) of the process, the number of parameters to estimate a Hawkes process on N assets is $N(2N+1) \sim O(N^2)$, which is clearly too high. Moreover, even when we consider $N = 2$ stocks, we empirically find that a bivariate Hawkes model is unable to describe the empirical data, especially to replicate the high number of *synchronous* jumps that we observe. This is due to the fact that the kernel structure of Hawkes is more suited to model lagged jumps rather than synchronous jumps.

For this reason, the main methodological contribution of this study is the introduction of Hawkes factor models to describe systemic cojumps. We postulate the presence of an unobservable point process describing a market factor. When this factor jumps, each asset jumps with a given probability, which is different for each stock. In general, an asset can jump also by

following an idiosyncratic point process. In order to capture also the time clustering of jumps, we model the point processes as Hawkes processes. We show how to estimate this model and discriminate between systemic and idiosyncratic jumps. We show that the model is able to reproduce both the longitudinal and the cross sectional properties of the multi-asset jump process.

Recent approaches sharing some aspects with the present study are discussed in [112, 124] and [109]. The former design a model of asset returns able to capture periods of crisis characterised by contagion and consider it to solve the problem of optimal investment-consumption for a log-utility investor. The jump diffusion component of the dynamics is described in terms of a class of multi-dimensional Hawkes models and the authors discuss an estimation methodology based on the Generalised Method of Moments. Not surprisingly, when they estimate the model on real data they face the problem of the curse of dimensionality and limit most of the study to the two asset case. Our one factor approach solves the calibration issue in a natural way and therefore represents a viable alternative to their model. By using data sampled at a frequency of eleven minutes the authors of [109] find empirical support to the hypothesis that stocks tend to be involved in systematic cojumps [1], rejecting an assumption of independence in jump arrival times among stocks. We find strong evidence of this result also at the frequency of one minute and furthermore we extensively investigate the properties of self and mutual excitation possessed by the jumps series of different stocks.

This chapter is organised as follows. In Section 8.2 we present our dataset and in Section 8.3 we summarise our jump detection method (detailed in Section 6.2). The identification of jumps is a delicate topic and therefore we are careful in using a robust identification method in order to minimise the number of false positives. In Section 8.4 we provide empirical evidence of the large number of systemic cojumps and we describe some simple statistical properties. In Section 8.5 we present the statistics we use to test our models in their ability of reproducing the multiple jumps of a single stock in a given time window and the cross jumps, i.e. the occurrences of jumps in two different stocks in a given time window. Section 8.6 discusses how Hawkes processes fit the jump process in a univariate and in a multivariate setting. In Section 8.7 we present the Hawkes factor model approach, showing how to estimate it and the results obtained in the investigated dataset. In Section 8.8 we detail a robustness analysis performed on our Hawkes factor

[1] From a terminological viewpoint, authors of [109] define a systematic cojump as an instance when one stock and a market index jump at the same time. In this work we use the term systemic cojump to indicate a sizeable number of stocks jumping simultaneously. In Section 8.7 we show how to identify systemic cojumps in a self consistent way.

model, investigating the dependence of its performances on the temporal and cross-sectional dimensions of the dataset, as well as on the specific market. Finally, in Section 8.9 we draw some conclusions.

8.2 Data description and data handling

The analyses reported in this work are performed on tick-by-tick transaction data for the period from 5th March 2012 to 9th July 2012 that have been made available to us by LIST S.p.A. [2]. In this work we investigate 20 among the most liquid stocks of the FTSE MIB index of the Italian stock exchange Borsa Italiana. We sample prices at a frequency of one minute, taking the last executed price, to obtain 1-minute logarithmic returns. The total number of days in the period is 88, with 505 intraday returns each day.

In Table 8.1 we list the twenty stocks, along with the ISIN code and the average number of trades per day. From the last column we see that the assets are characterised by a certain level of heterogeneity, since the trade activity varies between 2.8 and 27.4 thousand transactions per day. A low level of activity implies in general a higher probability for the absence of transactions inside a given sampling interval, and this effect has important implications for the methodology that we use for the detection of jumps.

The process from tick-by-tick data to jump identification can be summarised as follows. First, anomalous values in tick-by-tick price data are detected and removed. The algorithm for the outliers detection that we use is due to Brownlees and Gallo [80] and it is explained in Section 6.2.1. We take $k = 60$, $\delta = 10\%$, $c = 3$, $\gamma = 0.05$, where the parameters have the meaning of Section 6.2.1. We find no outliers at all for 13 stocks and few units for the other 7, for the vast majority concentrated on the very first minute of the day. Removing outliers can in principle introduce a distortion in the analysis. However, the number of anomalous prices that we identify is extremely low, and the probability of an actual distortion of returns, computed at the minute level, is negligible. Cleaned prices are then sampled at a frequency of one minute and logarithmic returns are calculated carrying forward to the next sampling instant the last price observed within the sampling window.

In the identification of jumps, an important care should be devoted to the way in which intervals without trades are treated. In fact, jump detection methods typically compare returns with local volatility. In periods of low liquidity (or missing observations) one risks to underestimate volatility and to identify relatively small price returns as jumps. There are two different situations which lead to having no price observations in a given time interval.

[2] `www.list-group.com`

Company	ISIN code	trades per day ($\times 10^3$)
Assicurazioni Generali	IT0000062072	7.5
Mediobanca	IT0000062957	3.6
Banca Popolare di Milano	IT0000064482	4.9
Saipem	IT0000068525	5.4
Intesa Sanpaolo	IT0000072618	18.1
Mediaset	IT0001063210	3.1
Banca Monte dei Paschi di Siena	IT0001334587	9.2
Fiat	IT0001976403	9.6
Enel	IT0003128367	10.7
Eni	IT0003132476	10.4
UBI Banca	IT0003487029	3.0
Telecom Italia	IT0003497168	6.3
Finmeccanica	IT0003856405	4.7
Prysmian	IT0004176001	2.8
Banco Popolare	IT0004231566	4.2
Pirelli & C.	IT0004623051	4.6
Fiat Industrial	IT0004644743	5.1
UniCredit	IT0004781412	27.4
Tenaris S.A.	LU0156801721	3.3
STMicroelectronics N.V.	NL0000226223	4.3

Table 8.1: List of the twenty investigated stocks, with the ISIN code and the average number of transactions per day.

The first one occurs in correspondence of volatility auction phases. According to the rules of the market [125, 126], whenever the price exits a reference interval, reaching too high or too low levels, the continuous double auction phase of the exchange is suspended and a volatility auction starts. In such a phase we consider the related returns as *not available*. We do not have direct information on volatility auctions, but we infer their presence from market data. Since the investigated stocks are among the most liquid ones, we are reasonably sure to be able to distinguish volatility auction phases from continuous double auction phases of low liquidity. For more details, see the discussion in Section 8.10.

The second mechanism for missing price observations is when the stock is available for trades, but still there are no transactions in that particular minute. There are several ways to treat these cases. To fix the notation, let i and $i+1$ label two consecutive sampling instants between which no trade has been made. The most common way to deal with the missing observation of

the price between times i and $i+1$ is to bring forward the last recorded price. This means setting the price $p_{i+1} = p_i$ and the log-return $r_{i+1} = 0$. Sampling intervals with no transactions in them are therefore given a zero return. This is our first method of dealing with Missing Observations, which we call MO1. An alternative method of treating minutes with no transactions is simply to consider the corresponding returns as not available data ($r_{i+1} = \texttt{NA}$), thus avoiding to give them a numerical value. The next available return will be at the first sampling time after the next available observation. It can be defined as the usual difference of the sampled logarithmic prices after and before the missing observations (method MO2) or as this quantity rescaled by the square root of the time between the two sampling times (MO3). For instance, for the price series $p_0, p_1, -, -, p_4, p_5$ we construct the following three return series:

$$\begin{aligned}
&\textbf{MO1} && \log\tfrac{p_1}{p_0},\ \ 0,\ \ 0,\ \ && \log\tfrac{p_4}{p_1}, \log\tfrac{p_5}{p_4};\\
&\textbf{MO2} && \log\tfrac{p_1}{p_0}, \texttt{NA}, \texttt{NA}, && \log\tfrac{p_4}{p_1}, \log\tfrac{p_5}{p_4};\\
&\textbf{MO3} && \log\tfrac{p_1}{p_0}, \texttt{NA}, \texttt{NA}, \tfrac{1}{\sqrt{3}}\log\tfrac{p_4}{p_1}, && \log\tfrac{p_5}{p_4}.
\end{aligned}$$

The reason for the rescaling of the return in method MO3 is that we do not want to identify as a jump a price change that is compatible with a diffusive random movement of the efficient price.

Once the returns are defined according to one of the three possible methodologies discussed above, they are checked in order to eliminate possible large values due to stock splits or merges. We check for the presence of stock splits in our data through the detection of returns greater than 0.2 in absolute value. This would detect, for example, a 3-for-2 split or a 4-for-5 merge. In our data we do not find any such return.

The last step to perform is the removal of daily seasonalities, by filtering out the intraday pattern from raw returns. As it is well known, intraday returns show significant seasonal behaviour, as the dynamics of markets is greatly variable during the day. Opening and closing periods generally show a higher volatility than the rest of the day, since traders are more active during these phases. If these daily seasonalities are not properly filtered out, spurious jump detection may happen in correspondence of these periods. We describe the details of how we deal with this issue in Section 8.2.1.

8.2.1 Intraday pattern

We show in Figure 8.1 the intraday volatility pattern (that is, the factors ζ_t) for Monte dei Paschi di Siena. We notice that, although the profile is

somewhat noisy because of the relatively small number of days on which the average pattern is calculated, a clear change in average intraday volatility is identifiable around time 15:30, that is, when the New York Stock Exchange opens.

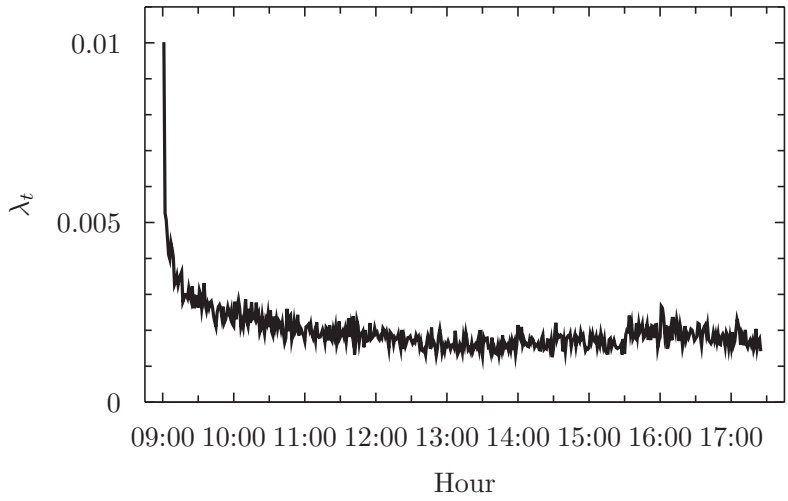

Figure 8.1: The intraday volatility pattern of Monte dei Paschi di Siena.

8.3 Jump identification

After performing the data cleaning procedures explained in Section 8.2, we proceed with the detection of jumps. Following a standard approach, as adopted also in [127], we estimate the local volatility σ and then test whether the ratio between absolute returns and local volatility is above a given threshold, that is, if

$$\frac{|r|}{\sigma} > \theta. \tag{8.1}$$

In order to make a good choice for θ, we performed a sensitivity analysis of the jump detection method with respect to θ, as we will explain in details in Section 8.4.2. Since modelling the dynamics of jumps asks for a well-populated dataset and at the same time we want to minimise the contribution coming from the continuous diffusive component of the process, in this work we choose $\theta = 4$ as a reasonable trade-off.

The return series in Equation (8.1) are obtained from historical prices as reported in Section 8.2, while in order to estimate the local volatility we can follow several strategies available in literature. We base our approach on the realised absolute variation and the realised bipower variation, whose

asymptotic theory is treated for example in [82]. Our estimators of local volatility based on these quantities are defined by the exponentially weighted moving averages

$$\hat{\sigma}_{\mathrm{abs},t} = \mu_1^{-1}\alpha \sum_{i>0}(1-\alpha)^{i-1}|r_{t-i}| \tag{8.2}$$

and

$$\hat{\sigma}^2_{\mathrm{bv},t} = \mu_1^{-2}\alpha \sum_{i>0}(1-\alpha)^{i-1}|r_{t-i}||r_{t-i-1}|, \tag{8.3}$$

where $\mu_1 = \sqrt{\frac{2}{\pi}} \simeq 0.797885$ and the parameter of the exponential averages is $\alpha = \frac{2}{M+1}$, with $M = 60$, which corresponds to a half-life time of nearly 21 minutes. In order to avoid biased estimates due to the presence of jumps in past returns, we actually use modified versions of the estimators (8.2) and (8.3) that use only returns where no jump is detected. In fact the presence of a jump in the estimation window leads to an overestimation of the volatility.

The asymptotic result (6.6) continues to hold when a jump component is added to the continuous process (6.5). However, for finite δ, price jumps are a source of bias in estimation of volatility through the realised bipower variation. A solution to this problem is to use the *threshold bipower variation* (see [92]), which takes into account only returns smaller than a certain threshold. We follow this idea and estimate volatility through returns which are not identified as jumps, that is, returns whose absolute value is not larger than 4 times the local volatility. Using exponentially weighted moving averages instead of flat ones, our estimators are recursively defined by the equations

$$\hat{\sigma}_{\mathrm{abs},t} = \mu_1^{-1}\alpha|r_{t'}| + (1-\alpha)\hat{\sigma}_{\mathrm{abs},t-1},$$

where t' is such that $t' \leq t-1$, $\frac{|r_{t'}|}{\hat{\sigma}_{\mathrm{abs},t'}} \leq 4$ and $\frac{|r_\tau|}{\hat{\sigma}_{\mathrm{abs},\tau}} > 4$ for each $t' < \tau < t$, and

$$\hat{\sigma}^2_{\mathrm{bv},t} = \mu_1^{-2}\alpha|r_{t''}||r_{t'}| + (1-\alpha)\hat{\sigma}^2_{\mathrm{bv},t-1},$$

where t'' and t' are such that $t'' < t' \leq t-1$, $\frac{|r_{t''}|}{\hat{\sigma}_{\mathrm{bv},t''}} \leq 4$, $\frac{|r_{t'}|}{\hat{\sigma}_{\mathrm{bv},t'}} \leq 4$ and $\frac{|r_\tau|}{\hat{\sigma}_{\mathrm{bv},\tau}} > 4$ for each $t'' < \tau < t'$ and for each $t' < \tau < t$. The value taken for the parameter of the exponentially weighted moving average is $\alpha = \frac{2}{M+1}$, with $M = 60$, corresponding to a half-life time of nearly 21 minutes.

Since we construct three series of returns (which differ only in correspondence and proximity of missing observations), for each one we obtain the volatility estimates $\hat{\sigma}_{\mathrm{abs}}$ and $\hat{\sigma}_{\mathrm{bv}}$, we have six volatility-normalised series $\left\{\frac{r_t}{\sigma_t}\right\}$, each of which with its own set of detected jumps. In order not to be sensitive to the choice of how to treat missing observations nor to the choice

of volatility estimation, we take as final series of jumps the intersection of the six sets of jumps. This should also minimise spurious jump detection, as suggested in [114].

8.3.1 Performance of the jump detection methods on simulated data

In Section 8.4 we discuss the results that we have found about identified jumps. Preliminarily, we conduct a simulation study to assess the effectiveness of our entire procedure from data cleaning to jump detection. The usual approach is to assume the observed log-price $p(t)$[3] as coming from the sum

$$p(t) = X(t) + \epsilon(t) \tag{8.4}$$

of an efficient log-price $X(t)$, following a simple jump diffusion model

$$\mathrm{d}X(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t) + \kappa(t)\,\mathrm{d}J(t), \tag{8.5}$$

where the three terms on the right hand side are the drift, diffusion and jump components, and a microstructure noise $\epsilon_t \sim$ i.i.d. $\mathcal{N}(0, \eta^2)$. Following Fan and Wang [88], for the dynamics of the volatility $\sigma(t)$ we take the Geometric Ornstein-Uhlenbeck model

$$\mathrm{d}\log\sigma^2(t) = -(0.6802 + 0.1\log\sigma^2(t))\,\mathrm{d}t + 0.25\,\mathrm{d}W'(t)\,, \tag{8.6}$$

with correlation between the Wiener processes $W(t)$ and $W'(t)$ equal to -0.62. Simulating model of Equations (8.4) and (8.5) we fix the drift term μ equal to 0, and the microstructure noise standard deviation η to $10^{-5}\mathrm{d}t$. As far as the process $J(t)$ is concerned, we set its rate equal to three jumps per day, and the size κ as a multiple of the spot volatility bootstrapped from the sample of the empirical data. Using the Euler discretisation scheme we obtain simulated high-frequency log-prices from the model at a frequency of one minute, which corresponds to the sampling frequency of our data. The time convention that we adopt fixes the one day horizon equal to one, so we have $\mathrm{d}t = 1/1440$. We simulate data for a period 50 times longer than that of the sample data, that is, $88 \times 50 = 4400$ days. We preliminarily draw a Monte Carlo simulation of the process (8.6) and then we rescale it by adding the intraday volatility pattern. The seasonal volatility is used in (8.5) to generate one realisation of the process $X(t)$. Finally, for each day we discard

[3] In this chapter we express the time dependence among parenthesis when we are dealing with continuous time variables, whilst we use indices for discrete time models and sampled data.

some observations in order to reproduce the observed intertrade times, thus introducing artificial missing observations to investigate the role of methods MO1, MO2, MO3 of Section 8.2 in the jump detection. This operation is performed through a random sampling from the empirical distribution of the intertrade times.

Table 8.2 shows the performance of our jump detection procedure for simulations calibrated on two stocks: Intesa Sanpaolo, the second most liquid stock among our data, and Banco Popolare, which has a much lower liquidity. The performance is presented in terms of size and power of the test, for the jump series coming from methods MO1, MO2, MO3, with the two different estimates of volatility $\hat{\sigma}_{\text{abs}}$ and $\hat{\sigma}_{\text{bv}}$. In the last column and row we report the intersections between the different methods. Simulations calibrated on Intesa Sanpaolo show small differences among the three methods of dealing with the missing observations in sampled data. This is expected, since there are relatively few points in which the corresponding return series differ. Also the intersection of jumps detected in methods MO1, MO2, MO3 has performances comparable to those of each of the three methods.

A benchmark simulation calibrated on the stock Banco Popolare shows instead a very different picture, namely it highlights the significance of the way of filling the missing observations for a stock of low liquidity. If such missing observations are filled with zeroes (method MO1), volatility estimates from past returns are lower than for methods MO2 and MO3. This translates into detecting many more jumps, both true and spurious. In absolute terms, average counts of right and false positives are respectively 189 and 143 for method MO1, 146 and 45 for method MO2, 147 and 10 for method MO3, 134 and 7 for the intersection of all methods. Thus, results of this simulation clearly show that taking the intersection of jumps detected in methods MO1, MO2, MO3 drastically reduces the number of false positives, although this is accompanied also with a significant reduction of right positives. Taking a conservative point of view, we are ready to miss the detection of some true jumps as long as this procedure allows minimising false positive hits. This simulation also demonstrates in a very clear way how the method MO3 is effective in detecting much fewer false positives than method MO2, still detecting the same amount of true jumps.

	MO1		MO2		MO3		$\bigcap_{\text{MO1,MO2,MO3}}$	
	Size	Power	Size	Power	Size	Power	Size	Power
σ_{abs}	0.036%	63.1%	0.033%	62.5%	0.029%	62.7%	0.028%	62.1%
σ_{bv}	0.039%	63.4%	0.037%	62.7%	0.033%	62.7%	0.032%	62.3%
$\bigcap_{\text{abs,bv}}$	0.031%	61.6%	0.029%	61.0%	0.026%	61.0%	0.025%	60.6%

	MO1		MO2		MO3		$\bigcap_{\text{MO1,MO2,MO3}}$	
	Size	Power	Size	Power	Size	Power	Size	Power
σ_{abs}	0.369%	72.9%	0.110%	56.4%	0.027%	58.0%	0.019%	52.3%
σ_{bv}	0.390%	74.7%	0.135%	59.7%	0.031%	58.3%	0.025%	55.1%
$\bigcap_{\text{abs,bv}}$	0.324%	71.7%	0.102%	55.2%	0.023%	55.8%	0.016%	50.9%

Table 8.2: Size and power of the jump tests, determined on a simulation calibrated on real data of Intesa Sanpaolo (top) and Banco Popolare (bottom). The size is the ratio between the number of false positives and the number of minutes when the jump is not present. The power is one minus the ratio between the number of false negatives and the number of real jumps.

8.4 Basic statistics of jumps of individual stocks and of systemic jumps

In this section we present some basic statistical properties of detected jumps. In particular we will consider the restricted set of jumps identified simultaneously by all methods, in order to minimise the number of false positives. We then consider some simple statistical characterisation of cojumps, i.e. instances in which the prices of at least two stocks jump in the same minute.

8.4.1 Jumps of individual stocks

In the previous sections we have introduced several methods to identify jumps in price time series. By using Monte Carlo simulations of a realistic model of price dynamics, we have shown also that different methods display different ability to identify jumps correctly, both in terms of false positives and in terms of false negatives. The use of the intersection among the different methods improves significantly the identification procedure, at least on simulated data.

Table 8.3 shows the number of jumps for the stock Monte dei Paschi di Siena, detected with the six different methods. For this specific stock, the

number of jumps ranges from 200 to 281, showing a significant dependence on the used method. The table shows also the size of the intersection of the sets of detected jumps. When one considers jumps detected by all methods the number of jumps falls to 178. This is the restricted set of events that we will consider. We notice that, in this case, by estimating the volatility as $\hat{\sigma}_{\text{abs}}$ we find more jumps than when using $\hat{\sigma}_{\text{bv}}$, regardless of what the method of treating missing observation is. However this is not always the case. For the stock Intesa Sanpaolo it is the other way round: more jumps are detected when we estimate local volatility by $\hat{\sigma}_{\text{bv}}$ than when we use $\hat{\sigma}_{\text{abs}}$, in methods MO1, MO2 and MO3. It is not clear to us whether there are features of the return series (such as quantity and position of the missing observations) which are systematically responsible for the first or the second scenario to happen.

	MO1	MO2	MO3	$\bigcap_{\text{MO1,MO2,MO3}}$
$\hat{\sigma}_{\text{abs}}$	281	228	217	205
$\hat{\sigma}_{\text{bv}}$	260	208	200	190
$\bigcap_{\text{abs,bv}}$	239	196	186	**178**

Table 8.3: Summary table of the number of detected jumps by using different methods and their intersection. The investigated stock is Monte dei Paschi di Siena.

We extend this analysis to the whole set of 20 stocks. In Table 8.4 we report simple statistics about the jumps identified on the 20 stocks by using the intersection of the six methods. The number of jumps detected in the 88 days of our sample varies between 59 and 188 across the twenty stocks (that is, between 0.67 and 2.14 jumps per day per stock), with an overall average value of 108, corresponding to 1.23 jumps per day per stock [4].

Five stocks show a statistically significant difference in the number of positive and negative jumps (at 10% significance level), when one assumes a null model in which stocks have the same probability of jumping up and down. In such a model, the number of jumps in either direction has a binomial distribution. Interestingly, in all five cases the number of negative jumps is greater than that of positive jumps. Statistically significant asymmetry in the jumping direction is also found in the overall jump count (at 1% significance

[4] It is worth noticing that in a recent study [127] authors found more than seven jumps per stock per day when investigating a relatively large set of US stocks. However several reasons might account for this difference, for instance we adopt the intersection of qualitatively different estimators for the realised volatility while they use only one method, they do not detail the data cleaning procedure, and they investigate a much larger universe of stocks.

level). The twenty stocks have also variable proportions of "single jumps", that is, jumps that do not occur simultaneously to jumps of other stocks, with two having even a greater number of cojumps than single jumps. Indeed, the complementary proportion of cojumps attains relatively high values, among 16% and 57%, suggesting that the cojumps play a relevant role in jumps' behaviour and motivating further analyses in this respect.

ISIN	jumps	jumps up	jumps down	single jumps	cojumps
IT0000062072	103	48 (47%)	55 (53%)	53 (51%)	50 (49%)
IT0000062957	63	29 (46%)	34 (54%)	38 (60%)	25 (40%)
IT0000064482	121	60 (50%)	61 (50%)	97 (80%)	24 (20%)
IT0000068525	93	46 (49%)	47 (51%)	56 (60%)	37 (40%)
IT0000072618	127	67 (53%)	60 (47%)	55 (43%)	72 (57%)
IT0001063210	59	28 (47%)	31 (53%)	44 (75%)	15 (25%)
IT0001334587	178	**73** (41%)	**105** (59%)	150 (84%)	28 (16%)
IT0001976403	123	61 (50%)	62 (50%)	76 (62%)	47 (38%)
IT0003128367	188	**81** (43%)	**107** (57%)	107 (57%)	81 (43%)
IT0003132476	155	**66** (43%)	**89** (57%)	95 (61%)	60 (39%)
IT0003487029	70	28 (40%)	42 (60%)	41 (59%)	29 (41%)
IT0003497168	129	74 (57%)	55 (43%)	79 (61%)	50 (39%)
IT0003856405	95	50 (53%)	45 (47%)	74 (78%)	21 (22%)
IT0004176001	74	41 (55%)	33 (45%)	46 (62%)	28 (38%)
IT0004231566	103	50 (49%)	53 (51%)	72 (70%)	31 (30%)
IT0004623051	115	**47** (41%)	**68** (59%)	85 (74%)	30 (26%)
IT0004644743	100	51 (51%)	49 (49%)	65 (65%)	35 (35%)
IT0004781412	118	**49** (42%)	**69** (58%)	57 (48%)	61 (52%)
LU0156801721	59	27 (46%)	32 (54%)	32 (54%)	27 (46%)
NL0000226223	86	39 (45%)	47 (55%)	51 (59%)	35 (41%)
total	2159	**1015**	**1144**		
average	108.0	50.8 (47%)	57.2 (53%)		

Table 8.4: Number of detected jumps for the twenty stocks, with direction and cojumping information. Bold values are inconsistent (at 10% significance level for single stocks, at 1% significance level for the total counts) with a null assumption of equal probability of jumping up and jumping down. Single jumps occur at a time when no other stock jumps, while cojumps occur simultaneously with jumps in other stocks.

We investigate the distribution of jumps during the trading day. The left panel of Figure 8.2 shows the histogram of the time of the day when a jump occurs. The figure is obtained by including all the jumps of the 20 stocks, but we count only once a minute where multiple stocks jump simultaneously. We

observe that there is no clear periodicity in the number of jumps, indicating that our intraday pattern removal is quite effective. Moreover, apart a spike observed at the beginning of the day, there is no evidence of minutes of the day when it is more likely that one stock jumps.

8.4.2 Systemic jumps

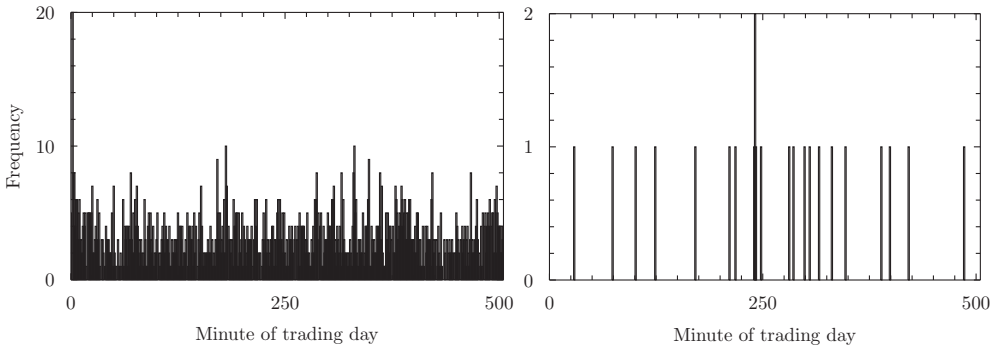

Figure 8.2: Left panel: intraday distribution of the time of the day when a jump occurs. The x axis is the number of minutes from the beginning of the trading day. Notice that if more than one stock jump in a given minute, we count the minute only once in the histogram. This is done in order to avoid to count N times an event when N stocks jump simultaneously. Right panel: same figure as the left panel, but considering only cojumps of more than 5 stocks.

The main topic of this chapter is systemic cojumps, i.e. minutes when the price of a (possibly large) number of stocks displays a jump. For this reason it is important to investigate how frequently cojumps occur and to compare the observed statistics with those expected under some null hypotheses.

We start with some simple visualisation of the occurrences of multiple cojumps. Figure 8.3 shows the dynamics of the number of cojumps, indicating also the number of stocks that display a jump in a given minute. We notice that there are several occurrences where more than 8 stocks jump simultaneously (big circles). For example, we observe one case each when 10, 12, 13, 15, 16, 17, and 20 stocks jump simultaneously. Also the number of cases where more than 3 stocks jump (medium-sized circles) is quite high. For example, we observe 22 cases with 4 stocks, 15 cases with 5, 2 cases with 6, 4 cases with 7, 5 cases with 8, 3 with 9, and 2 with 11. Finally, there is a significant background of cases where two (136) or three (44) stocks jump simultaneously. Thus multiple stock cojumps are relatively frequent.

By considering all the 240 events in which multiple stocks jump, we find only 7 cases in which not all the jumping stocks follow the same direction. There are 6 cases in which 2 stocks jump in opposite direction and 1 case in which two stocks jump down and one stock jumps up. The fact that when several stocks simultaneously jump they all move in the same direction suggests that a single common factor explains the jumping probability. In Section 8.7 we will develop this idea more formally by introducing models that capture this important feature of real data.

It is worth noticing that the Figure 8.3 also indicates that there are no specific times of the day (for example corresponding to pre-announced news or opening of other markets) where the systemic cojumps are more frequent. This qualitative statement can be made more precise by drawing the histogram of the time of the day when a systemic cojump with more than 5 stocks jumping simultaneously (see right panel of Figure 8.2). Therefore systemic jumps do not occur at preferential moments of the trading day.

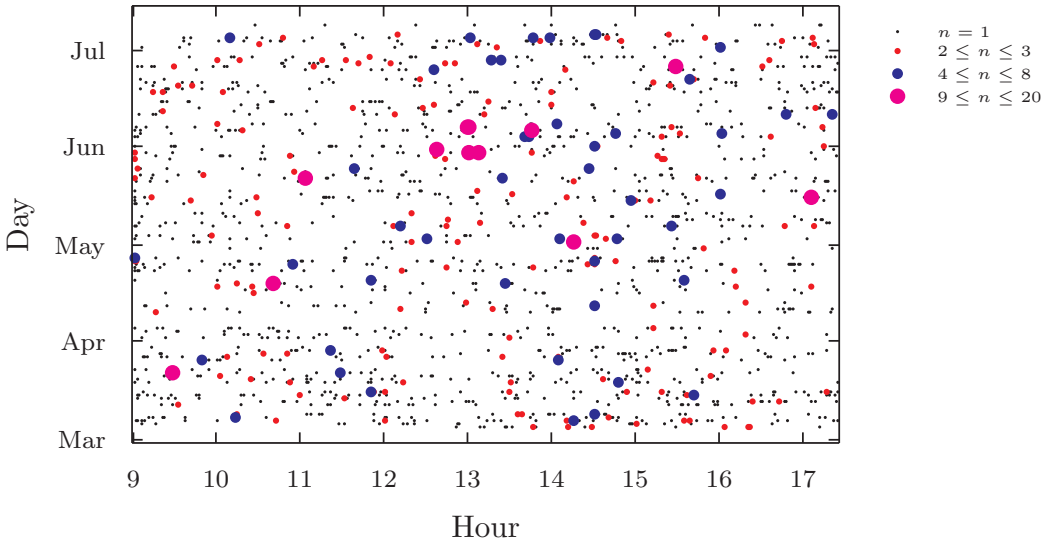

Figure 8.3: Time series of the cojumps observed in the set of 20 investigated stocks. The horizontal axis is the time of the day and the vertical axis is the day. The size of the circle codes the number of stocks simultaneously jumping in a given minute.

In order to compare the pattern observed in Figure 8.3 for the systemic cojumps with a null hypothesis, we perform a bootstrap analysis. Specifically, we construct a bootstrap replica independently for each stock. Therefore our replicas are consistent with a model of 20 independent but not identical

Poisson processes for the jumps. In the next section we will use this model as the simplest benchmark model. In the left panel of Figure 8.4 we show the analogous of Figure 8.3 for one of the bootstrap replicas of the real data. As it can be seen, in this replica there are no cases where more than three stocks jump simultaneously, a result that is clearly inconsistent with the real data. To be more quantitative, in the right panel of Figure 8.4 we show the histogram of the number of stocks jumping in a minute (in which at least one stock jumps) for real data (solid line). We compare it with the curve obtained by taking the 0.01% confidence interval from bootstrap analysis. It can be seen that already the observed number of cojumps of two stocks is incompatible with the one observed in bootstrap test at the 0.01% confidence. Moreover in $10,000$ bootstrap replicas we never observe a cojump with more than 4 stocks (observed only once) while in real data we have several cases with the price of many stocks simultaneously jumping. Since the focus of our analysis relies on the deviation from the null hypothesis of independence, we investigate if the results in Figure 8.4 are sensitive to the particular choice of the jump threshold θ by exploring the range from 4 to 8. In Figure 8.5 we show, for several values of θ, the cumulative versions of the distributions present in the right panel of Figure 8.4. We find that, while the number of detected jumps obviously varies in size, the Poisson null is systematically rejected for all the explored values.

In conclusion, our descriptive analysis shows that (i) jumps are relatively frequent, even when one considers relatively strict detection criteria, (ii) there is a large number of cojumps, i.e. minutes when a sizeable number of stocks (up to 20!) simultaneously jump, (iii) these cojumps show no clear timing inside the day, and, more important, (iv) they are absolutely not compatible with a null model of independent but not identical Poisson processes. In Section 8.7 we will introduce models able to describe the jump dynamics of a set of stocks and we will keep the jump threshold $\theta = 4$ for the rest of our analysis to preserve the quality of the statistics.

8.5 A multi-scale statistical test based on multiple jumps and cross jumps detection

The empirical observation of a large number of cojumps requires a rigorous statistical test to compare the observed behaviour of jumps with the prediction of multivariate point processes. The main problem in working with jumps is that their number is relatively small, and therefore the statistics must be suited to work on small samples. Moreover, we want to use a multi-

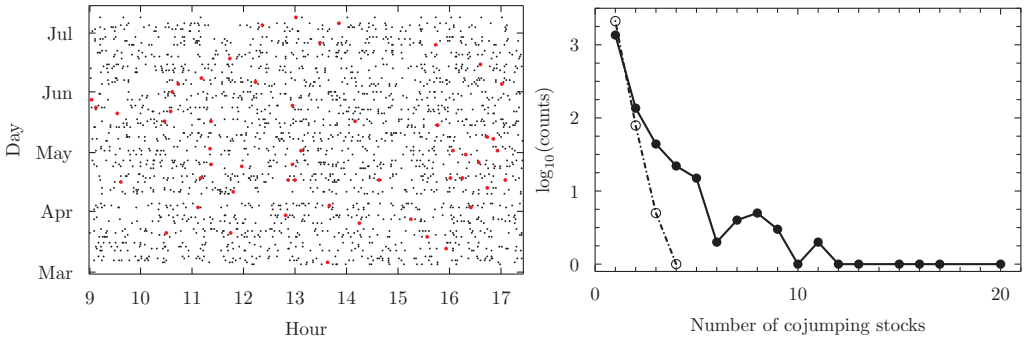

Figure 8.4: Left panel. Time series of the cojumps observed in a bootstrap replica of the real data preserving the jump intensity of each stock. The horizontal axis is the time of the day and the vertical axis is the day. The size of the circle codes the number of stocks simultaneously jumping in a given minute under the same convention of Figure 8.3. Right panel. Histogram of the number of stocks simultaneously cojumping in a minute (solid line). For comparison the dashed line shows the 0.01% confidence interval for the counts under the null hypothesis of independent but not identical Poisson processes.

scale statistical test, i.e. a test that is able to identify deviations from models at different time scales. If the observation time is much larger than the correlation time of the process and we only focus on the counting of events disregarding the interarrival times, then a correlated point process is undistinguishable from a Poisson process. Therefore we use a test that considers simultaneously different time scales. Finally, even if the main topic of this chapter is cojumps of different stocks, we are also interested in investigating deviation from a Poisson behaviour at the individual stock level. Therefore we shall introduce two related statistics, one for consecutive jumps of the same asset and one for jumps of different stocks occurring at close times.

Since the distributional and time correlation analysis is unfeasible with small samples of jumps we will consider the following statistics that measures the frequency of multiple jumps occurring in the same time windows.

Specifically, when considering individual stocks, we define a *multiple jump* (MJ) as an event which occurs when *at least two jumps of the same stock price* are observed inside a time window of a fixed length w. Let us call s_i the number of jumps inside the i-th window. An estimator of the MJ probability

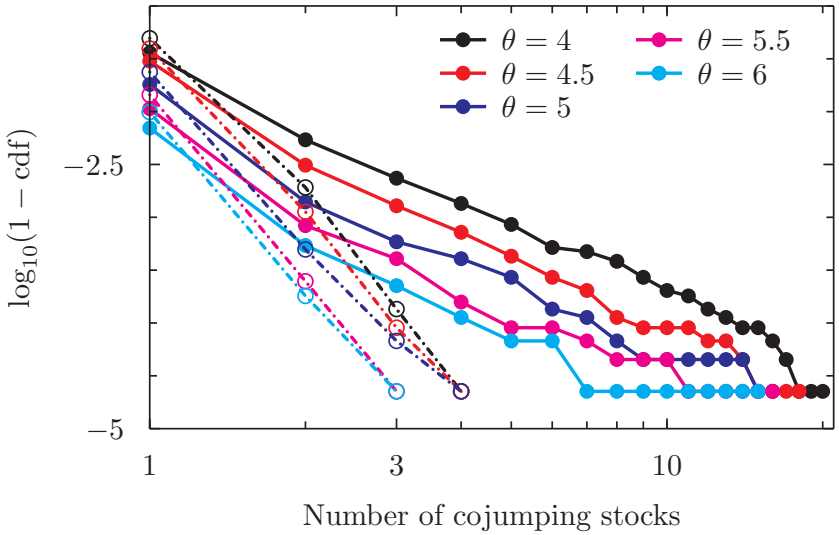

Figure 8.5: Cumulative distribution of the number of stocks simultaneously cojumping in a minute as a function of the jump threshold θ, in the range from 4 to 6; dashed lines correspond to the 0.01% confidence interval for the cumulatives, under the Poisson null hypothesis. Since the finite size of our sample (88 days) worsens the quality of the statistics, we omit the results for θ larger than 6.

in a window of length w over a sampling period of length N is given by

$$\hat{p}_w^{\mathrm{MJ}} = \frac{\sum_{i=1}^{\lfloor \frac{N}{w} \rfloor} \mathbf{1}_{s_i \geq 2}}{\lfloor \frac{N}{w} \rfloor}, \tag{8.7}$$

where $\mathbf{1}_A$ is the indicator function of the event A, and the symbol $\lfloor N/w \rfloor$ corresponds to the integer part of the ratio N/w.

The same idea can be extended to capture also the cross sectional clustering of jumps, in particular the evidence of a large number of simultaneous jumps of different stocks. The second notion which we will work with is therefore that of *cross jumps* (CJ) between two stocks defined when *both stocks jump at least once* inside a given time window of fixed length w. With the same notation as before, the estimator of the CJ probability between stock l and k is given by

$$\hat{p}_w^{\mathrm{CJ}} = \frac{\sum_{i=1}^{\lfloor \frac{N}{w} \rfloor} \mathbf{1}_{s_i^l \geq 1} \mathbf{1}_{s_i^k \geq 1}}{\lfloor \frac{N}{w} \rfloor}. \tag{8.8}$$

Our statistical procedure consists in estimating these quantities in real data and compare them, at each time scale, with the 99% and 95% confidence bands of the tested model obtained analytically (when possible) or via Monte Carlo simulations. It should be noted that, since the number of windows that we are considering is greater than one, we are indeed performing a multiple hypothesis test and we have to correct the significance level accordingly. Among the possible approaches discussed in literature, we decide to adopt the most conservative one, namely the Bonferroni correction. This correction amounts to divide the significance level of the single hypothesis by the total number of tested hypotheses in order to achieve a global significance at least of the pre-fixed level. Therefore when the empirical points of $\hat{p}_w^{\mathrm{MJ}}$ and $\hat{p}_w^{\mathrm{CJ}}$ fall inside the confidence bands of a given model we cannot reject the null hypothesis with the given confidence level.

8.5.1 A benchmark case: the Poisson model

In order to show how our testing procedure works, we consider here an important benchmark case, in which the jumps of each stock are described by an independent Poisson process. Under this model, the mean and variance of both previous estimators can be computed. For the MJ estimator we have

$$\mathbb{E}[\hat{p}_w^{\mathrm{MJ}}] = p_{w,\lambda}\,, \qquad \mathrm{Var}[\hat{p}_w^{\mathrm{MJ}}] = \frac{p_{w,\lambda} - p_{w,\lambda}^2}{\lfloor \frac{N}{w} \rfloor}\,,$$

where $p_{w,\lambda} = \mathbb{P}(\{s \geq 2\}) = 1 - \mathrm{e}^{\lambda w}(1 - \lambda w)$, and λ is the intensity of the Poisson process.

Analogously for the CJ estimator, we obtain

$$\mathbb{E}[\hat{p}_w^{\mathrm{CJ}}] = q_{w,\lambda_l} q_{w,\lambda_k}\,, \qquad \mathrm{Var}[\hat{p}_w^{\mathrm{CJ}}] = \frac{q_{w,\lambda_l} q_{w,\lambda_k} - q_{w,\lambda_l}^2 q_{w,\lambda_k}^2}{\lfloor \frac{N}{w} \rfloor}\,,$$

where $q_{w,\lambda_i} = \mathbb{P}(\{s^i \geq 1\}) = 1 - \mathrm{e}^{\lambda_i w}$.

Since both quantities in Equation (8.7) and (8.8) correspond to the sum of a large number of indicator functions, the Central Limit Theorem implies that their distribution is well approximated by a Normal law whose p-values are readily available, allowing the analytical computation of the confidence bands. As usual, the value of the intensity of the Poisson process given by the maximum likelihood estimator is $\hat{\lambda} = \#\mathrm{jumps}/N$.

In Figure 8.6 we show the result of our test on the Italian asset Assicurazioni Generali, for which $\lambda = 2.4 \times 10^{-3}\ \mathrm{min}^{-1}$. The filled circles correspond to the empirical values of the estimator, the solid line to the theoretical mean,

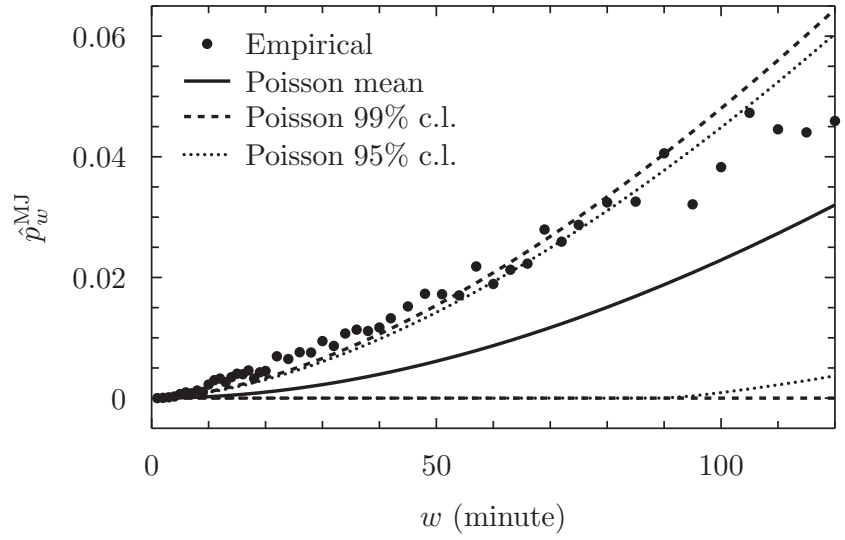

Figure 8.6: MJ probability test under Poisson null for the Italian asset Assicurazioni Generali.

while the dashed and dotted lines establish the boundaries of the 99% and 95% confidence bands, respectively, adjusted with the Bonferroni correction. The figure shows that the Poisson model is rejected for both levels. In our data sample this is the typical situation, which occurs for 18 stocks out of 20 (the two exceptions are represented by the assets Mediobanca and Finmeccanica). We therefore conclude that there is a strong evidence of time clustering of jumps and violation of the univariate Poisson model.

8.6 Modelling jumps with Hawkes processes

At this stage, the natural step to proceed is to assume a more sophisticated model for the jump process, and to test it as an alternative hypothesis. The Poisson hypothesis can be relaxed in several respects: among the possible alternatives we can weaken the assumption of identically and independently distributed waiting times, for example modelling them either as Markov processes or as realisations of a time inhomogeneous Poisson process. For instance the extension of the Poisson process that we consider in this study is the class of point processes known has *Hawkes processes*, where the intensity is itself stochastic and tends to increase when a new jump arrives.

8.6.1 Univariate case

In this section we provide the main results needed for the remaining of the chapter, but for a complete mathematical treatment of Hawkes and more general point processes we refer the reader to the comprehensive textbook [128].

A univariate point process $N(t)$ [5] is called a Hawkes process if it is a *linear self-exciting process*, defined by the intensity

$$I(t) = \lambda(t) + \int_{-\infty}^{t} \nu(t-u)\mathrm{d}N(u) = \lambda(t) + \sum_{t_i<t} \nu(t-t_i)\,, \tag{8.9}$$

where λ is a deterministic function called the base intensity, ν is a positive decreasing weight function, and t_i are the jumping times. The most common parametrisation of ν is given by $\nu(t) = \sum_{j=1}^{P} \alpha_j \mathrm{e}^{-\beta_j t}$, for $t > 0$, where $\alpha_j \geq 0$ are scale parameters, $\beta_j > 0$ control the strength of decay, and the positive integer P is the order of the process. A particular advantage of the linear Hawkes process of order $P = 1$ is that the log-likelihood function can be computed as

$$\mathcal{L}(t_1, \ldots, t_n) = (1-\lambda)t_n - \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(1 - \mathrm{e}^{-\beta(t_n - t_i)}\right) + \sum_{i=1}^{n} \ln(\lambda + \alpha R_i)\,,$$

where the R function satisfies the recursion $R_i = \mathrm{e}^{-\beta(t_i - t_{i-1})}(1 + R_{i-1})$ for $i \geq 2$ and $R_1 = 0$. In Ref. [40] it is shown that the stationarity of the process is guaranteed when $\int_0^\infty \nu(s)\mathrm{d}s < 1$, which in our case reduces to the requirement $\alpha/\beta < 1$. If stationarity holds and under the further constraint that the base intensity is constant, the expected number of jumps in an arbitrary time interval of length T is given by $\lambda T/(1-\alpha/\beta)$. The latter observation will be useful for the calibration of the factor models that we will discuss in Section 8.7. The characterisation of the asymptotic properties of the the maximum likelihood estimator of the Hawkes process parameters that we employ in this study has been provided in [129] and [130], while the simulation algorithm we use is based on the procedure discussed in [131], which directly derives from the Shedler-Lewis thinning algorithm, [132].

We estimate the Hawkes processes on the univariate series of jumps of the investigated stocks. In Table 8.5 we report the parameters values with the associated errors and significance values for Generali and Intesa Sanpaolo. All the values are statistically significant. We then test the Hawkes model in its

[5] If $\{t_i\}_{i=1,\ldots,n}$ represents the random sequence of increasing event times $0 < t_1 < \ldots < t_n$ associated with the point process, then $N(t) \doteq \sum_{i\geq 1} \mathbf{1}_{t_i \leq t}$ defines the right continuous counting function. In what follows we will refer equivalently to the process and its counting function.

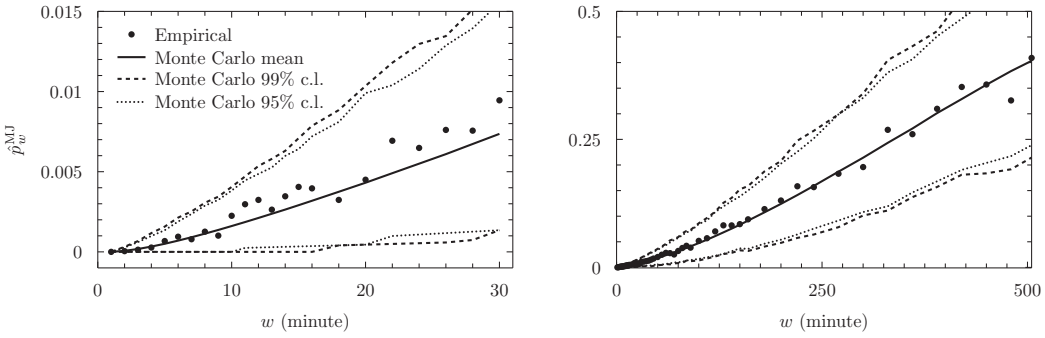

Figure 8.7: MJ probability test under Hawkes null for the assets Assicurazioni Generali (left) and Intesa Sanpaolo (right).

	Assicurazioni Generali	Intesa Sanpaolo
λ (min^{-1})	$(2.1 \pm 0.2) \times 10^{-3}$ ***	$(2.5 \pm 0.2) \times 10^{-3}$ ***
α (min^{-1})	$(3.1 \pm 1.3) \times 10^{-2}$ *	$(5.9 \pm 2.1) \times 10^{-2}$ **
β (min^{-1})	$(2.5 \pm 0.9) \times 10^{-1}$ **	$(4.3 \pm 1.2) \times 10^{-1}$ ***

Table 8.5: List of the parameters of the one dimensional Hawkes processes. Significance codes: $p_{\text{value}} < 0.001$ '***', $0.001 \leq p_{\text{value}} < 0.01$ '**', $0.01 \leq p_{\text{value}} < 0.05$ '*', and $p_{\text{value}} \geq 0.05$ ' '.

ability of reproducing the MJ probability at different time scales. The left and right plots in Figure 8.7 are obtained after the calibration of a one dimensional Hawkes process on the jump time series of the assets Assicurazioni Generali and Intesa Sanpaolo, respectively. The figure shows that we can not reject the null both at 1% and at 5% significance levels (obtained from $N_{MC} = 10^4$ Monte Carlo simulations). The result holds both for short window lengths and for longer horizons, an example of the former case is given by Generali for time windows ranging from one minute up to half an hour, while for the latter case we consider Intesa Sanpaolo with horizons running up to one day. In the sections which follow we present the result of the statistical tests only for short horizons, since they usually correspond to the time scales where the most interesting effects take place. However, an analysis extended to longer horizons would have been equally effective.

We therefore conclude that univariate Hawkes processes are able to capture the empirical dynamics and time clustering of jumps for individual

stocks.

8.6.2 Bivariate case

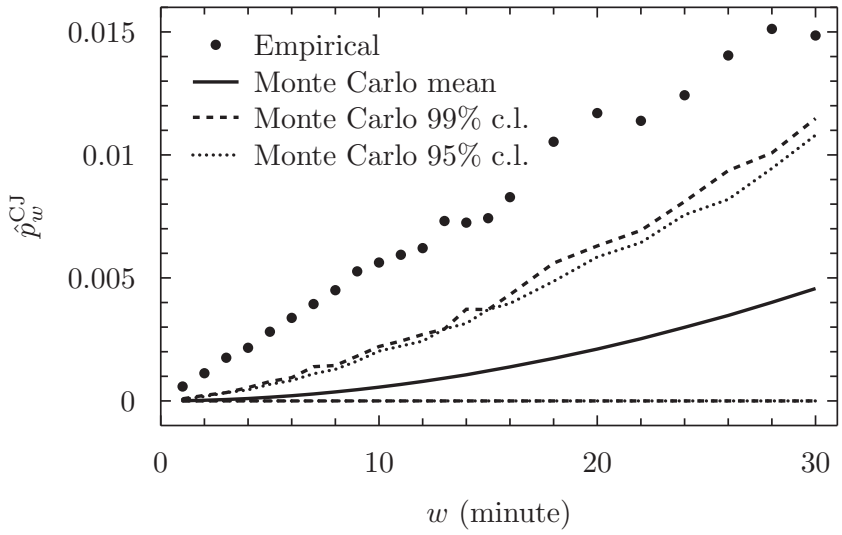

Figure 8.8: CJ probability test under independent Hawkes null for the pair of Italian assets Generali - Intesa Sanpaolo.

Are independent Hawkes processes able to describe the empirical behaviour of CJ probabilities? In order to answer this question, we compute the estimator (8.8) on each pair of assets and in Figure 8.8 we show an example of the results from the test computed on the pair Generali - Intesa Sanpaolo. The figure shows that the independent Hawkes process miserably fails in describing the CJ probability at all time scales. This is clearly due, at least in part, to a lack of coupling between the two processes. In order to capture such a dependence we calibrate a bivariate Hawkes process.

A K-dimensional Hawkes process is a linear self-exciting process defined by the multivariate intensity $\boldsymbol{I}(t) = \big(I^1(t), \dots, I^K(t)\big)'$, where the k-type intensity with an exponential kernel of order one is given by

$$I^k(t) = \lambda^k(t) + \sum_{m=1}^{K} \sum_{t_i^m < t} \alpha_{km} \mathrm{e}^{-\beta_{km}(t - t_i^m)} \,.$$

All the parameters which appear in the above expression are strictly positive, and the stationarity of the process is guaranteed if the spectral radius of the

matrix $\Gamma = \left(\frac{\alpha_{km}}{\beta_{km}}\right)_{k,m=1,\dots,K}$ is strictly smaller than one. The parameters α_{kk} and β_{kk} are responsible for the self-exciting property of the point process, while the remaining $2K(K-1)$ α_{km} and β_{km} capture the cross exciting effect of a jump in the stock m on the process of the asset k. When $K = 2$ the number of free parameters is equal to ten, and the maximisation of the likelihood becomes less trivial than the univariate case. To maximise the likelihood, we initialise the parameters with the values suggested by the one dimensional calibration, and we constrain their value to remain strictly positive. We preliminarily perform $N_{Ann} = 100$ searches of the maximum with the simulated annealing algorithm, then, with the optimal candidate supplied by the stochastic search, we initialise the deterministic search via conjugate gradient.

λ^1 (min^{-1})	$(2.0 \pm 0.1) \times 10^{-3}$ ***	λ^2 (min^{-1})	$(2.3 \pm 0.2) \times 10^{-3}$ ***
α_{11} (min^{-1})	$(1.6 \pm 0.9) \times 10^{-2}$ *	α_{21} (min^{-1})	$(4.5 \pm 0.1) \times 10^{-4}$ ***
α_{12} (min^{-1})	$(14 \pm 0.1) \times 10^{-4}$ ***	α_{22} (min^{-1})	$(3.4 \pm 1.3) \times 10^{-2}$ **
β_{11} (min^{-1})	$(3.7 \pm 1.6) \times 10^{-1}$ *	β_{21} (min^{-1})	$(7.0 \pm 18) \times 10^{-1}$
β_{12} (min^{-1})	$(3.2 \pm 3.4) \times 10^{-1}$	β_{22} (min^{-1})	$(4.9 \pm 1.6) \times 10^{-1}$ ***

Table 8.6: List of the parameters of the bivariate Hawkes process for Generali and Intesa Sanpaolo (labelled with the indices 1 and 2, respectively). Significance codes: $p_{\text{value}} < 0.001$ '***', $0.001 \leq p_{\text{value}} < 0.01$ '**', $0.01 \leq p_{\text{value}} < 0.05$ '*', and $p_{\text{value}} \geq 0.05$ ' '.

In Table 8.6 we present the results of the calibration of a bi-dimensional Hawkes on the joint jumps process of Generali-Intesa. We then use Monte Carlo simulations to compute confidence bands for CJ probability as shown in Figure 8.9. It is clear from the figure that bivariate exponential Hawkes fails to describe the cross sectional clustering of jumps observed in real data. This might be due to several facts: the value of the cross exciting constants α_{12} and α_{21} is at least of one order of magnitude smaller than that of the constants α_{11} and α_{22}. Moreover, the time constants β_{12} and β_{21} associated with the cross excitations are poorly statistically significant. One possible reason could be the unreliability of the parameters since the optimisation procedure is performed in a high dimensional space on a strongly non-linear objective function. However, in this respect, we perform numerous experiments in a Monte Carlo framework. One test is the simulation of a bivariate Hawkes process, whose parameters are chosen equal to the values of Table 8.6. On

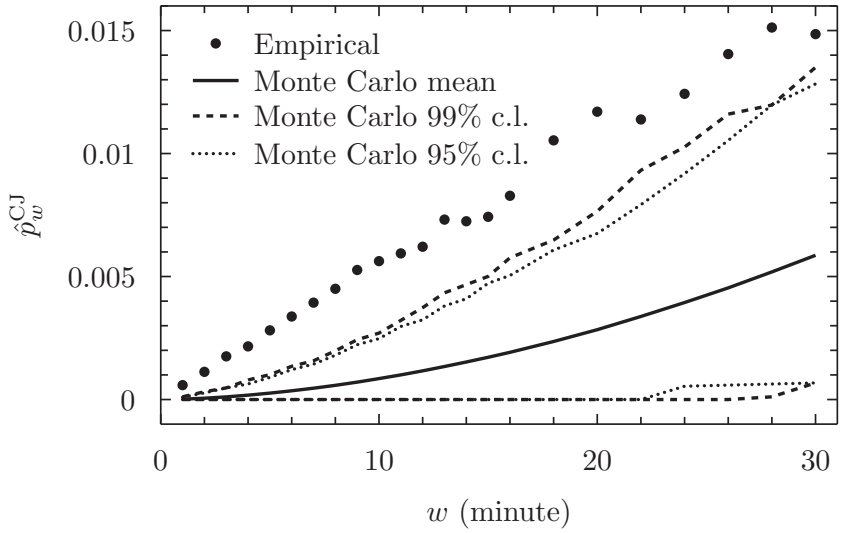

Figure 8.9: CJ probability test under bivariate Hawkes null for Generali - Intesa Sanpaolo.

each random realisation, we apply the above optimisation scheme and record the optimal values. With a statistics of 10^3 values per parameter, we measure no significant bias induced by the multidimensional maximisation procedure. Moreover, we simulate 10^3 independent pairs of univariate Hawkes of length 44440 minutes, whose parameters are given in Table 8.5. We then perform the ten dimensional optimisation on each copy, and find a confirmation of the tendency of the cross-exciting constants to significantly decrease. In light of these numerical evidences, we tend to marginalise the role played by the optimisation scheme. Finally, we believe that the main reason why Hawkes processes fail to describe the cross-sectional dependence is that they are not designed by construction to capture synchronous effects, which in the case under our consideration seem to dominate the dependence between jumps.

8.7 A factor model approach to systemic jumps

In this section we abandon the idea of an N dimensional point process, and we develop a different approach, which we will refer to as a jump factor model. The intuition which drives the modelling is that a stock jumps both because it is triggered by jumps of a market common factor and because of an idiosyncratic term.

The first issue that we want to clarify is if, even in a Poisson framework,

the factor mechanism is able to describe the empirical dependence structure. Then we will propose a scalable model, which ideally should be effective and also sufficiently simple and robust.

8.7.1 Bivariate Poisson factor model

In order to illustrate the main idea, we start with a toy model which enlightens the role of the market factor and clarifies the mechanism which generates the dependence structure. However, this is an unrealistic model, where the behaviour of the assets is completely determined by the evolution of the factor process.

We assume that there is one unobserved market factor point process. When the factor jumps, the stock S_1 jumps with probability p_1 and the stock S_2 jumps with probability p_2. If the factor does not jump, neither the first nor the second stock can jump. Obviously the converse is not true. If the market factor is described by a Poisson process of intensity λ_F, then the expected number of factor's jumps over the period T is given by $\lambda_F T$. In order to fix the free parameters of the model, λ_F, p_1, and p_2, we use the following relations

$$\begin{aligned} p_1\ \lambda_F T &= n_1\,, \\ p_2\ \lambda_F T &= n_2\,, \\ p_1\ p_2\ \lambda_F T &= n_{12}\,, \end{aligned} \tag{8.10}$$

where n_1 and n_2 are the observed number of jumps of the stock S_1 and S_2, respectively, while n_{12} represents the observed number of cojumps among S_1 and S_2 within the one minute sampling interval. The first two equations thus require that the expected number of jumps of the single assets matches the realised values. The latter relation guarantees that, on average, the realised number of cojumps among the assets corresponds to the theoretical expectation. The system of Equations (8.10) can be inverted and provides a direct way to express the parameters of the model in terms of the observable quantities T, n_1, n_2, and n_{12}:

$$\lambda_F = \frac{n_1 n_2}{n_{12}} \frac{1}{T}\,, \quad p_1 = \frac{n_{12}}{n_2}\,, \quad \text{and} \quad p_2 = \frac{n_{12}}{n_1}\,.$$

In Figure 8.10 we show the result of the test of the cojumps estimator $\hat{p}_w^{\mathrm{CJ}}$ against a null represented by the Poisson factor model that we have just described. The time period that we consider corresponds to the usual interval of 88 days, during which Generali and Intesa Sanpaolo jump 103 and 127 times, respectively, while the realised number of cojumps n_{12} is equal to

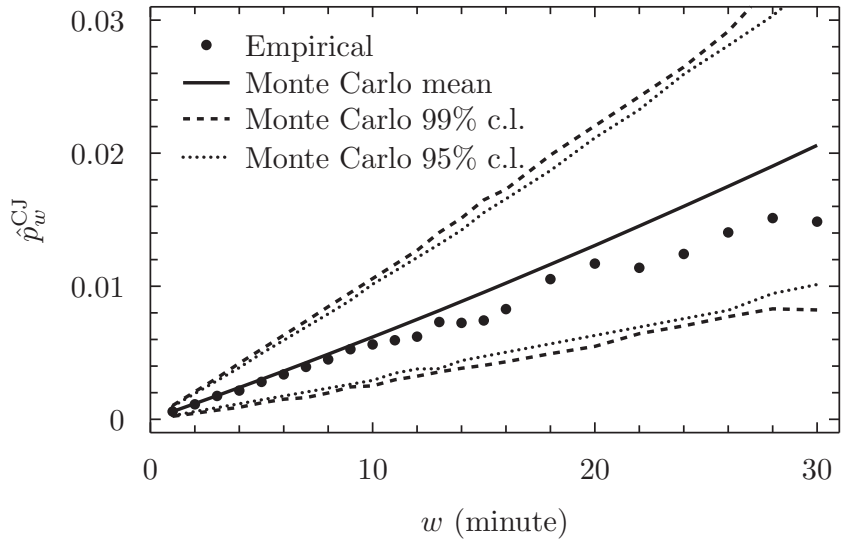

Figure 8.10: CJ probability test under Poisson factor null for Generali - Intesa Sanpaolo.

26. From these values we obtain a Poisson intensity equal to $\lambda_F = 1.1 \times 10^{-2}$ per minute, and probabilities $p_1 = 0.20$ and $p_2 = 0.25$. The confidence intervals for the null are estimated drawing $N_{MC} = 10^4$ paths from the Poisson factor process, and then thinning each realisation with Bernoulli variables of probability p_1 and p_2.

The results provided by this simple model are very satisfactory and should convince the reader that the proposed mechanism is adequate to capture the cross dependence between jumps. However, the model is quite unrealistic in several respects: (i) the assets can not jump independently of the factor, (ii) the Poisson nature of the process leads to a severe underestimation of the realised number of multiple jumps of the same stock, and (iii) the approach has not a straightforward extension to an arbitrary number of assets. In order to amend all these drawbacks, in the next section we will discuss a more general and flexible model, rooted on the factor idea, but able to achieve a higher level of realism and scalability.

8.7.2 N dimensional Hawkes factor model with idiosyncratic components

We now consider a set of N assets. With respect to the previous model, we need to introduce explicitly a proxy of the common factor process. The naif

idea of a proxy of the factor based on those events when a cojump among all the N stocks happens realistically does not work. In the data sample of twenty stocks that we consider in the current study, we experience just a single event when all the stocks jump simultaneously. The proxy that we propose is characterised by a counting function which increases by one unit whenever we detect a cojump which involves a group of assets whose number J conflicts with the null of independence. We therefore need to identify the threshold value $\bar{J}$ compatible with independence. In general it will be time dependent, $\bar{J}_t$, and will vary between one and N. In order to detect $\bar{J}$ we propose a rigorous statistical methodology. However, we preliminarily want to convince the reader that the range of variability of $\bar{J}$ for the case under study $N = 20$ is quite narrow. First of all, we can exclude the case $\bar{J} = 1$, since otherwise we would consider every jump that occurs on the market as potentially systemic. On the other side, if we fix $\bar{J}$, by definition then we would consider the events which involve $J < \bar{J}$ jumps as likely occurrences. As shown by the bootstrap experiment in Section 8.4.2, the probability associated with events involving more than two stocks is extremely low, and this would fix $\bar{J} = 3$. However, the null represented by the bootstrapped data is too extreme, since it does not take into account any effect of self excitation. Heuristically, we conclude that $\bar{J} \simeq 4$.

We now present a rigorous procedure that supports the above heuristic argument. To each stock in our sample we associate a finite number of event times $\{t_i^s\}$, where $s = 1, \ldots, N$ and $i = 1, \ldots, n_s$ label the assets and the asset idiosyncratic jumps, respectively. In equivalent terms, each asset is characterised by a counting function $N^s(t)$, and $n_s = N^s(T)$ is the observed number of jumps of the stock s. According to the procedure of jumps' identification, the t_i^s's are measured in minutes and take only integer values running from one to T. A counting process $N^s(t)$ is associated with an intensity function $I^s(t)$ defined by

$$\mathbb{P}\left(N^s(t) \text{ has a jump in } [t, t + \Delta t] | \mathcal{F}_t\right) = I^s(t)\Delta t\,,$$

for $\Delta t \to 0^+$, where $\mathbb{P}$ stands for the probability and $\mathcal{F}_t$ corresponds to the history of the process up to present time t. We then assume that the counting processes which describe our assets correspond to N independent univariate Hawkes processes and we estimate the parameters which characterise the intensities $I^s(t)$ for $s = 1, \ldots, N$ via maximum likelihood. We fix Δt equal to one minute and via Equation (8.9), for each $t = 1, \ldots, T$, we can compute the vector of probabilities $\boldsymbol{\pi}_t = \left(\pi_t^1, \ldots \pi_t^N\right)' = \left(I_t^1 \Delta t, \ldots, I_t^N \Delta t\right)'$. We test if the number of jumps that we observe at time t is compatible with the cross independence among the processes. Under the null hypothesis, the discrete

probability of the event $J_t = j$ reads

$$\mathbb{P}\left(J_t = j\right) = \sum_{1 \leq l_1 < \ldots < l_j \leq N} \pi_t^{l_1} \ldots \pi_t^{l_j} \prod_{k \in \{1,\ldots,N\} \setminus \{l_1,\ldots,l_j\}} \left(1 - \pi_t^k\right) .$$

Since we repeat the test T times, we adjust the significance level with the Bonferroni's correction. If at time t^F we reject the null, we attribute the event to a systemic shock, and we remove it from the set $\{t_i^s\}$. The procedure has to be iterated as many times as required in order to remove all the systemic jumps.

The advantages of the above procedure are manifold. Specifically, the set of events $\{t_i^F\}$ with $i = 1, \ldots, n_F$ identifies the n_F jumps of the common factor, but not less important each reduced set $\{t_{i'}^s\}_{i'=1,\ldots,n'_s} = \{t_i^s\}_{i=1,\ldots,n_s} \setminus \{t_j^F\}_{j=1,\ldots,n_F} \subseteq \{t_i^s\}_{i=1,\ldots,n_s}$ for $s = 1, \ldots, N$ corresponds in a natural way to the set of the n'_s jumps of the s-th idiosyncratic component. The latter result represents a second major improvement with respect to the bivariate model. Moreover, the entire procedure is easy scalable, since it does not depend critically on the dimension of the portfolio. Last but not least, we are not a priori fixing the nature of the point processes which describe the factor and the idiosyncratic components. We can apply the log-likelihood approach described in 8.6 to the univariate sequence of the event times, fix the value of the parameters of the factor, λ_F, α_F and β_F, and of the idiosyncratic components, λ_s, α_s and β_s, for $s = 1, \ldots, N$, estimate the associated p-values, and, eventually, if it is the case, reject the Hawkes process in favour of the Poisson description.

To sum up, the multivariate model describes the extreme events which occur in a portfolio of stocks as a superposition of systemic shocks, propagating to the s-th asset with probability p_s, and jumps specific of the single assets. In order to estimate the vector of probabilities $\boldsymbol{p} = (p_1, \ldots, p_N)'$, we replace the system of equations given by (8.10), with the new relations

$$\begin{aligned} p_1 \, \frac{\lambda_F}{1 - \alpha_F/\beta_F} T &= n_1 - n'_1 \, , \\ &\vdots \\ p_N \, \frac{\lambda_F}{1 - \alpha_F/\beta_F} T &= n_N - n'_N \, , \end{aligned}$$

where $\lambda^F T/(1 - \alpha_F/\beta_F)$ corresponds the average number of factor's jumps. The above equations force the expected number of cojumps among the factor and the s-th stock to balance the realised number of shocks.

We test our model on the data set of twenty stocks analysed in the first part of this chapter, and we show in Figure 8.11 the results of the tests

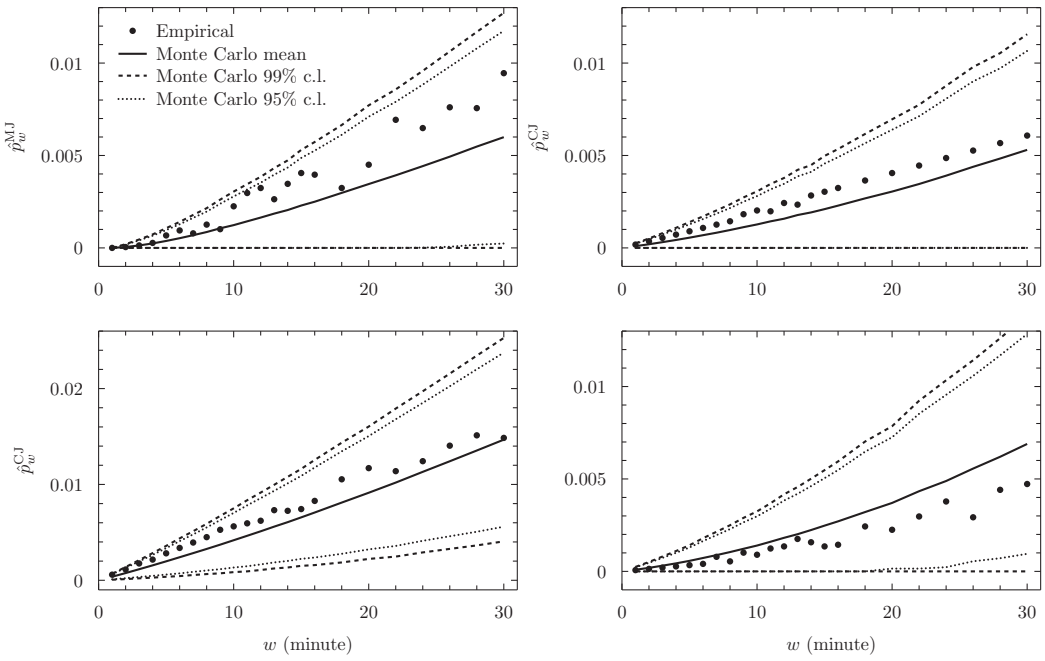

Figure 8.11: From the top left clockwise: MJ probability test under N factor model null for the asset Generali; CJ probability test for the pairs Generali-Mediobanca, Generali-Banca Popolare Milano, and Generali-Intesa Sanpaolo.

performed over the assets Generali, Mediobanca, Banca Popolare di Milano, and Intesa Sanpaolo. The plots clearly show the ability to capture both the self and the cross dependence among jumps, which is a remarkable feature of a factor model easy to implement and calibrate on the data. Moreover, by construction, the model is genuinely scalable, and this fact makes it a viable alternative to more sophisticated but computationally complex models. In Table 8.7 we present the value of the parameters of the common factor, with the associated standard errors and p-values. From the measured confidence for the scale and decay length we can reject the null hypothesis of a Poisson model for the factor. In Table 8.8 we report the remaining parameters for the idiosyncratic components.

It is important to notice that we have performed a Monte Carlo experiment by drawing $N_{MC} = 10^3$ scenarios with the values given in Tables 8.7 and 8.8 and we have then re-estimated the parameters on each copy. This experiment has shown that the entire procedure is quite robust, but the comparison with the process N_t^F, which is known in the artificial Monte Carlo framework, evidences a systematic underestimation of the number of jumps

of the common factor. The misidentification is due to the fact that even when the factor jumps, there is always a small, but finite, probability that only a small number of assets below the detectability threshold jumps too. In this case, our methodology does not detect a systemic event. The misidentification probability depends on the true values of $(p_1, \ldots, p_N)$, and for those fixed in the Monte Carlo experiment, it amounts approximately to 7.5%. Nonetheless, the bias reduces when the number of assets increases, and tends to zero for $N \to \infty$.

	Hawkes factor
λ_F	$(2.0 \pm 0.2) \times 10^{-3}$, p-value: < 0.001
α_F	$(4.9 \pm 1.9) \times 10^{-2}$, p-value: 0.0105
β_F	$(3.3 \pm 1.1) \times 10^{-1}$, p-value: 0.0021

Table 8.7: List of the parameters of the common factor process.

8.8 Robustness analysis

In this section we provide an extension of the study conducted on the FTSE MIB data, by addressing some questions about the robustness of our Hawkes one factor model with idiosyncratic components and its ability to well describe different market settings. In particular, we aim at answering the following questions. Does the model provide a good description of the jump clustering properties (both in time and cross sectional perspective) also in different markets? Is the model equally effective when a larger number of stocks is investigated? How does the model behave when wider ranges of time are considered?

As performance measures to assess the ability of our model to reproduce the observed MJ and CJ features, we use two kinds of statistical tests: simple ones, where for each window length w (the range of w is 1–30 minutes) we perform a hypothesis test separately and independently from other window lengths, and multiple ones, where we require that for every window length the observed quantity lies within the confidence bands. Thus, the null hypothesis underlying the multiple test is not rejected if *all* the points representing the observed quantities (please refer to Figure 8.13) lie inside the confidence bands of the simulated model. In the multiple tests we calculate the rejection regions for the various window lengths taking into account the Bonferroni

ISIN	p_s	λ_s ($\times 10^{-3}$), p-value	α_s ($\times 10^{-2}$), p-value	β_s ($\times 10^{-1}$), p-value
IT0000062072	0.31	1.4 ± 0.2, $^{(*)}$	2.6 ± 1.4, 0.067	2.2 ± 1.0, 0.030
IT0000062957	0.12	1.1 ± 0.2, $^{(*)}$	1.1 ± 1.3, 0.401	2.3 ± 2.6, 0.386
IT0000064482	0.10	2.2 ± 0.2, $^{(*)}$	3.7 ± 1.6, 0.020	2.6 ± 1.1, 0.014
IT0000068525	0.24	1.2 ± 0.2, $^{(*)}$	0.9 ± 0.4, 0.020	0.3 ± 0.1, 0.021
IT0000072618	0.48	1.6 ± 0.2, $^{(*)}$	0.6 ± 0.6, 0.277	1.0 ± 1.1, 0.365
IT0001063210	0.08	1.0 ± 0.2, $^{(*)}$	0.6 ± 0.5, 0.174	0.6 ± 0.4, 0.125
IT0001334587	0.17	2.9 ± 0.3, $^{(*)}$	6.2 ± 1.8, 0.001	3.0 ± 0.8, $^{(*)}$
IT0001976403	0.31	1.7 ± 0.2, $^{(*)}$	2.8 ± 1.2, 0.015	1.7 ± 0.6, 0.006
IT0003128367	0.39	2.6 ± 0.3, $^{(*)}$	2.4 ± 0.8, 0.004	1.2 ± 0.4, 0.002
IT0003132476	0.37	2.2 ± 0.2, $^{(*)}$	1.3 ± 0.5, 0.016	0.9 ± 0.3, 0.012
IT0003487029	0.18	1.0 ± 0.2, $^{(*)}$	5.1 ± 2.8, 0.069	3.9 ± 1.7, 0.019
IT0003497168	0.31	1.8 ± 0.2, $^{(*)}$	0.7 ± 0.4, 0.126	0.4 ± 0.3, 0.232
IT0003856405	0.11	1.6 ± 0.2, $^{(*)}$	0.8 ± 0.4, 0.087	0.7 ± 0.3, 0.058
IT0004176001	0.14	0.9 ± 0.2, $^{(*)}$	0.7 ± 0.3, 0.024	0.2 ± 0.1, 0.018
IT0004231566	0.16	1.7 ± 0.2, $^{(*)}$	1.8 ± 0.9, 0.056	1.5 ± 0.7, 0.031
IT0004623051	0.15	1.8 ± 0.2, $^{(*)}$	2.4 ± 0.9, 0.008	1.0 ± 0.4, 0.006
IT0004644743	0.21	1.4 ± 0.2, $^{(*)}$	3.6 ± 1.5, 0.014	1.9 ± 0.7, 0.003
IT0004781412	0.37	1.4 ± 0.2, $^{(*)}$	1.0 ± 0.4, 0.025	0.5 ± 0.2, 0.007
LU0156801721	0.17	0.8 ± 0.1, $^{(*)}$	1.0 ± 0.7, 0.147	0.7 ± 0.4, 0.092
NL0000226223	0.22	1.0 ± 0.2, $^{(*)}$	3.2 ± 1.6, 0.048	1.2 ± 0.7, 0.073

Table 8.8: List of the parameters of the idiosyncratic components; $^{(*)} < 0.001$.

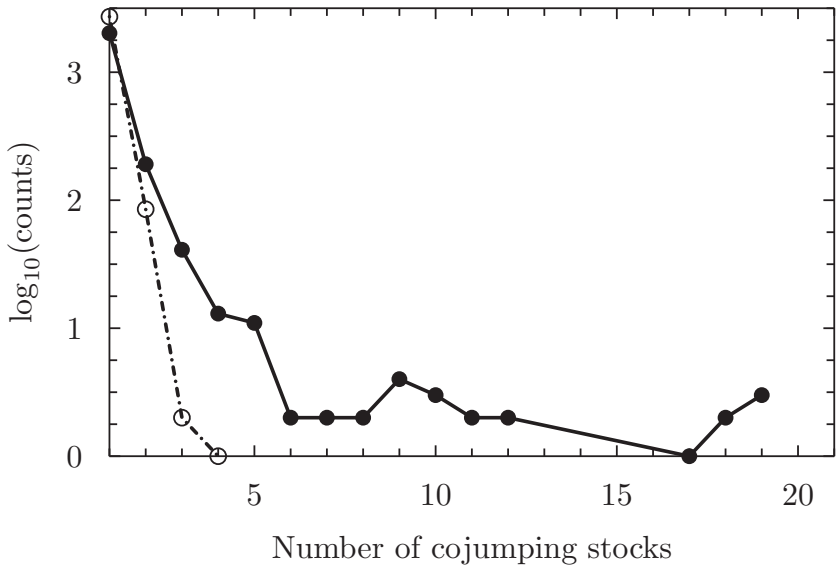

Figure 8.12: Histogram of the number of stocks simultaneously cojumping in a minute (solid line) for the twenty most liquid assets of the Russell 3000 index during year 2013. For comparison the dashed line shows the 0.01% confidence interval for the counts under the null hypothesis of independent but not identical Poisson processes.

correction, that is, by dividing the preset significance level by the number of the tested window lengths.

In the first row of Table 8.9, we present the results of the above mentioned statistical tests for the FTSE data. We see that 19 out of the 20 stocks and 172 out of the 190 pairs of stocks pass the multiple tests for the MJ and the CJ performed with a 99% confidence level with window lengths equal to 5, 10, 15, 20, 25, 30 minutes. The simple tests performed at different window lengths w show generally good results, with the best ones at higher values of w and the greatest number of exceptions occurring at small values. More precisely, the largest fraction of rejections occurs for window lengths shorter than 10 minutes, where the observed number of MJ and CJ is higher than that predicted under the null hypothesis. This indicates that, although the clustering of jumps is generally well captured, at very small time scales the model underestimates the real MJ and CJ frequency.

In order to test our model on a different market, we carry out the same analysis on a dataset formed by the most liquid stocks of the Russell 3000 index during the year 2013. The first analysis is done keeping the number of assets equal to 20 and the length of the time series approximately equal

to that of the FTSE MIB data (we take the first 114 days of the year).[6] The goal is to see if a different market exhibits analogous properties to those observed in the Italian market. As far as the jump distributional properties are concerned, we report in Figure 8.12 the same analysis presented in the right panel of Figure 8.4. We indeed get comparable results and the observed number of stocks jumping in the same minute strongly rejects the null model of twenty independent Poisson processes. We also confirm the tendency of the jumps to cluster in time rejecting the null of a Poisson process in favour of a self-exciting process for all the stocks in our sample.

Then, we test the sensitivity of our model to the jump threshold θ looking at the parameter estimates of the factor model for values of θ larger than 4. As θ increases the number of detected events rapidly decreases and thus the self-exciting features of the jump processes diminish. For $\theta = 5$ the null of a Poisson process is not rejected for both the factor and 18 out of the 20 idiosyncratic components. Moving to $\theta = 6$ we cannot reject the Poisson null for any of the idiosyncratic jump point processes. However, the cross-exciting features still persist, strongly rejecting a model of independent Poisson processes, and we conclude that the most adequate description of our datasets corresponds to a factor model where both the factor and the idiosyncratic components are Poisson processes.

The top left and top right panels in Figure 8.13, analogous to Figure 8.6 and Figure 8.7, show that MJ for Citigroup is well explained by a Hawkes process, while a Poisson process clearly fails. The bottom left and bottom right panels refer instead to our Hawkes one factor model and show that it is able to capture the MJ and CJ properties for Citigroup and the pair Citigroup-Pfizer. We report a detailed summary for the entire data set of twenty Russell 3000 assets in the second row of Table 8.9. The numerical results confirm that our factor model is adequate to describe the jump behaviour not only for Italian assets but also for the US market.

In order to investigate the sensitivity of the Hawkes factor model when the number of assets grows we estimate it on samples of increasing size. As more and more stocks are included in the analysis, we expect the description of the system based on a single factor driving the whole market to become less and less adequate. In the second column of Table 8.10 we report the fraction of total variance explained by the first eigenvalue of the jump covariance matrix and we indeed observe that it decreases from 12.53% for $N = 20$ to 7.72% for $N = 60$. Consistently with this, we see that the performance of the

[6]The list of the twenty company tickers sorted in decreasing order of liquidity is: C, PFE, JPM, WFC, T, KO, MRK, FCX, CVX, F, AIG, MO, EMC, HPQ, CAT, HD, WMT, DIS, SLB, PG.

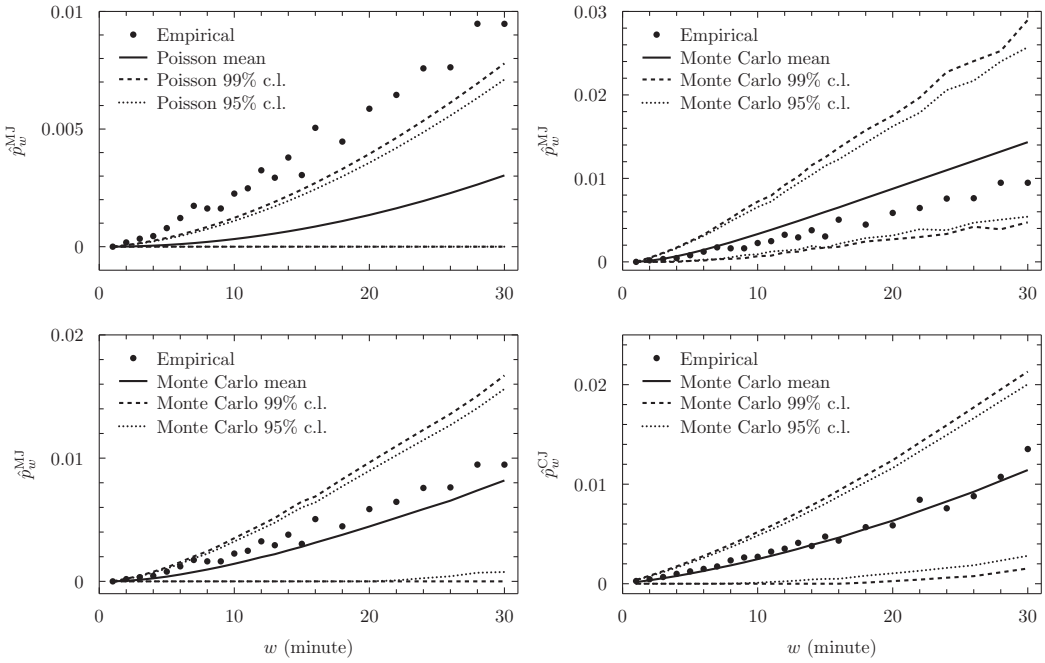

Figure 8.13: (Top left) Observed MJ values in relation to Poisson confidence bands. (Top right) Observed MJ values in relation to the one-dimensional Hawkes confidence bands. (Bottom left and bottom right) Observed MJ and CJ frequencies in relation to the Hawkes one factor model with idiosyncratic components. We perform MJ test for the asset Citigroup, while the CJ test refers to the pair Citigroup-Pfizer.

model in reproducing the MJ and CJ features tends to worsen as N increases, although only slightly so that reasonably good results are maintained for N up to 60.

Concerning the sensitivity to the length of the time series, Table 8.9 (rows 2–7) details the results of the MJ and CJ tests for the Russell 3000 dataset when T increases from 44346 minutes (nearly 5 months) up to one trading year. The performance of the CJ tests are overall quite good, slightly worsening for the shortest time windows. The behaviour of the MJ tests is more puzzling, since the test fails more frequently in particular for window lengths smaller than or equal to five minutes. As a first possible explanation we investigated the impact of non-stationarity of the model parameters on the statistical tests. Generating three four-month periods with different parameters,[7] gluing them together, and finally performing a single estimate on the

[7]The parameters are obtained by estimating the model on the three periods separately.

whole period, we nevertheless found that the MJ and CJ tests had an almost 100% positive outcome. Thus, we tend to exclude the non-stationarity as the main responsible for the worsening of our model performances. Since the one factor model fails more frequently for very short time windows while for windows larger than ten minutes it is always successful, we hypothesize that the single exponential specification of the process kernel may be inadequate at the shortest time scales. When T increases the statistical uncertainty associated with the model parameters decreases and we are able to better discriminate the model adequacy. As a future development we plan to extend our analysis to more complex kernel specifications.

In summary, the robustness analysis shows that the proposed model is robust to different datasets, to changes in the jump threshold parameter θ and to the increase of the number of assets, while some mild misspecifications arise when widening the time range of the analysed time series. Moreover, this robustness analysis confirms that the proposed Hawkes factor model outperforms competitor Poisson models and provides a much more realistic description of the collective jump dynamics.

8.9 Conclusions and perspectives

The detection techniques that we develop in Sections 8.2 and 8.3 show that a large number of jumps is present in financial time series. Even though the identification process of the extreme events suffers from some dependence on the details of the detection method, we believe that the idea of intersecting the different methodologies partially amends this drawback. Relying on the correct identification of jumps, we find that, as far as individual stocks are concerned, jumps are clearly not described by a Poisson process. The evidence of time clustering can be accounted for and modelled by means of linear self-exciting Hawkes processes. Moving to a cross-sectional perspective, we identify a significant number of systemic events, especially simultaneous cross jumps, that can not be reduced to a purely random effect. We have provided quantitative arguments against the idea of modelling this effect in terms of multidimensional Hawkes processes. The simultaneity of events is not captured by this class of point processes, and the increase of dimensionality of the parameters space associated with the multivariate model is discouraging. In Section 8.7.2 we propose a one factor model which is able to describe the main features that characterise the departure from a random behaviour of jumps, namely, the time clustering of jumps on individual stocks, the large number of simultaneous systemic jumps, and the time lagged cross excitation between different stocks.

T	MJ (multiple)	MJ (simple)							CJ (multiple)	CJ (simple)							
		$w = 2$	3	4	5	10	15	20		$w = 1$	2	3	4	5	10	15	20
44440	19	10	17	16	16	18	19	20	172	155	152	156	155	160	163	175	177
44346	19	11	18	17	18	20	20	20	183	164	176	176	180	178	182	186	186
54849	17	7	9	10	10	20	20	20	186	138	165	168	181	181	184	187	186
65352	18	7	6	11	9	19	20	20	186	158	174	173	179	182	183	187	186
75855	14	4	3	6	4	20	20	20	186	151	168	171	179	181	185	187	187
86358	13	3	2	7	3	20	19	20	185	149	164	170	176	178	185	188	186
96861	7	3	2	4	2	19	20	20	188	154	170	177	181	185	188	188	188

Table 8.9: (First row) Number of successful tests for the FTSE MIB data set. (Rows 2–7) Test results for the 20 most liquid stocks of the Russell 3000 Index data set, year 2013, for T regularly increasing from 44346 minutes (114 trading days) to 96861 (one trading year). We perform the MJ multiple tests on time windows of length equal to 5, 10, 15, 20, 25, and 30 minutes.

N	1st PC	MJ (multiple)(%)	MJ (simple)(%)							CJ (multiple)(%)	CJ (simple)(%)							
			$w = 2$	3	4	5	10	15	20		$w = 1$	2	3	4	5	10	15	20
20	12.53 %	95	55	90	85	90	100	100	100	96	86	92	93	95	94	96	98	98
30	10.16 %	90	43	67	70	90	100	100	100	93	74	85	87	86	87	94	96	95
40	9.01 %	90	45	65	73	83	97	95	100	91	61	80	81	81	84	91	95	95
50	8.33 %	86	44	56	72	78	96	94	100	89	52	71	76	77	80	91	95	94
60	7.72 %	87	45	57	67	72	95	95	100	89	51	69	74	75	79	91	94	94

Table 8.10: Test results for the most liquid stocks of the Russell 3000 Index data set, year 2013, for $T = 44346$ and N equal to 20, 30, 40, 50, and 60. We perform the MJ multiple tests on time windows of length equal to 5, 10, 15, 20, 25, and 30 minutes. The second column refers to the fraction of the total variance explained by the first principal component (PC) of the jump covariance matrix.

The robustness analysis detailed in Section 8.8 investigates the dependence of the results on the temporal and cross-sectional dimensions of the dataset, and on the specific market. We conclude that our model applies equally effectively to both the Italian and US markets and it is thus robust to different datasets. It performs well when the number of assets increases (up to 60 investigated assets), while some mild misspecifications arise widening the time range of the analysed time series. We hypothesize that a refined specification of the process kernel may further improve the performances of our factor model at the shortest time scales. We leave the answer to this interesting research question to future developments.

In order to capture possible contagion effects among distinct asset classes, an interesting development of the present study is the extension to a mixed portfolio with different but related securities, such as equities, futures, options and other derivative products. A second extension of this work is the study of the properties of cross jumps for a given group of securities in different periods, to assess for example whether changes in the regulations of a market have an impact on the frequency of systemic events.

Moving from a descriptive point of view to one that investigates the origin of the behaviour of jumps and cojumps, the research direction we consider more promising is the study of the order book in proximity of extreme systemic events, that is, cojumps involving a large number of assets. The perspective to take should be to explore the cross direction at a fixed time, more than the time direction for a single asset. The bid-ask spread, the depth of the book, the asymmetry of buy and sell volumes are all quantities whose dynamics may reveal interesting features related to systemic events.

The high level of synchronisation between the jumps of different assets that we empirically observe in our data calls for possible explanations. Even if our work is mainly methodological, we believe that some comments are needed in order to explain this fact observed in (modern) financial markets. Financial markets are becoming increasingly interconnected at a high speed due to several reasons, first of all the increased automation of the trading process and of the information processing. High frequency trading strategies, statistical arbitrageurs, hedging strategies could be partly responsible of the large number of cojumps we observe in our sample. Certainly, we believe, a proper modelling of the jump process in a systemic context is important for regulators and for investors in order to assess in a reliable way the level of risk of a market or of a large portfolio of assets.

8.10 Appendix: Volatility auctions

According to the rules stated in *Rules of the markets organised and managed by Borsa Italiana S.p.A.* and *Instructions accompanying the Rules of the markets organised and managed by Borsa Italiana S.p.A.* ([125, 126]), whenever a stock's price gets too far from a reference value, Borsa Italiana is obliged to start a volatility auction phase. During this phase, which has a duration between 10 and 11 minutes, trades are suspended and a new reference value for the price is sought. Possibly, if no valid price is reached during the volatility auction phase, another such phase starts immediately after the first one and so on. We treat all intertrade times of at least ten minutes as volatility auction periods (the distributions of intertrade times shown in Figure 8.14 suggest this is correct) and consider as *not available* the returns at the sampling times falling in these periods. In Figure 8.14 we report the empirical distributions of the intertrade times for the assets Fiat and Telecom Italia. In both histograms we can easily recognize the presence of volatility auctions, which manifest in terms of the peaks appearing on the far right tail in correspondence of intertrade times close to the multiples of eleven minutes.

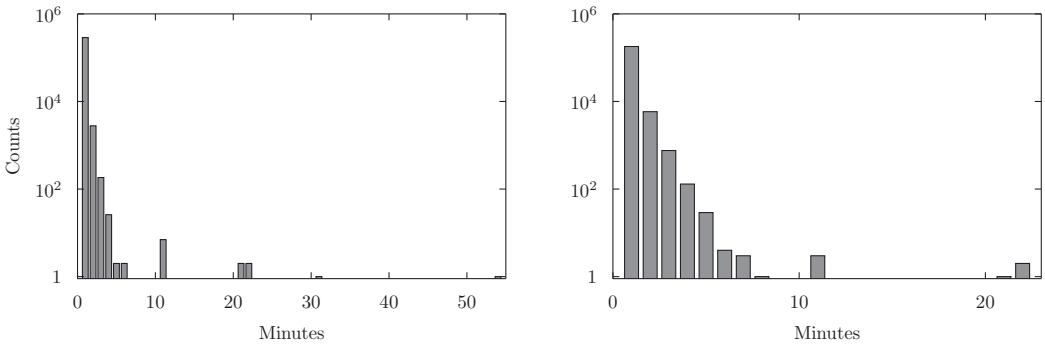

Figure 8.14: The distributions of intertrade times for stocks Fiat (left panel) and Telecom Italia (right panel).

Chapter 9

Collective synchronization and high frequency systemic instabilities in financial markets

9.1 Introduction

Quoting from Michael Lewis' Flash Boys "The world clings to its old mental picture of the stock market because it's comforting" [133]. But trading activity has profoundly changed from the old phone conversation or click and trade on a screen to software programming. Market statistics confirm that automated algorithms carry out a significant fraction of the trading activity on US and Europe electronic exchanges [134, 32]. As algos feed on financial and news data, the speed of information processing has dramatically increased and potentially allows large price movements to propagate very rapidly through different assets and exchanges [34].

The synchronization effect had its most spectacular appearance during the May 6th, 2010 Flash Crash. The crash started from a rapid price decline in the E-Mini S&P 500 market and in a very short time the anomaly became systemic and the shock propagated towards ETFs, stock indices and their components, and derivatives [35, 36]. The price of the Dow Jones Industrial Average plunged by 9% in less than 5 minutes but recovered the pre-shock level in the next 15 minutes of trading. The SEC reported that such a swing was sparked by an algorithm executing a sell order placed by a large mutual fund. Then high frequency traders, even though did not ignited the event, caused a "hot potato" effect amplifying the crash. In the aftermath of the crash, several studies have focused on events, evocatively named *Mini Flash Crashes*, concerned with the emergence of large price movements of an

asset in a very limited fraction of time and attributing their origin to the interaction between several automatic algorithms [37] or to the unexpected product of regulation framework and market fragmentation [38].

The Flash Crash, however, has also dramatically shown how strongly interconnected different markets and asset classes can become, especially during extreme events. In this chapter, by taking a different, yet complementary approach to the above literature, (i) we identify nonparametrically one-minute extreme events as over-threshold movements, studying how the frequency of collective instabilities at high frequency has changed in the last years, and (ii) we use Hawkes processes [40] to model parametrically the dynamics of these events and their mutual cross-excitation. Our approach shares some similarities with previous works employing non-parametric tests to identify extreme movements, see [135, 96, 113, 136, 114]. We perform our analysis on a yearly basis from 2001 to 2013 on a data sample of highly liquid US equities and we identify extreme events affecting a sizable fraction of the investigated assets. Remarkably, very little research has been devoted to the investigation of this kind of systemic events. Few noticeable exceptions are [103], who aim at the identification of common large movements between the market portfolio and individual stocks, and [109], who investigate the tendency of large movements to arrive simultaneously. A very recent non-parametric test of the occurrence of simultaneous jumps across multiple assets is discussed in [137]. Our research provides the empirical evidence that, while the total number of extreme movements has decreased along years, the occurrence of systemic events has significantly increased. As a terminology clarification, we prefer the use of the term *systemic* rather than *systematic*, since the latter has been used in the literature (see e.g. [109]) to define events where assets jump together with a market index, while the events we look at not necessarily imply this.

To identify the possible causes of such events we compare their time occurrences with a database of pre-scheduled macroeconomic announcements. Since macroeconomic news can be expected to have a market-level influence, they represent a natural candidate to explain market-wide events. For instance, literature has recognised the peculiar role played by Federal Open Market Committee (FOMC) meetings deciding the interest rate level [138, 139]. However, unexpectedly, only a minor fraction (less than 40%) of events involving a large fraction of assets has been preceded by the release of a macro news. This evidence opens the route to the more intriguing hypothesis that a genuinely endogenous dynamics is taking place. To the best of our knowledge, the association between extreme equity price movements and the news arrival has been previously investigated in [113, 140], finding a positive association, but the results have been challenged in [141]. Table 11 in [109]

suggests the existence of a particularly strong relationship between FOMC announcements and the arrival of a systemic event (defined as an event when the market index jumps). However, none of the previous works performs an analysis of the association between news and extreme movements conditional on the level of systemicity of the event.

Finally, we show that when an event affecting a significant fraction of assets occurs, the probability of a novel extreme event in the subsequent minutes increases. More interestingly, there is a clear evidence that the more systemic the conditioning event is, the larger the expected number of assets swinging synchronously in the immediate future will be. In order to reproduce such empirical evidences, we propose a model within the class of mutually exciting point processes, termed Hawkes processes [40] which in recent years have experienced an increasing popularity in mathematical finance and econometrics [43, 44, 118, 119, 121, 120, 122, 123, 45]. We present a multidimensional, yet parsimonious, Hawkes process capturing with remarkable realism the cross-excitation which affects extreme events identified non-parametrically as over-threshold returns.

9.2 Data

9.2.1 Financial data

We conduct our analysis on price time series of financial stocks belonging to the Russell 3000 Index, traded in the US equity markets (mostly NYSE and NASDAQ). We consider the thirteen years from 2001 to 2013 and for each year we select 140 highly liquid stocks. More explicitly, we take the 140 stocks with the highest percentage of minutes in which at least one trade was made. The 140th selected asset of each year has such a percentage equal to 93% for 2001, 96% for 2002, between 97% and 98.5% for the years 2003–2005, and always greater than 99% for the years from 2006 onwards. We use 1-minute closing price data during the regular US trading session, i.e. from 9:30 a.m. to 4:00 p.m and, as explained in Section 9.7, we remove the intraday pattern of volatility, which is a local measure of the diffusion rate of price.

9.2.2 News data

We use macroeconomic news data provided by Econoday, Inc. `www.econoday.com`. We consider the 42 most important news categories, which are classified into two large groups according to their capacity of influencing the financial markets: the Market Moving Indicator group and the Merit Extra Attention

group. Since we are concerned with matching news with market extreme events, we consider only the 27 categories whose announcement times occur during the trading session. The number of total news announcements ranges from around 150 in the first years to around 260 in the last years, for a total of 2,888 news. See Section 9.7 for more details.

9.3 Methods

9.3.1 Identification of extreme events

In order to detect extreme variations of the stock prices P_t, we compare price returns (defined as $r_t = \ln P_t/P_{t-1}$) with an estimate of the historical spot volatility, which sets the scale of local price fluctuations. Specifically, we calculate a volatility time series σ_t as an exponential-moving-average version of the bipower variation (see [135, 82, 92]) of the return time series and we finally say that an extreme return occurs when

$$\frac{|r_t|}{\sigma_t} > \theta, \tag{9.1}$$

for a certain threshold θ. In our main analyses we take $\theta = 4$, but we also investigate higher values of the threshold, namely $\theta = 6$, 8, 10, in some of our descriptive statistics.

9.4 Results

The main objective of this study is the modeling of the dynamics of synchronous large price variations at high frequency. We say that a stock *jumps* in a given one minute interval if condition of Equation 9.1 is observed for a given θ. Here we are mostly interested in *cojumps*, i.e. the simultaneous (inside the minute) occurrence of jumps for a subset of M stocks. The quantity M is termed the *multiplicity* of the cojump, and it gives a measure of the systemic nature of the event. In the following we consider three questions: (i) how has the high frequency instability changed in the last fifteen years? (ii) what fraction of the systemic instabilities can be attributed to macroeconomic news? (iii) how can we model the short term dynamics of market instabilities?

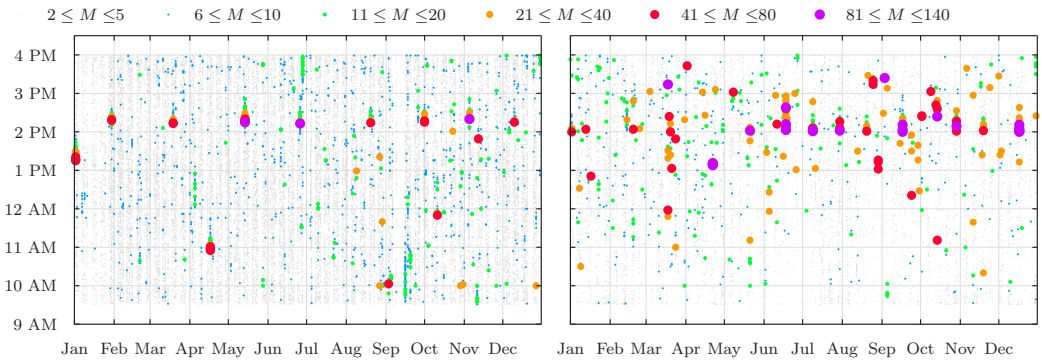

Figure 9.1: Time series of the cojumps detected for the dataset of 140 selected highly liquid stocks of the Russell 3000 Index during year 2001 (left panel) and 2013 (right panel). The size of the circles increases with the multiplicity of the cojump event.

9.4.1 Historical dynamics of jumps and cojumps

A visual representation of how instability of financial markets has changed in the last years is shown in Figure 9.1, which compares the dynamics of $\theta = 4$ cojumps in 2001 (left panel) and 2013 (right panel). The horizontal axis represents the trading day and the vertical axis indicates the hour of the day. The presence of a circle indicates the occurrence of a cojump and the color codifies the number of stocks simultaneously cojumping (i.e. the multiplicity). In 2001 there were many cojumps with low multiplicity and the high multiplicity cojumps are concentrated mostly at specific hours of the day (10 a.m. and 2:15 p.m.) corresponding to the release of important macro announcements, such as, for example, the FOMC announcements. On the contrary, in 2013 we observe less low multiplicity cojumps and many more high multiplicity cojumps, which are quite scattered during the day. This is an indication that modern financial markets have become more systemically unstable and that these instabilities are less related to macro news. In the following we show that this is the case with more quantitative analyses.

First, in the top left panel of Figure 9.2 we show the frequency of jumps per minute in each year, considering different values of θ. We observe that for all θs the number of jumps has actually *decreased* over time. The different lines are quite parallel one to each other (especially for $\theta \geq 6$) indicating that the tails of the one minute return distribution remained quite stable. A completely different pattern emerges when we consider the dynamics of cojumps. The top right panel of Figure 9.2 shows the frequency of cojumps of different multiplicity (normalised to its value in 2001). While the frequency of cojumps with any multiplicity ($M \geq 2$) has slightly declined, the frequency

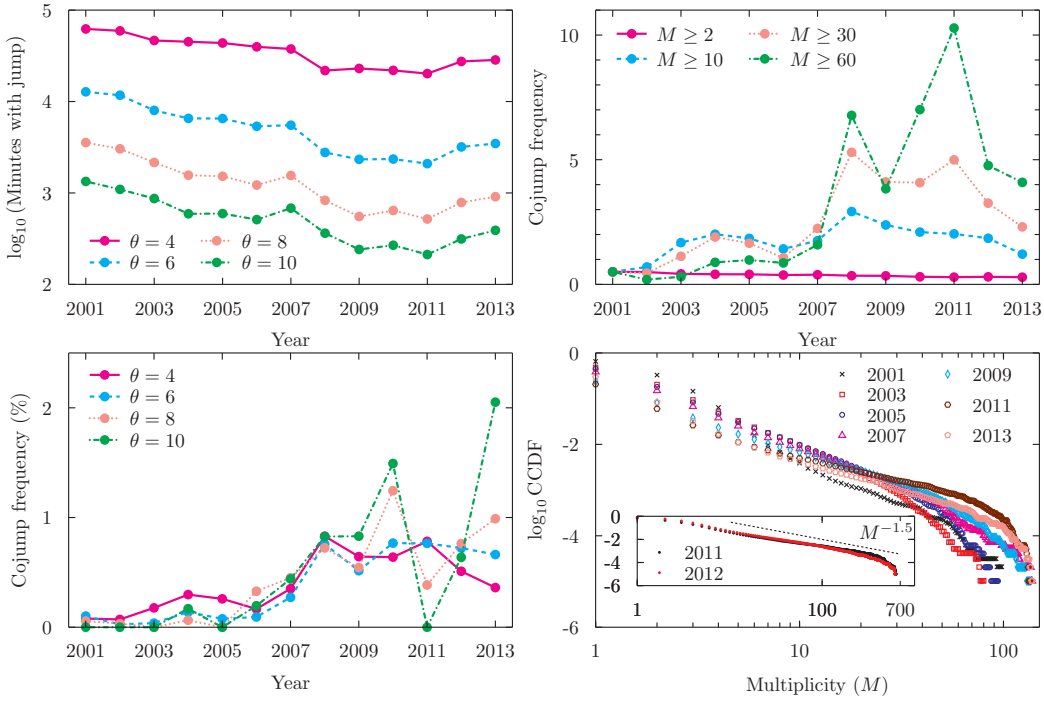

Figure 9.2: Top left panel: Semi-log plots of the total number of minutes where we detect at least one jump among the 140 selected assets of the Russell 3000 Index. Curves correspond to four different levels of the threshold parameter θ. Top right panel: For $\theta = 4$, yearly time evolution of the fraction of minutes with at least one event of multiplicity larger than or equal to 2, 10, 30, 60. All values are normalised by the corresponding 2001 values. Bottom left panel: Yearly evolution of the percentage fraction of cojumps with multiplicity at least equal to 30 for four different values of θ. Bottom right panel: Log-log plots of the Complementary of the Cumulative Distribution Function of the cojump multiplicity for seven different years. The panel reports the empirical evidence for a portfolio of 140 stocks, while the inset details results of the same analysis conducted with 700 liquid assets from Russell 3000 during years 2011 and 2012.

of high multiplicity cojumps has become in recent years up to 10 times more frequent than its value in 2001. The result is essentially unchanged when fixing the minimal multiplicity (e.g. $M \geq 30$) and computing the number of cojumps for different values of θ (bottom left panel of Figure 9.2). Clearly larger fluctuations are observed for larger values of M. The increase of frequency of high multiplicity events is not due to the fact that markets have become faster. In Section 9.8 we show the fraction of cojumps with $M \geq 30$ and $M \geq 60$ at $1, \ldots, 5$ minutes. It is clear that the variability with the time window defining the event is much smaller than the secular variability of the events. In fact the fraction of 1-min cojumps with $M \geq 30$ in 2013 is significantly larger than the fraction of 5-min cojumps with $M \geq 30$ in 2001. The same is true for cojumps with $M \geq 60$. Therefore, the increase in synchronization is a genuine phenomenon, not explained by the increase in market speed.

Finally, the bottom right panel of Figure 9.2 shows the distribution function of the cojump multiplicity for different years. Despite some variation is observed across the years, a clear power law tail behavior is evident. This means that the probability of systemic cojumps is quite large. Consistently with the observations above, the tail is thicker in recent years (even if in 2013 we observe a slightly thinner tail). It is important to notice that the bending of the distributions for large multiplicity is very likely due to the finite support of the distribution. Clearly for a set of N stocks the multiplicity cannot be larger than N, thus the distribution function is zero at $M = N$. To show the role of the finite support, in the inset we show the multiplicity distribution function for a larger set of 700 highly liquid assets. In this case the power law region extends for a wider range and close to $M = 700$ we observe the expected bending of the function. The tail exponent of these distributions is close to 1.5 (similarly to what observed in [127]).

In conclusion, at the beginning of 2000's individual jumps were more frequent and high frequency systemic instabilities, i.e. high multiplicity jumps, were rare and mostly concentrated on macro-news announcements. In recent years, on the contrary, markets display often systemic cojumps and these are scattered across the trading day.

9.4.2 Systemic cojumps and macroeconomic news

The second question is which fraction of these systemic cojumps has an exogenous or an endogenous origin. To answer this question we study how frequently a systemic cojump is preceded by a scheduled macroeconomic news. It is in fact unlikely that stock idiosyncratic news affect the whole market. We measure how frequently a systemic cojump with multiplicity

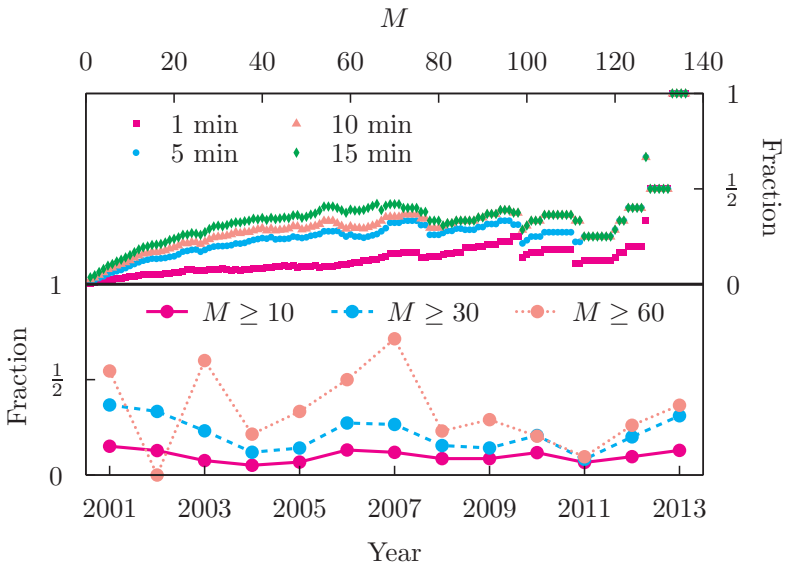

Figure 9.3: Top panel: Fraction of cojumps in 2012 with multiplicity larger than or equal to the value reported on the x axis for which a news occurred in the last 1, 5, 10, and 15 minutes. Bottom panel: Fraction of cojumps for different multiplicities M for which we observe at least one news in a time window of five minutes preceding the jump event.

larger than M is preceded by a macronews in the last $\tau = 1, 5, 10, 15$ minutes. The top graph of Figure 9.3 shows that only 40% of the high multiplicity cojumps are preceded by a macronews in the previous 15 minutes. Notice that the fractions of news-triggered systemic events in the 5, 10, and 15 minutes time windows are very close one to each other, indicating that if a macronews triggers a systemic cojump, this will typically happen within 5 minutes from the news.

For a historical perspective, the bottom graph of Figure 9.3 shows that the fraction of systemic cojumps triggered by macroeconomic news is quite constant across the years and, even for large M, clearly below 50%. Thus our empirical analysis shows that a relevant portion of systemic cojumps is not associated with scheduled macroeconomic announcements. Idiosyncratic company-specific news may play a role, but plausibly only for those events which involve a very limited number of assets. For high multiplicity cojump events, endogenous mechanisms are likely to play a determinant role.

9.5 Model

9.5.1 Hawkes process for multiplicity vector

The empirical evidence of the previous section suggests that a large fraction of the dynamics of the systemic cojumps is unrelated to macro news and is likely endogenously generated. Moreover, as observed for example in the 2010 Flash Crash, market instabilities tend to propagate quickly to other assets, markets, or asset classes. Thus it is important to model the self- and cross-dependence of instabilities, considering both synchronous and lagged dependence, by studying whether and how systemic instabilities trigger other instabilities in the short run.

However the estimation of the interaction among a set of 140 variables is extremely challenging and some sort of filtering is needed. A first step in this direction was taken in [135] where we modeled the multivariate point process describing the jumps with a Hawkes factor model. Each stock is represented by a point process, each count being a jump. The coupling between the stocks is given by a one factor model structure, i.e. the intensity is the sum of the intensity of a factor and the intensity of an idiosyncratic term. Finally in order to capture the temporal clustering of events we assumed that both the factor and the idiosyncratic term follow a Hawkes process.

As shown in [135] this type of modeling is very effective (and parsimonious) in describing the pairwise properties of cojumps, i.e. the probability that two stocks jump in the same time interval. However when considering cojumps of $M > 2$ stocks, the model shows its weakness. An important indication is given by the distribution of multiplicities. It is possible to show that in the large N limit, the factor model of [135] predicts a multiplicity distribution with Gaussian tails, at odds with the power law behavior observed empirically in the bottom right panel of Figure 9.2. Moreover the multiplicity of a systemic cojump is independent from the multiplicity of previous systemic cojumps, while the right panel of Figure 9.1 shows clear temporal clusters of high multiplicity cojumps.

For these reasons, in this chapter we propose a new modeling approach which preserves the parsimony and is able to overcome the problems of the model of [135]. The idea is to model directly the vector of multiplicities, losing information on the identity of the cojumping stocks.

Specifically, we consider an N-dimensional point process characterised by the vector of intensities $\boldsymbol{\lambda}_t$. An event in the i-component at time t means that at this time a systemic cojump of multiplicity i has occurred. Under this modeling assumption we know the total number of assets which have jumped, but we can no longer identify which companies among the N possible ones

have moved. To model the self- and cross-excitation of cojumps we use an N-dimensional Hawkes process with exponential kernels (see Section 9.9 for the definition and the most relevant features). In general, the model depends parametrically on the baseline intensity vector $\boldsymbol{\mu}$, and on the $N \times N$ matrices α_{ij} and β_{ij} of parameters characterizing the kernels. In order to reduce the dimension of the estimation problem from $N + 2N^2$ to a more manageable number of unknowns, we proceed as follows. Since an important goal of our model is the ability to reproduce the empirical stationary distribution of the multiplicity vector, we assume $\boldsymbol{\mu} = \eta \mathbb{E}[\boldsymbol{\lambda}_t]$, where $0 < \eta < 1$, and $\mathbb{E}[\boldsymbol{\lambda}_t]$ proportional to the observed multiplicity frequencies. Interestingly, it is possible to show that $1 - \eta$ is the spectral radius of the kernel matrix and therefore it measures the fraction of intensity explained by the self- and cross-excitation, while η is the fraction explained by the baseline intensity. We assume that all the parameters β_{ij} which characterize the decay time of the self- and cross-excitations are equal to a constant value β. Finally, we hypothesize that, for fixed $i = 1, \ldots, N$, the largest intensity shock is ascribable to the self-exciting term α_{ii}, while the cross-exciting effects as a function of the distance $|i - j|$ between multiplicities decrease hyperbolically with a tail exponent γ. This means that cojumps of a given multiplicity excite with higher probability cojumps with similar multiplicity. To sum up, the model is completely specified in terms of three parameters, η, β, and γ, and the empirical expected number of events with fixed multiplicity.

9.5.2 Model results

We apply the model to the dataset of 140 stocks in 2013. For the complete analysis on all years from 2001 to 2012 see Section 9.9. In order to calibrate and test the model we make use of two quantities, $f_\tau^{(1)}(M; J)$ and $f_\tau^{(2)}(M)$, defined in Section 9.9. The first one is the probability, conditional on the realization at time t of an event with multiplicity at least M, of a cojump with multiplicity at least J in the interval $(t, t + \tau]$. It measures how frequently a systemic cojump triggers other systemic cojumps in the short run. The second quantity is the average multiplicity of the cojumps inside a time interval of length τ after a cojump of multiplicity larger than or equal to M. It therefore measures the typical cojump multiplicity triggered by a cojump of multiplicity at least M. We consider here the case $\tau = 5$ minutes.

We use the $f_\tau^{(1)}(M; J)$ with $J = 10$ and $f_\tau^{(2)}(M)$ to calibrate the model (see Section 9.9 for details) and we test it on $f_\tau^{(1)}(M; J)$ with $J = 30$ and $J = 60$. The estimated parameters are $\eta = 0.15$, $\beta = 0.6$, $\gamma = 2.65$. Thus 85% of the cojump activity is explained by the excitation mechanism and only 15% can be attributed to the baseline intensity. The typical timescale of the

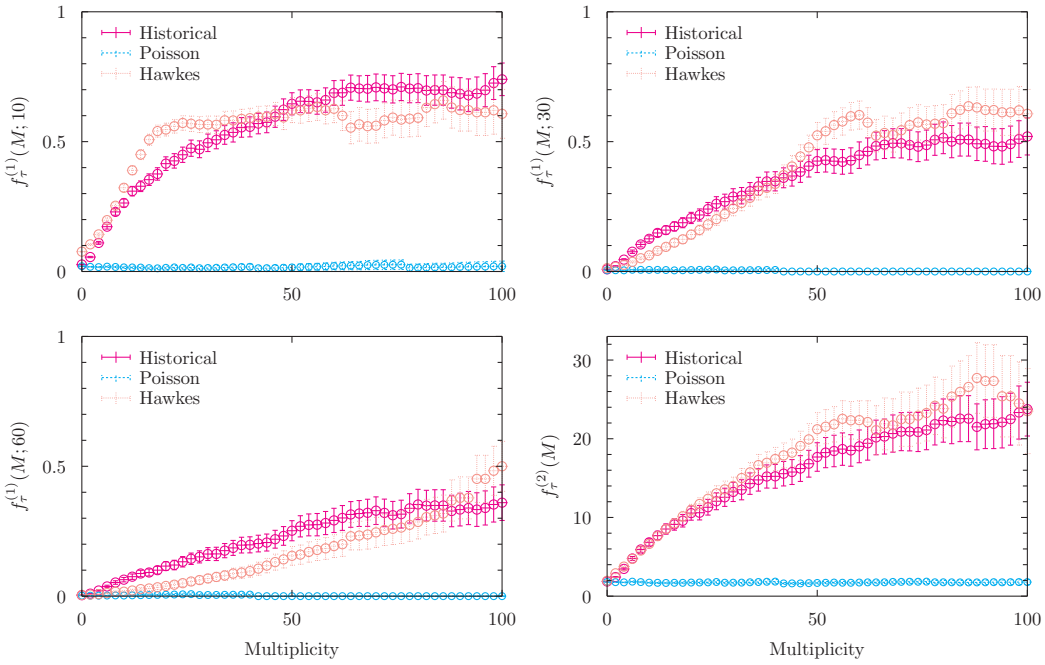

Figure 9.4: Top left panel: Probability that a cojump with multiplicity larger than or equal to 10 occurs in a $\tau = 5$ minute interval following a cojump at time t with multiplicity $M_t \geq M$. Plots are obtained from historical and simulated data. The error bars represent standard errors. Top right and bottom left panels: Threshold 10 replaced by 30 and 60, respectively. Bottom right panel: Expected amplitude of the cojumps in a $\tau = 5$ minute interval following a cojump with multiplicity $M_t \geq M$.

memory is $1/\beta \simeq 1.67$ minutes and the relatively low value of γ indicates a strong cross-excitation between different multiplicities. As expected, the model effectively reproduces the stationary distribution of the multiplicities observed in empirical data (see Figure 9.8 in Section 9.9). Figure 9.4 reports the quantities $f_\tau^{(1)}(M;J)$ and $f_\tau^{(2)}(M)$ in real and simulated data. The solid line corresponds to the empirical probabilities, the dotted line to the results from the Hawkes model, and as a benchmark case we also show the result of a shuffling experiment on the multiplicity time series (dashed line). It is evident that dropping the lagged correlations we obtain an unrealistic description of the multiplicity process. The Hawkes model, on the contrary, fits well the empirical data and therefore adequately describes the cross-excitation mechanism between systemic cojumps. Some discrepancies are observable for $J = 60$, but the general shape of the curve and its level are well reproduced and the Hawkes model is a huge improvement with respect to the benchmark case. This evidence confirms that the larger is the value of the conditioning multiplicity the greater is the probability that in the subsequent minutes an event with large multiplicity happens.

9.6 Discussion

By investigating a portfolio of highly liquid stocks, our research enlightens a remarkable evidence: Since 2001 the total number of extreme events has remarkably diminished, but the number of occurrences where a sizable fraction of assets jump together has increased. This trend is more and more pronounced as we consider events of higher and higher multiplicity. This evidence is a clear mark that markets are nowadays more and more interconnected and a strong synchronization between jumps of different assets is present.

What are the factors responsible for the appearance of extreme movements? The cause can be either exogenous or endogenous. The former case is linked to the release of macro-economic news impacting the price dynamics, while the latter may result from unstable market conditions, such as a temporary lack of liquidity. Quite unexpectedly, only a minor fraction (up to 40%) of the cojumps involving a large number of assets can be attributed to exogenous news. The remaining 60% suggests that a more intriguing endogenous mechanism is taking place. Why has the synchronization among different assets increased through the recent years? We hypothesize that a major role is played by the dramatic increase of algorithmic trading. Thanks to the technological innovation, faster information processing is responsible for the more rapid propagation of large price movements through different

assets. We also provide the evidence that highly systemic instabilities have the double effect of (i) increasing the probability that another systemic event takes place in the near future and (ii) increasing the degree of systemicity of short-term instabilities.

The low timescale of the memory of the exciting effects and the strong persistence of the cross-excitation among different multiplicities support the idea that, to achieve an accurate description of high frequency price dynamics, we should abandon conventional modeling assumptions. Coherently, we propose an innovative approach to the collective behavior of assets' prices based on the Hawkes description of the multiplicity process. Our model well describes the short term dynamics of systemic instabilities while preserving a remarkable parsimony in the number of parameters. Thus, it provides a realistic description of the market behavior which is of prime importance from several perspectives, from trading to risk control, and market designing.

9.7 Appendix A: Data

9.7.1 Market data

Data are provided by Kibot, `www.kibot.com`. We consider the thirteen years from 2001 to 2013 and for each year we select 140 highly liquid stocks in the Russell 3000 index. We exclude American Depositary Receipts, which are negotiable instruments representing ownership in non-US companies, since their dynamics is heavily influenced by their primary market and thus shows a peculiar intraday pattern. We use 1-minute closing price data during the regular US trading session, i.e. from 9:30 a.m. to 4:00 p.m. We discard early-closing days (typically, the eves of Independence Day, Thanksgiving and Christmas). Data are adjusted for splits and dividends.

Intraday returns are first filtered for the average intraday pattern, since price fluctuations are known to exhibit significant differences in absolute size depending on the time of the day, showing a typical U shape with larger movements at the beginning and at the end of the trading day. Returns at intraday time t are rescaled by a factor ζ_t, which is calculated as the average, over all days, of adjusted absolute returns at time t. More precisely, if $\tilde{r}_{d,t}$ is the raw return of day d and intraday time t, we define the rescaled return

$$r_{d,t} = \frac{\tilde{r}_{d,t}}{\zeta_t},$$

where

$$\zeta_t = \frac{1}{N_{\text{days}}} \sum_{d'} \frac{|r_{d',t}|}{s_{d'}},$$

with N_{days} indicating the number of days in the sample and $s_{d'}$ the standard deviation of absolute intraday returns of day d'. Scaled returns no longer possess any daily regularities and can thus be considered a unique time series with no periodic structure. For more details please refer to [135].

9.7.2 News data

The macronews dataset is provided by Econoday, Inc., `www.econoday.com`. Table 9.1 shows the number of news announcements, organised by year and news category.

9.8 Appendix B: Dependence of systemic co-jumps on time scale detection

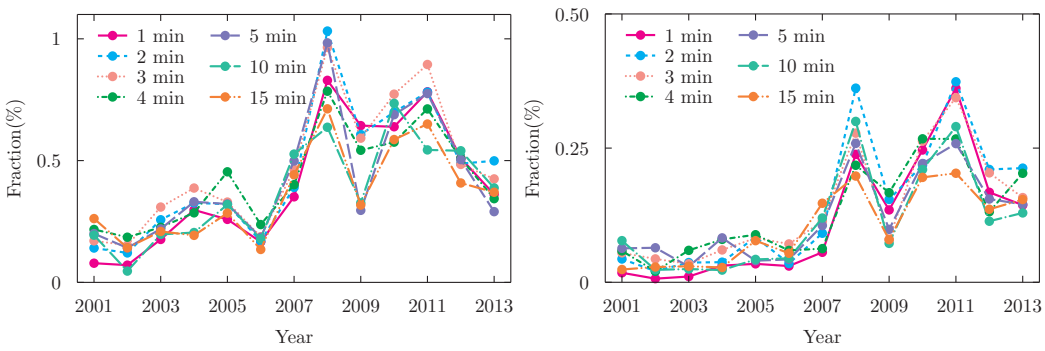

Figure 9.5: Yearly time evolution of the fraction of cojumps with multiplicity $M \geq 30$ (left) or $M \geq 60$ (right) over the total number of cojumps ($M \geq 1$) for $\theta = 4$ and different time horizons, namely $1, 2, 3, 4, 5, 10, 15$ minutes.

This study mostly considers one minute (co)jumps. However one minute in 2013 is not equivalent to one minute in 2001 in terms of market activity. Hence it is important to test whether the increase in number of high multiplicity cojumps is due to the fact that in older years synchronization occurred on a time scale longer than one minute. To test this possibility we have repeated the analysis varying the time scale for jump detection from one to five minutes. Analyses on the dynamics of cross-correlation between stocks data suggested us that the time scale over which stocks become correlated has decreased by a factor approximately equal to five from 2001 to 2013. Moreover, since at high sampling frequencies microstructure noise may be a source of bias and dominate the results, we take into account the findings of

[142] and, similarly to [143], we consider also sampling frequencies up to 15 minutes.

Figure 9.5 shows the yearly time evolution of the fraction of cojumps with multiplicity $M \geq 30$ (left) or $M \geq 60$ (right) over the total number of cojumps ($M \geq 1$) for $\theta = 4$ and different time scales, namely $1, 2, 3, 4, 5, 10, 15$ minutes. Except for the first two years, no clear sorting of this fraction with the time scale is detectable, while the global secular trend has a much larger variability. This is particularly evident for the $M \geq 60$ case. Hence the number of high multiplicity one minute cojumps in 2013 is much higher than the number of high multiplicity five minute cojumps in 2001, indicating that the increased speed of market activity is a minor cause of the increase of high multiplicity systemic cojumps in recent years. We also note that this increase is also evident at sampling frequencies of 10 and 15 minutes, which gives robustness to the results by providing the evidence that the increase of systemic cojumps is not due to microstructure noise. Finally, it is worth to point out that the fraction of cojumps with multiplicity equal to or larger than 30 reaches its maximum in 2008. When conditioning to events with even more extreme cojump events – with multiplicity equal to or larger than 60 – a second peak appears in 2011. Then, these results seem to suggest – maybe quite expectedly – the emergence of a tighter collective market dynamics during the sub-prime mortgage crisis and the collapse of Lehman Brothers in 2008, and the rise of the sovereign debt crisis in 2011.

9.9 Appendix C: Model

In this study we model the point process describing the cojumps of k stocks (independently from their identity) as the k-th component of a multivariate Hawkes process. These processes were introduced in the early Seventies [40], and have been widely employed to model earthquake data [117, 41, 42]. For a complete overview of the properties of Hawkes processes please refer to [128, 44], while for a review of their recent applications in a financial context see [45]. Here we detail how we build and estimate the model.

9.9.1 Multivariate Hawkes point processes

An N-dimensional Hawkes process is a point process characterised by the vector of intensities $\boldsymbol{\lambda}_t := \left(\lambda_t^1, \ldots, \lambda_t^N\right)^\intercal$, where the i-type intensity satisfies the relation

$$\lambda_t^i = \mu_t^i + \sum_{j=1}^{N} \sum_{t_k^j < t} \nu_j^i (t - t_k^j) \,,$$

where μ_t^i and ν_j^i are positive deterministic functions for all $i, j = 1, \ldots, N$. The set $\left\{t_k^j\right\}$ corresponds to the random sequence of increasing events associated with the j-component of the N-dimensional point process. If $\mu_t^i = \mu^i$ is a constant and the kernel function ν_j^i reduces identically to zero, then the Hawkes point process describing the i-component reduces to a Poisson process with constant intensity μ^i. On the contrary, if the kernel is positive, each time an event occurs for any component of the multidimensional process, the intensity λ_t^i increases by a positive amount.

9.9.2 Choice of the parametrization

As in most high-dimensional problems, the estimation of multivariate Hawkes processes is problematic because of the large number of parameters. In order to overcome the curse of dimensionality problem, in this study we choose a quite rigid parametrization of the kernel matrix, reducing significantly the number of free parameters. We also propose a method to estimate the model on data.

First of all, we assume that the vector $\boldsymbol{\mu} := \left(\mu_t^1, \ldots, \mu_t^N\right)^\intercal$ does not depend on time. Second, we consider the most common parametrization of the kernel in terms of exponential functions

$$\nu_j^i(t - t_k^j) := \alpha_{ij} \mathrm{e}^{-\beta_{ij}(t - t_k^j)} ,$$

with $\alpha_{ij} > 0$ and $\beta_{ij} > 0$ for all i, j. The parameter α_{ij} fixes the scale of the intensity process λ^i and provides the deterministic amount by which the j-type event at t_k^j shocks the intensity of the i-type process. The parameter β_{ij} describes the inverse of the time needed by the process i to lose memory of a count of process j.

The process is stationary if the spectral radius (i.e. the absolute value of the largest eigenvalue) of the matrix Γ of elements

$$\Gamma_{ij} = \frac{\alpha_{ij}}{\beta_{ij}}$$

is strictly smaller than one. In this case the unconditional expected intensities of the process reads

$$\mathbb{E}\left[\boldsymbol{\lambda}_t\right] = \left(\mathrm{I}_N - \Gamma\right)^{-1} \boldsymbol{\mu} , \tag{9.2}$$

where I_N is the N-dimensional identity matrix.

We make the following further assumptions:

- We assume that all the β_{ij} are equal to a constant value $\beta > 0$. This means that there is only one time scale characterizing the decay of the kernels.

- We impose the condition that $\boldsymbol{\mu} = \eta \mathbb{E}[\boldsymbol{\lambda}_t]$, with $0 < \eta < 1$. This means that the distribution of multiplicity in the observed process is the same as the distribution of the multiplicity in the baseline (or ancestor) process. In other words, the cross-excitation between the different components of the Hawkes process does not change the unconditional law of multiplicity. Notice that this assumption implies that

$$\Gamma \mathbb{E}[\boldsymbol{\lambda}_t] = (1-\eta)\mathbb{E}[\boldsymbol{\lambda}_t] \ ,$$

  i.e. $\mathbb{E}[\boldsymbol{\lambda}_t]$ (or $\boldsymbol{\mu}$) is the eigenvector of Γ with eigenvalue $1-\eta$.

- The generic matrix element Γ_{ij} describing the intensity of the excitation of variable j on variable i is the product of a term D_{ii} which depends on the excited variable and a term $\sigma(|i-j|)$ which depends on the absolute difference of the two multiplicities. Therefore we can rewrite $\Gamma = D\Sigma$, where D is a diagonal matrix of elements

$$D_{ii} := \frac{(1-\eta)\mu^i}{\sum_{j=1}^{N} \mu^j \sigma(|i-j|)},$$

  and $\Sigma_{ij} = \sigma(|i-j|)$.

- Finally, we parametrize the matrix Σ as

$$\Sigma_{ij} = \sigma(|i-j|) = (|i-j|+1)^{-\gamma}$$

  This hyperbolic decay is chosen to model with only one parameter γ the strong cross-excitation between two very different multiplicities.

The model is therefore parametrised by the vector $\boldsymbol{\mu}$ and the three parameters η, γ, and β.

Before presenting the estimation procedure, we discuss some properties of the model. As all the entries of Γ are strictly positive, the Perron-Frobenius Theorem applies. Then, there exists only one eigenvector with all strictly positive components, and the associated eigenvalue is the spectral radius. Since $\mathbb{E}[\lambda_t^i] > 0$ for all $i = 1, \ldots, N$, we conclude that the spectral radius is $1-\eta$. Incidentally, we notice that all the eigenvalues of Γ are real. This property readily follows from observing that Γ is the product of two symmetric matrices, and D is diagonal and positive definite. Indeed, denoting with $\sqrt{D}$ the square root of the matrix D, Γ is *similar* to $\sqrt{D}^{-1} D\Sigma\sqrt{D}$, which is by construction symmetric. Moreover, if Γ is diagonal dominant, i.e. if $|\Gamma_{ii}| > \sum_{j\neq i} |\Gamma ij|$ for $i = 1, \ldots, N$, the eigenvalues are also strictly positive.

9.9.3 Estimation of the model parameters

A rigorous estimation of our model's parameters through likelihood maximization poses several computational problems. We instead propose a heuristic and robust calibration procedure based on moments. In particular we consider the following two conditional expectations, whose values on real and simulated data are graphed in Figure 9.4:

$$f_\tau^{(1)}(M;J) := \mathbb{P}\Big[\exists t' \in (t, t+\tau] \text{ s.t. } M_{t'} \geq J \Big| M_t \geq M\Big], \tag{9.3}$$

$$f_\tau^{(2)}(M) := \mathbb{E}\Big[M_{t'} \Big| M_t \geq M,\, \exists t' \in (t, t+\tau] \text{ s.t. } M_{t'} > 0\Big]. \tag{9.4}$$

The first quantity, $f_\tau^{(1)}(M;J)$, is the probability of observing a systemic event with multiplicity at least J inside a time interval of length τ after a cojump of multiplicity M_t larger than or equal to M. It therefore measures the probability that a cojump of multiplicity at least M triggers a systemic cojump (J fixes the threshold for a systemic cojump). The second quantity, $f_\tau^{(2)}(M)$, is the average multiplicity of the cojumps inside a time interval of length τ after a cojump of multiplicity M_t larger than or equal to M. It therefore measures the typical cojump multiplicity triggered by a cojump of multiplicity at least M.

We use $f_\tau^{(1)}(M;J)$ (for fixed J and τ) and $f_\tau^{(2)}(M)$ (for fixed τ) to estimate via a weighted least squares approach the three model parameters η, γ, and β. Since we are not able to compute analytically the moments of $f_\tau^{(1)}(M;J)$ and $f_\tau^{(2)}(M)$ from the model, we perform Monte Carlo simulations with fixed parameters. Specifically, given a multiplicity M, let the data and the model conditional expectations of any of the quantity in Equations (9.3) and (9.4) be represented by their average values $a_{\mathrm{d}}(M)$, $a_{\mathrm{m}}(M)$ and standard errors $\delta_{\mathrm{d}}(M)$, $\delta_{\mathrm{m}}(M)$. Then, for the expectation $f_\tau^{(i)}$ $(i = 1, 2)$ we construct the loss function

$$\chi^2_{(i)} = \sum_{M \in S} \frac{(a_{\mathrm{d}} - a_{\mathrm{m}})^2}{\delta_{\mathrm{d}}^2 + \delta_{\mathrm{m}}^2}, \tag{9.5}$$

where the sum is taken over a set of multiplicities S. We then construct the total loss function $\chi^2_{(1)} + 0.5\chi^2_{(2)}$ and we search for the model parameters which minimize the loss function. Given the small number of parameters we explore a large region of the three-dimensional space of parameters on a 0.05-spaced grid.

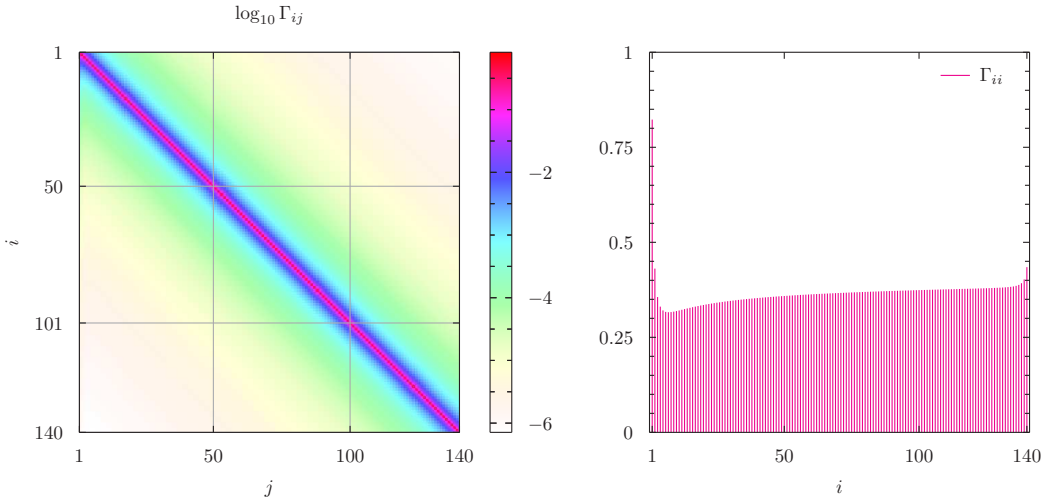

Figure 9.6: Left panel: Logarithmic entries of the matrix $\Gamma_{ij} := \alpha_{ij}/\beta_{ij}$ for $\beta_{ij} = \beta = 0.6$ for all $i, j = 1, \ldots, 140$, $\eta = 0.15$, and $\gamma = 2.65$. Right panel: Linear plot of the diagonal entries of Γ as a function of the multiplicity i.

9.9.4 Results for the investigated dataset

As an example of the estimation procedure and to discuss the properties of the fitted model, we consider in detail the case of $N = 140$ highly liquid assets of the Russell 3000 Index in 2013. The same set is used also in Figure 9.4. We fix $J = 10$ in Equation (9.3), $\tau = 5$ in Equations (9.3) and (9.4), $S = \{5, 10, 15, \ldots, 65, 70\}$ and look for the parameters that minimise the total loss function. Following this approach, we find a clear minimum corresponding to the values $\eta = 0.15$, $\beta = 0.6$, $\gamma = 2.65$.

The left panel of Figure 9.6 reports the logarithmic value of 140×140 entries of the Γ matrix. Coherently with the definitions given above, Γ_{ij} for fixed i, is the impact of past events with multiplicity j on the multiplicity i. The largest value corresponds to the diagonal term $\Gamma_{ii} = D_{ii}$ and quantifies the shock of the intensity due to a self-exciting effect. Then, moving away from the Γ_{ii}, the kernel matrix decreases symmetrically along the row according to a hyperbolic scaling with tail index $\gamma = 2.65$. The parameter η rescales the level of the main diagonal of the matrix Γ, reported in the right panel of Figure 9.6, and determines the degree of stationarity of the process. In Figure 9.7 we plot the complete spectrum of the matrix Γ. As expected, the largest value corresponds to $1 - \eta = 0.85$, while the positive definiteness of all the eigenvalues follows from the evidence, verified numerically, that the matrix is diagonal dominant. More specifically, for the chosen values of η, β, and γ the matrix Γ is determined uniquely through the specification of

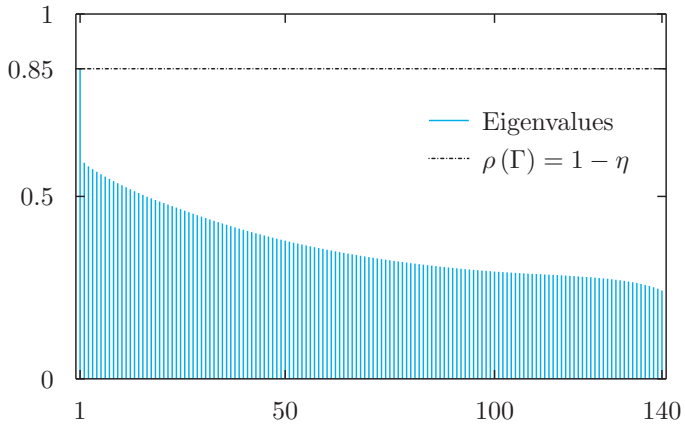

Figure 9.7: Eigenvalue spectrum of the matrix Γ. The spectral radius $\rho(\Gamma)$ corresponds to $1 - \eta$. Since $\eta = 0.15$, and more generally for $0 < \eta < 1$, the multidimensional Hawkes process describing the stochastic evolution of the multiplicity remains stationary. For the chosen parameter values, we verified numerically that Γ satisfies the diagonal dominant condition and so all its eigenvalues are strictly positive.

the vector of expected intensities, $\mathbb{E}[\boldsymbol{\lambda}_t]$. In our numerical experiment we replace the vector of expected intensities multiplied by the length of the time series, i.e. 96,861, with the empirical frequencies observed for the 140 assets from the Russell 3000 Index in 2013. Figure 9.8 conveys this information in terms of the Complementary of the Cumulative Distribution Function of the cojump multiplicities associated with the empirical data (bold line). We also report the same quantity measured from a synthetic time series corresponding to a Monte Carlo simulation of the 140-dimensional Hawkes process (dashed line).

9.9.5 Model results for the years 2001–2012

As supplement to the model results for the year 2013 presented in Section 9.5.2, we present here a complete view of our model's behaviour for the years 2001–2012. Figures 9.9–9.20 show the analogous of Figure 9.4 for the twelve years. Values of $f_\tau^{(1)}(M; J)$ equal to 0 or 1 have a standard error equal to 0 since they are the average of all zero or all one values, respectively. Values with undefined standard error appear in correspondence of points where the empirical conditioning selects only one event. Our model provides a good description of the empirical multiplicity dependence structure for some years, although for some others it does not accurately reproduce the empirical multiplicity dependencies. We note that our model is nonetheless

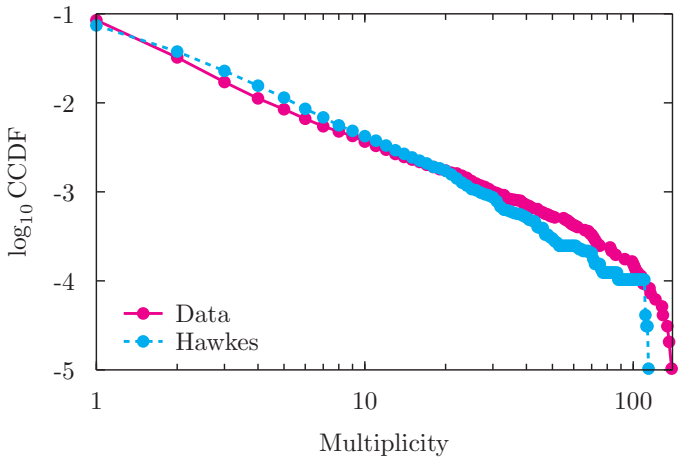

Figure 9.8: Log-log plot of the Complementary of the Cumulative Distribution Function of the cojump multiplicities. The bold line corresponds to the empirical distribution measured from the Russell 3000 data sample, 140 assets, during year 2013. The dashed line is the distribution obtained from a simulation of the multidimensional Hawkes process. The total number of minutes drawn from the simulation coincides with the length of the empirical time series and is equal to 96861.

a major improvement with respect to the shuffling benchmark for all years, and that it is always able to reproduce the multiplicity dependencies to a certain extent.

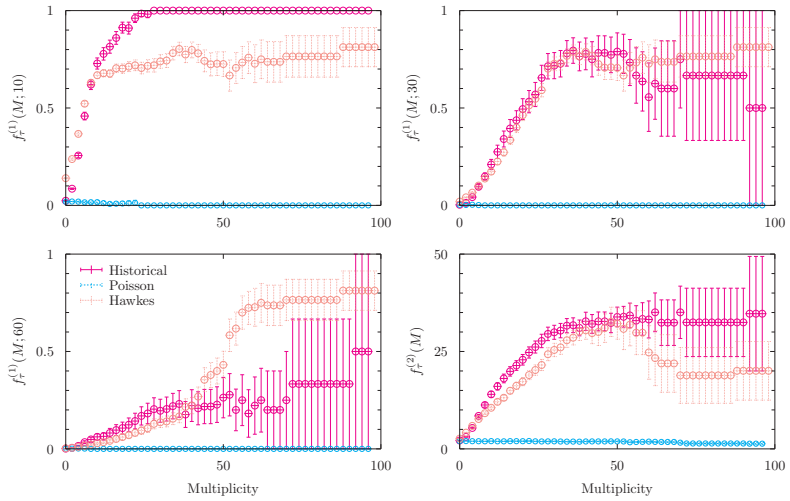

Figure 9.9: Analogous of Figure 9.4. Results of the model for the 2001 data. Estimated parameters are: $\eta = 0.1$, $\beta = 0.55$, $\gamma = 2.8$.

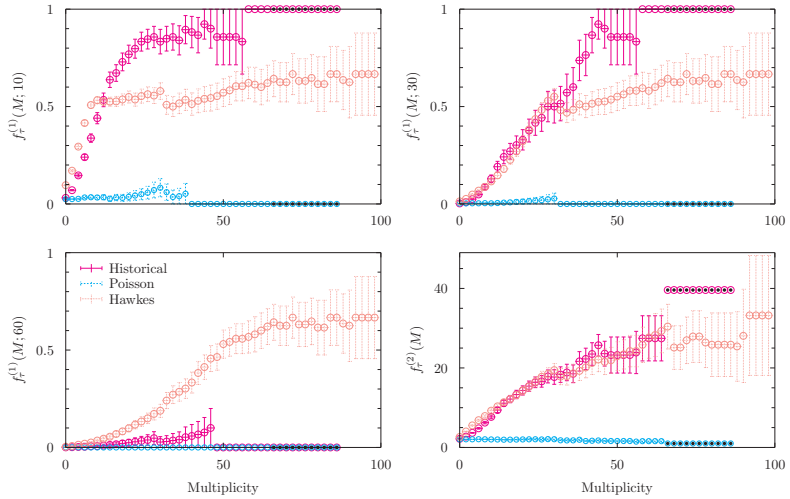

Figure 9.10: Analogous of Figure 9.4. Results of the model for the 2002 data. Estimated parameters are: $\eta = 0.25$, $\beta = 0.7$, $\gamma = 2.8$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

Table 9.1: Number of news announcements, organised by year and news category.

		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	all years
ADP Employment Report	MEA[a]	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Beige Book	MEA	8	8	8	7	8	8	8	8	8	8	8	8	8	103
Business Inventories	MEA	0	0	3	6	5	12	12	12	12	12	12	12	12	110
Chairman Press Conference	MMI[b]	0	0	0	0	0	0	0	0	0	0	3	5	4	12
Chicago PMI	MEA	12	12	12	12	12	12	12	12	12	12	12	12	11	155
Construction Spending	MEA	12	12	12	12	12	12	12	12	12	12	12	12	12	156
Consumer Confidence	MEA	12	12	12	12	12	12	12	12	12	13	12	12	11	156
Consumer Sentiment	MEA	12	23	24	23	24	24	24	24	24	24	24	24	24	298
Dallas Fed Mfg Survey	MEA	0	0	0	0	0	0	0	0	0	0	8	12	12	32
Durable Goods Orders	MMI	1	1	0	0	0	0	0	0	0	0	0	0	0	2
EIA Petroleum Status Report	MEA	0	0	0	14	51	52	52	53	52	52	52	52	52	482
Empire State Mfg Survey	MEA	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Existing Home Sales	MMI	12	12	12	12	12	12	12	12	12	12	12	12	12	156
Factory Orders	MEA	12	12	12	12	12	12	12	12	12	12	12	12	12	156
FOMC Forecasts	MMI	0	0	0	0	0	0	0	0	0	0	0	5	4	9
FOMC Meeting Announcement	MMI	8	8	8	8	8	8	8	8	8	8	8	8	8	104
FOMC Minutes	MMI	8	8	8	8	8	8	8	8	8	8	8	8	7	103
Housing Market Index	MEA	0	0	0	0	0	12	12	12	12	12	12	12	12	96
ISM Mfg Index	MMI	12	12	12	12	12	12	12	12	12	12	12	12	12	156
ISM Non-Manufacturing Employment Index	MEA	0	0	0	0	0	0	0	11	12	0	0	0	0	23
Motor Vehicle Sales	MEA	0	12	0	0	0	0	0	0	0	0	0	0	0	12
New Home Sales	MMI	13	12	12	12	12	12	12	12	12	12	12	12	12	157
Pending Home Sales Index	MEA	0	0	0	0	0	12	12	12	12	13	12	12	12	97
Personal Income and Outlays	MMI	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Philadelphia Fed Survey	MMI	12	12	12	12	12	12	12	12	12	12	12	12	12	156
Retail Sales	MMI	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Treasury Budget	MEA	12	12	12	12	12	12	12	11	12	12	12	12	10	153
all categories		147	169	159	174	212	245	244	256	256	246	255	266	259	2888

[a]Merit Extra Attention according to the classification provided by Econoday, Inc.

[b]Market Moving Indicator according to the classification provided by Econoday, Inc.

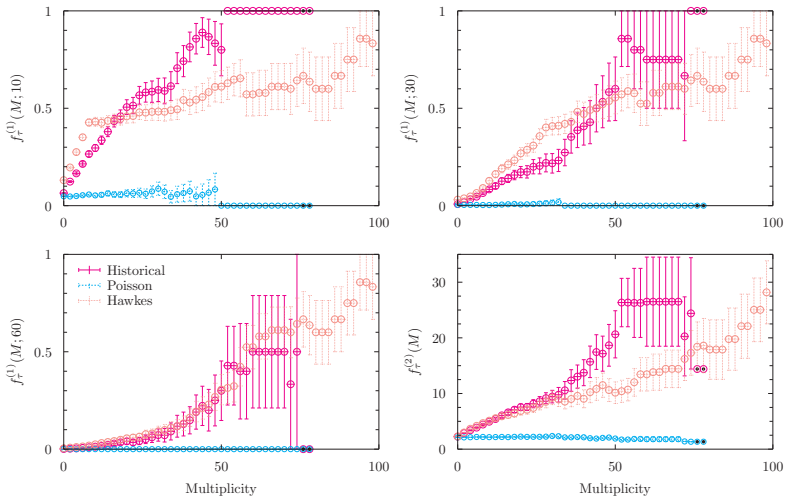

Figure 9.11: Analogous of Figure 9.4. Results of the model for the 2003 data. Estimated parameters are: $\eta = 0.1$, $\beta = 0.45$, $\gamma = 2.5$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

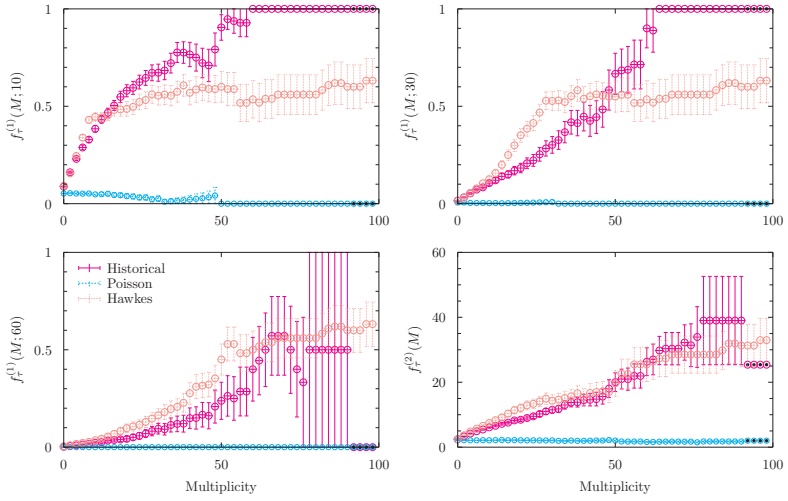

Figure 9.12: Analogous of Figure 9.4. Results of the model for the 2004 data. Estimated parameters are: $\eta = 0.2$, $\beta = 0.7$, $\gamma = 2.55$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

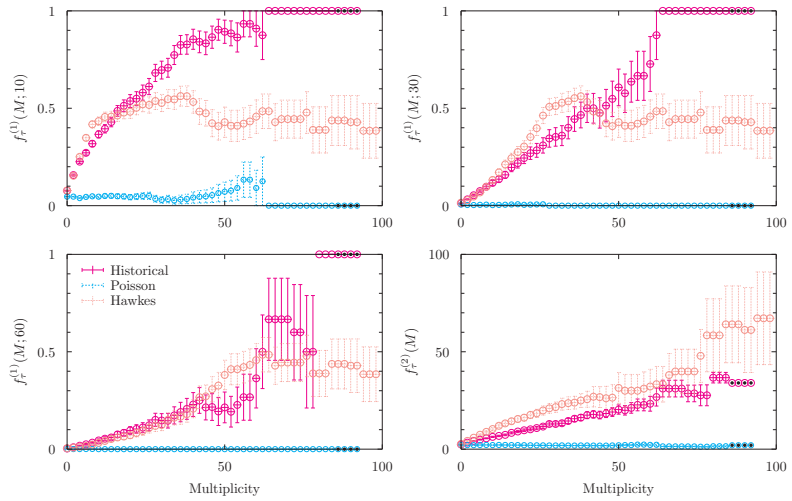

Figure 9.13: Analogous of Figure 9.4. Results of the model for the 2005 data. Estimated parameters are: $\eta = 0.25$, $\beta = 0.65$, $\gamma = 2.6$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

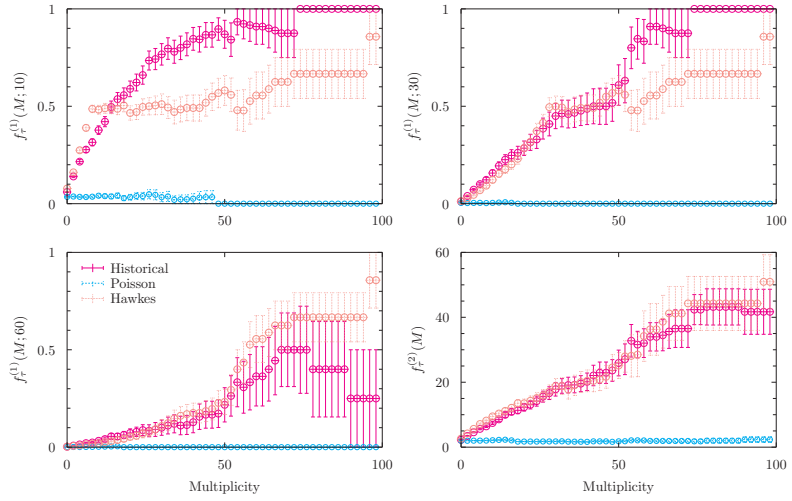

Figure 9.14: Analogous of Figure 9.4. Results of the model for the 2006 data. Estimated parameters are: $\eta = 0.25$, $\beta = 0.5$, $\gamma = 2.8$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

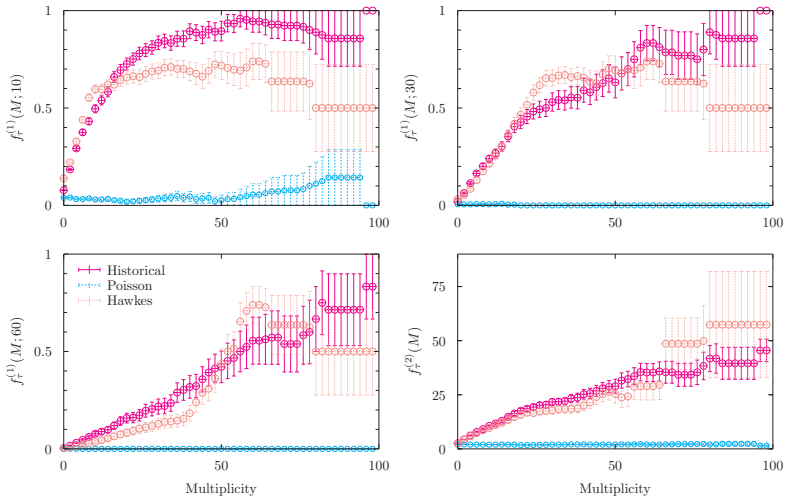

Figure 9.15: Analogous of Figure 9.4. Results of the model for the 2007 data. Estimated parameters are: $\eta = 0.1$, $\beta = 0.65$, $\gamma = 2.65$.

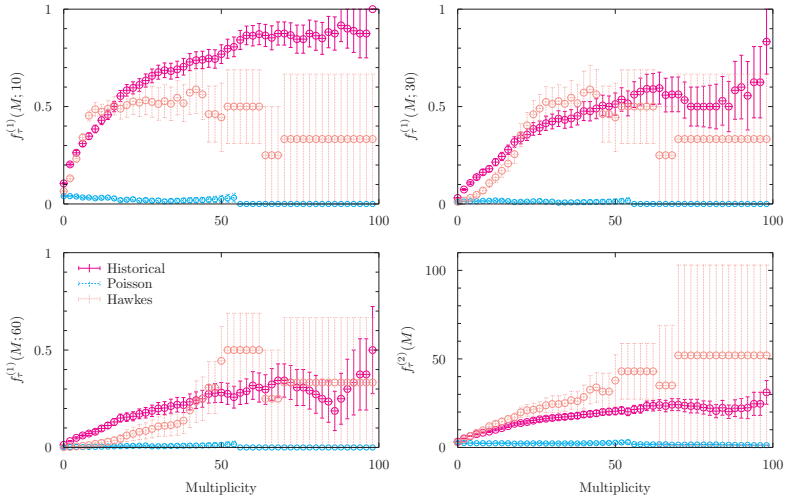

Figure 9.16: Analogous of Figure 9.4. Results of the model for the 2008 data. Estimated parameters are: $\eta = 0.25$, $\beta = 0.55$, $\gamma = 2.8$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

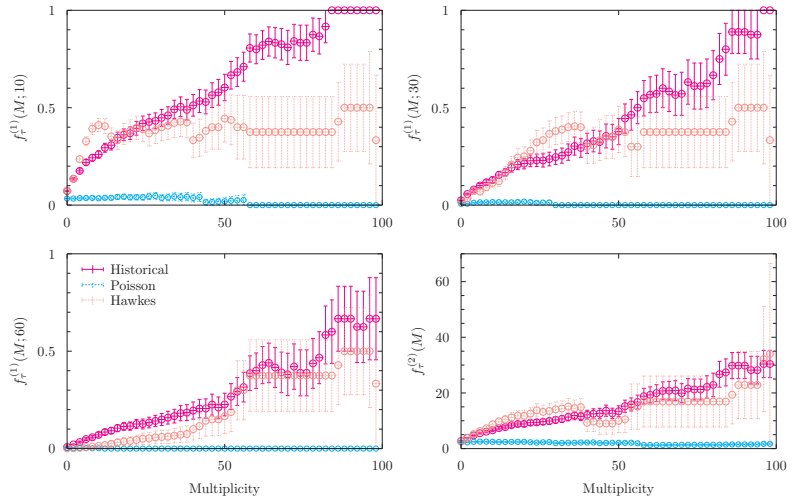

Figure 9.17: Analogous of Figure 9.4. Results of the model for the 2009 data. Estimated parameters are: $\eta = 0.15$, $\beta = 0.45$, $\gamma = 2.6$.

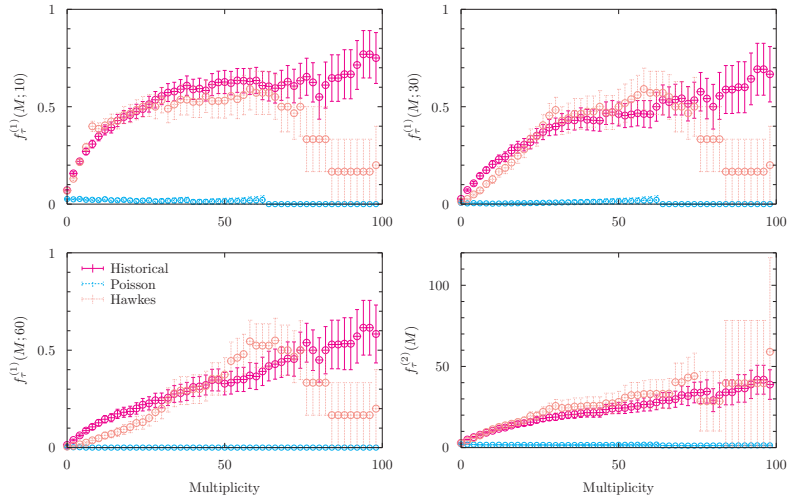

Figure 9.18: Analogous of Figure 9.4. Results of the model for the 2010 data. Estimated parameters are: $\eta = 0.25$, $\beta = 0.5$, $\gamma = 2.5$.

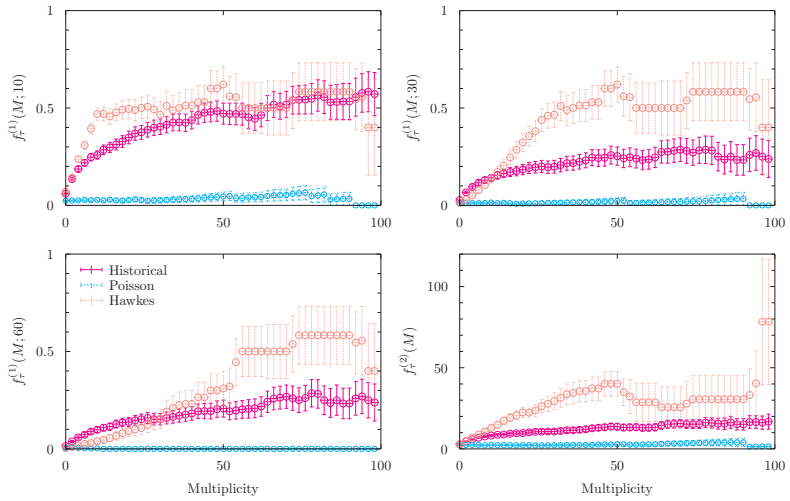

Figure 9.19: Analogous of Figure 9.4. Results of the model for the 2011 data. Estimated parameters are: $\eta = 0.25$, $\beta = 0.75$, $\gamma = 2.6$. Circles filled with a black dot correspond to single-event measurements for which standard deviation is not defined.

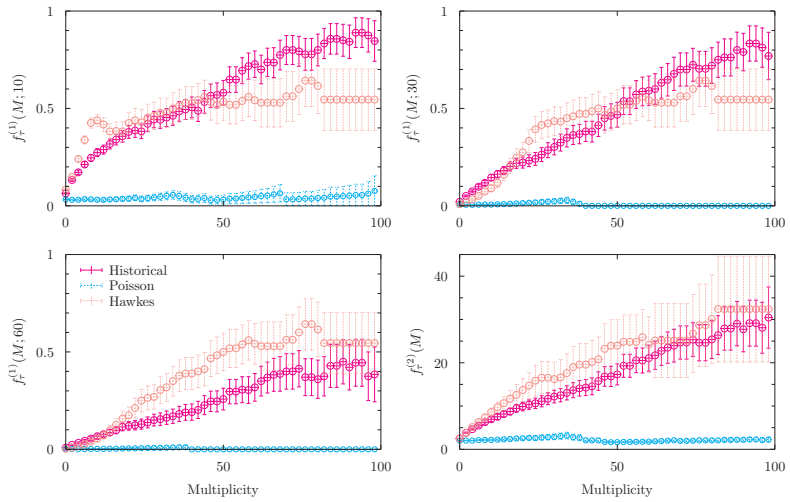

Figure 9.20: Analogous of Figure 9.4. Results of the model for the 2012 data. Estimated parameters are: $\eta = 0.15$, $\beta = 0.5$, $\gamma = 2.65$.

Bibliography

[1] Thomas M. Cover, Joy A. Thomas, *Elements of information theory*, Wiley-Interscience, 2nd edition (2006).

[2] Martin Sewell, *Chaos in financial markets* (2008).

[3] David A Hsieh, *Chaos and nonlinear dynamics: application to financial markets*, The journal of finance **46** (1991), 1839–1877.

[4] Benoit B Mandelbrot, *Fractal and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*, Springer (1997).

[5] George C Philippatos, Charles J Wilson, *Entropy, market risk, and the selection of efficient portfolios*, Applied Economics **4** (1972), 209–220.

[6] Jianshe Ou, *Theory of portfolio and risk based on incremental entropy*, The Journal of Risk Finance **6** (2005), 31–39.

[7] Ilhan Usta, Yeliz Mert Kantar, *Mean-variance-skewness-entropy measures: A multi-objective approach for portfolio selection*, Entropy **13** (2011), 117–133.

[8] Les Gulko, *Dart boards and asset prices: Introducing the entropy pricing theory*, Advances in Econometrics **12** (1997), 237–276.

[9] Les Gulko, *The entropy theory of stock option pricing*, International Journal of Theoretical and Applied Finance **2** (1999), 331–355.

[10] Les Gulko, *The entropy theory of bond option pricing*, International Journal of Theoretical and Applied Finance **5** (2002), 355–383.

[11] Peter W Buchen, Michael Kelly, *The maximum entropy distribution of an asset inferred from option prices*, Journal of Financial and Quantitative Analysis **31** (1996), 143–159.

[12] Gabjin Oh, Seunghwan Kim, Cheoljun Eom, *Market efficiency in foreign exchange markets*, Physica A **382** (2007), 209–212.

[13] Wiston Adrián Risso, *The informational efficiency: The emerging markets versus the developed markets*, Applied Economics Letters **16** (2009), 485–487.

[14] Leon Zingales, *Windows of predictability in financial markets: a shannon entropy approach*, Communications in Applied and Industrial Mathematics **2** (2011).

[15] Claude Elwood Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), 379–423, 623–656.

[16] John L. Kelly, Jr, *A new interpretation of information rate*, Bell System Technical Journal **35** (1956), 917–926.

[17] Henry Allen Latane, *Criteria for choice among risky ventures*, Journal of Political Economy **67** (1959), 144–155.

[18] L Breiman, *Optimal gambling systems for favorable games*, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 65–78.

[19] Andrew R Barron, Thomas M Cover, *A bound on the financial value of information*, Information Theory, IEEE Transactions on **34** (1988), 1097–1100.

[20] Louis Bachélier, *Théorie de la spéculation*, Ph.D. thesis (1900).

[21] Paul Anthony Samuelson, *Proof that properly anticipated prices fluctuate randomly*, Industrial Management Review **6** (1965), 41–49.

[22] Eugene F. Fama, *Random walks in stock-market prices*, Financial Analysts Journal **21** (1965), 55–59.

[23] Eugene Francis Fama, *Mandelbrot and the stable paretian hypothesis*, Journal of business **36** (1963), 420–429.

[24] Eugene Francis Fama, *The behavior of stock market prices*, Journal of business **38** (1965), 34–105.

[25] Eugene F. Fama, *Efficient capital markets: a review of theory and empirical work*, Journal of Finance **25** (1970), 383–417.

[26] Andrew W. Lo, *Efficient market hypothesis*, in L. Blume, S. Durlauf, editors, *The New Palgrave: A Dictionary of Economics*, Palgrave McMillan, New York, 2nd edition (2007).

[27] John Y. Campbell, Andrew W. Lo, A. Craig MacKinlay, *The econometrics of financial markets*, Princeton University Press (1997).

[28] R. Giglio, R. Matsushita, A. Figueiredo, I. Gleria, S. Da Silva, *Algorithmic complexity theory and the relative efficiency of financial markets*, EPL **84** (2008), 48005.

[29] Armin Shmilovici, Yael Alon-Brimer, Shmuel Hauser, *Using a stochastic complexity measure to check the efficient market hypothesis*, Computational Economics **22** (2003), 273–284.

[30] Armin Shmilovici, Yoav Kahiri, Irad Ben-Gal, Shmuel Hauser, *Measuring the efficiency of the intraday forex market with a universal data compression algorithm*, Computational Economics **33** (2009), 131–154.

[31] Daniel O. Cajueiro, Benjamin M. Tabak, *Ranking efficiency for emerging markets*, Chaos, Solitons and Fractals **22** (2004), 349–352.

[32] Jeffrey G MacIntosh, *High frequency traders: Angels or devils?*, CD Howe Institute Commentary **391** (2013).

[33] Terrence Hendershott, Ryan Riordan, *Algorithmic trading and information*, Manuscript, University of California, Berkeley (2009).

[34] Austin Gerig, *High-frequency trading synchronizes prices in financial markets*, Available at SSRN 2173247 (2013).

[35] *Findings regarding the market events of May 6, 2010. report of the staffs of the CFTC and SEC to the joint advisory committee on emerging regulatory issues*, Available online at: `http://www.sec.gov/news/studies/2010/marketevents-report.pdf` (2010).

[36] Andrei A Kirilenko, Albert S Kyle, Mehrdad Samadi, Tugkan Tuzun, *The Flash Crash: The impact of high frequency trading on an electronic market*, Available at SSRN 1686004 (2014).

[37] Neil Johnson, Guannan Zhao, Eric Hunsader, Hong Qi, Nicholas Johnson, Jing Meng, Brian Tivnan, *Abrupt rise of new machine ecology beyond human response time*, Scientific reports **3** (2013).

[38] Anton Golub, John Keane, Ser-Huang Poon, *High frequency trading and mini flash crashes*, arXiv:1211.6667 (2012).

[39] Rama Cont, P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall - CRC (2004).

[40] A. G. Hawkes, *Spectra of some self-exciting and mutually exciting point processes*, Biometrika **58** (1971), 83–90.

[41] D. Vere-Jones, T. Ozaki, *Some examples of statistical inference applied to earthquake data*, Annals of the Institute of Statistical Mathematics **34** (1982), 189–207.

[42] Y. Ogata, *Statistical models for earthquake occurrences and residual analysis for point processes*, Journal of the American Statistical Association **83** (1988), 9–27.

[43] Clive G Bowsher, *Modelling security market events in continuous time: Intensity based, multivariate point process models*, Journal of Econometrics **141** (2007), 876–912.

[44] Luc Bauwens, Nikolaus Hautsch, *Modelling financial high frequency data using point processes*, in T. Mikosch, J.-P. Kreiss, R. A. Davis, T. G. Andersen, editors, *Handbook of Financial Time Series*, Springer, Berlin, 953–979.

[45] E. Bacry, I. Mastromatteo, J. F. Muzy, *Hawkes processes in finance*, Market Microstructure and Liquidity **1** (2015), 1550005.

[46] P. C. Shields, *The ergodic theory of discrete sample paths*, American Mathematical Society (1996).

[47] P. Walters, *An introduction to ergodic theory*, Springer-Verlag, New York (1982).

[48] K. Petersen, *Ergodic theory*, Cambridge University Press (1983).

[49] G. D. Birkhoff, *Proof of the ergodic theorem*, Proceedings of the National Academy of Sciences USA **17** (1931), 656–660.

[50] A. I. Khinchin, *Mathematical foundations of information theory*, Dover, New York (1957).

[51] A. N. Kolmogorov, *A new metric invariant of transitive dynamical systems and automorphisms of lebesgue spaces*, Doklady Akademii Nauk SSSR **119** (1958), 861–864. In Russian.

[52] Yakov G. Sinai, *On the notion of entropy of a dynamical system*, Doklady Akademii Nauk SSSR **124** (1959), 768–771. In Russian.

[53] W. Krieger, *On entropy and generators of measure-preserving transformations*, Transactions of the American Mathematical Society **149** (1970), 453–464.

[54] W. Ebeling, M. A. Jiménez-Montaño, *On grammars, complexity, and information measures of biological macromolecules*, Mathematical Biosciences **52** (1980), 53–71.

[55] M. A. Jiménez-Montaño, *On the syntactic structure of protein sequences and the concept of grammar complexity*, Bulletin of Mathematical Biology **42** (1984), 641–659.

[56] P. E. Rapp, I. D. Zimmermann, E. P. Vining, N. Cohen, A. M. Albano, M. A. Jiménez-Montaño, *The algoritmic complexity of neural spike trains increases during focal seizures*, The Journal of Neuroscience **14** (1994), 4731–4739.

[57] Peter Grassberger, *Data compression and entropy estimates by non-sequential recursive pair substitution*, arXiv:physics/0207023 (2002).

[58] D. Benedetto, E. Caglioti, D. Gabrielli, *Non-sequential recursive pair substitution: some rigorous results*, Journal of Statistical Mechanics (2006), P09011.

[59] G. A. Miller, *Note on the bias of information estimates*, Information Theory in Psychology: Problems and Methods (1955), 95–100.

[60] Peter Grassberger, *Entropy estimates from insufficient samplings*, arXiv:physics/0307138v2 (2008).

[61] M. A. Jiménez-Montaño, W. Ebeling, T. Pohl, P. E. Rapp, *Entropy and complexity of finite sequences as fluctuating quantities*, BioSystems **64** (2002), 23–32.

[62] Juan A. Bonachela, Haye Hinrichsen, Miguel A. Muñoz, *Entropy estimates of small data sets*, Journal of Physics A: Mathematical and Theoretical **41** (2008), 202001.

[63] D. Ornstein, B. Weiss, *How sampling reveals a process*, The Annals of Probability **18** (1990), 905–930.

[64] F. Ledrappier, L.-S. Young, *The metric entropy of diffeomorphisms. ii. relations between entropy, exponents and dimension*, Annals of Mathematics **122** (1985), 540–574.

[65] F. Hofbauer, P. Raith, *The hausdorff dimension of an ergodic invariant measure for a piecewise monotonic map of the interval*, Canadian Mathematical Bulletin **35** (1992), 84–98.

[66] Marcelo Viana, *Stochastic Dynamics of Deterministic Systems*, Lecture Notes XXI Brazilian Mathematics Colloquium, Instituto de Matemática Pura e Aplicada, Rio de Janeiro (1997).

[67] Jérôme Buzzi, *A minicourse on entropy theory on the interval*, arXiv:math/0611337 (2006).

[68] T. Steinberger, *Local dimension of ergodic measures for two-dimensional lorenz transformations*, Ergodic Theory and Dynamical Systems **20** (2000), 911–923.

[69] J. C. Sprott, G. Rowlands, *Improved correlation dimension calculation*, International Journal of Bifurcation and Chaos **11** (2001), 1865–1880.

[70] D. Dolgopyat, *On mixing properties of compact group extensions of hyperbolic systems*, Israel Journal of Mathematics **130** (2002), 157–205.

[71] N. J. Larsson, A. Moffat, *Offline dictionary-based compression*, Proceedings of the IEEE **88** (2000), 1722–1732.

[72] V. Benci, C. Bonanno, S. Galatolo, G. Menconi, M. Virgilio, *Dynamical systems and computable information*, Discrete and Continuous Dynamical Systems - B **4** (2004), 935–960.

[73] Y. Gao, I. Kontoyiannis, E. Bienenstock, *Estimating the entropy of binary time series: Methodology, some theory and a simulation study*, Entropy **10** (2008), 71–99.

[74] Gideon Weiss, *Time-reversibility of linear stochastic processes*, Journal of Applied Probability (1975), 831–836.

[75] Alfred Cowles, 3rd, Herbert Jones, *Some a posteriori probabilities in stock market action*, Econometrica **5** (1937), 280–294.

[76] Abraham Lempel, Jacob Ziv, *On the complexity of finite sequences*, IEEE Transactions on Information Theory **22** (1976), 75–81.

[77] F. Kaspar, H. G. Schuster, *Easily calculable measure for the complexity of spatiotemporal patterns*, Physical Review A **36** (1987), 842–848.

[78] Jorma Rissanen, *A universal data compression system*, IEEE Transactions on Information Theory **29** (1983), 656–664.

[79] Stephen J. Taylor, *Asset price dynamics, volatility, and prediction*, Princeton University Press (2011).

[80] Christian T. Brownlees, Giampiero M. Gallo, *Financial econometric analysis at ultra-high frequency: Data handling concerns*, Computational Statistics & Data Analysis **51** (2006), 2232–2245.

[81] T. G. Andersen, T. Bollerslev, *Intraday periodicity and volatility persistence in financial markets*, Journal of Empirical Finance **4** (1997), 115–158.

[82] Ole E. Barndorff-Nielsen, Neil Shephard, *Power and bipower variation with stochastic volatility and jumps*, Journal of Financial Econometrics **2** (2004), 1–48.

[83] Yacine Aït-Sahalia, Per A Mykland, Lan Zhang, *Ultra high frequency volatility estimation with dependent microstructure noise*, Journal of Econometrics **160** (2011), 160–175.

[84] Hirotugu Akaike, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control **19** (1974), 716–723.

[85] Gideon Schwarz, *Estimating the dimension of a model*, Annals of Statistics **6** (1978), 461–464.

[86] Yacine Aït-Sahalia, *Disentangling diffusion from jumps*, Journal of Financial Economics **74** (2004), 487–528.

[87] Ole E. Barndorff-Nielsen, Neil Shephard, *Econometrics of testing for jumps in financial economics using bipower variation*, Journal of Financial Econometrics **4** (2006), 1–30.

[88] Jianqing Fan, Yazhen Wang, *Multi-scale jumps and volatility analysis for hig-frequency financial data*, Journal of the American Statistical Association **102** (2007), 1349–1362.

[89] G. J. Jiang, R. C. A. Oomen, *Testing for jumps when asset prices are observed with noise–a "swap variance" approach*, Journal of Econometrics **144** (2008), 352–370.

[90] Yacine Aït-Sahalia, J. Jacod, *Testing for jumps in a discretely observed process*, Annals of Statistics **37** (2009), 184–222.

[91] S. S. Lee, J. Hannig, *Detecting jumps from lévy jump diffusion processes*, Journal of Financial Economics **96** (2010), 271–290.

[92] Fulvio Corsi, Davide Pirino, Roberto Renò, *Threshold bipower variation and the impact of jumps on volatility forecasting*, Journal of Econometrics **159** (2010), 276–288.

[93] Rituparna Sen, *Jumps and microstructure noise in stock price volatility*, in Greg G. Gregoriou, editor, *Stock Market Volatility*, Chapman & Hall/CRC (2010), 163–178.

[94] John Maheu, Thomas McCurdy, *News arrival, jump dynamics and volatility components for individual stock returns*, Journal of Finance **59** (2004), 755–793.

[95] S. R. Das, R. Uppal, *Systemic Risk and International Portfolio Choice*, The Journal of Finance **59** (2004), 2809–2834.

[96] Torben G Andersen, Tim Bollerslev, Dobrislav Dobrev, *No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications*, Journal of Econometrics **138** (2007), 125–180.

[97] J. H. Wright, Hao Zhou, *Bond risk premia and realized jump volatility* (2007). Working paper, Federal Reserve Board.

[98] T. Bollerslev, T. H. Law, G. Tauchen, *Risk, jumps and diversification*, Journal of Econometrics **144** (2008), 234–256.

[99] T. Bollerslev, U. Kretschmer, C. Pigorsch, G. Tauchen, *A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects*, Journal of Econometrics **150** (2009), 151–166.

[100] D. Duffie, J. Pan, K. Singleton, *Transform analysis and asset pricing for affine jump-diffusions*, Econometrica **68** (2000), 1343–1376.

[101] Bjorn Eraker, Michael Johannes, Nicholas Polson, *The impact of jumps in equity index volatility and returns*, Journal of Finance **58** (2003), 1269–1300.

[102] B. Eraker, *Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices*, The Journal of Finance **59** (2004), 1367–1404.

[103] Tim Bollerslev, Viktor Todorov, Sophia Zhengzi Li, *Jump tails, extreme dependencies, and the distribution of stock returns*, Journal of Econometrics **172** (2013), 307–324.

[104] T. Bollerslev, V. Todorov, *Tails, fears, and risk premia*, The Journal of Finance **66** (2011), 2165–2211.

[105] Ole E. Barndorff-Nielsen, Neil Shephard, *Variation, jumps, market frictions and high frequency data in financial econometrics*, in *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, Cambridge University Press (2007), 328–372.

[106] M. Dungey, M. McKenzie, L. V. Smith, *Empirical evidence on jumps in the term structure of the us treasury market*, Journal of Empirical Finance **16** (2009), 430–445.

[107] J. Lahaye, S. Laurent, C. J. Neely, *Jumps, cojumps and macro announcements*, Journal of Applied Econometrics **26** (2011), 893–921.

[108] M. Dungey, L. Hvozdyk, *Cojumping: Evidence from the us treasury bond and futures markets*, Journal of Banking & Finance (2012).

[109] Dudley Gilder, Mark B Shackleton, Stephen J Taylor, *Cojumps in stock prices: Empirical evidence*, Journal of Banking & Finance **40** (2014), 443–459.

[110] F. Gobbi, C. Mancini, *Diffusion covariation and co-jumps in bidimensional asset price processes with stochastic volatility and infinite activity lévy jumps*, in *SPIE Fourth International Symposium on Fluctuations and Noise*, International Society for Optics and Photonics, 660111–660111.

[111] J. Jacod, V. Todorov, *Testing for common arrivals of jumps for discretely observed multidimensional processes*, The Annals of Statistics **37** (2009), 1792–1838.

[112] Y. Aït-Sahalia, J. Cacho-Diaz, R. J. A. Laeven, *Modeling financial contagion using mutually exciting jump processes*, Technical report, National Bureau of Economic Research (2012).

[113] Suzanne S Lee, Per A Mykland, *Jumps in financial markets: A new nonparametric test and jump dynamics*, Review of Financial Studies **21** (2008), 2535–2563.

[114] Ana-Maria Dumitru, Giovanni Urga, *Identifying jumps in financial assets: a comparison between nonparametric jump tests*, Journal of Business & Economic Statistics **30** (2012), 242–255.

[115] *Flash Crash Analysis Continuing Developments*. `http://www.nanex.net`.

[116] Neil Johnson, Guannan Zhao, Eric Hunsader, Jing Meng, Amith Ravindar, Spencer Carran, Brian Tivnan, *Financial black swans driven by ultrafast machine ecology*, Available at SSRN 2003874 (2012).

[117] D. Vere-Jones, *Stochastic models for earthquake occurrence*, Journal of the Royal Statistical Society, Series B **32** (1970), 1–62.

[118] I. Muni Toke, *"Market making" in an order book model and its impact on the spread*, in F. Abergel, B. K. Chakrabarti, A. Chakraborti, M. Mitra, editors, *Econophysics of Order-Driven Markets*, Springer-Verlag, Milan, 49–64.

[119] I. Muni Toke, F. Pomponio, *Modelling trades-through in a limit order book using Hawkes processes*, Economics: The Open-Access, Open-Assessment E-Journal **6** (2012), 1–23.

[120] E. Bacry, S. Delattre, M. Hoffmann, J. F. Muzy, *Modelling microstructure noise with mutually exciting point processes*, Quantitative Finance **13** (2013), 65–77.

[121] D. Filimonov, D. Sornette, *Quantifying reflexivity in financial markets: Toward a prediction of flash crashes*, Physical Review E **85** (2012), 056108–1–9.

[122] Stephen J Hardiman, Nicolas Bercot, Jean-Philippe Bouchaud, *Critical reflexivity in financial markets: a Hawkes process analysis*, The European Physical Journal B **86** (2013), 1–9.

[123] Marcello Rambaldi, Paris Pennesi, Fabrizio Lillo, *Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach*, Physical Review E **91** (2015), 012819.

[124] Y. Aït-Sahalia, T. R. Hurd, *Portfolio choice in markets with contagion*, arXiv:1210.1598 (2012).

[125] *Rules of the markets organised and managed by Borsa Italiana S.p.A.*, Borsa Italiana S.p.A. (2015). `http://www.borsaitaliana.it/borsaitaliana/regolamenti/rulesno16-02-2015.en.pdf`.

[126] *Instructions accompanying the Rules of the Markets organised and managed by Borsa Italiana S.p.A.*, Borsa Italiana S.p.A. (2016). `http://www.borsaitaliana.it/borsaitaliana/regolamenti/istruzioni/instr11072016.en_pdf.htm`.

[127] A. Joulin, A. Lefevre, D. Grunberg, J.-P. Bouchaud, *Stock price jumps: News and volume play a minor role*, Wilmott Magazine **Sep/Oct** (2008), 1–7.

[128] D. J. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods*, Springer, Heidelberg (2003).

[129] Yoshiko Ogata, *The asymptotic behaviour of maximum likelihood estimators for stationary point processes*, Annals of the Institute of Statistical Mathematics **30** (1978), 243–261.

[130] Tohru Ozaki, *Maximum likelihood estimation of hawkes' self-exciting point processes*, Annals of the Institute of Statistical Mathematics **31** (1979), 145–155.

[131] Yosihiko Ogata, *On lewis' simulation method for point processes*, IEEE Transactions on Information Theory **27** (1981), 23–31.

[132] PAW Lewis, GS Shedler, *Simulation of nonhomogeneous poisson processes with log linear rate function*, Biometrika **63** (1976), 501–505.

[133] Michael Lewis, *Flash boys: a Wall Street revolt*, WW Norton & Company (2014).

[134] Peter Gomber, Björn Arndt, Marco Lutat, Tim Uhle, *High-frequency trading*, Available at SSRN 1858626 (2011).

[135] Giacomo Bormetti, Lucio Maria Calcagnile, Michele Treccani, Fulvio Corsi, Stefano Marmi, Fabrizio Lillo, *Modelling systemic price cojumps with Hawkes factor models*, Quantitative Finance **15** (2015), 1137–1156.

[136] Torben G Andersen, Tim Bollerslev, Per Frederiksen, Morten Ørregaard Nielsen, *Continuous-time models, realized volatilities, and*

testable distributional implications for daily stock returns, Journal of Applied Econometrics **25** (2010), 233–261.

[137] Massimiliano Caporin, Alexey Kolokolov, Roberto Renò, *Multi-jumps*, Available at SSRN 2488603 (2014).

[138] Alexander M Petersen, Fengzhong Wang, Shlomo Havlin, H Eugene Stanley, *Quantitative law describing market dynamics before and after interest-rate change*, Physical Review E **81** (2010), 066121.

[139] Alexander M Petersen, Fengzhong Wang, Shlomo Havlin, H Eugene Stanley, *Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws*, Physical Review E **82** (2010), 036114.

[140] Suzanne S Lee, *Jumps and information flow in financial markets*, Review of Financial Studies **25** (2012), 439–479.

[141] Pierre Bajgrowicz, Olivier Scaillet, Adrien Treccani, *Jumps in high-frequency data: Spurious detections, dynamics, and news*, Swiss Finance Institute Research Paper (2013).

[142] T. G. Andersen, T. Bollerslev, F. X. Diebold, P. Labys, *Great realizations*, Risk **13** (2000), 105–108.

[143] Tim Bollerslev, Tzuo Hann Law, George Tauchen, *Risk, jumps, and diversification*, Journal of Econometrics **144** (2008), 234–256.