University of Genoa
Faculty of Engineering
Ph.D. Program in Information and Communication Technologies
Curriculum on Electronics and Computer Engineering

Dissertation

MODELING AND STATISTICAL ANALYSIS OF FINANCIAL MARKETS AND FIRM GROWTH

by

LINDA PONTA

Supervisor: Professor Silvano Cincotti

SSD ING-IND/35

ABSTRACT

The quantitative study of financial markets is more and more widespread due to their growing importance in the economy and everyday life. Financial markets can be viewed as real-world complex dynamical systems which are continually evolving, have significant practical importance and produce an enormous amount of data recording the aggregate action of many participants. In recent years, the quantitative approach to the modeling of stock markets has greatly benefitted from methods and tools developed in the domains of engineering and physics. Indeed, the recent availability of large and high quality data sets has transformed finance into the most quantitative social science.

A field called "Financial Engineering" has appeared, gathering scientists experienced in non-linear, deterministic and stochastic dynamical systems and interested in modeling and forecasting financial markets. Many world-class US universities have opened laboratories, departments or courses on financial engineering.[1] Moreover, the study by means of simulations of complex systems, characterized by the interaction of a large amount of heterogeneous units, needs one to develop flexible software frameworks based on advanced software engineering techniques. The growing interest in this kind of approach is further confirmed by an emerging field of statistical physics, called "Econophysics".

The first part of this Thesis presents an artificial financial market conceived as a computational experimental facility where different experiments can be performed, hypotheses verified and conjectures validated. An artificial financial market is an agent-based computer simulator of a financial market. Artificial financial markets model financial interactions from the bottom up by means of a large number of interacting agents. This approach relies heavily on computational tools in order to avoid the restrictions of analytical methods.

The analysis of financial markets through the construction of artificial markets aims to explain the emergence of the characteristic statistical properties of asset prices on the basis of hypotheses on traders behavior, market microstructure and economic environment. These problems are usually too complex to be treated analytically, thus computer simulations have to be employed. This approach involves the identification of an interactions system that allows the researcher to generate financial time series in order to study the relationships between the elements of the system and market results.

The dynamics of a financial market depends on the interactions between the rules defining the trading mechanism and the behavioral assumptions about the agents population. Therefore, building an artificial market means to determine the trading

[1] See, for instance, the Operations Research and Financial Engineering Department of Princeton University (`http://www.orfe.princeton.edu/`), the Laboratory for Financial Engineering at MIT (`http://lfe.mit.edu/`) and the Master degree on Financial Engineering at Berkeley (`http://www.haas.berkeley.edu/MFE/`).

rules defining the price formation process on one side, and to specify the trading strategies followed by the agents on the other side. Using computer simulation, this methodology opens the possibility to analyze the impact of interactions rules which are analytically intractable. Thus, this approach allows the investigation of issues fundamental to understand how financial markets work.

In details, the statistical properties of high-frequency data are investigated by means of computational experiments performed with the Genoa Artificial Stock Market - GASM. In the market model, heterogeneous agents trade one risky asset in exchange for cash. Agents have zero intelligence and issue random limit or market orders depending on their budget constraints. The price is cleared by means of a limit order book. The order generation is modelled with a renewal process where the distribution of waiting times between two consecutive orders is a Poisson process, a Weibull distribution and a mixture of Poisson processes. Results point out that, according to the empirical evidences, only a mixture of Poisson processes does not reject the hypothesis of Weibull distribution for the waiting times between two consecutive market orders. Moreover, the mechanism of the limit order book is able to recover fat tails in the distribution of price returns without ad-hoc behavioral assumptions regarding agents; moreover, the kurtosis of the return distribution depends also on the renewal process chosen for orders.
Furthermore, a multi-assets artificial financial market has been considered. The financial market is populated by zero-intelligence traders with finite financial resources. The market is characterized by different types of stocks representing firms operating in different sectors of the economy. Zero-intelligence traders follow a random allocation strategy which is constrained by finite resources, past market volatility and allocation universe. Within this framework, stock price processes exhibit volatility clustering, fat-tailed distribution of returns and reversion to the mean. Moreover, the cross-correlations between returns of different stocks is studied using methods of random matrix theory. The probability distribution of eigenvalues of the cross-correlation matrix shows the presence of outliers, similar to those recently observed on real data for business sectors. It is worth noting that business sectors have been recovered as only consequence of random restrictions on the allocation universe of zero-intelligence traders. Moreover, in the presence of dividend paying stocks and in the case of cash inflow added to the market, the artificial stock market points out the same structural results obtained in the simulation without dividends. Such results suggest a significative structural influence on statistical properties of multi-assets stock market.
Finally, an information-based multi-assets artificial stock market has been modeled. The market is populated by heterogeneous agents that are seen as nodes of sparsely connected graphs. The market is characterized by different types of stocks and agents trade risky assets in exchange for cash. Beside the amount of cash and of stocks owned, each agent is characterized by sentiments. Moreover, agents share their sentiments by means of interactions that are determined by direct graphs. A central market maker (clearing house mechanism) determines the price processes for each stock at the in-

tersection of the demand and the supply curves. Within this framework, stock price processes exhibit main univariate stylized facts, i.e., unitary root price processes, fat tails of return distribution and volatility clustering. Furthermore, the multivariate price processes exhibit both static and dynamic stylized facts, i.e., the presence of static factors and common trends.

The second part of this Thesis applies statistical physics approaches to investigate quantitatively the size and growth of the complex system of business firms. It is studied the logarithm of the one-year growth rate of firms $g \equiv \log(S(t+1)/S(t))$ where $S(t)$ and $S(t+1)$ are the sizes of firms in the year t and $t+1$ measured in monetary values.

These sections review some main empirical results of firm size and firm growth based on different databases. They are (i) the size distribution of firms $P(S)$ are found to be skewed (either log-normal or power-law depending on the different databases), (ii) the growth-rate distributions of firms $P(g)$ are of Laplace form with power-law tails, (iii) the standard deviation of firm growth rates is related by a negative power-law to the firm size. The distribution of firm growth rates conditioned on firm size collapses onto a single curve, which implies that a universal functional form may exist to describe the distribution of firm growth rate. Moreover it is modeled the *Entry & Exit* effect and firm proportional growth using a generalized preferential attachment model. The model assumes that a new firm enters the system with a constant rate; a new unit enters/exits one of existing firms preferentially, that is, the larger firms have bigger probability to obtain the new unit, and the larger firms have bigger probability to lose a unit. The model successfully explains the observations: (i) the distribution of unit number $P(K)$ in a firm is power law with exponential tails, (ii) $P(g)$ is of Laplace form with power-law tails with exponent 3.

CONTENT

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Why a quantitative approach to economics and finance?

Introduction

The computer simulation of financial markets is a multidisciplinary topic which requires the involvement of strong skills from very different fields such as model building, complex systems, econometric and statistical data analysis, and software engineering. This approach is continuing to gain a wider and wider interest in the scientific community. Its birth can be traced back to the 1950's with the introduction of Monte Carlo computer simulations [144]; however, a special mention in the history of this field is devoted to the pioneering work done at the Santa Fe Institute in the early nineties [11, 117, 154].

1.1 Historical background

The study of economic systems by means of agent-based computational models is a relatively new field. In order to model the interactions of large numbers of agents characterized by heterogeneous behaviors, computationally expensive experiments are required. Only in the last fifteen years, the increasing availability of cheap computing power made them possible. However, the study of agents heterogeneity in economics from an analytical point of view has a long history and can be traced back to the early eighties.

1.1.1 Beyond the representative agent model

In the past two decades, economics has witnessed an important paradigmatic change. Analytically tractable models of the economy based on rational expectations theory and the representative agent hypothesis have gradually moved to a boundedly rational heterogeneous agents framework with a computationally oriented and evolutionary approach. In the rational expectations framework, agents can be heterogeneous in the sense that they might have different utilities functions, however there is not heterogeneity of beliefs as all agents share the same full knowledge of processes followed by the economy [126, 148]. The cooperative actions of all agents can be summed up to the action of the single representative agent under appropriate mathematical conditions. In a boundedly rational environment, agents heterogeneity is relative not only to different utilities functions, as in the rational expectations theory, but also to different beliefs and decision making processes of individual agents [172, 182].

This change has been characterized by some important closely related aspects: the representative agent has been substituted with interacting heterogeneous agents systems, full rationality with bounded rationality and a mainly analytical approach with a mainly computational one. Obviously, heterogeneity complicates the modelling framework and may lead to analytical intractability. A computational approach is thus better suited for investigating an heterogeneous agents world.

In finance, a similar paradigmatic change has occurred. From a perfect rational world where asset allocations and prices are completely determined by the perfect knowledge of the dynamic process governing them, to a boundedly rational world where heterogeneous agents employ competing trading strategies and prices may, at least partially, be driven by market psychology. The counterpart in finance of the rational expectations theory is the Efficient Market Hypothesis (EMH) formulated in the 1960's by Samuelson [171] and Fama [72]. The EMH states that, at each time t, the current price $p(t)$ of every financial asset reflects all relevant information for judging the future returns of those assets. The intuitive idea is that rational individual traders process the information that is available to them and take optimal positions in assets on the basis of this information. The market price for an asset then aggregates this diverse trader information and, in this sense, reflects all the available information. An implication of the EMH is that asset prices from periods prior to period t will not be of any help in predicting asset prices for periods $t+1$ and beyond, since this past price information is already fully reflected in period-t asset prices. While throughout of the 1970's, many economists seemed to accept the presumption that financial markets are efficient, i.e., that the EMH was satisfied, in the 1980's, a growing concern about the EMH began. In the last years, many empirical studies appeared showing evidences of violation of the EMH: price volatility was seen to be strongly temporally correlated [64], the largest price movements often occurred with little or no news [54], prices were not accurately reflecting rational valuations and sometimes diverged systematically from fundamental values originating bubbles [37, 179, 193].

1.1.2 The heterogeneous agents framework

The new heterogeneous agents approach was born in order to address the empirical findings which were well outside the theoretical framework of the representative agent and of the EMH. First, in the late seventies, Grossman and Stiglitz[1] proposed a sort

[1] In 2001, Prof. Joseph Stiglitz has been awarded of the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel for his pioneering studies of asymmetric information in markets.

of heterogeneous agents rational expectation model [89, 90]. The authors tried to complement the EMH, introducing the so-called "noisy rational expectations", in the sense that each agent is endowed with private information which is not fully reflected in asset prices, but all investors nevertheless exhibit rational behavior. Later, models departing totally from the EMH appeared. The new models were characterized by heterogeneous interacting agents grouped in different coexisting populations according to different beliefs or expectations.

In 1980, Beja and Goldman wrote a seminal paper [19] which departed from the idea of perfect rational investors and introduced an agents behavioral model characterized by the dynamics of value investors (or fundamentalists) and trend followers (or chartists). Value investors hold an asset when they think it is undervalued and short it when it is overvalued; trend followers hold an asset when the price has been going up and sell it when it has been going down. Beja and Goldman, assuming linear trading rules for each type of trading, showed that equilibrium in unstable when the fraction of trend followers is sufficiently high and introduced the idea that the interaction of different categories of agents could explain some typical features of the dynamics of prices. In 1986, in his presidential address at the American Finance Association, Fisher Black introduced to the mainstream of financial economists the notion of "noise trading" [24]. He suggested that irrational behavior should be part of a realistic theory of financial markets. A noise trader is a market participant with incorrect information who implements trades on the basis of this information under the false belief that this information is correct. Black argued that the presence of noise traders was necessary to explain the large volume of trading activity that occurs in financial markets. Without noise traders, there would be virtually no trades in individual shares. Rational investors trading with each other would realize that any other trader willing to pay a higher price for the asset must have superior information about the asset return and so they would not trade. The presence of noise traders provides rational traders with an incentive to gather information. Sophisticated traders can bring correct information to a market and exploit the profit opportunities created by the presence of noise traders. Sophisticated traders tend to move asset prices toward fundamental values. However, the continual presence of noise traders (particularly, the continual entrance of new noise traders) makes it difficult for other traders to discern who is a noise trader and who is acting on correct information. This can give a sophisticated trader a chance to make profits from his superior information for an extended period of time.

Stimulated by the 1987 Wall Street market crash, a symposium on market bubbles by the American Economic Association held in 1989 increased the interest on this subject among the economists. In the early nineties, two papers by De Long at al. [59, 60] studied in detail the lasting effects that a misperception of future returns by some agents has on the behavior of a financial market. In their model, furthermore, irrational agents may earn greater returns that rational ones. This result supported the plausible claim that the existence of different categories of traders cannot be dismissed as a transient phenomenon that market forces would get rid of. In the same years, Shleifer and Summers argued that, if noise traders take larger risky positions on average, because of erroneous beliefs, some noise traders can still profit in the market despite their erroneous beliefs [180]. The essential idea is that some noise traders holding large risky positions with high expected returns are lucky and manage to earn a high enough return rate on their wealth to enable them to remain in the market.

In 1992, Chiarella extended the work by Beja and Goldman introducing, still in a purely deterministic framework, non-linear chartist demand functions [43]. Chiarella showed that when the fraction of trend followers is sufficiently low, the price process is characterized by a stable equilibrium, but when the fraction exceeds a critical value, the equilibrium becomes unstable and prices exhibit periodic limit cycles. In a similar fashion, a related work, reported in a paper by Day and Huang [58] that appeared in 1990, presented a simple model where a “fundamentalist” and a “noise” representative agent interact in discrete time and cause persistent deviations from the equilibrium price which can be interpreted as a sequence of chaotic switchings between bull and bear market regimes.

An agent-based model of stock market makes both structural assumptions, regarding the market design and the economic environment, and behavioral assumptions describing the rule by which traders take their decisions. In 1993, a seminal paper by Gode and Sunder [84] established the allocative efficiency of a double auction regardless of the specific behavioral assumptions, by comparing the experimental performance of human traders with the simulated choices of zero-intelligence traders. However, in the following years, agent-based modelling of stock markets has explored a very rich set of behavioral assumptions, but has paid comparatively little attention to structural assumptions.

The landmark of most of the recent research has become the introduction and explicit consideration of various degrees of heterogeneity among agents, concerning either their trading strategies, or their learning abilities, or their adaptive responses. Genuine multi-agent models in which trading strategies co-evolve were described for example in a paper by Margarita and Beltratti [20] and in a paper by Blume and Easley [25], both in 1992. In the second half of nineties, Brock and Hommes, studying the effect of switching between trend following and value investing behavior, focused their work on the bifurcation structure and conditions under which the dynamics become chaotic [31, 32]. Some researchers introduced agents with learning and optimization capabilities. Arifovic studied the behavior of the exchange rate between two currencies when decision rules by agents are updated using genetic algorithms [8]. Beltratti et al. developed a stock market model based on artificial neural network agents [21].

While these works were done in a purely deterministic setting, studies along somewhat different lines appeared in the late nineties by Lux [127, 128] and Lux and Marchesi [129, 130]. These papers present a disequilibrium method of price formation, and focus their work on demonstrating agreement with more realistic price series. They also assume a stochastic value process, and stress the role of the market as a signal processor. The model by Lux and Marchesi is based on the interaction of three different populations of traders where agents can switch between trend and value investing due to contagion effects. The issue of herding behavior was already addressed some years before in a famous paper by Kirman who introduced an interesting explanation of herding in financial markets that relied on an analogy with the behavior of ants choosing between different food sources [112]. Recently, Cont and Bouchaud modelled herding behavior in financial markets by means of random graphs and lattices and reproduced the important stylized fact of fat tails in the distributions of returns [51]. Garibaldi et al. derived a model of herding from a well-known model of statistical physics, the Ehrenfest urn model [80]. In 1997 and 1998, Youssefmir et al. [208, 209], while avoiding the introduction of adaptative or learning capabilities, present a continuous time model where chartist agents extrapolate trends using technical signals (moving averages of different lengths) and value investors expect price to return quickly to fundamentals. Such a market exhibits a periodic sequence of rising and falling prices. Moreover, while bubbles are stable when a small noise is added, stronger exogenous shocks can abruptly invert the direction of the market.

In the last years, the number of models on heterogeneous agents continuously increased. Among them, some models still deserve a special mention. The first is the model proposed in 1997 by Bak et al. [14] where agents are endowed with finite resources. Second, the so-called "Minority Game" [40] that Challet and Zhang proposed in the same year of the model by Bak et al.. The minority game is a repeated game where N (odd) players have to choose one out of two alternatives at each time step. Those who happen to be in the minority win. The minority game is an abstraction of the famous El-Farol's bar problem [10] where the problem was that 100 people would like to go to a bar (El Farol) which is too crowded if there are more than 60 people. Since its introduction, this model has gained quite a lot of interest. Indeed, since it is quite simple, this model is very suitable for detailed numerical studies and analytical descriptions. Moreover, the so-called Minority Game is a game where agents are characterized by partial information and bounded rationality and it is believed that a game with a minority mechanism captures an essential feature of systems where agents compete for limited resources, like financial markets [39, 104].

Recently, Iori proposes a model where a trade friction, by responding to price movements, creates a feedback mechanism on future trading and generates volatility clustering [102]; a paper by Hommes [96] presents an analytical evolutionary model able to explain the stylized facts of real financial time series; Farmer and Joshi [73] argue that traders can be thought as signal processing devices and explicitly raise the problem of testing for cointegration between prices and fundamental value as a benchmark for the ability of the market price to track the dividend process underlying the fundamental value. Aoki introduces an heterogeneous agents framework to the modelling of macroeconomic fluctuations [5, 6].

1.1.3 The fully computational approach

The main feature of models discussed in the previous section is that, in order to be handled analytically or by means of simple numerical simulations, agents are often treated as a statistical ensemble and they do not have their own individuality. Agents heterogeneity is then addressed by means of different populations of traders, each population or group of agents being characterized by a particular behavior, but agents belonging to a group are not tracked individually. The switching between different trading strategies is resolved by changing the number of agents belonging to a particular population. These models are then kept quite simple but, consequently, they are often able to reproduce only partially the complex statistical features of financial time series.

In order to deal with the problem of very complex heterogeneity, which leaves the boundary of what can be handled analytically, some researchers decided to abandon the goal of analytical tractability and to embrace a complete computational approach. The most influential project in this respect is the Santa Fe artificial stock market [11, 117, 154] developed in the early 1990's. This simulated market bypasses some pitfalls of the representative agent approach by endowing agents with non-trivial capabilities: the actors have an internal representation of the world and can try to figure out the best optimizing model through continuous testing of alternative demand rules. Agents do not need to share information (other than price) and, when intensive learning is assumed, the market exhibits a complex behavior where even technical trading can be profitable in the short-run.

This line of research initiated by Arthur and his coauthors opened two promising ways to study financial markets:

- creation of artificial markets where heterogeneous, boundedly rational and possibly optimizing agents co-evolve dynamically;
- analysis of the computer generated price series by statistical tests to check for consistency with the known stylized facts in real data that challenge traditional models (e.g., fat tails of returns and volatility clustering).

Following the Santa Fe artificial market, a number of computer simulated markets later appeared. In 1998, Steiglitz and Shapiro [189] introduced an auction-mediated computer simulated market which was able to produce price bubbles and subsequent crashes by means of the presence of value-based and trend-based traders. In the same year, an agent-based microeconomic simulation model of the U.S. economy was developed at Sandia National Laboratories [15]. It was an ambitious computer simulator of a capitalistic economy with a detailed financial sector including a banking system and a bond market; agents learning was simulated by means of genetic algorithms.

In the last years, the interest on the subject of agent-based computational economics increased worldwide. Recently, a project for the development of an artificial market started at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology [42]. Two books have been published on this field [67, 121] and a number of studies appeared in the scientific literature; LeBaron [118] and Wan et al. [204] provide good reviews. Recently, four special issues on the subject of agent-based computational economics have been published by influential scientific journals. See the guest editorials by Tesfatsion [197–199] and by Lux and Marchesi [131]. In October 2001, a colloquium titled "Adaptive agents, intelligence, and emergent human organization. Capturing complexity through agent-based modeling" has been held at the National Academy of Science and Engineering of the United States of America. A special issue of the Proceedings of the Academy followed; see therein, e.g., the paper by Nigel and Bankes [83] and the paper by Kephart [111] on software issues. In the same issue, the papers by Tesfatsion [200] and by LeBaron [119] also deserve a mention.

This body of recent literature variously support the claim that heterogeneity of agents can produce endogenous price fluctuations with the same statistical features of financial time series. These heterogeneous agents market models can be classified with respect to how they describe trading strategies and learning algorithms of agents. Up to now, however, the literature on artificially simulated financial markets has only rarely addressed an explicit modeling of the market microstructure, favoring instead unrealistic approximating devices for the price formation mechanisms like some sort of Walrasian auctioner or market maker with unbounded liquidity. In some recent papers, Maslov [139], Matassini [141], Chiarella and Iori [44] suggest that an accurately modeled order-driven market can produce remarkably fat-tailed returns in a variety of simple settings even without the recourse to complicated behavioral hypotheses on agents.

1.2 Key issues in artificial markets

An agent model of an artificial market is characterized by a number of independent agents which interact by trading stocks and cash. There must be a system that allows agents to buy and sell, either directly with one another, by means of a central clearing mechanism, or through special dealer agents. Agents must encapsulate some decision process which they use to determine whether or not to trade, and at what price. The decision process may be fixed, or may be adaptive (they may learn how to trade over time). Decisions are based on information, usually at least some information on previous prices (endogenous information) but also often some exogenous information giving a (noisy) estimate of future yield. The share price itself indirectly gives some indication of other agents information, as the price reflects their decision processes. Last but least, artificial price series have to be characterized by the same statistical features, the so-called "stylized facts", of real stock price series; this being a necessary condition for the validation of the artificial market.

1.2.1 Stylized facts of real stock price series

Stock price series exhibit a number of statistical features which are common in all financial markets and for which a satisfactory explanation is still lacking in standard theories of financial markets. Pagan provides an authoritative survey of these salient features [152]. The main empirical regularities characterizing the univariate prices time series which an agent model of speculative activity should incorporate are: the unit-root property, the fat tails phenomenon and the volatility clustering.

1.2.1.1 Unit-root property

Denoting by $x(t)$ the value of a time series at time t, $x(t)$ is said to follow an unit root process if $x(t)$ can be expressed as an autoregressive process: $x(t) = \rho \cdot x(t-1) + \epsilon(t)$, where $\rho = 1$ and $\epsilon(t)$ represents a stationary stochastic increment. Using standard statistical procedures such as the Dickey-Fuller test [61, 62], one is usually unable to reject the null hypothesis of $\rho = 1$ for the time series of logarithms of prices, i.e., $x(t) = \log p(t)$, where $p(t)$ is the stock price at time t. Moreover, if $\epsilon(t)$ represents a white noise process, i.e., the increments $\epsilon(t)$ are independent and identically distributed, one is unable to reject the hypothesis that stock price prices follow a geometric random walk. If logs of prices obey a unit-root dynamics, stock prices are characterized by non stationarity and lack of predictability, while returns or differences of logs should be stationary, independent and identically distributed random variables. These empirical findings served as the pillars for the efficient market hypothesis (EMH) and the view of arbitrage-free financial markets [72, 171].

1.2.1.2 Fat tails phenomenon

In the last two decades, a great increase in the calculation and data storage capabilities of computers has permitted a deeper empirical analysis of financial data. The empirical analysis of a huge amount of data shows significative departures from the pillars of the classical financial mathematics theory (EMH). The first important departure regards the probability distribution of returns.

Empirical studies show that returns at weekly, daily and higher frequencies exhibit more probability mass in the tails and in the center of the distribution than does the standard Normal [30, 136, 137]. It is perhaps also remarkable that, besides this deviation from the Gaussian, the shape of the distribution usually appears well-behaved: namely, histograms of stock price returns mostly show a unimodal bell shape with, in most cases, only modest levels of negative skewness. The fat tails phenomenon has been identified with excessive fourth moments (leptokurtosis). Recent literature [87] provides evidences that the decline of probability mass in the outer parts follows a power law with a exponent in the range $3 \div 5$, whereas the Normal and a number of other often-used distributions have an exponential decline.

1.2.1.3 Volatility clustering

Another striking feature is the intermittent nature of the fluctuations of stock prices. Volatility is a measure of the amplitude of price fluctuations over a given time interval. Localized bursts of volatility, i.e., of the amplitude of fluctuations, can be identified. This fact, known as volatility clustering, is also evident in the slow decay of the autocorrelation function of the magnitude of price fluctuations, whereas the time series of raw fluctuations, i.e., fluctuations with their sign, exhibits insignificant autocorrelations. Moreover, the autocorrelation function of the daily volatility can be fitted by an inverse power law with a rather small exponent in the range $0.1 \div 0.3$, see Ref. [123, 137]. This suggests that there is no characteristic scale for volatility fluctuations: outbursts of market activity can persist for short times (a few hours), but also for much longer times (weeks or even months).

1.2.2 Definition of agents

In agent models, the definition of agents structure and decision making process is fundamental to the design. Heterogeneity of agents is a key requirement and provides the main source of dynamics in the agent model, as agents will not trade unless their estimates of gain (implicit or explicit) differ.

1.2.2.1 The decision making process

The decision making process of agents can be modeled by means of a preference function regarding the proportions of cash and stock they prefer to hold. The preference function may be implemented by a range of techniques, including simple functions, probabilistic functions, and artificial intelligence models such as neural networks and classifier systems. Frequently, the function is based on a forecasting method that predicts the share price in the future. Share price prediction may be performed using a simple regression function, linear equations, neural networks or other predictive techniques. If agents are not constrained by their money and stock reserves, the forecasting process can be used fairly directly to provide the preference function.

1.2.2.2 Information

The preference function takes as input some parameters and produces as output a decision. Most of the preference functions used in agent modeling need, as input, some information about the market. Endogenous information includes the current trading price and possibly previous prices, interest rate and dividend, in addition to information about the agent itself (cash and stock holding). Exogenous information is supposed to correspond to facts that, in a real market, would lead investors to alter their perception of likely future yield. It can be incorporated as an estimate of future dividends, corrupted with noise. Thus agents are characterized by the amount of available information and endogenous or exogenous information distinguishes an informed agent from an uninformed one.

1.2.2.3 The individual endowment

The individual endowment is the money and stocks that agents hold. Some agent models ignore the endowment, allowing agents to trade an unbounded amount of stock and to incur negative stock holdings and/or cash reserves. The inclusion of individual endowments makes the agent system more realistic. This imposes a budget constraint on the agents, as they can only buy and sell within the finite amounts of money and stock they hold. Thus the terms endowment and budget constraint are synonyms in this context.

1.2.2.4 Learning

Agents may possess the ability to learn from the success or failure of their previous investment decisions. Adaptation of the agents can be implemented via changes to their preference function. The learning procedures used in adaptive agent models invariably require some measure of performance, which is used to guide adaptation.

1.2.3 Market design

1.2.3.1 Traded securities

In real stock markets, investors decide how to distribute their investment between different financial instruments, including interest bearing cash accounts, and stocks in various companies. Agent models of stock markets typically allow a choice between interest-bearing cash investment and stocks.

Most agent models allow only a single stock as an alternative to cash. The interest rate is fixed, but the dividend on the stock may vary. This is sufficient to simulate many of the interesting dynamic features of a stock market, including the effects of market sentiment, while keeping the simulated system as simple as possible. The ratio of dividend to interest assigns a yield to the stock that can be used as one fundamental measure of stock value, and variation in the stock price alone is equivalent to variation in both dividend and interest rate (as it is the ratio that indicates which is better to hold).

1.2.3.2 Clearing mechanism

A clearing mechanism is necessary in order to process transactions made by agents. In most of the literature, there is no explicit description of the clearing method. Clearing of bids and offers is typically instantaneous and is performed by assuming zero aggregate excess demand or by means of a central market authority matching supply and demand.

Recently, an increasing attention has being devoted to the clearing mechanism as a possible source of the stylized facts of financial time series. In particular, order-driven artificial markets have appeared with an explicit modeling of the clearing process by means of a limit orders book.

1.2.4 Validation

Validation is a critical issue. The problem of validation characterize all types of economic models, however, there are some key issues which in particular affect artificial financial markets. Artificial markets can be characterized by a great number of parameters which might be utilized to fit any feature of actual data but often make the model exceedingly complicated, almost to the point of losing tractability.

Indeed, the computer modeler faces a different set of problems from the analytical modeler. Often the latter is faced with constraints about what can be done analytically. This pushes to keep the framework simple, but the simplicity is due to analytical tractability rather than economic structure. The computer modeler is free of these constraints, but this can be both a blessing and a curse, in that it can lead to overly complicated structures which are difficult to examine. Thus, a move towards keeping the models relatively simple without cutting out their key components is often necessary.

The problem of validation can be addressed with the requirement of an high similarity with the statistical properties of empirical data, also regarding different data sets and different time horizons. However, the use of experimental data, and the potential design of experiments to falsify computer models remains a largely unexplored area.

1.3 Major contributions of the GASM

The Genoa Artificial Stock Market (GASM) is an unique and innovative artificial financial market. A number of features deserve special mention:

- the finiteness of agents financial resources;
- the clearing mechanism;
- the random background trading with the volatility feedback;
- the agents wealth distribution;
- the number of agents engaged in trading.

1.3.0.1 The finiteness of agents financial resources

In the present literature on artificial financial markets, the attention of researchers is mainly focused on modeling agent optimization and learning capabilities. Little effort has been devoted to study how the market microstructure and the macroeconomic environment affect market prices. The Genoa Artificial Stock Market has been conceived mainly to address these problems.

The hypothesis is that, in the long run, is the interplay between the flow of cash in and out financial markets and the creation/destruction of stocks that determine price trends. In order to address this issue by means of the GASM, agents are endowed with finite financial resources (cash and stocks) and the system is able to keep track of every agent's portfolio. The finiteness of agents' financial resources is an innovative and essential feature of the GASM. It poses significant constraints on possible trading strategies if different populations of agents should coexist indefinitely. Actually, the study of the interplay between different trading strategies in a changing market environment is one of the main results of this thesis. Results show that a trading strategy cannot be judged only on the basis of the strategy itself, but its success depends also on the market conditions.

1.3.0.2 The clearing mechanism

The price formation process is another essential feature of GASM. It is based on a realistic auction-type order matching mechanism that allows to define a demand-supply schedule. The demand-supply schedule is an essential feature since price fluctuations are due to an imbalance between demand and supply. As market must clear, this means that somehow the "intention" to buy or sell must be modified to allow orders to match. This is the essence of the demand-supply schedule. As demonstrated in many studies on market microstructure, e.g. see [151], the details of the order matching process have a bearing on both price setting and price-volume relationships.

1.3.0.3 The random background trading with the volatility feedback

In the GASM, the great majority of agents does not follow complex trading strategies but issue random buy and sell orders which are constrained by the limited financial resources and by the past price volatility. This gives birth to a a background trading which is able to produce a price process characterized by the main stylized facts observed in real markets, i.e., fat-tailed distribution of returns and volatility clustering.

The background trading is originated by the random buy and sell orders of agents. However, although random, the background trading is characterized by different states of volatility. Volatility is uncertainty and, in volatile markets, agent uncertainty on asset market prices grows. The link between nervous (i.e. volatile) markets and agent uncertainty is modeled through the ordering mechanism. To represent this, orders are issued at random, but their limit price exhibits a functional dependence on past price volatility. In periods of high volatility, traders are more nervous and therefore allow for wider price limits in order to get their trades done quickly. The assumption is that trades have to be done for exogenous reasons and therefore traders want to execute trades at the best possible price. Fearing large market movements, they allow more freedom in the setting of limit prices.

It is worth noting that a population of agents that issue random orders in a limited resource market produces a price process with a Gaussian distribution and mean-reverting behavior, but neither fat tails nor volatility clustering. The introduction of the volatility feedback mechanism results in a price process exhibiting fat tails, zero autocorrelation of returns and the serial autocorrelation of volatility.

1.3.0.4 The agents wealth distribution

The agents wealth distribution is another important structural element of the GASM; it follows a Pareto distribution and gives an important contribution to the fat tails of the price returns distribution. This wealth distribution is also a very realistic assumption. In fact, it is well known empirically that in an economy the wealth distribution tends to follow a Pareto inverse power law [122, 155]. It has been also demonstrated theoretically that auto-catalytic processes naturally lead to inverse power law distribution of wealth [28, 97, 98, 146]. The wealth of agents in the GASM is indeed governed by an auto-catalytic process which explains the emergence and conservation of inverse power law distributions.

1.3.0.5 The number of agents engaged in trading

The number of agents engaged in trading at each moment is a small fraction, randomly selected, of the total number of agents. This is a realistic feature also present in real markets where prices are set by transactions that involve only a small fraction of the market participants. The notion that the entire population of investors is continuously engaged in trading is simply unrealistic as trading cost would skyrocket. Even professional fund managers tend to limit their trading activity to only a few trades per day and this only to tune portfolios that remain substantially stable for periods of the order of weeks. The above implies that the consideration of the “thermodynamic limit” of markets, i.e. an infinite number of traders, is simply unrealistic and possibly misleading. Finite size effects in real markets are not an artifact but a real feature. The interaction of a small set of agents sets the “wealth” of the entire market.

1.4 The simulation software

Agent-based models are naturally implemented in object-oriented programming (OOP) languages [66, 83]. In the OOP framework, each agent is an object and the object related to each agent is a particular instance of the class describing all agents characterized by the same data structures and the same behavior. The wealth and the number of shares of each agent are naturally the instance variables of the object and the trading strategies are implemented by methods. The hierarchical organization of classes, the properties of inheritance and polymorphism of variables and methods belonging to classes of the same hierarchy allow to develop compact and logical programming framework which are also very flexible for variations and extensions.

From the early 1990's, most agent models have been developed in OOP languages such as C++, Java and Smalltalk. Recently, standard libraries that allow programmers to develop simulation environments have been made available. Among them, the most famous are REPAST [2] developed in Java at the University of Chicago, ASCAPE[3], also in Java, of The Brookings Institution, and SWARM [4] in C++ developed at the Santa Fe Institute.

The GASM simulator has been developed in Smalltalk language. Smalltalk proved very suited for this kind of application. It is also worth noting that the first version of the Santa Fe artificial stock market was written in Objective C, which is C with embedded Smalltalk instructions. Using Smalltalk, in fact, it is possible to develop complex systems and to make substantial modifications to them very quickly, not jeopardizing quality. As regarding performance, though Smalltalk is an interpreted language, the simulation speed is enough for the purposes of the simulator and there has not been any need to trade Smalltalk flexibility for further speed.

The simulator has been implemented following Extreme Programming (XP) as development process. Extreme Programming is a software engineering technique developed by Kent Beck in the late nineties [18, 192]. In the XP framework, the software development process is incremental and iterative. The software system is developed step by step, adding new features at each step. These features come from requirements gathering in the form of user stories, i.e., short sentences describing the behavior of the system and/or its interactions with the users. User stories are collected, sorted by importance and risk and then implemented. A working system, implementing a subset of the required features, is released as soon as possible and new releases are made every step.

Testing and refactoring are of paramount importance in the process [76]. Indeed, an automatic test suite for all objects present in the system has been developed. The suite has been continuously increased as new objects and new features were added to the system, and is run after every modification, to ensure that the modification does not break system integrity.

A revised version of XP has been used [9]. It proved effective and enabled to successfully deliver a first release of the system in a few months.

Finally, by means of OOP and XP, the system is not a stand-alone application optimized for the presented models, but is an evolving system, able to be continuously modified and updated. For instance, the system can manage a practically unlimited number of different kinds of securities, and could be used also as an engine for a trading game, or for implementing real online trading. As another example, the modification of adding more intelligence to traders, like the capability to learn and to take decisions by means of genetic algorithms as in the seminal stock market model developed at the

[2] `http://repast.sourceforge.net/`

[3] `http://www.brook.edu/dybdocroot/es/dynamics/models/ascape/`

[4] `http://www.swarm.org`

Santa Fe Institute, would amount simply to add the objects supporting this kind of computation and to introduce a new subclass of the superclass Agent. The remaining parts of the system would remain unaffected, and trading could start immediately with the new "intelligent" traders.

Outline

After this introductory chapter, the rest of this thesis will mainly discuss the mathematical formulation of the GASM and the computational experiments. The software implementation based on XP practices greatly simplified the growing complexity that characterized the development process. In the end the statistical analysis and the modeling of the firm growth problem is presented.

Chapter 2

Modeling and Statistical Analysis of a Single Asset Artificial Stock Market

Overview

In the first part of this chapter, the statistical properties of high-frequency data are investigated by means of computational experiments performed with the Genoa Artificial Stock Market (Raberto et al. 2001, 2003, 2004). In the market model, heterogeneous agents trade one risky asset in exchange for cash. Agents have zero intelligence and issue random limit or market orders depending on their budget constraints. The price is cleared by means of a limit order book. The order generation is modeled with a renewal process where the distribution of waiting times between two consecutive orders is a Weibull distribution. This hypothesis is based on recent empirical investigation made on high-frequency financial data (Mainardi at al. 2000, Raberto et al. 2002, Scalas et al. 2003). How the statistical properties of prices and of waiting times between transactions are affected by the particular renewal process chosen for orders is investigated. Results point out that the mechanism of the limit order book is able to recover fat tails in the distribution of price returns without ad-hoc behavioral assumptions regarding agents; moreover, the kurtosis of the return distribution depends also on the renewal process chosen for orders. As regarding the renewal process underlying trades, in the case of exponentially distributed order waiting times, also trade waiting times are exponentially distributed. Conversely, if order waiting times follow a Weibull, the same does not hold for trade waiting times.

In the second part of the chapter, empirical analysis and computational experiments are presented on high-frequency data for a double-auction (book) market. Main objective is to generalize the order waiting time process in order to properly model such empirical evidences.
The empirical study is performed on the best bid and best ask data of 7 U.S. financial markets, for 30-stock time series. In particular, statistical properties of trading wait-

ing times have been analyzed and quality of fits is evaluated by suitable statistical tests, i.e., comparing empirical distributions with theoretical models.
Starting from the statistical studies on real data, attention has been focused on the reproducibility of such results in an artificial market. The computational experiments have been performed within the Genoa Artificial Stock Market. In the market model, heterogeneous agents trade one risky asset in exchange for cash. Agents have zero intelligence and issue random limit or market orders depending on their budget constraints. The price is cleared by means of a limit order book. The order generation is modeled with a renewal process.
Based on empirical trading estimation, the distribution of waiting times between two consecutive orders is modeled by a mixture of exponential processes. Results show that the empirical waiting-time distribution can be considered as a generalization of a Poisson process. Moreover, the renewal process can approximate real data and implementation on the artificial stocks market can reproduce the trading activity in a realistic way.

2.1 The waiting-time distribution of trading activity in a double auction artificial financial market

2.1.1 Introduction

A model of trading in the Genoa Artificial Stock Market [46, 138, 160, 162, 163] characterized by a double auction clearing mechanism, i.e., the limit order book [45, 161] is presented. The limit order book is a snapshot at a given instant of the queues of all buy limit orders and sell limit orders, with their respective price and volume. Limit orders are organized in ascending order according to their limit prices. All buy limit orders are below the best buy limit order, i.e., the buy limit order with the highest limit price (the bid price). The best buy limit order is situated below the best sell limit order, i.e., the sell limit order with the lowest limit price (the ask price). All other sell limit orders are above the best sell limit order. Orders are stored in the book. A transaction occurs when a trader hits the quote (the bid or the ask price) on the opposite side of the market. If a trader issues a limit order, say a sell limit order, the order either adds to the book if its limit price is above the bid price, or generates a trade at the bid if it is below or equal to the bid price. In the latter case, the limit order becomes a marketable limit order or more simply a market order. Conversely, if the order is a buy limit order it becomes a market order and is executed if its limit price is above the ask price, otherwise it is stored in the book. Limit orders with the same limit price are prioritized by time of submission, with the oldest order given the highest priority. Order's execution often involves partial fills before it is completed, but partial fills do not change the time priority.

In recent years, some studies about the statistical properties of the limit order book have appeared in the literature [23, 29, 103, 124, 140]. An important empirical variable is the waiting time between two consecutive transactions. In fact, trading via the order book is asynchronous, i.e., a transaction occurs only if a trader issues a market order. For liquid stocks, waiting times vary in a range between some seconds to a few minutes. Raberto et al. [164] analyze the intra-day trades of General Electric stock prices and find that waiting times exhibit a 1-day periodicity, corresponding to the daily stock market activity, and a survival probability distribution which is well fitted by a stretched exponential. They also find a significative cross-correlation between waiting times and the absolute value of log-returns.

In this chapter, the effect of a more general distribution of order waiting times is investigated. In particular, the attention is focused on the Weibull distribution that admits the exponential distribution as a limit case. Results show that in the case of exponentially distributed order waiting times, also trade waiting times are exponentially distributed. Conversely, if order waiting times follow a Weibull, the same does not hold for trade waiting times.

2.1.2 The model

A model of artificial trading by means of a limit order book is presented in this section. Agents trade one single stock in exchange for cash. They are modeled as liquidity traders; as a consequence, the decision making process is nearly random and depends on the finite amount of financial resources (cash + stocks) they own. At the beginning of the simulation, cash and stocks are uniformly distributed among agents.

2.1.2.1 The order generation process

Trading is organized in M daily sections. Each trading day is subdivided in T elementary time steps, say seconds. During the trading day, at given time steps t_h, a trader i is randomly chosen for issuing an order. Order waiting times $\tau_h^O = t_h - t_{h-1}$ are determined according to a Weibull distribution, whose probability density function (PDF) $p(\tau)$ is:

$$p(\tau) = \frac{\beta}{\eta}\left(\frac{\tau}{\eta}\right)^{\beta-1} e^{-\left(\frac{\tau}{\eta}\right)^{\beta}}, \tag{2.1}$$

where η is the scale parameter and β is the shape parameter, also known as the slope, because the value of β is equal to the slope of the regressed line in a probability plot. The expected value $\langle\tau\rangle$ of a random variable following a Weibull PDF is given by:

$$\langle\tau\rangle = \eta \cdot \Gamma\Big(1/\beta + 1\Big), \tag{2.2}$$

where Γ is the Gamma function. The surival probability distrbutiion $P_>(\tau) = \int_\tau^\infty p(\tau)$ is given by:

$$P_>(\tau) = e^{-\left(\frac{\tau}{\eta}\right)^{\beta}}. \tag{2.3}$$

The exponential distribution is a particular case of the Weibull distribution for $\beta = 1$. In the case $\beta < 1$, the Weibull distribution assumes the form of the so-called stretched exponential and great values of τ occur with higher probability than in the case of $\beta = 1$. In the computational experiments, only values of the shape parameters β less or equal than one are considered.

The order generation process is then described as a general renewal process where the times between two consecutive orders τ_h^O are independent and identically distributed random variables following a Weibull distribution. In the case $\beta = 1$, the order generation process is assumed to be a Poisson process with an exponential waiting-time distribution. For further details on renewal processes, see Cox and Isham 1980 [52].

2.1.2.2 The trading decision making process

A trader issues a buy or a sell order with probability 50%. Denote with $a(t_{h-1})$ and with $d(t_{h-1})$ the values of the ask and of the bid prices stored in the book at time step t_{h-1}. Suppose that the order issued at time step t_h be a sell order, then the limit price $s_i(t_h)$ associated to the sell order is given by:

$$s_i(t_h) = n_i(t_h) \cdot a(t_{h-1})\,, \tag{2.4}$$

where $n_i(t_h)$ is a random draw by trader i at time step t_h from a Gaussian distribution with constant mean $\mu = 1$ and standard deviation σ. If $s_i(t_h) > d_i(t_{h-1})$, the limit order is stored in the book and no trades are recorded; else, the order becomes a market order and a transaction occurs at the price $p(t_h) = d(t_{h-1})$. In the latter case, the sell order is partially or totally fulfilled and the bid price is updated. The quantity of stocks offered for sale is a random fraction of the quantity of stocks owned by the trader.

If the order is a buy order, we assume that the associated limit price $b_i(t_h)$ is given by:

$$b_i(t_h) = n_i(t_h) \cdot d(t_{h-1})\,; \tag{2.5}$$

where $n_i(t_h)$ is determined as in the previous case. If $b_i(t_h) < a(t_{h-1})$, the limit order is stored in the book and no trades are recorded; otherwise the order becomes a market order and a transaction occurs at the price $p(t_h) = a(t_{h-1})$. The quantity of stocks ordered to buy depends on cash endowment of the trader and on the value of $b_i(t_h)$.

It is worth noting that, in this framework, agents compete for the provision of liquidity. If an agent issues a buy order, its benchmark is the best limit buy order given by the bid price. Being $\mu = 1$, half times, he offers a more competitive buy order (if $b_i(t_h) > d(t_{h-1})$), which may result in a trade if $b_i(t_h) \geq a(t_{h-1})$. The same applies for sell limit orders.

2.1.3 Computational experiments

The timing parameters of every simulation have been set as follows: $M = 50$ daily sections, each characterize by a length of $T = 25,200$ s (corresponding to 7 hours of trading activity). Each simulation is characterized by a particular value of the shape parameter β of the Weibull distribution modeling order waiting times. 17 different values of β ranging from 0.2 to 1 with step 0.05 have been chosen. The average order waiting time $\langle \tau^o \rangle$ has been set to 20 s for every simulation. The scale factor η is so determined by the choice of β and $\langle \tau^o \rangle$ according to Eq. 2.2. The orders lifespan has been set to 600 s $\gg \langle \tau^o \rangle$. Sell and buy limit prices are computed following Eq. 2.4 and Eq. 2.5 respectively. The random number $n_i(t_h)$ is a random draw by trader i from a Gaussian distribution with constant mean $\mu = 1$ and standard deviation $\sigma = 0.005$.

The number of agents is set to 10,000. At the beginning of the simulation, the stock price is set at 100.00 units of cash, say dollars and each trader is endowed with an equal amount of cash and of shares of the risky stocks. These amounts are 100,000 dollars and 1,000 shares, respectively.

Computational experiments produce realistic intraday price paths, see Ref. [161] for further details. Log-returns $r_{\Delta t}$ have been computed in homogeneous time windows Δt, according to the previous-tick interpolation technique [56, pag. 37]. The time window Δt has been set to 100 s, which is about two times the average value of trade waiting times. Figure 2.1 presents the average values of orders waiting times $\langle \tau^o \rangle$ and trade waiting times $\langle \tau^T \rangle$, respectively, as a function of the shape parameter β. $\langle \tau^o \rangle$ is nearly 20 s, as expected by model construction, whereas $\langle \tau^T \rangle$, i.e., an output of the GASM model, is around 55 s. $\langle \tau^T \rangle$ appears to be independent from β when $\beta \geq 0.3$; the pattern of increasing values of $\langle \tau^T \rangle$ when $\beta < 0.3$ may be due to numerical effects. The choice of $\Delta t > \langle \tau^T \rangle$ has been made in order to avoid spurious effect in the statistical properties of returns due to long period of trading inactivity, especially in the case of small value of β.

Figure 2.2 presents the values of kurtosis of log-returns as a function of the shape parameter β. The figure shows that the distribution of log-returns is characterized by increasing values of kurtosis as β decreases. In previous works [45, 161], it is already showed that the mechanism of the limit order book was able to recover fat tails in the distribution of log-returns without ad-hoc behavioral assumptions regarding agents. In that cases, β was set to 1, i.e., the order generation process was modeled as a Poisson process with exponentially distributed waiting times. Figure 2.2 generalizes such a result showing that the same conclusion also holds for a more general renewal process, i.e., Weibull distributed waiting times. Furthermore, the tails of the distribution of returns become fatter when β decreases from 1, i.e., the distribution of order waiting times becomes a stretched exponential.

Figure 2.3 and Figure 2.4 show estimates of the survival probability distribution of order waiting times (dots) and of the survival probability distribution of trade waiting times (crosses) in the case $\beta = 0.4$ (Fig. 2.3) and $\beta = 1$ (Fig. 2.4). As expected by the assumptions of the model, survival probability distributions of order waiting times follow the corresponding Weibull distribution, represented by the continuous line.

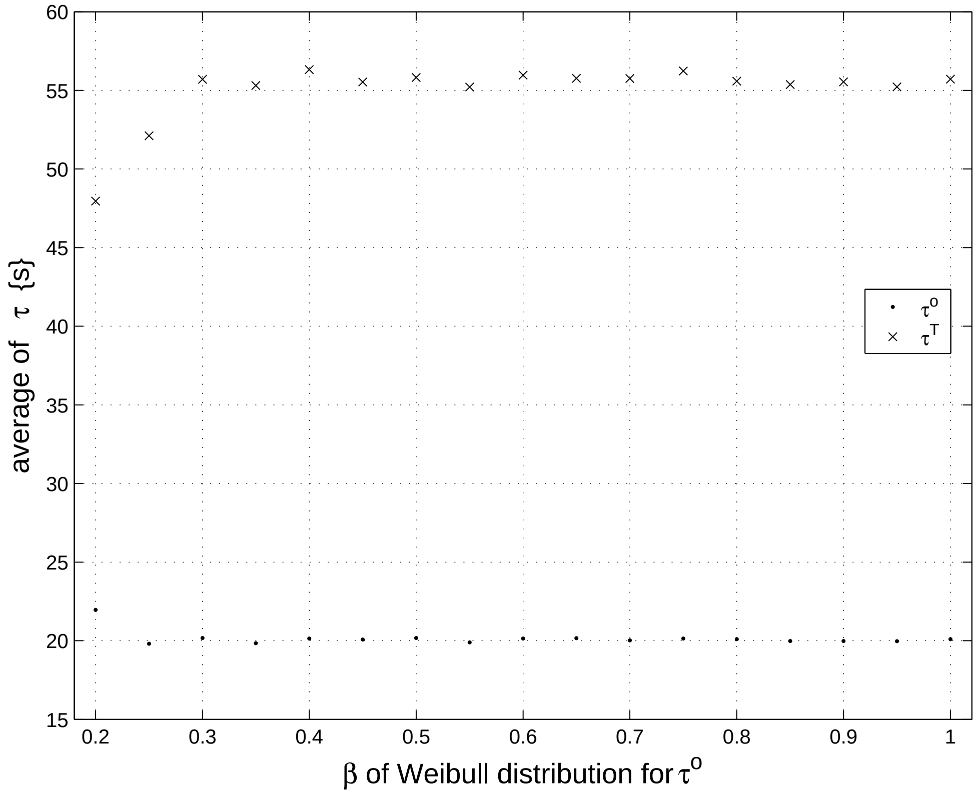


Figure 2.1: Average values of order waiting times and of trade waiting times as a function of the shape parameter β of the Weibull distribution for orders.

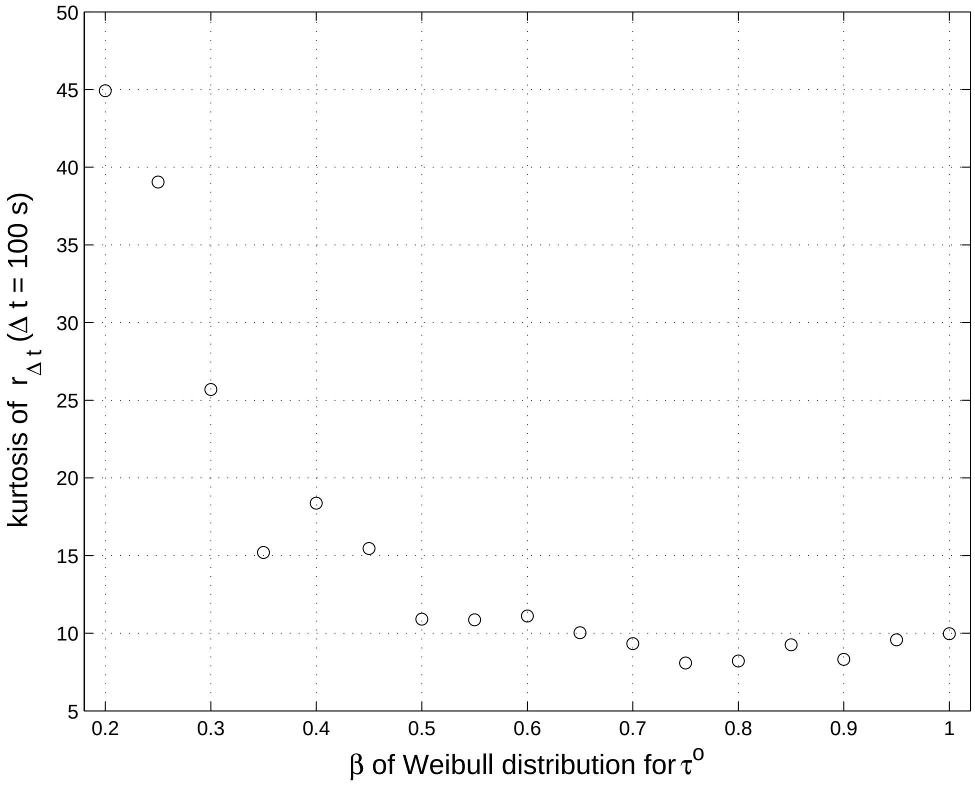


Figure 2.2: Kurtosis of the distribution of log-returns as a function of the shape parameter β.

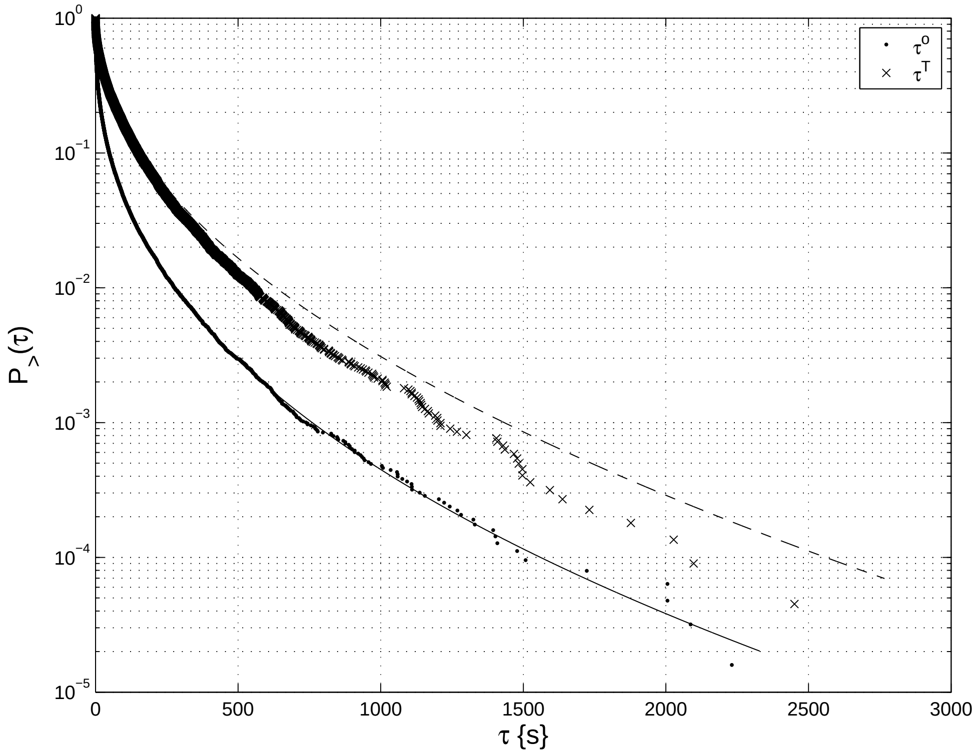


Figure 2.3: Survival probability distribution of order waiting times (dots) and of trade waiting times (crosses) in the case $\beta = 0.4$. The two curves represent the corresponding Weibull fits.

The survival probability distribution of trade waiting times is well fitted by a Weibull distribution only in the case $\beta = 1$. Figure 2.3 shows that, in the case $\beta = 0.4$, the survival probability distribution of trade waiting times departs from a Weibull distribution determined from data by means of the maximum likelihood principle and represented in the Figure with the dashed curve. In this case, the Kolmogorov-Smirnov test rejects the null hypothesis of Weibull distribution at the significance level of 5 %. Generally speaking, trade waiting times are Weibull distributed only in the case $\beta = 1$.

In the case $\beta = 1$, the process of trading is a Poisson process as the order arrival process. An identical conclusion follows from theoretical considerations. In fact, consider that every transaction occurs when a new order that arrives in the book finds a matching order in the queue of orders of the opposite type. Therefore, any new order will be satisfied or not in function of the state of the book in that moment. The state of the book varies for each moment and for each simulation path. Given the absence of significant feedbacks in the market, however, it is reasonable to assume that the average state of the book is time invariant. Therefore, each incoming order will be satisfied on average with a constant probability and the trading process can

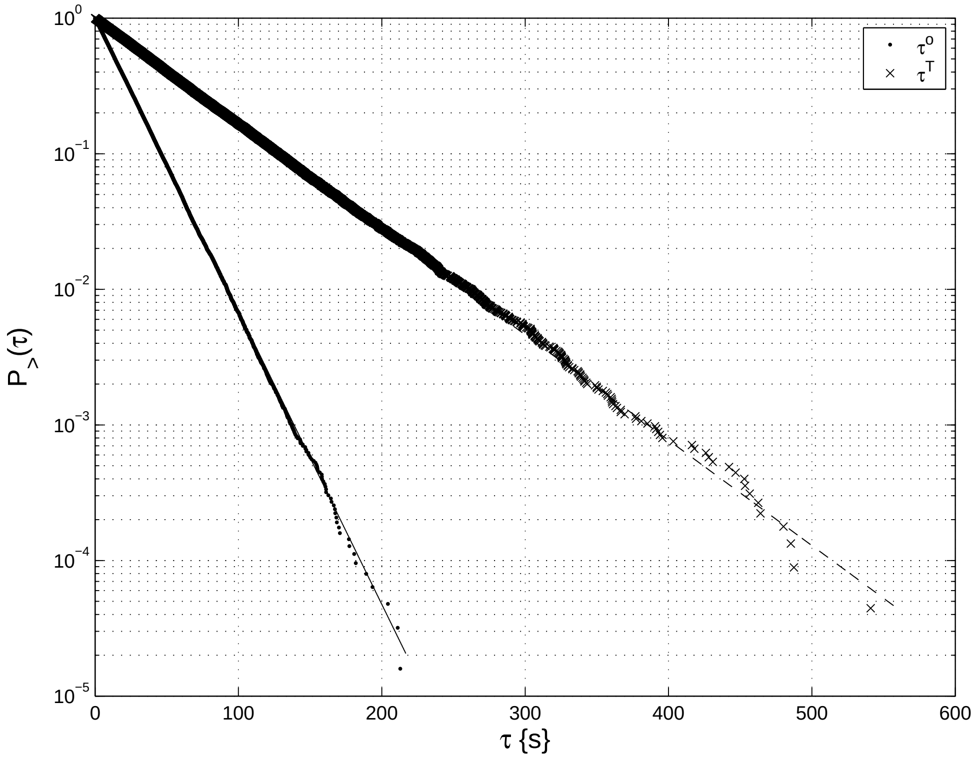


Figure 2.4: Survival probability distribution of order waiting times (dots) and of trade waiting times (crosses) in the case $\beta = 1$ (exponential distribution). The two lines represent the corresponding exponential fits.

be regarded as a random extraction form the Poisson process of orders issuing. The procedure of random extraction from a Poisson process is called "thinning" and it is well known that a thinning from a Poisson process is a new Poisson process; see Ref. [52] for further details.

2.2 Poisson-process generalization for the trading waiting-time distribution in a double-auction mechanism

2.2.1 Introduction

The dynamics of a stock market depends on the interaction between trading mechanism and trading behaviors. The behavior of the participants is the outcome of their trading strategies, which include how they form expectations or interpret signals. Conversely, the trading mechanism defines the rules of the market, which specify how orders are placed and how the price changes.

A model of the stock market makes structural assumptions, related to the trading mechanism. In order-driven systems, competing market makers supply liquidity by quoting bid and ask prices and volumes at which they are willing to trade. Investors demand liquidity through the submission of market orders. A limit order book is a snapshot at a given instant of the queues of all buy limit orders and sell limit orders, with their respective price and volume.

All buy limit orders are below the best buy limit order, i.e., the buy limit order with the highest limit price (the bid price). The best buy limit order is situated below the best sell limit order, i.e., the sell limit order with the lowest limit price (the ask price). All other sell limit orders are above the best sell limit order. A transaction occurs when a trader hits the quote (the bid or the ask price) on the opposite side of the market. If a trader issues a limit order, say a buy limit order, the order either adds to the book if its limit price is below the ask price, or generates a trade at the ask if it is larger or equal to the ask price. In the latter case, the limit order becomes a marketable limit order, or more simply, a market order. Conversely, if the order is a sell limit order it becomes a market order and is executed if its limit price is below the bid price, otherwise it is stored in the book. Limit orders with the same limit price are prioritized by time of submission, with the oldest order given highest priority. The execution of orders often involves partial fills before it is completed, but partial fills do not change the time priority.

In recent years, some studies on the statistical properties of the limit order book have proposed by the scientific community [23, 29, 103, 124, 140]. An important empirical variable is the waiting time between two consecutive transactions. Indeed, trading via the order book is asynchronous, i.e., a transaction occurs only if a trader issues a market order. For liquid stocks, waiting times may vary in a range between some seconds to a few minutes. Generally speaking, the survival of order waiting times is modeled by an exponential[46]. However, analysis of the intra-day trades of General Electric stock prices pointed out that trading waiting times exhibit a 1-day periodic-

ity, corresponding to the daily stock market activity, and that the survival probability distribution is properly fitted by a stretched exponential[132, 164, 174]. Moreover, a significative cross-correlation between waiting times and the absolute value of log-returns was also found which suggested other memory effects.
Starting from these results, the effects of a more general distribution of order waiting times have been investigated. In particular, attention has been focused on the Weibull distribution that admits the exponential distribution as a limit case [47]. Results showed that in the case of exponentially distributed order waiting times, also trade waiting times are exponentially distributed. Conversely, if order waiting times follow a Weibull, the same does not hold for trading waiting times. Thus, such studies concluded that a single Weibull distribution of the order waiting time could not reproduce the empirical evidences.
In this chapter, order and trading waiting times are studied in the general framework of the Genoa Artificial Stock Market [46, 138, 160, 162, 163]. A trading mechanism characterized by a double auction clearing mechanism, i.e., the limit order book, is considered[45, 161]. In particular, mixture of Poisson process is used to describe the order generation process. The characteristics of the Poisson process in the mixture are estimated by high frequency real data that points out changes of the average waiting time during the trading day[22]. Results pointed out that a mixture of Poisson process can reproduce the behavior of real stock market.

2.2.2 The artificial stock market

In this subsection, a model of artificial trading by means of a limit order book is presented. The model makes reference to the Genoa Artificial Stock Market - GASM[47, 138, 160–163]. In the GASM, agents trade a single stock in exchange for cash. They are modeled as liquidity traders, and the decision making process is nearly random constrained by the finite amount of financial resources (cash + stocks) they own. At the beginning of the simulation, cash and stocks are uniformly distributed among agents.

2.2.2.1 Trading decision making process

The GASM is populated by random trader, i.e., a trader issues a buy or a sell order with probability 50%. Denote with $a(t_{h-1})$ and with $d(t_{h-1})$ the values of ask and of bid prices stored in the book at time step t_{h-1}. Now assume that the order issued at time step t_h is a sell order. Thus, the quantity of stocks offered for sale is a random fraction of the quantity of stocks owned by the trader whereas the limit price $s_i(t_h)$ associated to the sell order is given by:

$$s_i(t_h) = n_i(t_h) \cdot a(t_{h-1})\,, \tag{2.6}$$

where $n_i(t_h)$ is a random draw by trader i at time step t_h from a Gaussian distribution with constant mean $\mu = 1$ and standard deviation σ. If $s_i(t_h) > d_i(t_{h-1})$, the limit order is stored in the book, otherwise the order becomes a market order and a transaction occurs at the price $p(t_h) = d(t_{h-1})$. In the latter case, the sell order is partially or totally fulfilled and the bid price is updated.

If the order is a buy order, the associated limit price $b_i(t_h)$ is given by:

$$b_i(t_h) = n_i(t_h) \cdot d(t_{h-1})\,; \tag{2.7}$$

where $n_i(t_h)$ is determined as in the previous case. If $b_i(t_h) < a(t_{h-1})$, the limit order is stored in the book, otherwise the order becomes a market order and a transaction occurs at the price $p(t_h) = a(t_{h-1})$. The quantity of stocks ordered to buy depends on cash endowment of the trader and on the value of $b_i(t_h)$.

It is worth noting that, in this framework, agents compete for the provision of liquidity. If an agent issues a buy order, its benchmark is the best limit buy order given by the bid price. Being $\mu = 1$, he offers in average a more competitive buy order (if $b_i(t_h) > d(t_{h-1})$), which may result in a trade if $b_i(t_h) \geq a(t_{h-1})$. The same applies for sell limit orders.

2.2.2.2 Order generation process

Trading is organized in M daily sections and each trading day is subdivided in T elementary time steps, say seconds. During the trading day, at given time steps t_h, a trader i is randomly chosen for issuing an order. The trading day can be divided into L subintervals. Within each subinterval, waiting times $\tau_h = t_h - t_{h-1}$ follow an exponential distribution with different average waiting times $\tau_0^1, ..., \tau_0^L$. Recalling that the rate μ_i is the inverse of the average waiting time, i.e. $\mu_i = 1/\tau_0^i$, the survival function of the i-th subinterval is given by

$$p_i(\tau) = e^{-\mu_i \tau}, \tag{2.8}$$

where $i = 1, ..., L$.
In recent years, several author have shown that, in average, trading waiting time are not uniformly distributed during a trading day. In particular, Bertram pointed out that the number of trading in time interval of 600 seconds (i.e, 10 minutes) are variable in time with a typical *smile* pattern[22] (see Figure 2.5). According to his result, it was used the average number of trading to estimate the rate parameter μ_i of the exponential distribution that represents the order waiting time distribution. It is worth noting that such estimation is based on trades instead of orders. However, one can assume that a linear relationship generally exist between the number of issue orders and the number of transactions.
Generally speaking, the average number of trading in a period of 600 seconds is a function of time (see Figure 2.5). In this study, the average number of transactions is modeled by a polynomial approximation i.e.,

$$n(t) = \sum_{j=0}^{J} a_j t^j \tag{2.9}$$

Moreover, least squared fit on the empirical data pattern reported by Bertram[22] points out that Eq. 2.9 with $J = 8$ is in good agreement with the empirical results (as shown in Figure 2.5). Thus, Eq. 2.9 allows one to evaluate the rate of the exponential distribution in the sub-interval as

$$\mu_i = \frac{n(t_i)}{600} \tag{2.10}$$

where $i = 1, ..., L$. It is worth noting that $n(t_i)$ denotes the value of the polynomial calculated in the middle of i-th interval (see circle in Figure 2.5). As a consequence, in any time interval, agents issue orders according to an exponential distribution whose rate changes in time, i.e., order waiting times are distributed according to a mixture of Poisson processes.

2.2.3 Empirical Analysis

In this section, the trading waiting time for 30 DJIA titles traded at NYSE in October 1999 have been considered. The time-series have been statistically analyzed and Table 2.1 points out main results. The first column reports the Anderson-Darling statistics A^2. In these cases, the critical value is 1.9 and, as clearly stated, the null hypothesis of exponential distribution is rejected for all real time-series. This conclusion is in perfect agrement with previous results, thus pointing out that the point process of trading is not a Poisson process [47, 164].

The second column shows estimation of Weibull exponent parameter β. The Weibull process generalizes the Poisson process. Indeed, the exponential distribution is a particular case of the Weibull distribution for $\beta = 1$. In the case $\beta < 1$, the Weibull distribution assumes the form of the so-called stretched exponential and great values of trading waiting times occur with higher probability than in the case of $\beta = 1$[47, 164]. As shown, the estimated β are lower than 1, i.e., the trading waiting times distribution are stretched exponentials.

The third column reports Ljung-Box test statistics Q at lag 15. The critical value of the test is 24.99 at the 5% significance level. As clearly stated, the null hypothesis of no serial correlation is generally rejected, except for the case of Citigroup Inc. (denoted by $*$), that is the only case in which test is not rejected.

Furthermore, Table 2.2 summarizes serial correlation results. First, second and third columns reports autocorrelation value of trading waiting times for different lags, measured in seconds, for IBM stock in December 1990, January 1991, October 1999. Fourth column shows the same statistics of the trading times for Citigroup Inc. (C), i.e., the only stock for which the null hypothesis of no serial correlation is not rejected. The data sets used to produce Table 2.2 have different characteristics. First sample exhibits an average time interval between trades of 26.2 seconds, with minimum and maximum interval of 0 and of 4,592 seconds (i.e., about 1 hour and 15 minutes), respectively. Moreover, standard deviation results of 53.9 seconds and skewness is equal to 42.3. Second sample has an average of 27.07 seconds with minimum and maximum interval of 0 and of 426 seconds (i.e., about 7 minutes), respectively. The standard

deviations is 37.69 seconds. The third sample has an average time interval between trades of 9.8 seconds, minimum interval is 0 seconds, and maximum interval is $3,600$ seconds. The standard deviation is 34.3 seconds and skewness is 91.4. The fourth sample has an average time interval between trades of 9.2 seconds, with minimum and maximum interval of 0 seconds and $3,655$ seconds respectively, the standard deviation is 41.8 seconds and skewness is 82.2.
The autocorrelations of trading waiting times in Table 2.2 point out the presence of correlation, i.e., autocorrelation functions decrease slowly with exception of C asset whose autocorrelation function decrease faster. This result is confermed by the Ljung-Box statistical test. Indeed, Ljung-Box statistics are very large, except for the stock C. Thus, the null hypothesis of white noise is generally rejected based on the critical value of 24.99 at the 5% significance level.

2.2.4 Computational experiments

Besides real data analysis, some computational experiments have been considered. The timing parameters of every simulation have been set as follows: $M = 50$ daily sections, each characterized by a length of $T = 21,000$ s. Each simulation is characterized by a different number of exponential distributions used in the mixture. In the simulation how many exponential distributions use can be chosen. 7 different value for $L : 2, 3, 4, 5, 15, 25, 35$ have been chosen. The orders lifespan has been set to 600 s $\gg \langle \tau^o \rangle$.

Sell and buy limit prices are computed following Eq. 2.6 and Eq. 2.7 respectively. The random number $n_i(t_h)$ is a random draw by trader i from a Gaussian distribution with constant mean $\mu = 1$ and standard deviation $\sigma = 0.005$.

The number of agents is set to 10,000. At the beginning of the simulation, the stock price is set at 100.00 units of cash, say dollars and each trader is endowed with an equal amount of cash and of shares of the risky stocks. These amounts are 100,000 dollars and 1,000 shares, respectively.

The effects of the number of exponential distribution on the order and trading waiting times have been considered. In particular, mixtures of $2, 3, 4, 5, 15, 25, 35$ exponential distributions have been studied, being 35 the maximum time resolution allowed by empirical data.[22]
Figure 2.6 shows the survival functions of order and trading waiting times. As clearly stated, the effect of large numbers of exponential distribution on the survivals of order and trading waiting times results almost negligible, thus suggesting that already a mixture of two Poisson processes can properly represent the empirical evidences. It is worth noting that a mixture of two exponential process per trading day corresponds to order waiting times identically distributed at the beginning and at the end of the trading day. In fact, it can be thought that the Poisson process at the end of one day is exactly the same at the beginning of the next day.
Figure 2.6 points out that waiting times generated by the GASM are not distributed according to a Poisson process. This is further confirmed by the Anderson-Darling

test, shown in Table 2.3. According to the critical value of 1.9, we reject the null hypothesis of exponential distribution[47, 173]. Conversely, Weibull distribution has been verified with the Kolmogorov-Smirnov test. As shown in Table 2.4 the null hypothesis is never rejected, thus allowing the possibility of Weibull processes for order and trading waiting times.

In addition to Kolmogorov-Smirnov test, estimation of β (i.e., the Weibull parameter) for GASM data points out values quite close to those obtained in the case of 30 DJIA stocks traded at NYSE in October 1999 (see Tables 2.5 and 2.1 for a comparison). This suggest applicability of the proposed model in order to reproduce empirical evidences. Moreover, in the case of β estimation, results point out again an almost negligible effect of the number of exponential distribution in the mixture.

Concerning serial correlation, Ljung-Box test statistics Q at lag 15 for the order τ^O and trading τ^T waiting times have been calculated. As shown in Table 2.6, the null hypothesis is rejected. Moreover, fifth column in Table 2.2 points out the autocorrelation of trading waiting times for different lags in the case of 35 exponential distribution in the mixture.

It is worth noting that GASM data are referred to 50 trading days, instead of the one month trading data set for columns 1-4. Moreover, the GASM data are characterized by an average time interval between trades of 130.61 seconds, with minimum and maximum intervals of 0 and 3109 seconds, respectively. Standard deviation is equal to 181.06 seconds and skewness is 5.08. The statistical properties of GASM intertrade durations are not in contradiction with those measured for real data. This confirm that the GASM can reproduce the beavior of real stock market.

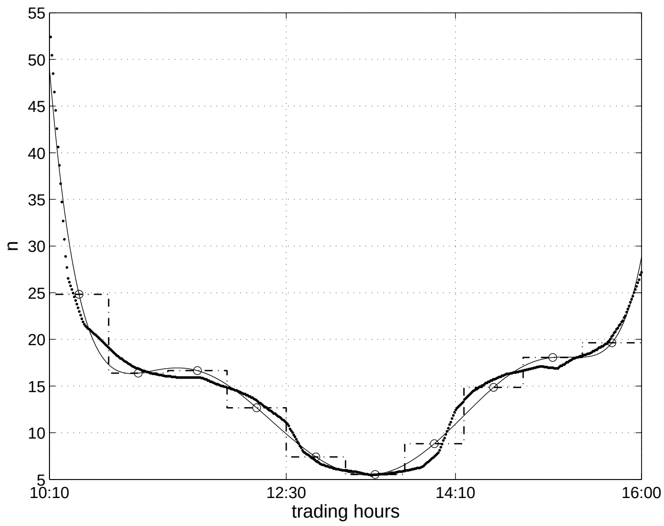


Figure 2.5: Average number of transactions during the trading day. Dotted line represents real data[22], continuous line the polynomial fit and dashed curve the discretization of polynomial fit in 15 intervals.

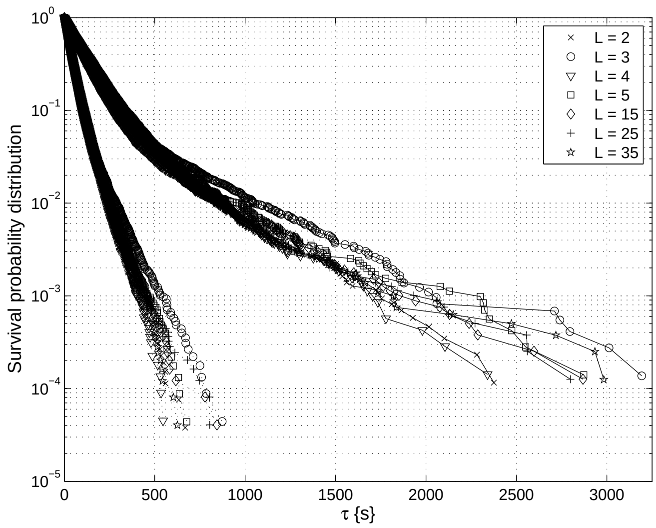


Figure 2.6: Survival probability distributions of order and trading waiting times for different mixtures of exponential distributions. Continuous line: order distributions; dashed curves: trading distributions.

Table 2.1: Anderson-Darling test statistics A^2, β estimated and Ljung-Box test statistics Q at lag 15 for trading waiting times τ^T of the 30 DJIA titles traded at NYSE in October 1999.

	A^2	β	Q
Stocks	τ^T		
AA	inf	0.74	569
ALD	111	0.72	1058
AXP	inf	0.58	401
BA	inf	0.73	815
C	inf	0.48	20*
CAT	291	0.76	754
CHV	inf	0.74	403
DD	inf	0.68	63
DIS	inf	0.64	238
EK	123	0.85	997
GE	inf	0.55	110
GM	144	0.84	1131
GT	262	0.78	3098
HWP	inf	0.57	634
IBM	inf	0.51	208
IP	inf	0.76	439
JNJ	107	0.64	1812
JPM	inf	0.66	797
KO	134	0.61	2537
MCD	177	0.75	2007
MMM	211	0.80	1264
MO	inf	0.60	551
MRK	inf	0.55	601
PG	126	0.61	2252
S	inf	0.75	655
T	inf	0.55	347
UK	182	0.99	1529
UTX	169	0.80	777
WMT	inf	0.58	58
XON	399	0.68	3469

Table 2.2: Trading-interval autocorrelations function and Ljung-Box statistics for 15 lags. The first two columns are taken from Engle and Russell (1994)[65].

	IBM (Dec 1990)	IBM (Jan 1991)	IBM (Oct 1999)	C (Oct 1999)	GASM (L=35)
lag 1	0.168	0.129	0.061	0.003	0.165
lag 2	0.090	0.120	0.009	0.011	0.157
lag 3	0.068	0.106	0.005	0.005	0.140
lag 4	0.074	0.119	0.005	0.005	0.133
lag 5	0.059	0.107	0.007	0.004	0.151
lag 6	0.069	0.096	0.003	0.006	0.144
lag 7	0.051	0.100	0.002	0.003	0.104
lag 8	0.046	0.099	0.003	0.004	0.102
lag 9	0.045	0.123	0.004	0.005	0.114
lag 10	0.042	0.085	0.003	0.005	0.074
lag 11	0.043	0.105	0.002	0.003	0.101
lag 12	0.045	0.087	0.001	0.004	0.060
lag 13	0.047	0.089	0.005	0.004	0.075
lag 14	0.037	0.089	0.001	0.004	0.054
lag 15	0.025	0.083	0.002	0.002	0.058
Q	1272	2423	208	20	1583

Table 2.3: Anderson-Darling test statistics, A^2 for order τ^O and trading τ^T waiting times. Critical value of the test is 1.9. In all cases, the null hypothesis of exponential distribution is rejected.

	A^2	
L	τ^O	τ^T
2	52	53
3	66	71
4	20	22
5	31	34
15	61	57
25	84	84
35	73	64

Table 2.4: Kolmogorov-Smirnov test statistics D for order waiting times τ^O and trade waiting times τ^T.

	D		
L	τ^O	τ^T	Critical Value
2	0.033	0.029	0.052
3	0.030	0.039	0.050
4	0.025	0.018	0.051
5	0.026	0.024	0.051
15	0.032	0.026	0.052
25	0.034	0.025	0.051
35	0.033	0.026	0.51

Table 2.5: Value of β for trading waiting times and orders waiting times for GASM data.

L	β^O	β^T
2	0.93	0.87
3	0.92	0.86
4	0.96	0.91
5	0.95	0.89
15	0.92	0.87
25	0.91	0.85
35	0.92	0.86

Table 2.6: Ljung-Box test statistics Q at lag 15 for order waiting times τ^O and trade waiting times τ^T. The critical value of the test is 24.99. In all the cases, the null hypothesis of no serial correlation up to lag 15 is rejected.

	Q	
L	τ^O	τ^T
2	6405	1543
3	7970	1155
4	2371	562
5	3429	526
15	7442	1380
25	7662	1792
35	8717	1583

Chapter 3

Modeling and Statistical Analysis of a Multi-Assets Artificial Stock Market

Overview

At the beginning of this chapter, a multi-assets artificial financial market populated by zero-intelligence traders with finite financial resources is presented. The market is characterized by different types of stocks representing firms operating in different sectors of the economy. Zero-intelligence traders follow a random allocation strategy which is constrained by finite resources, past market volatility and allocation universe. Within this framework, stock price processes exhibit volatility clustering, fat-tailed distribution of returns and reversion to the mean. Moreover, the cross-correlations between returns of different stocks is studied using methods of random matrix theory. The probability distribution of eigenvalues of the cross-correlation matrix shows the presence of outliers, similar to those recently observed on real data for business sectors. It is worth noting that business sectors have been recovered in our framework without dividends as only consequence of random restrictions on the allocation universe of zero-intelligence traders. Furthermore, in the presence of dividend paying stocks and in the case of cash inflow added to the market, the artificial stock market points out the same structural results obtained in the simulation without dividends. These results suggest a significative structural influence on statistical properties of multi-assets stock market.

In the second part of the chapter, an information-based multi-assets artificial stock market is presented. The market is populated by heterogeneous agents that are seen as nodes of sparsely connected graphs. The market is characterized by different types of stocks and agents trade risky assets in exchange for cash. Beside the amount of cash and of stocks owned, each agent is characterized by sentiments. Moreover, agents share their sentiments by means of interactions that are determined by graphs. A central market maker (clearing house mechanism) determines the price processes for each stock at the intersection of the demand and the supply curves. Within this framework, stock price processes exhibit main univariate stylized facts, i.e., unitary root price processes, fat tails of return distribution and volatility clustering. Furthermore, the multivariate price processes exhibits both static and dynamic stylized facts, i.e., the presence of static factors and common trends. These results suggest a significative structural influence on statistical properties of multi-assets stock market.

3.1 A multi-assets artificial stock market with zero-intelligence traders

3.1.1 Introduction

Agent based simulation of financial markets is a rapidly growing field. The Genoa Artificial Stock Market (GASM)[1] is a computer simulator of financial markets which reproduce main stylized facts of the real markets, i.e., volatility clustering and fat-tailed distribution of returns with very simple assumptions about the behavior of agents and realistic market microstructure. Based on recent results on the Genoa Artificial Stock Market with a single risky stock traded in exchange of cash [138, 162, 163], an extension of the GASM in a multi-assets environment is presented. In previous researches [46], a multi-assets environment has been investigated with agents characterized by different trading strategies e.g.,random traders, mean-variance traders, relative chartist traders and mean-reversion traders. In this case, only random traders are assumed, i.e., agents that do not use the information about dividends to choose the stocks on which to invest. Both stocks paying dividends and non-paying dividends are considered. Results point out statistical properties of the multi-assets artificial market similar to those of real market. This suggests that behavioral assumptions may not be necessary to obtain the main stylized facts of the market.

3.1.2 The model

In this subsection, the model of a multi-assets artificial market is presented. The model makes reference to the Genoa Artificial Stock Market - GASM. In the GASM, agents trade stocks in exchange for cash. They are modeled as liquidity traders, i.e., decision making process is random constrained by the finite amount of financial resources (cash + stocks) they own. At the beginning of the simulation, cash and stocks are uniformly distributed among agents.
Let S be the number of sectors that characterize the economy, e.g., construction, information technology, manufacturing, etc. Let s denote the particular economic sector and the pair s, j the j-th firm operating in the sector s. For each firm, a number of stock securities is traded sequentially in the market by N agents. Let the subscript i denote the i-th agent. At every discrete time step h, agent i is characterized by a cash $c_i(h)$ and number of stocks $a_i^{s,j}(h)$ of firm s, j held in his portfolio. If $p^{s,j}(h)$ is

[1] The name is devoted to the city where the project has been developed.

the market price of the risky asset s, j at time step h, the risky wealth $W_i^r(h)$ owned by trader i at time step h is:

$$W_i^r(h) = \sum_{s,j} a_i^{s,j}(h) p^{s,j}(h)\,, \tag{3.1}$$

whereas $W_i(h) = c_i(h) + W_i^r(h)$ represents the total wealth. The weight $\omega_i^{s,j}(h)$ of asset s, j in the portfolio of agent i at time step h is given by:

$$\omega_i^{s,j}(h) = p^{s,j}(h) \cdot a_i^{s,j} / W_i^r\,. \tag{3.2}$$

It is worth noting that $\sum_{s,j} \omega_i^{s,j} = 1$ for all i by definition. Furthermore, it is imposed that $\omega_i^{s,j} \geq 0$ for all i and for all asset types s, j (i.e., short positions are not allowed).

At each step, every zero-intelligence agent is activated with a given probability. An agent i, if activated, tries to allocate in the risky assets a random fraction γ_i of his total wealth, i.e., $\widehat{W}_i^r(h+1) = \gamma_i W_i(h)$, where γ_i is a random draw from an uniform distribution between 0 and 1. The symbol $\widehat{.}$ means that $\widehat{W}_i^r(h+1)$ is the amount that agent i desires to allocate in the risky investment, whereas the real amount $W_i(h+1)$ effectively allocated in stocks will depend on the trading process with other agents. Zero-intelligence traders follow a random allocation strategy, i.e., the desired weight $\widehat{\omega}_i^{s,j}(h+1)$ for stock s, j in the portfolio of agent i at time step $h+1$ is a random draw from an uniform distribution between 0 and 1 with the above constraints on normalization and absence of short positions. The number of stocks $\widehat{a}_i^{s,j}(h+1)$ desired by trader i for stock s, j at time step $h+1$ is determined accordingly, i.e.,

$$\widehat{a}_i^{s,j}(h+1) = \left\lfloor \widehat{\omega}_i^{s,j}(h+1) \cdot \frac{\widehat{W}_i^r(h+1)}{p^{s,j}(h)} \right\rfloor, \tag{3.3}$$

where the symbol $\lfloor .. \rfloor$ denotes the integer part. In order to fulfill the prescription of their random allocation strategies, agents issue buy or sell orders regarding all the assets traded in the market. The amount $\Delta_i^{s,j}(h+1)$ of the order issued by trader i at time step $h+1$ relative to stock s, j is:

$$\Delta_i^{s,j}(h+1) = \hat{a}_i^{s,j}(h+1) - a_i^{s,j}(h)\,. \tag{3.4}$$

$\Delta_i^{s,j}$ is the difference between the desired amount of stock s, j at time step $h+1$ and the real amount held in the portfolio by agent i. If $\Delta_i^{s,j} > 0$ the order is a buy order, conversely if $\Delta_i^{s,j} < 0$ the agent issues a sell order.

Every order is associated with a limit price. According with the previous models [46, 138, 162, 163], it is stipulated that buy (sell) orders cannot be executed at prices above (below) their limit price $d_i^{s,j}$ ($o_i^{s,j}$), where:

$$d_i^{s,j}(h+1) = p^{s,j}(h) \cdot N_i(\mu, \sigma_i^{s,j})\,, \tag{3.5}$$

$$o_i^{s,j}(h+1) = \frac{p^{s,j}(h)}{N_i(\mu, \sigma_i^{s,j})}\,. \tag{3.6}$$

$N_i(\mu, \sigma_i^{s,j})$ is a random draw from a Gaussian distribution with average $\mu = 1.01$ and standard deviation $\sigma_i^{s,j}$ that is proportional to the historical volatility $\sigma^{s,j}(T_i)$

of the price $p^{s,j}(k)$ of stock s, j through the equation $\sigma_i^{s,j} = \alpha\sigma^{s,j}(T_i)$. Linking limit orders to volatility takes into account a realistic aspect of trading psychology: when volatility is high, uncertainty on the "true" price of a stock grows and traders place orders with a broader distribution of limit prices. In this model, α is a constant for all agents, whereas $\sigma^{s,j}(T_i)$ is the standard deviation of log-price returns of asset s, j, computed in a time window T_i proper for agent i-th. All buy and sell orders issued at time step $h + 1$ are collected and the demand and supply curves are consequently computed. The intersection of the two curves determines the new price $p^{s,j}(h + 1)$ of stock s, j (see [163] for more details on market clearing).

Let us now consider dividends paying stocks. Each stock is characterized by a dividend $q_n^{s,j}$ which is payed every Q time steps. The dividend process $q_n^{s,j}, \quad n = 1, 2, 3, ...$ follows an autoregressive process of order 1, i.e.,

$$q_n^{s,j} = a_0^s + a_1 q_{n-1}^{s,j} + b^{s,j}\epsilon_n^s, \tag{3.7}$$

where a_0^s is a coefficient specific to the economic sector s, a_1 assumes the same value for all sector s, $b^{s,j}$ is a number specific to each stock and ϵ_n^s is a random draw from a normal distribution and is specific to each economic sectors s. The initial condition $q_0^{s,j}$ for each dividend is set to the mean value of the autoregressive process, i.e.,

$$q_0^{s,j} = \frac{a_0^s}{1 - a_1}. \tag{3.8}$$

As traders are zero-intelligent, the information about dividends is not used. In fact they choose the stocks j randomly, i.e., without considering the dividends.

3.1.3 Computational experiments

Two different market conditions characterized by constant financial resources, i.e., without dividend paying stocks and without cash inflow, and by varying financial resources, i.e., with dividend paying stocks and external cash inflow are considered. Both market conditions are characterized by 100 different stocks, each related to a particular firm. Assets are traded sequentially in the market in exchange of cash and the economy is characterized by 10 sectors each constituted by 10 firms. Both market conditions consider two different cases. In the first, agents invest on stock belonging to all the sectors ($L = 10$), whereas in the latter, each agent randomly restricts his allocation universe to 5 business sectors among the 10 available ($L = 5$). In the simulation with varying financial resources, stocks pay dividends every $Q = 63$ time steps, i.e., about three financial months. Thus, every three months agents increase their cash according the number of stocks owned. Dividends are calculated according to Eq. 3.7. The parameter a_0^s is specific to each sector and ranges from 0.05 to 0.095, with steps of 0.005 in the 10 sectors considered, a_1 is a constant set to 0.9, whereas $b^{s,j}$ is a random draw from an uniform distribution in the range [0, 0.02].

Agents are initially characterized by an equal endowment of cash and number of

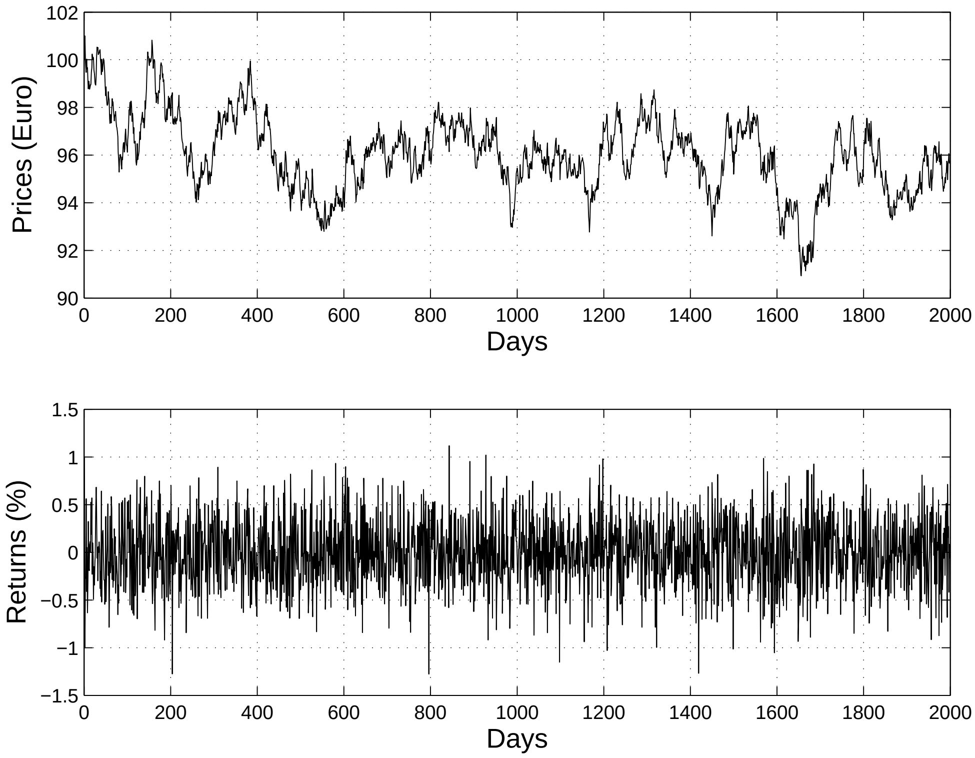


Figure 3.1: Prices and Returns of a stocks in simulations without dividends paying stocks.

stocks, set to € 10,000,000 and to 1,000 shares for each stock. Figure 3.1 shows the dynamic of prices and returns in the absence of dividend paying stock, whereas Figure 3.2 presents the same processes for a dividend paying stock. In both cases, price fluctuations are characterized by volatility clustering of returns and fat-tailed distributions.

The normal distribution of returns has been tested by Jarque-Bera test and the null hypothesis of no serial correlation has been tested by the Ljung Box Q test. These features are recovered by means of the volatility feedback effect which has been modeled explicitly in the limit price associated to orders according to Eqs. 3.5 and 3.6 for buy and sell order, respectively. Due to the finiteness of financial resources, price processes exhibit strong reversion to the mean. In the simulation without dividends paying stocks, the variance-ratio test [36] rejects the null hypothesis of random walk and the augmented Dickey-Fuller test [62], with a specification of a constant term in the deterministic part, rejects the null hypothesis of unit root. Conversely, in simulation with dividends paying stocks, the variance-ratio test rejects the null hypothesis of random walk, but the Dickey-Fuller test does not reject the hypothesis of unit root. Both tests have been performed at the significance level of

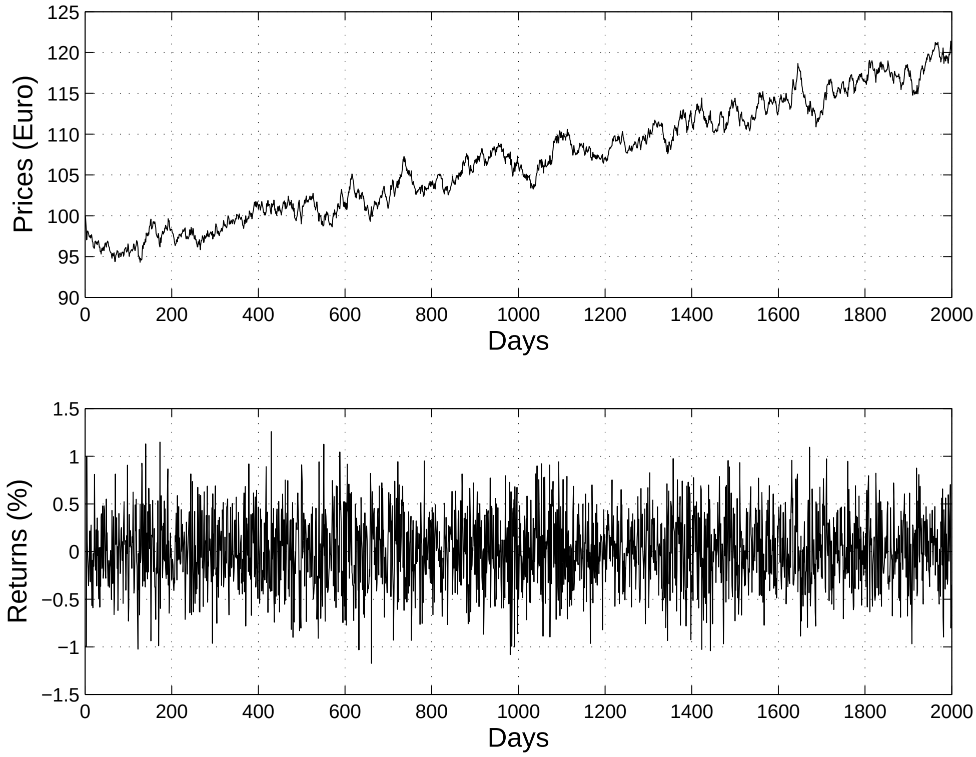


Figure 3.2: Prices and Returns of a stocks in simulations with dividends paying stocks.

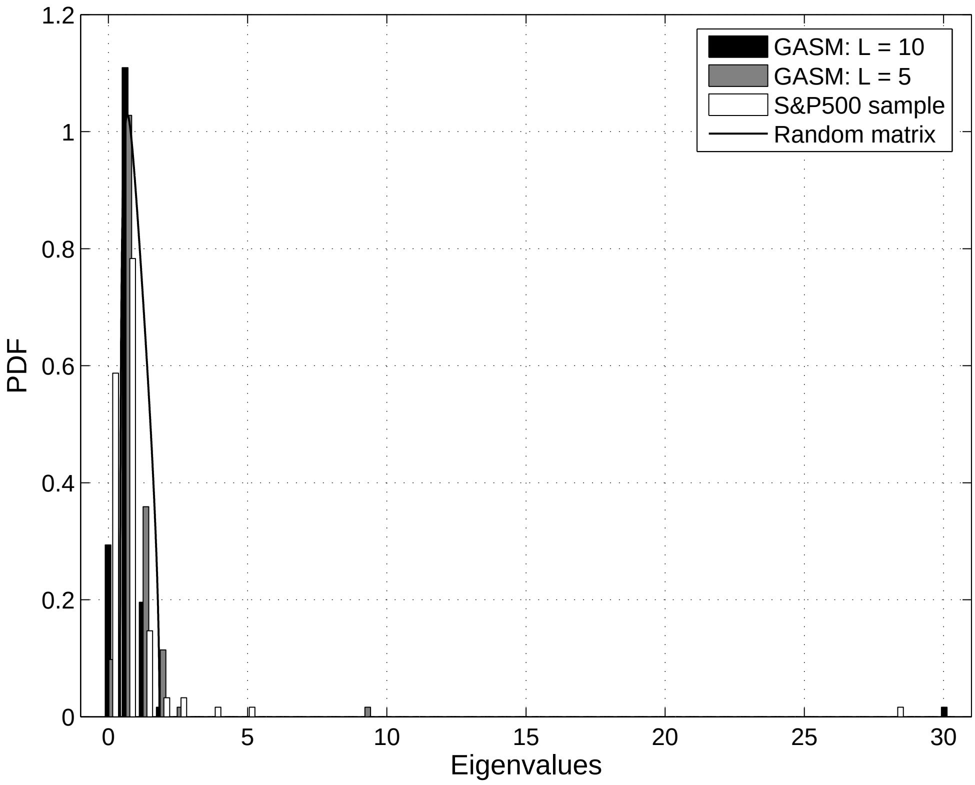


Figure 3.3: Probability density function (PDF) for eigenvalues cross-correlation matrix of returns in simulation without dividends paying stocks.

5%.

Stated these price process properties, in this study the attention has been focused on statistical properties of the multivariate process of prices and returns. In particular, following the approach recently introduced in the econophysics literature [116, 158, 159], the cross-correlations between returns of different stocks have been studied using methods of random matrix theory (RMT). Figures 3.3 and 3.4 show the probability density function (PDF) of eigenvalues of the cross-correlation matrix for the two cases considered in simulations, i.e., $L = 10$ and $L = 5$. These two cases have been represented by the black colored and the gray colored histograms, respectively.

Figure 3.3 is referred to a simulation with non dividends paying stocks, Figure 3.4 to a simulation with dividends paying stocks. In Figures 3.3 and 3.4, the theoretical PDF for random matrices is represented by the continuous line. In this case, i.e., 100 series of returns and 800 time steps, the largest eigenvalue results equal to 1.83. For the sake of comparison, the white colored histogram in Figures 3.3 and 3.4 shows the PDF of eigenvalues for a sample of 100 stocks included in the S&P 500 index in a time window of 800 days (i.e., close prices from the year 2001 to the year 2004 are considered). Results point out the presence of outliers well above the bounds deter-

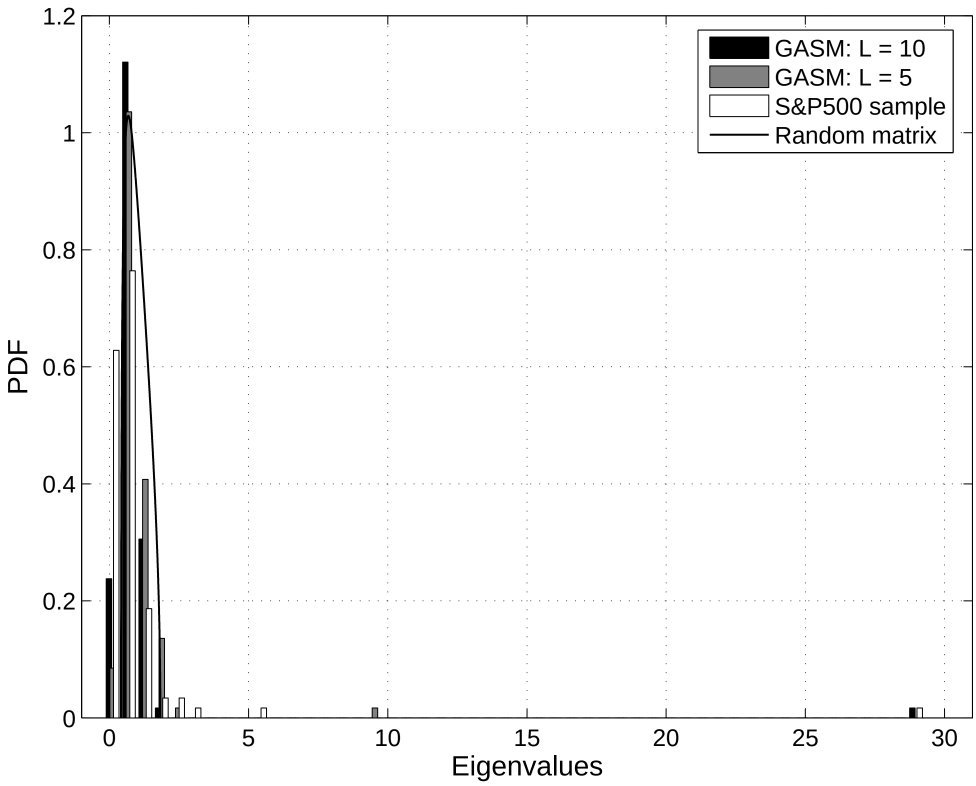


Figure 3.4: Probability density function (PDF) for eigenvalues cross-correlation matrix of returns in simulation with dividends paying stocks.

mined according to RMT.
Figure 3.5 reports the eigenvalues for the three cases considered, i.e., the simulated series with $L = 10$ and $L = 5$ number of business sectors, and the S&P 500 sample. The horizontal continuous line represents the largest eigenvalues according to the RMT. In the case of $L = 10$, only one large deviation from RMT is pointed out, whereas in the case of S&P 500 and of $L = 5$ more than 5 deviations from the largest bound set by RMT are pointed out. These deviations have been studied and interpreted recently in the literature [158]. The eigenvector related to the largest eigenvalue represents the entire market, whereas the other eigenvectors, whose eigenvalues deviate from RMT distribution, are related to the business sectors existing in the economy. In the computational experiment with a single business sector, only one deviation from RMT is present, whereas in the experiment where the number of sectors is set to 10, there are nearly 10 eigenvalues larger than the largest eigenvalue determined by RMT. Figure 3.6 shows the eigenvalues for the three cases considered, i.e., $L = 10$ and $L = 5$ number of business sectors,in the presence of dividends, and the S&P 500 sample. As clearly shown, the result is similar to the previous one. This can be explained observing that, as the traders are random the information about the dividends does not provide useful insight for trading.

In all cases, the analysis of the inverse participation ratio [158] show that the eigenvectors related the largest eigenvalue represents the entire market. Furthermore, for the simulated series with $L = 5$, the deviating eigenvalues corresponds to business sectors.
Finally, it is worth noting that the same results can be obtained also adding an external cash inflow.

3.2 Information-based multi-assets artificial stock market with heterogeneous agents

3.2.1 Introduction

The increasing interest towards complex systems characterized by a large number of simple interacting units has carried to the birth of co-operations between the fields of economics, physics, mathematics and engineering. The large availability of financial data has allowed to improve the knowledge about the price processes and many so-called stylized facts have been discovered, e.g., the fat tails of return distributions, the absence of autocorrelation of returns, the autocorrelation of volatility, the distribution of trading volumes and of intervals of trading, etc. [88, 133, 134, 136, 147]. Generally speaking, these features cannot be reproduced within the context of a single representation agent, thus resulting in a great interest in developing of artificial financial markets based on interacting agents.
In fact, in the classical approach, simple analytically tractable models with a representative, perfectly rational agent have been the main corner stones and mathematics

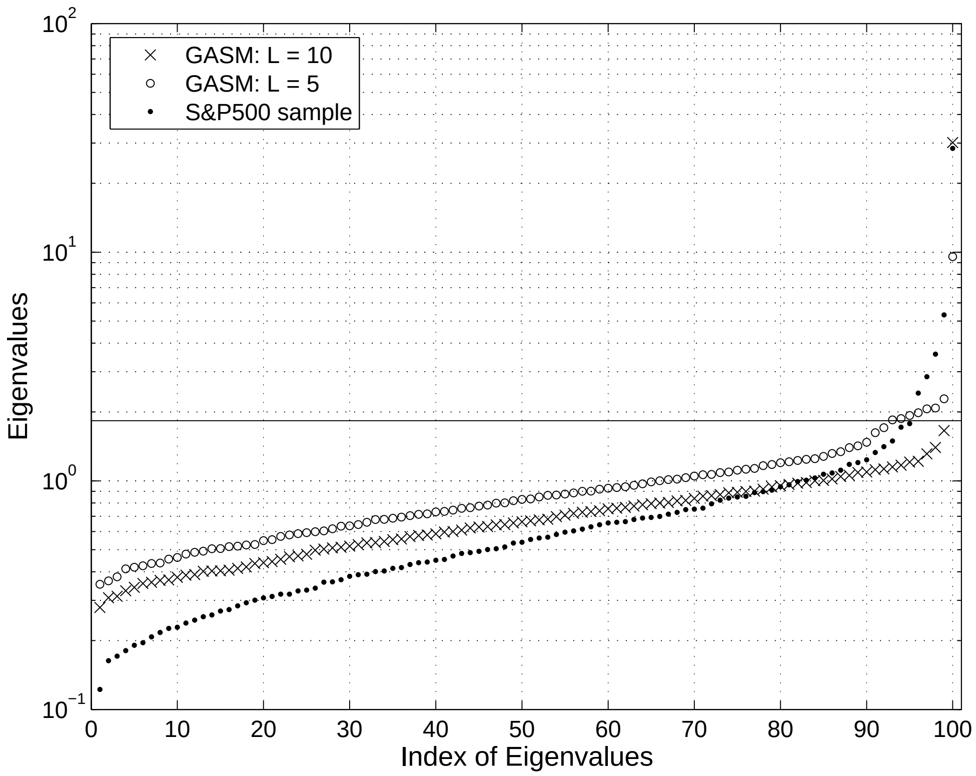


Figure 3.5: Eigenvalues of the cross-correlation matrix of returns without dividends paying stocks.

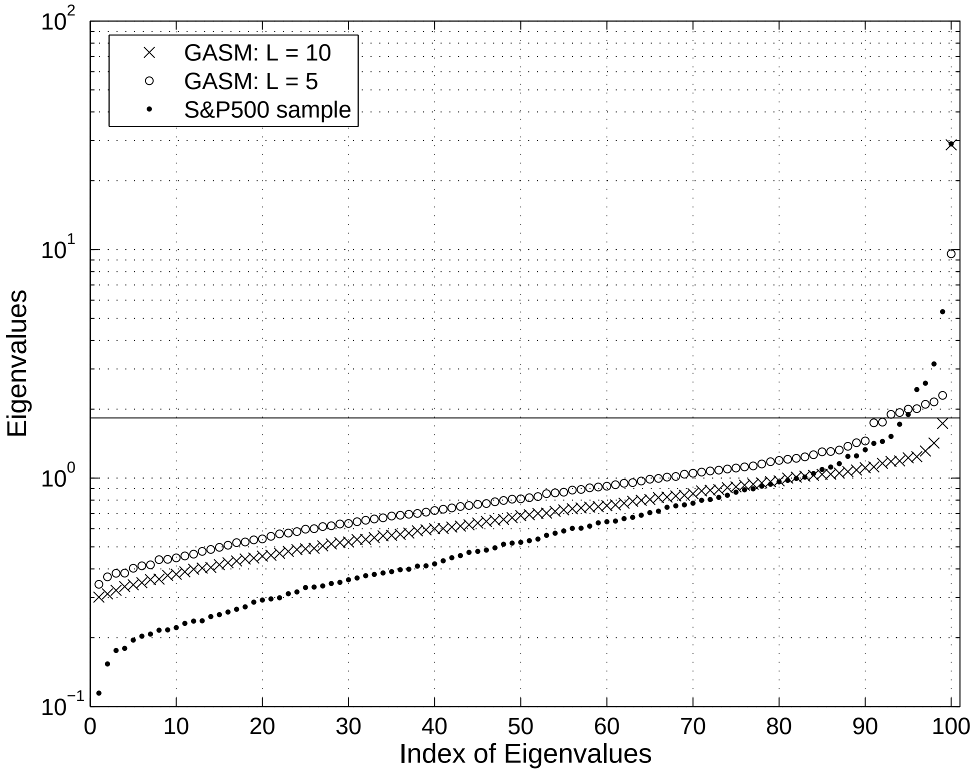


Figure 3.6: Eigenvalues of the cross-correlation matrix of returns with dividends paying stocks.

has been the main tool of analysis. Conversely, new behavioral approach, characterized by markets populated with boundedly rational, heterogeneous agents using rule of thumb strategies, fits much better with agent-based simulation models and computational and numerical methods have become an important tool of analysis. Over the last 15 years, a number of computer-simulated, artificial financial markets have been built; LeBaron [118] offers a review of recent work in this field. Following the pioneering work done at the Santa Fe Institute [117, 154], a large number of researchers have proposed artificial markets populated with heterogeneous agents endowed with learning and optimization capabilities.

This lead to several examples of artificial stock markets proposed in the literature, e.g., Santa Fe Institute Artificial Stock Market [11] and the Genoa Artificial Stock Market (GASM) [138, 162, 163]. Generally speaking, in the framework of artificial stock market, attention has been focused on single asset artificial stock markets. This in order to understand and to reproduce the main stylized facts of an univariate price process. Only recently, an extension of the GASM to multi-assets environment has been proposed [49]. In that case, the GASM was populated by zero-intelligence traders and computational experiments pointed out the possibility to reproduce some stylized facts both in terms of the single price process and of the aggregate behavior. However, results suggested a reduced capability in reproducing the well known unitary root stylized fact, as it was obtained only in the presence of exogenous cash inflow.

This limitation can be overcame employing recent results on a single-asset artificial stock market based on information propagation [48]. Indeed, the information-based artificial stock market proposed in [48] was able to reproduce unitary root also in an endogenous framework.

This framework deals with a multi-assets framework, where the market is populated by heterogeneous agents that are seen as nodes of sparsely connected graphs. The market is characterized by different types of stocks and agents trade risky assets in exchange for cash. Beside the amount of cash and of assets owned, each agent is characterized by sentiments. Moreover, agents share their sentiments by means of interactions that are determined by graphs. Agents are subject to a portfolio choice on number and type of risky securities. The allocation strategy is based on sentiments and wealth. A central market maker (clearing house mechanism) determines the price processes for each stock at the intersection of the demand and the supply curves.

The validation method followed in this paper is the capability of the information-based artificial stock market to reproduce the stylized facts for univariate and multivariate price processes. Concerning univariate processes, the three main stylized facts are taken as reference, i.e., unitary root of price processes, fat-tails distribution of returns and volatility clustering.

The multi-assets environment offers a new set of stylized facts for validation, i.e., the statistical properties of cross-correlation matrices of returns [116, 157, 159, 201] and of variance-covariance matrices of prices [191], that make reference to static and dy-

namic factors, respectively.
The results show that the main statistical properties of univariate and multivariate price processes are reproduced in an endogenously framework. This points out the importance of connection structure among the agents.

3.2.2 The model

In this section, the model of an informed multi-assets artificial stock market is presented. The model makes reference to the Genoa Artificial Stock Market - GASM. Heterogeneous and informed agents trade risky assets in exchange for cash depending on the interactions among agents. They are modeled as liquidity traders, i.e., decision making process is constrained by the finite amount of financial resources (cash + stocks) they own. At the beginning of the simulation, cash and stocks are distributed randomly among agents.
Let N be the number of traders and K the number of assets. Let k denote the particular asset.
For each asset, the traders of the market are organized according to a directed random graph, where the agents are the nodes and the branches represent the interactions among agents. The graphs are responsible of the changes in agent's sentiments. The graphs are directed, i.e., the interactions are assumed unidirectional (i.e., agent $j-th$ influences agent $i-th$ but not necessarily vice versa) and characterized by a strength g_{ji}^k, assumed a positive real number. Generally speaking, due to the presence of a directed graph, both an output node degree, related to the output branches of a given node, and an input node degree, related to the input branches, should be defined.
At each time step h, information is propagated through the market and sentiments of agent $i-th$ is updated. Let $\Im_i^k$ the set of agents that influences the behavior of trader $i-th$ for the asset k. The new sentiments S_i^k of agent $i-th$ for each asset are functions of the previous sentiments, of the influence of interacting agents, of the log return (market feedback) and of average sentiment of the agent about the market behavior according to the equation

$$S_i^k(h+1) = F(\alpha_{1,i}S_i^k(h) + \alpha_{2,i}\hat{S}_i^k(h) + \alpha_{3,i}r^k(h) + \alpha_{4,i}\tilde{S}_i(h)) \tag{3.9}$$

where

$$\hat{S}_i^k(h) = \frac{\sum_{j\in\Im_i^k} g_{ji}^k S_j^k(h)}{\sum_{j\in\Im_i^k} g_{ji}^k} \tag{3.10}$$

represents the influence of interacting agents,

$$r^k(h) = log[p^k(h)] - log[p^k(h-1)] \tag{3.11}$$

represents the market feedback and

$$\tilde{S}_i(h) = \frac{\sum_k S_i^k}{K} \tag{3.12}$$

models the global vision of agent $i-th$ for the market trend. This last term is a stabilizing element for the sentiment, so that the coefficient $\alpha_{4,i}$ in Eq. 3.9 is always

negative. Moreover the non-linear function in Eq. 3.9 is an hyperbolic tangent, i.e., $F(x) = tanh(2x)$ that constraints agent sentiments in the range $(-1, 1)$. Finally, a constraint on graph interaction is considered $|\alpha_{2,i}| = (0.6 - |\alpha_{1,i}|)$ with randomly change in sign at each time step. The meaning is that sometimes an agent changes idea about the sentiments of neighbor, and so he changes his reaction.
The agents are ranked according to a Zipf law, i.e., the importance of each agents is approximately inversely proportional to its rank. All the parameters of the agents are calculated according to such a ranking. Moreover, for each stock an agent is randomly connected to a set of other agents whose number and strength g_{ij}^k are inversely proportional to his rank, i.e., richer agents influences a larger number of agents with a higher strength. Consequently, the output degree distributions over the nodes are set to power laws and the input degree distributions result power laws too. Furthermore, agent's trading decision is based on cash and stocks owned and on sentiment. Thus, the stock price processes depend on the propagation of information among the interacting agents, on budget constraints and on market feedbacks. In this respect, also the α coefficients in Eq. 3.9 are inversely proportional to agent's rank, i.e., richer agents have stronger beliefs.
Let $S_i^k(h)$ be the sentiment, $C_i(h)$ the amount of cash, $q_i^k(h)$ the amount of asset k owned by the i-th trader at time h.
If $p^k(h)$ is the market price of the risky asset k at time step h, the risky wealth $W_i^r(h)$ owned by trader i at time step h is:

$$W_i^r(h) = \sum_k q_i^k(h) p^k(h) \tag{3.13}$$

whereas $W_i(h) = c_i(h) + W_i^r(h)$ represents the total wealth of agent $i - th$.
At each simulation step, trader $i - th$ tries to allocate in risky assets a fraction γ_r of his total wealth related to his vision of the market trend, i.e.,

$$\widehat{W}_i^r(h+1) = \gamma_r(h) W_i(h), \tag{3.14}$$

where $\gamma_r = \frac{1+\tilde{S}_i(h)}{2}$.
$\tilde{S}_i$ is the average sentiments of all assets described by Eq. 3.12. The symbol $\widehat{\cdot}$ denotes that $\widehat{W}_i^r(h+1)$ is the amount that agent $i - th$ desires to allocate in the risky investment, whereas the real amount $W_i(h+1)$ effectively allocated in stocks will depend on the trading process with the other agents. For each assets, a positive sentiment denotes a propensity to buy, whereas a negative sentiment corresponds a propensity to sell. In this model only long positions are allowed. Thus, if the agent $i - th$ is characterized by a positive sentiment for asset k , the quantity desired of risky asset k is given by:

$$\widehat{q}_i^k(h+1) = \left\lfloor \frac{\gamma_a^k \widehat{W}_i^r(h+1)}{p^k(h)} \right\rfloor \tag{3.15}$$

where γ_a is given by:

$$\gamma_a^k = \frac{S_i^k}{\sum_{k \in A_i} S_i^k}. \tag{3.16}$$

A_i is the subset of assets in which the sentiments of agent $i-th$ are positive, i.e.,$\sum_k \gamma_a^k = 1$. The symbol $\lfloor .. \rfloor$ in Eq. 3.15 denotes the integer part. Conversely, if the sentiment relatives to asset k is negative, the agent $i-th$ is characterized by a desired quantity $\widehat{q}_i^k(h+1) = 0$.
The amount $\Delta_i^k(h+1)$ of the order issued by trader $i-th$ at time step $h+1$ relative to stock k is:

$$\Delta_i^k(h+1) = \hat{q}_i^k(h+1) - q_i^k(h)\,. \tag{3.17}$$

Δ_i^k is the difference between the desired amount of stock k at time step $h+1$ and the real amount held in the portfolio by agent $i-th$. If $\Delta_i^k > 0$ the order is a buy order. Conversely if $\Delta_i^k < 0$ the agent issues a sell order. Every order is associated with a limit price. According to previous models [46, 138, 162, 163], it is stipulated that buy (sell) orders cannot be executed at prices above (below) their limit price d_i^k, i.e.,

$$d_i^k(h+1) = p^k(h) \cdot N_i(\mu_i^k, \sigma_i^k) \tag{3.18}$$

$N_i(\mu_i^k, \sigma_i^k)$ is a random draw from a Gaussian distribution with average

$$\mu_i^k = (1 + sgn(\Delta_i^k)|S_i^k|). \tag{3.19}$$

It is worth noting that for a buy order (i.e., $\Delta_i^k > 0$) in average $d_i^k(h+1) > p^k(h)$, otherwise for a sell order (i.e., $\Delta_i^k < 0$) in average $d_i^k(h+1) < p^k(h)$. Furthermore, the standard deviation σ_i^k is proportional to the historical volatility $\sigma^k(T_i)$ of the price $p^k(h)$ of stock k through the equation $\sigma_i^k = \xi\sigma^k(T_i)$. Linking limit orders to volatility takes into account a realistic aspect of trading psychology: when volatility is high, uncertainty on the "true" price of a stock grows and traders place orders with a broader distribution of limit prices. In this model, ξ is a constant for all agents, whereas $\sigma^k(T_i)$ is the standard deviation of log-price returns of asset k, computed in a time window T_i proper for agent $i-th$ randomly associated to the agent.
All buy and sell orders issued at time step $h+1$ are collected and the demand and supply curves are consequently computed. The intersection of the two curves determines the new price(clearing price) $p^k(h+1)$ of stock k (see [163] for more details on market clearing).
Buy and sell orders with limit prices compatible with $p^k(h+1)$ are executed. After any transactions, traders' cash, portfolio and sentiments are updated. Orders that do not match the clearing price are discarded.

3.2.3 Computational experiments

Generally speaking, the main objective of an artificial market is to reproduce the statistical features of the price process with minimal hypotheses about the intelligence of agents. This general philosophy is used in the model described in the previous section to reproduced the main stylized facts of the real markets both for univariate and multivariate cases. As regarding the single-asset market three main stylized facts, i.e.,

the integrated I(1) property of prices processes, the presence of volatility clustering in the returns time-series and the presence of fat tails distributions of the returns are considered. Concerning the multi-assets market two statistical properties have been recently assumed as stylized facts. The first is related to the returns processes, i.e., static factors evidenced by the random matrix property of the cross-correlation matrix of returns [158, 159], whereas the second concerned the common trends of the price processes in terms of the variance-covariance matrix of prices[191].
As it will be shown, the proposed model is able to reproduce in an endogenous framework all these stylized facts, i.e., univariate, multivariate, statics and dynamic. It is worth remarking the importance of this result, as for the first time, an artificial stock market reproduce endogenously all this features.
The market conditions adopted for the computational experiments are characterized by 100 different stocks. The number of agents is set to 2,278 that are initially endowed by a random distribution of cash and number of stocks. Furthermore, time window T_i for the calculation of the historical standard deviation is randomly chosen by a uniform distribution in the range $(10, 100)$.
Figure 3.7 shows the prices processes for the $k = 100$ assets. The price processes point out relevant differences, depending on the specific nature of the asset. Indeed, starting from identical initial prices, after a transient (not shown in the Figure 3.7 for the sake of compactness) price levels result significantly different. This suggest a possible herding behavior induced by the graphs that drives the agent propensities to buy/sell the assets.

Moreover, focussing attention to a single asset, it allows one to verify the main stylized facts of a univariate price time-series. Figure 3.8 shows the prices and the return process of asset number 9.

Moreover, the corresponding autocorrelation function of raw returns, of absolute value of returns and of the square returns are plotted in Figure 3.9. Volatility clusters are well underlined in Figure 3.8 and a long memory effect is further pointed out in Figure 3.9 by the autocorrelation function of absolute value of returns and square returns. As clearly pointed out in Figure 3.9, autocorrelation of raw returns shows immediate decay within noise level of the correlation just after one lag, whereas absolute value of returns and of square returns exhibits slow decay of the autocorrelation.

Focussing attention on statistical evidences, the normal distribution of returns has been tested by the Jarque-Bera test and the unitary root has been tested by the Augmented Dickey-Fuller test at the significance level of 5%. Table 3.1 summarized the corresponding results. Moreover, for the sake of comparison, the results obtained for the case of 100 assets randomly selected within the S&P500 are also included. NrI1 refers to the number of assets that do not reject the hypothesis of unitary root according the Augmented Dickey-Fuller test (ADF test). NrFTJB refers the number of assets whose returns process does not follow a normal distribution according to Jarque-Bera test. It is worth noting that the value of GASM data are in very close agreement with those of the S&P500. These evidences, together with the volatility

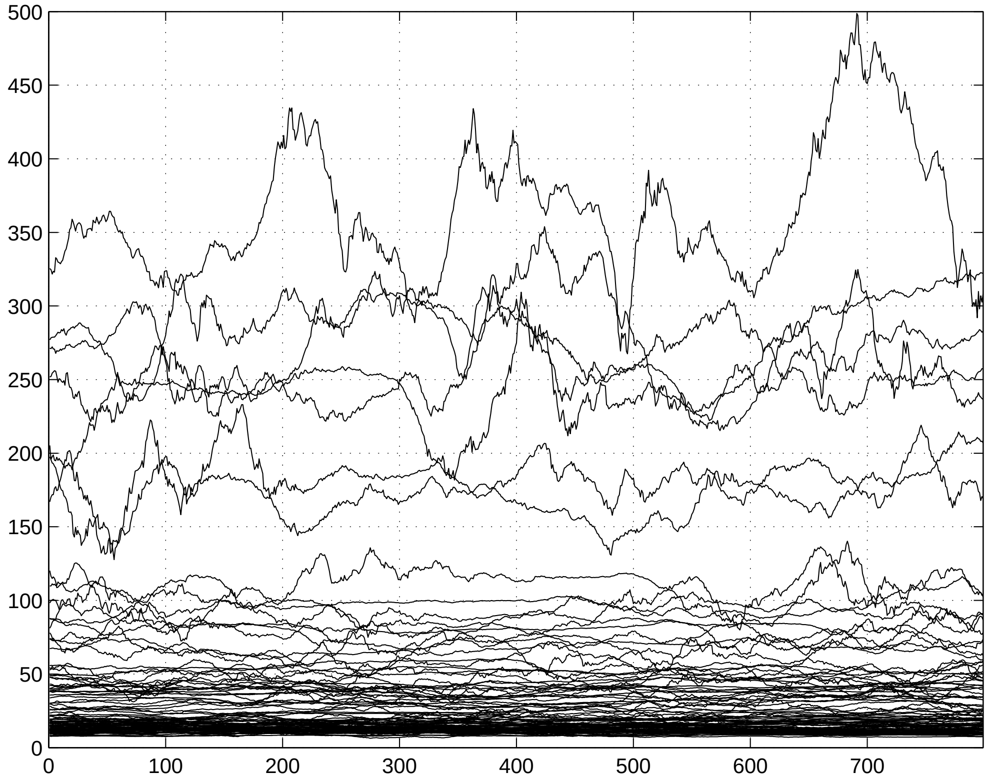


Figure 3.7: Price processes of k=100 assets

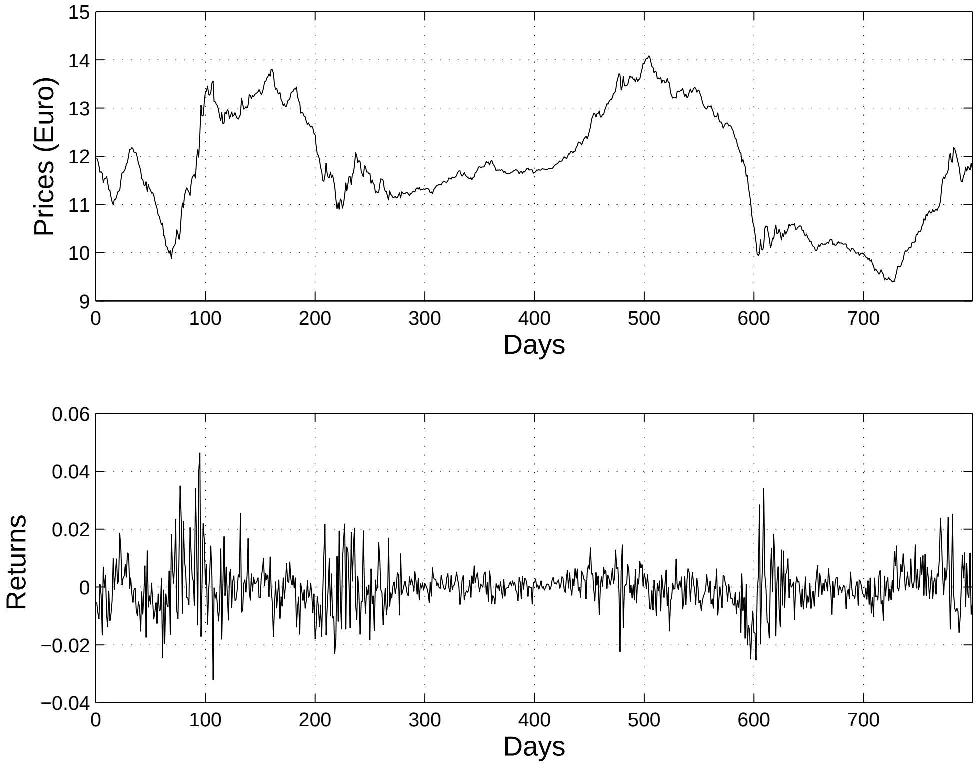


Figure 3.8: Price process and returns of stock number 9

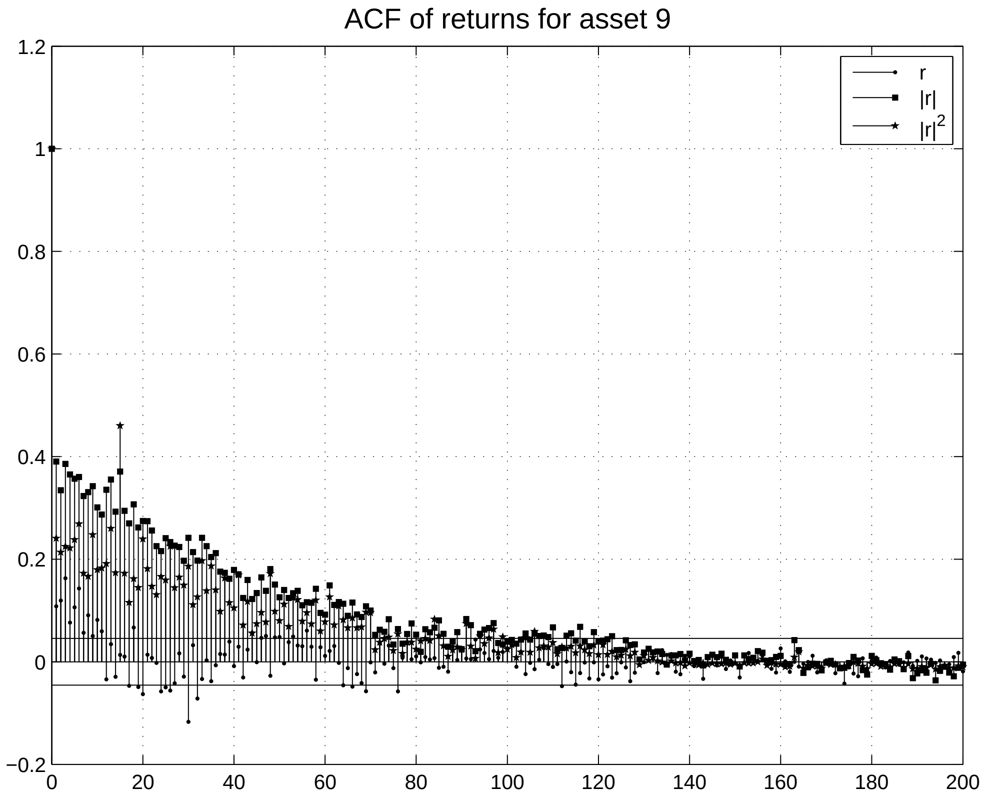


Figure 3.9: Autocorrelation of stock number 9

Table 3.1: ADF test and Jarque-Bera test for GASM data and real data(S&P500)

Data	NrI1	NrFTJB
SASM	99	100
S&P500	83	100

clustering property previously pointed out, allow, one to conclude that the GASM is able to reproduce endogenously the main stylized facts of univariate processes.

Stated the univariate price process properties, the attention has been focused on the statistical properties of the multivariate process of prices and returns. Generally speaking, the analysis of multivariate stylized facts leads to the definition of factor models. Furthermore, in the context of factor models, two main classes can be identified, i.e., static factor and dynamic factor models. Concerning the former class, attention is payed to returns as the return processes result (in the first approximation) stationary. In particular, the risk of a security can be described as superposition of different source of risks (also described by stationary processes) and this general formulation is the basic for classical portfolio theory and risk management, e.g., CAMP, multifactors CAMP, APT, etc.[75, 143, 170, 178].
Conversely, in the case of dynamic factors attention is payed to prices and the main employed concept is co-integration. In particular, it is assumed that in a large market it is not possible to reject the hypothesis of integrated univariate price processes, but these price processes are not independent. Indeed, only few independent integrated processes can be identified, whereas all the others are co-integrated, i.e., it is possible to identify linear (lagged) combinations that result stationary (so called cointegration equations) [75, 143, 170, 178].
In this chapter, the static factors are studied according to the cross-correlation matrix of returns. In particular, following the approach recently introduced in the econophysics literature [116, 158, 159], the cross-correlations between returns of different stocks have been studied using methods of random matrix theory (RMT). Figure 3.10 shows the probability density function (PDF) of eigenvalues of the cross-correlation matrix. The PDF of eigenvalues of GASM data is presented in black colored histogram and the theoretical PDF for random matrices is represented by the continuous line. In this case, i.e., 100 series of returns and 800 time steps, the largest eigenvalue results equal to 1.83. For the sake of comparison, the white colored histogram in Figure 3.10 shows the PDF of eigenvalues for a random sample of 100 stocks included in the S&P 500 index in a time window of 800 business days (i.e., close prices from the year 2001 to the year 2004 are considered). Results point out the presence of outliers well above the bounds determined according to RMT in perfect agreement to the empirical evidence shown by S&P500 data.

Furthermore, Figure 3.11 plots the GASM data and the S&P 500 sample. The horizontal continuous line represents the largest eigenvalues according to the RMT. In this case, some deviations from the largest bound set by RMT are pointed out. These

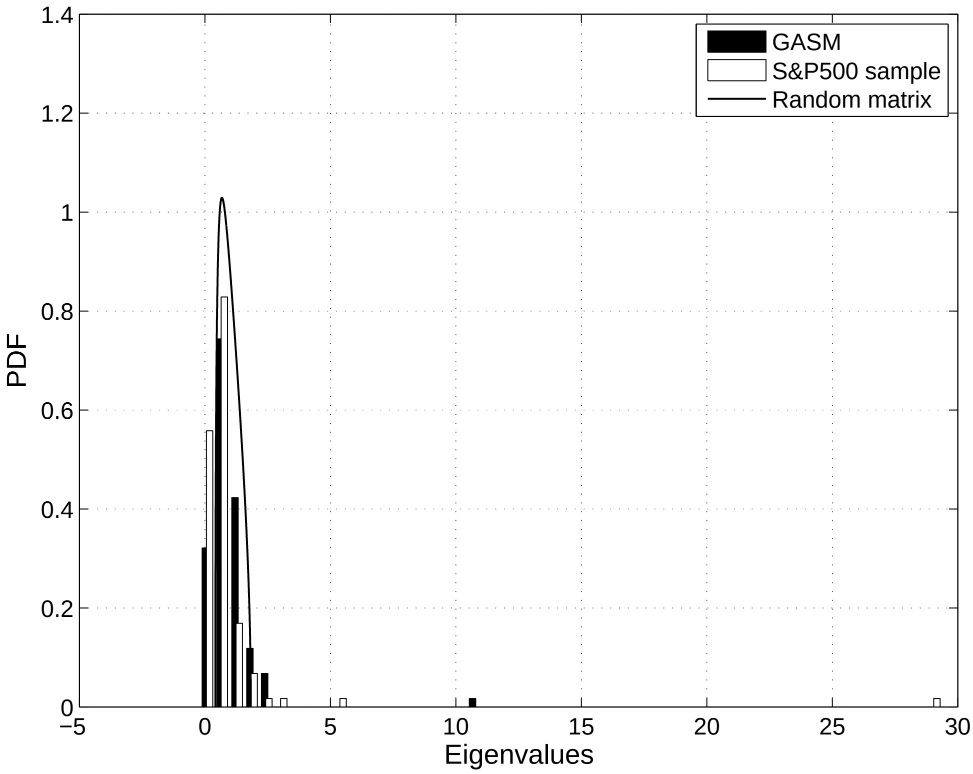


Figure 3.10: Probability density function (PDF) for eigenvalues cross-correlation matrix of returns

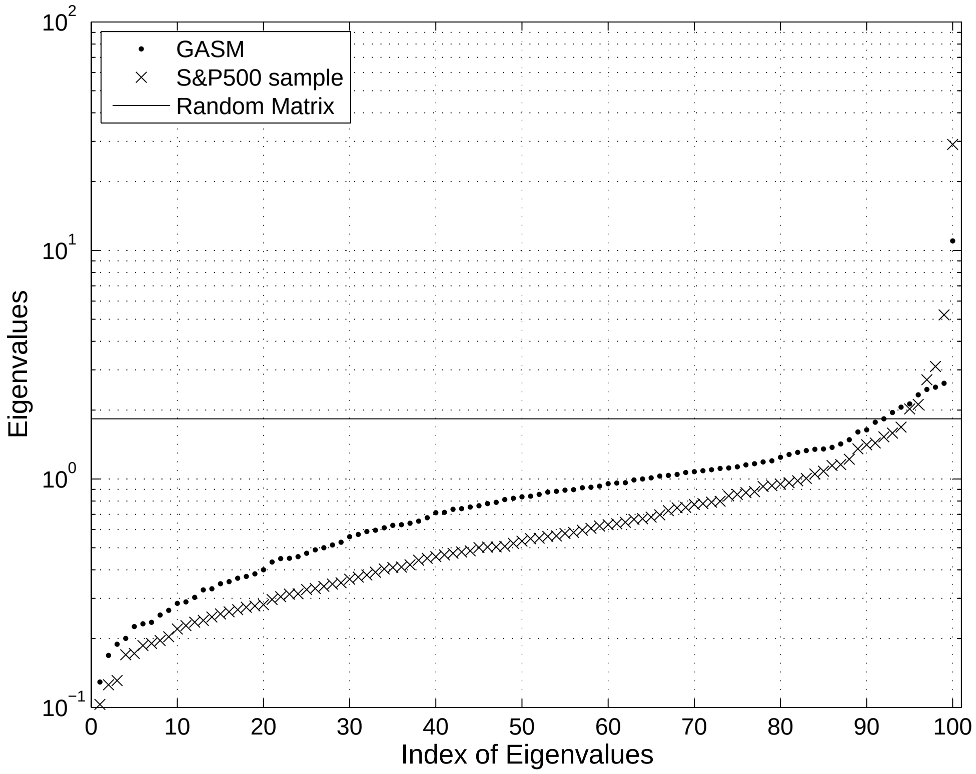


Figure 3.11: Eigenvalues of the cross-correlation matrix of returns

deviations have been studied and interpreted recently in the literature [158]. The eigenvector related to the largest eigenvalue represents the entire market, whereas the other eigenvectors, whose eigenvalues deviate from RMT distribution, are related to the business sectors existing in the economy. Also this feature is pointed out in the computer experiments, but is not included here for the sake of compactness.

As regarding the dynamic factors, they have been studied according to the variance-covariance matrix of prices. According to real data, only a reduced number of assets prices series in a market are independent integrated processes [191]. In fact, the analysis of prices processes shows that financial assets are random walk, i.e., I(1) processes, but aggregate of financial assets exhibits cointegration, i.e., linear combination of I(1) process results a stationary stochastic I(0) process. The analysis of this data has been carried on following the procedure described by Stock and Watson [191]. In particular, the PCA analysis on the variance-covariance matrix of prices allows one to identify portfolios with minimum variance. Conversely to price processes, these portfolios, i.e., linear combination of prices, generally accept the hypothesis of stationarity [191]. This feature can be verified by the ADF test at significance level of 5%. Figure 3.12 shows the results of the ADF test for the series of a random sample of 100 assets within the S&P500. As clearly stated, only a reduced number of portfolios (i.e., equal to 9) reject the hypothesis of stationarity. These series are the

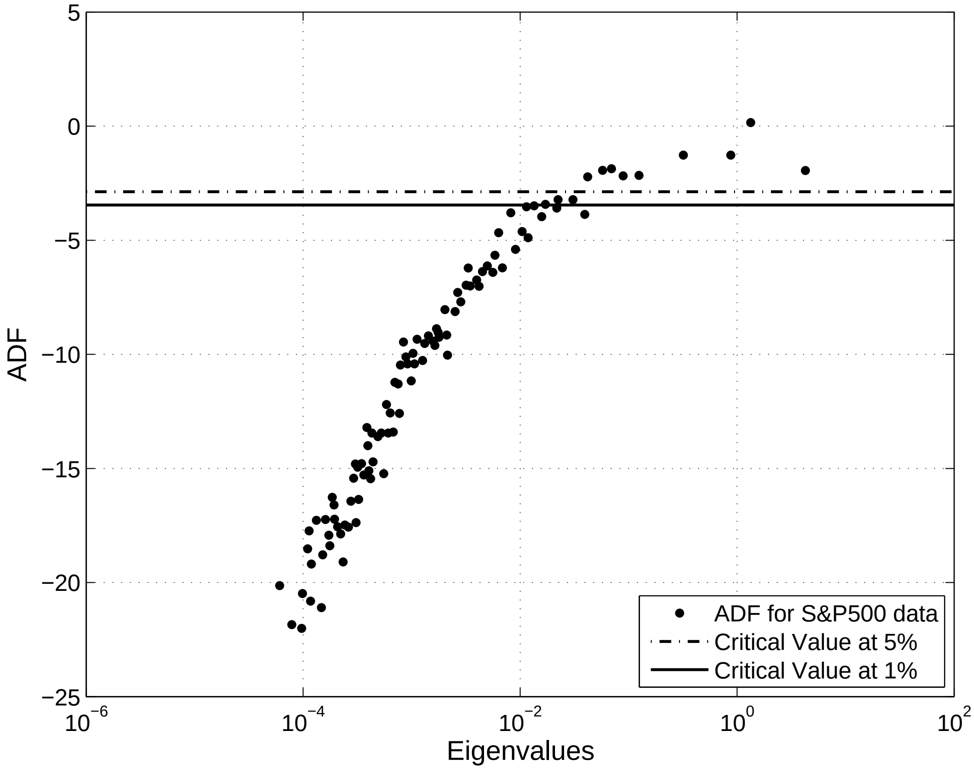


Figure 3.12: ADF test statistics of the cointegration portfolios in the case of S&P500

only independent I(1) processes, i.e., the common trends of the aggregate, whereas there exist 91 cointegration equations (i.e., the I(0) portfolios). Figure 3.13 shows the same analysis in the case of GASM data. As clearly pointed out, also in the case of the artificial stock market, only a reduced number of portfolios (i.e., equal to 12) reject the hypothesis of stationarity. These series that are the only independent I(1) processes, i.e., the common trends of the aggregate.

Thus, the proposed information-based artificial stock market is able to reproduce the statistical properties of single-asset environmental as well as the stylized facts of multi-assets. Moreover, the multi-assets properties make reference to static and dynamic factors, in close agreement with empirical evidences for real data. Finally, it is worth noting that both the stylized facts on returns and on prices are obtained in endogenous way, i.e., without any external framework.

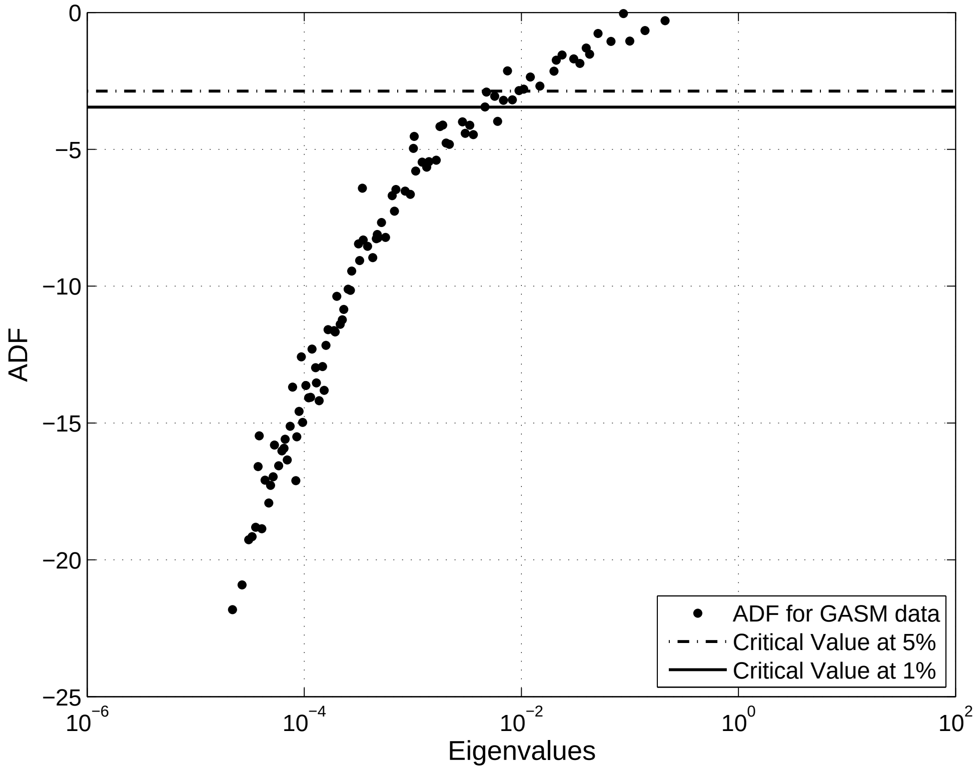


Figure 3.13: ADF test statistics of the cointegration portfolios in the case of GASM

Chapter 4

Modeling and Statistical Analysis of the Firms Growth Problem

Overview

4.0.4 What is the Firm Growth Problem

The central task of complex system and statistical physics is to understand macroscopic phenomena that result from microscopic interactions among many individual components driven by competing forces. A firm clearly is an extremely complex system which includes lots of competing divisions.

The firm growth problem is based on statistical quantification of size changes of firms. The size of a firm can be measured by the number of employees, the expenditures of firm, the total assets of a company, or any other measurable quantity that can describe the size of a company. Firm size is a random variable and it fluctuates with time due to different reasons: sometimes a company grows by obtaining a new patent or contract due to the invention of a new technique, or as an effect of new management staffs or techniques; but sometimes the company shrinks because of some changes of industry policy, or bad news in the market. Firm growth rate is a derived quantity from firm size and it usually is defined as

$$g \equiv \ln(S(t+1)/S(t)), \tag{4.1}$$

where $S(t)$ is firm size at time t. In this work, firm size S is primarily measured by the sales of organizations.

It has been discovered that these seemingly random and complex phenomena exhibit some universal behaviors no matter what industry a company gets involved with: first, the distribution of firm size exhibits a stable shape (usually claimed as a log-normal distribution); second, the distribution of firm growth rates (the quantity derived from firm size) has a Laplace-distributed shape; third, the standard deviation of growth rate is power-law related to firm size with power-law exponent about 0.2. These findings above are independent of the specific database, which implies that there is a common principle that manipulates the dynamics of firm growth, and that this common principle may be independent of the particular business that a firm does. Therefore, the above empirical findings deserve deeper understanding and explanation. Economists have not anticipated them, and this study attempts to reproduce them with stochastic models based on ideas from complex system and statistical physics. Interpretation of these models provides new insights into how the empirical regularities may arise.

The firm growth problem focuses on the statistical properties of firm growth rate, e.g. the distribution of growth rate, the mean growth rate, the standard deviation of growth rate, and the relationship between growth rate and other economic quantities. Firm growth rates interest economists because growth rate means the change of firm size or firm scale, which may be related to the profits of a firm. Whether a firm is profitable or not is a standard to measure whether it performs well in the market. So far, the distribution of growth rates of business firms is an unsolved topic in economics [17, 26, 50, 55, 57, 63, 68, 69, 77, 79, 81, 82, 85, 86, 92–95, 99–101, 105, 108–110, 125, 145, 149, 153, 165, 176, 177, 181, 184, 185, 188, 190, 194, 202, 205, 210]. However, the complex system theory and the statistical physics approaches can make some contributions, because (i) each firm is a complex system and that includes many individual components, and (ii) the firms usually are driven by competing forces, such as entry and exit processes of companies or products, and merging and splitting processes for bigger profit or survival. To explain the empirical findings by utilizing basic theories of random process and mature statistical physics models, let to understand the growth of firms more deeply, and may develop better strategies to manage or even control the growth of companies, so as to improve the economy of a country. Because companies are the fundamental building-bricks of the whole economy of a country, people may live better if firms perform well.

4.0.5 The Distribution of Firm Size S

In this section the empirical findings on the distribution of firm size $P(S)$ is presented. The concept of firm can be extended to another concept—"organization" because many non-economic subjects also share the same characteristics as companies. Organizations here are the classes whose units usually interact with each other and may agglomerate or split to form a new class. The size S here is a dynamic and measurable quantity of some classes, e.g., it can be the GDP for a country, the annual sales for a company, the size of a city measured by its area or population, or the income for individuals. Thus, A organization can be a country (Lee 1998; Canning, 1998), a business firm (Stanley, 1996; Amaral, 1997; Buldyrev, 1997; Takayasu, 1998; Sutton, 2000; Wyart, 2002), a university or research institute (Plerou, 1999), a voluntary social organization, or a bird species (Keitt, 1998; Keitt, 2002).

The size distributions of organizations are typically 1) either Pareto (power law) distribution whose probability density function (PDF) is given by

$$P(S) = \left(\frac{S}{S_{\min}}\right)^{-k} \tag{4.2}$$

where S is any number greater than $S_{\min}$, which is the (necessarily positive) minimum possible value of S, and k is a positive parameter, or log-normal distribution whose PDF is

$$P(S) = \frac{e^{-(\ln S)^2/2\sigma^2}}{\sigma\sqrt{2\pi}S} \quad (S \geq 0;\ \sigma > 0). \tag{4.3}$$

On a log-log plot, the PDF for the Pareto distribution $\rho(S)$ is a straight line but the one for the log-normal distribution, $\rho(\ln S)$, is like a parabola.

Since Pareto and log-normal are two typical skewed distributions, we conclude that the size distributions of organizations are skewed. In the following sections, we briefly show the size distribution of different subjects.

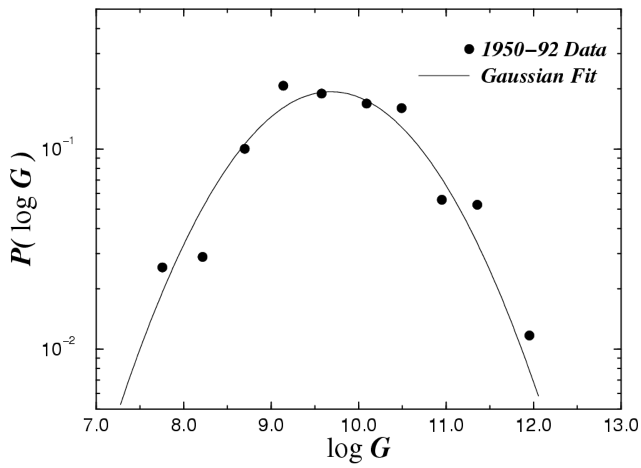


Figure 4.1: Probability distribution of the logarithm of GDP, $P(log(G))$ where G is as S here. The data have been detrended by the average growth rate, so values for different years are comparable. The data points are the average over the entire period, 1950-1992, and the continuous line is a Gaussian fit to the data. The bins were chosen equally spaced on a logarithmic scale with bin size 0.5. The distribution is claimed stationary — i.e., remains the same for different time intervals. (From Lee et al. [120])

4.0.5.1 Country

Countries are clearly complex organizations because they are usually composed of millions of people and companies. The size of a country can be measured by the total population or the area of the country. Lee and Canning *et al.* [38, 120] firstly analyzed the gross domestic product (GDP) of 152 countries during the period 1950–1992. As Fig. 4.1 shows, the distribution of GDP can be well fitted by Gaussian function of logarithmic scale which means that country GDP follows a log-normal distribution.

Buldyrev [34] *et al* find that the size distribution of countries (PDF), if measured by the quantity area or population, follows an inverse power-law with an exponent about 1 (Fig. 4.2).

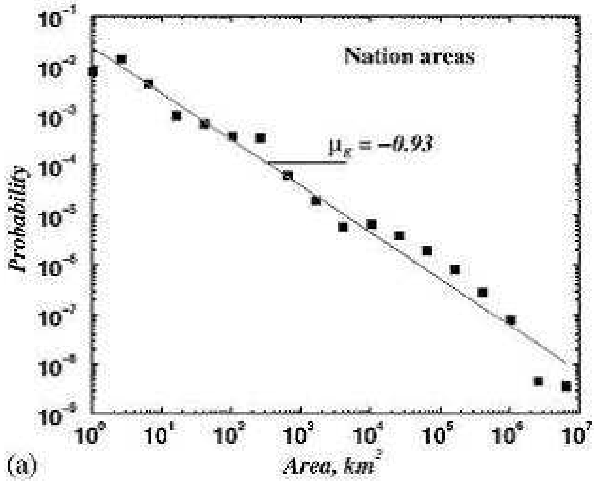


Figure 4.2: The area and population distributions of nations, $P(S)$. Double-logarithmic plot of the histogram of areas, A, of the 255 nations of the world in 1998; the linear regression coefficient is $\mu_N \approx 0.93$. The bins were chosen equally spaced on a logarithmic scale with bin size 0.5. Using population as alternative measure of size, almost identical exponents are found. (From Buldyrev et al. [34])

4.0.5.2 City

Compared with countries, cities are complex organizations on a lower level because a country contains many cities. Zipf [210] observed that the population distribution of cities follows a power-law behavior with exponent about 1 (Fig. 4.3).

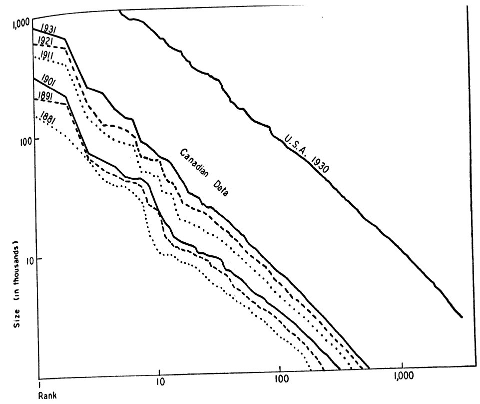


Figure 4.3: The population for Canadian communities of 1,000 or more inhabitants in 1881-1931, ranked in decreasing order of population size. (From Zipf et al. [210])

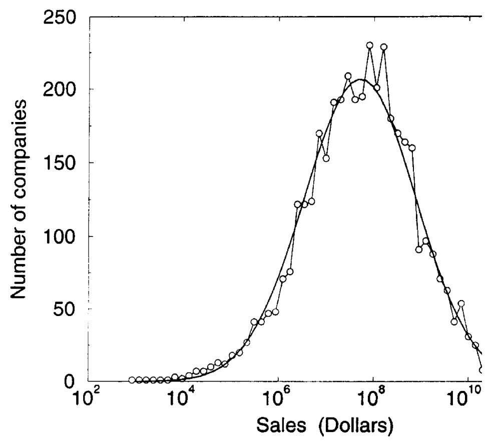


Figure 4.4: Distribution of firm size. The circles are a histogram showing the number of firms having 1993 sales of X dollars as a function of $\log X$. The data are for the 4701 Compustat firms in SIC code 2000-3999. The values of the sales are binned in powers of $\sqrt{2}$. The solid curve is a log-normal fit to the data using the mean of the log of sales and the standard deviation of the log of sales as fitting parameters. (From Stanley et al. [188])

4.0.5.3 Company in Different Industries

Firms are also complex systems since a firm is composed of different divisions or production units. In 1995, M. Stanley, *at al.* [188] tested the 1993 sales of 4071 manufacturing firms (SIC codes 2000-3999) from the Compustat database. As Fig. 4.4 shows, the distribution of firm sizes is well fit by a log-normal function. The data in other years also leads to similar conclusions.

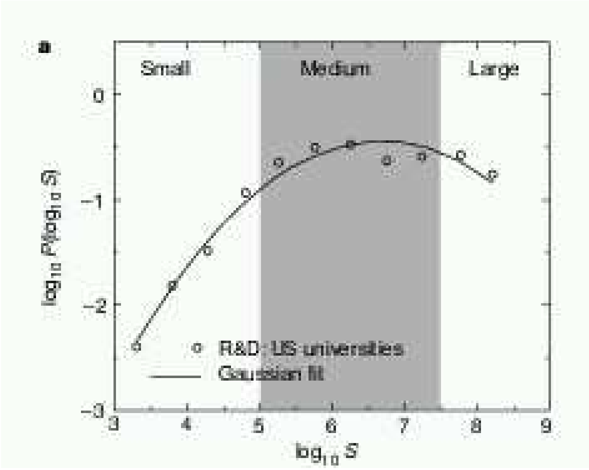


Figure 4.5: Histogram of the logarithm of the annual R&D expenditures of 719 US universities, expressed in 1992 US dollars. Here, S denotes the R&D expenditures deflated by the Consumer Price Index so that values for different years are comparable. The data for individual universities are the average over the entire period, 1979–1995, and the continuous line is a Gaussian fit to the data. The bins were chosen equally spaced on a logarithmic scale with bin size 0.5. The form of the size distribution is similar for different measures of academic performance such as the number of papers published and the number of patents filed each year. The functional form of the size distribution holds also for the external income of English universities and the grants for Canadian universities. (From Plerou et al. [156])

Plerou [156] *et al.* studied a different industry and analyzed the size distribution of universities based on the production of academic research and development (R&D) funding. The database includes: (i) a National Science Foundation database of the R&D expenditures for science and engineering of 719 United States (US) research universities for the 17-year period 1979–1995, and (ii) an Institute of Scientific Information database of the research publications of the top 112 US research universities for the 17 year period 1981–1997. Fig. 4.5 shows that university size defined by R&D expenditure is log-normal distributed. Other measures produced similar results.

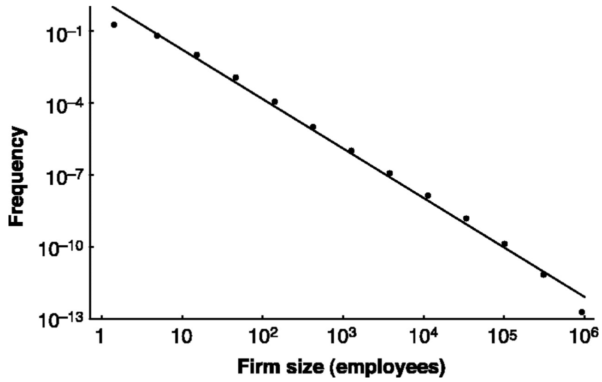


Figure 4.6: Histogram of US firm sizes, by employees. Data are for 1997 from the US Census Bureau, tabulated in bins having width increasing in powers of three. The solid line is the ordinary least squares (OLS) regression line through the data, and it has a slope of 2.059, meaning that probability density function of growth rate $P(S) \sim S^{-2.059}$. (From Axtell et al. [13])

In 2001, Axtell [13] tested U.S. Census Bureau data from 1997. Data from the U.S. Census Bureau put the total number of firms that had employees sometime during 1997 at about 5.5 million, including over 16,000 having more than 500 employees. Compared with Compustat database, the Census Bureau data has more records on smaller firms whose employee number is less than 500. In Fig. 4.6, the distribution of firm size (measured by the number of employees) is well fit by a power law.

How about an even lower level, products, since companies are composed of products? The Pharmaceutical Industry Database (PHID) analyzed by Matia *et al.* [142] records quarterly sales figures of 55624 pharmaceutical products commercialized by 3939 firms in the European Union and North America from September 1991 to June 2001. They find that the size distribution of products also approximately follows a log-normal function (Fig. 4.7).

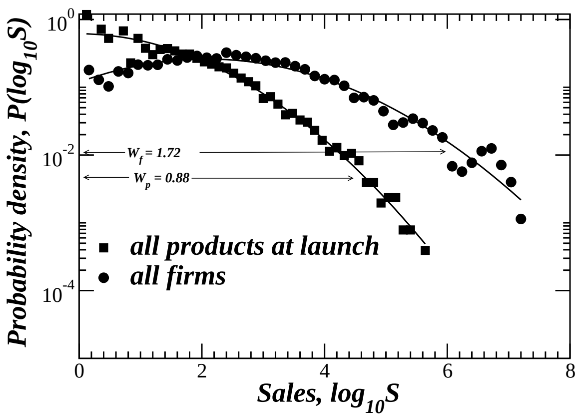


Figure 4.7: The size distribution of products sales (squares) and firm sales (circles) in the pharmaceutical industry during 1990-2001. The variance of the PDF's of products sales at launch and firm sales are estimated to be $W_p = 0.88$ and $W_f = 1.72$ respectively. The PDF of sales is observed to be log-normal. (From Matia et al. [142])

4.0.5.4 Income

Aoyama *et al.* [7] has studied the data on income and income-tax of individuals in Japan for the fiscal year 1998. The income data contains all $6,224,254$ workers who filed tax returns and among them, there are $84,515$ individuals who paid income tax because their incomes are 10 million yen or more in 1998.

Figure. 4.8 shows the rank-size plot, which is the log-log plot of the rank (R) as a function of income (S). Because $P(S) = R(S)/R(0)$, and $R(0)$ is a constant, y axis represents the distribution of income. It is found that a power law, $R \propto S^{-2.06}$, fits the data over three magnitudes of income, $10 < S < 10^4$.

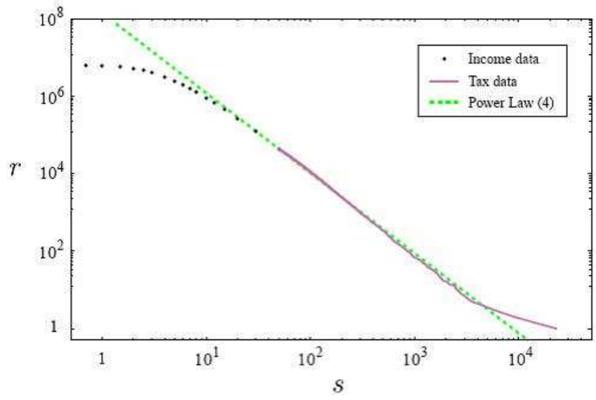


Figure 4.8: The rank-size plot of the income. The raw income data is shown by dots, whereas deduced income data from the income-tax database are connected by the solid line-segments. The dash line is given by Eq. 4.2 with $k = 2.06$. (From Aoyama et al. [7])

This result that diverse organizations such as countries, companies, universities and products follow log-normal or power-law distributions is much more remarkable than modern economists generally acknowledge. Perhaps the reason it has not previously generated much interest is that systematic differences in firm sizes across markets are related to industry characteristics such as economics of scale and the opportunities to advertise. If the distribution of firm sizes within an industry are driven by such factors, then the over-all distribution of firm size would seem to be determined by the composition of output, time dependent and diverse factors. As a consequence, there was no reason to believe that this distribution would remain stable over time or that it would necessarily be subject to modeling.

4.0.6 The Distribution of Growth Rate

Previous studies (described below) have found that g follows a distribution that looks like a tent, hence it is called the "tent-shaped" distribution. This distribution has a Laplace shape in the body, therefore, on a log-linear plot, the shape of $P(g)$ is two symmetric lines. Specifically this $P(g)$ can be written by:

$$P(g) = \frac{1}{\sqrt{2}\sigma_o} \exp\left(-\frac{\sqrt{2}\,|g-\bar{g}|}{\sigma_o}\right), \tag{4.4}$$

where σ_o is the standard deviation. The empirical conditional probability density of g given the size S therefore follows as

$$P(g|S) = \frac{1}{\sqrt{2}\sigma(S)} \exp\left(-\frac{\sqrt{2}\,|g-\bar{g}|}{\sigma(S)}\right), \tag{4.5}$$

where $\sigma(S)$ is the standard deviation for size S.

In the next subsection, the empirical findings on $P(g)$ and $P(g|S)$ of different subjects are showed.

4.0.7 Cou '

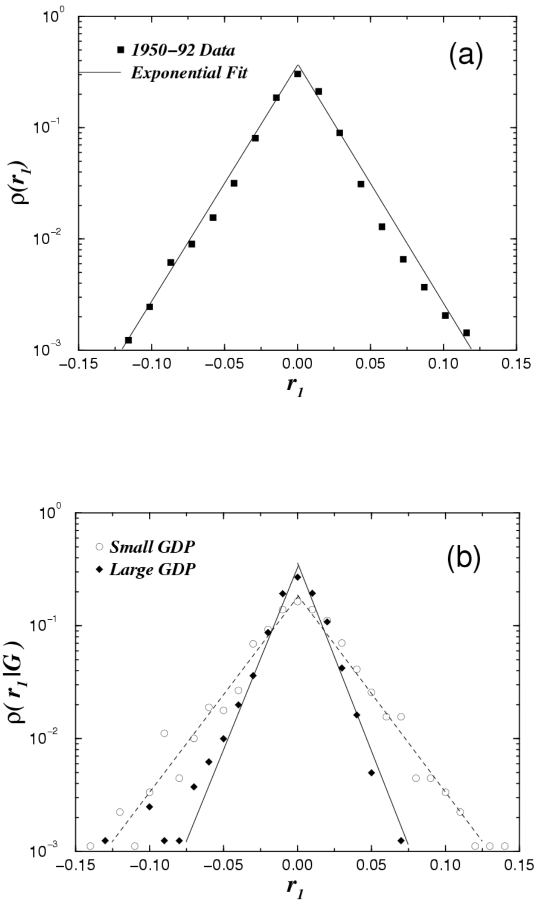


Figure 4.9: (a) PDF of annual growth rate r of GDP for 152 countries where r is the same as g here. Shown are the average annual growth rates for the entire period 1950–1992 together with an exponential fit, as indicated in Eq. (4.4). (b) PDF of annual growth rate for two subgroups with different ranges of S. The entire database was divided into 3 groups: $6.9 \times 10^7 \leq S < 2.4 \times 10^9$, $2.4 \times 10^9 \leq S < 2.2 \times 10^{10}$, and $2.2 \times 10^{10} \leq S < 7.6 \times 10^{11}$, and the figure shows the distributions for the smallest and largest groups. (From Lee et al. [120])

Lee *et al.* [120] analyzed the fluctuations in the growth rate of the gross domestic product (GDP) of 152 countries during the period 1950–1992. In Fig. 4.9, $P(g)$ and $P(g|S)$ can be fit by the Laplace distribution function (Eq. 4.4 and Eq. 4.5).

In the limit of small annual changes in S, $g(t)$ is the *relative* change in S. For all countries and all years, it was proposed that the probability density of g is consistent, for a certain range of $|g|$, with a double-exponential decay (see Fig. 4.9a), that is, Eq. 4.4 is a good option for the fitting function of $P(g)$.

The countries were divided into groups according to their GDP. It was found that the empirical *conditional* probability density of g for countries with approximately the same GDP is also consistent in a given range with the double-exponential form (see Fig. 4.9b) given by Eq. 4.5.

4.0.7.1 Company in Different Industries

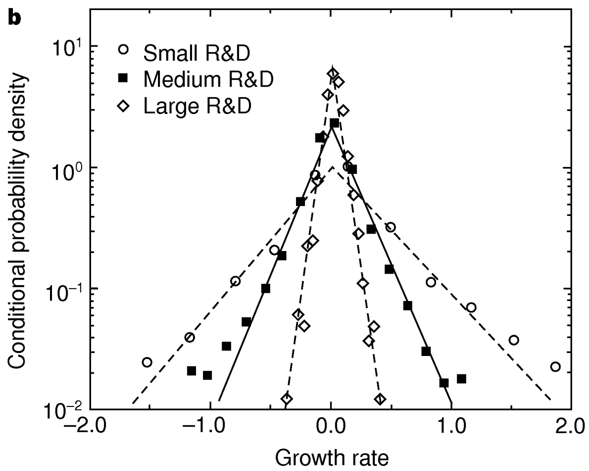


Figure 4.10: Conditional probability density function $P(g|S)$ of the annual growth rates g in R&D industry. For this plot the entire database is divided into three groups (depicted in a by different shades). (From Plerou et al. [156])

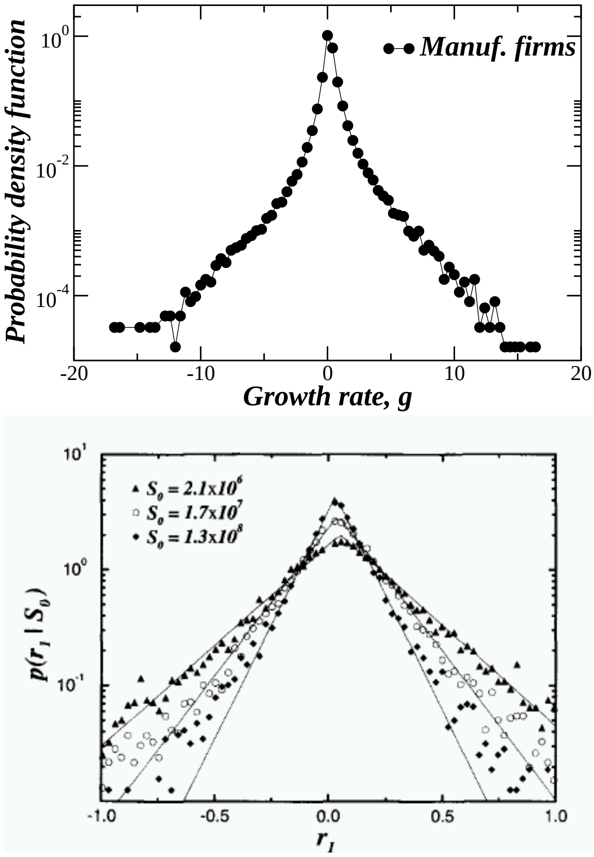
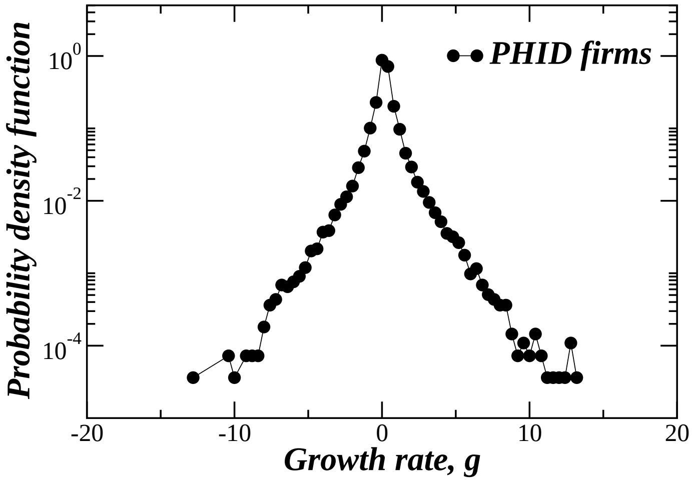


Figure 4.11: (a) Probability density $P(g)$ of the growth rate for all publicly-traded US manufacturing companies in the 1994 Compustat database with Standard Industrial Classification index of 2000-3999. The distribution represents all annual growth rates observed in the 19-year period 1974-1993.(b) Conditional probability density $P(g|S)$ of the growth rate. The data for three different bins of initial sales (with sizes increasing by powers of 8: $8^7 < S < 8^8, 8^8 < S < 8^9$, and $8^9 < S < 8^{10}$. The solid lines are exponential fits to the empirical data close to the peak. It is seen that the wings are somewhat "fatter" than what is predicted by an exponential dependence. (From Amaral et al. [3])



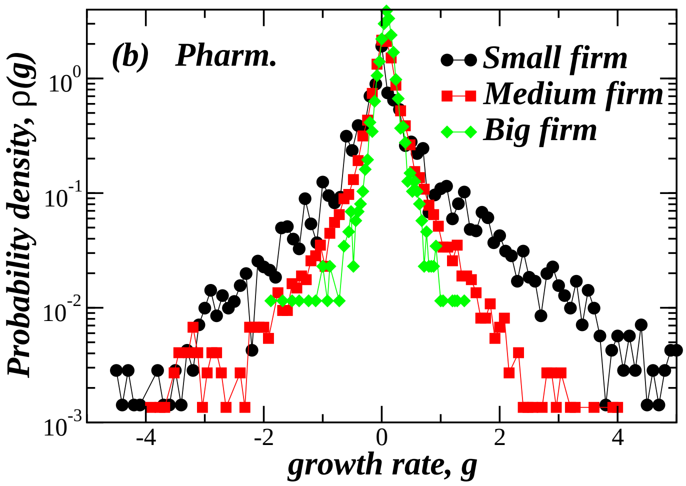


Figure 4.12: (a) Distribution of firm growth rate $P(g)$ in the pharmaceutical industry. Only the central part of the distribution can be fit by a Laplace function. (b) Conditional probability density function $P(g|S)$ of the annual growth rates g in PHID. Firm are divided into 3 categories: small $[S < 10^2]$, medium $[10^2 < S < 10^4]$, and large $[10^4 < S]$ firms. Again, each of them gives a Laplace central part and fat tails. (From Matia et al. [142])

For all the US manufacturing companies studied by Amaral *et al.* [3], Fig. 4.11 shows that $P(g|S)$ can be fit by Laplace distribution (see Eq. 4.5). Compared with manufacturing industry, the R&D industry also shows that the Laplace distribution for $P(g|S)$ (Fig. 4.10) is a good choice. Figure 4.11 shows that, for the pharmaceutical industry, (a) the central part of $P(g)$ is Laplace distributed, but the tails clearly deviate from it; (b) shows $P(g|S)$ have similar shapes as $P(g)$ and they also have fat tails.

4.0.8 Size-variance Relation

If the conditional distribution of growth rates has a functional form dependent on S, the standard deviation $\sigma(S)$—which is a measure of the width of $P(g|S)$—should be dependent on S. Thus, when the scaled quantities

$$\sigma(S)P(g/\sigma(S)|S) \qquad \text{versus} \qquad g/\sigma(S) \tag{4.6}$$

are plotted, all σ curves from the different size groups collapse onto a single curve. Then $p(g|S)$ follows a universal scaling (Amaral, 1997, Buldyrev, 1997)

$$P(g|S) \sim \frac{1}{\sigma(S)} \ f\left(\frac{g}{\sigma(S)}\right) . \tag{4.7}$$

Interestingly, studies by Matia *et al.* reveal that $\sigma(S)$ decays as a power law Stanley (1996), Buldyrev (1997)

$$\sigma(S) \sim S^{-\beta}, \tag{4.8}$$

where β is known as a *scaling exponent* for size-variance relation.

Various kinds of classes such as country, company, university and product *etc.*, are all composed of units. If it is assumed that the units are growing independently, according to the law of large numbers, the relation between their size and the variance of their corresponding growth rate should follow $S \sim g^{-\beta}$ where β is 0.5. But the empirical results give different value of β.

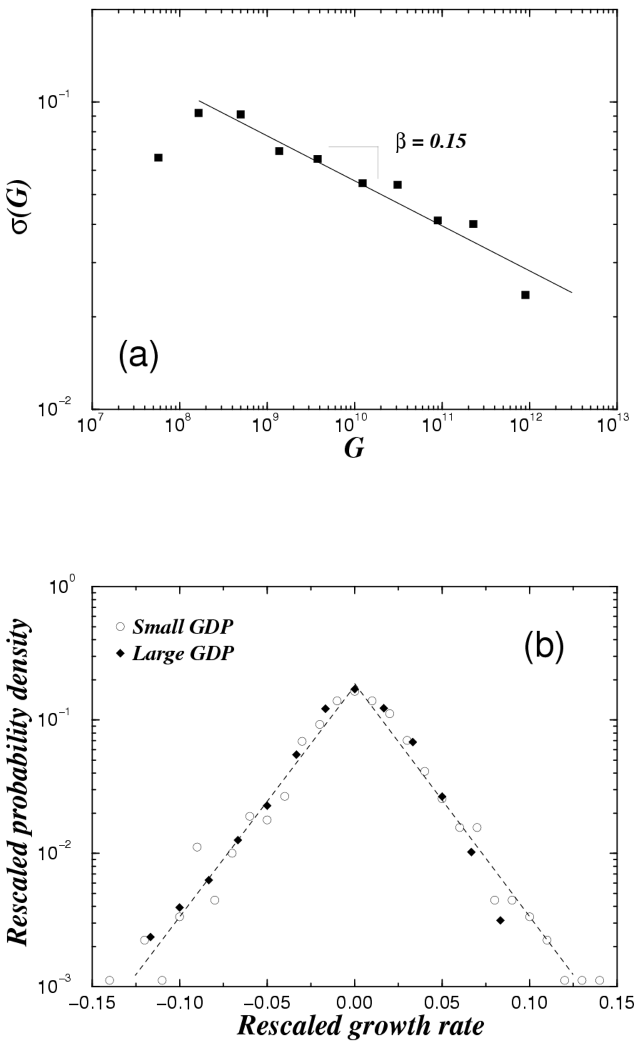


Figure 4.13: (a) Plot of the standard deviation $\sigma(G)$ of the distribution of annual growth rates as a function of S where G is the same as S here, together with a power law fit (obtained by a least square linear fit to the logarithm of σ vs the logarithm of S). The slope of the line gives the exponent β, with $\beta = 0.15$. (b) Rescaled probability density function, $\sigma(S)\rho(g|S)$, of the rescaled annual growth rate, $g/\sigma(S)$. Note that all data collapse onto a single curve. (From Lee et al. [120])

4.0.8.1 Country

Lee *et al.* [120] show that the width of the distribution scales as a power law of GDP with a scaling exponent $\beta \approx 0.15$. In the limit of small annual changes in S, $g(t)$ is the *relative* change in S. The countries were divided into groups according to their GDP. It was found that the empirical *conditional* probability density of g for countries with approximately the same GDP is also consistent in a given range with the double-exponential form given by Eq. 4.5. These data are shown in Figure 4.13.

4.0.8.2 Company in Different Industries

It is apparent from Fig. 4.14, the width of the distribution of growth rates in manufacturing industry decreases with increasing firm size S_0. It is found that $\sigma(S_0)$ is well approximated for eight orders of magnitude (from sales of less than 10^3 dollars up to sales of more than 10^{11} dollars) by the law $\sigma_1(S_0) \sim exp(-\beta S_0)$, where $\beta = 0.17 \pm 0.03$.

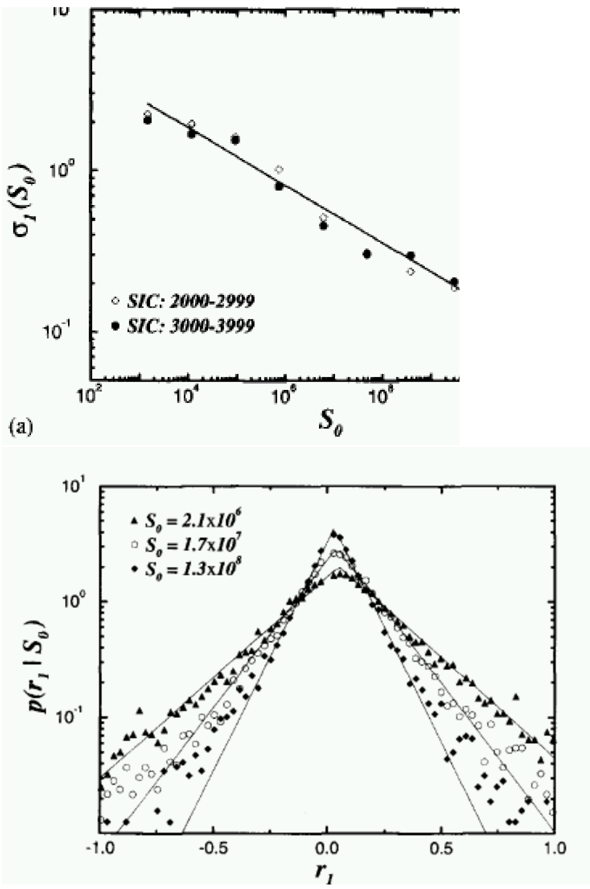


Figure 4.14: (a) Dependence of σ_1 on S_0 for two subsets of the data corresponding to different values of the SIC codes for manufacturing industry. In principle, companies in different subsets operate in different markets. The figure suggests that these results are universal across markets. (b) Scaled probability density $P_{\text{scal}}(r) \equiv \sqrt{2}\sigma(S_0)P(r|S_0)$ as a function of the scaled growth rate $r_{\text{scal}} \equiv \sqrt{2}[g - \bar{r}(S_0)]/\sigma(S_0)$ where r is the same as the g defined. The values were rescaled using the measured values of $\bar{g}(S0)$ and $\sigma(S0)$. All the data collapse upon the universal curve $P_{scal}(g) = exp(-|g_{scal}|)$ as predicted by Eqs. 4.5 and 4.8. Again, it is seen that all the data collapse onto a single curve. (From Amaral et al. [3])

The size-variance relation is also verified to exist by the databases of R&D and pharmaceutical industries in figures 4.15 and 4.16. The values of β are 0.25 for R&D industry and 0.20 for pharmaceutical industry.

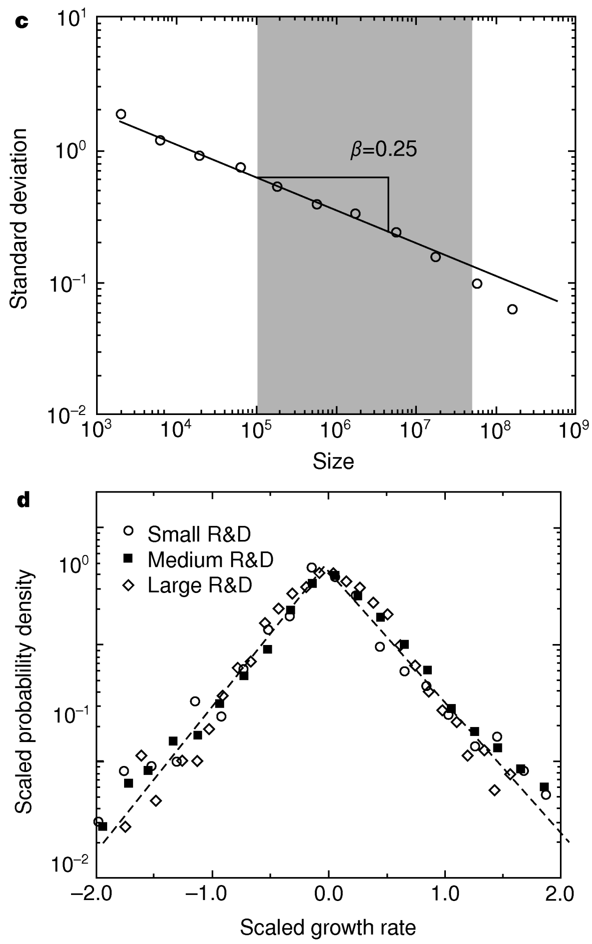
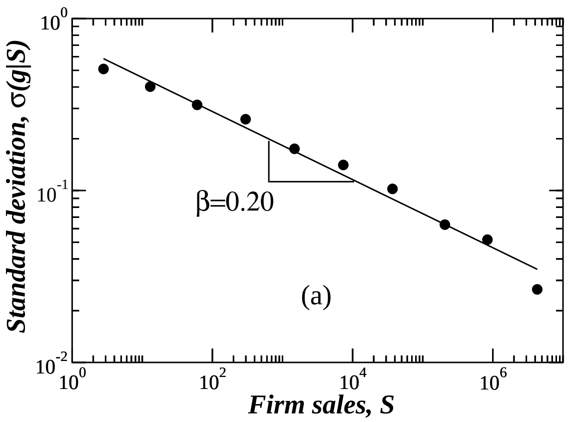


Figure 4.15: (c) The size-variance relation for universities in R&D industry. It shows a power-law shape with exponent 0.25. (d) Rescaled distribution of growth rates after applying the size-variance relationship. The overlap of $P(g|S)$ implies that different universities follow the same growth dynamics. (From Plerou et al. [156])



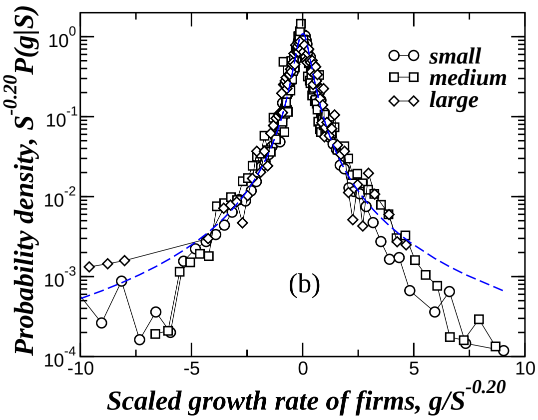


Figure 4.16: (a) Firms in the pharmaceutical industry are divided into 10 groups according to sales S. The standard deviation $\sigma(g|S)$ of the growth rates scales is a power law, $\sigma(g|S) \sim S^{-\beta}$ with $\beta = 0.20 \pm 0.01$. (b) PDF of the growth rates for small $[S < 10^2]$, medium $[10^2 < S < 10^4]$, and large $[10^4 < S]$ values of S is scaled by their standard deviation. Note the collapse of the histograms of the three groups which confirms the scaling exponent β. (From Matia et al. [142])

What is remarkable about Eqs. 4.5 and 4.8 is that they approximate the growth rates of a diverse set of companies. They differ not only in their size but also in what they manufacture. The conventional economic theory of the company is based on production technology, which varies from product to product. Conventional theory does not suggest that the processes governing the growth rate of car companies should be the same as those governing, e.g., pharmaceutical or paper companies. Indeed, these findings are reminiscent of the concept of universality found in statistical physics, where different systems can be characterized by the same fundamental laws, independent of "microscopic" details. Thus, it can be posed the question of the universality of these results: Is the measured value of the exponent β due to some averaging over the different industries, or is it due to a universal behavior valid across all industries? As a "robustness check", Amaral *et al.* split the entire company sample into two distinct intervals of SIC codes. It is visually apparent in Fig. 4.14 that the same behavior holds for the different industries. Thus, it can be concluded that these results are indeed universal across different manufacturing industries in the US. In statistical physics, scaling phenomena are sometimes represented graphically by plotting a suitably "scaled" dependent variable as a function of a suitably "scaled" independent variable. If scaling holds, then the data for a wide range of parameter values are said to "collapse" upon a single curve. From the following empirical results, the data collapse upon the single straight line shows small but consistent deviations for large growth rates from the exponential distribution in Eq. (4.5). Thus, Eq. (4.5) can be regarded only as a first-order approximation to reality.

4.1 Previous Models

All of the empirical results presented above not only provide a justification for developing a model to explain the statistical properties of firm growth, but also have implications for what statistical properties are most important to explain. In 1931, Gibrat presented striking evidence [81, 82] that the distribution of firm size at different times and within different "populations" was approximately log-normal; and showed the log-normality could be generated by a process in which the distribution of growth rates is independent of initial size. Gibrat's stochastic model is an early example of the use of statistical physics in economics. His model is one of a random stochastic process, and the objective of his model was to explain the shape of a distribution that emerged from it.

The assumptions Gibrat proposed are : (1) the growth rate R of a company is independent of its size (this assumption is usually referred to by economists as the *law of proportionate effect*), (2) the successive growth rates of a company are uncorrelated in time, and (3) the companies do not interact.

In mathematics, Gibrat's model is expressed by the stochastic process:

$$S_{t+\Delta t} = S_t(1 + \epsilon_t), \tag{4.9}$$

where $S_{t+\Delta t}$ and S_t are, respectively, the size of the company at times $(t + \Delta t)$ and t, and ϵ_t is an uncorrelated random number with some bounded distribution and variance much smaller than one (usually assumed to be Gaussian). Hence,

$$S_t = S_0(1 + \epsilon_1)(1 + \epsilon_2) \cdots (1 + \epsilon_t). \tag{4.10}$$

If it is assumed that all companies are born at approximately the same time and have approximately the same initial size, then the distribution of company sizes is also log-normal. This prediction from the Gibrat model is approximately correct.

However, under Gibrat's assumptions, it can be easily derived the expression of growth rate g as follows: at time T (T is much larger than t, for example a year)

$$\log S(t = T) = \log S(t = 0) + \sum_{t'=1}^{M} \log(\eta_{t'}), \tag{4.11}$$

where M is the total number of time steps. Because of $g(T) = \log(S(T)/S(0))$, we get

$$g = \sum_{t'=1}^{M} \log(\eta_{t'}). \tag{4.12}$$

Since M is large, and η_t is a random variable following a certain distribution, using the Central Limit Theorem the distribution of g is Gaussian. Now we know that this prediction is wrong and this model needs some changes.

Buldyrev [33] et al. (1997) present a model of organizational hierarchy in which decisions get passed down through successive layers. At each decision point in the hierarchy, a manager can either accept the decision of his immediate superior or reject it and make his own independent decision. Amaral [4] et al. (1998) describe a model of Gibrat-type growth processes at the business unit level and firms consisting of multiple units with uncorrelated growth processes. This model predicts the key empirical findings about firm growth for a wide variety of parameters. The Buldyrev and Amaral et al. models qualitatively justify why the distribution of firm growth rates shows a "tent" shape, and also numerically give the size-variance relation.

Sutton [194] postulates that all partitions of a company of size S into smaller subpieces are equiprobable . This is similar to the corresponding hypothesis in statistical physics that all microstates which a physical system can attain are equiprobable. More precisely Sutton assumes that S is a large integer, and uses known mathematical results on the number of partitions to compute $\sigma(S)$. Finally, Sutton analytically gives $\beta \approx 0.24$ as a universal power-law exponent. Following Sutton, Bouchaud et al. (2002) present a model in which firms consist of independent, divisions that are varying groupings of a basic unit size [206]. He assumes that all partitions of the firm are equally likely. For example, a firm of size 4 could consist of four one-unit divisions, two one-unit divisions and one two-unit division, a two two-unit division, one three-unit and one one-unit division, or one four-unit division. This model is similar to the Amaral et. al. model in that it views firms as consisting of units whose growth rates are independent of each other. But, Sutton model discusses little about the distribution of firm growth rates, and its relationship with firm size.

Axtell (1999) presents a model in which firms are teams of individuals who select effort levels and share the output [12]. Holding effort levels constant, adding individuals (or collections of individuals) to a firm increases output. As firms become larger, however, each individual has more of an incentive to shirk because of the sharing rule. The growth dynamics come from having individuals randomly joining and leaving firms. This model implies a Pareto distribution of firm size and the scaling relationships in growth dynamics described above.

Given that there are already four models that predict the same empirical findings, one might question the need for a fifth. However, there is no reason to believe that these findings sufficiently identify the "true" model of the firm. Models that are not consistent with empirical findings can be rejected, but a model just because it is consistent with them cannot be accepted. The primary feature that distinguishes the preferential attachment is that it is explicitly designed to address the question of which activities are organized within a given firm.

4.2 The distribution of unit number, $P(K)$

4.2.1 Introduction

The preferential attachment model originally comes from an explanation for the distribution of unit number $P(K)$. By analyzing the pharmaceutical dataset PHID, $P(K)$ exhibits a power-law shape for small K and exponential tails for large K. It is argued that the exponential shape is not due to a finite-size effect as others had previously assumed, but rather is the characteristic of the growth of a complex system. PHID is a very good dataset which records the values of K and gives possible empirical controls for different boundary conditions of the model, such that, whether the model can give a good explanation can be tested. In the model, it is used "class" to denote the upper-level subject and "unit" for the lower-level subject that the class contains.

Many complex systems of interest to physicists, biologists and economists [16, 35, 91, 106, 210] share two basic similarities in their dynamics: (i) The system does not have a steady state but is growing. (ii) Basic units are born, they agglomerate to form classes, and classes grow in size according to a rule of proportional growth [81, 82]. In biological systems, units could be bacteria, and classes would be bacterial colonies. In the context of economic systems, units could be products, and classes would be firms; in social systems units could be human beings, and classes would be cities.

The probability distribution function $P(K)$ of the class size K of the systems mentioned above has been shown to follow a universal scale-free behavior $P(K) \sim K^{-\tau}$ with $\tau \approx 2$ [16, 106, 115, 210]. Other possible values of τ are discussed and reported in [150]. Also, for most of the systems $P(K)$ has an exponential cut-off, which is often assumed to be a finite size effect of the databases analyzed. Several models [35, 41, 74, 79, 101, 168] explain $\tau \approx 2$ but none explains the exponential cut-off of $P(K)$. Moreover, the models describing $P(K) \sim K^{-\tau}$ are not suitable to describe simultaneously systems or ranges of K for which $P(K) \sim \exp(-\gamma K)$.

4.2.2 The model

In this section, a model with a simple set of rules to describe $P(K)$ for the entire range of K, *i.e.*, a power law with an exponential cut-off is presented. The exponential cut-off of the power law is not due to finite size, but is an effect of the finite time interval of the evolution. How the functional form of $P(K)$ is determined by different scenarios of the model, changing from a pure exponential to a pure power law (with $\tau \approx 2$), via a power law with an exponential cut-off is showed. The predictions of the model are then tested through the analysis of a unique industrial database [70, 142], which covers both elementary units (products) and classes (markets, firms) in a given industry (pharmaceutical).

The model in this section consists of the following rules:

1. At time $t = 0$ there exist N classes, each with a single unit [1].

2. At each step:

   - (a) With probability b $(0 \le b \le 1)$ a new class with a single unit is born.
   - (b) With probability λ $(0 < \lambda \le 1)$ a randomly selected class grows one unit in size. The selection of the class that grows is made with probability proportional to the number of units it already has ["preferential attachment"].
   - (c) With probability μ $(0 < \mu < \lambda)$ a randomly selected class shrinks one unit in size. The selection of the class that shrinks is done with probability proportional to the number of units it already has ["preferential detachment"].

After M steps one expects $N + bM$ classes and $N + (\lambda - \mu + b)M$ units, since each class starts with one unit. Note that, for now, it is not included a Gibrat process to control the growth of each unit because the distribution of unit number $P(K)$ is our focus and a Gibrat formula has no effect on $P(K)$. In the continuum limit, the above rules give rise to a master equation for the class size PDF $P(K, t_i, t)$, which is the probability at time t, for a class i introduced at time t_i, to have K units,

$$\frac{\partial P(K,t_i,t)}{\partial t} = \lambda \frac{(K-1)}{n(t)} P(K-1,t_i,t) + \mu \frac{(K+1)}{n(t)} P(K+1,t_i,t) - (\lambda+\mu) \frac{K}{n(t)} P(K,t_i,t), \tag{4.13}$$

where $n(t) \equiv N + (\lambda - \mu + b)t$ is the expected total number of units at simulation step t. Each class has its own master equation. Eq. (4.13) is transformed to the master equation of birth and death processes [53] by a new variable s, where $dt/ds = n(t)$ and $\bar{P}(K, s_i, s) \equiv P(K, t_i(s_i), t(s))$. The master equation for $\bar{P}(K, s_i, s)$ has the same form as Eq. (4.13) after replacing t by s, $P(K, t_i, t)$ by $\bar{P}(K, s_i, s)$ and $n(t)$ by 1 respectively. From the well-known solution of birth and death processes under the initial condition $\bar{P}(1, s_i, s_i) = 1$ [2], the solution after transforming back from s to t is,

$$P(K,t_i,t) = \begin{cases} \frac{\mu}{\lambda}\eta_{t_i,t} & [K=0] \\ (1-\eta_{t_i,t})(1-\frac{\mu}{\lambda}\eta_{t_i,t})\eta_{t_i,t}^{K-1} & [K>0] \end{cases} \tag{4.14}$$

with

$$\eta_{t_i,t} = \frac{1 - R^{\alpha}}{1 - \frac{\mu}{\lambda} R^{\alpha}}, \tag{4.15}$$

where $R = [t_i + N(\lambda - \mu + b)^{-1}]/[t + N(\lambda - \mu + b)^{-1}]$ and $\alpha = (\lambda - \mu)(\lambda - \mu + b)^{-1}$. $P(K, t_i, t) \propto \eta_{t_i,t}^{K-1}$ is obviously an exponential function of K. Finally, one can obtain $P(K, t)$, by averaging $P(K, t_i, t)$ over units introduced at different t_i as follows:

$$P(K,t) = \frac{N}{N + bt} P(K, 0, t) \quad + \quad \frac{b}{N + bt} \int_0^t dt_i \, P(K, t_i, t). \tag{4.16}$$

The first term, I_1, is

$$I_1 \propto \exp(-\gamma K) \qquad \text{with} \quad \gamma = -\log \eta_{0,t} \sim t^{-\alpha}. \tag{4.17}$$

Note that this represents the contribution from the original firms. To obtain the second term I_2 we first substitute $P(K, t_i, t)$ from Eq. (4.14) in Eq. (4.16), then change the variable of integration from t_i to η. Hence

$$I_2 = \frac{b(t + \frac{N}{\lambda - \mu + b})(1 + \frac{b}{\lambda - \mu})(1 - \frac{\mu}{\lambda})}{(N + bt)} \int_0^{\eta_{0,t}} d\eta \left(\frac{1 - \eta}{1 - \frac{\mu}{\lambda}\eta} \right)^{1 + \frac{b}{\lambda - \mu}} \eta^{K-1}. \tag{4.18}$$

In the limit of $t \to \infty$, $\eta_{0,t} \to 1$. Since $(1 - \mu\eta/\lambda)^{-1} \approx 1 + \mu\eta/\lambda$, Eq. (4.18) can be integrated giving the Yule distribution [184]

$$I_2 = (1 + \frac{b}{\lambda - \mu})(1 - \frac{\mu}{\lambda}) \sum_{m=0}^{\infty} \frac{(\frac{\mu}{\lambda})^m}{m!} \frac{(\frac{b}{\lambda - \mu} + m)!}{(\frac{b}{\lambda - \mu})!} \int_0^1 d\eta \, (1 - \eta)^{1 + \frac{b}{\lambda - \mu}} \eta^{m + K - 1}. \tag{4.19}$$

In the limit of $t \to \infty$, from Eq. (4.19)

$$I_2 \propto K^{-(2 + \frac{b}{\lambda - \mu})}, \tag{4.20}$$

in which an exponential function has been transformed into a power law function by integration. This situation is analogous to the one described by the standard preferential attachment model [106], where the power law distribution also follows from the Yule distribution. In the limit of fixed time t and $K \to \infty$, $I_2 \propto \exp(-\gamma K)/K$ which decays faster than Eq. (4.17), implying that the distribution of class size K for new classes has an exponential cut-off faster than for the old classes. Thus, the full solution of Eq. (4.13) is a power law (Eq. (4.20)) with an exponential cut-off (Eq. (4.17)). These two terms are of the same order in the range $K \geq t^{\alpha}$ for large finite t.

A mean-field interpretation of the result $\tau \approx 2$ is next presented. At any time t_0, the number of units in the already-existing classes is $n(t_0)$. Suppose a new class consisting of one unit is born at time t_0. According to Rules 2b and 2c, its growth rate is proportional to $1/n(t_0)$. Neglecting the effect of the influx of new classes on $n(t_0)$, the average size K of this class born at t_0 is proportional to $1/n(t_0)$. So the classes which were born at times $t > t_0$ tend to have an average size measured in terms of K which is smaller than the one of older classes. If the classes are sorted according to their size, the rank $R(K)$ of a class is proportional to its age $R(K) \propto t_0$. Thus, $K \sim 1/n(t_0) \sim 1/t_0 \sim 1/R(t_0)$, coherently with the standard formulation of the Zipf law [210] according to which the size of a class K is inversely proportional to its rank. If it is taken into account the decrease of the growth rate with the influx of new classes, one can show after some algebra that $K \sim R^{-(\lambda-\mu)/(\lambda-\mu+b)}$, which includes $K \sim R^{-1}$ as a limiting case for $b \to 0$. Since $R(K)$ is the number of classes whose sizes are larger than K, it can be written in the continuum limit $R(K) \sim \int_K^\infty P(K)dk$, and hence $P(K) \sim K^{-2-b/(\lambda-\mu)}$.

The full solution of Eq. (4.13), a power law with an exponential cut-off, can be interpreted as follows. Starting with N classes which are colored red, and let the newly born classes be colored blue. Due to the preferential attachment rule, the red classes have on average a number of units which is larger than the blue classes. Thus for large K, $P(K)$ is governed by the exponential distribution of the red classes (*Case i*), while for small K, $P(K)$ is governed by the power law distribution of the blue classes (*Case ii*).

4.2.3 Comparison of model with data

The predictions of the model has been tested using the pharmaceutical industry database (PHID), a micro-level economic database which allows a fine grained decomposition of the statistical properties of growth dynamics of business firms in a given industry. PHID records quarterly sales figures of 48,819 pharmaceutical products commercialized in the European Union and North America from September 1991 to June 2001. The products are then classified into different hierarchical levels based on the Anatomic and Therapeutic Classification (ATC) (Table 4.1). Each level has a specific number of classes (Table 4.2).

At all different levels, there are positive correlations between the number of units (products) which enter or exit and the number of units in the classes of a given level $L = A, B, C, D$ (Table 4.3). This empirical observation is consistent with a preferential birth and death process (Rules 2b and 2c), as described.

Level	Type	N	Code	Content
A	Anatomical main group	13	M	Musculo-Skeletal System
B	Therapeutic subgroup	84	M05	Drugs for treatment of bone diseases
C	Pharmacological subgroup	259	M05B	Drugs affecting bone structure and mineralization
D	Chemical subgroup	432	M05BA	Bisphosphonates

Table 4.1: The ATC hierarchical classification. The ATC categorizes drugs at four levels of aggregation according to the organ or system on which they act and their chemical, pharmacological and therapeutic properties. There are 13 main groups (level A) and 84 pharmacological subgroups (level B). The levels C and D are pharmacological/therapeutic subgroups. Medicinal products, such as Bisphosphonates in the example, are classified according to the main therapeutic use of the main active ingredient. The basic principle is one ATC code for each pharmaceutical formulation. The WHO is responsible to manage the ATC. Over the period of our empirical analysis, the number of classes of levels A and B have remained constant, while the number of classes in levels C and D increased by 3% and 5% respectively.

Level	A	B	C	D	firms	products
N_f	13	84	259	432	3913	48819
N_b	0	0	8	20	458	12645
N_d	0	0	0	0	252	3361

Table 4.2: The evolution of the number of classes N for different levels of the PHID over 10 years. There are three different cases of N in 6 levels: (i) For levels A and B there is no birth or death of classes (*i.e.*, the number of newly born classes N_b is 0 and the number of dead classes N_d is also 0. (ii) For levels C and D system grows not only with birth and death of units inside classes but also with the birth of classes. The system grows with the birth of new classes to the final N_f classes (259 for level C and 432 for level D). (iii) For firm level, there also exists the death of classes which is not considered in the model. From the table, the values of $b/(\lambda - \mu)$ estimated to be $N_{b,L}/(N_{b,p} - N_{d,p})$ are 0.0009 (level C), 0.002 (level D) and 0.049 (firms).

Level	A	B	C	D	firms
$C(k_b, k_e)$	0.93	0.87	0.84	0.82	0.70
$C(k_d, k_e)$	0.88	0.86	0.80	0.78	0.75

Table 4.3: Correlation coefficient $C(K_b, K_e)$ between the number of born units K_b and existing number of units K_e in classes and $C(K_d, K_e)$ between the number of dead units K_d and K_e in classes for each level in the PHID. The observed correlations justify the assumptions in the model: the preferential birth and death of units (Rules 2b and 2c).

(Level)	γ (A)	γ (B)	γ (C)	γ (D)	γ (firms)	τ (firms)
data	0.00031	0.0015	0.0039	0.0044	0.0054	1.97
model	0.00020	0.0013	0.0033	0.0050	0.0173	2.05

Table 4.4: Comparison of values of the parameters in the model using the data and from the model. γ of the data is estimated by regression in Fig. 4.17, and γ of the model is estimated using $\gamma = -\log \eta_{0,t}$ ($\eta_{0,t}$ in Eq. (4.15) is estimated by the value of $b/(\lambda - \mu)$ and N which is the solution of two equations: $N + (\lambda - \mu + b)t = 48819$ and $N + bt = N_{f,L}$, based on Table 4.2). τ of the data is estimated by regression in Fig. 4.18a and τ of the model is estimated using $\tau = 2 + b/(\lambda - \mu)$ with the numbers of Table 4.2.

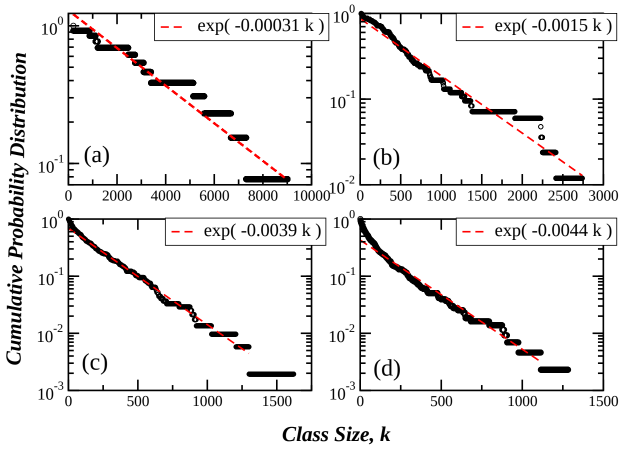


Figure 4.17: Empirical results for the cumulative probability distribution, $P(K)$, of class size K at different levels. Figures (a)-(d) correspond to levels A-D respectively. Symbols represent data points in each level (a)$\sim$(d), while solid lines are predictions of the model. The cumulative probability distributions for all levels are reasonably well fit by pure exponentials, as predicted by the model.

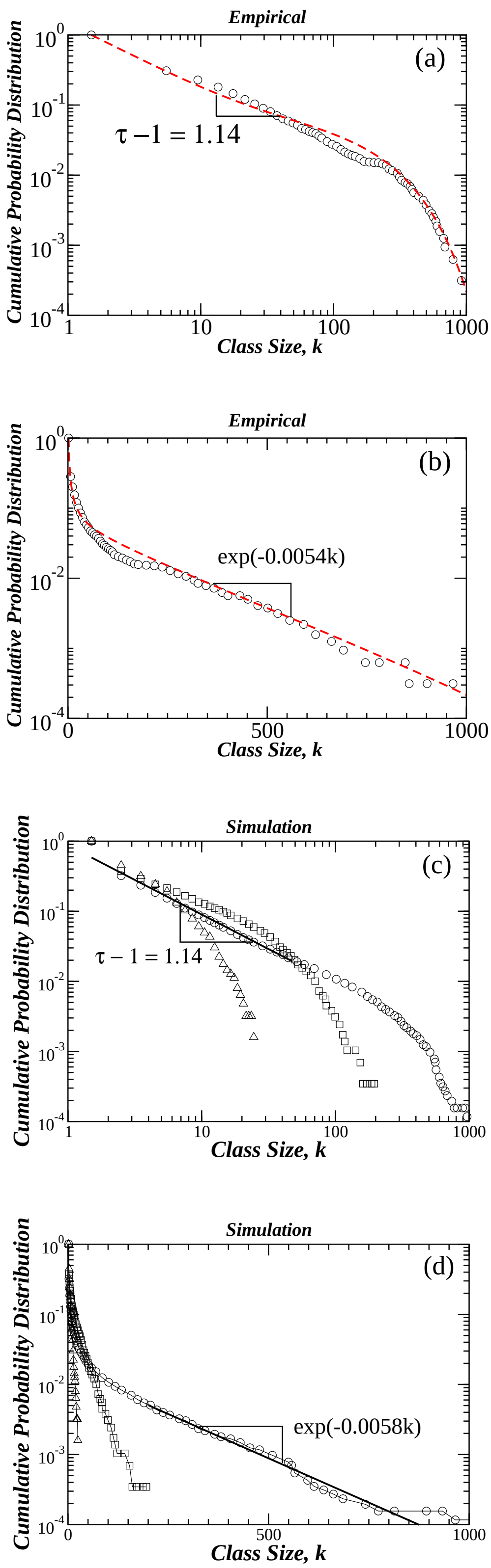


Figure 4.18: Comparison of empirical results for firms of the PHID and the simulation results. (a) and (c) Log-log plots of the cumulative probability distribution' of the class sizes show a power law decay $K^{-(\tau-1)}$ with $\tau \approx 2$ for $K < 200$. (b) and (d) Log-linear plots of the cumulative probability distribution, showing exponential decay for $K > 200$. In (c) and (d) $\bigcirc$, $\square$ and $\triangle$ show the distribution for $t = 200,000$, $t = 20,000$ and $t = 2000$ respectively. Note that the exponential function gradually changes into a power law function.

For levels A and B, the number of classes did not change during the period of observation and it is found an exponential distribution (Figs. 4.17a and 4.17b) as predicted by the limiting *Case i* of the model. For levels C and D, a weak departure from the exponential functional form (Fig. 4.17c and Fig. 4.17d) can be accounted for within the model, if it is considered a slight growth in the number of classes.

The full solution predicted by the model, *i.e.*, power law followed by the exponential decay of $P(K)$, is observed empirically for firms (Fig. 4.18), displaying a power law with exponent $\tau = 1.97$ for $K < 200$, and an exponential cut-off for $K > 200$. Coherently with the predictions of the model, the exponential part of $P(K)$ arises from large, diversified, "old" firms, while the power law part of $P(K)$ is produced by young firms. The reason for the departure of the empirically observed τ from the prediction of the model $\tau > 2$ comes from the fact that the distribution is the sum of the distributions of the new and old classes. The latter create a "bump" on the power law distribution caused by the new classes in the region of $K \geq t^{\alpha}$. We have tested numerically that the value of τ measured for the distribution of the new classes is consistent with the analytical predictions $\tau > 2$ while the effective value of $\tau < 2$ measured for the entire distribution is in agreement with the empirical data.

The estimated parameters are given based on Table 4.2: $b/(\lambda - \mu)$ is estimated to be $N_{b,L}/(N_{b,p} - N_{d,p})$, $\lambda/\mu = N_{b,p}/N_{d,p}$, $N + (\lambda - \mu + b)t = 48819$ and $N + bt = N_{f,L}$ (where the subscripts 'p', 'b', 'd' and 'f' denote 'product', 'birth', 'death' and 'final' respectively, and 'L' means either of level A to D or firms). Using $\gamma = -\log \eta_{0,t}$ and $\eta_{0,t}$ in Eq. (4.15) by eliminating t, γ and τ can be estimated (Table 4.4).

The oldest firms within the industry entered it almost 150 years ago, while this data cover only the last decade. Nonetheless, the theoretical estimations of γ and τ based on Eq. 4.17 and Eq. 4.20 are surprisingly good, except for the γ of firms. This departure can be accounted for if it is considered that the real data on firms are shaped not only by *firm entry*, but also by *firm exit*, *mergers & acquisitions*, which are not considered by the model [101, 169]. Additional computer simulations show, that if a possible exit of classes is included in the model, the value of γ estimated from the parameters of the model comes to an agreement with the actual one. Simulation results are showed in Fig. 4.18c and Fig. 4.18d, and they are in good agreement with the empirical results in Fig. 4.18a and Fig. 4.18b.

4.3 The Distribution of Growth Rates, $P(g)$, and its Modeling

4.3.1 Introduction

The growth-rate distributions $P(g)$ exhibit many similar behaviors among different subjects and fields but without a general model towards the justification. In particular, growth rate interests economists and entrepreneurs far more than other quantities because growth rate means the change of firm size which relates to the profit of a firm, and positive growth rate means that a given company can live in the market.

The first prediction of the Gibrat model, log-normal distribution of firm size, was supported by many subsequent researchers [3, 34, 120, 142, 187, 207], but the second one has been challenged by many researchers [27, 120, 142, 156, 187] who found that the firm growth distribution is not Gaussian but displays a tent shape. Here a mathematical framework that provides an unifying explanation for the growth of business firms based on the number and size distribution of their elementary constituent components [3, 4, 33, 34, 38, 71, 195, 196] is introduced. Specifically a model of proportional growth in both the number of units and their size is presented and some general implications on the mechanisms which sustain business firm growth [71, 92, 100, 109, 135, 194] is drawn. According to the model, the probability density function (PDF) of growth rates, $P(g)$ is Laplace [113] in the center [187] with power law tails [166, 168] decaying as $g^{-\zeta}$ where $\zeta = 3$.

Two key sets of assumptions in the model are described in Sec. 4.3.2.1 (the number of units K in a class grows in proportion to the existing number of units) and Sec. 4.3.2.2 (the size of each unit fluctuates in proportion to its size). Our objective is to first find $P(K)$ (by a different approach from last chapter), the probability distribution of the number of units in the classes at large t, and then find $P(g)$ using the convolution of $P(K)$ and the conditional distribution of the class growth rates $P(g|K)$, which for large K converges to a Gaussian.

4.3.2 The Model

4.3.2.1 Case 1: Mean-field solution of $P(K)$

In this subsection, a different approach to solve the distribution of unit number $P(K)$ by mean-field approximation, compared to the method that it is used in Section 4.2 by solving partial differential equations is given. The two different methods starting from the same assumptions show the same results for $P(K)$.

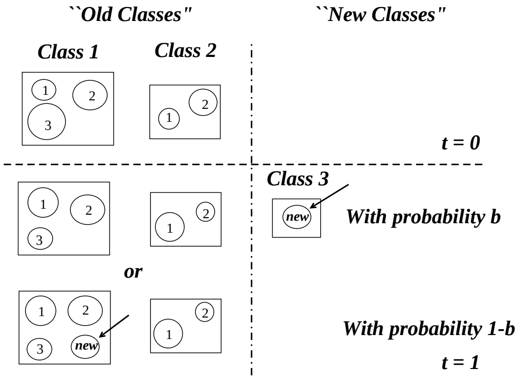


Figure 4.19: Schematic representation of the model of proportional growth. At time $t = 0$, there are $N(0) = 2$ classes (□) and $n(0) = 5$ units (◯) (Assumption A1). The area of each circle is proportional to the size ξ of the unit, and the size of each class is the sum of the areas of its constituent units (see Assumption B1). At the next time step, $t = 1$, a new unit is created (Assumption A2). With probability b the new unit is assigned to a new class (class 3 in this example) (Assumption A3). With probability $1 - b$ the new unit is assigned to an existing class with probability proportional to the number of units in the class (Assumption A4). In this example, a new unit is assigned to class 1 with probability $3/5$ or to class 2 with probability $2/5$. Finally, at each time step, each circle i grows or shrinks by a random factor η_i (Assumption B2).

The set of assumptions [207] described in the previous section can be simplified as

(A1) Each class α consists of $K_\alpha(t)$ number of units. At time $t = 0$, there are $N(0)$ classes consisting of $n(0)$ total number of units. The initial average number of units in a class is thus $n(0)/N(0)$.

(A2) At each time step a new unit is created. Thus the number of units at time t is $n(t) = n(0) + t$.

(A3) With birth probability b, this new unit is assigned to a new class, so that the average number of classes at time t is $N(t) = N(0) + bt$.

(A4) With probability $1 - b$, the new unit is assigned to an existing class α with probability $P_\alpha = (1 - b)K_\alpha(t)/n(t)$, so $K_\alpha(t + 1) = K_\alpha(t) + 1$.

This model can be generalized to the case when the units are born at any unit of time t' with probability μ, die with probability λ, and in addition a new class consisting of one unit can be created with probability b' [207]. This model can be reduced to the present model if one introduce time $t = t'(\mu - \lambda + b')$ and probability $b = b'/(\mu - \lambda + b')$. The simplified model is shown in Fig. 4.19.

In two limiting cases (i) $b = 0$, $K_\alpha = 1$ ($\alpha = 1, 2 \ldots N(0)$) and (ii) $b \neq 0$, $N(0) = 1$, $n(0) = 1$ this model has exact analytical solutions $P(K) = N(0)/t(t/(t+N(0)))^K(1+O(1/t))$ [107, 114] and $\lim_{t\to\infty} P(K) = (1+b)\Gamma(K)\Gamma(2+b)/\Gamma(K+2+b)$ [167] respectively. In general, an exact analytical solution of this problem cannot be presented in a simple closed form. Accordingly, it is sought for an approximate mean-field type [186] solution which can be expressed in simple integrals and even in elementary functions in some limiting cases. First it will be presented a known solution of the preferential attachment model in the absence of the influx of new classes [53]:

$$P_{\text{old}}(K) = \lambda^K \frac{1}{K(t) - 1} \approx \frac{1}{K(t)} \exp(-K/K(t))[1 + O(t^{-1})], \tag{4.21}$$

where $\lambda = 1 - 1/K(t)$, and $K(t) = [n(0) + t]/N(0)$ is the average number of units in the old classes at time t. Note that the form of the distribution of units in the old classes remains unchanged even in the presence of the new classes, whose creation does not change the preferential attachment mechanism of the old classes and affects only the functional form of $K(t)$.

Now the problem in the presence of the influx of the new classes will be treated. Assume that at the beginning there are $N(0)$ classes with $n(0)$ units. Because at every time step, one unit is added to the system and a new class is added with probability b, at moment t there are

$$n(t) = n(0) + t \tag{4.22}$$

units and approximately

$$N(t) = N(0) + bt \tag{4.23}$$

classes, among which there are approximately bt new classes with n_{new} units and $N(0)$ old classes with n_{old} units, such that

$$n_{old} + n_{new} = n(0) + t. \tag{4.24}$$

Because of the preferential attachment assumption (A4), it can be written, neglecting fluctuations [186] and assuming that t, n_{old}, and n_{new} are continuous variables:

$$\frac{dn_{new}}{dt} = b + (1 - b)\frac{n_{new}}{n(0) + t}, \tag{4.25}$$

$$\frac{dn_{old}}{dt} = (1 - b)\frac{n_{old}}{n(0) + t}. \tag{4.26}$$

Solving the second differential equation and taking into account the initial condition $n_{old}(0) = n(0)$, $n_{old}(t) = (n(0) + t)^{1-b}\, n(0)^b$. Analogously, the number of units at time t in the classes existing at time t_0 is

$$n_e(t_0, t) = (n(0) + t)^{1-b}(n(0) + t_0)^b \tag{4.27}$$

where the subscript 'e' means "existing". Accordingly, the average number of units in old classes is

$$K(t) = \frac{n_{old}(t)}{N(0)} = \frac{(n(0)+t)^{1-b}}{N(0)}\ n(0)^b. \tag{4.28}$$

Now using Eq. (4.21) to calculate the distribution of units in the old classes

$$P_{old}(K) \approx \frac{N(0)}{(n(0)+t)^{1-b}n(0)^b}\ \exp\left(-\frac{K\,N(0)}{(n(0)+t)^{1-b}n(0)^b}\right). \tag{4.29}$$

and the contribution of the old classes to the distribution of all classes is

$$\tilde{P}_{old}(K) = P_{old}(K)N(0)/(N(0)+bt). \tag{4.30}$$

Next, the same strategy is used: first to figure out the average number of units in new classes, and second to calculate the $P_{new}(K)$. In an infinitesimal time dt, the number of units in the classes that appear at t_0 is $b\,dt$ and the number of these classes is $b\,dt$. Because the probability that a class captures a new unit is proportional to the number of units it has already gotten at time t, the number of units in the classes that appear at time t_0, observed at time t, is

$$n_{new}(t_0,t) = n_e(t_0,t)bdt/[n(0)+t_0]. \tag{4.31}$$

The average number of units in these classes is

$$K(t_0,t) = n_{new}(t_0,t)/b\,dt = (n(0)+t)^{1-b}/(n(0)+t_0)^{1-b}. \tag{4.32}$$

Assuming that the distribution of units in these classes is given by a continuous approximation (4.21)

$$P_{new}(K,t_0) \approx \frac{1}{K(t_0,t)}\ \exp\left(-K/K(t_0,t)\right). \tag{4.33}$$

Thus, their contribution to the total distribution is

$$\frac{b\,dt_0}{N(0)+b\,t}\,\frac{1}{K(t_0,t)}\ \exp\left(-K/K(t_0,t)\right)$$

The contribution of all new classes to the distribution $P(K)$ is

$$\tilde{P}_{new}(K) \approx \frac{b}{N(0)+b\,t}\int_0^t \frac{1}{K(t_0,t)}\ \exp\left(-K/K(t_0,t)\right)\,dt_0. \tag{4.34}$$

If $y = K/K(t_0,t)$ then $\tilde{P}_{new}(K) = P_{new}(K)bt/(N(0)+bt)$ where

$$P_{new}(K) \quad \approx \quad \frac{n(0)/t+1}{1-b}\ K^{\left(-\frac{1}{1-b}-1\right)}\ \int_{K'}^{K} e^{-y}\ y^{\frac{1}{1-b}}\ dy. \tag{4.35}$$

and the low limit of integration, K' is given by

$$K' = K\left(\frac{n(0)}{n(0)+t}\right)^{1-b} \tag{4.36}$$

Finally the distribution of units in all classes is given by

$$P(K) = \frac{N(0)}{N(0)+bt} P_{old}(K) + \frac{bt}{N(0)+bt} P_{new}(K). \tag{4.37}$$

Now the asymptotic behavior of the distribution in Eq. (4.35) is investigated and it can be described by the Pareto power law tail with an exponential cut-off.

1. At fixed K when $t \to \infty$, $K' \to 0$, thus

$$\begin{aligned} P_{new}(K) &= \frac{1}{1-b} K^{-\frac{1}{1-b}-1} \int_0^K e^{-y} \, y^{\frac{1}{1-b}} \, dy, \\ &= \frac{1}{1-b} \left[\Gamma\left(1+\frac{1}{1-b}\right) - \int_K^{\infty} e^{-y} \, y^{\frac{1}{1-b}} \, dy \right] K^{-1-\frac{1}{1-b}}. \end{aligned} \tag{4.38}$$

As $K \to \infty$, $P_{new}(K)$ converges to a finite value:

$$P_{new}(K) = K^{-1-\frac{1}{b}} \left(\frac{1}{1-b}\right) \Gamma\left(1+\frac{1}{1-b}\right). \tag{4.39}$$

Thus for large $K \gg 1$, but such that $K' \ll 1$ or $K \ll (1+t/n(0)^{1-b}$, an approximate power-law behavior is :

$$P_{new}(K) \sim K^{-\varphi}, \tag{4.40}$$

where $\varphi = 2 + b/(1-b) \geq 2$.

As $K \to 0$,

$$P_{new}(K) = \frac{1}{1-b} K^{\left(-\frac{1}{1-b}-1\right)} \frac{K^{\left(1+\frac{1}{1-b}\right)}}{1+\frac{1}{1-b}} = \frac{1}{2-b}. \tag{4.41}$$

2. At fixed t when $K \to \infty$, using partial integration to evaluate the incomplete Γ function:

$$\int_x^{\infty} e^{-y} \, y^{\alpha} \, dy = -e^{-y} \, y^{\alpha}|_x^{\infty} + \alpha \int_x^{\infty} e^{-y} \, y^{\alpha-1} \, dy \approx e^{-x} \, x^{\alpha}.$$

Therefore, from Eq. (4.35)

$$\begin{aligned} \tilde{P}_{new}(K) &\approx \frac{n(0)+t}{N(0)+bt} \frac{b}{1-b} K^{-\frac{1}{1-b}-1} \int_{K\left(\frac{n(0)}{n(0)+t}\right)^{1-b}}^{\infty} e^{-y} \, y^{\frac{1}{1-b}} \, dy, \\ &= \frac{n(0)}{N(0)+bt} \frac{b}{1-b} \frac{1}{K} \exp\left(-K\left(\frac{n(0)}{n(0)+t}\right)^{1-b}\right), \end{aligned} \tag{4.42}$$

which always decays faster than Eq. (4.29) because $n(0) \geq N(0)$ and there is an additional factor K^{-1} in front of the exponential. Thus the behavior of the distribution of all classes is dominated for large K by the exponential decay of the distribution of units in the old classes.

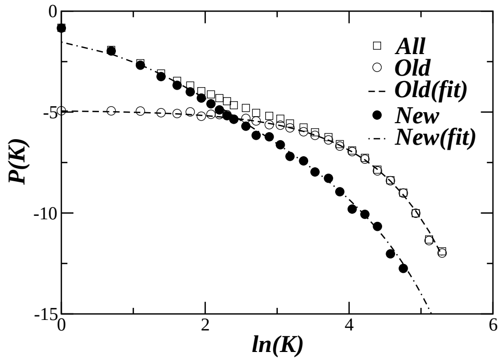


Figure 4.20: Comparison of the distributions $P(K)$ for the new and old classes obtained by numerical simulations of the model with the predictions of Eq. (4.34) and Eq. (4.30) respectively. For large K the agreement is excellent. The discrepancy exists only for $\tilde{P}_{new}$ at small K, e.g. Eq. (4.34) significantly underestimates the $\tilde{P}_{new}(1)$ and $\tilde{P}_{new}(2)$.

Note that Eq. (4.29) and Eq. (4.35) are not exact solutions but continuous approximations which assume K is a real number. This approximation produces the most serious discrepancy for small K. To test this approximation, numerical simulations of the model for $b = 0.1$ is performed, $N(0) = n(0) = 10^4$ and $t = 4 \times 10^4$. The results are presented in Fig. 4.20. While the agreement is excellent for large K, Eq. (4.35) significantly underestimates the value of $\tilde{P}_{new}(K)$ for $K = 1$ and $K = 2$. Note that in reality the power-law behavior of $\tilde{P}_{new}(K)$ extends into the region of very small K.

The predictions of the model is tested using the pharmaceutical industry database (PHID), a micro-level economic database which allows a fine grained decomposition of the statistical properties of growth dynamics of business firms in a given industry. PHID records quarterly sales figures of 48,819 pharmaceutical products commercialized in the European Union and North America from September 1991 to June 2001. The empirical data (Fig. 4.18) have the same shape and power-law exponent (1.14) as the simulation results (Fig. 4.20).

4.3.2.2 Case 2: The Proportional Growth of Size of Units

The model in Sec. 4.3.2.1 can be extended for the purpose of obtaining the growth rate distribution $P(g)$ by introducing the Gibrat process onto the growth of each units. The resulting distribution of the growth rates of all classes is determined by

$$P(g) \equiv \sum_{K=1}^{\infty} P(K)P(g|K), \tag{4.43}$$

where $P(K)$ is the distribution of the number of units in the classes, computed in the previous stage of the model and $P(g|K)$ is the conditional distribution of growth rates of classes with given number of units. Because the form of $P(K)$ is already obtained in Sec. 4.3.2.1, an expression of $P(g)$ can be derived if $P(g|K)$ can be calculated from a new set of assumptions, and they are:

(B1) At time t, each class α has $K_\alpha(t)$ units of size $\xi_i(t)$, $i = 1, 2, ...K_\alpha(t)$ where K_α and $\xi_i > 0$ are independent random variables taken from the distributions $P(K_\alpha)$ and $P_\xi(\xi_i)$ respectively. $P(K_\alpha)$ is defined by Eq. (4.37) and $P_\xi(\xi_i)$ is a given distribution with finite mean and standard deviation and $\ln \xi_i$ has finite mean $\mu_\xi = \langle \ln \xi_i \rangle$ and variance $V_\xi = \langle (\ln \xi_i)^2 \rangle - \mu_\xi^2$. The size of a class is defined as $S_\alpha(t) \equiv \sum_{i=1}^{K_\alpha} \xi_i(t)$.

(B2) At time $t+1$, the size of each unit is decreased or increased by a random factor $\eta_i(t) > 0$ so that

$$\xi_i(t+1) = \xi_i(t)\, \eta_i(t), \tag{4.44}$$

where $\eta_i(t)$, the growth rate of unit i, is an independent random variable taken from a distribution $P_\eta(\eta_i)$, which has finite mean and standard deviation. $\ln \eta_i$ has finite mean $\mu_\eta \equiv \langle \ln \eta_i \rangle$ and variance $V_\eta \equiv \langle (\ln \eta_i)^2 \rangle - \mu_\eta^2$.

Assuming that due to the Gibrat process, both the size and growth of units (ξ_i and η_i respectively) are distributed log-normally

$$p(\xi_i) = \frac{1}{\sqrt{2\pi V_\xi}} \frac{1}{\xi_i} \exp\left(-(\ln \xi_i - m_\xi)^2 / 2V_\xi\right), \tag{4.45}$$

$$p(\eta_i) = \frac{1}{\sqrt{2\pi V_\eta}} \frac{1}{\eta_i} \exp\left(-(\ln \eta_i - m_\eta)^2 / 2V_\eta\right). \tag{4.46}$$

If units grow according to a multiplicative process, the size of units $\xi_i' = \xi_i \eta_i$ is distributed log-normally with $V_{\xi'} = V_\xi + V_\eta$ and $m_{\xi'} = m_\xi + m_\eta$.

The n^{th} moment of the variable x distributed log-normally is given by

$$\begin{aligned} \mu_x(n) &= \int_0^\infty \frac{1}{\sqrt{2\pi V}} \frac{x^n}{x}\, dx \exp\left(-(\ln x - m)^2 / 2V\right) \\ &= \exp\left(n m_x + n^2 V_x / 2\right). \end{aligned} \tag{4.47}$$

Thus, its mean is $\mu_x \equiv \mu_x(1) = \exp(m_x + V_x/2)$ and its variance is $\sigma_x^2 \equiv \mu_x(2) - \mu_x(1)^2 = \mu_x(1)^2 (\exp(V_x) - 1)$.

Remember our goal is to find an analytical approximation for $P(g|K)$ by the distribution $P_\xi(\xi)$ and $P_\eta(\eta)$. To get an exact analytical solution for $P(g|K)$ is difficult, and its expression is approximated by a Gaussian distribution when K is large. In order to do so by the central limit theorem, first it must be shown that, for large K, the random variable g has finite mean and variance.

According to the central limit theorem, the sum of K independent random variables with mean $\mu_\xi \equiv \mu_\xi(1)$ and finite variance σ_ξ^2 is

$$\sum_{i=1}^{K} \xi_i = K\mu_\xi + \sqrt{K}\nu_K, \tag{4.48}$$

where ν_K is the random variable with a distribution converging to a Gaussian

$$\lim_{K\to\infty} P(\nu_K) \to \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left(-\nu_K^2/2\sigma_\xi^2\right). \tag{4.49}$$

Accordingly, $\ln(\sum_{i=1}^K \xi_i)$ can be replaced by its Taylor's expansion $\ln K + \ln\mu_\xi + \nu_K/(\mu_\xi\sqrt{K})$, neglecting the terms of order K^{-1}.

The distribution of g growth rate of classes is now found. From the definition Eq. 4.1 and the assumption (B1),

$$\begin{aligned} g &= \ln\sum_{i=1}^K \xi_i' - \ln\sum_{i=1}^K \xi_i, \\ &= \ln(K\mu_{\xi'}) + \frac{\nu_K'}{\sqrt{K}\mu_{\xi'}} - \ln(K\mu_\xi) - \frac{\nu_K}{\sqrt{K}\mu_\xi}, \\ &= m_\eta + \frac{V_\eta}{2} + \frac{\nu_K'\mu_\xi - \nu_K\mu_{\xi'}}{\sqrt{K}\mu_\xi\mu_{\xi'}}. \end{aligned} \tag{4.50}$$

Here, neglecting the influx of new units, $K_\alpha = K_\alpha(t+1) = K_\alpha(t)$. For large K the last term in Eq. (4.50) is the difference of two Gaussian variables and that is a Gaussian variable itself. Thus for large K, g converges to a Gaussian with mean, $m = m_\eta + V_\eta/2$. Next, whether g has finite variance will be tested.

In order to do this,

$$\frac{\nu_K'}{\sqrt{K}\,\mu_{\xi'}} = \frac{\sum_{i=1}^K(\xi_i' - \mu_{\xi'})}{K\,\mu_{\xi'}},$$

and

$$\frac{\nu_K}{\sqrt{K}\,\mu_\xi} = \frac{\sum_{i=1}^K(\xi_i - \mu_\xi)}{K\,\mu_\xi}$$

are rewritten. Thus

$$\begin{aligned} g &= m_\eta + \frac{V_\eta}{2} + \frac{\sum_{i=1}^K \xi_i(\eta_i\mu_\xi - \mu_{\xi'})}{K\mu_\xi\mu_{\xi'}}, \\ &= m_\eta + \frac{V_\eta}{2} + \frac{\sum_{i=1}^K \xi_i(\eta_i - \mu_\eta)}{K\mu_{\xi'}}. \end{aligned} \tag{4.51}$$

Since $\mu_{\xi'} = \mu_\xi\mu_\eta$, the average of each term in the sum is $\mu_{\xi'} - \mu_\xi\,\mu_\eta = 0$. The variance of each term in the sum is $\langle(\xi_i\,\eta_i)^2\rangle - \langle 2\xi_i^2\,\eta_i\,\mu_\eta\rangle + \langle\xi_i^2\,\mu_\eta^2\rangle$ where $\xi_i\eta_i$, $\xi_i^2\eta_i$ and ξ_i^2 are all log-normal independent random variables. Particularly, $(\xi_i\eta_i)^2$ is log-normal with $V = 4V_\eta + 4V_\xi$ and $m = 2m_\eta + 2m_\xi$; $\xi_i^2\eta_i$ is log-normal with $V = 4V_\xi + V_\eta$ and $m = 2m_\xi + m_\eta$; ξ_i^2 is log-normal with $V = 4V_\xi$ and $m = 2m_\xi$. Using Eq. (4.47)

$$\langle(\xi_i\eta_i)^2\rangle = \exp(2m_\eta + 2m_\xi + 2V_\eta + 2V_\xi), \tag{4.52a}$$

$$\langle\xi_i^2\eta_i\rangle = \exp(m_\eta + 2m_\xi + 2V_\xi + V_\eta/2), \tag{4.52b}$$

$$\langle\xi_i^2\rangle = \exp(2m_\xi + 2V_\xi). \tag{4.52c}$$

Collecting all terms in Eqs. (4.52a-4.52c) together and using Eq. (4.51) the variance of g is:

$$\sigma^2 = \frac{K\,\exp(2m_\xi + 2V_\xi + 2m_\eta + V_\eta)(\exp(V_\eta) - 1)}{K^2\exp(2m_\xi + V_\xi + 2m_\eta + V_\eta)}, \quad = \frac{1}{K}\exp(V_\xi)\,(\exp(V_\eta) - 1). \tag{4.53}$$

Therefore, for large K, g has a Gaussian distribution

$$P(g|K) = \frac{\sqrt{K}}{\sqrt{2\pi V}}\exp\left(-\frac{(g-m)^2 K}{2V}\right), \tag{4.54}$$

where

$$m = m_\eta + V_\eta/2 \tag{4.55}$$

and

$$V \equiv K\sigma^2 = \exp(V_\xi)(\exp(V_\eta) - 1). \tag{4.56}$$

Note, that the convergence of the sum of log-normals to the Gaussian given by Eq. (4.48) is a very slow process, achieving reasonable accuracy only for $K \gg \mu_\xi(2) \sim \exp(2V_\xi)$. For the PHID which is introduced in Sec. 4.3.4 [78, 142], we have $V_\xi = 5.13$, $m_\xi = 3.44$, $V_\eta = 0.36$, and $m_\eta = 0.16$.

Now, obtained the approximation of $P(g|K)$, as Eq. 4.54 shows, using Eq. 4.43 the closed-form approximations for the distribution of growth rate is got. Since in Eq. 4.37, $P(K)$ is the sum of two terms, the $P_{old}(g)$ and $P_{new}(g)$ for old classes and new classes based on $P_{old}(K)$ and $P_{new}(K)$ respectively will be calculated. The shapes of $P_{old}(g)$ and $P_{new}(g)$ should give a similar shape to each other.

The distribution of the growth rate of the old classes can be found by Eq. (4.43). In order to find a closed-form approximation, the summation in Eq. (4.43) is replaced by integration and the distributions $P(K)$ by Eq. (4.29) and $P(g|K)$ by Eq. (4.54). Assuming $m = 0$, we have

$$P_{old}(g) \approx \frac{1}{\sqrt{2\pi V}}\int_0^\infty \frac{1}{K(t)}\exp(\frac{-K}{K(t)})\exp(-\frac{g^2\,K}{2\,V})\sqrt{K}\;dK, \quad = \frac{\sqrt{K(t)}}{2\sqrt{2\,V}}\left(1 + \frac{K(t)}{2V}\,g^2\right)^{-\frac{3}{2}}, \tag{4.57}$$

where $K(t)$ is the average number of units in the old classes (see Eq. (4.28)). This distribution decays as $1/g^3$ and thus does not have a finite variance. In spite of the drastic assumptions, Eq. (4.57) correctly predicts the shape of the convolution $P_{old}(g)$.

For the new classes, when $t \to \infty$ the distribution of the number of units is approximated by

$$P_{\rm new}(K) \approx \frac{1}{1-b}K^{-1-\frac{1}{1-b}}\int_0^K y^{\frac{1}{1-b}}\,e^{-y}\;dy. \tag{4.58}$$

Again replacing summation in Eq. (4.43) by integration and $P(g|K)$ by Eq. (4.54) and after switching the order of integration:

$$P_{new}(g) \approx \frac{1}{1-b}\,\frac{1}{\sqrt{2\pi V}} \int_0^\infty \exp(-y)\, y^{\frac{1}{1-b}}\, dy \int_y^\infty \exp(-g^2\, K/2V)\, K^{(-\frac{1}{2}-\frac{1}{1-b})}\, dK. \tag{4.59}$$

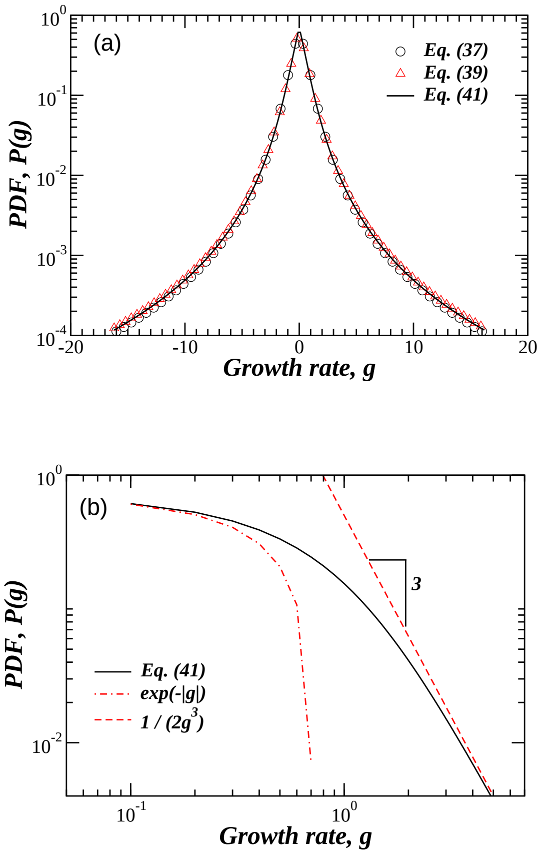


Figure 4.21: (a) Comparison of three different approximations for the growth rate PDF, $P_g(g)$, given by Eq. (4.57), mean field approximation Eq. (4.59) for $b = 0.1$, and by Eq. (4.61). Each $P_g(g)$ shows similar tent shape behavior in the central part. There is little difference between the three cases, $b = 0$ (no entry), $b = 0.1$ (with entry) and the mean field approximation. This means that entry of new classes ($b > 0$) does not perceptibly change the shape of $P_g(g)$. Note that it is used $K(t)/V_g = 2.16$ for Eq. (4.57) and $V_g = 1$ for Eq. (4.61). (b) The crossover of $P_g(g)$ given by Eq. (4.61) between the Laplace distribution in the center and power law in the tails. For small g, $P_g(g)$ follows a Laplace distribution $P_g(g) \sim \exp(-|g|)$, and for large g, $P_g(g)$ asymptotically follows an inverse cubic power law $P_g(g) \sim g^{-3}$.

As $g \to \infty$, the second integral in Eq. (4.59) by partial integration can be evaluated:

$$P_{new}(g) \approx \frac{1}{1-b}\int_0^\infty \frac{1}{\sqrt{2\pi V}}\frac{2V}{g^2}\, y^{-\frac{1}{1-b}-\frac{1}{2}}\, y^{\frac{1}{1-b}}\ \exp(-y)\ \exp(-y\,g^2/2V)\ dy,$$
$$= \frac{1}{1-b}\frac{1}{\sqrt{2\pi V}}\frac{2V}{g^2}\frac{1}{\sqrt{g^2/2V+1}}\sqrt{\pi} \sim \frac{1}{g^3}. \qquad (4.60)$$

It can be computed the first derivative of the distribution (4.59) by differentiating the integrand in the second integral with respect to g. The second integral converges as $y \to 0$, and we find the behavior of the derivative for $g \to 0$ by the substitution $x = Kg^2/(2V)$. As $g \to 0$, the derivative behaves as $g\,g^{2[-(3/2)+1/(1-b)]} \sim g^{2b/(1-b)}$, which means that the function itself behaves as $C_2 - C_1|g|^{2b/(1-b)+1}$, where C_2 and C_1 are positive constants. For small b this behavior is similar to the behavior of a Laplace distribution with variance V: $\exp(-\sqrt{2}|g|/\sqrt{V})/\sqrt{2V} = 1/\sqrt{2V} - |g|/V$.

When $b \to 0$, Eq. (4.59) can be expressed in elementary functions:

$$P_{new}(g)|_{b\to 0} \approx \frac{1}{\sqrt{2\pi V}}\int_0^\infty K^{-3/2}\exp(-K\,g^2/2\,V)\,dK\int_0^K \exp(-y)\,y\ dy,$$
$$\approx \frac{1}{\sqrt{2\,V}}\left(-\frac{1}{\sqrt{1+g^2/2\,V}} + \frac{2}{|g|/\sqrt{2\,V}+\sqrt{g^2/2\,V+1}}\right).$$

Simplifying, the main result is:

$$P_{\text{new}}(g)|_{b\to 0} \approx \frac{2V}{\sqrt{g^2+2V}\,(|g|+\sqrt{g^2+2V})^2}. \qquad (4.61)$$

which behaves for $g \to 0$ as $1/\sqrt{2V} - |g|/V$ and for $g \to \infty$ as $V/(2g^3)$. Thus the distribution is well approximated by a Laplace distribution in the body with power-law tails. Because of the discrete nature of the distribution of the number of units, when $g \gg \sqrt{2V}$ the behavior for $g \to \infty$ is dominated by $\exp(-g^2/2V)$.

In Fig. 4.21a the distributions given by Eq. (4.57), the mean field approximation Eq. (4.59) for $b = 0.1$ and Eq. (4.61) for $b \to 0$ is compared. All three distributions have very similar tent-shaped behavior in the central part. In Fig. 4.21b the distribution Eq. (4.61) with its asymptotic behaviors for $g \to 0$ (Laplace cusp) and $g \to \infty$ (power law) is also compared, and it is found the crossover region between these two regimes.

4.3.3 Analytical Results

The analytical solution of this model can be obtained only for certain limiting cases but a numerical solution can be easily computed for any set of assumptions. The model is investigated numerically and analytically and it is found:

(1) In the presence of the influx of new classes ($b > 0$), the distribution of units converges for $t \to \infty$ to a power law $P(K) \sim K^{-\varphi}$, $\varphi = 2 + b/(1-b) \geq 2$. Note that this behavior of the power-law probability density function leads to a

power law rank-order distribution where the rank of a class R is related to the number of its units K as

$$R = N(t)\int_K^\infty P(K)dk \sim K^{-\varphi+1}. \tag{4.62}$$

Thus $K \sim R^{-\zeta}$, where $\zeta = 1/(\varphi - 1) = 1 - b \leq 1$, which leads in the limit $b \to 0$ to the celebrated Zipf's law[210] for cities' populations, $K \sim 1/R$. Note that this equation can be derived for the model using elementary considerations. Indeed, due to proportional growth the rank of a class, R, is proportional to the time of its creation t_0. The number of units $n(t_0)$ existing at time t_0 is also proportional to t_0 and thus also proportional to R. According to the proportional growth, the ratio of the number of units in this class to the number of units in the classes that existed at time t_0 is constant: $K(t_0, t)/n_e(t_0, t) = 1/n(t_0)$. If it is assumed that the total number of units in the classes created after t_0, can be neglected since the influx of new classes b is small, $n_e(t_0, t)$ can be approximated by $\approx n(t) \sim t$. Thus for large t, $n_e(t_0, t)$ is independent of t_0 and hence $K(t_0, t) \sim 1/R$. If the influx of new classes is not neglected, Eq. (4.27) gives $n_e(t_0, t) \sim t_0^b$, hence $K(t_0, t) \sim 1/R^{1-b}$.

(2) The conditional distribution of the logarithmic growth rates $P(g|K)$ for the classes consisting of a fixed number K of units converges to a Gaussian distribution (4.54) for $K \to \infty$. Thus the width of this distribution, $\sigma(K)$, decreases as $1/K^\beta$, with $\beta = 1/2$. Note that due to slow convergence of log-normals to the Gaussian in the case of a wide log-normal distribution of unit sizes $V_\xi = 5.13$, computed from the empirical data [78], and $\beta = 0.2$ for relatively small classes. This result is consistent with the observation that large firms with many production units fluctuate less than small firms [3, 4, 99, 194]. Interestingly, in case of large V_ξ, $P(g|K)$ converges to the Gaussian in the central interval which grows with K, but outside this interval it develops tent-shape wings, which become increasingly wider, as $K \to \infty$. However, they remain limited by the distribution of the logarithmic growth rates of the units, $P_\eta(\ln \eta)$.

(3) For $g \gg V_\eta$, the distribution $P(g)$ coincides with the distribution of the logarithms of the growth rates of the units:

$$P(g) \approx P_\eta(\ln \eta). \tag{4.63}$$

In the case of a power law distribution $P(K) \sim K^{-\varphi}$ which dramatically increases for $K \to 1$, the distribution $P(g)$ is dominated by the growth rates of classes consisting of a single unit $K = 1$, thus the distribution $P(g)$ practically coincides with $P_\eta(\ln \eta)$ for all g.

(4) If the distribution $P(K) \sim K^{-\varphi}$, $\varphi > 2$ for $K \to \infty$, as happens in the presence of the influx of new units $b \neq 0$, $P(g) = C_1 - C_2|g|^{2\varphi-3}$, for $g \to 0$ which in the limiting case $b \to 0$, $\varphi \to 2$ gives the cusp $P(g) \sim C_1 - C_2|g|$ (C_1 and

C_2 are positive constants), similar to the behavior of the Laplace distribution $P_L(g) \sim \exp(-|g|C_2)$ for $g \to 0$.

(5) If the distribution $P(K)$ weakly depends on K for $K \to 1$, the distribution of $P(g)$ can be approximated by a power law of g: $P(g) \sim g^{-3}$ over a wide range $\sqrt{V_g/K(t)} \ll g \ll \sqrt{V}$, where $K(t)$ is the average number of units in a class. This case is realized for $b = 0$, $t \to \infty$ when the distribution of $P(K)$ is dominated by the exponential distribution and $K(t) \to \infty$ as defined by Eq. (4.21). In this particular case, $P(g)$ for $g \ll \sqrt{V_g}$ can be approximated by Eq.(4.57)

(6) In the case in which the distribution $P(K)$ is not dominated by one-unit classes but for $K \to \infty$ behaves as a power law, which is the result of the mean field solution for our model when $t \to \infty$, the resulting distribution $P(g)$ has three regimes, $P(g) \sim C_1 - C_2|g|^{2\varphi-3}$ for small g, $P(g) \sim g^{-3}$ for intermediate g, and $P(g) \sim P(\ln \eta)$ for $g \to \infty$. The approximate solution of $P(g)$ in this case is given by Eq. (4.59) For $b \neq 0$ Eq. (4.59) can not be expressed in elementary functions. In the $b \to 0$ case, Eq. (4.59) yields the main result Eq.(4.61). which combines the Laplace cusp for $g \to 0$ and the power law decay for $g \to \infty$. Note that due to replacement of summation by integration in Eq. (4.43), the approximation Eq. (4.61) holds only for $g < \sqrt{V_\eta}$.

In conclusion although the derivations of the distributions (4.57), (4.59), and (4.61) are not rigorous they satisfactory reproduce the shape of empirical data, especially the $1/g^3$ behavior of the wings of the distribution of the growth rates and the sharp cusp near the center.

4.3.4 Empirical Evidence

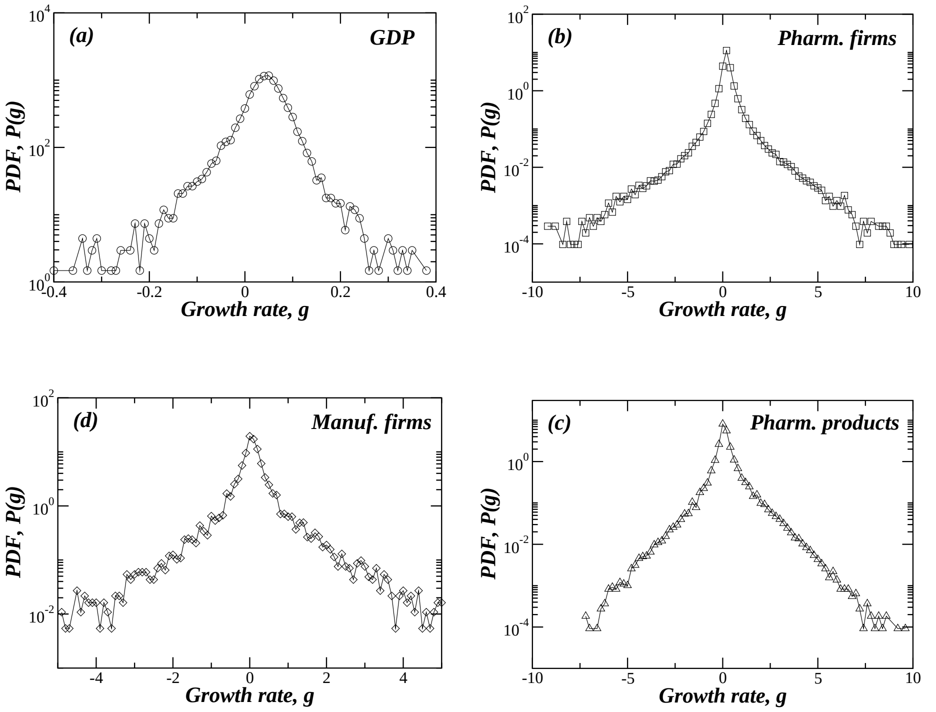


Figure 4.22: Empirical results of the probability density function (PDF) $P(g)$ on different subjects. (a) GDP for 195 countries from 1960 to 2004. (b) all firms in PHID database. (c) All U.S. publicly-traded manufacturing firms from 1973 to 2004 in Compustat database. (d) all products in PHID database.

To test the model, different levels of aggregation of economic systems are analyzed, from the micro level of products to the macro level of industrial sectors and national economies. The empirical data for several examples are shown in Fig. 4.22. First, a new and unique database is analyzed, the pharmaceutical industry database (PHID), which records sales figures of the $189,303$ products commercialized by $7,184$ pharmaceutical firms in 21 countries from 1994 to 2004, covering the whole size distribution for products and firms and monitoring the flows of entry and exit at both levels. The database was kindly provided by the EPRIS program. Then, the growth rates of all U.S. publicly-traded firms from 1973 to 2004 in all industries, based on Security Exchange Commission filings (Compustat) is studied. Finally, at the macro level, the growth rates of the gross domestic product (GDP) of 195 countries from 1960 to 2004 (World Bank) is studied.

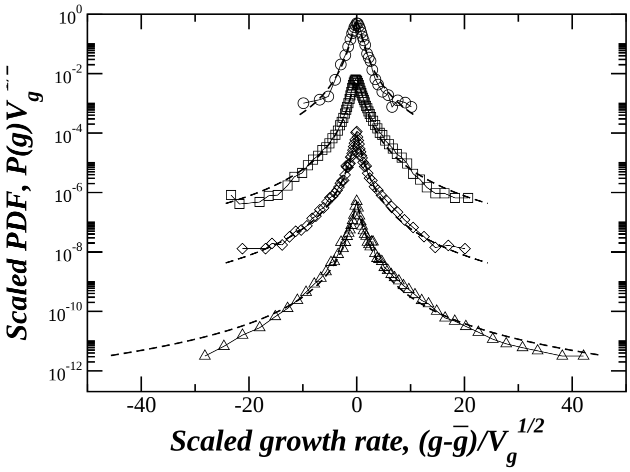


Figure 4.23: Empirical tests of Eq. (4.61) for the probability density function (PDF) $P(g)$ of growth rates rescaled by $\sqrt{V}$. The shapes of $P(g)$ for all four levels of aggregation are well approximated by the PDF predicted by the model (dashed lines). Dashed lines are obtained based on Eq. (4.61) with $V \approx 4 \times 10^{-4}$ for GDP, $V \approx 0.014$ for pharmaceutical firms, $V \approx 0.019$ for manufacturing firms, and $V \approx 0.01$ for products. After rescaling, the four PDFs can be fit by the same function. For clarity, the pharmaceutical firms are offset by a factor of 10^2, manufacturing firms by a factor of 10^4 and the pharmaceutical products by a factor of 10^6. Note that the data for pharmaceutical products extend from $P(g) = 1$ to $P(g) \approx 10^{-4}$ and the mismatch in the tail parts is because $P(g)$ for large g is mainly determined by the logarithmic growth rates of units $\ln \eta$.

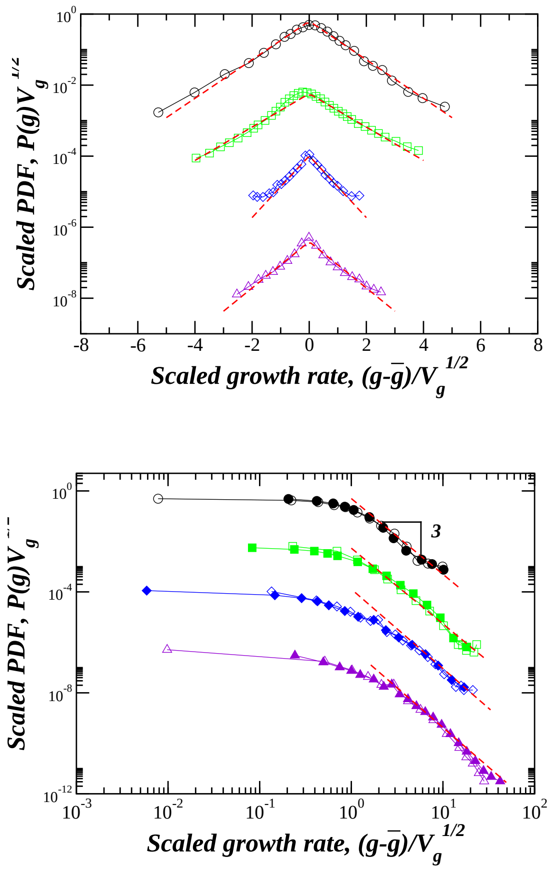


Figure 4.24: (a) Empirical tests of Eq. (4.61) for the *central* part in the PDF $P(g)$ of growth rates rescaled by $\sqrt{V}$. The shape of central parts for all four levels of aggregation can be well fit by a Laplace distribution (dashed lines). Note that a Laplace distribution can fit $P(g)$ only over a restricted range, from $P(g) = 1$ to $P(g) \approx 10^{-1}$. (b) Empirical tests of Eq. (4.61) for the *tail* parts of the PDF of growth rates rescaled by $\sqrt{V}$. The asymptotic behavior of g at any level of aggregation can be well approximated by power laws with exponents $\zeta \approx 3$ (dashed lines). The symbols are as follows: Country GDP (left tail: ◯, right tail: •), pharmaceutical firms (left tail: □, right tail: ■), manufacturing firms (left tail: ◊, right tail: ⧫), pharmaceutical products (left tail: △, right tail: ▲).

Figure 4.23 shows that the growth distributions of countries, firms, and products are well fitted by the distribution in Eq. (4.61) with different values of V_g. Indeed, the growth distributions at any level of aggregation depict marked departures from Gibrat's Gaussian shape. Moreover, even if the $P_g(g)$ of GDP can be approximated by a Laplace distribution, the $P_g(g)$ of firms and products are clearly more leptokurtic than Laplace. Based on the model, the growth distribution is Laplace in the body, with power-law tails. In fact, Fig. 4.24a shows that the central body part of the growth rate distributions at any level of aggregation is well approximated by a double exponential fit. Fig. 4.24b reveals that the asymptotic behaviors of g at any level of aggregation can be well fitted by a power-law with an exponent $\zeta = 3$.

The analysis in Sec. 4.3.2.1 predicts that the power law regime of $P_g(g)$ may vary depending on the behavior of $P(K)$ for $K \to 1$, and the distribution of the growth rates of units. In case of the PHID, for which $P(1) \gg P(2) \gg P(3) \ldots$ the growth rate distribution of firms must be almost the same as the growth rate distribution of products, as it was stated in Sec. II. Hence the power law wings of $P_g(g)$ for firms originate on the level of products. Because the PHID does not contain information on the subunits of products it is not possible test the prediction directly, but it can be hypothesized that the distribution of the product subunits (number of customers or shipping ways) is less dominated by small K, but has a sufficiently wide power law regime due to the influx of new products. These rather plausible assumptions are sufficient to explain the shape of the distribution of the product growth rates, which is well described by Eq. (4.61).

To further test the universality of fitting function Eq. 4.61, a new database from pharmaceutical industry in two major developing countries: China and India is used. In Fig. 4.25, before the scaling transformation, the firm growth rate distributions in two countries are quite different because it is clear that India has a bigger spread or volatility of growth rates than China; but after scaling with appropriate V_g, $P(g)$ for the different countries overlapped with each other and they clearly follow the prediction of Eq. 4.61.

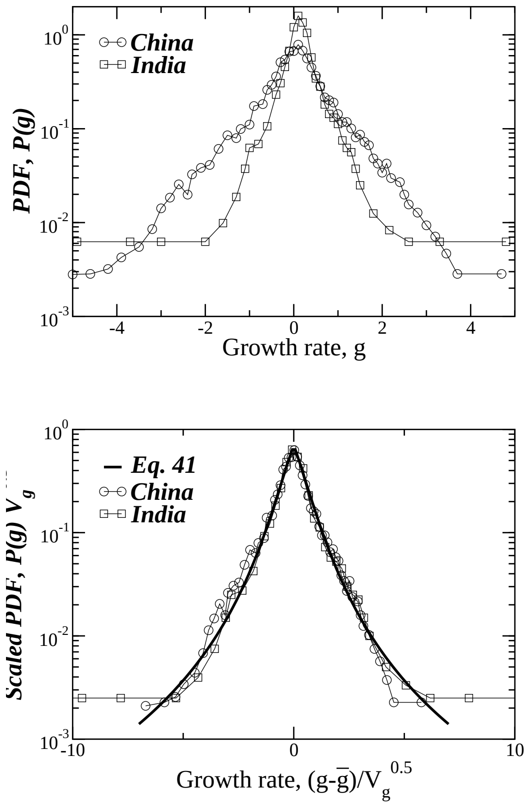


Figure 4.25: (a) Empirical $P(g)$ of pharmaceutical companies in China and India. As it shows, the spread of $P(g)$ for China is wider than for India which implies that the growth in the pharmaceutical companies in China is more volatile than in India. (b) After rescaling the $P(g)$ with V, rescaled $P(g)$ for China and India pharmaceutical companies overlaps and both of them are well fitted by the Eq. 4.61. It supports the claim that Eq. 4.61 can be a universal fitting function of $P(g)$, even in developing countries.

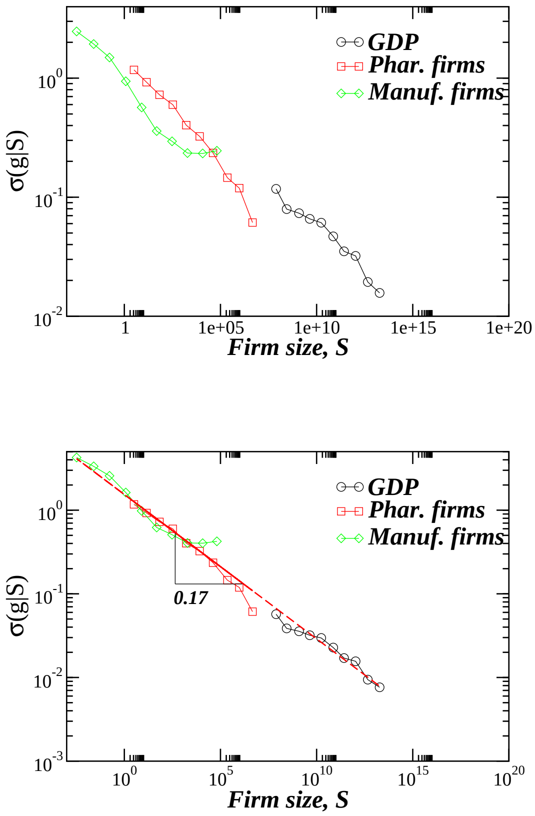


Figure 4.26: (a) The empirical results of the relationship between the standard deviation of growth rates and firm size. Note that they have similar shape with a similar slope. (b) The β curves of GDP and Manufacturing firms, pharmaceutical firms are offset a little in order to get a fit by one power-law function.

Finally another important empirical result is tested: the power-law relationship between standard deviation of firm growth rates, $\sigma(g|S)$, and firm size S. Based on previous empirical findings, the power-law exponent β is 0.15 for country's GDP [120], 0.17 for American Manufacturing Industries [3], and 0.20 for PHID [142]. In FIG. 4.26, all of those power-law curves are put together, and it is offset some to fit them by a power law with $\beta = 0.17$. A numerical simulation gives $\beta = 0.18$ as Fig. 4.27 shows.

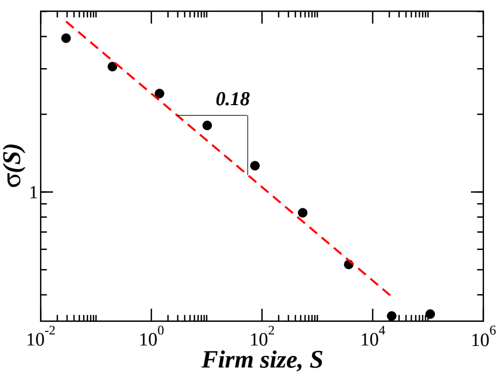


Figure 4.27: The simulation result of the relationship between the standard deviation of growth rates and firm size. The model runs about 1 million time steps. This simulation has $\alpha = 0.05$, $\lambda = -0.52$, $\mu = 0.50$, a log-normal distribution of product size ξ with mean value 3.44 and standard deviation 5.13.

The existence of β implies a scaling phenomenon. Following [142], it is separated all firms by size into three categories: small, medium, and large firms. For each category, the distribution of firm growth rates as Fig. 4.28 is plotted. As it shows, the conditional distribution of growth rates $P(g|S)$ gives similar shape and the spread of the conditional distribution for small firms is indeed much larger than the one for large firms. Finally, scaling transformation in Fig. 4.28 is done, that is, the y axis—$P(g|S)$— is multiplied by the standard deviation of growth rate $\sigma(g|S)$ and divide the x axis—g—by $\sigma(g|S)$. In Fig. 4.29 the three conditional growth-rate distributions collapse onto a single curve.

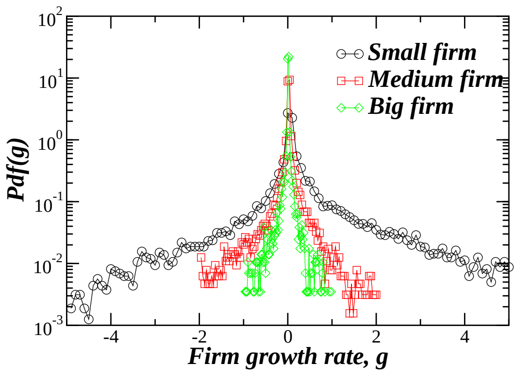


Figure 4.28: The simulation result of $P(g)$ for different firm size. All firms are divided into three categories: small, medium and large. The three conditional distribution of growth rates give a similar shape which implies that they can be fit by the same function.

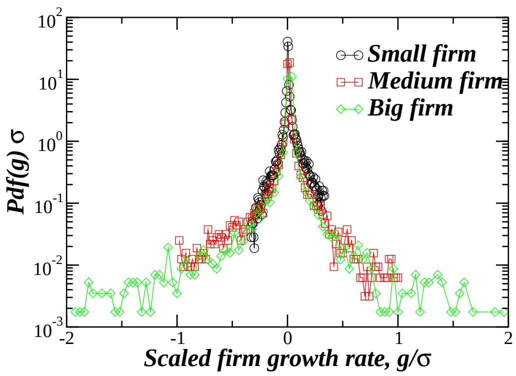


Figure 4.29: The simulation result of $P(g|S)$ rescaled by standard deviation of growth rates. As it shows, the three different curves collapse onto a single curve.

4.4 The size-variance relationship

Fig. 4.30a shows the size-variance relationship for firms. Although the slope $\beta(S)$ increases with S, it does not display a strong crossover predicted by the preferential attachment model described here. This discrepancy may arise from the inaccuracy of two assumptions made in the model. The first assumption is that the new units attached to a class are taken at random from a general distribution of unit sizes which does not depend on the size of a particular class. The second assumption is that the

growth rates of units are taken at random from a general distribution of unit growth rates which do not depend on a size of a unit.
To test the first possibility, the products are randomly reassigned to the firms keeping the number of the products in each firm unchanged, and keeping the history of the fluctuation of each product sales unchanged.
This randomization practically does not change the size-variance dependence of the firms. (Fig. 4.30a).
This test demonstrates that although there is a small correlation between the number of products in the firm and their sizes (see Fig. 4.30b), this correlation is not responsible for the origin of the power-law size-variance relationship observed for the empirical data.
To test the second possibility, the sizes of products ξ_i and their number K_α at year t for each firm are kept the same as in the original data, so $S_t = \sum_{i=1}^{K_\alpha} \xi_i$ is the same as in the empirical data. However, to compute the sales of a firm in the following year $\widetilde{S}_{t+1} = \sum_{i=1}^{K_\alpha} \xi'_i$, it is assumed that $\xi'_i = \xi_i \eta_i$, where η_i is an annual growth rate of a randomly selected product. The surrogate growth rate $\widetilde{g} = ln\frac{\widetilde{S}_{t+1}}{S_t}$ obtained in this way does not display any size-variance relationship (Fig. 4.30c). The second test shows that the size-variance relationship for the firm growth rates in the pharmaceutical data base is generated on the level of products. Indeed, the size-variance relationship of the growth rates of products $g_\xi = ln(\frac{\xi_{t+1}}{\xi_t})$ shows a large range of approximate power law behavior with $\beta = 0.096$ (Fig. 4.30b). The origin of this size-variance relationship cannot be determined from the present data set. It can come from the fact that small experimental products prescribed by few physicians to few patients are less stable then well established products prescribed to large number of patients.
Comparing the size-variance relationship for firms and products, for small firms this two distribution almost coincide, however for large firms,the exponent β is much larger than for the products.

Figure 4.31 shows the survivor function for ρ_i^t. ρ_i^t is the number of products that represent the 50% of the whole company i size at time t. Figure 4.31 tells how many products give the 50% of the whole company size. This curve is the survivor function, and for $\rho < 10$ the curve is a power law with exponent around 1.7. In only few cases the company size is determined by more than 10 products. In the most cases the firms size is only due at 2 or 3 larger products. The fluctuations of the firms are due to the fluctuations of the largest or the few larger products. This data set tells only that the distribution of the number of products is power law, and there is no more information. So the shape of the firms growth rate is not due to the preferential attachment model, that only supports this evidence, but to the shape of the products growth rate. The mystery is why products have such kind of distribution, but this data set has not enough information to understand this behavior.
The origin of the size-variance relationship of the firms comes from the analogous

size-variance relationship of the products and not from the preferential attachment model. To test this, the data have been left almost unchanged, randomizing only one parameter of the data at a time. In the first case, the products are reassigned to the firms in random order (keeping the number of products in each firm and the history of each products unchanged). Figure 4.30a shows β for firms left almost unchanged. This lets to conclude that β is not only due to the allocation process of products to the firms. In the other case, the product growth rate is reassigned in a random order, keeping the number of the products in each firm, K, and the size of each product ξ_i in the year t the same as in original data. The firm size at $t+1$ is computed as $\sum_{i=1}^{K} \xi_i \eta_k$ where η_k is the growth rate of a randomly selected product. Figure 4.30b and Figure 4.30c show the results. In this case beta becomes -0.028. There is no size-variance relationship. These two analysis clearly show that β originates already on the level of products but not due to the crossover in the preferential attachment. This database has not enough information to understand more deeply the origin of β.

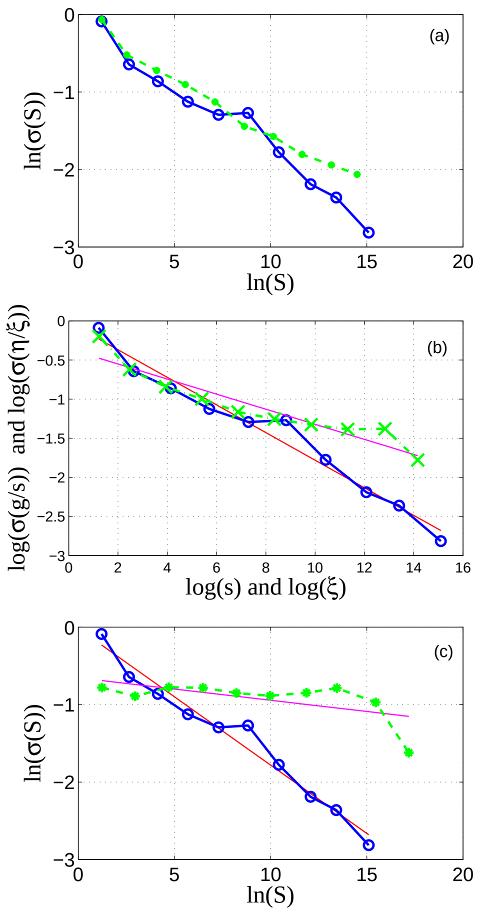


Figure 4.30: (a) Size-variance relationship at the firm level. Continuous line with circle represents the empirical data. The slope of the linear fit is -0.18 ($\beta = -0.18$). Dashed line with asterisks shows the size-variance relationship for the surrogate data set in which it is randomly allocated the products to the firms, keeping the number of products in each firms and the historical records for each product unchanged. In this case the slope is -0.14 ($\beta = -0.14$). (b) Size-variance relationship at the product and firms level. The continuous line with circle plots the empirical data for firms, whereas the dashed line with cross the data for products.The slope for the firms data is $\beta = -0.18$ and for products is $\beta = -0.096$. (c) Size-variance relationship at the firm level. The continuous line with circle plots the empirical data, whereas the dashed line with asterisks the surrogate data t but after random reassignment of products growth rates. The slope for the empirical data is $\beta = -0.18$, but after random reassignment of the products growth rates explained in the text the slope becomes $\beta = -0.028$

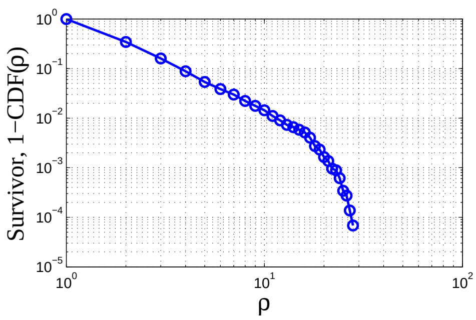


Figure 4.31: Survivor function of ρ. ρ_i^t is the number of products that represent the 50% of the whole company i size at time t.

.

Chapter 5

Remarks and Conclusions

5.1 Single asset model and empirical results

Empirical analysis and computational experiments of high-frequency data for a double-auction (book) market have been presented. The results of Raberto et al. 2004 [161] are confirmed and strengthened. Indeed, the return distribution is leptokurtic not only when the order generation process is Poisson but also in presence of memory, as in the case of Weibull-distributed waiting times. Moreover, with memory, return tails are fatter.

This result deserves attention because previous empirical analysis [132, 164, 175] has shown that the distribution of trade waiting times is non-exponential. Conversely, exponentially distributed trade waiting times only result from a finite thinning of a Poisson order process. Consequently, the distribution of order waiting times should be a more general distribution, e.g., a Weibull, than an exponential distribution underlying a simple memoryless Poisson process.

The hypothesis of non exponentially distributed order waiting times is empirically accessible and can be directly checked if full book information were available. Moreover, one can try to solve the inverse problem: given a trade waiting time distribution, which is the originating order waiting time distribution?

Empirical analysis confirmed that trading waiting times are not exponentially distributed, whereas a Weibull process cannot be rejected. This result deserves attention because exponentially distributed trade waiting times only result from a finite thinning of a Poisson order process. Consequently, the distribution of order waiting times should be a more general distribution than an exponential distribution underlying a simple memoryless Poisson process.

In order to better understand the relationship between order and trading waiting times, computer experiments on the Genoa Artificial Stock Market have been considered. In particular, a double-auction mechanism (i.e., limit order book) with order waiting times characterized by a mixture of Poisson process has been modeled and implemented. The characteristics of Poisson process in the mixture have been properly estimated by real data. It has been shown that, both order and trading waiting times, reject hypothesis of exponentially distributed point process. Conversely, the hypothesis of Weibull distribution cannot be rejected and estimation of the corresponding Weibull parameter β pointed out values quite close to those obtained in the case of 30 DJIA stocks traded at NYSE in October 1999. Finally, the influence of the number of exponential distribution of the mixture appears almost negligible as far as survivals of order and trading waiting times and β do not change significantly. This allows one to conclude that already a mixture of two Poisson process can be sufficient in order to reproduce the behavior of real stock market.

5.2 Multi assets model and empirical results

The ability to recover these stylized facts in a framework where only zero-intelligence traders operate is an interesting and surprising result. Moreover, it is worth noting that for the difference of the business sectors dividends or cash inflow are not important, in fact also in simulation without these exogenous parameters business sectors are differentiated. So only the restrictions on agents' allocation strategies produce this difference. In the standard economic and finance theory, agents usually follow an utility of profit maximizing behavior with adaptive or rational expectations about the future. Only in the last decade, following the pioneering work by Gode and Sunder [84], agents endowed with zero intelligence have been take into account. Zero-intelligence behavior still deserve much interest for its simplicity and the possibility to focus the attention more on the structural aspects than on the behavioral features.

Moreover an artificial stock market characterized by heterogeneous and interacting agents has been studied. In this complex system, agents are characterized by cash, stocks and sentiments. Sentiments denote the propensities to buy or to sell of agent. Agents are seen as nodes of sparsely connected graph, so that each agent is influenced by a subset of agents, the only ones that are "near" to him. The statistical properties of the univariate and the multivariate process of prices and returns are studied. In particular, concerning univariate price processes, the proposed approach was able to endogenously reproduce the property of unitary root, of volatility clustering and of fat tail distribution of returns. Furthermore, concerning the multivariate price process, the evidence of static factors in the returns and the presence of common trends in the prices have been investigated. The presence of static factors has been studied making reference to the cross-correlations between returns of different stocks, whereas the presence of common trends has been carried on considering the variance-covariance matrix of prices. The computational experiments pointed out the possibility to endogenously reproduce the multivariate stylized facts on cross-correlation matrix and on variance-covariance matrix. Finally, it is worth remarking the importance of this result, as for the first time, an artificial stock market reproduce endogenously all this features.

5.3 The firm growth problem

Business firms grow by increasing their scale and scope. The scope of a firm is given by the number of its products. The scale of a firm is given by the size of its products. A firm like Microsoft has a few big products while Amazon sells a huge variety of goods, each of small size in terms of sales. In this article it is argued that both mechanisms of growth are proportional. The number of products a firm can successfully launch is proportional to the number of products it has already commercialized. Once a product has been launched its success depends on the number of customers who buy it and the price they are willing to pay. To a large extent, if products are different enough, the success of a product is independent of other products commercialized by the same company. Hence, the sales of products can be modeled as independent stochastic processes. Moreover, sometimes, new products are commercialized by new companies. As a result, small companies with few products can experience sudden jerks of growth due to the successful launch of a new product.

It has been found that the empirical distribution of firm growth rates exhibits a central part which is distributed according to a Laplace distribution and power-law wings $P_g(g) \sim g^{-\zeta}$ where $\zeta = 3$. If the distribution over classes of the number of units K is dominated by single unit classes, the tails of firm growth rate are primarily due to smaller firms that have one or few products. The Laplace center of the distribution is shaped by big multiproduct firms. The shape of the distribution of firm growth is almost the same whether there is a small entry rate or zero entry. The model predictions are accurate also in the case of product growth rates, which implies that products can be considered as composed of elementary sale units, which evolve according to a random multiplicative process [190]. Although there are several plausible explanations for the Laplace body of the distribution [3, 113], the power law decay of the tails has not previously been explained. A simple and general model that accounts for both the central part and the tails of the distribution is introduced. The shape of the business growth rate distribution is due to the proportional growth of both the number and the size of the constituent units in the class. This result holds in the case of an *open* economy (with entry of new firms) as well as in the case of a *closed* economy (with no entry of new firms).

This model can numerically duplicate the size-variance relationship. The power-law relation implies that for different firm sizes, firm growth rates follow the same dynamics, and the simulation results of our model verify this point.

This study provides a framework for the firm growth problem and it can act as a direction of future study of the research. Although the model justifies some useful findings, some questions still remain open to investigate. For example, (1) the model does not consider a common economic phenomenon — the merging and splitting of firms. Can this model be modified to include that? (2) the model emphasizes the effect of the preferential attachment mechanism, and from it, a simple fitting function for $P(g)$ is got and unfortunately, the expression of $P(g)$ does not include the term S. Thus, so far, the size-variance relationship from $P(g)$ is not derived and we only showed the numerical results of size-variance relationship. These open questions require to further modify and refine the model in order to more deeply understand the firm growth problem.

5.4 Open questions of future studies

The next step can be to construct a unique model to reproduce the main statistical properties of real financial markets and of the firm growth problem. A bottom-up agent-based approach allows one to represent the economy as a large complex adaptive system consisting of a huge number of independent agents, that interact in various ways and that change their state or actions as a result of the events in the process of interaction. Reliable agent-based software simulators represent the computational laboratories to perform and to test the impact of different economic measures, e.g., the effects of a tax scheme on economic welfare and equality or the effectiveness of industrial policies aimed at increasing the average firm size in order to boost innovation.

In the top-down approach of traditional neoclassical models the bottom level typically comprises a representative individual which is constrained by strong consistency requirements associated with equilibrium and rationality. Conversely, a fundamental characteristic of agent-based models is heterogeneity of agents, that can range from initial endowments, psychological attitudes, social dimensions, to behavioral rules, preferences and degree of rationality. Therefore, agent-based models are perfectly fit to take into account the realistic diversity of human behaviors. Within this approach, a central issue is the theoretical and algorithmic foundation of a technological platform for scalable modeling and simulation, where a new formalism capable of representing and modeling complex multilevel systems can be developed.

The empirical validation of agent-based computational models of a real-world system is a central issue in order to provide reliable what-if scenarios. The large amount of data available to economists, particularly in the finance domain, opens the possibility to set up data driven agent-based simulations, where the predictions are continuously adjusted by the assimilation of new real data.

The complex strategic interaction of the huge number of actors in the market is further strongly affected by an underling physical network, which implies constraints in generation, transmission, distribution, storage, etc.

A top-down approach to manage such kind of systems is unrealistic, and the bottom-up complex artificial world in one-to-one correspondence with a real economic system should be adopted in order to design, to validate, to infer and to control the behavior of service and utility markets. This represents a grand multidisciplinary challenging research that involves competencies originally developed in engineering, computer science, economics, mathematics and physical sciences.

BIBLIOGRAPHY

[1] The assumption here is necessary to estimate the theoretical value of γ, considering the limited historical information of the phid.

[2] For more general initial conditions, the model predicts the same behavior for the sufficiently large t, but with different value of the exponential decay constant γ.

[3] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, and . M. H. R. Stanley, Journal de Physique I (France) **7**. 1997.

[4] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, M. A. Salinger, and . H. E. Stanley, Physical Review Letter **80**. 1998.

[5] M. Aoki. *New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria, and Externalities as Field Effects.* Cambridge University Press, 1996.

[6] M. Aoki. *Modeling Aggregate Behavior and Fluctuations in Economics: Stochastic Views of Interacting Agents.* Cambridge University Press, 2002.

[7] H. Aoyama, Y. Nagahara, M. Okasaki, W. Souma, H. Takayasu, and M. Takayasu. *Fractals* **8***, 293*, 2000.

[8] J. Arifovic. The behavior of the exchange rate in the genetic algorithm and experimental economies. *J Polit Econ*, 104(3):510–541, Jun 1996.

[9] G. Armano and M. Marchesi. A rapid development process with UML. *ACM SIGAPP Applied Computing Review*, 8(1):4–11, April 2000.

[10] W. B. Arthur. Inductive reasoning and bounded rationality. *Am Econ Rev*, 84(2):406–411, May 1994.

[11] W. B. Arthur, J. H. Holland, B. D. LeBaron, R. Palmer, and P. Tayler. Asset pricing under endogeneous expectations in an artificial stock market. In W. Arthur, S. Durlauf, and D. Lane, editors, *The Economy as an Evolving Complex System II*, SFI Studies in the Sciences of Complexity. Addison Wesley Longman, 1997.

[12] R. Axtell. The emergence of firms in a population of agents: Local increasing returns, unstable nash equilibria, and power law size distributions,.

[13] R. L. Axtell. *Science* **293***, 1818*, 2001.

[14] P. Bak, M. Paczuski, and M. Shubik. Price variations in a stock market with many agents. *Physica A*, 246(3-4):430–453, December 1997.

[15] N. Basu, R. Pryor, and T. Quint. Aspen: a microsimulation model of the economy. *Computational Economics*, 12(3):223–241, December 1998.

[16] M. Batty and P. Longley. Fractal cities. (Academic Press, San Diego, 1994).

[17] W. Baumol. Business behavior, value, and growth.

[18] K. Beck. *Extreme Programming Explained. Embrace Change.* Addison-Wesley, Reading, Massachusetts, 1999.

[19] A. Beja and M. B. Goldman. On the dynamic behavior of prices in disequilibrium. *J Financ*, 35:235–247, May 1980.

[20] A. Beltratti and S. Margarita. *Evolution of trading strategies among heterogeneous artificial economic agents.* MIT Press, Cambridge, MA, 1992.

[21] A. Beltratti, S. Margarita, and P. Terna. *Neural networks for economic and financial modelling.* Thomson Computer Press, 1997.

[22] W. K. Bertram. An empirical investigation of australian stock exchange data. *Physica A*, 341:533–546, October 2004.

[23] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris bourse. *J Financ*, 50(5):1655–1689, December 1995.

[24] F. Black. Noise. *J Financ*, 41(3):529–543, July 1986.

[25] L. Blume and D. Easley. Evolution and market behaviour. *J Econ Theor*, 58:9–40, 1992.

[26] G. Bottazzi, G. Dosi, M. Lippi, F. Pammolli, and . M. Riccaboni, International Journal of Industial Organization **19**. 2001.

[27] G. Bottazzi, G. Dosi, M. Lippi, F. Pammolli, and . M.Riccaboni, International Journal of Industrial Organization **19**. 2001.

[28] J.-P. Bouchaud and M. Mezard. Wealth condensation in a simple model of economy. *Physica A*, 282(3-4):536–545, June 2000.

[29] J.-P. Bouchaud, M. Mezard, and M. Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, August 2002.

[30] J.-P. Bouchaud and M. Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, UK, 2000.

[31] W. A. Brock and C. H. Hommes. Rational route to randomness. *Econometrica*, 65(5):1059–1095, September 1997.

[32] W. A. Brock and C. H. Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *J Econ Dyn Control*, 22(8-9):1235–1274, August 1998.

[33] S. V. Buldyrev, L. A. N. Amaral, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, and . M. H. R. Stanley, Journal de Physique I **7**. 1997.

[34] S. V. Buldyrev, N. V. Dokholyan, S. Erramilli, M. Hong, J. Y. Kim, G. Malescio, and . H. E. Stanley, Physica A **330**. 2003.

[35] S. V. Buldyrev, N. V. Dokholyan, S. Erramilli, M. Hong, J. Y. Kim, G. Malescio, and . H. E. Stanley, Physica A **330**. 2003.

[36] J. T. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton Univ Pr, USA, 1997.

[37] J. Y. Campbell and R. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review Financial Studies*, 1:195–227, 1988.

[38] D. Canning, L. A. N. Amaral, Y. Lee, M. Meyer, and . H. E. Stanley, Economics Letters **60**. 1998.

[39] D. Challet, M. Marsili, and Y.-C. Zhang. Minority games and stylized facts. *Physica A*, 299(1-2):228–233, October 2001.

[40] D. Challet and Y.-C. Zhang. On the minority game: Analytical and numerical studies. *Physica A*, 256(3-4):514–532, August 1997.

[41] D. Champernowne. *Economic Journal* ***63**, 318*, 1953.

[42] T. Chan. *Artificial markets and intelligent agents*. PhD thesis, Massachusetts Intitute of Technology, 2001.

[43] C. Chiarella. The dynamics of speculative behavior. *Ann Oper Res*, 37:101–123, 1992.

[44] C. Chiarella, W. Semmler, S. Mittnik, and P. Y. Zhu. Stock market, interest rate and output: a model and estimation for us time series data. *Stud Nonlinear Dyn E*, 6(1):art no. 2, April 2002.

[45] S. Cincotti, C. Dose, S. M. Focardi, M. Marchesi, and M. Raberto. Analysis and simulation of a double auction artificial financial market. EURO/INFORM 2003, 6-10 July 2003, Istanbul, Turkey.

[46] S. Cincotti, S. M. Focardi, M. Marchesi, and M. Raberto. Who wins? study of long-run trader survival in an artificial stock market. *Physica A*, 324(1-2):227–233, June 2003.

[47] S. Cincotti, S. M. Focardi, L. Ponta, M. Raberto, and E.Scalas. The waiting-time distribution of trading activity in a double auction artificial financial market. In A. Namatame, T. Kaizouji, and Y. Aruka, editors, *Economics and Heterogeneous Interacting Agents*. Springer-Verlag, Berlin, 2005. in press.

[48] S. Cincotti and S. Pastore. Heterogeneous information-based artificial stock market. WEHIA 2005, 13-15 June 2005, Essex, United Kingdom.

[49] S. Cincotti, L. Ponta, and M. Raberto. A multi-assets artificial stock market with zero-intelligence traders. WEHIA 2005, 13-15 June 2005, Essex, United Kingdom.

[50] R. H. Coase. *Economica* **4**, *386*, 1937.

[51] R. Cont and J.-P. Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroecon Dyn*, 4(2):170–196, June 2000.

[52] D. R. Cox and V. Isham. *Point processes*. Chapman and Hall, 1980.

[53] D. R. Cox and H. D. Miller. The theory of stochastic processes. (Chapman and Hall, London, 1968)

[54] D. M. Cutler, J. M. Poterba, and L. H. Summers. What moves stock prices? *J Portfolio Management*, pages 4–12, Aprl 1989.

[55] R. Cyert and J. March. A behavioral theory of the firm.

[56] M. M. Dacorogna, R. Gencay, U. Muller, R. B. Olsen, and O. V. Pictet. *An Introduction to High Frequency Finance*. Academic Press, 2001.

[57] S. J. Davis and . J. Haltiwanger, The Quarterly Journal of Economics **107**. 1992.

[58] R. H. Day and W. H. Huang. Bulls, bears and market sheep. *J Econ Behav Organ*, 14(3):299–329, December 1990.

[59] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Noise trader risk in financial markets. *J Polit Econ*, 98(4):703–738, August 1990.

[60] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. The survival of noise traders in financial markets. *J Bus*, 64(1):1–19, January 1991.

[61] D. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc*, 74:427–431, 1979.

[62] D. Dickey and W. A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057–1072, July 1981.

[63] T. Dunne, M. Roberts, and . L. Samuelson, The Quarterly Journal of Economics **104**. 1989.

[64] R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation. *Econometrica*, 50:987–1008, 1982.

[65] R. F. Engle and J. R. Russell. Forecasting transaction rates: the autoregressive conditional duration model. Working Paper No. 4966, National Bureau of Economic Research, 1994.

[66] J. M. Epstein. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60, 1999.

[67] J. M. Epstein and R. L. Axtell. *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, 1996.

[68] R. E. Ericson and . A. Pakes, Review of Economic Studies **62**. 1995.

[69] D. S. Evans. *Journal of Political Economics* **95**, *657*, 1987.

[70] G. D. Fabritiis, F. Pammolli, and . M. Riccaboni, Physica A **324**. 2003.

[71] G. D. Fabritiis, F. Pammolli, and . M.Riccaboni, Physica A **324**. 2003.

[72] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *J Financ*, 25:383–417, 1970.

[73] J. D. Farmer and S. Joshi. The price dynamics of common trading strategies. *J Econ Behav Organ*, 49(2):149–171, October 2002.

[74] J. Fedorowicz. *Journal of American Society for Information Science* **33**, *223*, 1982.

[75] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, 119:231255, 2004.

[76] M. Fowler, K. Beck, J. Brant, W. Opdyke, and D. Roberts. *Refactoring: Improving the Design of Existing Code*. Addison Wesley Longman, Reading, Massachusetts, 1999.

[77] D. Fu, S. V. Buldyrev, M. A. Salinger, and . H. E. Stanley, Physical Review E **74**. 2006.

[78] D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, and . H. E. Stanley, Proceedings of the National Academy of Sciences **102**. 2005.

[79] X. Gabaix. *The Quarterly Journal of Econmics* **114**, *739*, 1999.

[80] U. Garibaldi, M. A. Penco, and P. Viarengo. *An exact physical approach to market participation models*, volume 521 of *Lecture notes in economics and mathematical systems*, pages 91–103. Springer, February 2003.

[81] R. Gibrat. *Bull. Stat. Gén,* **19**, *469*, 1930.

[82] R. Gibrat. *Les Ingalits Economique*. Sirely, Paris, 1931.

[83] N. Gilbert and S. C. Bankes. Platforms and methods for agent-based modeling. *P Natl Acad Sci USA*, 99(Suppl. 3):7197–7198, May 2002.

[84] D. K. Gode and S. Sunder. Allocative efficiency of markets with zero intelligence traders. market as a partial substitute for individual rationality. *J Polit Econ*, 101(1):119–137, February 1993.

[85] A. Golan. A discrete stochastic model of economic production and a model of fluctuations in production – theory and empirical evidence.

[86] A. Golan. *Advances in Econometrics* **10**, *1*, 1994.

[87] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, and H. E. Stanley. Inverse cubic law for the distribution of stock price variations. *Eur Phys J B*, 3(2):139–140, May 1998.

[88] P. Gopikrishnan, V. Plerou, X. Gabaix, and H. E. Stanley. Statistical properties of share volume traded in financial markets. *Phys rev E*, 62:4493–4496, 2000.

[89] S. J. Grossman. On the efficiency of competitive stock markets when trades have diverse information. *J Financ*, 31:573–585, 1976.

[90] S. J. Grossman and J. E. Stiglitz. On the impossibility of informationally efficient markets. *Am Econ Rev*, 79(3):393–408, June 1980.

[91] . H. A. Makse et al., Physical Review E **58**. 1998.

[92] B. H. Hall. *Journal of Industrial Economics* **35**, *583*, 1987.

[93] P. Hart and . N. Oulton, The Economic Journal **106**. 2006.

[94] P. E. Hart and . S. J. Prais, Journal of the Royal Statistical Society **119**. 1956.

[95] B. R. Holmstrom and J. Tirole. Handbook of industrial organization.

[96] C. H. Hommes. Modeling the stylized facts in finance through simple nonlinear adaptive systems. *P Natl Acad Sci USA*, 99(Suppl. 3):7221–7228, May 2002.

[97] Z. F. Huang and S. Solomon. Finite market size as a source of extreme wealth inequality and market instability. *Physica A*, 294(3-4):503–513, 2001.

[98] Z. F. Huang and S. Solomon. Power, levy, exponential and gaussian-like regimes in auto catalytic financial systems. *Eur Phys J B*, 20(4):601–607, 2001.

[99] S. Hymer and . P. Pashigian, Journal of Political Economics **52**. 1962.

[100] Y. Ijiri and . H. A. Simon, Proceedings of the National Academy of Sciences **72**. 1975.

[101] Y. Ijiri and H. A. Simon. Skew distributions and the sizes of business firms.

[102] G. Iori. A microsimulation of trading activity in the stock market: the role of heterogeneity, agents' interactions and trade frictions. *J Econ Behav Organ*, 49(2):269–285, October 2002.

[103] G. Iori, M. G. Daniels, J. D. Farmer, L. Gillemot, S. Krishnamurthy, and E. Smith. An analysis of price impact function in order-driven markets. *Physica A*, 324(1-2):146–151, June 2003.

[104] P. Jefferies, M. L. Hart, P. M. Hui, and N. F. Johnson. From market games to real-world markets. *Eur Phys J B*, 20(4):493–501, April 2001.

[105] M. C. Jensen and . W. H. Meckling, Journal Financial Economics **3**. 1976.

[106] H. Jeong, B. Tomber, R. Albert, Z. N. Oltvai, and **407**. . A. L. Barabàsi, Nature. 2000.

[107] N. L. Johnson and S. Kotz. Urn models and their applications (wiley, new york, 1977).

[108] B. Jovanovic. *Econometrica* **50**, *649*, 1982.

[109] M. Kalecki. *Econometrica* ***13***, *161*, 1945.

[110] J. Kapteyn and M. J. Uven. Skew frequency curves in biology and statistics (hoitsema brothers, groningen, 1916).

[111] J. O. Kephart. Software agents and the route to the information economy. *P Natl Acad Sci USA*, 99(Suppl. 3):7207–7213, May 2002.

[112] A. Kirman. Ants, rationality and recruitment. *Q J Econ*, 108(1):137–156, February 1993.

[113] S. Kotz, T. J. Kozubowski, and K. Podgórski. The laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance (birkhauser, boston, 2001).

[114] S. Kotz, H. Mahmoud, S. P. Robert, and . Probability Letters **49**. 2000.

[115] R. Kumar, P. Raghavan, S. Rajagopalan, and . A. Tomkins, Computer Networks **31**. 1999.

[116] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Phys Rev Lett*, 83(7):1467–1470, August 1999.

[117] B. D. LeBaron. Time series properties of artificial stock market. *J Econ Dyn Control*, 23(9-10):1487–1516, June 1999.

[118] B. D. LeBaron. Agent-based computational finance: Suggested readings and early research. *J Econ Dyn Control*, 24(5-7):679–702, June 2000.

[119] B. D. LeBaron. Short-memory traders and their impact on group learning in financial markets. *P Natl Acad Sci USA*, 99(Suppl. 3):7201–7206, May 2002.

[120] Y. Lee, L. A. N. Amaral, D. Canning, M. Meyer, and . H. E. Stanley. Physical Review Letter **81**. 1998.

[121] M. Levy, H. Levy, and S. Solomon. *Microscopic Simulation of Financial Markets*. Academic Press, New York, 2000.

[122] M. Levy and S. Solomon. New evidence in the power-law distribution of wealth. *Physica A*, 242(1-2):90–94, August 1997.

[123] Y. H. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C. K. Peng, and H. E. Stanley. Statistical properties of the volatility of price fluctuations. *Phys Rev E*, 60(2):1390–1400, August 1999.

[124] A. W. Lo, A. C. MacKinlay, and J. Zhang. Econometric models of limit-order executions. *J Financ Econ*, 65(1):31–71, July 2002.

[125] R. Lucas. *The Bell Journal of Economics* **9**, *508*, 1978.

[126] R. E. Lucas. Expectations and the neutrality of money. *J Econ Theory*, 4:103–124, 1972.

[127] T. Lux. Time variation of second moments from a noise trader infection model. *J Econ Dyn Control*, 22(1):1–38, November 1997.

[128] T. Lux. The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. *J Econ Behav Organ*, 33(2):143–165, January 1999.

[129] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6718):498–500, February 1999.

[130] T. Lux and M. Marchesi. Volatility clustering in financial markets: a microsimulation of interacting agents. *Int J Theor Appl Finance*, 3(4):675–702, 2000.

[131] T. Lux and M. Marchesi. Special issue on heterogeneous interacting agents in financial markets. *J Econ Behav Organ*, 49(2):143–147, October 2002.

[132] F. Mainardi, M. Raberto, R. Gorenflo, and E. Scalas. Fractional calculus and continuous-time finance ii: the waiting-time distribution. *Physica A*, 287(3-4):468–481, December 2000.

[133] B. Mandelbrot. The variation of certain speculative prices. *J Business*, 36:394–419, 1963.

[134] B. Mandelbrot. *Fractals and Scaling in Finance.* Berlin: Springer, 1997.

[135] D. E. Mansfield. *American Economic Review* **52**, *1024*, 1962.

[136] R. N. Mantegna and H. E. Stanley. Scaling behavior in the dynamics of an economic index. *Nature*, 376(6535):46–49, July 1995.

[137] R. N. Mantegna and H. E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, Cambridge, UK, 1999.

[138] M. Marchesi, S. Cincotti, S. M. Focardi, and M. Raberto. *The Genoa artificial stock market: microstructure and simulation*, volume 521 of *Lecture Notes in Economics and Mathematical Systems*, pages 277–289. Springer, February 2003.

[139] S. Maslov. Simple model of a limit order-driven market. *Physica A*, 278(3-4):571–578, April 2000.

[140] S. Maslov and M. Mills. Price fluctuations from the order book perspective: empirical facts and a simple model. *Physica A*, 299(1-2):234–246, October 2001.

[141] L. Matassini. The trading rectangle strategy within book models. *Physica A*, 301(1-4):449–456, December 2001.

[142] K. Matia, D. Fu, S. V. Buldyrev, F. Pammolli, M. Riccaboni, and . H. E. Stanley, Europhysics Letter **67**. 2004.

[143] R. C. Merton. An intertemporal capital asset pricing model. *Econometrica*, pages 867–888, 1973.

[144] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J Chem Phys*, 21:1087–1092, 1953.

[145] P. Milgrom and J. Roberts. Economics, organization, and management.

[146] E. W. Montroll and M. F. Schlesinger. On 1/f noise and other distributions with long tails. *P Natl Acad Sci USA*, 79:3380–3383, 1982.

[147] U. A. Mueller, M. Dacorogna, R. B. Olsen, O. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates. *Empirical Journal of Banking and Finance*, 14, 1990.

[148] J. F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29:315–335, 1961.

[149] R. R. Nelson and S. G. Winter. An evolutionary theory of technical change.

[150] M. E. J. Newman. preprint condmat/0412004.

[151] M. O'Hara. *Market Microstructure Theory*. Blackwell, Oxford, UK, 1995.

[152] A. Pagan. The econometrics of financial markets. *J Empirical Finance*, 3(1):15–102, May 1996.

[153] A. Pakes and . P. McGuire, Rand Journal of Economics **25**. 1994.

[154] R. G. Palmer, W. B. Arthur, J. H. Holland, B. D. LeBaron, and P. Tayler. Artificial economic life: a simple model of a stock market. *Physica D*, 75:264–274, 1994.

[155] V. Pareto. *Course d'Economie Politique*. Lausanne, CH, 1897.

[156] V. Plerou, L. A. N. Amaral, P. Gopikrishnan, M. Meyer, and . H. E. Stanley, Nature (London) **400**. 1999.

[157] V. Plerou, P. Gopikrishnan, X. Gabaix, and H. E. Stanley. Quantifying stock price response to demand fluctuations. *Phys Rev E*, 66(2), August 2002.

[158] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley. Random matrix approach to cross correlations in financial data. *Phys Rev E*, 65(6):066126–1:18, June 2002.

[159] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Phys Rev Lett*, 83(7):1471–1474, August 1999.

[160] M. Raberto. *Modelling and implementation of an artificial financial market using object oriented technology: the Genoa artificial stock market*. PhD thesis, University of Genoa, Italy, 2003.

[161] M. Raberto and S. Cincotti. Modeling and simulation of a double auction artificial financial market. *Physica A*, 355(1):34–45, September 2005.

[162] M. Raberto, S. Cincotti, S. Focardi, and M. Marchesi. Traders' long-run wealth in an artificial financial market. *Computational Economics*, 22(2-3):255–272, October-December 2003.

[163] M. Raberto, S. Cincotti, S. M. Focardi, and M. Marchesi. Agent-based simulation of a financial market. *Physica A*, 219:319–327, October 2001.

[164] M. Raberto, E. Scalas, and F. Mainardi. Waiting-times and returns in high-frequency financial data: an empirical study. *Physica A*, 314(1-4):749–755, November 2002.

[165] R. Radner. *Econometrica* **61***, 1109*, 1993.

[166] W. J. Reed. The Pareto, Zipf and other power laws. *Econ Lett*, 74(1):15–19, December 2001.

[167] W. J. Reed and . B. D. Hughes, Mathematical Biosciences **189**. 2004.

[168] W. J. Reed and . B. D. Hughes, Physical Review E **66**. 2002.

[169] M. Riccaboni and . F. Pammolli, Research Policy **31**. 2002.

[170] S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, pages 343–362, December 1976.

[171] P. A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2):41–49, 1965.

[172] T. Sargent. *Bounded rationality in macroeconomics.* The Arne Ryde Memorial Lectures. Claredon Press, Oxford, UK, 1993.

[173] E. Scalas, R. Gorenflo, H. Lucklock, F. Mainardi, M. Mantelli, and M. Raberto. Anomalous waiting times in high-frequency financial data. *Quantitative Finance*, 4:1–8, December 2004.

[174] E. Scalas, R. Gorenflo, and F. Mainardi. Fractional calculus and continuous-time finance. *Physica A*, 284:376–384, 2000.

[175] E. Scalas, R. Gorenflo, F. Mainardi, M. Mantelli, and M. Raberto. Anomalous waiting times in high-frequency financial data. submitted to Quantitative Finance, `http://xxx.lanl.gov/abs/cond-mat/0310305`, 2003.

[176] F. M. Scherer and D. R. Ross. Industrial market structure and economic performance.

[177] D. Shapiro, R. D. Bollman, and . P. Ehrensaft, American Journal of Agricultural Economics **69**. 1987.

[178] W. F. Sharpe. Capital asset prices: a theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, September 1964.

[179] R. Shiller. *Market volatility.* MIT Press, Cambridge, MA, 1989.

[180] A. Shleifer and L. H. Summers. The noise trader approach to finance. *J Economic Perspectives*, 4:19–33, Spring 1990.

[181] H. A. Simon. *Biometrika* **42***, 425*, 1955.

[182] H. A. Simon. *The science of the Artificial.* MIT Press, 3 edition, 1996.

[183] H. A. Simon. *Models of Bounded Rationality: Rmpirically Grounded Economic Reason*, volume 3. MIT Press, 1997.

[184] H. A. Simon and . C. P. Bonini, American Economical Review **48**. 1958.

[185] A. Singh and . G. Whittington, Review Economical Studies **42**. 1975.

[186] H. E. Stanley. Introduction to phase transitions and critical phenomena.

[187] M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, and . H. E. Stanley, Nature (London) **379**. 1996.

[188] M. H. R. Stanley, S. V. Buldyrev, R. Mantegna, S. Havlin, M. A. Salinger, and H. E. Stanley. *Economics Letters* **49**, *453*, 1995.

[189] K. Steiglitz and D. Shapiro. Simulating the madness of crowds: price bubbles in an auction -mediated robot market. *Computational Economics*, 12(1):35–59, August 1998.

[190] J. Steindl. Random processes and the growth of firms: A study of the pareto law.

[191] J. H. Stock and M. W. Watson. Tesing for common trends. *J Am Stat Assoc*, 83(404):1097–1107, December 1988.

[192] G. Succi and M. Marchesi. *Extreme Programming Examined*. Addison-Wesley, Reading, Massachusetts, 2001.

[193] L. H. Summers. Does the stock market rationally reflect fundamentals. *J Financ*, 41:591–601, 1986.

[194] J. Sutton. *Journal of Economic Literature* **35**, *40*, 1997.

[195] J. Sutton. *Physica A* **312**, *577*, 2002.

[196] H. Takayasu and . K. Okuyama, Fractals **6**. 1998.

[197] L. Tesfatsion. Agent-based modeling of evolutionary economic systems. *IEEE T Evolut Comput*, 5(5):437–441, October 2001.

[198] L. Tesfatsion. Introduction. *Computational Economics*, 18(!):1–8, August 2001.

[199] L. Tesfatsion. Introduction to the special issue on agent-based computational economics. *J Econ Dyn Control*, 25(3-4):281–293, 2001.

[200] L. Tesfatsion. Economic agents and markets as emergent phenomena. *P Natl Acad Sci USA*, 99(Suppl. 3):7191–7192, May 2002.

[201] M. Tumminello, T. Aste, T. D. Matteo, and R. N. Mantegna. A tool for filterininformation in complex systems. *PNAS*, 102(30):10421–10426, July 2005.

[202] H. R. Varian. Microeconomics analysis. 1978.

[203] H. A. Wan. *On the agent market model of stock markets*. PhD thesis, School of Computing and Information Systems, Sunderland University, UK, 1997.

[204] H. A. Wan, A. Hunter, and P. Dunne. Autonomous agent models and stock markets. *Artif Intell Rev*, 17(2):87–128, April 2002.

[205] O. E. Williamson. Handbook of industrial organization.

[206] M. Wyart and . J. P. Bouchaud, Physica A **326**. 2003.

[207] K. Yamasaki, K. Matia, S. V. Buldyrev, D. Fu, F. Pammolli, M. Riccaboni, and R. H. E. Stanley, Physical Review E **74**. 2006.

[208] M. Youssefmir and B. A. Huberman. Clustered volatility in multiagent dynamics. *J Econ Behav Orgam*, 32(1):101–118, January 1997.

[209] M. Youssefmir, B. A. Huberman, and T. Hogg. Bubbles and market crashes. *Computational Economics*, 12(2):97–114, October 1998.

[210] G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Hafner Publishing Company, 1949.