Econophysics: Empirical facts and agent-based models

Anirban Chakraborti[a], Ioane Muni Toke[a], Marco Patriarca[b,c], Frédéric Abergel[a]

[a] *Chair of Quantitative Finance, Ecole Centrale Paris, Grande Voie des Vignes, 92290 Châtenay-Malabry, FRANCE*
[b] *National Institute of Chemical Physics and Biophysics, Rävala 10, 15042 Tallinn, ESTONIA*
[c] *Instituto de Fisica Interdisciplinaire y Sistemas Complejos (CSIC-UIB), E-07122 Palma de Mallorca, SPAIN.*

Abstract

This article aims at reviewing recent empirical and theoretical developments usually grouped under the term *Econophysics*. Since its name was coined in 1995 by merging the words "Economics" and "Physics", this new interdisciplinary field has grown in various directions: theoretical macroeconomics (wealth distributions), microstructure of financial markets (order book modeling), econometrics of financial bubbles and crashes. We give a brief introduction in the first part and begin with discussing interactions between Physics, Mathematics, Economics and Finance that led to the emergence of Econophysics in the second part. Then the third part is dedicated to empirical studies revealing statistical properties of financial time series. We begin the presentation with the widely acknowledged "stylized facts" describing the distribution of the returns of financial assets: fat-tails, volatility clustering, etc. Then we show that some of these properties are directly linked to the way "time" is taken into account, and present some new remarks on this account. We continue with the statistical properties observed on order books in financial markets. For the sake of illustrations in this review, (nearly) all the stated facts are reproduced using our own high-frequency financial database. Contributions to the study of correlations of assets such as random matrix theory and graph theory are finally presented in this part. The fourth part of our review deals with models in Econophysics through the point of view of agent-based modeling. Using previous work originally presented in the fields of behavioural finance and market microstructure theory, econophysicists have developed agent-based models of order-driven markets that are extensively reviewed here. We then turn to models of wealth distribution where an agent-based approach also prevails: kinetic theory models, and continue with game theory models and review the now classic minority games. We end this review by providing an outlook on possible directions of research.

Keywords: Econophysics; Stylized facts; Financial returns; Correlations; Order book; Agent-based models; Wealth distributions; Game Theory; Minority Games; Pareto Law.

PACS Nos.: 05.45.Tp, 02.50.Sk, 05.40.-a, 05.45.Ra, 89.75.Fb

Contents

Email addresses: `anirban.chakraborti@ecp.fr` (Anirban Chakraborti), `ioane.muni-toke@ecp.fr` (Ioane Muni Toke), `marco.patriarca@kbfi.ee` (Marco Patriarca), `frederic.abergel@ecp.fr` (Frédéric Abergel)

Part I
Introduction

What is Econophysics? Many years after the word “Econophysics” was coined by H. E. Stanley by a peculiar merging of the words ‘Economics’ and ‘Physics’ at an international conference on Statistical Physics held in Kolkata in 1995, this is still a commonly asked question. Many still wonder how theories aimed at explaining the physical world in terms of particles could be applied to understand complex structures, such as those found in the social and economic behavior of human beings. In fact physics is supposed to be precise or specific and to base its prediction power on the use of a few but universal properties of matter which are sufficient to explain many physical phenomena. But are there analogous precise universal properties known for human beings, who, on the contrary of fundamental particles, are certainly not identical in any respect to each other? And what little amount of information is sufficient to infer their complex behavior? Even if there is no evidence yet, there exists a positive attitude towards these questions. During the past decade physicists have made many successful attempts to approach problems in various fields of social science (1; 2; 3). In particular, Economics and its various subfields have been the target of various investigations — based on methods imported from or also used in physics — which are the subject of the present paper.

Economics deals with how societies efficiently use their resources to produce valuable commodities and distribute them among different people or economic agents (4; 5). It is a discipline related to almost everything around us, starting from the marketplace through the environment to the fate of nations. At first sight this may seem a complementary situation, with respect to that of physics, whose birth as a well defined scientific theory is usually associated with the study of particular mechanical objects moving with negligible friction, such as falling bodies and planets. However, a deeper comparison shows many more analogies than differences. On a general level, both economics and physics deal with “everything around us”, despite with different perspectives. On a practical level, the goals of both disciplines can be either purely theoretical in nature or strongly oriented toward the improvement of the quality of life. On a more technical side, analogies often become equivalences. As a representative example, it is worth mentioning that the frictionless systems which mark the early history of physics were recognized soon to be rare cases: not only at microscopic scale, where they obviously represent an exception due to the unavoidable interactions with the environment, but also at the macroscopic scale, where fluctuations of internal or external origin make a prediction of their time evolution impossible. Thus equilibrium and nonequilibrium statistical mechanics, the theory of stochastic processes, and the theory of chaos, became main tools for studying real systems as well as an important part of the theoretical framework of modern physics. But this was no news for Economics, in which the theory of the random walk had been already introduced by Bachelier some years before the works on Brownian motion by Einstein, Smolukowski, and Langevin. Mathematics, and in particular the theory of stochastic processes, certainly represents the main link, common language, and work framework, for both Economics and Physics. It is hard to find other applied disciplines besides physics — apart from probably chemistry — where the mathematical language of stochastic processes has a central role.

This review aims at reviewing recent empirical and theoretical developments that use tools from Physics in the fields of Economics and Finance. The review is organized as follows. A short historical view on the relationships between Mathematics, Physics, Economics and Finance, with an emphasis on the intermediary role of statistical mechanics between various disciplines is given in part II. Econophysics research is then reviewed in the following two parts. In part III, empirical studies revealing statistical properties of financial time series are reviewed. We begin the presentation with the widely acknowledged “stylized facts” describing the distribution of the returns of financial assets and continue with the statistical properties observed on order books in financial markets. For the sake of illustrating this review, (nearly) all the stated facts are reproduced using our own high-frequency financial database. In part IV, Econophysics models are reviewed through the point of view of agent-based modeling. Using previous work originally presented in the fields of behavioural finance and market microstructure theory, econophysicists have developed agent-based models of order-driven markets that are extensively reviewed here. We then turn to models of wealth distribution where an agent-based approach also prevails: kinetic theory models are presented. We continue with game theory models and review the now classic minority games. A final discussion, including an outline emerging from these interesting recent interactions between economics and econophysics, is given in part V.

Part II

Economics, Markets, Mathematics, Physics

1. Classic modeling in Finance: Probability and Stochastic Calculus

Mathematical finance has benefited a lot in the past thirty years from modern probability theory: Brownian motion, martingale theory are mandatory courses for any quantitative finance student. Financial mathematicians are often proud to recall the most well-known source of the interactions between the two fields: five years before Einstein's seminal work, the theory of the Brownian motion was formulated first by the French mathematician Bachelier in his doctoral thesis (6; 7; 8) where he used this model to describe price fluctuations at the Paris Bourse. Bachelier had even given a course as a "free professor" at the Sorbonne University with the title: "Probability calculus with applications to financial operations and analogies with certain questions from physics" (see the historical articles in Refs. (9; 10; 11)).

Then Itô, motivated by the works of Bachelier, formulated the presently known Itô calculus (12) and a variant called the geometric Brownian motion. Later, geometric Brownian motion became an important ingredient of models in Economics (13; 14), and in the well-known theory of option pricing (15; 16). In fact, stochastic calculus of diffusion processes combined with classical hypotheses in Economics led to the development of the *arbitrage pricing theory* ((17), (18)). The deregulation of financial markets at the end of the 1980's led to the exponential growth of the financial industry. Mathematical finance followed the trend: stochastic finance with diffusion processes and exponential growth of financial derivatives have had intertwined developments. Finally, this relationship was carved in stone when the Nobel prize was given to M.S. Scholes and R.C. Merton in 1997 (F. Black died in 1995) for their contribution to the theory of option pricing and their celebrated "Black-Scholes" formula.

However, this whole theory is closely linked to classical economics hypotheses and has not been grounded with empirical studies of financial time series. The Black-Scholes hypothesis of Gaussian log-returns of prices is in strong disagreement with empirical evidence. Mandelbrot (19; 20) was one of the firsts to observe a clear departure from Gaussian behaviour for these fluctuations. Within the framework of stochastic finance and martingale modeling, more complex processes have been considered (e.g. jumps processes (21), stochastic volatility (22; 23; 24)), but recent events on financial markets and the succession of financial crashes (e.g. (25) for a historical perspective) should lead scientists to re-think basic concepts of modelization. This is where Econophysics is expected to come to play (see e.g. (26; 27; 28) for reactions within the Econophysics community to the 2008 crash).

2. Statistical Mechanics bridging Physics and Economics

Statistical mechanics has been defined as that

> "branch of physics that combines the principles and procedures of statistics with the laws of both classical and quantum mechanics, particularly with respect to the field of thermodynamics. It aims to predict and explain the measurable properties of macroscopic (bulk) systems on the basis of the properties and behaviour of their microscopic constituents." (29)

The tools of statistical mechanics or statistical physics (30; 31; 32) that include extracting the average properties of a macroscopic system from the microscopic dynamics of the systems, are believed to prove useful for an economic system, because even though it is difficult or almost impossible to write down the "microscopic equations of motion" for an economic system with all the interacting entities, in general, economic systems can be investigated at various size scales. The understanding of the global behaviour of economic systems seems to need such concepts as stochastic dynamics, correlation effects, self-organization, self-similarity and scaling, and for their application we do not have to go into the detailed "microscopic" description of the economic system.

Besides the chaos theory (which had some impact in Economics modelling - e.g. Brock-Hommes-Chiarella), the theory of disordered systems has played a core role in Econophysics and study of "complex systems". The term "complex systems" was coined to cover the great variety of such systems which include examples from physics, chemistry, biology and also social sciences. The concepts and methods of statistical physics turned out to be extremely useful in application to these diverse complex systems including economic systems. Many complex systems in natural and social environments share the characteristics of competition among interacting agents for resources and their adaptation to dynamically changing environment (33; 34). The concept of disordered systems for instance allowed to go beyond the representative agent approach prevailing in much of (macro)-economics.

This deep interest in the nature of fluctuations already links the communities of physicists and economists. To make the connections even more evident we mention a few important contributions. The Italian social economist Pareto investigated a century ago, the wealth of individuals in a stable economy (35; 36) by modelling them with the distribution, $P(>x) \sim x^{-\alpha}$, where $P(>x)$ is the number of people having income greater than or equal to x and α is an exponent (known now as the Pareto exponent) which he estimated to be 1.5.

In the physics community it was Majorana who first took scientific interest in financial and economic systems. He wrote a pioneering paper on the essential analogy between statistical laws in physics and in social sciences (37; 38; 39). However, prior to the 1990's very few physicists like Kadanoff (40) and Montroll (41) had an explicit interest for research in social or economic systems. It was not until around 1990 that physicists started turning to this interdisciplinary subject and their research activity to begin to be complementary to the most traditional approaches of mathematical finance.

The chronological development of Econophysics has been well covered in the book of B. Roehner (42). Here it is worth mentioning a few landmarks. The first article on analysis of finance data which appeared in a physics journal was that of Mantegna (43) in 1991. The first conference in Econophysics was held in Budapest in 1997 and was followed by a series of meetings, such as the APFA — Application of Physics to Financial Analysis, and the WEHIA — Workshop on Economic Heterogeneous Interacting Agents, Econophys-Kolkata series of conferences should be mentioned. In the recent years the number of papers has increased dramatically, the community has grown rapidly and several new directions of research have opened. By now such renowned physics journals like Physical Review Letters, Physical Review E, Physica A, Europhysics Letters, European Physical Journal B, International Journal of Modern Physics C regularly publish papers in this area. The interested reader can also follow the developments quite well from the cond-mat preprint server and should visit the Web Sites of the *Econophysics Forum* (44) and *Econophysics.Org* (45).

Part III
Some statistical properties of financial data

Prices of commodities or assets produce what is called time series. Financial time-series analysis has been of great interest not only to the practitioners (an empirical discipline) but also to the theoreticians for making inferences and predictions. The inherent uncertainty in the financial time-series and its theory makes it specially interesting to economists, statisticians and physicists (46).

Different kinds of financial time series have been recorded and studied for decades, but the scale changed twenty years ago. The computerization of stock exchanges that took place all over the world in the mid 1980's and early 1990's has lead to the explosion of the amount of data recorded. Nowadays, all transactions on a financial market are recorded *tick-by-tick*, i.e. every event on a stock is recorded with a timestamp defined up to the millisecond, leading to huge amounts of data. For example, as of today (2009), the Reuters Datascope Tick History (RDTH) database records roughly 7 gigabytes of data *every trading day*.

Prior to this improvement in recording market activity, statistics could be computed with daily data at best. Now scientists can compute intraday statistics in high-frequency. This allows to check known properties at new time scales (see e.g. 1.2 below), but also implies special care in the treatement (see e.g. the computation of correlation on high-frequency 3.1 below).

It is a formidable task to make an exhaustive review on this topic but we try to give a flavour of some of the aspects in this section.

1. Time series of financial returns

1.1. "Stylized facts" of financial time series

The concept of "stylized facts" was introduced in macroeconomics around 1960 by Nicholas Kaldor, who advocated that a scientist studying a phenomenon "should be free to start off with a stylized view of the facts" (47). In his work, Kaldor isolated several statistical facts characterizing macroeconomic growth over long periods and in several countries, and took these robust patterns as a starting point for theoretical modeling.

This expression has thus been adopted to describe empirical facts that arose in statistical studies of financial time series and that seem to be persistent across various time periods, places, markets, assets, etc. One can find many different lists of these facts in several reviews (e.g. (48), (49; 50; 51)). We choose in this article to present a minimum set of facts now widely acknowledged, at least for the prices of equities.

1.1.1. Fat-tailed empirical distribution of returns

Let p_t be the price of a financial asset at time tt. We define its return over a period of time τ to be:

$$r_\tau(t) = \frac{p(t+\tau) - p(t)}{p(t)} \approx \log(p(t+\tau)) - \log(p(t)) \tag{1}$$

It has been largely observed – see e.g. (52) in 1963, (53) for tests on more recent data – and it is the first stylized fact, that the empirical distributions of financial returns and log-returns are fat-tailed. On figures 1 and 2 we plot these distribution for a liquid French stock (BNP Paribas) with $\tau = 5$ minutes, and for the index S& P500 with $\tau = 1$ day.

Many studies obtain same observations on different sets of data. For example, using two years of data on more than a thousand US stocks, (54) finds that returns asymptotically follow a power law $F(r_\tau) \sim |r|^{-\alpha}$ with $\alpha > 2$ ($\alpha \approx 2.8 - 3$).

With $\alpha > 2$, the second moment (the variance) is well-defined, excluding stable laws with infinite variance. There has been various suggestions for the form of the distribution: Student's-t, hyperbolic, normal inverse Gaussian, exponentially truncated stable, and others, but no general consensus exists on the exact form of the tails.

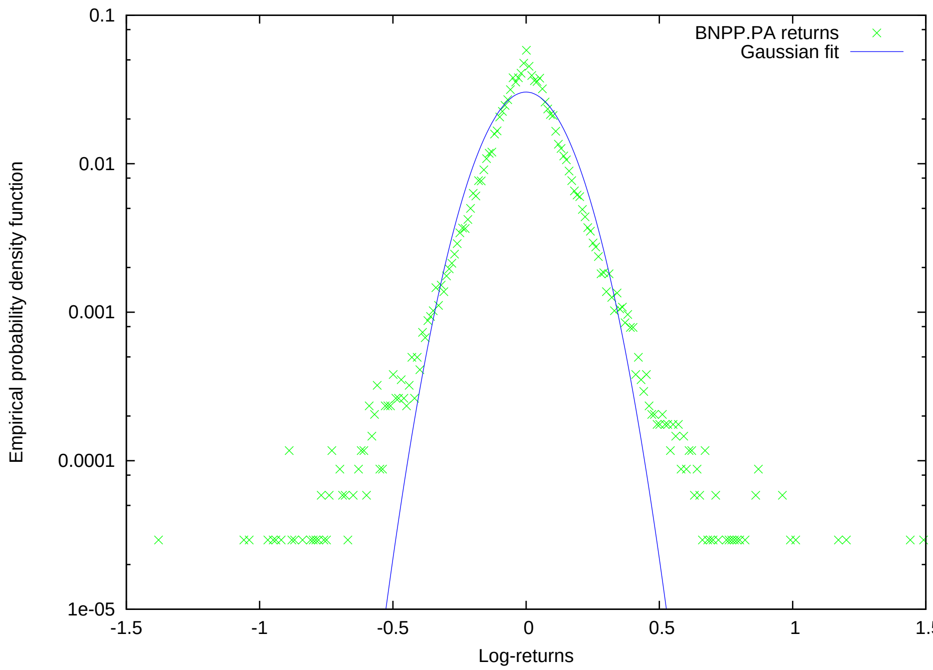

Figure 1: Empirical distribution of BNP Paribas log-returns over a period of time $\tau = 5$ minutes. This empirical probability density function is computed by sampling tick-by-tick data from 9:05am til 5:20pm between January 1st, 2007 and May 3Oth, 2008, i.e. 15664 values. Gaussian fit is provided for comparison.

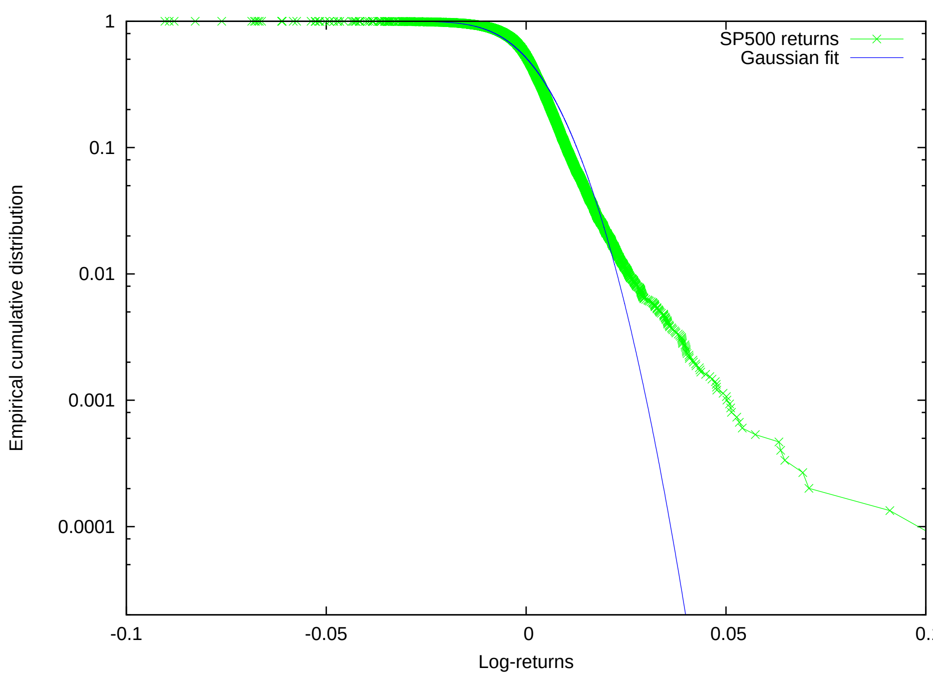

Figure 2: Empirical cumulative distribution of S& P 500 daily returns. This empirical distribution is computed using official daily close price between January 1st, 1950 and June 15th, 2009, i.e. 14956 values. Gaussian fit is provided for comparison.

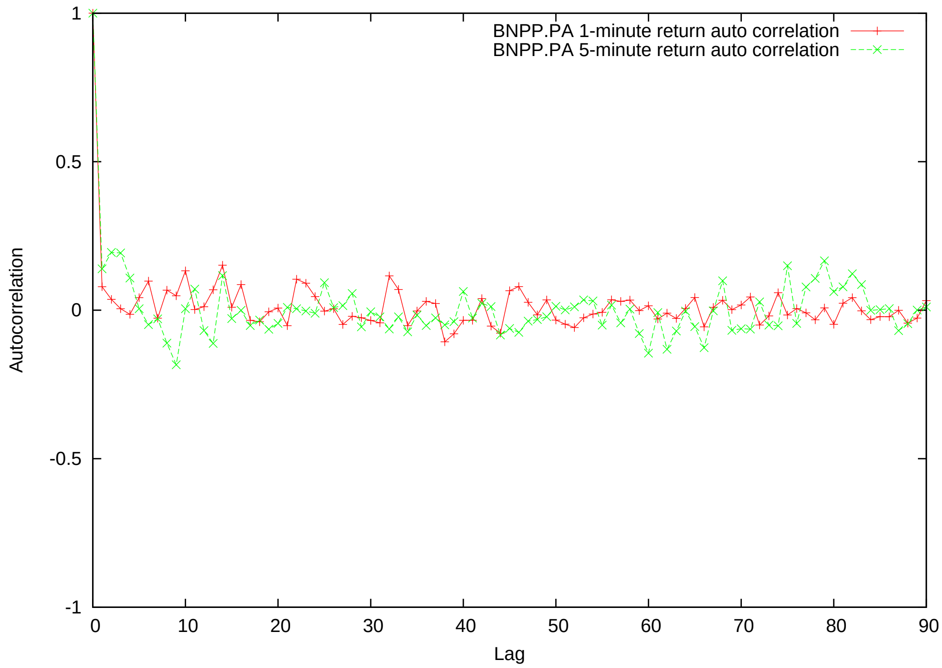

Figure 3: Autocorrelation function of BNPP.PA returns. This empirical function is computed by sampling tick-by-tick data from 9:05am til 5:20pm between January 1st, 2007 and May 30th, 2008.

1.1.2. Absence of auto-correlations of returns

On figure 3 we plot the auto-correlation of log-returns defined as $\rho(T) \sim \langle r_\tau(t+T) r_\tau(t) \rangle$ with τ =1 minute and 5 minutes. We observe here, as it is widely known (see e.g. (49; 55)), that there is no evidence of correlation between successive returns. The autocorrelation function decays very rapidly to zero, even for a few lags of 1 minute.

1.1.3. Volatility clustering

The third "stylized-fact" that we present here is of primary importance. Absence of correlation between returns must no be mistaken for a property of independance and identical distribution: price fluctuations are not identically distributed and the properties of the distribution change with time.

In particular, "stylized fact" number three states that absolute returns or squared returns exhibit a long-range slowly decaying auto correlation function. This phenomena is widely known as "volatility clustering", and was formulated by Mandelbrot in 1963 (52) as "large changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes".

On figure 4, the autocorrelation function of absolute returns is plotted for $\tau = 1$ minute and 5 minutes. The levels of aurocorrelations at the first lags vary wildly with the parameter τ. On our data, it is found to be maximum (more than 70% at the first lag) for a returns sampled every five minutes. However, whatever the sampling frequency, autocorrelation is still above 10% after several hours of trading. On this data, we can grossly fit a power law decay with exponent 0.4. Other empirical tests report exponents between 0.1 and 0.3 ((55; 56; 57)).

1.1.4. Aggregational normality

It has been observed that as one increases the time scale over which the returns are calculated, the fat-tail property becomes less pronounced, and their distribution approaches the Gaussian form.

On figure 5, we plot these standardized distributions for S& P 500 index between January 1st, 1950 and June 15th, 2009. It is clear that the larger the time scale increases, the more Gaussian the distribution. The

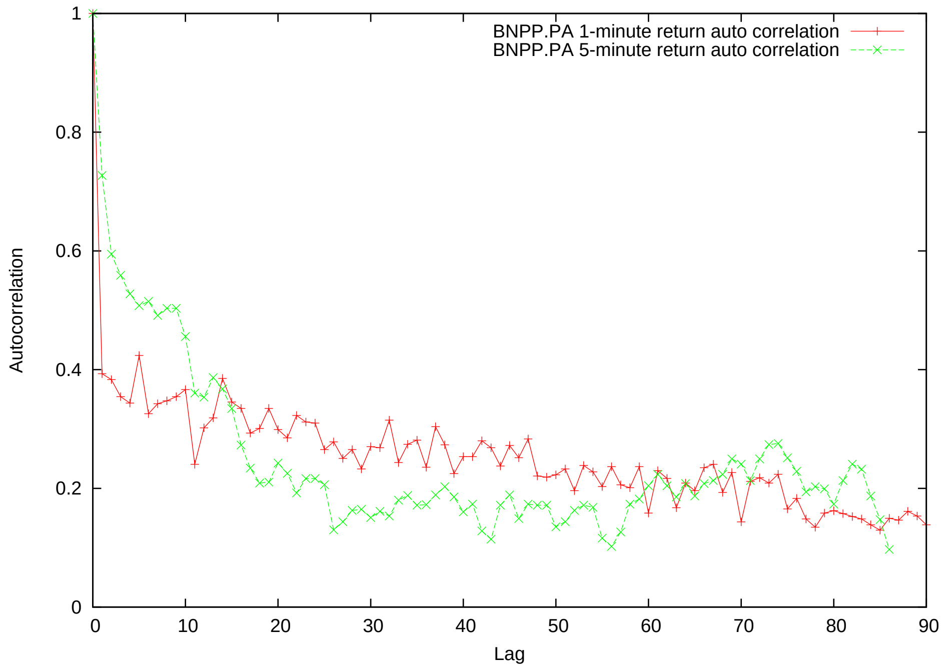

Figure 4: Autocorrelation function of BNPP.PA absolute returns. This empirical function is computed by sampling tick-by-tick data from 9:05am til 5:20pm between January 1st, 2007 and May 3Oth, 2008.

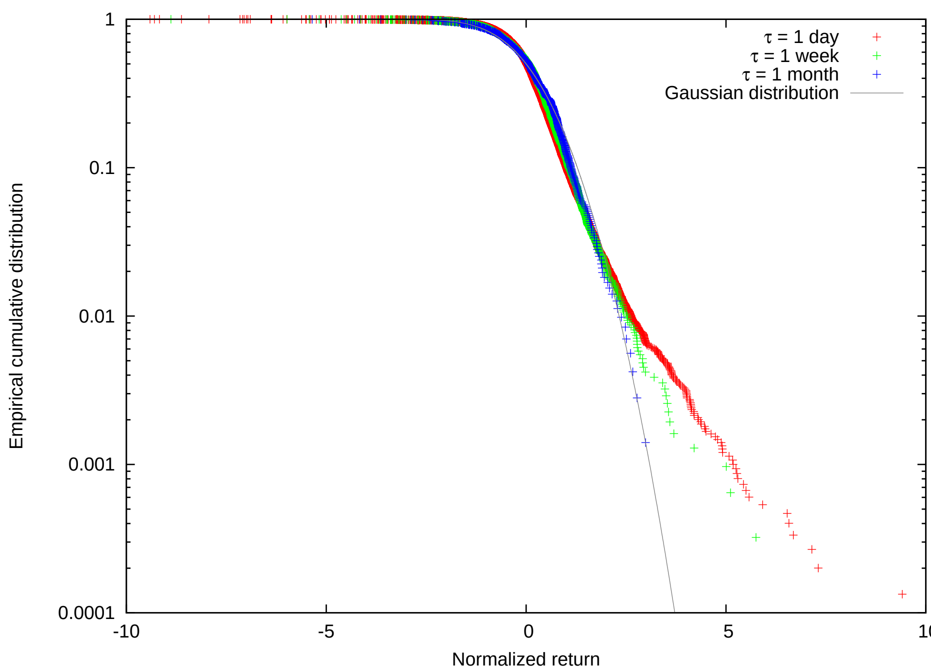

Figure 5: Distribution of log-returns of S& P 500 daily, weekly and monthly returns.

fact that the shape of the distribution changes with τ makes it clear that the random process underlying prices must have non-trivial temporal structure.

1.2. Getting the right "time"

1.2.1. Four ways to measure "time"

In the previous section, all "stylized facts" have been presented in *calendar time*, i.e. time series were indexed, as we expect them to be, in hours, minutes, seconds, milliseconds. Let us recall here that tick-by-tick data available on financial markets all over the world is timestamped up to the millisecond, but that the order of magnitude of the guaranteed precision is much larger, usually one second.

Calendar time is the time usually used to compute statistical properties of financial time series. This means that computing these statistics involves sampling, which might be a delicate thing to do when dealing for example with several stocks with different liquidity. Therefore, three other ways to keep track of time may be used.

Let us first introduce *event time*. Using this count, time is increased by one unit each time one order is submitted to the observed market. This framework is natural when dealing with the simulation of financial markets, as it will be showed in section 1. The main characteristic of event time lies in its "smoothing" of data. In event time, intraday seasonality (lunch break) or outburst of activity consequent to some news are smoothened in the time series, since we always have one event per time unit.

Now, when dealing with time series of prices, another count of time might be relevant, and we call it *trade time* or *transaction time*. Using this count, time is increased by one unit each time a transaction happens. The advantage of this count is that limit orders submitted far away in the order book, and may thus be of lesser importance with respect to the price series, do not increase the clock by one unit.

Finally, going on with focusing on important events to increase the clock, we can use *tick time*. Using this count, time is increased by one unit each time the price changes. Thus consecutive market orders that progressively "eat" liquidity until the first best limit is removed in an order book are counted as one unit time.

Let us finish by noting that with these definition, when dealing with mid prices, or bid and ask prices, a time series in event time can easily be extracted from a time series in calendar time. Furthermore, one can always extract a time series in trade time or in price time from a time series in event time. However, one cannot extract a series in price time from a series in trade time, as the latter ignores limit orders that are submitted inside the spread, and thus change mid, bid or ask prices without any transaction taking place.

1.2.2. Revisiting "stylized facts" with a new clock

Now, using the right clock might be of primary importance when dealing with statistical properties and estimators. For example, (58) investigates the standard realized variance estimator (see section 3.1) in trade time and tick time. In this section we compute some statistics complementary to the ones we've presented in the previous section 1.1 and show the role of the clock in the studied properties.

Aggregational normality in trade time. We have seen above that when the sampling size increases, the distribution of the log-returns tends to be more gaussian. This property is much better seen using trade time. On figure 6, we plot the distributions of the log-returns for BNP Paribas stock using 2-month-long data in calendar time and trade time. Over this period, the average number of trade per day is 8562, so that 17 trades (resp.1049 trades) corresponds to an average calendar time step of 1 minute (resp. 1 hour). We observe that the distribution of returns sampled every 1049 trades is much more gaussian than the one sampled every 17 trades (aggregational gaussianity), and that it is also more gaussian that the one sampled every 1 hour (quicker convergence in trade time).

Note that this gaussianity in trade time is observed only when shorter time-series are studied. When dealing with much longer time series, this is no longer the case as observed on figure 7. We compute the same distribution of returns in trade time for a 17-month long series et and a 2-month long series. It appears that the distribution computed on the longer time series has much fatter tails than the other one. Our understanding of this phenomena is that much longer time series include much rarer events, and periods

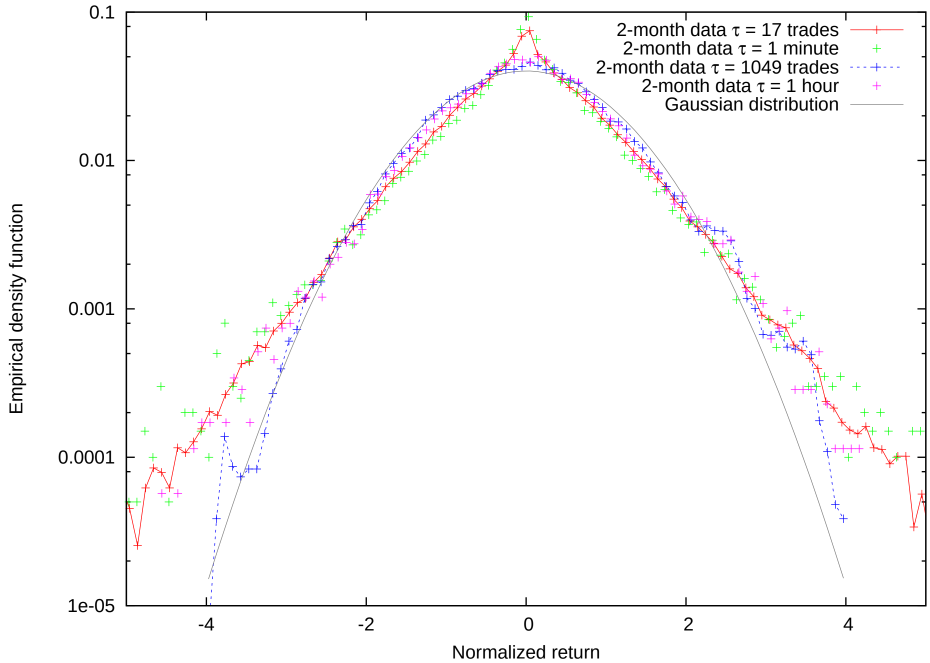

Figure 6: Distribution of log-returns of stock BNPP.PA. This empirical distribution is computed using data from 2007, April 1st until 2008, May 31st.

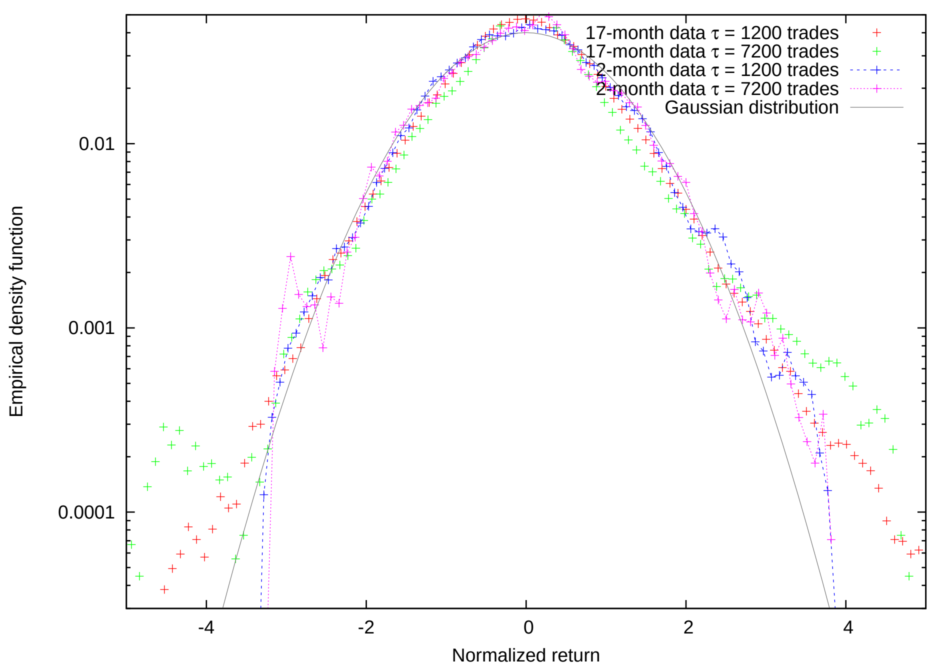

Figure 7: Distribution of log-returns of stock BNPP.PA. This empirical distribution is computed using data from 2007, April 1st until 2008, May 31st.

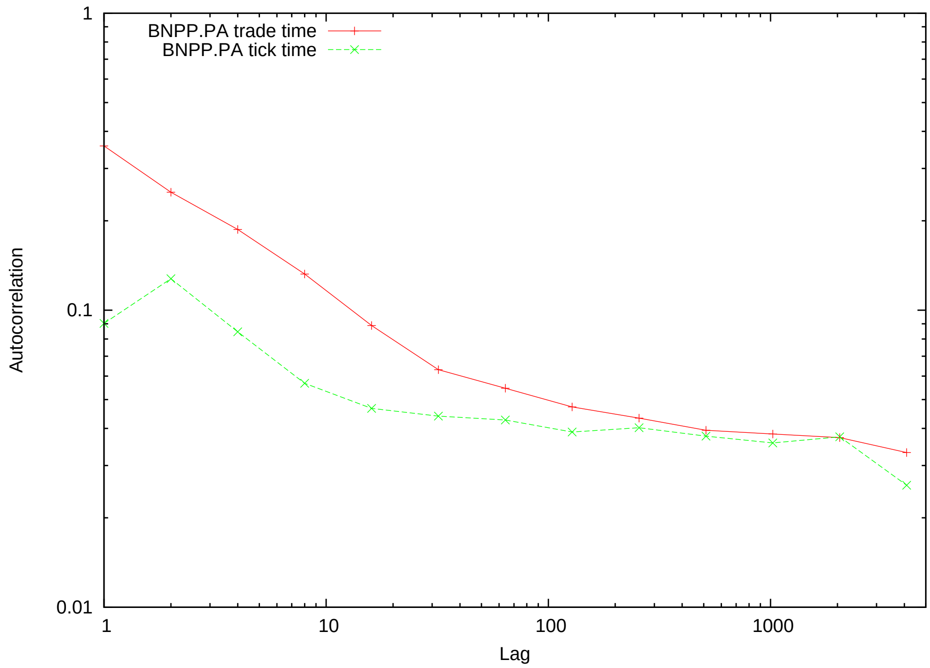

Figure 8: Auto-correlation of trade signs for stock BNPP.PA. This empirical distribution is computed using data from 2007, January 1st until 2008, May 31st.

with very different levels of volatility. When dealing with smaller series, we have a more homogeneous sample, hence the gaussianity.

Note that this property appears to be valid in a multidimensional setting, see (59).

Auto-correlation of trade signs in tick time. It is well-known that the series of the signs of the trades on a given stock (usual convention: +1 for a transaction at the ask price, −1 for a transaction at the bid price) exhibit large autocorrelation. It has been observed in (60) for example that the autocorrelation function of the signs of trades (ϵ_n) was a slowly decaying function in $n^{-\alpha}$ with α roughly 0.5. We compute this statistics for the trades on BNP Paribas stock from 2007, January 1st until 2008, May 31st. We plot the result in figure 8. We find that the first values for short lags are about 0.3, and that he log-log plot clearly shows some power-law decay with roughly $\alpha = 0.7$.

A very plausible explanation of this phenomenon relies on the execution strategies of some major brokers on a given markets. These brokers have large transaction to execute on the account of some clients. In order to avoid to make the market move because of an inconsiderately large order (see below section 2.6 on market impact), they tend to split large orders into small ones. We think that these strategies explains, at least partly, the large auto-correlation observed. Using data on markets where orders are publicly identified and linked to a given broker, it can be shown that the auto-correlation function of the order signs *of a given broker* is even higher. See (61) for a review of these facts and some associated theories.

We propose here another hint supporting this explanation. We compute the auto-correlation function of order signs *in tick time*, i.e. taking only into account transactions that make the price change. Results is plotted on figure 8. We find that the first values for short lags are about 0.10, which is much smaller than the values observed on the previous graph. This support the idea that many small transactions progressively “eat” the available liquidity at the best quotes. Note however that even in tick time the correlation remains positive even for large lags.

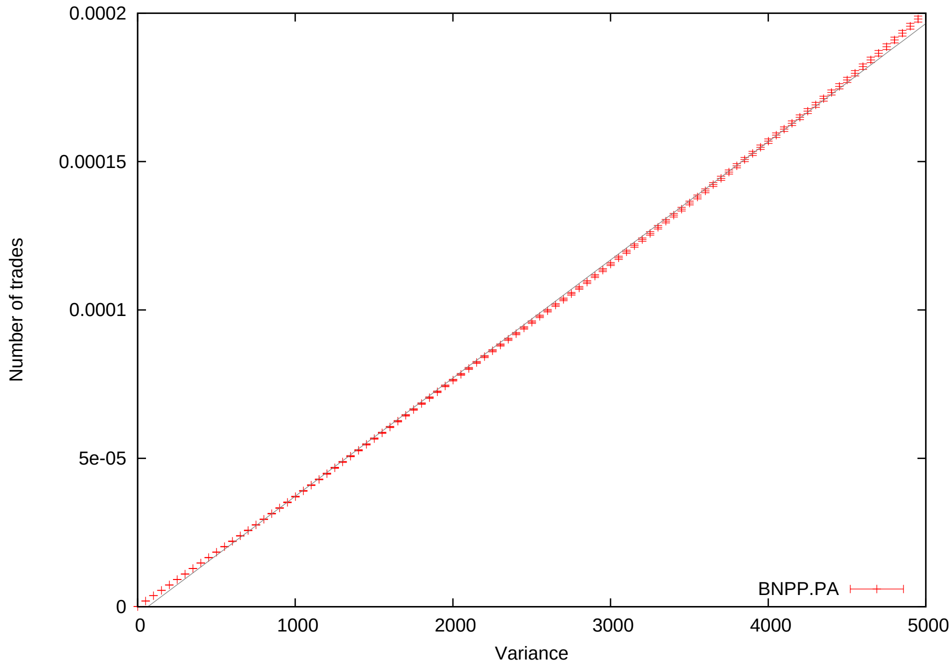

Figure 9: Second moment of the distribution of returns over N trades for the stock BNPP.PA. This empirical function is computed by sampling tick-by-tick data from 9:05am til 5:20pm between January 1st, 2007 and May 31st, 2008, i.e. 356 treated days of trading.

1.2.3. Correlation between volume and volatility

Investigating time series of cotton prices, Clark (22) notes that "trading volume and price change variance seem to have a curvilinear relationship". *Trade time* allows us to have a better view on this property: Plerou et al. (62), Silva et al. (63) among others, shows that the variance of log-returns after N trades, i.e. over a time period of N in trade time is proprtional to N.

We confirm this observation by plotting the second moment of the distribution of log-returns after N trades as a function of N for our data, as well as the average number of trades and the average volatility on a given time interval. The results are shown on figure 9 and 10.

This results are to be put in relationship with the one presented in Gopikrishnan et al. (64), where the statistical properties of the number of shares traded $Q_{\Delta t}$ for a given stock in a fixed time interval Δt is studied. They analyzed transaction data for the largest 1000 stocks for the two-year period 1994-95, using a database that recorded every transaction for all securities in three major US stock markets. They found that the distribution $P(Q_{\Delta t})$ displayed a power-law decay as shown in Fig. 11, and that the time correlations in $Q_{\Delta t}$ displayed long-range persistence. Further, they investigated the relation between $Q_{\Delta t}$ and the number of transactions $N_{\Delta t}$ in a time interval Δt, and found that the long-range correlations in $Q_{\Delta t}$ were largely due to those of $N_{\Delta t}$. Their results are consistent with the interpretation that the large equal-time correlation previously found between $Q_{\Delta t}$ and the absolute value of price change $|G_{\Delta t}|$ (related to volatility) were largely due to $N_{\Delta t}$.

Therefore, studying variance of price changer in *trade time* suggests that the number of trade might be a good proxy for the unobserved volatility.

1.2.4. A link with stochastic processes: subordination

These empirical facts (aggregational normality in trade time, relationship between volume and volatility) reinforce the interest for models based on the subordination of stochastic processes, which has been introduced in financial modeling by Clark (22).

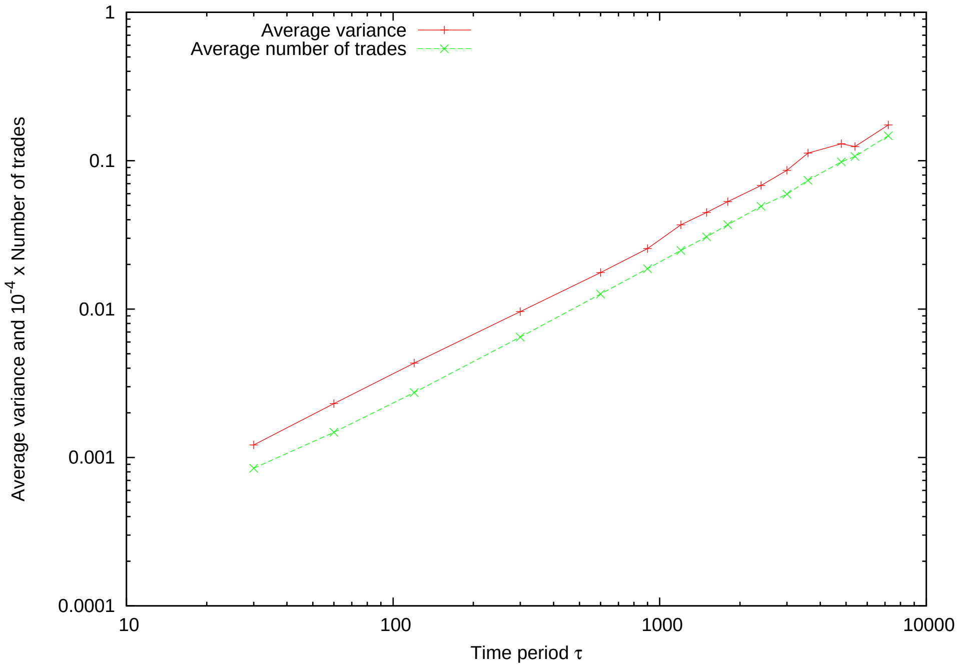

Figure 10: Average number of trades and average volatility on a time period τ for the stock BNPP.PA. This empirical function is computed by sampling tick-by-tick data from 9:05am til 5:20pm between January 1st, 2007 and May 31st, 2008, i.e. 356 treated days of trading.

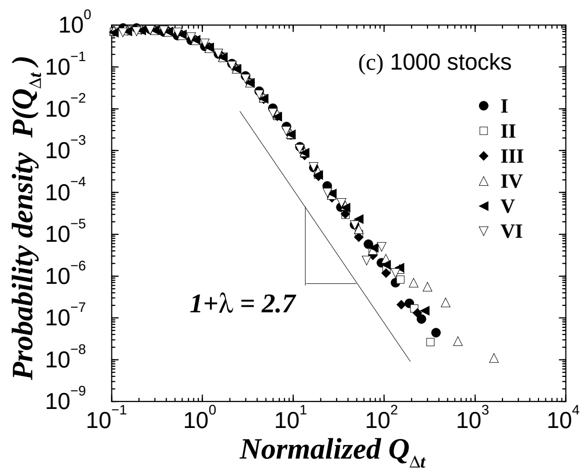

Figure 11: Distribution of the number of shares traded $Q_{\Delta t}$. Adapted from `arXiv:cond-mat/0008113`.

Let us introduce it simply here. Assuming the proportionality between the variance $\langle x \rangle_\tau^2$ of the returns x and the number of trades N_τ over a time period τ, we can write:

$$\langle x \rangle_\tau^2 = \alpha N_\tau. \tag{2}$$

Therefore, assuming the normality in trade time, we can write the density function of log-returns after N trades as

$$f_N(x) = \frac{e^{\frac{-x^2}{2\alpha N}}}{\sqrt{2\pi\alpha N}}, \tag{3}$$

Finally, denoting $K_\tau(N)$ the probability density function of having N trades in a time period τ, the distribution of log returns in calendar time can be written

$$P_\tau(x) = \int_0^\infty \frac{e^{\frac{-x^2}{2\alpha N}}}{\sqrt{2\pi\alpha N}} K_\tau(N) dN. \tag{4}$$

This is the subordination of the Gaussian process x_N using the number of trades N_τ as the *directing process*, i.e. as the new clock. With that kind of modelization, it is expected, since P_N is gaussian, that observed non-gaussian behavior will comme from $K_\tau(N)$. For example, some specific choice of directing processes may lead to a symetric stable distribution (see (65), to be checked). (22) tests empirically a lognormal subordination with time series of prices of cotton. In a similar way, (63) tests the hypothesis of an exponential subordination:

$$K_\tau(N) = \frac{1}{\eta\tau} e^{-\frac{N}{\eta\tau}}. \tag{5}$$

If the orders were submitted to the market in a independant way and at a constant rate η, then the distribution of the number of trade per time period τ should be a Poisson process with intensity $\eta\tau$. Therefore, the empirical fit in equation 5 is inconsistent with such a simplistic hypothesis of distribution of time of arrivals of orders. We will suggest in the next section some possible distribution that fit our empirical data.

2. Statistics of order books

The computerization of financial markets in the second half of the 1980's provided the empirical scientists with easier access to extensive data on order books. (66) is an early study of the new data flows on the newly (at that time) computerized Paris Bourse. Variables crucial to a fine modeling of order flows and dynamics of order books are studied: time of arrival of orders, placement of orders, size of orders, shape of order book, etc. Many subsequent papers offers complementary empirical findings, e.g. (67), (68), (69), (70), (71). Before going further in our review of available models, we try to summarize some of these empirical facts.

For each of the enumerated properties, we propose our own new empirical plots. We use Reuters tick by tick data on the Paris Bourse. We select four stocks: France Telecom (FTE.PA) , BNP Paribas (BNPP.PA), Societe Générale (SOGN.PA) and Renault (RENA.PA). For any given stocks, the data displays timestamps, traded quantities, traded prices, the first five best bid limits and the first five best ask limits. from now on, we will denote $a_i(t)$ (resp. $(b_j(t))$ the price of the i-th limit at ask (resp. j-th limit at bid). Except when mentioned otherwise, all statistics are computed using all trading days from Oct, 1st 2007 to May, 30th 2008, i.e. 168 trading days. On a given day, orders submitted between 9:05am and 5:20pm are taken into account, i.e. first and last minutes of each trading days are removed.

Note that we do not deal in this section with the correlations of the signs of trades, since statistical results on this fact have already been treated in section 1.2.2.

Note also that although most of these facts are wildly acknowledged, we will not describe them as new "stylized facts for order books" since their range of validity is still to be checked among various products/stocks, markets and epochs, and strong properties have to be properly extracted and formalized from these observations. However, we'll keep them in mind as we go through the new trend of "empirical modeling" of order books.

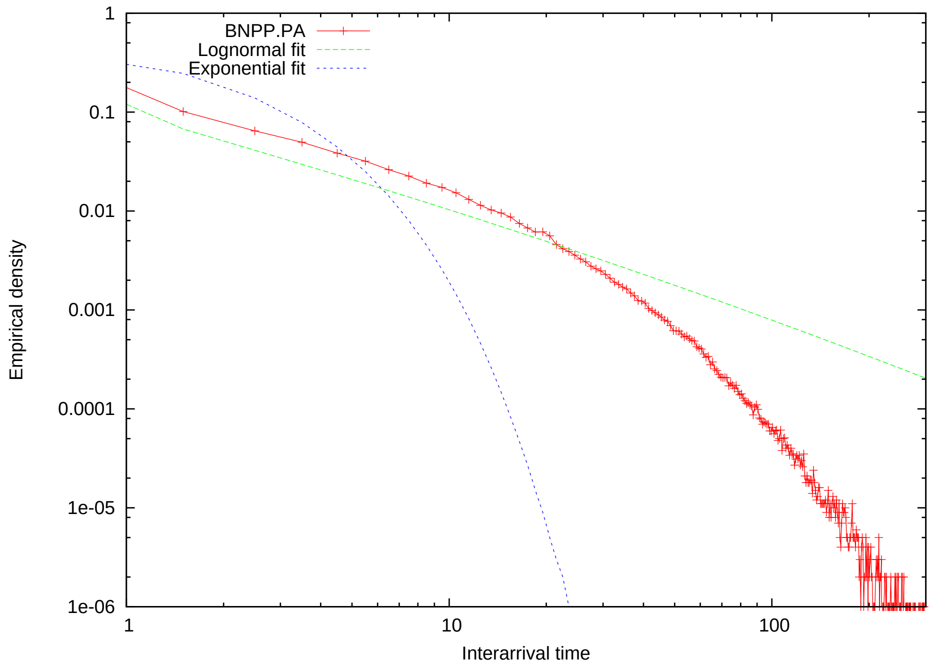

Figure 12: Distribution of interarrival times for stock BNPP.PA. This empirical distribution is computed using data from 2007, January 1st until 2008, May 31st.

Finally, let us recall that the markets we are dealing with are electronic order books with no official market maker and where orders are submitted in a double auction and executions follow price/time priority. This type of exchange is now adopted nearly all over the world, but this was not obvious as long as computerization was not complete. Different market mechanisms have been widely studied in the microstructure literature: see e.g. (72; 73; 74; 75; 76; 77). We will not review this literature here (except (72) reviewed in part IV, as this would be too large a digression, however such a literature is linked in many aspects to the problems reviewed in this paper.

2.1. Time of arrivals of orders

As explained in the previous section, the choice of the time count might be of prime importance when dealing with "stylized facts" of empirical financial time series. When reviewing the subordination of stochastic processes (22; 63), we have seen that the Poisson hypothesis for the arrival times of orders is not empirically verified.

We compute the empirical distribution for interarrival times of market orders on the stock BNP Paribas using data from 2007, October 1st until 2008, May 31st. The result is plotted in figure 12 and 13, both in linear and log scale. It is clearly observed that the exponential fit is not a good one. However, it seems that a the fit of a gamma distribution is potentially a very good one.

Using the same data, we compute the empirical distribution of the number of transactions in a given time period τ. Results are plotted in figure 14. It seems that the lognormal and the gamma distributions are both good candidates, however none of them really describes the empirical result, suggesting a complex structure of arrival of orders. A similar result on Russian stocks was presented in (78).

2.2. Volume of orders

Empirical studies show that the unconditional distribution of order size is very complex to characterize. (67) and (69) observe a power law decay with an exponent $1 + \mu =\approx 2.3 - 2.7$ for market orders and

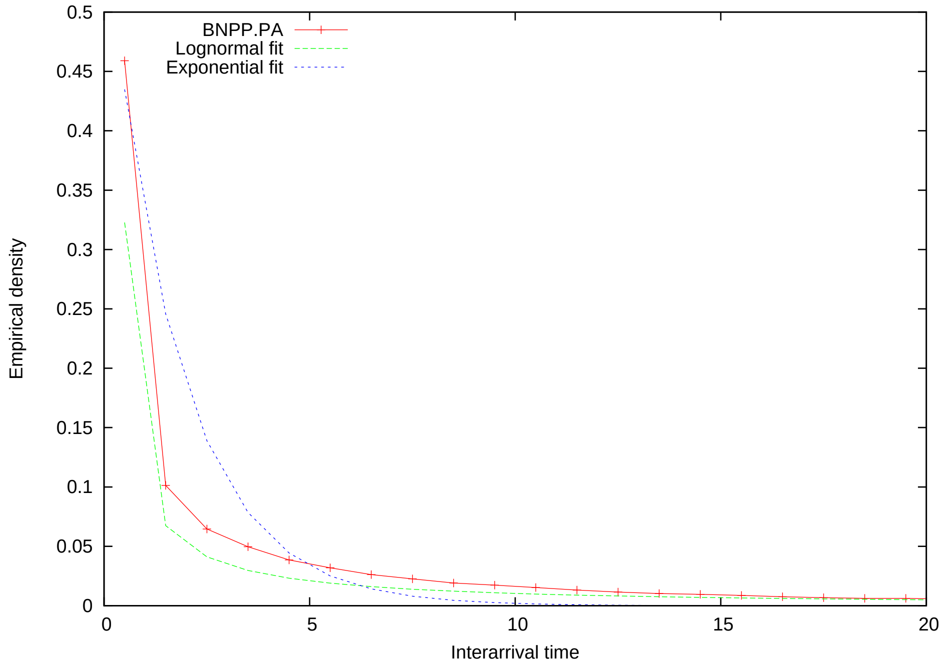

Figure 13: Distribution of interarrival times for stock BNPP.PA (Main body, linear scale). This empirical distribution is computed using data from 2007, January 1st until 2008, May 31st.

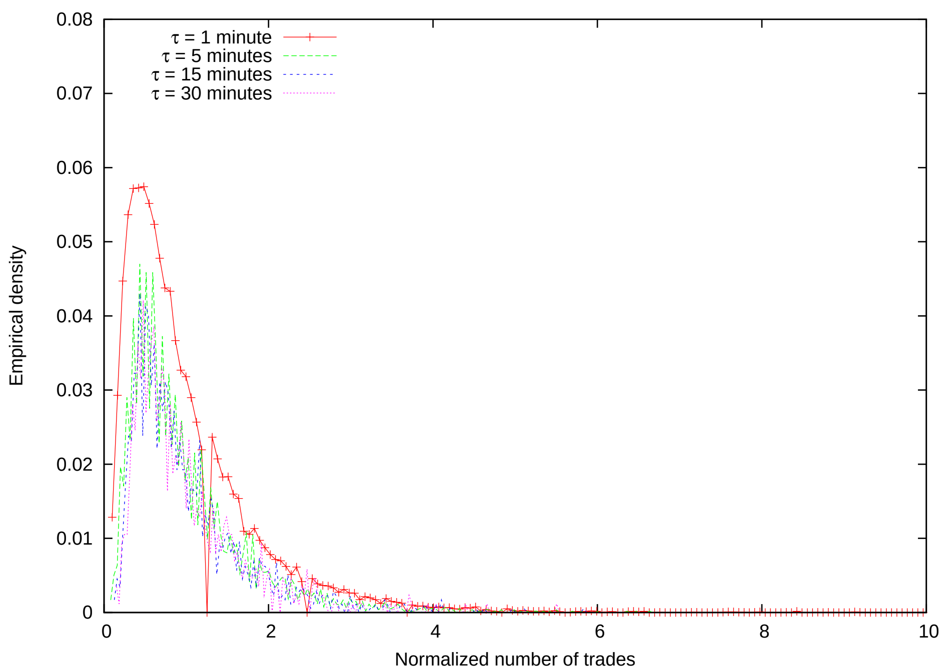

Figure 14: Distribution of the number of trades in a given time period τ for stock BNPP.PA. This empirical distribution is computed using data from 2007, October 1st until 2008, May 31st.

$1+\mu \approx 2.0$ for limit orders. (68) emphasize on a clustering property: orders tend to have a "round" size in packages of shares, and clusters are observed around 100's and 1000's. As of today, no consensus emerges in proposed models, and it is plausible that such a distribution varies very wildly with products and markets.

In figure 15, we plot the distribution of volume of market orders for the four stocks composing our benchmark. Quantities are normalized by their mean. Power-law coefficient is estimated by a Hill estimator (see e.g. (79; 80)). We find a power law with exponent $1+\mu = 2.7$ which confirms studies previously cited. Figure 16 displays the same distribution for limit orders (of all available limits). We find an average value of $1+\mu =\approx 2.1$, consistent with previous studies. However, we note that the power law is a much poorer fit in the case of limit orders: data normalized by their mean collapse badly on a single curve, and computed coefficients vary with stocks.

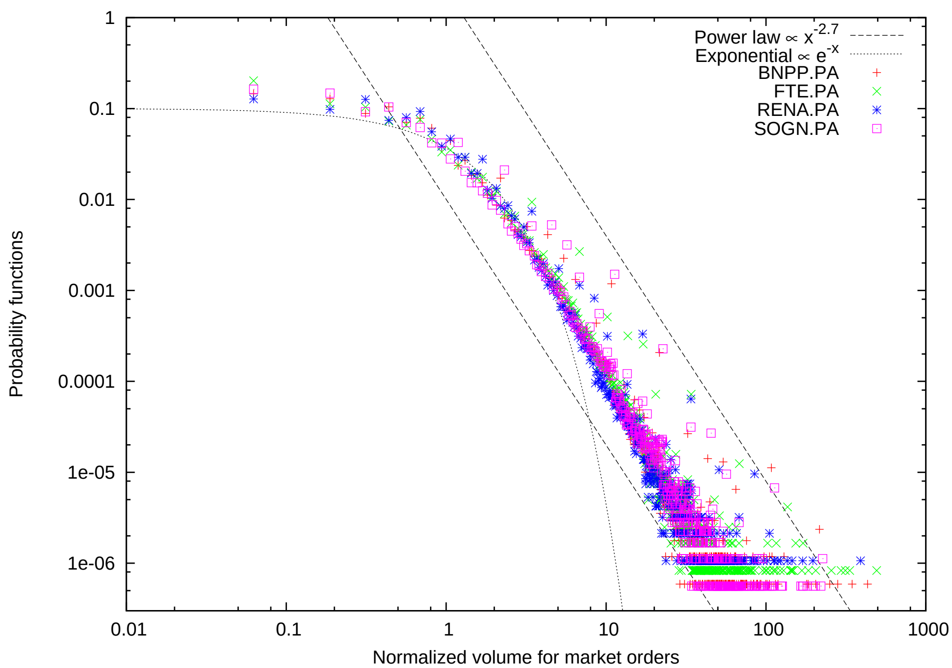

Figure 15: Distribution of volumes of market orders. Quantities are normalized by their mean.

2.3. Placement of orders

Placement of arriving limit orders. (70) observe a broad power-law placement arount the best quotes on French stocks, confirmed in (71) on US stocks. Observed exponents are quite stable across stocks, but place dependant: $1+\mu \approx 1.6$ on the Paris Bourse, $1+\mu \approx 2.0$ on the New York Stock Exchange, $1+\mu \approx 2.5$ on the London Stock Exchange. (81) propose to fit the empirical distribution with a Student distribution with 1.3 degree of freedom.

We plot the distribution of the following quantity computed on our data set, i.e. using only the first five limits of the order book:

$$\Delta p = \begin{cases} b(t) - b_0(t-) & \text{if an bid order arrives at price } b(t), \\ & b_0(t) \text{ being the best bid before the arrival of this order} \\ a_0(t-) - a(t) & \text{if an ask order arrives at price } a(t), \\ & a_0(t) \text{ being the best ask before the arrival of this order} \end{cases} \quad (6)$$

We use data on the BNP.Paribas order book from 2007, September 1st until 2008, May 31st. Results are plotted on figure 17 in semilog scale and 18 in linear scale. This graphs being computed with incomplete

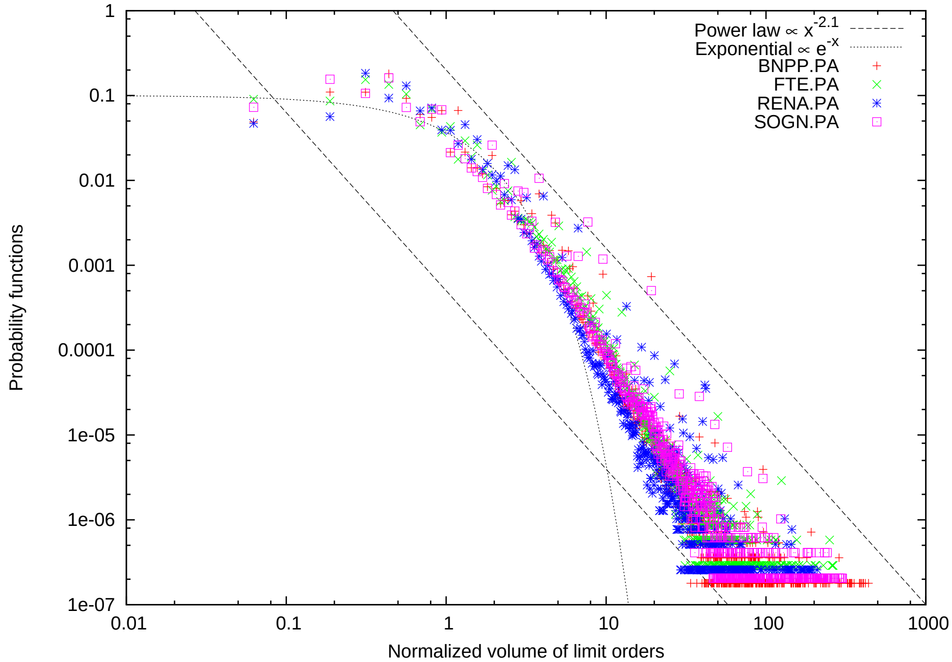

Figure 16: Distribution of normalized volumes of limit orders. Quantities are normalized by their mean.

data (five best limits), we do not observe a placement as broad as in (70). However, our data makes it clear that fat tails are observed. We also observe a slight asymetry in the empirical distribution: the left side is less broad than the right side. Since the left side represent limit orders submitted *inside* the spread, this was expected.

Thus, the empirical distribution of the placement of arriving limit orders is maximum at zero (same best quote). How is it translated in terms of shape of the order book ?

Average shape of the order book. Contrary to what one might expect, it seems that the maximum of the average offered volume in an order book is located away from the best quotes (see e.g. (70)). Our data confirms this observation: the average quantity offered on the five best quotes grows with the level. This result is presented in figure 19. We also compute the average price of these levels in order to plot a cross-sectional graph similar to the ones presented in (66). Our result is presented for stock BNP.PA in figure 20 and displays the expected shape. Results for other stocks are similar. We find that the average gap between to levels is constant among the five best bids and asks (less than one tick for FTE.PA, 1.5 tick for BNPP.PA, 2.0 ticks for SOGN.PA, 2.5 ticks for RENA.PA). We also find that the average spread is roughly twice as large the average gap (factor 1.5 for FTE.PA, 2 for BNPP.PA, 2.2 for SOGN.PA, 2.4 for RENA.PA).

2.4. Cancelation of orders

(68) show that the distribution of the average lifetime of limit orders fits a power law with exponent $1+\mu \approx 2.1$ for cancelled limit orders, and $1+\mu \approx 1.5$ for executed limit orders. (81) finds that in any case the exponential hypothesis (Poisson process) is not satisfied on the market.

We compute the average lifetime of cancelled and exectued orders on our dataset. Since our data does not include a unique identifier of a given order, we reconstruct life time orders as follows: each time a cancellation is detected, we go back through the history of limit order submission and look for a matching order with same price and same quantity. If an order is not matched, we discard the cancellation from our lifetime data. Results are presented in figure 21 and 22. We observe a power law decay with coefficients

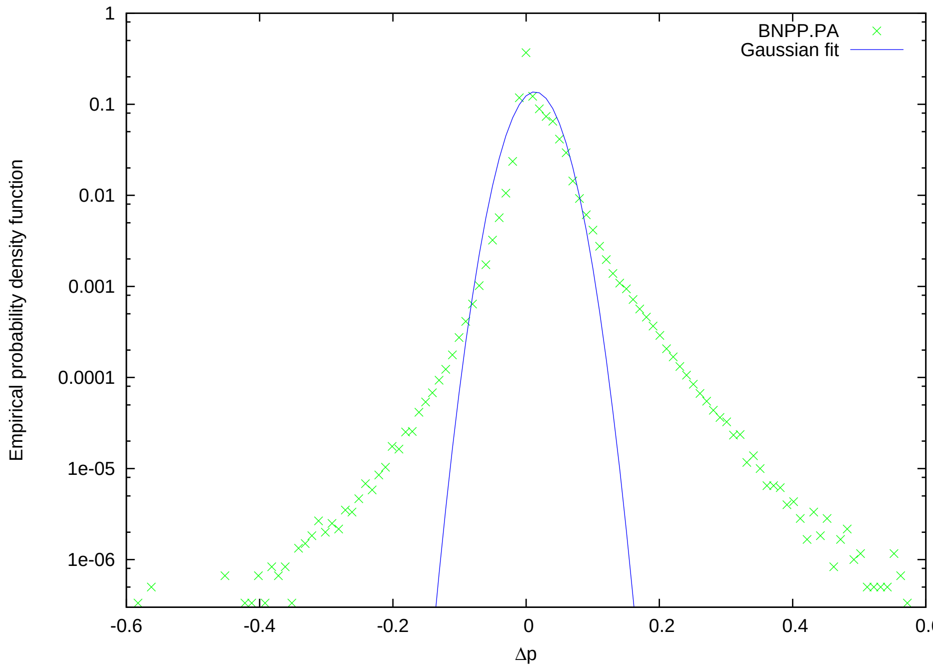

Figure 17: Placement of limit orders using the same best quote reference in semilog scale. Data used for this computation is BNP Paribas order book from 2007, September 1st until 2008, May 31st.

$1 + \mu \approx 1.3 - 1.6$ for both cancelled and executed limit orders, with little variations among stocks. These results are a bit different than the one presented in previous studies: similar for executed limit orders, but our data exhibits a lower decay as for cancelled orders. Not that the observed cut-off in the distribution for lifetimes above 20000 seconds is due to the fact that we do not take into account execution or cancellation of orders submitted on a previous day.

2.5. Intraday seasonality

Activity on financial markets is of course not constant throughout the day. Figure 23 (resp. 24) plots the (normalized) number of market (resp. limit) orders arriving in a 5-minute interval. It is clear that a U-shape is observed (an ordinary least-square quadratic fit is plotted): the observed market activity is larger at the beginning and the end of the day, and more quiet aroud lunchtime. Such a U-shaped curve is well-known, see (66) for example. On our data, we observe that the number of order on a 5-minute interval can vary with a factor 10 throughout the day.

(68) note that the average number of orders submitted to the market in a period ΔT vary wildly during the day. The authors also observe that these quantities for market orders and limit orders are highly correlated. Such a type of intraday variation of the global market activity is a well-known fact, already observed in (66) for example.

2.6. Market impact

The statistics we have presented may help to understand a phenomenon of primary importance for any financial market practitioner: the market impact, i.e. the relationship between the volume traded and the expected price shift once the order has been executed. On a first approximation, one understands that it is closely linked with many items described above: the volume of market orders submitted, the shape of the order book (how much pending limit orders are hit by one large market orders), the correlation of trade signs (one may assume that large orders are splitted in order to avoid a large market impact), etc.

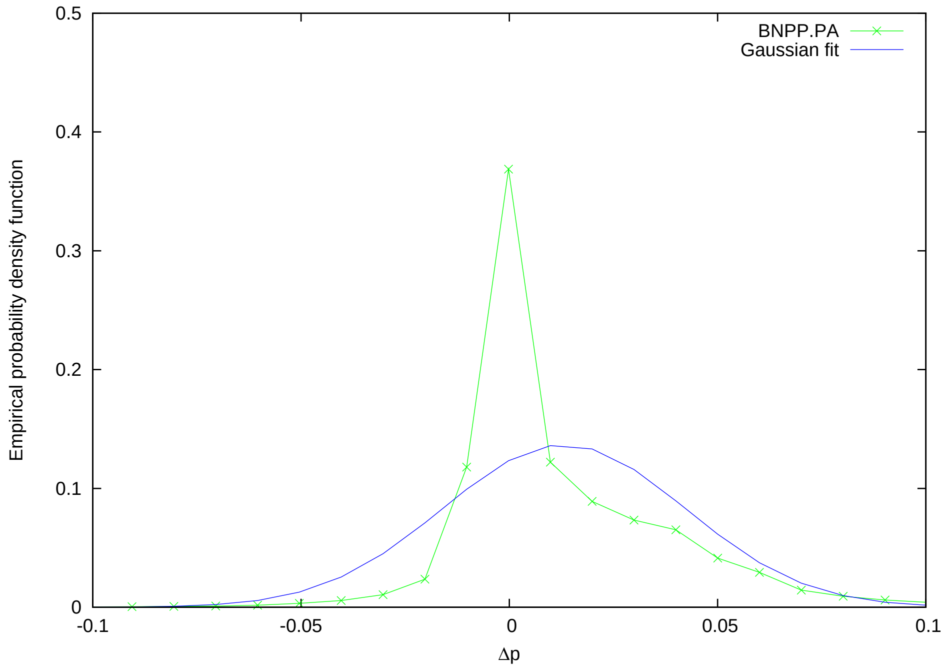

Figure 18: Placement of limit orders using the same best quote reference in linear scale. Data used for this computation is BNP Paribas order book from 2007, September 1st until 2008, May 31st.

Many empirical studies are available. An empirical study on the price impact of individual transactions on 1000 sotcks on the NYSE is conducted in (82). It is found that proper rescaling make all the curve collapse onto a single concave master curve. This function increases as a power that is the order of $1/2$ for small volumes, but then increases more slowly for large volumes. They obtain similar results in each year for the period 1995 to 1998.

We will not review any further the large literature of market impact, but rather refer the reader to the recent exhaustive synthesis proposed in (61), where different types of impacts, as well as some theoretical models are discussed.

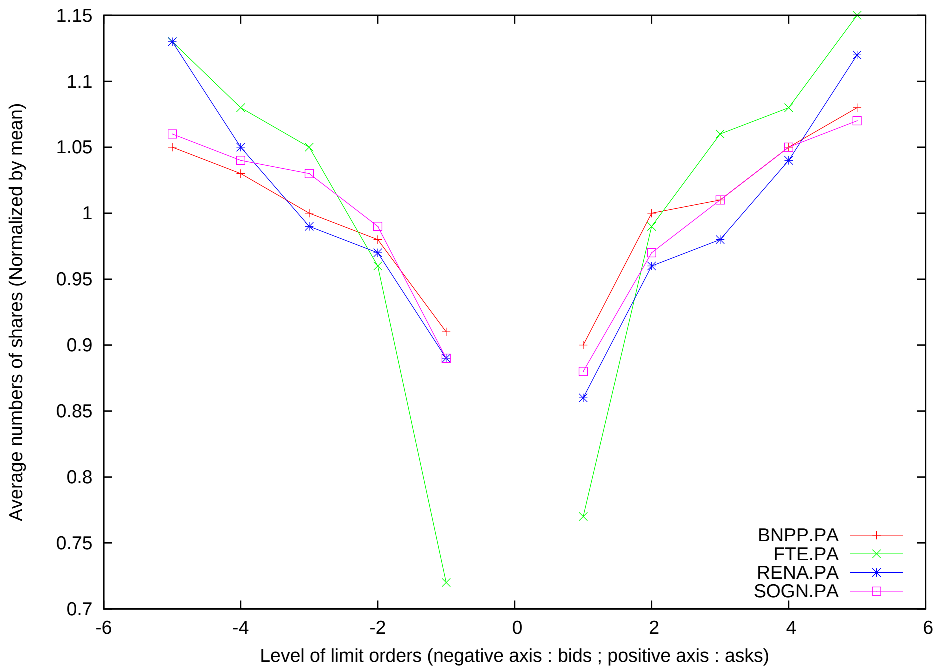

Figure 19: Average quantity offered in the limit order book.

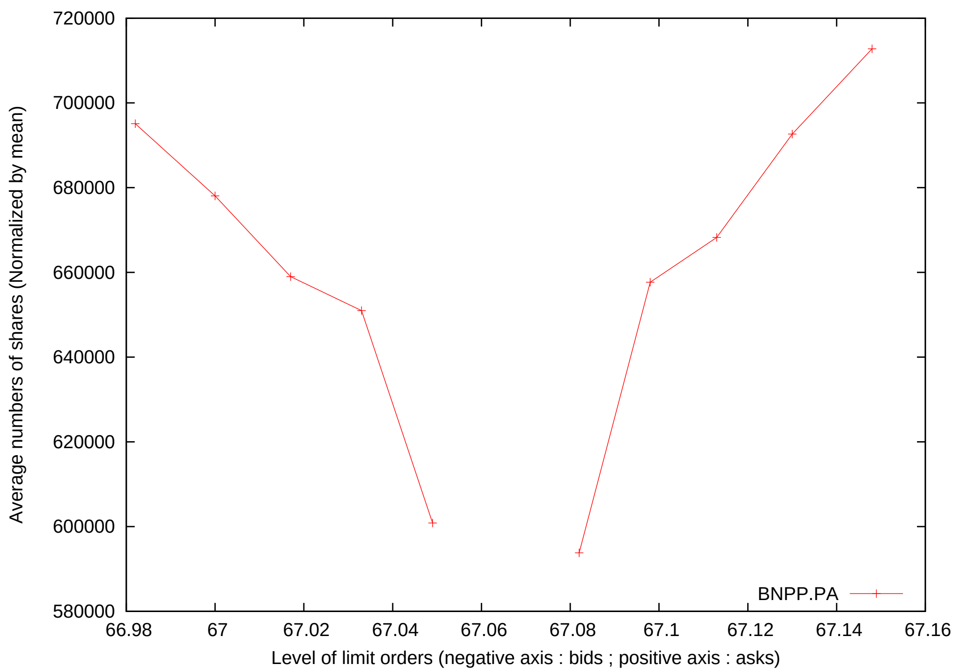

Figure 20: Average limit order book: price and depth.

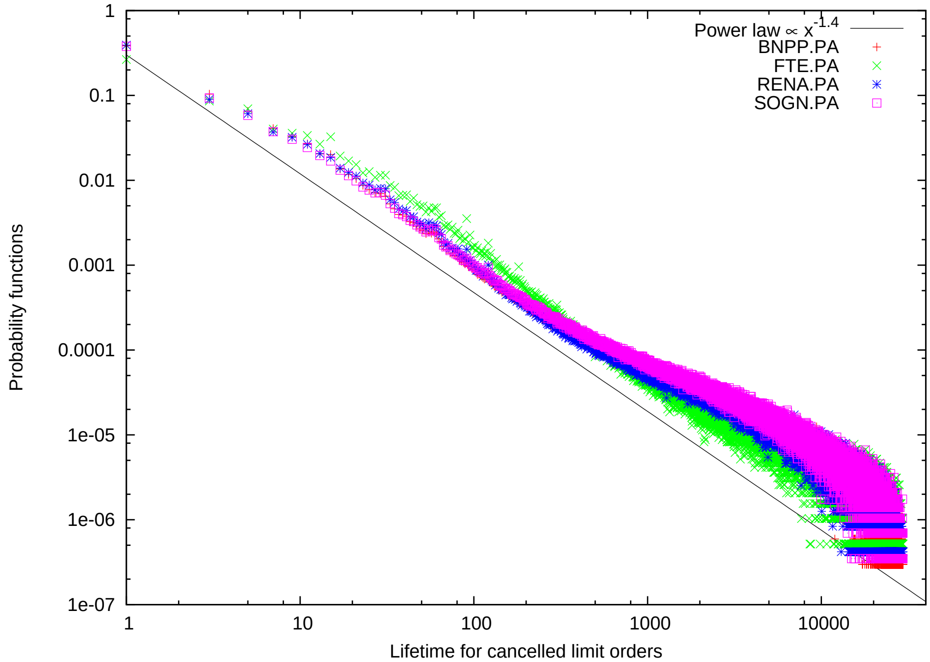

Figure 21: Distribution of estimated lifetime of cancelled limit orders.

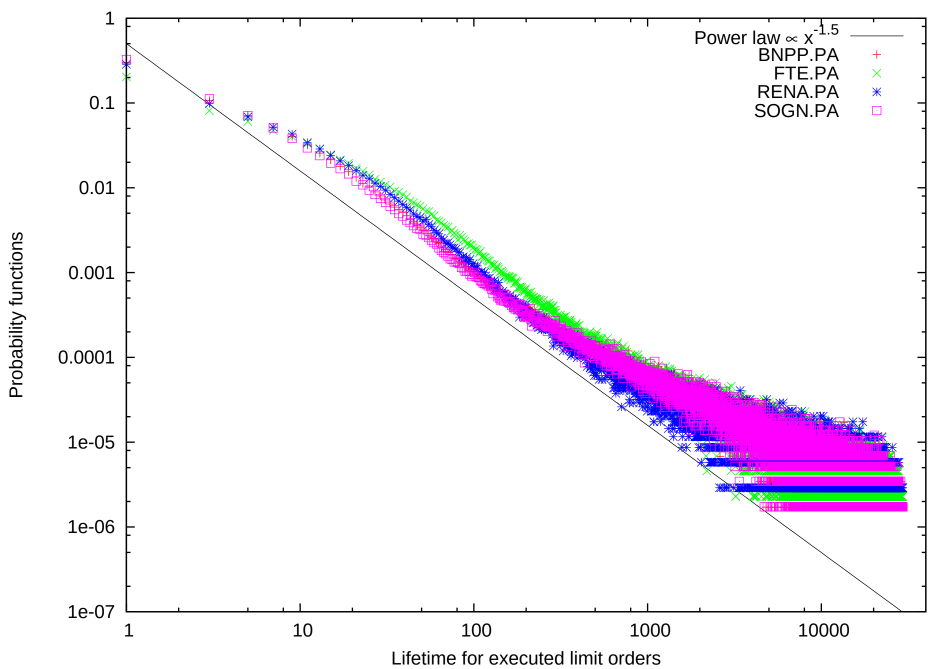

Figure 22: Distribution of estimated lifetime of executed limit orders.

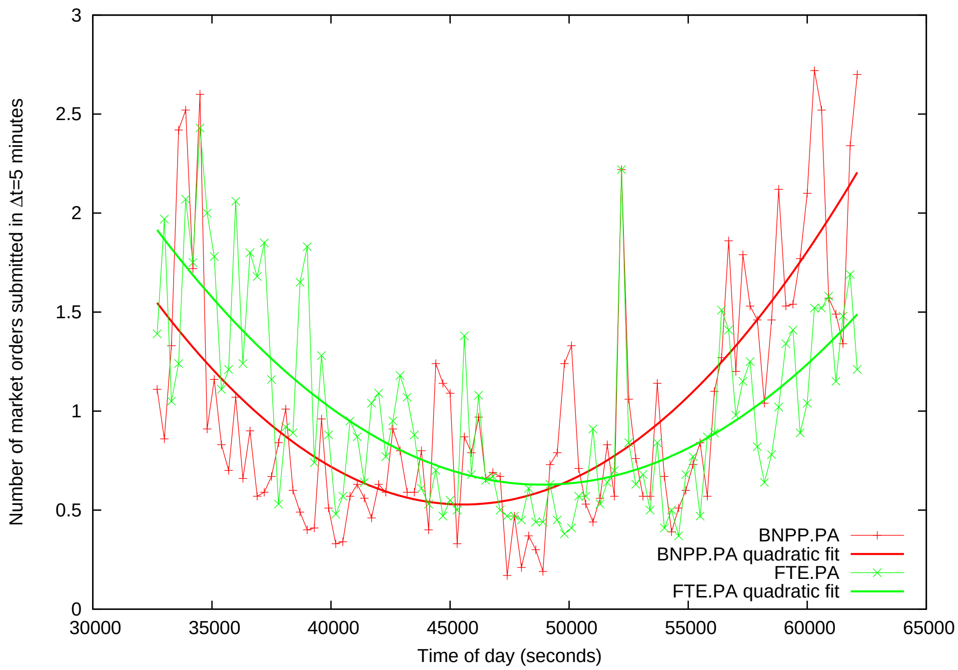

Figure 23: Normalized average number of market orders in a 5-minute interval.

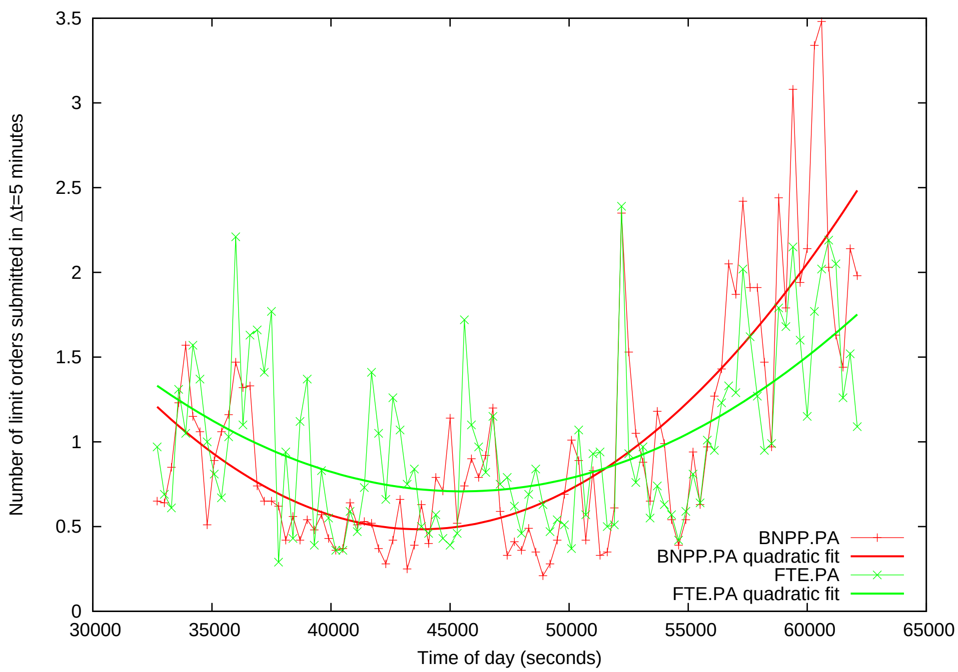

Figure 24: Normalized average number of limit orders in a 5-minute interval.

3. Correlations of assets

The word "correlation" is defined as "a relation existing between phenomena or things or between mathematical or statistical variables which tend to vary, be associated, or occur together in a way not expected on the basis of chance alone" (see http://www.m-w.com/dictionary/correlations). When we talk about correlations in stock prices, what we are really interested in are such relations with other variables, such as stock transaction volumes, as well as in the statistical distributions and laws which govern the price time series, the nature of the time series of stock prices, and whether the time series is random or not.

This section deals with several topics concerning linear correlation observed in financial data. The first part deals with the important issue of computing correlations in high-frequency. As mentioned earlier, the computerization of financial exchanges has lead to the availability of huge amount of tick-by-tick data, and computing correlation using these intraday data raises lots of issues concerning usual estimators. The second and third parts deals with the use of correlation in order to cluster assets with potential applications in risk management problems.

3.1. Estimating covariance on high-frequency data

Let us assume that we observe d time series of (log-)prices $p_i, i = 1, \dots, d$, observed at times $t_m, m = 0, \dots, M$. The usual estimator of the covariance of prices i and j is the *realized covariance estimator*, which is computed as:

$$\hat{\Sigma}_{ij}^{RV}(t) = \sum_{m=1}^{M} (p_i(t_m) - p_i(t_{m-1}))(p_j(t_m) - p_j(t_{m-1})). \tag{7}$$

The trouble is that high-frequency tick-by-tick data record changes of prices at random times, non-synchronously (contrary to daily close prices for example, that are recorded at the same time for all the assets on a given exchange). Using standard estimators without caution could be one cause for the "Epps effect", first observed in (83), which stated that "[c]orrelations among price changes in common stocks of companies in one industry are found to decrease with the length of the interval for which the price changes are measured." This has largely been verified since, e.g. in (84; 85). Non-synchronicity of tick-by-tick data and necessary sampling of time series in order to compute the usual realized covariance estimator partially explain this phenomenon (86). We very briefly review here two covariance estimators that do not need any synchronicity (hence, sampling) in order to be computed.

3.1.1. The Fourier estimator

The Fourier estimator has been introduced in (87). Let us assume that we have d time series of log-prices that are observations of Brownian semi-martingales:

$$dp_i = \sum_{j=1}^{K} \sigma_{ij} dW_j + \mu_i dt, i = 1, \dots, d. \tag{8}$$

The coefficient of the covariance matrix are then written $\Sigma_{ij}(t) = \sum_{k=1}^{K} \sigma_{ik}(t)\sigma_{jk}(t)$. (87) show that the Fourier coefficient of $\Sigma_{ij}(t)$ are, with n_0 a given integer:

$$a_k(\Sigma_{ij}) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{s=n_0}^{N} \frac{1}{2}(a_s(dp_i)a_{s+k}(dp_j) + a_{s+k}(dp_i)a_s(dp_j)), \tag{9}$$

$$b_k(\Sigma_{ij}) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{s=n_0}^{N} \frac{1}{2}(a_s(dp_i)b_{s+k}(dp_j) + b_{s+k}(dp_i)a_s(dp_j)), \tag{10}$$

where the Fourier coefficients $a_k(dp_i)$ and $b_k(dp_i)$ of dp_i can be directly computed on the time series. Indeed, rescaling the time window on $[0, 2\pi]$ and using integration by parts, we have:

$$a_k(dp_i) = \frac{p(2\pi) - p(0)}{\pi} - \frac{k}{\pi} \int_0^{2\pi} \sin(kt) p_i(t) dt. \tag{11}$$

This last integral can be discretized and approximately computed using the times t^i_m of observations of the process p_i. Therefore, fixing a sufficiently large N, one can compute an estimator Σ^F_{ij} of the covariance of the processes i and j. See (85; 88) for empirical studies using this estimator.

3.1.2. The Hayashi-Yoshida estimator

(86) have proposed a simple estimator in order to compute covariance/correlation without any need for synchronicity of time series. As in the Fourier estimator, it is assumed that the observed process is a Brownian semi-martingale. The time window of observation is easily partitioned into d family of intervals $\Pi^i = (J^i_m), i = 1, \ldots, d$, where $t^i_m = \inf\{U^i_{m+1}\}$ is the time of the m-th observation of the process i. Let us denote $\Delta p_i(U^i_m) = p_i(t^i_m) - p_i(t^i_{m-1})$. The *cumulative covariance estimator* as the authors named it, or the *Hayashi-Yoshida estimator* as it has been largely refered to, is then built as follows:

$$\hat{\Sigma}^{HY}_{ij}(t) = \sum_{m,n} \Delta p_i(U^i_m) \Delta p_j(U^j_n) \mathbf{1}_{\{U^i_m \cap U^j_n \neq \emptyset\}}. \tag{12}$$

3.2. Correlation matrix and Random Matrix Theory

The stock-market data is essentially a *multivariate* time-series data, we construct correlation matrix to study its spectra to contrast it with the random multivariate data from coupled map lattice. It is known from previous studies that the empirical spectra of correlation matrices drawn from time-series data, for most part, follow random matrix theory (89).

3.2.1. Correlation matrix and Eigenvalue density

Correlation matrix.

1. Financial Correlation matrix.
   If there are N assets with price $P_i(t)$ for asset i at time t, then the logarithmic return of stock i is $r_i(t) = \ln P_i(t) - \ln P_i(t-1)$, which for a certain consecutive sequence of trading days forms the return vector r_i. In order to characterize the synchronous time evolution of stocks, the equal time correlation coefficients between stocks i and j is defined as

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{[\langle r_i^2 \rangle - \langle r_i \rangle^2][\langle r_j^2 \rangle - \langle r_j \rangle^2]}}, \tag{13}$$

   where $\langle ... \rangle$ indicates a time average over the trading days included in the return vectors. These correlation coefficients form an $N \times N$ matrix with $-1 \leq \rho_{ij} \leq 1$. If $\rho_{ij} = 1$, the stock price changes are completely correlated; if $\rho_{ij} = 0$, the stock price changes are uncorrelated, and if $\rho_{ij} = -1$, then the stock price changes are completely anti-correlated.
2. Correlation matrix from spatio-temporal series from coupled map lattices.
   Consider a time-series of the form $z'(x,t)$, where $x = 1, 2, ...n$ and $t = 1, 2....p$ denote the discretised space and time, respectively. In this, the time-series at every spatial point is treated as a different variable. We define the normalised variable as

$$z(x,t) = \frac{z'(x,t) - \langle z'(x) \rangle}{\sigma(x)}, \tag{14}$$

   where the brackets $\langle . \rangle$ represent temporal averages and $\sigma(x)$ the standard deviation of z' at position x. Then, the equal-time cross-correlation matrix that represents the spatial correlations can be written as

$$S_{x,x'} = \langle z(x,t)\, z(x',t) \rangle, \qquad x, x' = 1, 2, \ldots, n. \tag{15}$$

   The correlation matrix is symmetric by construction. In addition, a large class of processes are translationally invariant and the correlation matrix can contain that additional symmetry, too. We will use this property for our correlation models in the context of coupled map lattice. In time-series analysis, the averages $\langle . \rangle$ have to be replaced by estimates obtained from finite samples. As usual, we will use the maximum likelihood estimates, $\langle a(t) \rangle \approx \frac{1}{p} \sum_{t=1}^{p} a(t)$. These estimates contain statistical uncertainties, which disappears for $p \to \infty$. Ideally, one requires $p \gg n$ to have reasonably correct correlation estimates. See arxiv:0704.1738 for details of parameters, etc.

Eigenvalue Density. The interpretation of the spectra of empirical correlation matrices should be done carefully if one wants to be able to distinguish between system specific signatures and universal features. The former express themselves in the smoothed level density, whereas the latter usually are represented by the fluctuations on top of this smooth curve. In time-series analysis, the matrix elements are not only prone to uncertainty such as measurement noise on the time-series data, but also statistical fluctuations due to finite sample effects. When characterizing time series data in terms of random matrix theory, one is not interested in these trivial sources of fluctuations which are present on every data set, but one would like to identify the significant features which would be shared, in principle, by an "infinite" amount of data without measurement noise. The eigenfunctions of the correlation matrices constructed from such empirical time-series carry the information contained in the original time-series data in a "graded" manner and they also provide a compact representation for it. Thus, by applying a random matrix theory based approach, one tries to identify non-random components of the correlation matrix spectra as deviations from random matrix theory predictions (89).

We will look at the eigenvalue density that has been studied in the context of applying random matrix theory methods to time-series correlations. Let $\mathcal{N}(\lambda)$ be the integrated eigenvalue density which gives the number of eigenvalues less than a given value λ. Then, the eigenvalue or level density is given by $\rho(\lambda) = \frac{d\mathcal{N}(\lambda)}{d\lambda}$. This can be obtained assuming random correlation matrix and is found to be in good agreement with the empirical time-series data from stock market fluctuations. From RMT considerations, the eigenvalue density for random correlations is given by

$$\rho_{rmt}(\lambda) = \frac{Q}{2\pi\lambda}\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}\,, \tag{16}$$

where $Q = N/T$ is the ratio of number of variables to length of each time-series. Here, λ_{max} and λ_{min} represent the maximum and minimum eigenvalues of the random correlation matrix respectively and are given by $\lambda_{max,min} = 1 + 1/Q \pm 2\sqrt{1/Q}$. However, due to presence of correlations in the empirical correlation matrix, this eigenvalue density is often violated for a certain number of dominant eigenvalues. They often correspond to system specific information in the data. In Fig. 25 we show the eigenvalue density for S&P500 data and also for the chaotic data from coupled map lattice. Clearly, both curves are qualitatively different. Thus, presence or absence of correlations in data is manifest in the spectrum of the corresponding correlation matrices.

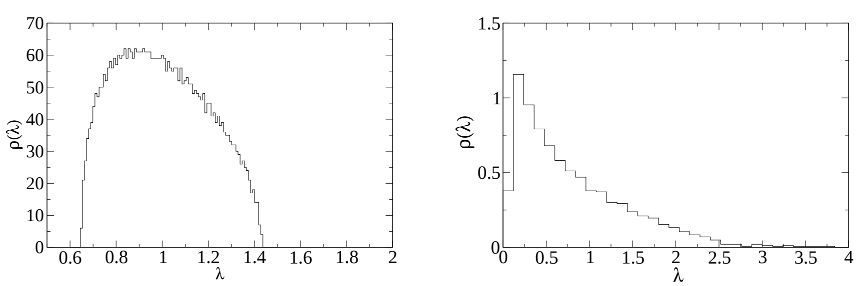

Figure 25: The left panel shows spectral density for multivariate spatio-temporal time-series drawn from coupled map lattices The right panel shows the eigenvalue density for the return time-series of the S&P500 stock market data (8938 time steps).

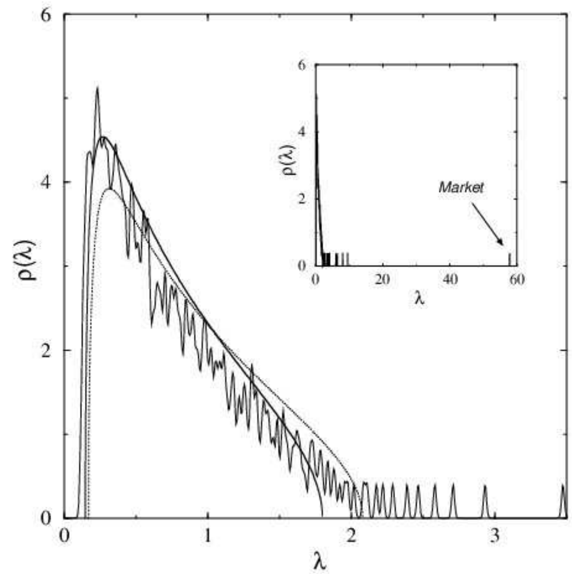

Figure 26: Eigenvalue spectrum of the correlation matrices. Adapted from cond-mat/9810255.

3.2.2. Earlier estimates and studies using Random Matrix Theory

Laloux et al. (90) showed that results from the random matrix theory (RMT) were useful to understand the statistical structure of the empirical correlation matrices appearing in the study of price fluctuations. The empirical determination of a correlation matrix is a difficult task. If one considers N assets, the correlation matrix contains $N(N-1)/2$ mathematically independent elements, which must be determined from N time-series of length T. If T is not very large compared to N, then generally the determination of the covariances is noisy, and therefore the empirical correlation matrix is to a large extent random. The smallest eigenvalues of the matrix are the most sensitive to this 'noise'. But the eigenvectors corresponding to these smallest eigenvalues determine the minimum risk portfolios in Markowitz theory. It is thus important to distinguish "signal" from "noise" or, in other words, to extract the eigenvectors and eigenvalues of the correlation matrix containing real information (those important for risk control), from those which do not contain any useful information and are unstable in time. It is useful to compare the properties of an empirical correlation matrix to a "null hypothesis"— a random matrix which arises for example from a finite time-series of strictly uncorrelated assets. Deviations from the random matrix case might then suggest the presence of true information. The main result of their study was the remarkable agreement between the theoretical prediction (based on the assumption that the correlation matrix is random) and empirical data concerning the density of eigenvalues (shown in Fig. 26) associated to the time-series of the different stocks of the S&P 500 (or other stock markets). Cross-correlations in financial data were also studied by Plerou et al. (91; 92). They analyzed cross-correlations between price fluctuations of different stocks using methods of RMT. Using two large databases, they calculated cross-correlation matrices of returns constructed from (i) 30-min returns of 1000 US stocks for the 2-yr period 1994–95, (ii) 30-min returns of 881 US stocks for the 2-yr period 1996–97, and (iii) 1-day returns of 422 US stocks for the 35-yr period 1962–96. They also tested the statistics of the eigenvalues λ_i of cross-correlation matrices against a "null hypothesis". They found that a majority of the eigenvalues of the cross-correlation matrices were within the RMT bounds, $[\lambda_{min}, \lambda_{max}]$, as defined above, for the eigenvalues of random correlation matrices. They also tested the eigenvalues of the cross-correlation matrices within the RMT bound for universal properties of random matrices and found good agreement with the results for the Gaussian orthogonal ensemble (GOE) of random matrices — implying a large degree of randomness in the measured cross-correlation coefficients. Furthermore, they found that the distribution of eigenvector components for the eigenvectors corresponding to the eigenvalues outside the RMT bound displayed systematic deviations from the RMT prediction and that these "deviating eigenvectors" were stable in time. They analyzed the components of the deviating eigenvectors and found that the largest eigenvalue corresponded to an influence common to all stocks. Their analysis of the remaining deviating eigenvectors showed distinct groups, whose identities corresponded to conventionally-identified business sectors.

3.3. Analyses of correlations and economic taxonomy

3.3.1. Models and theoretical studies of financial correlations

Podobnik et al. (93) studied how the presence of correlations in physical variables contributes to the form of probability distributions. They investigated a process with correlations in the variance generated by a Gaussian or a truncated Levy distribution. For both Gaussian and truncated Levy distributions, they found that due to the correlations in the variance, the process "dynamically" generated power-law tails in the distributions, whose exponents could be controlled through the way the correlations in the variance were introduced. For a truncated Levy distribution, the process could extend a truncated distribution beyond the *truncation cutoff*, leading to a crossover between a Levy stable power law and their "dynamically-generated" power law. It was also shown that the process could explain the crossover behavior observed in the S&P 500 stock index.

Jae Dong Noh (94) proposed a model for correlations in stock markets in which the markets were composed of several groups, within which the stock price fluctuations were correlated. The spectral properties of empirical correlation matrices (91; 90) were studied in relation to this model and the connection between the spectral properties of the empirical correlation matrix and the structure of correlations in stock markets was established.

The correlation structure of extreme stock returns were studied by Cizeau et al. (95). It has been commonly believed that the correlations between stock returns increased in high volatility periods. They investigated how much of these correlations could be explained within a simple non-Gaussian one-factor description with time independent correlations. Using surrogate data with the true market return as the dominant factor, it was shown that most of these correlations, measured by a variety of different indicators, could be accounted for. In particular, their one-factor model could explain the level and asymmetry of empirical exceeding correlations. However, more subtle effects required an extension of the one factor model, where the variance and skewness of the residuals also depended on the market return.

Burda et al. (96) provided a statistical analysis of three S & P 500 covariances with evidence for raw tail distributions. They studied the stability of these tails against reshuffling for the S&P 500 data and showed that the covariance with the strongest tails was robust, with a spectral density in remarkable agreement with random Levy matrix theory. They also studied the inverse participation ratio for the three covariances. The strong localization observed at both ends of the spectral density was analogous to the localization exhibited in the random Levy matrix ensemble. They showed that the stocks with the largest scattering were the least susceptible to correlations and were the likely candidates for the localized states.

3.3.2. Analyses of correlations and economic taxonomy

Mantegna introduced a method (97) for finding a hierarchical arrangement of stocks traded in financial market, through studying the clustering of companies by using correlations of asset returns. With an appropriate metric, based on the earlier explained correlation matrix coefficients ρ_{ij}'s computed between all pairs of stocks i and j in Eq. 13, of the portfolio by considering the synchronous time evolution of the difference of the logarithm of daily stock price, a fully connected graph was defined in which the nodes are companies, or stocks, and the "distances" between them were obtained from the corresponding correlation coefficients. The minimum spanning tree (MST) was generated from the graph by selecting the most important correlations and it was used to identify clusters of companies. The hierarchical tree of the sub-dominant ultrametric space associated with the graph provided information useful to investigate the number and nature of the common economic factors affecting the time evolution of logarithm of price of well defined groups of stocks. Several other attempts have been made to obtain clustering from the huge correlation matrix.

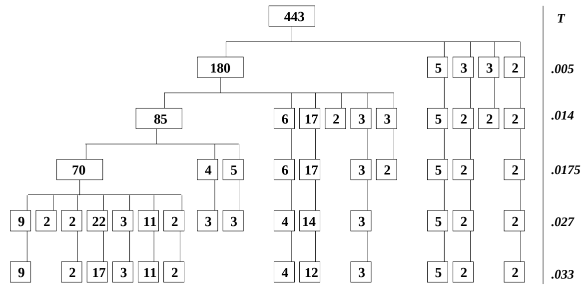

Figure 27: The hierarchical structure of clusters of the S&P 500 companies in the ferromagnetic case. In the boxes the number of the elements of the cluster are indicated. The clusters consisting of single companies are not indicated. Adapted from L. Kullmann, J. Kertesz, and R. N. Mantegna, arXiv.org:cond-mat/0002238.

Bonanno et al. (98) studied the high-frequency cross-correlation existing between pairs of stocks traded in a financial market in a set of 100 stocks traded in US equity markets. A hierarchical organization of the investigated stocks was obtained by determining a metric distance between stocks and by investigating the properties of the sub-dominant ultrametric associated with it. A clear modification of the hierarchical organization of the set of stocks investigated was detected when the time horizon used to determine stock returns was changed. The hierarchical location of stocks of the energy sector was investigated as a function of the time horizon. The hierarchical structure explored by the minimum spanning tree also seemed to give information about the influential power of the companies.

It also turned out that the hierarchical structure of the financial market could be identified in accordance with the results obtained by an independent clustering method, based on Potts super-paramagnetic transitions as studied by Kullmann et al. (99), where the spins correspond to companies and the interactions are functions of the correlation coefficients determined from the time dependence of the companies' individual stock prices. The method is a generalization of the clustering algorithm by Domany et. al. to the case of anti-ferromagnetic interactions corresponding to anti-correlations. For the Dow Jones Industrial Average, no anti-correlations were observed in the investigated time period and the previous results obtained by different tools were well reproduced. For the S&P 500, where anti-correlations occur, repulsion between stocks modified the cluster structure of the $N = 443$ companies studied, as shown in Fig. 3.3.2. The efficiency of the method is represented by the fact that the figure matches well with the corresponding result obtained by the minimal spanning tree method, including the specific composition of the clusters. For example, at the lowest level of the hierarchy (highest temperature in the super-paramagnetic phase) the different industrial branches can be clearly identified: Oil, electricity, gold mining, etc. companies build separate clusters. The network of influence was recently investigated by means of a time-dependent correlation method by Kullmann et al. (99). They studied the correlations as the function of the time shift between pairs of stock return time series of tick-by-tick data of the NYSE. They investigated whether any "pulling effect" between stocks existed or not, i.e., whether at any given time, the return value of one stock influenced that of another stock at a different time or not. They found that, in general, two types of mechanisms generated significant correlation between any two given stocks. One was some kind of external effect (say, economic or political news) that influenced both stock prices simultaneously, and the change for both prices appeared at the same time, such that the maximum of the correlation was at zero time shift. The second effect was that, one of the companies had an influence on the other company indicating that one company's operation depended on the other, so that the price change of the influenced stock appeared latter because it required some time to react on the price change of the first stock displaying a "pulling effect". A weak but significant effect with the real data set was found, showing that in many cases the maximum correlation was at non-zero time shift indicating directions of influence between the companies, and the characteristic time was of the order of a few minutes, which was compatible with efficient market hypothesis. In the pulling effect, they found that in general, more important companies (which were traded more) pulled the relatively smaller companies.

The time dependent properties of the minimum spanning tree (introduced by Mantegna), called a 'dynamic asset tree', were studied by Onnela et al. (100). The nodes of the tree were identified with stocks and the distance between them was a unique function of the corresponding element of the correlation matrix. By using the concept of a central vertex, chosen as the most strongly connected node of the tree, the mean occupation layer was defined which was an important characteristic of the tree. During crashes the strong global correlation in the market manifested itself by a low value of the mean occupation layer. The tree seemed to have a scale free structure where the scaling exponent of the degree distribution was different for 'business as usual' and 'crash' periods. The basic structure of the tree topology was very robust with respect to time. Let us discuss in more details how the dynamic asset tree was applied to studies of economic taxonomy.

Financial Correlation matrix and constructing Asset Trees. Two different sets of financial data were used: The first set from the Standard & Poor's 500 index (S&P500) of the New York Stock Exchange (NYSE) from July 2, 1962 to December 31, 1997 containing 8939 daily closing values, and the second set from the split-adjusted daily closure prices for a total of $N = 477$ stocks traded at the New York Stock Exchange (NYSE) over the period of 20 years, from 02-Jan-1980 to 31-Dec-1999 were used. This amounted a total of 5056 price quotes per stock, indexed by time variable $\tau = 1, 2, \ldots, 5056$. For analysis and smoothing purposes, the data was divided time-wise into M *windows* $t = 1, 2, ..., M$ of width T, where T corresponded to the number of daily returns included in the window. Note that several consecutive windows overlap with each other, the extent of which is dictated by the window step length parameter δT, which describes the displacement of the window and is also measured in trading days. The choice of window width is a trade-off between too noisy and too smoothed data for small and large window widths, respectively. The results presented here were calculated from monthly stepped four-year windows, i.e. $\delta T = 250/12 \approx 21$ days and $T = 1000$ days. A large scale of different values for both parameters were explored, and the cited values were found optimal (101). With these choices, the overall number of windows is $M = 195$.

The earlier definition of correlation matrix, given by Eq. 13 is used. These correlation coefficients form an $N \times N$ correlation matrix $\mathbf{C}^t$, which serves as the basis for trees discussed below.

An asset tree is then constructed according to the methodology by Mantegna (97). For the purpose of constructing asset trees, a distance is defined between a pair of stocks. This distance is associated with the edge connecting the stocks and it is expected to reflect the level at which the stocks are correlated. A simple non-linear transformation $d_{ij}^t = \sqrt{2(1-\rho_{ij}^t)}$ is used to obtain distances with the property $2 \geq d_{ij} \geq 0$, forming an $N \times N$ symmetric distance matrix $\mathbf{D}^t$. So, if $d_{ij} = 0$, the stock price changes are completely correlated; if $d_{ij} = 2$, the stock price changes are completely anti-uncorrelated. The trees for different time windows are not independent of each other, but form a series through time. Consequently, this multitude of trees is interpreted as a sequence of evolutionary steps of a single *dynamic asset tree*. An additional hypothesis is required about the topology of the metric space: the ultrametricity hypothesis. In practice, it leads to determining the minimum spanning tree (MST) of the distances, denoted $\mathbf{T}^t$. The spanning tree is a simply connected acyclic (no cycles) graph that connects all N nodes (stocks) with $N-1$ edges such that the sum of all edge weights, $\sum_{d_{ij}^t \in \mathbf{T}^t} d_{ij}^t$, is minimum. We refer to the minimum spanning tree at time t by the notation $\mathbf{T}^t = (V, E^t)$, where V is a set of vertices and E^t is a corresponding set of unordered pairs of vertices, or edges. Since the spanning tree criterion requires all N nodes to be always present, the set of vertices V is time independent, which is why the time superscript has been dropped from notation. The set of edges E^t, however, does depend on time, as it is expected that edge lengths in the matrix $\mathbf{D}^t$ evolve over time, and thus different edges get selected in the tree at different times.

Market characterization. We plot the distribution of (i) distance elements d_{ij}^t contained in the distance matrix $\mathbf{D}^t$ (Fig. 28), (ii) distance elements d_{ij} contained in the asset (minimum spanning) tree $\mathbf{T}^t$ (Fig. 29). In both plots, but most prominently in Fig. 28, there appears to be a discontinuity in the distribution between roughly 1986 and 1990. The part that has been cut out, pushed to the left and made flatter, is a manifestation of Black Monday (October 19, 1987), and its length along the time axis is related to the choice of window width T (102; 100). Also, note that in the distribution of tree edges in Fig. 29 most edges included in the tree seem to come from the area to the right of the value 1.1 in Fig. 28, and the largest distance element is $d_{max} = 1.3549$.

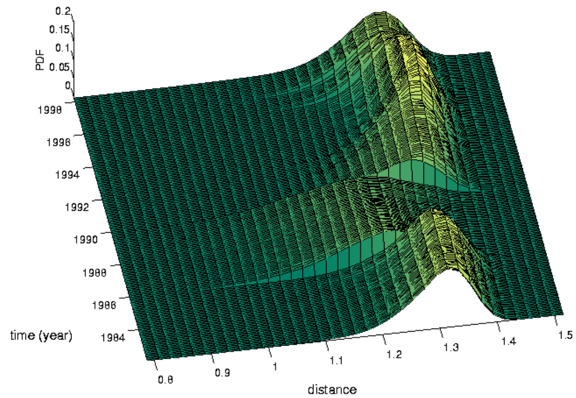

Figure 28: Distribution of all $N(N-1)/2$ distance elements d_{ij} contained in the distance matrix $\mathbf{D}^t$ as a function of time.

1. Tree occupation and central vertex
   Let us focus on characterizing the spread of nodes on the tree, by introducing the quantity of *mean occupation layer*
   $$l(t, v_c) = \frac{1}{N} \sum_{i=1}^{N} \text{lev}(v_i^t)\,, \tag{17}$$
   where $\text{lev}(v_i)$ denotes the level of vertex v_i. The levels, not to be confused with the distances d_{ij} between nodes, are measured in natural numbers in relation to the *central vertex* v_c, whose level is taken to be zero. Here the mean occupation layer indicates the layer on which the mass of the tree, on average, is conceived to be located. The central vertex is considered to be the parent of all other nodes in the tree, and is also known as the root of the tree. It is used as the *reference* point in the tree, against which the locations of all other nodes are relative. Thus all other nodes in the tree are children of the central vertex. Although there is an *arbitrariness* in the choice of the central vertex, it is proposed that the vertex is central, in the sense that any change in its price strongly affects the course of events in the market on the whole. Three alternative definitions for the central vertex were proposed in the studies, all yielding similar and, in most cases, identical outcomes. The idea is to find the node that is most strongly connected to its nearest neighbors. For example, according to one definition, the central node is the one with the highest *vertex degree*, i.e. the number of edges which are incident with (neighbor of) the vertex. Also, one may have either (i) static (fixed at all times) or (ii) dynamic (updated at each time step) central vertex, but again the results do not seem to vary significantly. The study of the variation of the topological properties and nature of the trees, with time were done.
2. Economic taxonomy
   Mantegna's idea of linking stocks in an ultrametric space was motivated *a posteriori* by the property of such a space to provide a meaningful economic taxonomy (103). Mantegna examined the meaningfulness of the taxonomy, by comparing the grouping of stocks in the tree with a third party reference grouping of stocks e.g. by their industry classifications (97). In this case, the reference was provided by Forbes [www forbes com], which uses its own classification system, assigning each stock with a sector (higher level) and industry (lower level) category. In order to visualize the grouping of stocks, a sample asset tree is constructed for a smaller dataset (shown in Fig. 30), which consists of 116 S&P 500 stocks, extending from the beginning of 1982 to the end of 2000, resulting in a total of 4787 price quotes per stock (**?**). The window width was set at $T = 1000$, and the shown sample tree is located time-wise at $t = t^*$, corresponding to 1.1.1998. The stocks in this dataset fall into 12 *sectors*, which are

Basic Materials, Capital Goods, Conglomerates, Consumer/Cyclical, Consumer/Non-Cyclical, Energy, Financial, Healthcare, Services, Technology, Transportation and Utilities. The sectors are indicated in the tree (see Fig. 30) with different markers, while the industry classifications are omitted for reasons of clarity. The term sector is used exclusively to refer to the given third party classification system of stocks. The term *branch* refers to a subset of the tree, to all the nodes that share the specified common parent. In addition to the parent, it is needed to have a reference point to indicate the generational direction (i.e. who is who's parent) in order for a branch to be well defined. Without this reference there is absolutely no way to determine where one branch ends and the other begins. In this case, the reference is the central node. There are some branches in the tree, in which most of the stocks belong to just one sector, indicating that the branch is fairly homogeneous with respect to business sectors. This finding is in accordance with those of Mantegna (97) , although there are branches that are fairly heterogeneous, such as the one extending directly downwards from the central vertex (see Fig. 30).

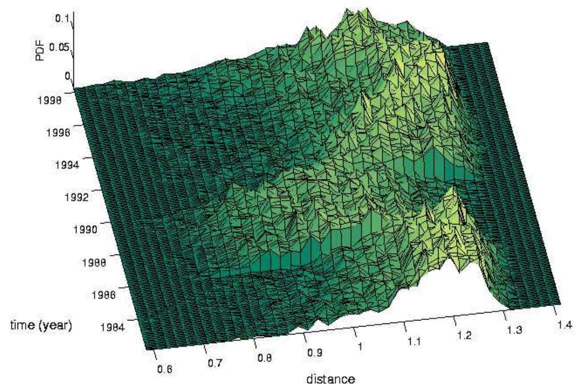

Figure 29: Distribution of the $(N-1)$ distance elements d_{ij} contained in the asset (minimum spanning) tree $\mathbf{T}^t$ as a function of time.

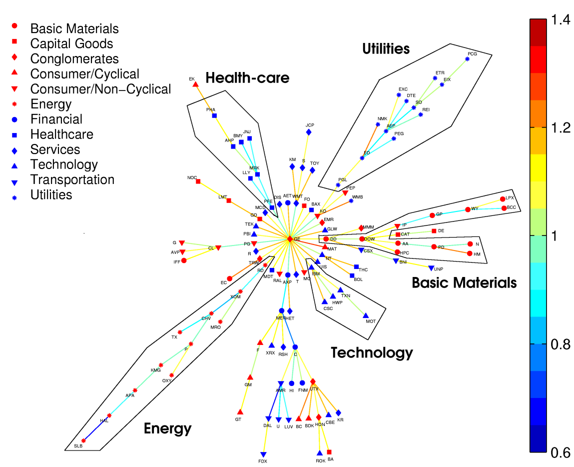

Figure 30: Snapshot of a dynamic asset tree connecting the examined 116 stocks of the S&P 500 index. The tree was produced using four-year window width and it is centered on January 1, 1998. Business sectors are indicated according to Forbes [www forbes com]. In this tree, General Electric (GE) was used as the central vertex and eight layers can be identified.

Part IV
Agent-based Modeling

1. Agent-based modeling of order books

1.1. Introduction

Although known – at least partly – for a long time (in his 1963 paper (52), Mandelbrot gives a reference for a paper dealing with non-normality of price time series in 1915, followed by several in the 1920's), "stylized facts" have often been left aside when modeling financial markets. They were even often refered to as "anomalous" characteristics as if observations failed to comply with theory.

Much has been done these past fifteen years in order to address this challenge and provide new models that can reproduce these facts. These recent developments have been built on top of early attempts at modeling mechanisms of financial markets with agents. Stigler (104) investigating some rules of the SEC (Security Exchange Commission), or Garman (72) investigating double-auction microstructure belong to those historical works.

It seems that the first modern attempts at that type of models were made in the field of behavioral finance. Agent-based models in financial economics are built with numerous agents who can exchange shares of stocks according to exogenously defined utility functions reflecting their preferences and risk aversions. (105) provides a recent review of that type of model. Although achieving some of their goals, these models suffer from many drawbacks: first, they are very complex, and it would be a very difficult task to try to identify the role of their numerous parameters and the types of dependance to these parameters; second, the chosen utility functions do not reflect what is observed on the mechanisms of a financial market.

A sensible change in modelisation appears with much simpler models implementing only well-identified and presumably realistic "behavior": (106) uses noise traders that are subject to "herding", i.e. form random clusters of traders sharing the same view on the market. The idea is used in (107) as well. A complementary approach is to characterize traders as fundamentalists, chartists or noise traders. (108) propose an agent-based model in which these types of traders interact. In all these models, the price variation directly results from the excess demand: at each time step, all agents submit orders and the resulting price is computed. Therefore, everything is cleared at each time step and there is no structure of order book to keep track of orders.

One big step is made with models really taking into account limit orders and keeping them in an order book once submitted and not executed. (109) build an agent-based model where all traders submit orders depending on the three elements identified in (108): chartists, fundamentalists, noise. Orders submitted are then stored in a persistent order book. In fact, one of the first simple models with this feature was proposed in (110). In this model, orders are particles moving along a price line, and each collision is a transaction. Due to numerous caveats in this model, the authors propose in the same paper an extension wiht fundamentalist and noise traders in the spirit of the models previously evoked. (111) goes further in the modeling of trading mechanisms by taking into account fixed limit orders and market orders that trigger transactions. The order book is here really simulated. This model was analytically solved using a mean-field approximation in (112).

Following this trend of modeling, the – more or less – rational agents composing models in economics tends to vanish and be replaced by the notion of flows: orders are not submitted any more buy an agent following a strategic behavior, but are viewed as an arriving flows whose properties are to be determined by empirical observations on market mechanisms. Thus, the modeling of order books calls for more "stylized facts", i.e. empirical properties that could be observed on a large number of order-driven markets. (66) is a thorough empirical study of the order flows in the Paris Bourse a few years after its complete computerization. Market orders, limit orders, time of arrivals and placement are studied. (70) and (71) provides statistical features on the order book itself. These empirical studies are the fundation for Zero-Intelligence models, where "stylized facts" are expected to be reproduced by the properties of the order flows and the structure of order book itself, without considering exogenous "rationality". (68) propose a simple model of order flows: limit orders are deposited in the order book and can be removed if not executed, in a simple deposition-evaporation process. (70) use this type of model with empirical distribution as inputs. As of today, the most complete empirical model is to our knowledge (81), where order placement and cancellation models are proposed and fitted on empirical data. Finally, new challenges arise as scientists try to identify simple mechanisms that allow an agent-based model to reproduce non-trivial behaviours: herding behaviour in(106), dynamic price placement in (113), threshold behaviour in (114), etc.

In this paper we review some of these models. This survey is of course far from exhaustive, and we have just selected models that we feel are representative of a specific trend of modeling.

1.2. Early order-driven market modeling: Market microstructure and policy issues

The pioneering works in simulation of financial markets were aimed to study market regulations. The very first one, (104), tries to investigate the effect of regulations of the SEC on American stock markets, using empirical data from the 20's and the 50's. Twenty years later, at the start of the computerization of financial markets, (115) implements a simulator in order to test the feasibility of automated market making. This section gives a few details on those early models.

1.2.1. Controversy in regulations issues: Stigler (1964)

Presentation of the model. To our knowledge, the first attempt to simulate a financial market is done by (104). This paper is a biting and controversial reaction to the Report of the Special Study of the Securities Markets of the SEC ((116)), whose aim was to "study the adequacy of rules of the exchange and that the New York stock exchange undertakes to regulate its members in all of their activities" ((117)). According to the author this study lacks rigorous tests when investigating the effects of regulation on financial markets. Stating that "demand and supply are [...] erratic flows with sequences of bids and asks dependent upon the random circumstances of individual traders", Stigler proposes a simple simulation model to investigate the evolution of the market. In this model, contrained by simulation capability in 1964, price is constrained within $L = 10$ ticks. (Limit) orders are randomly drawn, in trade time, as follows: they can be bid (resp.ask) orders with probability 1/2 (resp. 1/2), and their price level is uniformly distributed on the price grid. Each time an order crosses the opposite best quote, it is a market order. All orders are of size one. Orders not executed $N = 25$ time steps after their submission are cancelled. Thus, N is the maximum number of orders available in the order book.

Results. In the original paper, a run of a hundred trades is manually computed using tables of random numbers. Of course, no particular results concerning the "stylized facts" of financial time series is expected at that time. However, let us mention that in his review of some order book models, (118) makes simulations of a similar model, with parameters $L = 5000$ and $N = 5000$. It appears that price returns are not gaussian, and that their distribution exhibits power law with exponent -0.3 (far from empirical data). As expected, the limitation L is responsible for a sharp cut-off of the tails of this distribution.

1.2.2. Microstructure of the double auction: Garman (1976)

Presentation of the model. (72) provides an early study of the double auction market with a point of view that doesn't ignore temporal structure, and really defines order flows. Price is discrete and constrained to be within $\{p_1, p_L\}$. Buy and sell orders are assumed to be submitted according to two Poisson processes of intensities λ and μ. Each time an order crosses the best opposite quote, it is a market order. All quantities are asssumed to be equal to one. With its fixed contrained prices, (72) also defines the state of the order book at a given time as the vector $(n_i)_{i=1,...,L}$ of awaiting orders (negative quantity for bid orders, positive for ask orders).

Results. The aim of (72) is to provide an empirical study of the market microstructure. The main result of their Poisson model is to support the idea that negative correlation of consecutive price changes is linked the microstructure of the double auction exchange. The author points out that analytical solution of this simple Poisson model is not easy and provide Monte Carlo simulation. In its conclusion, the paper ask a visionary question: "Several topics for future investigations are suggested. For example, does the auction-market model imply the characteristic leptokurtosis seen in empirical security price changes ?"

1.2.3. Investigating clearing prices mechanisms: Hakansson, Beja and Kale (1985)

Presentation of the model. (115) consists of N agents trading S different securities. Let us point out already that it is the first model we know of to tackle (in theory at least) multidimensionality. Price p of securities is discrete on grid of ticks: $p = p_k, k \in \mathbf{N}$. The simulation in this model has two steps. First, traders generate randomly their demand function (a set of limit and market orders) by adding to their previous demand a random shift following a geometric random walk with a drift. Short positions are allowed. The condition that demand decreases with the price is ensured, so that the aggregated demand function $\bar{q}(p)$ is decreasing with price as well. The second step of the simulation is then to choose a price of execution. The demand being decreasing, the "equilibrium" price of the Walrasian auction is somewhere in the interval $[p_k; p_k + 1]$ where $k = \max\{k|\bar{q}(p_k) > 0\}$. Once the price is chosen, the market is cleared and traders informed of the results.

Results. The aim of the model is to study different mechanisms to choose the clearing price. Many rules are investigated: chosing p_k or $p_k + 1$; minimize price changes; minimize market maker costs; etc. Simulations are here used to compare different types of policies for an automated market maker. Although no attention is paid to time series, this model is an interesting early work in agent-based models with many agents.

1.2.4. Introducing zero-intelligence agents: Gode and Sunder (1993)

Presentation of the model. To our knowledge, (119) is the first paper to introduce the expression "zero-intelligence" in order to describe non-strategic behavior of traders. It is applied to traders that submit random orders in a double auction market. The expression has since been widely used in agent-based modeling (see more recent models described in this review). In this model, the price is an integer $p \in \{1, \ldots, L\}, L = 200$. Each trader is either a seller or a buyer. At the start of a trading period, trader i is given an allowed quantity q_i to buy or sell and a reference price p_i^R. This price is a limit price above which (resp. below which) a buyer (resp. seller) will loose money. These indications q_i and p_i^R ensure that the model has an increasing supply function and a decreasing demand function that are in line with neoclassic (Walrasian) economic theory, and hence imply the existence of an equilibrium price.

Two types of zero-intelligence traders are studied. First are unconstrained zero-intelligence traders. These agents can submit random order at random prices, within the allowed price range $\{1, \ldots, L\}$. Second are constrained zero-intelligence traders. These agents submit random orders as well, but with the constraint that they cannot cross their reference price p_i^R given: constrained zero-intelligence traders are not allowed to buy or sell at loss. Each time a transaction occurs, all pending orders are cancelled.

Results. The aim of the authors is to show that double auction markets exhibit an intrinsic "allocative efficiency" (ratio between the total profit earned by the traders divided by the maximum possible profit) even with zero-intelligence traders. In particular, in their experiment, it appears that price series resulting from actions by unconstrained zero-intelligence traders are highly volatile, while prices series in a market of constrained agents exhibits convergence towards an equilibrium price, similar to the one obtain when ("rational") human traders operate on the same market.

These results have been criticized in (120). This paper shows with details that the observed convergence of the simulated price towards the theoretical equilibrium price is an artifact of the model. More precisely, the choice of traders' demand carry a lot of constraints that alone explain the observed results. Convergence towards an equilibrium price might not appear in all situations.

1.2.5. Partial conclusion on early models

Much modern works owe a lot to those four contributions. Let us mention one big difference between these papers. In the last two papers (115) and (119), variations of the price are computed by aggregating demand of traders and finding an intersection between supply and demand function, therefore assuming some sort of "Walrasian" auction mechanism. Paradoxically, the two older contribution are much more in line with recent Econophysics developments, where prices are defined by the "meeting" of individual orders. The next subsection describes these models.

1.3. Early order-driven market modeling in Econophysics

Starting in the mid-90's, physicists have propose order book models direclty inspired from Physics, where the analogy "order = particle" is emphasized. Three main contributions are presented here.

1.3.1. The order book as a reaction-diffusion model: Bak,Paczuski and Shubik (1997)

Presentation of the model. A very simple model directly taken from Physics is presented in (110). The authors consider a market with N noise traders able to exchange one share of stock at a time. Price $p(t)$ at time t is constrained to be an integer (i.e. price is quoted in number of ticks) with a upper bound $\bar{p}$: $\forall t, \quad p(t) \in \{0, \ldots, \bar{p}\}$. Simulation is initiated at time 0 with half of the agents asking for one share of stock (buy orders, bid) with price:

$$p_b^j(0) \in \{0, \bar{p}/2\}, \qquad j = 1, \ldots, N/2, \tag{18}$$

and the other half offering one share of stock (sell orders, ask) with price:

$$p_s^j(0) \in \{\bar{p}/2, \bar{p}\}, \qquad j = 1, \ldots, N/2. \tag{19}$$

At each time step t, agents revise their offer by exactly one tick, with equal probability to go up or down. Therefore, at time t, each seller (resp. buyer) agent chooses his new price as:

$$p_s^j(t+1) = p_s^j(t) \pm 1 \qquad (\text{resp. } p_b^j(t+1) = p_b^j(t) \pm 1\ ). \tag{20}$$

A transaction occurs when there exists $(i,j) \in \{1, \ldots, N/2\}^2$ such that $p_b^i(t+1) = p_s^j(t+1)$. In such a case the orders are removed and the transaction price is recorded as the new price $p(t)$. Once a transaction has been recorded, two orders are placed at the extreme positions on the grid: $p_b^i(t+1) = 0$ and $p_s^j(t+1) = \bar{p}$. As a consequence, the number of orders in the order book remains constant and equal to the number of agents.

Results. As pointed out by the authors, this process of simulation is similar the reaction-diffusion model $A + B \to \emptyset$ in Physics. In such a model, two types of particles are inserted at each side of a pipe of length $\bar{p}$ and move randomly with steps of size 1. Each time two particles collide, they're annihilated and two new particles are inserted. The analogy is summarized in table 1. Following this analogy, it thus can be showed that the variation $\Delta p(t)$ of the price $p(t)$ verifies :

$$\Delta p(t) \sim t^{1/4} (ln(\frac{t}{t_0}))^{1/2}. \tag{21}$$

Thus, at long time scales, the series of price increments simulated in this model exhibit a Hurst exponent $H = 1/4$. As for the stylized fact $H \approx 0.7$, this sub-diffusive behavior appears to be a step in the wrong direction compared to the random walk $H = 1/2$. Moreover, (118) points out that no fat tails are observed in the distribution of the returns of the model, but rather fits the empirical distribution with an exponential decay. Many more drawbacks of the model could be mentioned, the main one being that "moving" orders is highly unrealistic as for modeling an order book.

Table 1: Analogy between the $A + B \to \emptyset$ reaction model and the order book in Bak,Paczuski and Shubik (110).

Physics	(110)
Particles	Orders
Finite Pipe	Order book
Collision	Transaction

However, we feel that such a model is very interesting because of its simplicity and its representation of an order-driven market. The authors have attempted to build many extension of this simple framework, in order to reproduce "stylized facts", adding fundamental traders, strategies, trends, etc. These extensions are not of interest for us in this review: as in (108), they make the model highly complex and prevent any understanding of the mechanisms at stake.

1.3.2. Introducing market orders: Maslov (2000)

Presentation of the model. In (111), the author keeps the zero-intelligence structure of the (110) model but adds more realistic features in the order placement and evolution of the market. First, limit orders are submitted and stored in the model, without moving. Second, limit orders are submitted around the best quotes. Third, market orders are submitted to trigger transactions.

More precisely, at each time step, a trader is chosen to perform an action. In contrast to previous models, the number of traders is not fixed, but one trader enters the market at each time step. This trader can either submit a limit order with probability q_l or submit a market order with probability $1 - q_l$. Once this choice is made, the order is a buy or sell order with equal probability. All orders have a one unit volume.

As usual, we denote $p(t)$ the current price. In case the submitted order at time step $t+1$ is a limit ask (resp. bid) order, it is placed in the book at price $p(t) + \Delta$ (resp. $p(t) - \Delta$), Δ being a random variable uniformly distributed in $]0; \Delta^M = 4]$. In case the submitted order at time step $t+1$ is a market order, one order at the opposite best quote is removed and the price $p(t+1)$ is recorded.

In order to prevent the number of orders in the order book from large increase, two mechanisms are proposed by the author: either keeping a fixed maximum number of orders (new limit orders are the discarded), or removing them after a fixed lifetime if they haven't been executed.

Results. Let us first note that (112) propose an analytical study of the model thanks to a mean-field approximation (See below section 1.5).

As for numerical simulations, results show that this model exhibits non-gaussian heavy-tailed distributions. For a time scale $\delta t = 1$, the author fit the tails distribution with a power law with exponent 3.0, i.e. reasonable compared to empirical value. However, the Hurst exponent of the price series is still $H = 1/4$ with this model.

This model brings very interesting innovations in order book simulation: order book with (fixed) limit orders, market orders, necessity to cancel orders waiting too long in the order book. These features are of prime importance and open a new trend in order book modelling.

1.3.3. The order book as a deposition-evaporation process: Challet and Stinchcombe (2001)

Presentation of the model. (68) continue the work of (110) and (111) with expliciting the analogy between dynamics of an order book and an infinite one dimensional grid where particles of two types (ask and bid) are subject ot three types of events: *deposition* (limit orders), *annihilation* (market orders) and *evaporation* (cancellation). Note that annihilation occurs when a particle is deposited on a site occupied by a particle of another type. The analogy is summarized in table 2.

Therefore, the model goes as follows. At each time step, a bid (resp. ask) order is deposited with probability λ at a price $n(t)$ drawn according to a Gaussian distribution centered on the best ask $a(t)$ (resp. best bid $b(t)$) and with variance depending linearly on the spread $s(t) = a(t) - b(t)$: $\sigma(t) = Ks(t) + C$. If $n(t) > a(t)$ (resp. $n(t) < b(t)$), then it is a market order: annihilation takes place and the price is recorded. Otherwise, it is a limit order and it is stored in the book. Finally, each limit order stored in the book has a probability δ to be canceled (evaporation).

Table 2: Analogy between the deposition-evaporation process and the order book in (68).

Physics	(68)
Particles	Orders
Infinite lattice	Order book
Deposition	Limit orders submission
Evaporation	Limit orders cancelation
Annihilation	Transaction

Results. The series of price returns simulated with this model exhibit a Hurst exponent $H = 1/4$ for short time scales, and that tends to $H = 1/2$ for larger time scales. This behavior might be the consequence of the random evaporation process (which was not modeled in (111), where $H = 1/4$ for large time scales). Although some modifications of the process (more than one order per time step) seem to shorten the sub-diffusive region, it il clear that no over-diffusive behavior is observed.

1.4. Empirical zero-intelligence models

The three models presented in the previous section 1.3 have successively isolated essential mechanisms that are to be used when simulating a "realistic" market: one order is the smallest entity of the modelisation; the submission of one order is the time dimension (i.e. we use event time, not an exogenous time defined by market clearing and "tatonnement" on exogenous supply and demand functions), see section 1.2; submission of market orders (as such in Maslov (111), as "crossing limit orders" in Challet and Stinchcombe (68)) and cancellation of orders are taken into account.

On the one hand, one may try to describe these mechanisms using a small number of parameters, using Poisson process with constant rates for order flows, constant volumes, etc. This might lead to some analytically tractable models, as will be described in section 1.5. On the other hand, one may try to fit more complex empirical distributions to market data without analytical concern. (81) propose as of today the most complete empirical zero-intelligence agent-based model fitted on market data.

1.4.1. Calibrating a zero-intelligence model on the market: Mike and Farmer (2008)

Presentation of the model. (81) is the first model to propose an advanced calibration on the market data as for order placement and cancellation methods. As for volume and time of arrivals, assumptions of previous models still hold: all orders have the same volume, discrete event time is used for simulation, i.e. one order is submitted per time step.

As in (68), there is no distinction between market and limit orders. At each time step, one trading order is simulated: an ask (resp. bid) trading order is randomly placed at $n(t) = a(t) + \delta a$ (resp. $n(t) = b(t) + \delta b$) according to a Student distribution with scale and degrees of freedom calibrated on market data. If an ask (resp. bid) order satisfies $\delta a < -s(t) = b(t) - a(t)$ (resp.$\delta b > s(t) = a(t) - b(t)$), then it is a buy (resp. sell) market order and a transaction occurs at price $a(t)$ (resp. $b(t)$.

During a time step, several cancellations of orders may occur. The authors propose an empirical distribution for cancellation based on three components for a given order :

- the position in the order book, measured as the ratio $y(t) = \frac{\Delta(t)}{\Delta(0)}$ where $\Delta(t)$ is the distance of the order from the opposite best quote at time t,
- the order book imbalance, measured by the indicator $N_{imb}(t) = \frac{N_a(t)}{N_a(t)+N_b(t)}$ (resp. $N_{imb}(t) = \frac{N_b(t)}{N_a(t)+N_b(t)}$) for ask (resp. bid) orders, where $N_a(t)$ and $N_b(t)$ are the number of orders at ask and bid in the book at time t,
- the total number $N(t) = N_a(t) + N_b(t)$ of orders in the book.

Their empirical study leads them to assume that the cancellation probability has an exponential dependance on $y(t)$, a linear one in N_{imb} and finally decreases approximately as $1/N_t(t)$ as for the total number of orders. Thus, the probability $P(C|y(t), N_{imb}(t), N_t(t))$ to cancel an ask order at time t is formally written :

$$P(C|y(t), N_{imb}(t), N_t(t)) = A(1 - e^{-y(t)})(N_{imb}(t) + B)\frac{1}{N_t(t)}, \quad (22)$$

where the constants A and B are to be fitted on market data.

Finally, the author mimick the observed long memory of order signs by simulating a fractional Brownian motion. The autocovariance function $\Gamma(t)$ of the increments of such a process exhibits a slow decay :

$$\Gamma(k) \sim H(2H - 1)t^{2H-2} \quad (23)$$

and it is therefore easy to reproduce exponent β of the decay of the empirical autocorrelation function of order signs observed on the market with $H = 1 - \beta/2$.

Results. Let us first note that when dealing with empirical parameters in the simulations, the stability of the order book is far from ensured. Thus, the author require that there is at least two orders in each side of the book.

This being said, the results of this empirical model are quite satisfying as for return and spread distribution. The distribution of returns exhibit fat tails which are in agreement with empirical data. The spread distribution is also very well reproduced, both quantitatively and qualitatively. As their empirical model has been built on the data of only one stock, the authors test their model on 24 other data sets of stocks on the same market and find for half of them a good agreement between empirical and simulated properties. However, the bad results of the other half suggest that such a model is still far from being "universal".

We feel that the main drawback of the model is the way order signs are simulated. As noted by the authors, using an exogenous fractional Brownian motion leads to correlated price returns, which is in contradiction with empirical stylized facts. We also find that at long time scales it leads to an dramatic increase of volatility. As we have seen in section 1.2.2, the correlation of trade signs can be at least partly seen as an artifact of exection strategies. Therefore this element is one of the numerous that should be taken into account when "programming" the agents of the model. In order to do so, we have to leave the "zero-intelligence" world and see how heterogeneous agent based modeling might help to reproduce non-trivial behaviors. Prior to this development below in 1.6, we briefly review some analytical works on the "zero-intelligence" models.

1.5. Attempts at analytical treatment of zero-intelligence models

In this section are presented some analytical results obtained on zero-intelligence models where processes are kept sufficiently simple so that a mean-field approximation may be derived ((112)) or probabilities conditionaly to the state of the order book may be computed ((121)).

1.5.1. Mean-field theory: Slanina (2001)

(112) proposes an analytical treatment of the model introduced by (111) and reviewed above. Let us briefly described the formalism used. The main hypothesis is the following: on each side of the current price level, the density of limit orders is uniform and constant (and ρ_+ on the ask side, ρ_- on the bid side). In that sense, this is a "mean-field" approximation since the individual position of a limit order is not taken into account. Assuming we are in a stable state, the arrival of a market order of size s on the ask (resp. bid) side will make the price change by $x_+ = s/\rho_+$ (resp. $x_- = s/\rho_-$). It is then observed that the transformations of the vector $X = (x_+, x_-)$ occuring at each event (new limit order, new buy market order, new sell market order) are linear transformation that can easily and explicitly be written. Therefore, an equation satisfied by the probability distribution of X can be obtained. Finally, assuming further simplifications (such as $\rho_+ = \rho_-$), one can solve this equation for a tail exponent and find that $P(x) \equiv x^{-2}$ for large x.

This analytical result is slightly different from the one obtained by simulation in (111). However, the numerous approximations make the comparison difficult. The main point here is that some sort of mean-field approximation is natural if we assume the existence of a stationary state of the order book, and thus may help handling order book models.

Note that (122) also propose some sort of mean-field approximation for zero-intelligence models. In a similar model (but including cancellation process), mean field theory and dimensional analysis produces interesting results. For example, it is easy to see that the book depth (i.e. number of orders) $N_e(p)$ at a price p far away from the best quotes is given by $N_e(p) = \lambda/\delta$, where λ is the rate of arrival of limit orders per unit of time and per unit of price, and δ the probability for an order to be cancelled per unit of time. Indeed, far from the best quotes no market orders occurs, so that if a steady-state exists, the number of limit orders par time step λ must be balanced by the number of cancellation $\delta N_e(p)$ per unit of time, hence the result.

1.5.2. Computing probabilities conditionaly on the state of the order book: Cont, Stoikov and Talreja (2008)

(121) is an original attempt at analytical treatments of limit order books. In their model, the price is contrained to be on a grid $\{1, \ldots, N\}$. The state of the order book can then be described by a vector

$X(t) = (X_1(t), \ldots, X_N(t))$ where $|X_i(t)|$ is the quantity offered in the order book at price i. Conventionaly, $X_i(t), i = 1, \ldots, N$ is positive on the ask side and negative on the bid side. As usual, limit orders arrive at level i at a constant rate λ_i, and market orders arrive at a constant rate μ. Finally, at level i, each order can be cancelled at a rate θ_i.

Using this setting, (121) show that each event (limit order, market order, cancellation) transforms the vector X in a simple linear way. Therefore, it is shown that under reasonable conditions, X is an ergodic Markov chain, and thuis admits a stationary state. The original idea is then to use this formalism to compute conditional probabilities on the processes. More precisely, it is shown that using Laplace transform, one may explicitly compute the probability of the an increase of the mid price conditionaly on the current state of the order book.

This original contribution could allow explicit evaluation of strategies and open new perspectives in high-frequency trading. However, it is based on a simple model that does not reproduce empirical observations such as volatility clustering. Complex models trying to include market interactions will not fit into these analytical frameworks. We review some of these models in the next section.

1.6. Towards non-trivial behaviors: modeling market interactions

In all the models we have reviewed until now, flows of orders are treated as independant processes. Under some (strong) modeling constraints, we can see the order book as a Markov chain and look for analytical results ((121)). In any case, even if the process is empirically detailed and not trivial ((81)), we work with the assumption that orders are independant and identically distributed. This very strong (and false) hypothesis is similar to the “representative agent” hypothesis in Economics: orders being successively and independantly submitted, we may not expect anything but regular behaviors. Following the work of economists such as Kirman ((123; 124; 125)), one has to translate the heterogenous property of the markets into the agent-based models. Agents are not identical, and not independant.

In this section we present some toy models implementing mechanisms that aim at bringing heterogeneity: herding behavior on markets in (106), heterogenous agents in (108), dynamic order submission in (113), threshold behaviour (114).

1.6.1. Herding behaviour: Cont-Bouchaud (2000)

Presentation of the model. The model presented in (106) considers a market with N agents trading a given stock with price $p(t)$. At each time step, agents choose to buy or sell one unit of stock, i.e. their demand is $\phi_i(t) = \pm 1, i = 1, \ldots, N$ with probability a or are idle with probability $1 - 2a$. The price change is assumed to be linearly linked with the excess demand $D(t) = \sum_{i=1}^{N} \phi_i(t)$ with a factor λ measuring the liquidity of the market :

$$p(t+1) = p(t) + \frac{1}{\lambda} \sum_{i=1}^{N} \phi_i(t). \tag{24}$$

λ can also be interpreted as a market depth, i.e. the excess demand needed to move the price by one unit.

In order to evaluate the distribution of stock returns from Eq.(24), we need to know the joint distribution of the individual demands $(\phi_i(t))_{1 \le i \le N}$. As pointed out by the authors, if the distribution of the demand ϕ_i is independent and identically distributed with finite variance, then the Central Limit Theorem stands and the distribution of the price variation $\Delta p(t) = p(t+1) - p(t)$ will converge to a Gaussian distribution as N goes to infinity.

The idea of the model is to model the diffusion of the information among traders by randomly linking their demand through clusters. At each time step, agents i and j can be linked with probability $p_{ij} = p = \frac{c}{N}$, c being a parameter measuring the degree of clustering among agents. Therefore, an agent is linked to an average number of $(N-1)p$ other traders. Once clusters are determined, the demand are forced to be identical among all members of a given cluster. Denoting $n_c(t)$ the number of cluster at a given time step t, W_k the size of the k-th cluster, $k = 1, \ldots, n_c(t)$ and $\phi_k = \pm 1$ its investement decision, the price variation is then straightforwardly written :

$$\Delta p(t) = \frac{1}{\lambda} \sum_{k=1}^{n_c(t)} W_k \phi_k. \tag{25}$$

Results. This modeling is a direct application to the field of finance of the random graph framework as studied in (126). (127) previously suggested it in economics. Using these previous theoretical works, and assuming that the size of a cluster W_k and the decision taken by its members $\phi_k(t)$ are independent, the author are able to show that the distribution of the price variation at time t is the sum of $n_c(t)$ independent identically distributed random variables with heavy-tailed distributions :

$$\Delta p(t) = \frac{1}{\lambda} \sum_{k=1}^{n_c(t)} X_k, \tag{26}$$

where the density $f(x)$ of $X_k = W_k \phi_k$ is decaying as :

$$f(x) \sim_{|x| \to \infty} \frac{A}{|x|^{5/2}} e^{\frac{-(c-1)|x|}{W_0}}. \tag{27}$$

Thus, this simple toy model exhibits fat tails in the distribution of prices variations, with a decay reasonably close to empirical data. Therefore, (106) show that taking into account a naive mechanism of communication between agents (herding behavior) is able to drive the model out of the Gaussian convergence and produce non-trivial shapes of distributions of price returns.

1.6.2. Heterogenous agents: Lux and Marchesi (2000)

Presentation of the model. (108) propose model very much in line with agent-based models in behavioral finance, but where trading rules are kept simple enough so that they can be identified with a presumably realistic behavior of agents. This model considers a market with N agents that can be part of two distinct groups of traders: n_f traders are "fundamentalists", who share an exogenous idea p_f of the value of the current price p; and n_c traders are "chartists", who make assumptions on the price evolution based on the observed trend (mobile average). The total number of agents is constant, so that $n_f + n_c = N$ at any time.

At each time step, the price can be moved up or down with a fixed jump size of ± 0.01 (a tick). The probability to go up or down is directly linked to the excess demand ED through a coefficient β. The demand of each group of agents is determined as follows :

- Each fundamentalist trades a volume V_f accordingly (through a coefficient γ) to the deviation of the current price p from the perceived fundamental value p_f: $V_f = \gamma(p_f - p)$.
- Each chartist trades a constant volume V_c. Denoting n_+ the number of optimistic (buyer) chartists and n_- the number of pessimistic (seller) chartists, the excess demand by the whole group of chartists is written $(n_+ - n_-)V_c$.

Finally, assuming that there exists some noise traders on the market with random demand μ, the global excess demand is written :

$$ED = (n_+ - n_-)V_c + n_f \gamma(p_f - p) + \mu. \tag{28}$$

The probability that the price goes up (resp. down) is then defined to be the positive (resp. negative) part of βED.

As in (106), the authors expect non-trivial features of the price series to results from herding behavior and transitions between groups of traders. Refering to Kirman's work as well, a mimicking behavior among chartists is thus proposed. The n_c chartists can change their view on the market (optimistic, pessimistic), their decision being based on a clustering process modelized buy an opinion index $x = \frac{n_+ - n_-}{n_c}$ representing the weight of the majority. The probabilities π_+ and π_- to switch from one group to another are formally written :

$$\pi_\pm = v \frac{n_c}{N} e^{\pm U} \qquad U = \alpha_1 x + \alpha_2 p / v, \tag{29}$$

where v is a constant, and α_1 and α_2 reflect respectively the weight of the majority's opinion and the weight of the observed price in the chartists' decision.

Finally, transitions between fundamentalists and chartists are also allowed, decided by comparison of expected returns. We do not reproduce here the explicit probabilities of transitions of the model (see (108) for details), as this would need to introduce more parameters that are not useful for the purpose of our review. They are formally similar to $\pi_\pm$ with of course a different function U.

Results. The authors present both theoretical and empirical results. First, they are able to derive approximate differential equation in continuous time governing mean values of the model, thus derive stationary solution for these mean values. Second, they provide simulation results and show that the distribution of returns generated by their model have excess kurtosis. Using a Hill estimator, they fit a power law to the fat tails of the distribution and observe exponents grossly ranging from 1.9 to 4.6. The authors also check hints for volatility clustering: absolute returns and squared returns exhibit a slow decay of autocorrelation, while raw returns do not.

So it appears that such a model can grossly fit some "stylized facts". However, the number of parameters involved, as well as the complicated rules of transition between agents makes clear identification of sources of phenomenas and calibration to market data difficult, if not impossible.

1.6.3. Dynamic placement of limit orders: Preis (2007)

Presentation of the model. (113) studies a model similar to (68) and (122) and adds a link between order placement and market trend as an attempt to obtain an over-diffusive behavior in the price series. Doing so, it is in line with some extensions of the previous models we've already mentioned.

The simulation process goes similarly to the one (68) with deposition, annihiliation and evaporation events. Price $p(t)$ is stored as an integer (number of ticks). Every oder has a volume of one unit of stock. As in previous presentations, $a(t)$ (resp. $b(t)$) is the best ask (resp. best bid) at time t.

First, at each time step, N liquidity providers submit limit orders at rate λ, i.e. on average λN limit orders are inserted in the order book per time step. The price of a new limit ask (resp. bid) orders is $n(t) = b(t) + \Delta$ (resp. $n(t) = a(t) - \Delta$) where Δ is a random variable with an exponential distribution with parameter α. Second, N liquidity takers submit market orders at rate μ, with an equal probability $q = 1/2$ to buy or sell. Finally, each order staying in the order book is canceled with probability δ.

Note that this formalism allows for a direct computation of the average number order of orders N_e in the order book if an equilibrium exists :

$$N_e = N\lambda \left(\frac{1}{\delta} - 1 \right) - \frac{\mu}{\delta}. \tag{30}$$

Results. As expected, this zero-intelligence framework does not produce fat tails in the distribution of (log-) returns nor an overdiffusive Hurst exponent. As in the previous similar models, the authors observe a underdiffusive behavior for short time scales, and a Hurst exponent $H = 1/2$ for larger time scales.

However, an interesting result appears when extending the basic model with random demand perturbation and dynamic limit order placement depth. The author propose that the probability q for a buy/sell market order evolves at each time step according to a feedback random walk:

$$q(t+1) = \begin{cases} q(t) - \Delta q & \text{with probability } q(t), \\ q(t) + \Delta q & \text{with probability } 1 - q(t), \end{cases} \tag{31}$$

where δq is a constant (feedback random walk).

This stochastic pertubation alone is not sufficient to produce non-trivial Hurst exponent. Thus, it is then combined with a dynamic placement order depth, i.e.the author propose that the parameter α vary during the simulation as follows:

$$\alpha(t) = \alpha(0) \left(1 + C|q(t) - 1/2|\right) \tag{32}$$

This equation formalizes the assumption that in a market with an observed trend in prices, liquidity providers expect larger variations and submit limit orders at a wider depth in the order book. Although the assumption behind such a mechanism may not be empirically confirmed (e.g. the assumed symetry) and should be further discussed, it is interesting enough that it directly provides fat tails in the log-return distributions and an over-diffusive Hurst exponent $H \approx 0.6 - 0.7$ for medium time-scales.

1.6.4. Volatility clustering with a threshold behavior: Cont (2007)

Let us conclude this part with a model focusing primarily on reproducing the stylized fact of volatility clustering, while most of the previous models we have reviewed were mostly focused on fat tails of log returns. (114) proposes a model with a rather simple mechanism to create volatility clustering. The idea is that volatility clustering characterizes several regimes of volatility (quite periods vs bursts of activity). Instead of implementing an exogenous change of regime, the author defines the following trading rules.

Presentation of the model. At each period, an agent $i \in \{1, \ldots, N\}$ can issue a buy or a sell order: $\phi_i(t) = \pm 1$. Information is represented by a series of i.i.d Gaussian random variables. (ϵ_t). This public information ϵ_t is a forecast for the value r_{t+1} of the return of the stock. Each agent $i \in \{1, \ldots, N\}$ decides whether to follow this information according to a threshold $\theta_i > 0$ representing its sensibility to the public information:

$$\phi_i(t) = \begin{cases} 1 & \text{if} \quad \epsilon_i(t) > \theta_i(t) \\ 0 & \text{if} \quad |\epsilon_i(t)| < \theta_i(t) \\ -1 & \text{if} \quad \epsilon_i(t) < -\theta_i(t) \end{cases} \tag{33}$$

The, once every choice is made, the price evolves according to the excess demand $D(t)D(t) = \sum_{i=1}^{N} \phi_i(t)$, in a way similar to (106).

A the end of each time step t, threshold are asynchronously updated. Each aget has a probability s to update its theshold $\theta_i(t)$. In such a case, the new threshold $\theta_i(t+1)$ is defined to be the absolute value $|r_t|$of the return just observed. In short:

$$\theta_i(t+1) = \mathbf{1}_{\{u_i(t)<s\}}|r_t| + \mathbf{1}_{\{u_i(t)>s\}}\theta_i(t). \tag{34}$$

Results. The author shows that the time series simulated with such a model do exhibit some realistics facts on volatility. In particular, long range correlations of absolute returns is observed. The strength of this model is that it directly links the state of the market with the decision of the trader. Such a feedback mechanism is essential in order to obtain non trivial characteristics. The main drawback of the model presented in (114) is that its simplicity prevents it to be fully calibrated on empirical data. However, this simple mechanism could be used in a more elaborate agent-based model in order to reproduce the empirical evidence of volatility clustering.

2. Agent-based modeling for wealth distributions: Kinetic theory models

2.1. General remarks

The distributions of money, wealth or income, i.e., how such quantities are shared among the population of a given country and among different countries, is a topic which has been studied by economists for a long time. The relevance of the topic to us is twofold: From the point of view of the science of Complex Systems, wealth distributions represent a unique example of a quantitative outcome of a collective behavior which can be directly compared with the predictions of theoretical models and numerical experiments. Also, there is a basic interest in wealth distributions from the social point of view, in particular in their degree of (in)equality. To this aim, the Gini coefficient (or the Gini index, if expressed as a percentage), developed by the Italian statistician Corrado Gini, represents a concept commonly employed to measure inequality of wealth distributions or, more in general, how uneven a given distribution is. For a cumulative distribution function $F(y)$, that is piecewise differentiable, has a finite mean μ, and is zero for $y < 0$, the Gini coefficient is defined as

$$G = 1 - \frac{1}{\mu}\int_0^\infty dy\,(1 - F(y))^2 = \frac{1}{\mu}\int_0^\infty dy\,F(y)(1 - F(y))\,. \tag{35}$$

It can also be interpreted statistically as half the relative mean difference. Thus the Gini coefficient is a number between 0 and 1, where 0 corresponds with perfect equality (where everyone has the same income) and 1 corresponds with perfect inequality (where one person has all the income, and everyone else has zero income). Some values of G for some countries are listed in Table 3.

Table 3: Gini indices of some countries (from *Human Development Indicators of the United Nations Human Development Report* 2004, pp.50-53, available at http://hdr.undp.org/en/reports/global/2004/)

Denmark	24.7
Japan	24.9
Sweden	25.0
Norway	25.8
Germany	28.3
India	32.5
France	32.7
Australia	35.2
UK	36.0
USA	40.8
Hong Kong	43.4
China	44.7
Russia	45.6
Mexico	54.6
Chile	57.1
Brazil	59.1
South Africa	59.3
Botswana	63.0
Namibia	70.7

Let us start by considering the basic economic quantities: money, wealth and income.

2.2. Money, wealth and income

A common definition of *money* suggests that money is the “Commodity accepted by general consent as medium of economics exchange” (29). In fact, money circulates from one economic agent (which can represent an individual, firm, country, etc.) to another, thus facilitating trade. It is “something which all other goods or services are traded for” (for details see Ref. (128)). Throughout history various commodities have been used as money, for these cases termed as “commodity money”, which include e.g. rare seashells or beads, and cattle (such as cow in India). Recently, "commodity money" has been replaced by other forms referred to as "fiat money", which have gradually become the most common ones, such as metal coins and paper notes. Nowadays, other forms of money, such as electronic money, have become the most frequent form used to carry out transactions. In any case the most relevant points about money employed are its basic functions, which according to standard economic theory are

- to serve as a medium of exchange, which is universally accepted in trade for goods and services;
- to act as a measure of value, making possible the determination of the prices and the calculation of costs, or profit and loss;
- to serve as a standard of deferred payments, i.e., a tool for the payment of debt or the unit in which loans are made and future transactions are fixed;
- to serve as a means of storing wealth not immediately required for use.

A related feature relevant for the present investigation is that money is the medium in which prices or values of all commodities as well as costs, profits, and transactions can be determined or expressed. *Wealth* is usually understood as things that have economic utility (monetary value or value of exchange), or material goods or property; it also represents the abundance of objects of value (or riches) and the state of having accumulated these objects; for our purpose, it is important to bear in mind that wealth can be measured in terms of money. Also *income*, defined as "The amount of money or its equivalent received during a period of time in exchange for labor or services, from the sale of goods or property, or as profit from financial investments" (see www.answers.com), is also a quantity which can be measured in terms of money (per unit time).

2.3. Modeling wealth distributions

It was first observed by Pareto (35) that in an economy the higher end of the distribution of income $f(w)$ follows a power-law,

$$f(x) \sim x^{-1-\alpha} \,, \tag{36}$$

with α, now known as the Pareto exponent, estimated by Pareto to be $\alpha \approx 3/2$. For the last hundred years the value of $\alpha \sim 3/2$ seems to have changed little in time and across the various capitalist economies.

In 1931, Gibrat (129) clarified that Pareto's law is valid only for the high income range, whereas for the middle income range he suggested that the income distribution is described by a log-normal probability density

$$f(x) \sim \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\log^2(x/x_0)}{2\sigma^2}\right\} , \tag{37}$$

where $\log(x_0) = \langle \log(x) \rangle$ is the mean value of the logarithmic variable and $\sigma^2 = \langle [\log(x) - \log(x_0)]^2 \rangle$ the corresponding variance. The factor $\beta = 1/\sqrt{2\sigma^2}$, also know an as Gibrat index, measures the equity of the distribution.

More recent empirical studies on income distribution have been carried out by physicists, e.g. those by Dragulescu and Yakovenko for UK and US (130; 131), and Fujiwara et al. for Japan (132), for an overview see Refs. (133). The distributions obtained have been shown to follow either the Gibbs or power-law types, depending on the range of wealth, as shown in Fig. 31.

One of the current challenges is to write down the "microscopic equation" which governs the dynamics of the evolution of wealth distributions, possibly predicting the observed shape of wealth distributions, including the exponential law at intermediate values of wealth as well as the century-old Pareto law. To this aim, several studies have been made to investigate the characteristics of the real income distribution and provide theoretical models or explanations, for a review see Refs. (134; 135; 133).

The model of Gibrat (129) mentioned above and other models formulated in terms of a Langevin equation for a single wealth variable, subjected to multiplicative noise (19; 136; 137; 138), can lead to equilibrium wealth distributions with a power law tail, since they converge toward a log-normal distribution. However,

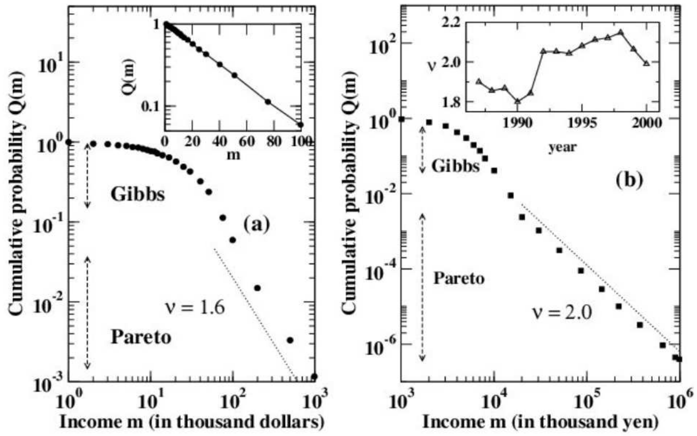

Figure 31: Income distributions in the US (left)and Japan (right). Reproduced and adapted from `arXiv:cond-mat/0302147`.

the fit of real wealth distributions does not turn out to be as good as that obtained using e.g. a Γ- or a β-distribution, in particular due to too large asymptotic variances (139).

Other models describe the wealth dynamics as a wealth flow due to exchanges between (pairs of) basic units. In this respect, such models are basically different from the class of models formulated in terms of a Langevin equation for a single wealth variable. For example, Levy and Solomon studied the generalized Lotka-Volterra equations in relation to power-law wealth distribution (140; 141). Ispolatov et al. (142) studied random exchange models of wealth distributions. Other models describing wealth exchange have been formulated using matrix theory (143), the master equation (144; 145; 146), the Boltzmann equation approach (145; 147; 148; 149; 150; 151; 152), or Markov chains (153; 154; 155).

In the two following sections we consider in greater detail a class of models usually referred to as *kinetic wealth exchange models* (KWEM), formulated through finite time difference stochastic equations (139; 156; 157; 158; 145; 159; 160; 161; 162; 163; 164; 165; 166). From the studies carried out using wealth-exchange models, it emerges that it is possible to use them to predict power law distributions. However, a general understanding of the dependence of the shape of the equilibrium distribution on the underlying mechanisms and parameters is still missing.

2.4. Homogeneous kinetic wealth exchange models

Here and in the next section we consider KWEMs, which are statistical models of closed economy, in which N agents exchange a quantity x, defined as wealth in some models and as money in others. As explained above, money can be interpreted as the unity of measure for all the goods that constitute the agents' wealth, so that it is possible, in order to avoid confusion, to use only the term wealth.

In these models the states of agents are defined in terms of the wealth variables $\{x_n\},\ n = 1, 2, \ldots, N$. The evolution of the system is carried out according to a trading rule between agents which, for obtaining the final equilibrium distribution, can be interpreted as the actual time evolution of the agent states as well as a Monte Carlo optimization. The algorhythm is based on a simple update rule performed at each time step t, when two agents i and j are extracted randomly and an amount of wealth Δx is exchanged,

$$\begin{aligned} x_i' &= x_i - \Delta x \,, \\ x_j' &= x_j + \Delta x \,. \end{aligned} \tag{38}$$

Notice that the quantity x is conserved during single transactions, $x_i' + x_j' = x_i + x_j$, where $x_i = x_i(t)$ and $x_j = x_j(t)$ are the agent wealths before, whereas $x_i' = x_i(t+1)$ and $x_j' = x_j(t+1)$ are the final ones after the transaction. Several rules have been studied for the model defined by Eqs. (38). We refer to the Δx there defined in illustrating the various models.

2.4.1. Exchange models without saving

In a first version of the model, considered in early works by Bennati (167; 168; 169) as well as by Dragulescu and Yakovenko (145), the money difference Δx in Eqs. (38) is assumed to have a constant value,

$$\Delta x = \Delta x_0 \ . \qquad (39)$$

This rule, together with the constraint that transactions can take place only if $x'_i > 0$ and $x'_j > 0$, provides an equilibrium exponential distribution, see the curve for $\lambda = 0$ in Fig. 32.

A unidirectional model had previously been considered by Angle (139), in which Δx is set as a random fraction of the wealth of one of the two agents,

$$\Delta x = \epsilon x_i \ \text{ or } \ \Delta x = \epsilon x_j \ . \qquad (40)$$

Here ϵ is a random number uniformly distributed between 0 and 1 and the flow direction (from i to j or vice versa) depends on the wealths of the two agents, see Ref. (139).

Various other trading rules were studied by Dragulescu and Yakovenko (145), choosing Δx as a random fraction of the average money between the two agents, $\Delta x = \epsilon(x_i + x_j)/2$, or of the average money of the whole system, $\Delta x = \epsilon\langle x\rangle$.

In the *reshuffling* model, which can be obtained as a particular case from the model with saving introduced in Ref. (158) for a saving parameter $\lambda = 0$ (see below), the wealths of the two agents are reshuffled randomly,

$$\begin{aligned} x'_i &= \epsilon(x_i + x_j)\,, \\ x'_j &= (1-\epsilon)(x_i + x_j)\,. \end{aligned} \qquad (41)$$

In this case the Δx appearing in the trading rule (38) is given by

$$\Delta x = (1-\epsilon)x_i - \epsilon x_j \ . \qquad (42)$$

All the models mentioned here, as well as some more complicated models (145), lead to a robust equilibrium Boltzmann distribution,

$$f(x) = \beta \exp(-\beta x)\,, \qquad (43)$$

with the effective temperature of the system equal to the average wealth, $\beta^{-1} = \langle x\rangle$. This result is largely independent of the details of the models, e.g. the multi-agent nature of the interaction, the initial conditions, and the random or consecutive order of extraction of the interacting agents. The Boltzmann distribution is sometimes referred to as an "unfair distribution", since it is characterized by a majority of poor agents and a few rich agents (sue to the exponential tail). The exponential distribution is characterized by a Gini coefficient of 0.5.

Note that despite the Boltzmann distribution is robust respect to the variation of some parameters, the way it depends on the details of the trading rule is subtle. Sometimes little modfication in the trading rule can also lead to drastic changes in the resulting distribution. For instance, in the model studied in Ref. (159), the equilibrium distribution can have a very different shape. In that toy model it is assumed that both the economic agents i and j invest the same amount x_{min}, which is taken as the minimum wealth between the two agents, $x_{min} = \min\{x_i, x_j\}$. The wealths after the trade are $x'_i = x_i + \Delta x$ and $x'_j = x_j - \Delta x$, where $\Delta x = (2\epsilon - 1)x_{min}$. We note that once an agent has lost all his wealth, he is unable to trade because x_{min} has become zero. Thus, a trader is effectively driven out of the market once he loses all his wealth. In this way, after a sufficient number of transactions only one trader survives in the market with the entire amount of wealth, whereas the rest of the traders have zero wealth. In this toy model, only one agent has the entire money of the market and the rest of the traders have zero money, which corresponds to a distribution with Gini coefficient equal to unity.

Now, a situation is said to be Pareto-optimal "if by reallocation you cannot make someone better off without making someone else worse off". In Pareto's own words:

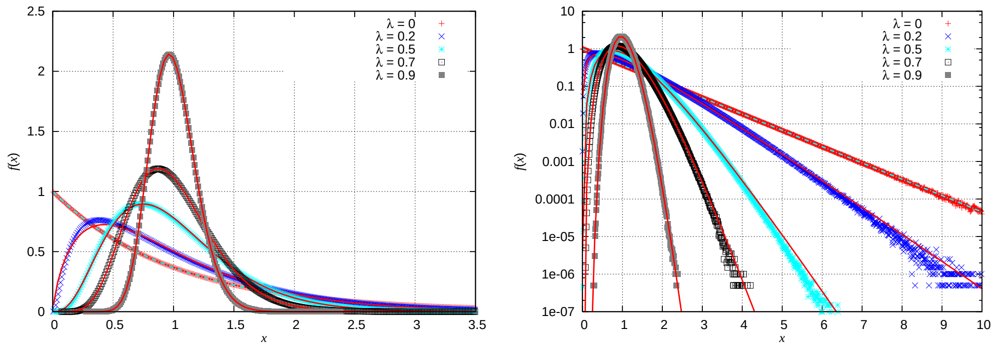

Figure 32: Probability density for wealth x. The curve for $\lambda = 0$ is the Boltzmann function $f(x) = \langle x \rangle^{-1} \exp(-x/\langle x \rangle)$ for the basic model of Sec. 2.4.1. The other curves correspond to a global saving propensity $\lambda > 0$, see Sec. 2.4.2.

> "We will say that the members of a collectivity enjoy maximum ophelimity in a certain position when it is impossible to find a way of moving from that position very slightly in such a manner that the ophelimity enjoyed by each of the individuals of that collectivity increases or decreases. That is to say, any small displacement in departing from that position necessarily has the effect of increasing the ophelimity which certain individuals enjoy, and decreasing that which others enjoy, of being agreeable to some, and disagreeable to others."
>
> — Vilfredo Pareto, Manual of Political Economy (1906), p.261.

However, as the Nobel Laureate Amartya Sen notes, *an economy can be Pareto-optimal, yet still "perfectly disgusting" by any ethical standards* (Sen, A.K., 1970, Collective Choice and Social Welfare, San Fransisco, Holden Day, also London: Oliver and Boyd (reprinted Amsterdam: North-Holland)). It is important to note that Pareto-optimality, is merely a descriptive term, a property of an "allocation", and there are no ethical propositions about the desirability of such allocations inherent within that notion. Thus, in other words there is nothing inherent in Pareto-optimality that implies the maximization of social welfare.

This simple toy model thus also produces a Pareto-optimal state (it will be impossible to raise the well-being of anyone except the *winner*, i.e., the agent with all the money, and vice versa) but the situation is economically undesirable as far as social welfare is concerned!

2.4.2. Exchange models with saving.

As a generalization and more realistic version of the basic exchange models without saving, a saving criterion can be introduced. As early as 1983, motivated by the surplus theory, John Angle introduced a unidirectional model of wealth exchange, in which only a fraction of wealth smaller than one can pass from one agent to the other, with a $\Delta x = \epsilon Z x_i$ or $(-\omega Z x_j)$, where the variable Z is determined by the agent wealths in a way depending on the version of the model (170; 139). Angle also introduced a model where a minimum constant fraction $1 - \omega$ is saved before the transaction (156) by the agent whose wealth decreases, with a trading rule defined by an exchanged wealth amount

$$\Delta x = \omega \epsilon x_i \qquad \text{or} \qquad -\omega \epsilon x_j \,, \tag{44}$$

where the direction of the transaction (from the i-th or j-th agent) is determined by the relative difference between the agent wealths.

A saving parameter $0 < \lambda < 1$ representing the fraction of wealth saved and not reshuffled, similarly to $(1 - \omega)$ of the Angle model, was introduced in the model introduced in Ref. (158). In this model (CC) wealth flows simultaneously toward and from each agent during a single transaction, the dynamics being defined by the equations

$$\begin{aligned} x'_i &= \lambda x_i + \epsilon(1-\lambda)(x_i + x_j)\,, \\ x'_j &= \lambda x_j + (1-\epsilon)(1-\lambda)(x_i + x_j)\,, \end{aligned} \tag{45}$$

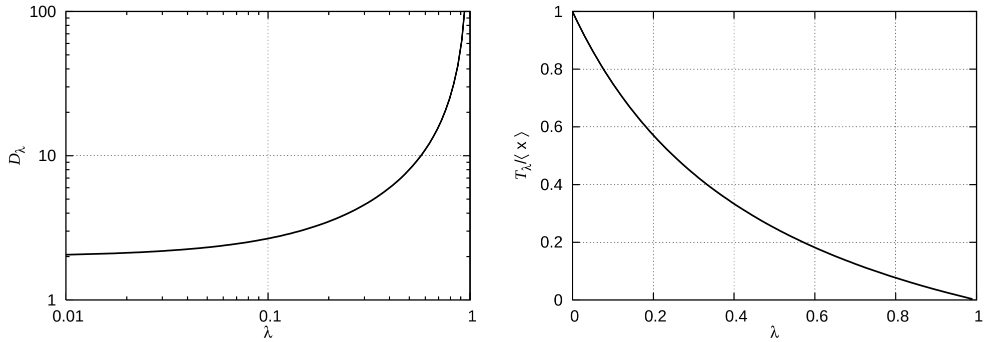

Figure 33: Effective dimension D_λ and temperature T as a function of the saving parameter λ.

Table 4: Analogy between kinetic the theory of gases and the kinetic exchange model of wealth

	Kinetic model	Economy model
variable	K (kinetic energy)	x (wealth)
units	N particles	N agents
interaction	collisions	trades
dimension	integer D	real number D_λ
temperature definition	$k_\mathrm{B}T = 2\langle K\rangle/D$	$T_\lambda = 2\langle x\rangle/D_\lambda$
reduced variable	$\xi = K/k_\mathrm{B}T$	$\xi = x/T_\lambda$
equilibrium distribution	$f(\xi) = \gamma_{D/2}(\xi)$	$f(\xi) = \gamma_{D_\lambda/2}(\xi)$

or, equivalently, by a Δx in (38) given by

$$\Delta x = (1-\lambda)[(1-\epsilon)x_i - \epsilon x_j]\,. \tag{46}$$

All these models lead to an equilibrium distribution qualitatively different from the purely exponential distribution of models without saving. In fact, there is now a mode $x_m > 0$, in agreement with real data of wealth and income distributions(131; 146; 171; 172; 173; 174). Furthermore, the limit for small x is zero, i.e. $P(x \to 0) \to 0$, see the example in Fig. 32. The functional form of such a distributions has been conjectured to be a Gamma distribution, on the base of the excellent fitting provided to numerical data (170; 139; 156; 175; 176; 177),

$$f(x) = n\langle x\rangle^{-1}\gamma_n(nx/\langle x\rangle) = \frac{1}{\Gamma(n)}\frac{n}{\langle x\rangle}\left(\frac{nx}{\langle x\rangle}\right)^{n-1}\exp\left(-\frac{nx}{\langle x\rangle}\right)\ , \tag{47}$$

$$n(\lambda) \equiv D_\lambda = 1 + \frac{3\lambda}{1-\lambda}\ . \tag{48}$$

where $\gamma_n(\xi)$ is the standard Γ-distribution. The analogy between models is more than qualitative. For instance, it has been found (177) that the only difference between the equilibrium solution above and that of the symmetrical model of Angle is in the effective dimension, which in the Angle model is $n_\mathrm{A}(\lambda) = n(\lambda)/2$, where n is given in Eq. (48).

The ubiquitous presence of Γ-functions in the solutions of kinetic models (see also below heterogeneous models) suggests a close analogy with kinetic theory of gases. In fact, interpreting $D_\lambda = 2n$ as an effective dimension, the variable x as kinetic energy, and introducing the effective temperature $\beta^{-1} \equiv T_\lambda = \langle x\rangle/2D_\lambda$ according to the equipartition theorem, Eqs. (47) and (48) define the canonical distribution $\beta\gamma_n(\beta x)$ for the kinetic energy of a gas in $D_\lambda = 2n$ dimensions, see Ref. (176) for details. The analogy is illustrated in Table 4 and the dependences of $D_\lambda = 2n$ and of $\beta^{-1} = T_\lambda$ on the saving parameter λ are shown in Fig. 33.

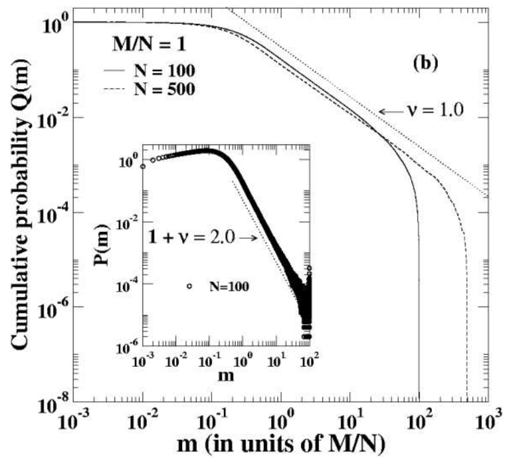

Figure 34: Results for randomly assigned saving parameters. Reproduced and adapted from `arXiv:cond-mat/0302147`.

The exponential distribution is recovered as a special case, for $n = 1$. In the limit $\lambda \to 1$, i.e. for $n \to \infty$, the distribution $f(x)$ above tends to a Dirac δ-function, as shown in Ref. (176) and qualitatively illustrated by the curves in Fig. 32. This shows that a large saving criterion leads to a final state in which economic agents tend to have similar amounts of money and, in the limit of $\lambda \to 1$, exactly the same amount $\langle x \rangle$.

The equivalence between a kinetic wealth-exchange model with saving propensity $\lambda \geq 0$ and an N-particle system in a space with dimension $D_\lambda \geq 2$ is suggested by simple considerations about the kinetics of collision processes between two molecules. In one dimension, particles undergo head-on collisions in which the whole amount of kinetic energy can be exchanged. In a larger number of dimensions the two particles will not travel in general exactly along the same line, in opposite verses, and only a fraction of the energy can be exchanged. It can be shown that during a binary elastic collision in D dimensions only a fraction $1/D$ of the total kinetic energy is exchanged on average for kinematic reasons, see Ref. (178) for details. The same $1/D$ dependence is in fact obtained inverting Eq. (48), which provides for the fraction of exchanged well $1 - \lambda = 6/(D_\lambda + 4)$.

2.5. Heterogeneous kinetic wealth exchange models

2.5.1. Formulation

The basic models considered above, giving equilibrium wealth distributions with an exponential tail, interpolate well real data at small and intermediate values of wealth (170; 139; 179; 130; 131; 146; 171). However, more realistic generalized models have been studied, in which agents are diversified by assigning different values of the saving parameter. For instance, Angle studied a model with a trading rule where diversified parameters $\{\omega_i\}$ occurr,

$$\Delta x = \omega_i \epsilon x_i \qquad \text{or} \qquad -\omega_j \epsilon x_j \,, \tag{49}$$

again with the direction of wealth flow determined by the wealths of agents i and j (156). Diversified saving parameters were independently introduced in Refs. (161; 180) by generalizing the model introduced in Ref. (158),

$$\begin{aligned} x_i' &= \lambda_i x_i + \epsilon[(1-\lambda_i)x_i + (1-\lambda_j)x_j] \,, \\ x_j' &= \lambda x_j + (1-\epsilon)[(1-\lambda_i)x_i + (1-\lambda_j)x_j] \,, \end{aligned} \tag{50}$$

corresponding to a

$$\Delta x = (1-\epsilon)(1-\lambda_i)x_i - \epsilon(1-\lambda_j)x_j \,. \tag{51}$$

The surprising result is that if the parameters $\{\lambda_i\}$ are suitably diversified, a power law appears in the equilibrium wealth distribution, see Fig. 34. In particular if the λ_i are uniformly distributed in $(0, 1)$ the wealth distribution exhibits a robust power-law tail,

$$f(x) \propto x^{-\alpha-1}\,, \tag{52}$$

with the Pareto exponent $\alpha = 1$ largely independent of the details of the λ-distribution. This result is supported by independent theoretical considerations based on different approaches, such as a mean field theory approach (181) (see below for further details) or the Boltzmann equation(162; 182; 183; 148; 184).

2.5.2. Power-law distribution as an overlap of Gamma distributions

A remarkable feature of the equilibrium wealth distribution obtained from heterogeneous models, noticed in Ref. (180), is that the individual wealth distribution $f_i(x)$ of the generic i-th agent with saving parameter λ_i has a well defined mode and exponential tail, in spite of the resulting power-law tail of the marginal distribution $f(x) = \sum_i f_i(x)$. In fact, it was found by numerical simulation that the marginal distribution $f(x)$ can be resolved as an overlap of individual Gamma distributions with λ-dependent parameters(185); furthermore, the mode and the average value of the distributions $f_i(x)$ both diverge for $\lambda \to 1$ as $\langle x(\lambda)\rangle \sim 1/(1-\lambda)$ (180; 185). This fact was justified theoretically by Mohanty (181). Consider the evolution equations (50). In the mean field approximation one can consider that each agents i has an (average) wealth $\langle x_i\rangle = y_i$ and replace the random number ϵ with its average value $\langle\epsilon\rangle = 1/2$. Indicating with y_{ij} the new wealth of agent i, due to the interaction with agent j, from Eqs. (50) one obtains

$$y_{ij} \;=\; (1/2)(1+\lambda_i)y_i + (1/2)(1-\lambda_j)y_j\,. \tag{53}$$

At equilibrium, for consistency, average over all the interaction must give back y_i,

$$y_i = \sum_j y_{ij}/N\,. \tag{54}$$

Then summing Eq. (53) over j and dividing by the number of agents N, one has

$$(1-\lambda_i)y_i = \langle(1-\lambda)y\rangle\,, \tag{55}$$

where $\langle(1-\lambda)y\rangle = \sum_j(1-\lambda_j)y_j/N$. Since the right hand side is independent of i and this relation holds for arbitrary distributions of λ_i, the solution is

$$y_i = \frac{C}{1-\lambda_i}\,, \tag{56}$$

where C is a constant. Besides proving the dependence of $y_i = \langle x_i\rangle$ on λ_i, this relation also demonstrates the existence of a power law tail in the equilibrium distribution. If in the continuous limit λ is distributed in $(0, 1)$ with a density $\phi(\lambda)$, $(0 \le \lambda < 1)$, then using (56) the (average) wealthdistribution is given

$$f(y) = \phi(\lambda)\frac{d\lambda}{dy} = \phi(1-C/x)\frac{C}{y^2}\,. \tag{57}$$

Figure 35 illustrates the phenomenon for a system of $N = 1000$ agents with random saving propensities uniformly distributed between 0 and 1. The figure confirms the importance of agents with λ close to 1 for producing a power-law probability distribution (180; 177).

However, when considering values of λ close enough to 1, the power law can break down at least for two reasons. The first one, illustrated in Fig. 35-bottom right, is that the power-law can be resolved into almost disjoint contributions representing the wealth distributions of single agents. This follows from the finite number of agents used and the fact that the distance between the average values of the distributions corresponding to two consecutive values of λ grows faster than the corresponding widths (185; 183). The second reason is due to the finite cutoff λ_{M}, always present in a numerical simulation. However, to study this

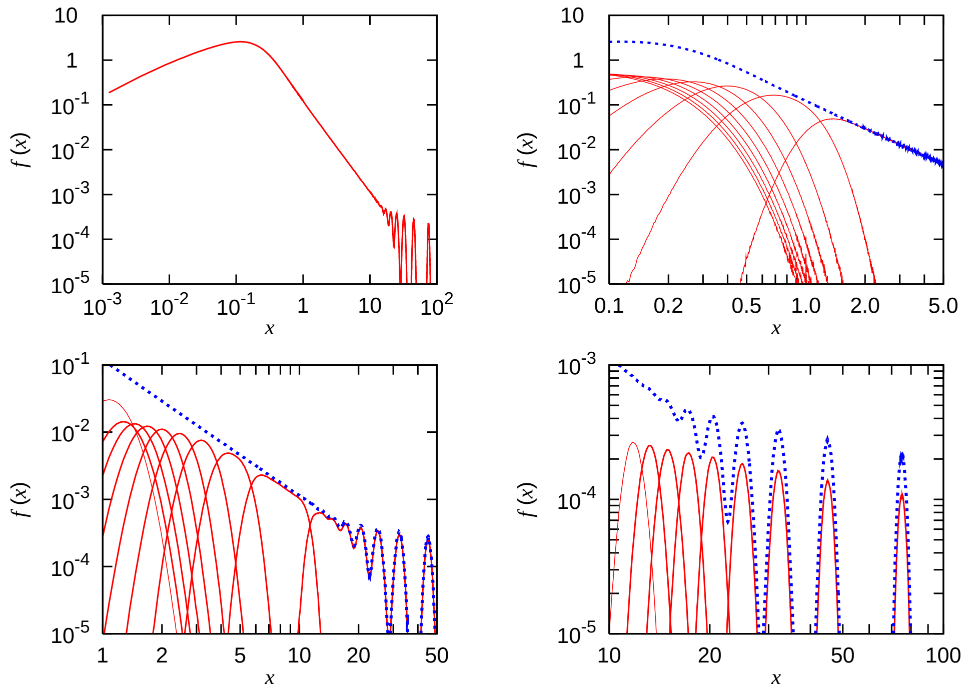

Figure 35: Wealth distribution in a system of 1000 agents with saving propensities uniformly distributed in the interval $0 < \lambda < 1$. Top left: marginal distribution. Top right: marginal distribution (dotted line) and distributions of wealth of agents with $\lambda \in (j\Delta\lambda, (j+1)\Delta\lambda)$, $\Delta\lambda = 0.1$, $j = 0, \ldots, 9$ (continuous lines). Bottom-left: the distribution of wealth of agents with $\lambda \in (0.9, 1)$ has been further resolved into contributions from subintervals $\lambda \in (0.9+j\Delta\lambda, 0.9+(j+1)\Delta\lambda)$, $\Delta\lambda = 0.01$. Bottom-right: the partial distribution of wealth of agents with $\lambda \in (0.99, 1)$ has been further resolved into those from subintervals $\lambda \in (0.99 + j\Delta\lambda, 0.99 + (j+1)\Delta\lambda)$, $\Delta\lambda = 0.001$.

effect, one has to consider a system with a number of agents large enough that it is not possible to resolve the wealth distributions of single agents for the sub-intervals of λ considered. This was done in Ref. (186) using a system with $N = 10^5$ agents with saving parameters distributed uniformly between 0 and λ_{M}. Results are shown in Fig. 36, in which curves from left to right correspond to increasing values of the cutoff λ_{M} from 0.9 to 0.9997. The transition from an exponential to a power-law tail takes place continuously as the cut-off λ_{M} is increased beyond a critical value $\lambda_{\mathrm{M}} \approx 0.9$ toward $\lambda_{\mathrm{M}} = 1$, through the enlargement of the x-interval in which the power-law is observed.

2.5.3. Relaxation process

The relaxation time scale of a heterogeneous system has been studied in Ref. (187). The system is observed to relax toward the same equilibrium wealth distribution from any given arbitrary initial distribution of wealth. If time is measured by the number of transactions n_t, the time scale is proportional to the number of agents N, i.e. defining time t as the ratio $t = n_t/N$ between the number of trades and the total number of agents N (corresponding to one Monte Carlo cycle or one sweep in molecular dynamics simulations) the dynamics and the relaxation process become independent of N. The existence of a natural time scale independent of the system size provides a foundation for using simulations of systems with finite N in order to infer properties of systems with continuous saving propensity distributions and $N \to \infty$.

Relaxation in systems with constant λ has already been studied in Ref. (158), where a systematic increase of the relaxation time with λ, and eventually a divergence for $\lambda \to 1$, was found. In fact, for $\lambda = 1$ no exchanges occurs and the system is frozen. In a system with uniformly distributed λ, the wealth distributions of each agent i with saving parameter λ_i relaxes toward different states with characteristic shapes $f_i(x)$ (185; 183; 186) with different relaxation times τ_i (187). The differences in the relaxation

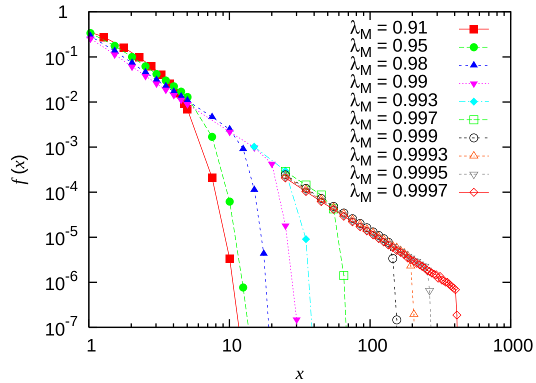

Figure 36: Wealth distribution obtained for the uniform saving propensity distributions of 10^5 agents in the interval $(0, \lambda_{\rm M})$.

process can be related to the different relative wealth exchange rates, that by direct inspection of the evolution equations appear to be proportional to $1 - \lambda_i$. Thus, in general, higher saving propensities are expected to be associated to slower relaxation processes with a rate $\propto 1/(1 - \lambda)$.

It is also possible to obtain the relaxation time distribution. If the saving parameters are distributed in $(0, 1)$ with a density $\phi(\lambda)$, it follows from probability conservation that $\tilde{f}(\bar{x})d\bar{x} = \phi(\lambda)d\lambda$, where $\bar{x} \equiv \langle x \rangle_\lambda$ and $\tilde{f}(\bar{x})$ the corresponding density of average wealth values. In the case of uniformly distributed saving propensities, one obtains

$$\tilde{f}(\bar{x}) = \phi(\lambda)\frac{d\lambda(\bar{x})}{d\bar{x}} = \phi\left(1 - \frac{k}{\bar{x}}\right)\frac{k}{\bar{x}^2}\,, \tag{58}$$

showing that a uniform saving propensity distribution leads to a power law $\tilde{f}(\bar{x}) \sim 1/\bar{x}^2$ in the (average) wealth distribution. In a similar way it is possible to obtain the associated distribution of relaxation times $\psi(\tau)$ for the global relaxation process from the relation $\tau_i \propto 1/(1 - \lambda_i)$,

$$\psi(\tau) = \phi(\lambda)\frac{d\lambda(\tau)}{d\tau} \propto \phi\left(1 - \frac{\tau'}{\tau}\right)\frac{\tau'}{\tau^2}\,, \tag{59}$$

where τ' is a proportionality factor. Therefore $\psi(\tau)$ and $\tilde{f}(\bar{x})$ are characterized by power law tails in τ and $\bar{x}$ respectively *with the same Pareto exponent*.

In conclusion, the role of the λ-cutoff is also related to the relaxation process. This means that the slowest convergence rate is determined by the cutoff and is $\propto 1 - \lambda_{\rm M}$. In numerical simulations of heterogeneous KWEMs, as well as in real wealth distributions, the cutoff is necessarily finite, so that the convergence is fast (188). On the other hand, if considering a hypothetical wealth distribution with a power law extending to infinite values of x, one cannot find a fast relaxation, due to the infinite time scale of the system, due to the agents with $\lambda = 1$.

2.6. Microeconomic formulation of Kinetic theory models

Very recently, Chakrabarti and Chakrabarti have studied in a paper (`arXiv:cond-mat/0905.3972`) the framework based on microeconomic theory from which the kinetic theory market models could be addressed. They derived the moments of the CC model (158) and reproduced the exchange equations used in the CC model (with fixed savings parameter). In their framework considered, the utility function deals with the behavior of the agents in an exchange economy. The details are summarized below.

They considered an N-agent exchange economy, where each agent produces a single perishable commodity. Each of these goods is different from all other goods and money exists in the economy to simply facilitate transactions. These agents care for their future consumptions and hence they care about their savings in the period considered as well. Each of these agents are endowed with an initial amount of money which is assumed to be unity for every agent for simplicity. At each time step t, two agents meet randomly to carry out transactions according to their utility maximization principle. It is also assumed that the agents have time dependent preference structure, i.e., the parameters of the utility function can vary over time.

Let agent 1 produces Q_1 amount of commodity 1 only and agent 2 produces Q_2 amount of commodity 2 only and the amounts of money in their possession at time t are $m_1(t)$ and $m_2(t)$ respectively. It is believed that both of them are willing to trade and buy the other good by selling a fraction of their own productions as well as with the money that they hold. In general, at each time step there would be a net transfer of money from one agent to the other due to trade.

The utility functions as defined as follows: For agent 1, $U_1(x_1, x_2, m_1) = x_1^{\alpha_1} x_2^{\alpha_2} m_1^{\alpha_m}$ and for agent 2, $U_2(y_1, y_2, m_2) = y_1^{\alpha_1} y_2^{\alpha_2} m_2^{\alpha_m}$ where the arguments in both of the utility functions are consumption of the first (i.e. x_1 and y_1) and second good (i.e. x_2 and y_2) and amount of money in their possession respectively. For simplicity, they assume that the utility functions are of the Cobb-Douglas form with the sum of the powers normalized to 1 i.e. $\alpha_1 + \alpha_2 + \alpha_m = 1$.

Let the commodity prices to be determined in the market be denoted by p_1 and p_2. Now, the budget constraints are as follows: For agent 1 the budget constraint is $p_1x_1 + p_2x_2 + m_1 \leq M_1 + p_1Q_1$ and similarly, for agent 2 the constraint is $p_1y_1 + p_2y_2 + m_2 \leq M_2 + p_2Q_2$, which mean that the amount that agent 1 can spend for consuming x_1 and x_2 added to the amount of money that he holds after trading at time $t+1$ (i.e. m_1) cannot exceed the amount of money that he has at time t (i.e. M_1) added to what he earns by selling the good he produces (i.e. Q_1), and the same is true for agent 2.

Then the basic idea is that both of the agents try to maximize their respective utility subject to their respective budget constraints and the *invisible hand* of the market that is the price mechanism works to clear the market for both goods (i.e. total demand equals total supply for both goods at the equilibrium prices), which means that agent 1's problem is to maximize his utility subject to his budget constraint i.e. maximize $U_1(x_1, x_2, m_1)$ subject to $p_1.x_1 + p_2.x_2 + m_1 = M_1 + p_1.Q_1$. Similarly for agent 2, the problem is to maximize $U_1(y_1, y_2, m_2)$ subject to $p_1.y_1 + p_2.y_2 + m_2 = M_2 + p_2.Q_2$. Solving those two maximization exercises by Lagrange multiplier and applying the condition that the market remains in equilibrium, the competitive price vector $(\hat{p}_1, \hat{p}_2)$ as $\hat{p}_i = (\alpha_i/\alpha_m)(M_1 + M_2)/Q_i$ for $i = 1$, 2 is found.

The outcomes of such a trading process are then:

1. At optimal prices $(\hat{p}_1, \hat{p}_2)$, $m_1(t) + m_2(t) = m_1(t+1) + m_2(t+1)$, i.e., demand matches supply in all market at the market-determined price in equilibrium. Since money is also treated as a commodity in this framework, its demand (i.e. the total amount of money held by the two persons after trade) must be equal to what was supplied (i.e. the total amount of money held by them before trade).
2. If a restrictive assumption is made such that α_1 in the utility function can vary randomly over time with α_m remaining constant. It readily follows that α_2 also varies randomly over time with the restriction that the sum of α_1 and α_2 is a constant (1-α_m). Then in the money demand equations derived, if we suppose α_m is λ and $\alpha_1/(\alpha_1 + \alpha_2)$ is ϵ, it is found that money evolution equations are

$$m_1(t+1) = \lambda m_1(t) + \epsilon(1-\lambda)(m_1(t) + m_2(t))$$

$$m_2(t+1) = \lambda m_2(t) + (1-\epsilon)(1-\lambda)(m_1(t) + m_2(t)).$$

   For a fixed value of λ, if α_1 (or α_2) is a random variable with uniform distribution over the domain $[0, 1-\lambda]$, then ϵ is also uniformly distributed over the domain $[0, 1]$. It may be noted that then λ (i.e. α_m in the utility function) is the savings propensity used in the CC model.
3. For the limiting value of α_m in the utility function (i.e. $\alpha_m \to 0$ which implies $\lambda \to 0$), the money transfer equation describing the random sharing of money without saving is obtained, which has been used in the model of Dragulescu and Yakovenko.

It should also be mentioned that by fixing the average amount of money per agent to unity, they then give a simple derivation why in the CC model the steady state distribution is approximately a Gamma distribution.

3. Agent-based modelling based on Games

3.1. El Farol Bar Problem

In 1994, Brian Arthur introduced the 'El Farol Bar' problem (189) as a paradigm of complex economic systems: a population of agents have to decide whether to go to the bar opposite Santa Fe, every Thursday night. The bar has limited number of seats and can entertain at most $X\%$ of the population. If less than $X\%$ of the population go to the bar, the time spent in the bar is considered to be satisfying and it is better to attend the bar rather than staying at home. On the other hand, if more than $X\%$ of the population go to the bar, then it is too crowded and all the people in the bar would have an unsatisfying time and staying at home is considered to be better choice than attending the bar. So in order to optimise its own utility each agent has to try and predict what everybody else will do.

In particular Arthur was also interested in agents who have bounds on "rationality":

- do not have perfect information about their environment, in general they will only acquire information through interaction with the dynamically changing environment;
- do not have a perfect model of their environment;
- have limited computational power, so they can't work out all the logical consequences of their knowledge;
- have other resource limitations (e.g. memory).

Arthur modelled all this by randomly giving each agent a fixed menu of potentially suitable models to predict the number who will go given past data (e.g. the same as two weeks ago, the average of the past few weeks, etc). Each week each agent evaluates these models against the past data and chooses the one that was the best predictor on this data and then uses this to predict the number who will go this time. It will go if this prediction is less than X and not if it is more than X. Thus, in order to make decisions on whether to attend the bar, all the individuals are equipped with certain number of "strategies", which provide them the predictions of the attendance in the bar next week, based on the attendance in the past few weeks. As a result the number who go to the bar oscillates in an apparently random manner around the critical $X\%$ mark.

3.2. Basic Minority game

The Minority Games (abbreviated, MGs) (190) refer to the multiagent models of financial markets with the original formulation introduced by Challet and Zhang in 1997 (191), and all other variants (192; 193), most of which share the principal features that the models are repeated games and agents are inductive in nature. The original formulation of the Minority Game by Challet and Zhang (191) is sometimes referred as the "Original Minority Game" or the "Basic Minority Game".

The basic minority game consists of N (odd natural number) agents, who choose between one of the two decisions at each round of the game, using their own simple inductive strategies. The two decisions could be, for example, "buying" or "selling" commodities/assets, denoted by 0 or 1, at a given time t. An agent wins the game if it is one of the members of the minority group, and thus at each round, the minority group of agents win the game and rewards are given to those strategies that predict the winning side. All the agents have access to finite amount of public information, which is a common bit-string "memory" of the M most recent outcomes, composed of the winning sides in the past few rounds. Thus the agents are said to exhibit "bounded rationality" (189).

Consider for example, memory $M = 2$; then there are $P = 2^M = 4$ possible "history" bit strings: 00, 01, 10 and 11. A "strategy" consists of a response, i.e., 0 or 1, to each possible history bit strings; therefore, there are $G = 2^P = 2^{2^M} = 16$ possible strategies which constitute the "strategy space". At the beginning of the game, each agent randomly picks k strategies, and after the game, assigns one "virtual" point to a strategy which would have predicted the correct outcome. The actual performance r of the player is measured by the number of times the player wins, and the strategy, using which the player wins, gets a "real" point. Unlike

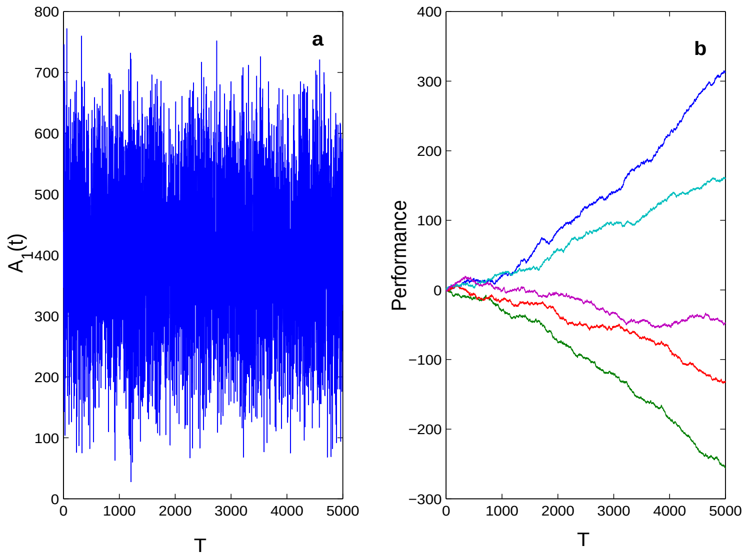

Figure 37: Attendance fluctuation and performances of players in Basic Minority Game. Plots of (a) attendance and (b) performance of the players for the basic minority game with $N = 801$; $M = 6$; $k = 10$ and $T = 5000$ (194).

most economics models which assume agents are "deductive" in nature, here a "trial-and-error" *inductive* thinking approach is implicitly implemented in process of decision-making when agents make their choices in the games. A record of the number of agents who have chosen a particular action, say, "selling" denoted by 1, $A_1(t)$ as a function of time is kept (see Fig. 37). The fluctuations in the behaviour of $A_1(t)$ actually indicate the system's total utility. For example, we can have a situation where only one player is in the minority and all the other players lose. The other extreme case is when $(N-1)/2$ players are in the minority and $(N+1)/2$ players lose. The total utility of the system is obviously greater for the latter case and from this perspective, the latter situation is more desirable. Therefore, the system is more efficient when there are smaller fluctuations around the mean than when the fluctuations are larger.

3.3. Evolutionary minority games

Challet generalized the basic minority game (191; 195) mentioned above to include the Darwinist selection: the worst player is replaced by a new one after some time steps, the new player is a "clone" of the best player, i.e. it inherits all the strategies but with corresponding virtual capitals reset to zero (analogous to a new born baby, though having all the predispositions from the parents, does not inherit their knowledge). To keep a certain diversity they introduced a mutation possibility in cloning. They allowed one of the strategies of the best player to be replaced by a new one. Since strategies are not just recycled among the players any more, the whole strategy phase space is available for selection. They expected this population to be capable of "learning" since bad players are weeded out with time, and fighting is among the so-to-speak the "best" players. Indeed in Fig. 38, they observed that the learning emerged in time. Fluctuations are reduced and saturated, this implies the average gain for everybody is improved but never reaches the ideal limit.

Li, Riolo and Savit (196; 197) studied the minority game in the presence of evolution. In particular, they examined the behavior in games in which the dimension of the strategy space, m, is the same for all agents and fixed for all time. We find that for all values of m, not too large, evolution results in a substantial improvement in overall system performance. They also showed that after evolution, results obeyed a scaling relation among games played with different values of m and different numbers of agents, analogous to that found in the non-evolutionary, adaptive games. Best system performance still occurs, for a given number of agents, at m_c, the same value of the dimension of the strategy space as in the non-evolutionary case,

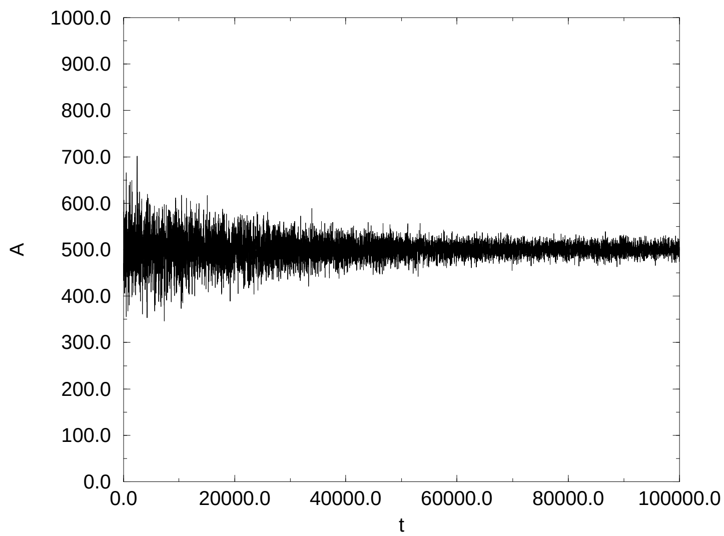

Figure 38: Temporal attendance of A for the genetic approach showing a learning process. Reproduced from (191; 195)

but system performance is now nearly an order of magnitude better than the non-evolutionary result. For $m < m_c$, the system evolves to states in which average agent wealth is better than in the random choice game, despite (and in some sense because of) the persistence of maladaptive behavior by some agents. As m gets large, overall systems performance approaches that of the random choice game.

They continued the study of evolution in minority games by examining games (196; 197) in which agents with poorly performing strategies can trade in their strategies for new ones from a different strategy space. In the context of the games which meant allowing for strategies that use information from different numbers of time lags, m. They found, in all the games, that after evolution, wealth per agent is high for agents with strategies drawn from small strategy spaces (small m), and low for agents with strategies drawn from large strategy spaces (large m). In the game played with N agents, wealth per agent as a function of m was very nearly a step function. The transition is at $m = m_t$, where $m_t \simeq m_c - 1$, and m_c is the critical value of m at which N agents playing the game with a fixed strategy space (fixed m) have the best emergent coordination and the best utilization of resources. They also found that overall system-wide utilization of resources is independent of N. Furthermore, although overall system-wide utilization of resources after evolution varied somewhat depending on some other aspects of the evolutionary dynamics, in the best cases, utilization of resources was on the order of the best results achieved in evolutionary games with fixed strategy spaces.

3.4. Adaptive minority games

Sysi-Aho et al. (198; 199; 194; 200) presented a simple modification of the basic minority game where the players modify their strategies periodically after every time interval τ, depending on their performances: if a player finds that he is among the fraction n (where $0 < n < 1$) who are the worst performing players, he adapts himself and modifies his strategies. They proposed that the agents use hybridized one-point genetic crossover mechanism (as shown in Fig. 39), inspired by genetic evolution in biology, to modify the strategies and replace the bad strategies. They studied the performances of the agents under different conditions and investigate how they adapt themselves in order to survive or be the best, by finding new strategies using the highly effective mechanism. They also studied the measure of total utility of the system $U(x_t)$, which is the number of players in the minority group; the total utility of the system is maximum U_{max} as the highest number of players win is equal to $(N-1)/2$. The system is more efficient when the deviations from

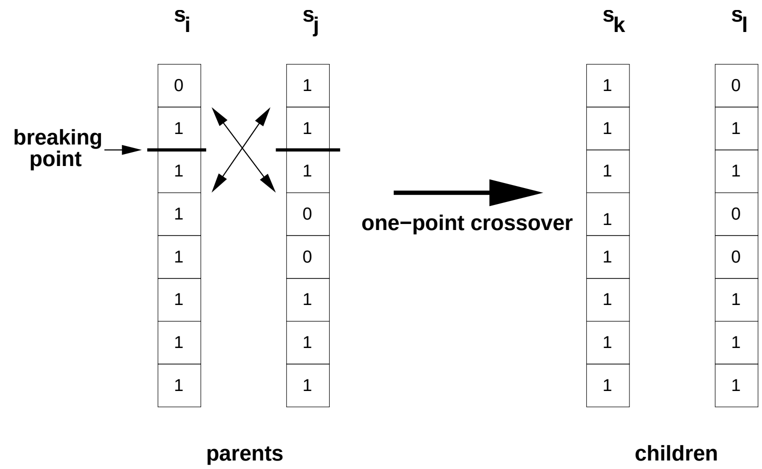

Figure 39: Schematic diagram to illustrate the mechanism of one-point genetic crossover for producing new strategies. The strategies s_i and s_j are the parents. We choose the breaking point randomly and through this one-point genetic crossover, the children s_k and s_l are produced and substitute the parents. Reproduced from (194).

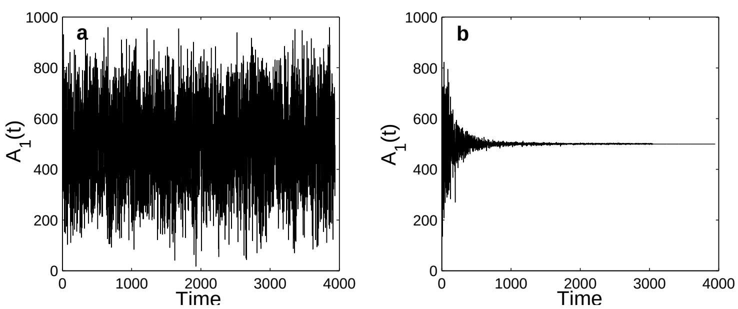

Figure 40: Plot to show the time variations of the number of players A_1 who choose action 1, with the parameters $N = 1001$, $m = 5$, $s = 10$ and $t = 4000$ for (a) basic minority game and (b) adaptive game, where $\tau = 25$ and $n = 0.6$. Reproduced from (194).

the maximum total utility U_{max} are smaller, or in other words, the fluctuations in $A_1(t)$ around the mean become smaller.

Interestingly, the fluctuations disappear totally and the system stabilizes to a state where the total utility of the system is at maximum, since at each time step the highest number of players win the game (see Fig. 40). As expected, the behaviour depends on the parameter values for the system (194; 200). They used the utility function to study the efficiency and dynamics of the game as shown in Fig. 41. If the parents are chosen randomly from the pool of strategies then the mechanism represents a "one-point genetic crossover" and if the parents are the best strategies then the mechanism represents a "hybridized genetic crossover". The children may replace parents or two worst strategies and accordingly four different interesting cases arise: (a) one-point genetic crossover with parents "killed", i.e. parents are replaced by the children, (b) one-point genetic crossover with parents "saved", i.e. the two worst strategies are replaced by the children but the parents are retained, (c) hybridized genetic crossover with parents "killed" and (d) hybridized genetic crossover with parents "saved".

In order to determine which mechanism is the most efficient, we have made a comparative study of the four cases, mentioned above. We plot the attendance as a function of time for the different mechanisms in Fig. 42. In Fig. 43 we show the total utility of the system in each of the cases (a)-(d), where we have plotted results of the average over 100 runs and each point in the utility curve represents a time average

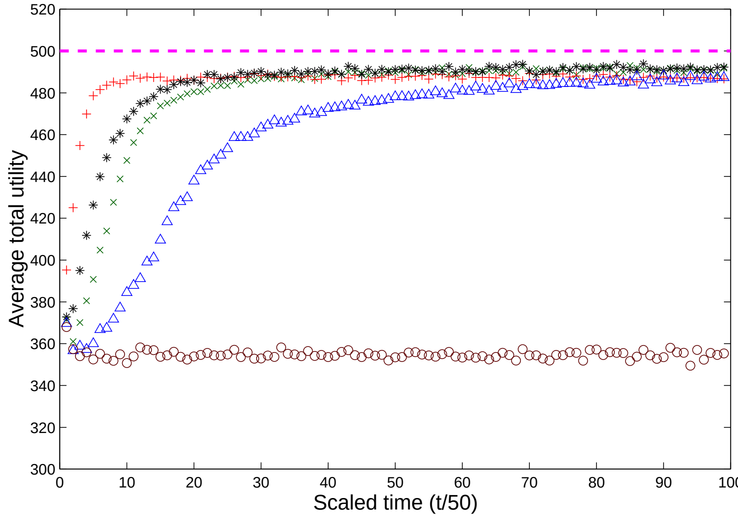

Figure 41: Plot to show the variation of total utility of the system with time for the basic minority game for $N = 1001$, $m = 5$, $s = 10$, $t = 5000$, and adaptive game, for the same parameters but different values of τ and n. Each point represents a time average of the total utility for separate bins of size 50 time-steps of the game. The maximum total utility ($= (N-1)/2$) is shown as a dashed line. The data for the basic minority game is shown in circles. The plus signs are for $\tau = 10$ and $n = 0.6$; the asterisk marks are for $\tau = 50$ an $n = 0.6$; the cross marks for $\tau = 10$ and $n = 0.2$ and triangles for $\tau = 50$ and $n = 0.2$. The ensemble average over 70 different samples was taken in each case. Reproduced from (194).

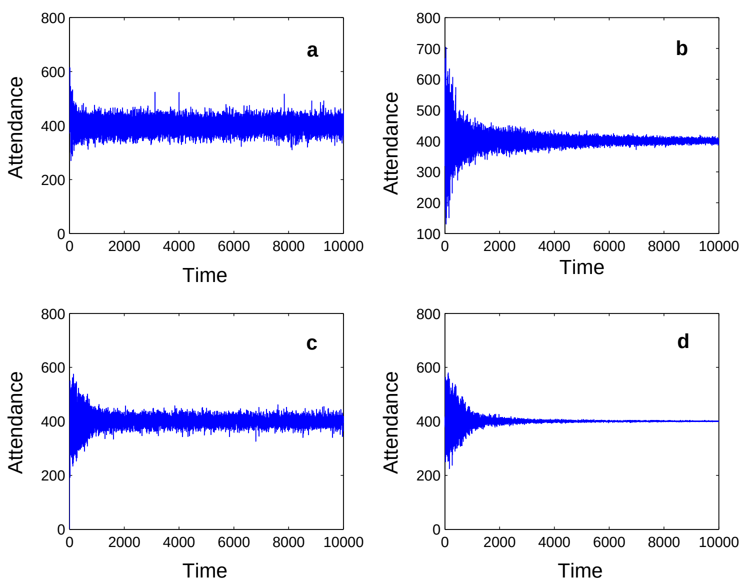

Figure 42: Plots of the attendances by choosing parents randomly (a) and (b), and using the best parents in a player's pool (c) and (d). In (a) and (c) case parents are replaced by children and in (b) and (d) case children replace the two worst strategies. Simulations have been done with $N = 801$, $M = 6$, $k = 16$, $t = 40$, $n = 0.4$ and $T = 10000$.

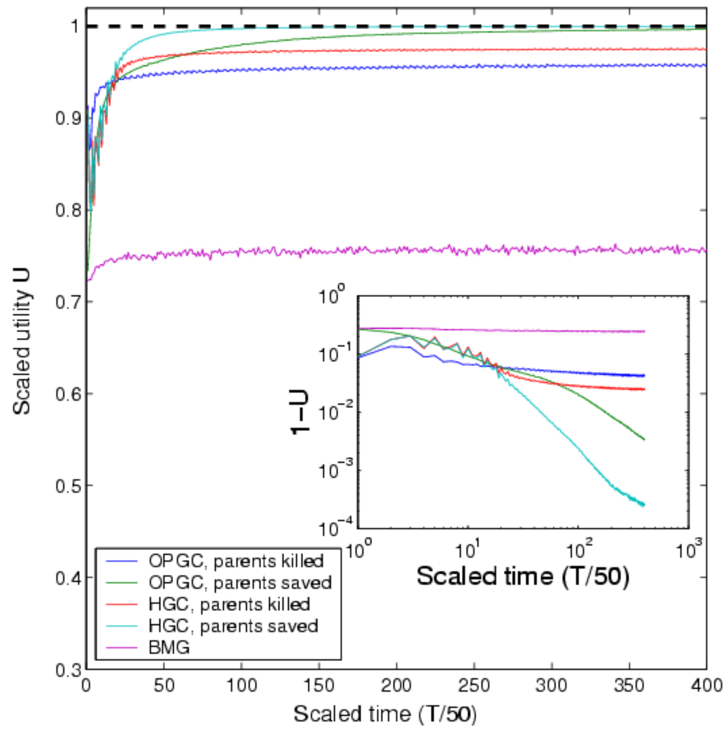

Figure 43: Plots of the scaled utilities of the four different mechanisms in comparison with that of the basic minority game. Each curve represents an ensemble average over 100 runs and each point in a curve is a time average over a bin of length 50 time-steps. In the inset, the quantity $(1 - U)$ is plotted against scaled time in the double logarithmic scale. Simulations are done with $N = 801$, $M = 6$, $k = 16$, $t = 40$, $n = 0.4$ and $T = 20000$. Reproduced from (194).

taken over a bin of length 50 time-steps. The simulation time is doubled from those in Fig. 42, in order to expose the asymptotic behaviour better. On the basis of Figs. 42 and 43, we find that the case (d) is the most efficient.

It should be noted that the mechanism of evolution of strategies is considerably different from earlier attempts (201; 196; 197). This is because in this mechanism the strategies are changed by the agents themselves and even though the strategy space evolves continuously, its size and dimensionality remain the same.

The Hamming distance d_H between two bit-strings is defined as the ratio of the number of uncommon bits to the total length of the bit strings. It is a measure of the correlation between two strategies:

$$d_H = \begin{cases} 0 & \text{correlated} \\ 0.5 & \text{uncorrelated} \\ 1 & \text{anti-correlated} \end{cases}$$

which can be plotted as the game evolves. In Fig. 44 (a) one can see the evolution of the average Hamming distance of all the strategies of a player in a game, where the player adapts using one-point genetic crossover and the two worst strategies are replaced by the children and the parents are also saved. It should be noted that the Hamming distance can change only when the worst strategies are replaced by the children and the parents are saved, where the bits in a strategy pool can change over time. Otherwise the bits in the pool of strategies remain the same. We observe that the curves tend to move downwards from around 0.5 towards zero, which means that as the time evolves, the correlation amongst the strategies increases and the strategies in the pool of a particular agent converges towards one strategy. The nature of the curves depend a lot on the parameters of the game. In Fig. 44 (b) one can see the evolution of the average Hamming distance

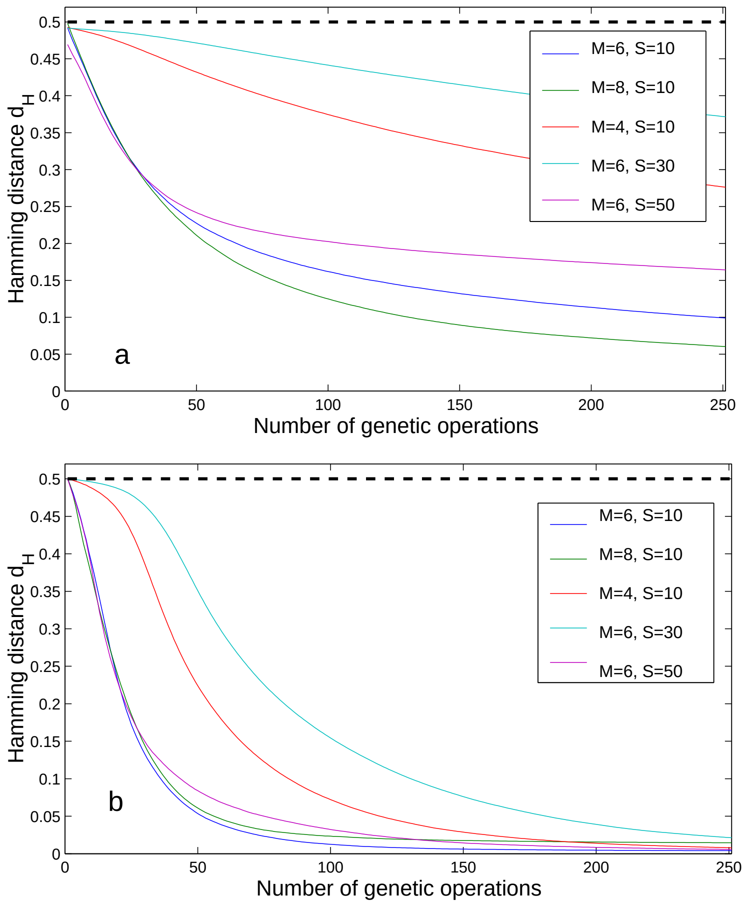

Figure 44: Plot of the average Hamming distance of all the strategies in a pool of a player with time, where the player adapts using (a) one-point genetic crossover and (b) hybridized genetic crossover, and in both cases the two worst strategies are replaced by the children and the parents are also saved. Each curve is an ensemble average over 20 runs. Reproduced from (194).

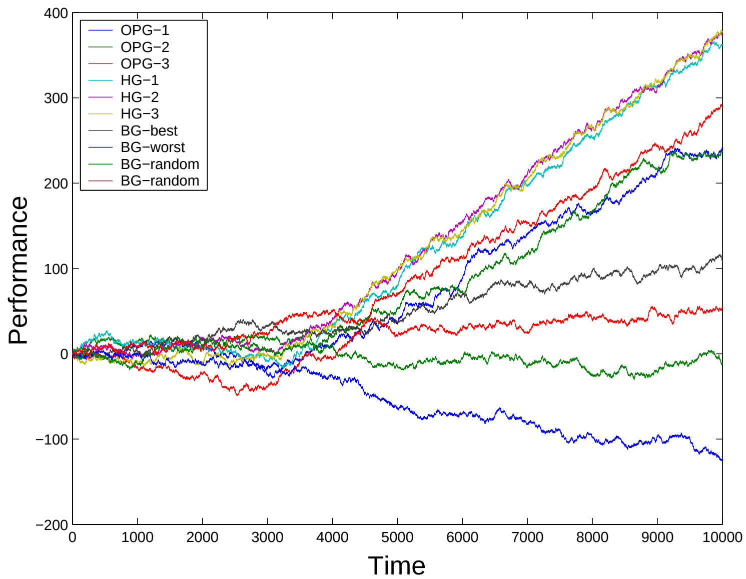

Figure 45: Plot of the performance of the players where after $T = 3120$ time-steps, six players begin to adapt and modify their strategies: three using hybridized genetic crossover mechanism and the other three using one point genetic crossover, where children replace the parents. Other players play the basic minority game all the time and do not adapt. The simulations are done with $N = 801$, $M = 8$, $k = 16$, $n = 0.3$, $t = 80$, and $T = 10000$. Reproduced from (194).

of all the strategies of a player in the game, where the player adapts using hybridized genetic crossover and the two worst strategies are replaced by the children and the parents are also saved. Here too, the strategies in the pool of a particular agent converges towards one strategy, and at a faster rate than with the previous mechanism. We observe that increasing memory M does not change dramatically the convergence rate, but as we increase the number of strategies in the pools, the convergence slows down. In order to investigate what happens in the level of an individual agent, we created a competitive surrounding– "test" situation where after $T = 3120$ time-steps, six players begin to adapt and modify their strategies such that three are using hybridized genetic crossover mechanism and the other three one point genetic crossover, where children replace the parents. The rest of the players play the basic minority game. In this case it turns out that in the end the best players are those who use the hybridized mechanism, second best are those using the one-point mechanism, and the bad players those who do not adapt at all. In addition it turns out that the competition amongst the players who adapt using the hybridized genetic crossover mechanism is severe.

Due to the simplicity of these models (198; 199; 194; 200), a lot of freedom is found in modifying the models to make the situations more realistic and appliable to many real dynamical systems, and not only financial markets. Many details in the model can be fine-tuned to imitate the real markets or behaviour of other complex systems. Many other sophisticated models based on these games can be setup and implemented, which show a great potential over the commonly adopted statistical techniques in analyses of financial markets.

3.5. Some important remarks

For modeling purposes, Minority Games were meant to serve as a class of simple models which could produce some of macroscopic features being observed in the real financial markets (193; 202), usually termed as "stylized facts" which included the fat-tail price return distribution and volatility clustering (190; 192). Inspite the initial furore (203; 204; 205) they have however failed to capture or reproduce most important stylized facts. However, in the physicists' community, the Minority Games became an interesting and established class of complex disordered systems (202; 206), lending a large amount of deep physical insights (207; 208). Since in the BMG model a Hamiltonian function can be defined and analytic solutions can be

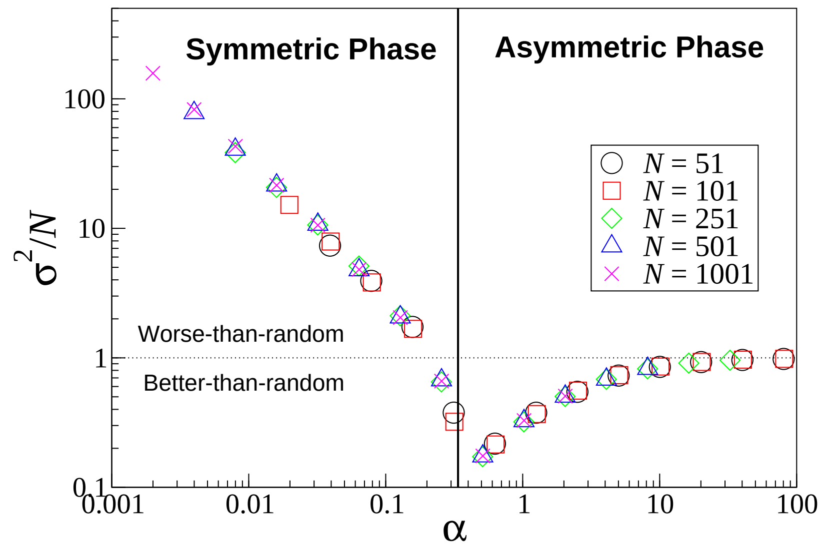

Figure 46: The simulation results of the variance in attendance σ^2/N as a function of the control parameter $\alpha = 2^M/N$ for games with $k = 2$ strategies for each agent, ensemble averaged over 100 sample runs. Dotted line shows the value of volatility in random choice limit. Solid line shows the critical value of $\alpha = \alpha_c \approx 0.3374$. Resolution of the curve can be improved to shows σ^2/N attains minimum at $\alpha \approx \alpha_c$. Reproduced from `arxiv: 0811.1479`.

developed in some regimes of the model, the model may be viewed with a more complete physical picture (209). It is characterized by a clear two-phase structure with very different collective behaviours in the two phases, as in many known conventional physical systems (208; 210).

Savit et al (208) first found that the macroscopic behaviour of the system does not depend independently on the parameters N and M, but instead depends on the ratio

$$\alpha \equiv \frac{2^M}{N} = \frac{P}{N} \tag{60}$$

which serves as the most important control parameter in the game. The variance in the attendance (see also Ref. (199) or volatility σ^2/N, for different values of N and M depend only on the ratio α. Fig. 46 shows a plot of σ^2/N against the control parameter α, where the data collapse of σ^2/N for different values of N and M is clearly evident. The dotted line in Fig. 46 corresponds to the "coin-toss" limit (random choice or pure chance limit), in which agents play by simply making random decisions (by coin-tossing) at every rounds of the game. This value of σ^2/N in coin-toss limit can be obtained by simply assuming a binomial distribution of the agents' binary actions, with probability 0.5, where $\sigma^2/N = 0.5(1-0.5)\cdot 4 = 1$. When α is small, the value of σ^2/N of the game is larger than the coin-toss limit which implies the collective behaviours of agents are worse than the random choices. In the early literature, it was popularly called as the *worse-than-random* regime. When α increases, the value of σ^2/N decreases and enters a region where agents are performing better than the random choices, which was popularly called as the *better-than-random* regime. The value of σ^2/N reaches a minimum value which is substantially smaller than the coin-toss limit. When α further increases, the value of σ^2/N increases again and approaches the coin-toss limit. This allowed one to identify two phases in the Minority Game, as separated by the minimum value of σ^2/N in the graph. The value of α where the rescaled volatility attended its minimum was denoted by α_c, which represented the phase transition point; α_c has been shown to have a value of $0.3374\ldots$ (for $k = 2$) by analytical calculations (209).

Besides these collective behaviours, physicists became also interested in the dynamics of the games such as crowd-anticrowd movement of agents, periodic attractors, etc (211; 212; 213). In this way, the Minority Games serve as a useful tool and provide a new direction for physicists in viewing and analyzing the underlying dynamics of complex evolving systems such as the financial markets.

Part V
Conclusions and outlook

Agent-based models of order books are a good example of interactions between ideas and methods usually linked to Economics and Finance (microstructure of markets, agent interaction) and Physics (reaction-diffusion processes, deposition-evaporation process, kinetic theory of gases). As of today, existing models exhibit a trade-off between "realism" and calibration in its mechanisms and processes (empirical models such as (81)), and explanatory power of simple observed behaviors ((106; 114) for example). In the first case, some of the "stylized facts" may be reproduced, but using empirical processes that may not be linked to any behaviour observed on the market. In the second case, these are only toy models that cannot be calibrated on data. The mixing of many features, as in (108) and as is usually the case in behavioral finance, leads to untractable models where the sensitivity to one parameter is hardly understandable.

Therefore, no empirical model can tackle properly empirical facts such as volatility clustering. Importing toy model features explaining volatility clustering or market interactions in order book models is yet to be done. Finally, let us also note that to our knowledge, no agent-based model of order books deals with the multidimensional case. Implementing agents trading simultaneously several assets in a way that reproduces empirical observations on correlation and dependance remains an open challenge.

We believe this type of modeling is crucial for future developments in finance. The financial crisis that occurred in 2007-2008 is expected to create a shock in classic modelization in Economics and Finance. Many scientists have expressed their views on this subject (e.g. (26; 27; 28)) and we believe as well that agent-based models we have presented here will be at the core of future modelizations. As illustrations, let us mention (214), which modelizes the interbank market and investigates systemic risk, (215), which investigates the effects of use of leverage and margin calls on the stability of a market and (133), which provides a brief overview of the study of wealth distributions and inequalities. No doubt these will be followed by many other contributions.

Acknowledgements

The authors would like to thank their collaborators. AC is grateful to B.K. Chakrabarti, K. Kaski, J. Kertesz, T. Lux, M. Marsili, D. Stauffer and V. Yakovenko for invaluable suggestions and criticisms.

References

[1] S. M. de Oliveira, P. M. C. de Oliveira, D. Stauffer, Evolution, Money, War and Computers, B. G. Teubner, Stuttgart-Leipzig, 1999.
[2] D. Stauffera, S. M. de Oliveira, P. M. C. de Oliveira, J. S. de Sa Martins, Biology, Sociology, Geology by Computational Physicists, Elsevier Science, 2006.
[3] B. K. Chakrabarti, A. Chakraborti, A. Chatterjee (Eds.), Econophysics and Sociophysics: Trends and Perspectives, 1st Edition, Wiley - VCH, Berlin, 2006.
[4] P. A. Samuelson, Economics, Mc Grow Hill, Auckland, 1998.
[5] J. M. Keynes, The general theory of employment, Interest and Money, The Royal Economic Society, Macmillan Press, London, 1973.
[6] L. Bachelier, Theorie de la speculation, Annales Scientifiques de l'Ecole Normale Superieure III-17 (1900) 21–86.
[7] A. J. Boness, in: P. H. Cootner (Ed.), The Random Character of Stock Market Prices, MIT, Cambridge, MA, 1967, Ch. English translation of: L. Bachelier, Theorie de la Speculation, Annales scientifiques de l'Ecole Normale Superieure III-17.
[8] T. A. S. S Haberman (Ed.), History of Actuarial Science, Vol. 7, Pickering and Chatto Publishers, London, 1995, Ch. English translation of: Louis bachelier, Théorie de la spéculation, Annales scientifiques de l'Ecole Normale Superieure.
[9] J. Courtault, Y. Kabanov, B. Bru, P. Crepel, I. Lebon, A. Le Marchand, Louis Bachelier on the centenary of Theorie de la Speculation, Mathematical Finance 10 (3) (2000) 339–353.
[10] M. Taqqu, Bachelier and his times: A conversation with Bernard Bru, Finance and Stochastics 5 (1) (2001) 3–32.
[11] E. F. R. J. O'Connor, The mactutor history of mathematics archive, http://www-history.mcs.st-andrews.ac.uk/Biographies/Bachelier.html (2000).
[12] K. Itō, H. McKean, Diffusion Processes and Their Sample Paths, Springer, Berlin, 1996.

[13] M. F. M. Osborne, Brownian motion in the stock market, OPERATIONS RESEARCH 7 (2) (1959) 145–173. doi:10.1287/opre.7.2.145.
URL http://or.journal.informs.org/cgi/content/abstract/7/2/145
[14] P. Samuelson, Proof that properly anticipated prices fluctuate randomly, Management Review 6 (2).
[15] F. Black, M. Scholes, The pricing of options and corporate liabilities, Journal of political economy 81 (3) (1973) 637.
[16] R. Merton, Theory of rational option pricing, The Bell Journal of Economics and Management Science (1973) 141–183.
[17] D. Duffie, Dynamic asset pricing theory, Princeton University Press Princeton, NJ, 1996.
[18] H. Föllmer, A. Schied, T. Lyons, Stochastic finance. An introduction in discrete time, The Mathematical Intelligencer 26 (4) (2004) 67–68.
[19] B. Mandelbrot, The Pareto-Levy law and the distribution of income, International Economic Review (1960) 79–106.
[20] B. Mandelbrot, The variation of certain speculative prices, Journal of business 36 (4) (1963) 394.
[21] R. Cont, P. Tankov, Financial modelling with jump processes, Chapman & Hall/CRC, 2004.
[22] P. K. Clark, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41 (1) (1973) 135–55.
URL http://ideas.repec.org/a/ecm/emetrp/v41y1973i1p135-55.html
[23] S. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, Rev. Financ. Stud. 6 (2) (1993) 327–343. doi:10.1093/rfs/6.2.327.
URL http://rfs.oxfordjournals.org/cgi/content/abstract/6/2/327
[24] J. Gatheral, The volatility surface: a practitioner's guide, Wiley, 2006.
[25] C. P. Kindleberger, R. Z. Aliber, Manias, Panics, And Crashes: A History Of Financial Crises, 5th Edition, John Wiley & Sons, 2005.
[26] J. Bouchaud, Economics needs a scientific revolution, Nature 455 (7217) (2008) 1181.
[27] T. Lux, F. Westerhoff, Economics crisis, Nature Physics 5 (1) (2009) 2–3.
[28] J. D. Farmer, D. Foley, The economy needs agent-based modelling, Nature 460 (7256) (2009) 685–686. doi:10.1038/460685a.
URL http://dx.doi.org/10.1038/460685a
[29] E. Britannica, Encyclopædia Britannica, http://www.britannica.com/ (2009).
[30] F. Reif, Fundamentals of Statistical and Thermal Physics, Mc Grow-Hill, Singapore, 1985.
[31] R. K. Pathria, Statistical Mechanics, 2nd Edition, Butterworth-Heinemann, Oxford, 1996.
[32] L. D. Landau, Statistical Physics, Vol. 5 of Theoretical Physics, Pergamon Press, Oxford, 1965.
[33] G. Parisi, Complex systems: a physicist's viewpoint, Physica A: Statistical Mechanics and its Applications 263 (1-4) (1999) 557 – 564, proceedings of the 20th IUPAP International Conference on Statistical Physics. doi:DOI:10.1016/S0378-4371(98)00524-X.
URL http://www.science-direct.com/science/article/B6TVG-3YGKSSR-20/2/c57b6ecbd49f09ef19275c42ef8cda6e
[34] W. Arthur, Complexity and the economy, Science 284 (5411) (1999) 107.
[35] V. Pareto, Cours d'economie politique, Rouge, Lausanne, 1897.
[36] V. Pareto, Cours d'economie politique, Droz, Geneva, 1897, reprinted as a volume of Oeuvres Completes.
[37] N. di Ettore Majorana, Il valore delle leggi statistiche nella fisica e nelle scienze sociali, Scientia 36 (1942) 58–66.
[38] R. Mantegna, Presentation of the English translation of Ettore Majorana's paper: The value of statistical laws in physics and social sciences, Quantitative Finance 5 (2) (2005) 133–140.
[39] R. Nunzio Mantegna, The Tenth Article of Ettore Majorana, Europhysics News 37.
[40] L. Kadanoff, From simulation model to public policy: An examination of Forrester's" Urban Dynamics", Simulation 16 (6) (1971) 261.
[41] E. Montroll, W. Badger, Introduction to quantitative aspects of social phenomena, Gordon and Breach New York, 1974.
[42] B. Roehner, Patterns of speculation: a study in observational econophysics, Cambridge University Press, 2002.
[43] R. NUNZIO MANTEGNA, Levy walks and enhanced diffusion in Milan stock exchange, Physica. A 179 (2) (1991) 232–242.
[44] Http://www.unifr.ch/econophysics/ (2009).
[45] Http://www.econophysics.org/ (2009).
[46] R. Tsay, Analysis of financial time series, Wiley-Interscience, 2005.
[47] N. Kaldor, Economic growth and capital accumulation, The Theory of Capital, Macmillan, London.
[48] T. Bollerslev, R. F. Engle, D. B. Nelson, Chapter 49 arch models, in: R. F. Engle, D. L. McFadden (Eds.), Handbook of Econometrics 4, Vol. 4 of Handbook of Econometrics, Elsevier, 1994, pp. 2959 – 3038. doi:DOI:10.1016/S1573-4412(05)80018-2.
URL http://www.sciencedirect.com/science/article/B7GX7-4FPWV09-M/2/08aeecb0508a0a1093e074eeb0686e91
[49] A. Pagan, The econometrics of financial markets, Journal of empirical finance 3 (1) (1996) 15–102.
[50] D. Guillaume, M. Dacorogna, R. Davé, U. Müller, R. Olsen, O. Pictet, From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets, Finance and Stochastics 1 (2) (1997) 95–129.
[51] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance 1 (2) (2001) 223–236.
[52] B. Mandelbrot, The variation of certain speculative prices, Journal of business 36 (4) (1963) 394.
[53] P. Gopikrishnan, V. Plerou, L. A. Nunes Amaral, M. Meyer, H. E. Stanley, Scaling of the distribution of fluctuations of financial market indices, Phys. Rev. E 60 (5) (1999) 5305–5316. doi:10.1103/PhysRevE.60.5305.
[54] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, H. E. Stanley, Inverse cubic law for the probability distribution of stock price variations, The European Physical Journal B 3 (1998) 139.

URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/9803374
[55] R. Cont, M. Potters, J.-P. Bouchaud, Scale Invariance and Beyond, SSRN eLibrarydoi:10.2139/ssrn.39420.
[56] Y. Liu, P. Cizeau, M. Meyer, C. K. Peng, H. E. Stanley, Correlations in economic time series, Physica A: Statistical and Theoretical Physics 245 (3-4) (1997) 437 – 440. doi:DOI:10.1016/S0378-4371(97)00368-3.
URL http://www.sciencedirect.com/science/article/B6TVG-3SPXB8W-1C/2/87aab1158cd697b58d6813b1fb34098f
[57] P. Cizeau, Y. Liu, M. Meyer, C. K. Peng, H. E. Stanley, Volatility distribution in the s& p500 stock index, Physica A: Statistical and Theoretical Physics 245 (3-4) (1997) 441 – 445. doi:DOI:10.1016/S0378-4371(97)00417-2.
URL http://www.sciencedirect.com/science/article/B6TVG-3SPXB8W-1D/2/9690acf6fed0562007691e1fd89b879c
[58] J. E. Griffin, R. C. A. Oomen, D. AG, Sampling returns for realized variance calculations: tick time or transaction time?, Econometric Reviews 27 (1) (2008) 230–253.
[59] N. Huth, F. Abergel, The times change: Multivariate subordination, empirical facts, SSRN eLibrary.
[60] F. Lillo, J. D. Farmer, The long memory of the efficient market, Studies in Nonlinear Dynamics & Econometrics 8 (3). doi:10.2202/1558-3708.1226.
URL http://www.bepress.com/snde/vol8/iss3/art1
[61] J. Bouchaud, J. D. Farmer, F. Lillo, T. Hens, K. R. Schenk-Hopp, How markets slowly digest changes in supply and demand, in: Handbook of Financial Markets: Dynamics and Evolution, North-Holland, San Diego, 2009, pp. 57–160.
URL http://www.sciencedirect.com/science/article/B8N8N-4W6Y2CK-C/2/25f44d99fa0db293e933d2510e84d2c3
[62] V. Plerou, P. Gopikrishnan, L. A. Nunes Amaral, X. Gabaix, H. Eugene Stanley, Economic fluctuations and anomalous diffusion, Phys. Rev. E 62 (3) (2000) R3023–R3026. doi:10.1103/PhysRevE.62.R3023.
[63] A. C. Silva, V. M. Yakovenko, Stochastic volatility of financial markets as the fluctuating rate of trading: An empirical study, Physica A: Statistical Mechanics and its Applications 382 (1) (2007) 278 – 285, applications of Physics in Financial Analysis - Proceedings of the 5th International Conference 'Applications of Physics in Financial Analysis' (APFA 5), Torino, Italy, 29 June 2006 – 1 July 2006, 5th International Conference 'Applications of Physics in Financial Analysis'. doi:DOI:10.1016/j.physa.2007.03.051.
URL http://www.sciencedirect.com/science/article/B6TVG-4NDMNBH-3/2/2db88b37597ea91b42eefcb81a8c48a7
[64] P. Gopikrishnan, V. Plerou, X. Gabaix, H. E. Stanley, Statistical properties of share volume traded in financial markets, Physical Review E 62 (4) (2000) R4493, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevE.62.R4493}.
URL http://link.aps.org/abstract/PRE/v62/pR4493
[65] W. Feller, Introduction to the Theory of Probability and its Applications, Vol. 2, Wiley, New York, 1968.
[66] B. Biais, P. Hillion, C. Spatt, An empirical analysis of the limit order book and the order flow in the Paris Bourse, Journal of Finance (1995) 1655–1689.
[67] P. Gopikrishnan, V. Plerou, X. Gabaix, H. E. Stanley, Statistical properties of share volume traded in financial markets, Physical Review - Series E- 62 (4; PART A) (2000) 4493–4496.
[68] D. Challet, R. Stinchcombe, Analyzing and modeling 1+ 1d markets, Physica A: Statistical Mechanics and its Applications 300 (1-2) (2001) 285–299.
[69] S. Maslov, M. Mills, Price fluctuations from the order book perspectiveempirical facts and a simple model, Physica A: Statistical Mechanics and its Applications 299 (1-2) (2001) 234–246.
[70] J. P. Bouchaud, M. Mézard, M. Potters, Statistical properties of stock order books: empirical results and models, Quantitative Finance 2 (2002) 251.
[71] M. Potters, J. P. Bouchaud, More statistical properties of order books and price impact, Physica A: Statistical Mechanics and its Applications 324 (1-2) (2003) 133–140.
[72] M. B. Garman, Market microstructure, Journal of Financial Economics 3 (3) (1976) 257 – 275. doi:DOI:10.1016/0304-405X(76)90006-4.
URL http://www.sciencedirect.com/science/article/B6VBX-458WN91-6/2/7ffe49abba201b52ca4980230456876c
[73] A. S. Kyle, Continuous auctions and insider trading, Econometrica 53 (6) (1985) 1315–1335, ArticleType: primary_article / Full publication date: Nov., 1985 / Copyright 1985 The Econometric Society. doi:10.2307/1913210.
URL http://www.jstor.org/stable/1913210
[74] L. R. Glosten, Is the electronic open limit order book inevitable?, The Journal of Finance 49 (4) (1994) 1127–1161, ArticleType: primary_article / Full publication date: Sep., 1994 / Copyright 1994 American Finance Association. doi:10.2307/2329182.
URL http://www.jstor.org/stable/2329182
[75] M. O'Hara, Market Microstructure Theory, second edition Edition, Blackwell Publishers, 1997.
[76] P. H. B. Biais, T. Foucault, Microstructure des marchs financiers : Institutions, modles et tests empiriques, Presses Universitaires de France - PUF, 1997.
[77] J. Hasbrouck, Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading, Oxford University Press, USA, 2007.
[78] I. Dremin, A. Leonidov, On distribution of number of trades in different time windows in the stock market, Physica A: Statistical Mechanics and its Applications 353 (2005) 388 – 402. doi:DOI:10.1016/j.physa.2004.12.048.
URL http://www.science-direct.com/science/article/B6TVG-4FB91CD-5/2/eb8c8c9a1b1ef385a9b134c3ade8711c
[79] B. M. Hill, A simple general approach to inference about the tail of a distribution, The Annals of Statistics 3 (5) (1975) 1163–1174, ArticleType: primary_article / Full publication date: Sep., 1975 / Copyright 1975 Institute of Mathematical Statistics. doi:10.2307/2958370.
URL http://www.jstor.org/stable/2958370
[80] L. de Haan, S. Resnick, H. Drees, How to make a hill plot, The Annals of Statistics 28 (1) (2000) 254–274.

[81] S. Mike, J. D. Farmer, An empirical behavioral model of liquidity and volatility, Journal of Economic Dynamics and Control 32 (1) (2008) 200–234.

[82] F. Lillo, D. Farmer, R. Mantegna, Econophysics: Master curve for price-impact function, Nature 421 (6919) (2003) 130, 129.
URL http://dx.doi.org/10.1038/421129a

[83] T. W. Epps, Comovements in stock prices in the very short run, Journal of the American Statistical Association 74 (366) (1979) 291–298, ArticleType: primary_article / Full publication date: Jun., 1979 / Copyright 1979 American Statistical Association. doi:10.2307/2286325.
URL http://www.jstor.org/stable/2286325

[84] G. Bonanno, F. Lillo, R. N. Mantegna, High-frequency cross-correlation in a set of stocks, Quantitative Finance 1 (1) (2001) 96. doi:10.1080/713665554.
URL http://www.informaworld.com/10.1080/713665554

[85] R. Reno, A closer look at the epps effect, International Journal of Theoretical and Applied Finance 6 (1) (2003) 87–102.

[86] T. Hayashi, N. Yoshida, On covariance estimation of non-synchronously observed diffusion processes, Bernoulli 11 (2) (2005) 359–379.

[87] P. Malliavin, M. E. Mancino, Fourier series method for measurement of multivariate volatilities, Finance and Stochastics 6 (1) (2002) 49–61. doi:10.1007/s780-002-8400-6.
URL http://dx.doi.org/10.1007/s780-002-8400-6

[88] G. Iori, O. V. Precup, Weighted network analysis of high-frequency cross-correlation measures, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) 75 (3) (2007) 036110–7. doi:{10.1103/PhysRevE.75.036110}.
URL http://link.aps.org/abstract/PRE/v75/e036110

[89] P. Gopikrishnan, B. Rosenow, V. Plerou, H. E. Stanley, Quantifying and interpreting collective behavior in financial markets, Physical Review E 64 (3) (2001) 035106, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevE.64.035106}.
URL http://link.aps.org/abstract/PRE/v64/e035106

[90] L. Laloux, P. Cizeau, J. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Physical Review Letters 83 (7) (1999) 1467, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevLett.83.1467}.
URL http://link.aps.org/abstract/PRL/v83/p1467

[91] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Physical Review Letters 83 (7) (1999) 1471, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevLett.83.1471}.
URL http://link.aps.org/abstract/PRL/v83/p1471

[92] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, H. E. Stanley, Random matrix approach to cross correlations in financial data, Physical Review E 65 (6) (2002) 066126, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevE.65.066126}.
URL http://link.aps.org/abstract/PRE/v65/e066126

[93] B. Podobnik, P. C. Ivanov, Y. Lee, A. Chessa, H. E. Stanley, Systems with correlations in the variance: Generating power law tails in probabilit
EPL (Europhysics Letters) 50 (6) (2000) 711–717.
URL http://www.iop.org/EJ/abstract/0295-5075/50/6/711

[94] J. D. Noh, Model for correlations in stock markets, Physical Review E 61 (5) (2000) 5981, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevE.61.5981}.
URL http://link.aps.org/abstract/PRE/v61/p5981

[95] P. Cizeau, M. Potters, J. Bouchaud, Correlation structure of extreme stock returns, Quantitative Finance 1 (2) (2001) 217. doi:10.1080/713665669.
URL http://www.informaworld.com/10.1080/713665669

[96] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, I. Zahed, Levy matrices and financial covariances, cond-mat/0103108.
URL http://arxiv.org/abs/cond-mat/0103108

[97] R. Mantegna, Hierarchical structure in financial markets, The European Physical Journal B - Condensed Matter and Complex Systems 11 (1) (1999) 193–197. doi:10.1007/s100510050929.
URL http://dx.doi.org/10.1007/s100510050929

[98] G. Bonanno, F. Lillo, R. N. Mantegna, High-frequency cross-correlation in a set of stocks, Quantitative Finance 1 (1) (2001) 96. doi:10.1080/713665554.
URL http://www.informaworld.com/10.1080/713665554

[99] L. Kullmann, J. Kertesz, R. N. Mantegna, Identification of clusters of companies in stock indices via potts super-paramagnetic transitions, Physica A: Statistical Mechanics and its Applications 287 (3-4) (2000) 412419.

[100] J. Onnela, A. Chakraborti, K. Kaski, J. Kertsz, A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, Physical Review E 68 (5) (2003) 056110, copyright (C) 2009 The American Physical Society; Please report any problems to prola@aps.org. doi:{10.1103/PhysRevE.68.056110}.
URL http://link.aps.org/abstract/PRE/v68/e056110

[101] J.-P. Onnela, Taxonomy of financial assets, Master's thesis, Helsinki University of Technology, Espoo, Finland (2000).

[102] J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertsz, Dynamic asset trees and black monday, Physica A: Statistical Mechanics and its Applications 324 (1-2) (2003) 247–252. doi:{10.1016/S0378-4371(02)01882-4}.
URL http://www.sciencedirect.com/science/article/B6TVG-47MCCB5-8/2/1edb1d15bfacd37b5f7c0d8a36307745

[103] J. Onnela, A. Chakraborti, K. Kaski, J. Kertisz, Dynamic asset trees and portfolio analysis, The European Physical Journal B - Condensed Matter and Complex Systems 30 (3) (2002) 285–288. doi:10.1140/epjb/e2002-00380-9.
URL http://dx.doi.org/10.1140/epjb/e2002-00380-9
[104] G. Stigler, Public regulation of the securities markets, Journal of Business (1964) 117–142.
[105] B. LeBaron, Agent-based computational finance, Handbook of computational economics 2 (2006) 1187–1233.
[106] R. Cont, J. P. Bouchaud, Herd behavior and aggregate fluctuations in financial markets, Macroeconomic dynamics 4 (02) (2000) 170–196.
[107] M. Raberto, S. Cincotti, S. Focardi, M. Marchesi, Agent-based simulation of a financial market, Physica A: Statistical Mechanics and its Applications 299 (1-2) (2001) 319–327.
[108] T. Lux, M. Marchesi, G. Italy, Volatility clustering in financial markets, Int. J. Theo. Appl. Finance 3 (2000) 675–702.
[109] C. Chiarella, G. Iori, A simulation analysis of the microstructure of double auction markets, Quantitative Finance 2 (5) (2002) 346–353.
[110] P. Bak, M. Paczuski, M. Shubik, Price variations in a stock market with many agents, Physica A: Statistical and Theoretical Physics 246 (3-4) (1997) 430 – 453. doi:DOI:10.1016/S0378-4371(97)00401-9.
URL http://www.sciencedirect.com/science/article/B6TVG-3SPXB8W-29/2/a3cbdbf6bb831a3c77d7275f25ce911d
[111] S. Maslov, Simple model of a limit order-driven market, Physica A: Statistical Mechanics and its Applications 278 (3-4) (2000) 571–578.
[112] F. c. v. Slanina, Mean-field approximation for a limit order driven market model, Phys. Rev. E 64 (5) (2001) 056136. doi:10.1103/PhysRevE.64.056136.
[113] T. Preis, S. Golke, W. Paul, J. J. Schneider, Statistical analysis of financial returns for a multiagent order book model of asset trading, Physical Review E 76 (1).
[114] R. Cont, Volatility clustering in financial markets: Empirical facts and Agent-Based models, in: Long Memory in Economics, 2007, pp. 289–309.
[115] N. Hakansson, A. Beja, J. Kale, On the feasibility of automated market making by a programmed specialist, Journal of Finance (1985) 1–20.
[116] M. Cohen, Report of the special study of the securities markets of the securities and exchanges commission, u.S. Congress, 88th Cong., 1st sess., H. Document 95 (1963).
[117] M. H. Cohen, Reflections on the special study of the securities markets, speech at the Practising Law Intitute, New York (May 1963).
[118] F. Slanina, Critical comparison of several order-book models for stock-market fluctuations, The European Physical Journal B-Condensed Matter and Complex Systems 61 (2) (2008) 225–240.
[119] D. K. Gode, S. Sunder, Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality, The Journal of Political Economy 101 (1) (1993) 119–137.
URL http://www.jstor.org/stable/2138676
[120] D. Cliff, J. Bruten, Zero is not enough: On the lower limit of agent intelligence for continuous double auction markets, Tech. Rep. HPL-97-141, Hewlett-Packard Laboratories, Bristol, UK (1997).
[121] R. Cont, S. Stoikov, R. Talreja, A Stochastic Model for Order Book Dynamics, SSRN eLibrary.
[122] E. Smith, J. D. Farmer, L. Gillemot, S. Krishnamurthy, Statistical theory of the continuous double auction, Quantitative Finance 3 (2003) 481–514.
[123] A. Kirman, Whom or what does the representative individual represent?, The Journal of Economic Perspectives (1992) 117–136.
[124] A. Kirman, Ants, rationality, and recruitment, The Quarterly Journal of Economics (1993) 137–156.
[125] A. Kirman, Reflections on interaction and markets, Quantitative Finance 2 (5) (2002) 322–326.
[126] P. Erdos, A. Renyi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci 5 (1960) 17–61.
[127] A. Kirman, in: R. Frydman, E. Phelps (Eds.), Individual Forecasting and Aggregate Outcomes: Rational Expectations Examined, Cambridge University Press, 1983, Ch. On mistaken beliefs and resultant equilibria.
[128] F. Shostak, Quart. J. Aust. Economics 3 (2000) 69.
[129] R. Gibrat, Les inégalités économiques.
[130] A. Dragulescu, V. M. Yakovenko, Exponential and power-law probability distributions of wealth and income in the united kingdom and the un
Physica A: Statistical Mechanics and its Applications 299 (1-2) (2001) 213–221. doi:{10.1016/S0378-4371(01)00298-9}.
URL http://www.sciencedirect.com/science/article/B6TVG-447MXW6-11/2/0799e7f0069143fb28798986c89631c2
[131] A. Drgulescu, V. Yakovenko, Evidence for the exponential distribution of income in the USA, The European Physical Journal B 20 (4) (2001) 5 pages. doi:{10.1007/PL00011112}.
[132] Y. Fujiwara, W. Souma, H. Aoyama, T. Kaizoji, M. Aoki, Growth and fluctuations of personal income, Physica A: Statistical Mechanics and its Applications 321 (3-4) (2003) 598–604. doi:{10.1016/S0378-4371(02)01663-1}.
URL http://www.sciencedirect.com/science/article/B6TVG-47CWR8T-G/2/3dcceaf30b1f46eb512529937272ca38
[133] V. M. Yakovenko, J. J. Barkley Rosser, Statistical mechanics of money, wealth, and income, arXiv:0905.1518.
[134] T. Lux, Emergent statistical wealth distributions in simple monetary exchange models: A critical review, in: A. Chatterjee, S.Yarlagadda, B. K. Chakrabarti (Eds.), Econophysics of Wealth Distributions, Springer, 2005, p. 51.
[135] A. Chatterjee, B. Chakrabarti, Kinetic exchange models for income and wealth distributions, Eur. Phys. J. B 60 (2007) 135. doi:10.1140/epjb/e2007-00343-8.
[136] M. Levy, S. Solomon, Power laws are logarithmic boltzmann laws, Int. J. Mod. Phys. C 7 (1996) 595.
[137] D. Sornette, Multiplicative processes and power laws, Phys. Rev. E 57 (1998) 4811.
[138] Z. Burda, J. Jurkiewics, M. A. Nowak, Is Econophysics a solid science?, Acta Physica Polonica B 34 (2003) 87.
[139] J. Angle, The surplus theory of social stratification and the size distribution of personal wealth, Social Forces 65 (1986)

293.
URL http://www.jstor.org
[140] S. Solomon, M. Levy, Spontaneous scaling emergence in generic stochastic systems.
URL arXiv:adap-org/9609002
[141] M. Levy, S. Solomon, G. Ram, Dynamical, Int. J. Mod. Phys. C 7 (1996) 65.
[142] S. Ispolatov, P. L. Krapivsky, S. Redner, Wealth distributions in asset exchange models, Eur. Phys. J. B 2 (1998) 267.
[143] A. K. Gupta, Money exchange model and a general outlook, Physica A 359 (2006) 634.
[144] J. P. Bouchaud, M. Mezard, Wealth condensation in a simple model of economy, Physica A 282 (2000) 536.
[145] A. Drgulescu, V. Yakovenko, Statistical mechanics of money, The European Physical Journal B 17 (4) (2000) 7 pages. doi:10.1007/s100510070114.
[146] J. C. Ferrero, The statistical distribution of money and the rate of money transference, Physica A 341 (2004) 575.
[147] F. Slanina, Inelastically scattering particles and wealth distribution in an open economy, Phys. Rev. E 69 (2004) 046102.
[148] P. Repetowicz, S. Hutzler, P. Richmond, Dynamics of money and income distributions, Physica A 356 (2005) 641.
[149] S. Cordier, L. Pareschi, G. Toscani, On a kinetic model for a simple market economy, J. Stat. Phys. 120 (2005) 253.
[150] D. Matthes, G. Toscani, On steady distributions of kinetic models of conservative economies, J. Stat. Phys. 130 (2007) 1087.
[151] B. Düring, G. Toscani, Hydrodynamics from kinetic models of conservative economies, Physica A 384 (2007) 493.
[152] B. Düring, D. Matthes, G. Toscani, Kinetic equations modelling wealth redistribution: A comparison of approaches, Phys. Rev. E 78 (2008) 056103.
[153] E. Scalas, U. Garibaldi, S. Donadio, Statistical equilibrium in simple exchange games i - methods of solution and application to the bennati-dragulescu-yakovenko (bdy) game, Eur. Phys. J. B 53 (2006) 267.
[154] E. Scalas, U. Garibaldi, S. Donadio, Erratum. statistical equilibrium in simple exchange games i, Eur. Phys. J. B 60 (2007) 271.
[155] U. Garibaldi, E. Scalas, P. Viarengo, Statistical equilibrium in simple exchange games ii. the redistribution game, Eur. Phys. J. B 60 (2007) 241.
[156] J. Angle, The statistical signature of pervasive competition on wage and salary incomes, J. Math. Sociol. 26 (2002) 217.
[157] J. Angle, The inequality process as a wealth maximizing process, Physica A 367 (2006) 388.
[158] A. Chakraborti, B. Chakrabarti, Statistical mechanics of money: how saving propensity affects its distribution, The European Physical Journal B - Condensed Matter and Complex Systems 17 (1) (2000) 167–170. doi:10.1007/s100510070173.
URL http://dx.doi.org/10.1007/s100510070173
[159] A. Chakraborti, Distributions of money in model markets of economy, International Journal of Modern Physics C-Physics and Computer 13 (10) (2002) 1315–1322.
[160] B. Hayes, Follow the money, Am. Sci. 90 (5).
[161] A. Chatterjee, B. K. Chakrabarti, S. S. Manna, Money in Gas-Like markets: Gibbs and pareto laws, Physica Scripta T106 (2003) 36–38.
URL http://www.iop.org/EJ/abstract/1402-4896/2003/T106/008/
[162] A. Das, S. Yarlagadda, A distribution function analysis of wealth distribution.
URL arxiv.org:cond-mat/0310343
[163] N. Scafetta, S. Picozzi, B. J. West, An out-of-equilibrium model of the distributions of wealth (2004).
URL cond-mat/0403045
[164] J. R. Iglesias, S. Goncalves, S. Pianegonda, J. L. Vega, G. Abramson, Wealth redistribution in our small world, Physica A 327 (2003) 12.
[165] J. R. Iglesias, S. Goncalves, G. Abramsonb, J. L. Vega, Correlation between risk aversion and wealth distribution, Physica A 342 (2004) 186.
[166] M. Ausloos, A. Pekalski, Model of wealth and goods dynamics in a closed market, Physica A 373 (2007) 560.
[167] E. Bennati, La simulazione statistica nell'analisi della distribuzione del reddito: modelli realistici e metodo di Monte Carlo, ETS Editrice, Pisa, 1988.
[168] E. Bennati, Un metodo di simulazione statistica nell'analisi della distribuzione del reddito, Rivista Internazionale di Scienze Economiche e Commerciali 35 (1988) 735–756.
[169] E. Bennati, Il metodo Monte Carlo nell'analisi economica, in: Rassegna di lavori dell'ISCO X, Vol. 31, 1993.
[170] J. Angle, The surplus theory of social stratification and the size distribution of personal wealth, in: Proceedings of the American Social Statistical Association, Social Statistics Section, Alexandria, VA, 1983, p. 395.
[171] A. C. Silva, V. M. Yakovenko, Temporal evolution of the 'thermal' and 'superthermal' income classes in the usa during 1983-2001, Europhysics Letters 69 (2005) 304.
[172] X. Sala-i Martin, S. Mohapatra, Poverty, inequality and the distribution of income in the g20, Columbia University, Department of Economics, Discussion Paper Series.
URL http://www.columbia.edu/cu/economics/discpapr/DP0203-10.pdf
[173] I. Salai-Martin, The world distribution of income, NBER Working Paper Series.
URL http://www.nber.org/papers/w8933
[174] H. Aoyama, W. Souma, Y. Fujiwara, Growth and fluctuations of personal and company's income, Physica A 324 (2003) 352.
[175] M. Patriarca, A. Chakraborti, K. Kaski, Gibbs versus non-Gibbs distributions in money dynamics, Physica A: Statistical Mechanics and its Applications 340 (1-3) (2004) 334–339. doi:10.1016/j.physa.2004.04.024.
URL http://www.sciencedirect.com/science/article/B6TVG-4C8DCSG-B/2/79e72f35c2092734291e2d1133e990db

[176] M. Patriarca, A. Chakraborti, K. Kaski, Statistical model with a standard gamma distribution, Phys. Rev. E 70 (2004) 016104.
[177] M. Patriarca, E. Heinsalu, A. Chakraborti, The abcd's of statistical many-agent economy models.
URL arxiv.org:physics/0611245
[178] A. Chakraborti, M. Patriarca, Gamma-distribution and wealth inequality, Pramana J. Phys 71 (2) (2008) 233.
[179] J. Angle, Deriving the size distribution of personal wealth from *the rich get richer, the poor get poorer*, J. Math. Sociol. 18 (1993) 27.
[180] A. Chatterjee, B. K. Chakrabarti, S. S. Manna, Pareto law in a kinetic model of market with random saving propensity, Physica A: Statistical Mechanics and its Applications 335 (1-2) (2004) 155–163. doi:10.1016/j.physa.2003.11.014.
URL http://www.sciencedirect.com/science/article/B6TVG-4B6TTVK-5/2/10e6f843c0360ec0e2dd9f33456b4d4b
[181] P. K. Mohanty, Generic features of the wealth distribution in ideal-gas-like markets, arXiv:physics/0603141.
[182] A. Das, S. Yarlagadda, An analytic treatment of the Gibbs–Pareto behavior in wealth distribution, Physica A 353 (2005) 529.
[183] A. Chatterjee, S. Yarlagadda, B. K. Chakrabarti (Eds.), Econophysics of Wealth Distributions, Springer, 2005.
[184] A. Chatterjee, B. K. Chakrabarti, R. B. Stinchcombe, Master equation for a kinetic model of trading market and its analytic solution, Phys. Rev. E 72 (2005) 026126.
[185] M. Patriarca, A. Chakraborti, K. Kaski, G. Germano, Kinetic theory models for the distribution of wealth: Power law from overlap of exponentials, in: A. Chatterjee, S.Yarlagadda, B. K. Chakrabarti (Eds.), Econophysics of Wealth Distributions, Springer, 2005, p. 93.
[186] M. Patriarca, A. Chakraborti, G. Germano, Influence of saving propensity on the power law tail of wealth distribution, Physica A 369 (2006) 723.
[187] M. Patriarca, A. Chakraborti, E. Heinsalu, G. Germano, Relaxation in statistical many-agent economy models, Eur. J. Phys. B 57 (2007) 219.
[188] A. K. Gupta, Relaxation in the wealth exchange models, Physica A 387 (2008) 6819.
[189] W. B. Arthur, Inductive reasoning and bounded rationality, The American Economic Review 84 (2) (1994) 406–411, ArticleType: primary_article / Issue Title: Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association / Full publication date: May, 1994 / Copyright 1994 American Economic Association. doi:10.2307/2117868.
URL http://www.jstor.org/stable/2117868
[190] D. Challet, M. Marsili, Y.-C. Zhang, Minority Games, Oxford University Press, Oxford, UK, 2005.
[191] D. Challet, Z. Y.-C., Physica A 246 (1997) 407.
[192] A. A. C. Coolen, The Mathematical Theory of Minority Games, Oxford University Press, Oxford, UK, 2004.
[193] N. F. Johnson, P. Jefferies, P. M. Hui, Financial Market Complexity, Oxford University Press, Oxford, UK, 2003.
[194] M. Sysi-Aho, A. Chakraborti, K. Kaski, Physica Scripta T 106 (2003) 32.
[195] D. Challet, Y.-C. Zhang, Physica A 256 (1998) 514.
[196] Y. Li, R. Riolo, R. Savit, Evolution in minority games. (I). games with a fixed strategy space, Physica A: Statistical Mechanics and its Applications 276 (1-2) (2000) 234–264. doi:{10.1016/S0378-4371(99)00436-7}.
URL http://www.sciencedirect.com/science/article/B6TVG-403799X-G/2/d68cbf33e7fb39890e84d700d3f2eb17
[197] Y. Li, R. Riolo, R. Savit, Evolution in minority games. (II). games with variable strategy spaces, Physica A: Statistical Mechanics and its Applications 276 (1-2) (2000) 265–283. doi:{10.1016/S0378-4371(99)00435-5}.
URL http://www.sciencedirect.com/science/article/B6TVG-403799X-H/2/82794e404faeaf9833ba66518a8c81c3
[198] M. Sysi-Aho, A. Chakraborti, K. Kaski, Physica A 322 (2003) 701.
[199] M. Sysi-Aho, A. Chakraborti, K. Kaski, Eur Phys J B 34 (2003) 701.
[200] M. Sysi-Aho, A. Chakraborti, K. Kaski, Phys. Rev. E 69 (2004) 036125.
[201] D. Challet, Y. C. Zhang, Emergence of cooperation and organization in an evolutionary game, Physica A: Statistical and Theoretical Physics 246 (3-4) (1997) 407–418. doi:{10.1016/S0378-4371(97)00419-6}.
URL http://www.sciencedirect.com/science/article/B6TVG-3SPXB8W-27/2/22d76fc5d413e3fdc9095ca7093654ca
[202] M. Marsili, Physica A 299 (2001) 93.
[203] D. Challet, M. Marsili, Y.-C. Zhang, Physica A 299 (2001) 228.
[204] D. Challet, A. Chessa, M. Marsili, Y.-C. Zhang, Quantitative Finance 1 (2000) 168.
[205] D. Challet, M. Marsili, Y.-C. Zhang, Physica A 276 (2000) 284.
[206] A. D. De Martino, M. Marsili, Journal of Physics A: Mathematics and General 34 (2001) 2525.
[207] D. Challet, M. Marsili, Physical Review E 60 (1999) R6271.
[208] R. Savit, R. Manuca, R. Riolo, Physical Review Letter 82 (1999) 2203.
[209] D. Challet, M. Marsili, R. Zecchina, Physical Review Letter 84 (2000) 1824.
[210] A. Cavagna, J. P. Garrahan, I. Giardina, D. Sherrington, Physical Review Letter 83 (1999) 4429.
[211] N. F. Johnson, P. M. Hui, R. Johnson, T. S. Lo, Physical Review Letter 82 (1999) 3360.
[212] N. F. Johnson, M. Hart, P. M. Hui, Physica A 269 (1999) 1.
[213] M. L. Hart, P. Jefferies, P. P. M., Hui, N. F. Johnson, European Physical Journal B 20 (2001) 547.
[214] G. Iori, S. Jafarey, F. G. Padilla, Systemic risk on the interbank market, Journal of Economic Behavior & Organization 61 (4) (2006) 525–542.
URL http://www.sciencedirect.com/science/article/B6V8F-4KF781F-2/2/043e12c00bdd40ebc40c2b79e63b3533
[215] S. Thurner, J. D. Farmer, J. Geanakoplos, Leverage causes fat tails and clustered volatility, Preprint.